

STATE OF THE ART IN NEURAL NETWORKS AND THEIR APPLICATIONS

VOLUME 2

Edited by
AYMAN S. EL-BAZ
JASJIT S. SURI



State of the Art in Neural Networks and Their Applications

This page intentionally left blank

State of the Art in Neural Networks and Their Applications

Volume 2

Edited by

Ayman S. El-Baz

*University of Louisville, Louisville, KY, United States;
University of Louisville at Alamein International University (UofL-AIU)
New Alamein City, Egypt*

Jasjit S. Suri

*Stroke Monitoring and Diagnosis Division, ATHEROPOINT,
Roseville, CA, United States*



ACADEMIC PRESS

An imprint of Elsevier

Academic Press is an imprint of Elsevier
125 London Wall, London EC2Y 5AS, United Kingdom
525 B Street, Suite 1650, San Diego, CA 92101, United States
50 Hampshire Street, 5th Floor, Cambridge, MA 02139, United States
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, United Kingdom

Copyright © 2023 Elsevier Inc. All rights reserved.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Details on how to seek permission, further information about the Publisher's permissions policies and our arrangements with organizations such as the Copyright Clearance Center and the Copyright Licensing Agency, can be found at our website: www.elsevier.com/permissions.

This book and the individual contributions contained in it are protected under copyright by the Publisher (other than as may be noted herein).

Notices

Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.

Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

ISBN: 978-0-12-819872-8

For Information on all Academic Press publications
visit our website at <https://www.elsevier.com/books-and-journals>

Publisher: Mara E. Conner
Acquisitions Editor: Chris Katsaropoulos
Editorial Project Manager: Emily Thomson
Production Project Manager: Surya Narayanan Jayachandran
Cover Designer: Miles Hitchen

Typeset by MPS Limited, Chennai, India



Dedication

*With love and affection to my mother and father, whose loving
spirit sustains me still*

–Ayman S. El-Baz

To my late loving parents, immediate family, and children

–Jasjit S. Suri

This page intentionally left blank

Contents

List of contributors	xv
About the editors	xxi
Acknowledgments	xxiii

CHAPTER 1 Microscopy Cancer Cell Imaging in B-lineage Acute Lymphoblastic Leukemia

Anubha Gupta, Shiv Gehlot and Ritu Gupta

1.1 Introduction	1
1.2 Building a computer-assisted solution	2
1.3 Data preparation	3
1.3.1 Preparation of slide for microscopic imaging	3
1.3.2 Capture of microscopic images from healthy and cancer subjects for B-acute lymphoblastic leukemia cancer	4
1.4 Normalization of color stain to correct for abnormalities during the staining process	5
1.4.1 Quantitative results	8
1.5 Segmentation of cells of interest (in B-lineage ALL cancer).....	8
1.5.1 Method-1 of cell segmentation using traditional image processing techniques	11
1.5.2 Method-2 of cell segmentation using deep belief network.....	12
1.5.3 Method-3 of cell segmentation using novel convolutional neural network architecture	12
1.6 Classification of cancer and healthy cells	19
1.6.1 C-NMC 2019 challenge dataset.....	20
1.6.2 Classification on C-NMC 2019 dataset.....	20
1.6.3 SDCT-AuxNet ^θ CNN architecture for C-NMC 2019 dataset	21
1.7 Conclusions	24
References.....	24

CHAPTER 2 Computational imaging applications in brain and breast cancer

Aimilia Gastounioti, Saima Rathore, Omid Haji Maghsoudi, Emily F. Conant, Despina Kontos and Spyridon Bakas

2.1 Introduction	29
2.2 Building upon current clinical standards.....	30
2.2.1 Clinical standards.....	30

2.2.2	Tissue segmentation.....	30
2.3	Deep learning applications in brain cancer	31
2.3.1	Tumor grading.....	31
2.3.2	Survival analysis	32
2.3.3	Radiogenomics.....	33
2.3.4	Pseudoprogession.....	35
2.4	Deep learning applications in breast cancer	36
2.4.1	Increasing accuracy in breast cancer risk assessment	36
2.4.2	Reproducible breast density assessment for improved breast cancer risk prediction	37
2.4.3	Improving performance in breast cancer diagnosis	38
2.4.4	Enhancing efficacy in breast cancer clinical practice	39
2.5	Conclusion	39
	Acknowledgments	40
	References.....	40

CHAPTER 3 Deep neural networks and advanced computer vision algorithms in the early diagnosis of skin diseases

*Joanna Jaworek-Korjakowska, Moi Hoon Yap,
Debotosh Bhattacharjee, Pawel Kleczek,
Andrzej Brodzicki and Marek Gorgon*

3.1	Introduction and motivation for the early diagnosis of melanoma	47
3.2	Artificial intelligence and computer vision in melanoma diagnosis	50
3.3	Medical diagnostic procedures for screening of skin diseases	53
3.4	State-of-the-art survey on skin mole segmentation methods.....	55
3.4.1	Comparison of the state of the art.....	59
3.4.2	Summary	60
3.5	Improved local and global patterns detection algorithms by deep learning algorithms.....	60
3.6	Early classification of skin melanomas in dermoscopy	64
3.6.1	Diagnostic algorithms	64
3.6.2	Approaches to detect the diagnostic criteria	65
3.6.3	Approaches to directly classify skin conditions.....	66
3.7	Conclusions	69

3.8	How to speed up the classification process with field-programmable gate arrays?	70
3.9	Challenges and future directions.....	72
3.10	Teledermatology.....	73
	References.....	74
CHAPTER 4	An accurate deep learning-based computer-aided diagnosis system for early diagnosis of prostate cancer	83
	<i>Islam R. Abdelmaksoud, Ahmed Shalaby, Mohammed Ghazal, Mohammed Elmogy, Ahmed AbouElfetouh, Ali Mahmoud and Ayman S. El-Baz</i>	
4.1	Introduction	83
4.2	Methods	85
	4.2.1 Feature Extraction	86
	4.2.2 CNN-based classification.....	88
4.3	Experimental results	89
4.4	Conclusion	91
	References.....	91
CHAPTER 5	Adaptive graph convolutional neural network and its biomedical applications.....	105
	<i>Junzhou Huang and Ruoyu Li</i>	
5.1	Introduction	105
5.2	Related work	109
	5.2.1 Evolution of graph convolutional neural networks.....	109
	5.2.2 Neural network on molecular graph.....	111
	5.2.3 Attention on graph	112
	5.2.4 Neural network for survival analysis.....	113
5.3	Method.....	114
	5.3.1 Spectral graph convolution-LL layer.....	114
	5.3.2 Adaptive graph convolution network architecture	117
	5.3.3 Graph attention network on adaptive graph.....	119
	5.3.4 DeepGraphSurv framework	120
5.4	Experiment	122
	5.4.1 Drug-property prediction	122
	5.4.2 DeepGraphSurv and survival prediction	124
5.5	Conclusion.....	128
	References.....	129
	Further reading	132

CHAPTER 6	Deep slice interpolation via marginal super-resolution, fusion, and refinement	133
	<i>Cheng Peng, Wei-An Lin, Haofu Liao, Rama Chellappa and Shaohua Kevin Zhou</i>	
6.1	Introduction	133
6.2	Related work	135
	6.2.1 Traditional slice interpolation methods	135
	6.2.2 Learning-based super-resolution methods	135
6.3	Problem formulation and baseline convolutional neural networks approaches	136
6.4	The proposed algorithm	136
	6.4.1 Marginal super-resolution	137
	6.4.2 Two-view fusion and refinement.....	138
	6.4.3 Comparison with baseline convolutional neural networks approaches	139
6.5	Experiments	139
	6.5.1 Implementation details.....	139
	6.5.2 Dataset.....	139
	6.5.3 Evaluation metrics	140
	6.5.4 Visual comparisons	140
	6.5.5 Ablation study	141
6.6	Conclusion	143
	References.....	144
CHAPTER 7	Explainable deep learning approach to predict chemotherapy effect on breast tumor’s MRI	147
	<i>Mohammed El Adoui, Mohammed Amine Larhman, Stylianos Drisis and Mohammed Benjelloun</i>	
7.1	Introduction	147
7.2	Materials and developed methods.....	148
	7.2.1 Study population	148
	7.2.2 Magnetic resonance imaging protocol.....	148
	7.2.3 Image preprocessing	149
	7.2.4 Convolution neural network architecture development	149
7.3	Results	151
	7.3.1 Quantitative results	151
	7.3.2 Qualitative results	152
7.4	Discussion.....	153
7.5	Conclusion	153
	Aknowledgments	154
	References.....	154

CHAPTER 8	Deep learning interpretability: measuring the relevance of clinical concepts in convolutional neural networks features	157
	<i>Mara Graziani, Vincent Andrearczyk and Henning Müller</i>	
8.1	Introduction	157
8.2	Related work on interpretable artificial intelligence	159
	8.2.1 Motivations.....	159
	8.2.2 Related terminology.....	160
	8.2.3 Related work on explainable artificial intelligence	161
	8.2.4 Evaluation of explainable artificial intelligence methods	166
8.3	Methods	168
	8.3.1 Retinopathy of prematurity.....	168
	8.3.2 Concept attribution with regression concept vectors.....	170
8.4	Experiments and results	176
	8.4.1 Network performance on the retinopathy of prematurity task	176
	8.4.2 Results of concept attribution.....	177
8.5	Discussion of the results	186
8.6	Conclusions	187
	Acknowledgments	188
	References.....	188
CHAPTER 9	Computational lung sound classification: a review.....	193
	<i>Truc Nguyen and Franz Pernkopf</i>	
9.1	Introduction	193
9.2	Data processing	195
	9.2.1 Audio signal preprocessing.....	195
	9.2.2 Feature extraction.....	197
	9.2.3 Data augmentation	201
9.3	Data modeling	203
	9.3.1 Machine learning	203
	9.3.2 Learning paradigm	205
9.4	Recent public lung sound datasets.....	207
	9.4.1 ICBHI 2017 dataset	207
	9.4.2 The Abdullah University Hospital 2020 dataset.....	208
	9.4.3 HF_Lung_V1 dataset	208
9.5	Conclusion.....	208
	References.....	209

CHAPTER 10	Clinical applications of machine learning in heart failure	217
	<i>Xinmu Li, Sharen Lee, George Bazoukis, Gary Tse and Tong Liu</i>	
10.1	Introduction	217
10.2	Diagnosis	218
	10.2.1 Automatic diagnosis, classification, and phenotyping of heart failure	218
	10.2.2 Detection of heart failure-associated arrhythmia	220
10.3	Management	221
	10.3.1 Prognostic prediction	221
	10.3.2 Development of therapy	224
	10.3.3 Optimal patient selection for specific therapies or recommendation of optimal therapy	224
10.4	Prevention	225
10.5	Conclusion	226
	References	226
CHAPTER 11	Role of artificial intelligence and radiomics in diagnosing renal tumors: a survey	235
	<i>Mohamed Shehata, Ahmed Elmahdy, Ahmed Alksas, Rasha Abouelkheir, Ali Mahmoud, Mohamed Abou El-Ghar, Mohammed Ghazal and Ayman S. El-Baz</i>	
11.1	Introduction	235
11.2	Basic background	236
	11.2.1 Deep learning	236
	11.2.2 Machine learning	236
	11.2.3 Radiomics	237
11.3	Steps of artificial intelligence-based diagnostic systems	237
	11.3.1 Image acquisition	237
	11.3.2 Image segmentation	237
	11.3.3 Feature extraction and qualifications	238
	11.3.4 Diagnostic analysis	238
11.4	Texture analysis	238
	11.4.1 Principles	238
	11.4.2 Statistical techniques	239
	11.4.3 Model-based methods	239
	11.4.4 Transform methods	239
	11.4.5 Texture parameters	240
11.5	Clinical applications of artificial intelligence and radiomics	242
	11.5.1 Benign versus malignant renal tumors	242

11.5.2 Renal cell carcinoma versus angiomyolipoma..... 242
 11.5.3 Renal cell carcinoma versus oncocytoma 243
 11.5.4 Renal cell carcinoma versus renal cyst 243
 11.5.5 Subtyping of renal cell carcinoma..... 243
 11.5.6 Grading of renal cell carcinoma 244
 11.5.7 Staging of renal cell carcinoma 244
 11.5.8 Characterization of small renal mass 244
11.6 Merits and limitations245
 11.6.1 Merits 245
 11.6.2 Limitations 245
11.7 Future directions.....245
11.8 Conclusion246
 References..... 246

CHAPTER 12 A review of texture-centric diagnostic models for thyroid cancer using convolutional neural networks and visualized texture patterns..... 265

Ahmed Naglah, Fahmi Khalifa, Reem Khaled, Ahmed Abdel Khalek Abdel Razeq, Mohammed Ghazal, Guruprasad Giridharan, Ali Mahmoud and Ayman S. El-Baz

12.1 Introduction265
12.2 Materials and collection protocols.....268
 12.2.1 Study participants and raw data collection 268
 12.2.2 Nodule segmentation and apparent diffusion coefficient calculations 268
12.3 Statistical analysis270
12.4 2D texture model.....270
12.5 3D texture model.....272
12.6 Texture analysis.....272
12.7 Results273
 12.7.1 Statistical results 273
 12.7.2 Diagnostic accuracy of 2D model 273
 12.7.3 Diagnostic accuracy of 3D model 276
 12.7.4 Texture pattern visualization 278
12.8 Discussion.....279
12.9 Conclusion280
 References..... 281

Index297

This page intentionally left blank

List of contributors

Islam R. Abdelmaksoud

University of Louisville, Louisville, KY, United States; Faculty of Computers and Information, Mansoura University, Mansoura, Egypt

Ahmed AbouElfetouh

Faculty of Computers and Information, Mansoura University, Mansoura, Egypt

Rasha Abouelkheir

Radiology Department, Urology and Nephrology Center, University of Mansoura, Mansoura, Egypt

Ahmed Alksas

University of Louisville, Louisville, KY, United States

Vincent Andrearczyk

Institute of Information Systems, University of Applied Sciences Western Switzerland (HES-SO), Sierre, Switzerland

Spyridon Bakas

Center for Biomedical Image Computing and Analytics, University of Pennsylvania, Philadelphia, PA, United States; Department of Radiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States; Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States

George Bazoukis

University of Nicosia, Medical School, Nicosia, Cyprus

Mohammed Benjelloun

Professor, Faculty Of Engineering, ILIA Department, University of Mons, Belgium

Debotosh Bhattacharjee

Department of Computer Science and Engineering, Jadavpur University, Kolkata, West Bengal, India

Andrzej Brodzicki

Department of Automatic Control and Robotics, AGH University of Science and Technology, Krakow, Lesser Poland, Poland

Rama Chellappa

Electrical and Computer Engineering (ECE), University of Maryland, College Park, MD, United States

Emily F. Conant

Department of Radiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States

Stylianos Drisis

Radiologist - Hospital Marie Curie, Charleroi, Belgium

Mohammed El Adoui

Postdoctoral researcher, ILIA Department, Faculty of Engineering, University of Mons, Belgium

Ayman S. El-Baz

University of Louisville, Louisville, KY, United States; University of Louisville at Alamein International University (UofL-AIU), New Alamein City, Egypt

Mohamed Abou El-Ghar

Radiology Department, Urology and Nephrology Center, University of Mansoura, Mansoura, Egypt

Ahmed Elmahdy

Radiology Department, Urology and Nephrology Center, University of Mansoura, Mansoura, Egypt

Mohammed Elmogy

Faculty of Computers and Information, Mansoura University, Mansoura, Egypt

Aimilia Gastouniotti

Center for Biomedical Image Computing and Analytics, University of Pennsylvania, Philadelphia, PA, United States; Department of Radiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States

Shiv Gehlot

Department of ECE, Indraprastha Institute of Information Technology-Delhi (IIIT-D), IIIT Delhi, Delhi, India

Mohammed Ghazal

Electrical, Computer, and Biomedical Engineering Department, Abu Dhabi University, Abu Dhabi, United Arab Emirates

Guruprasad Giridharan

University of Louisville, Louisville, KY, United States

Marek Gorgon

Department of Automatic Control and Robotics, AGH University of Science and Technology, Krakow, Lesser Poland, Poland

Mara Graziani

Institute of Information Systems, University of Applied Sciences Western Switzerland (HES-SO), Sierre, Switzerland

Anubha Gupta

Department of ECE, Indraprastha Institute of Information Technology-Delhi (IIIT-D), IIIT Delhi, Delhi, India

Ritu Gupta

All India Institute of Medical Sciences (AIIMS), New Delhi, India

Junzhou Huang

The University of Texas at Arlington, Arlington, TX, United States

Joanna Jaworek-Korjakowska

Department of Automatic Control and Robotics, AGH University of Science and Technology, Krakow, Lesser Poland, Poland

Reem Khaled

Department of Diagnostic Radiology, Faculty of Medicine, Mansoura University, Mansoura, Egypt

Fahmi Khalifa

University of Louisville, Louisville, KY, United States

Pawel Kleczek

Department of Automatic Control and Robotics, AGH University of Science and Technology, Krakow, Lesser Poland, Poland

Despina Kontos

Center for Biomedical Image Computing and Analytics, University of Pennsylvania, Philadelphia, PA, United States; Department of Radiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States

Mohammed Amine Larhman

VP of AI, Artificial Intelligence (AI) Department, Sol One, Bruges, Belgium

Sharen Lee

Cardiovascular Analytics Group, Hong Kong, China-UK Collaboration

Ruoyu Li

The University of Texas at Arlington, Arlington, TX, United States

Xinmu Li

Tianjin Key Laboratory of Ionic-Molecular Function of Cardiovascular Disease, Department of Cardiology, Tianjin Institute of Cardiology, Second Hospital of Tianjin Medical University, Tianjin, P.R. China

Haofu Liao

Computer Science, University of Rochester, Rochester, NY, United States

Wei-An Lin

Electrical and Computer Engineering (ECE), University of Maryland, College Park, MD, United States

Tong Liu

Tianjin Key Laboratory of Ionic-Molecular Function of Cardiovascular Disease, Department of Cardiology, Tianjin Institute of Cardiology, Second Hospital of Tianjin Medical University, Tianjin, P.R. China

Omid Haji Maghsoudi

Center for Biomedical Image Computing and Analytics, University of Pennsylvania, Philadelphia, PA, United States; Department of Radiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States

Ali Mahmoud

University of Louisville, Louisville, KY, United States

Henning Müller

Institute of Information Systems, University of Applied Sciences Western Switzerland (HES-SO), Sierre, Switzerland

Ahmed Naglah

University of Louisville, Louisville, KY, United States

Truc Nguyen

Signal Processing and Speech Communication Lab., Graz University of Technology, Graz, Austria

Cheng Peng

Electrical and Computer Engineering (ECE), University of Maryland, College Park, MD, United States

Franz Pernkopf

Signal Processing and Speech Communication Lab., Graz University of Technology, Graz, Austria

Saima Rathore

Center for Biomedical Image Computing and Analytics, University of Pennsylvania, Philadelphia, PA, United States; Department of Radiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States

Ahmed Abdel Khalek Abdel Razek

Department of Diagnostic Radiology, Faculty of Medicine, Mansoura University, Mansoura, Egypt

Ahmed Shalaby

University of Louisville, Louisville, KY, United States

Mohamed Shehata

University of Louisville, Louisville, KY, United States

Gary Tse

Tianjin Key Laboratory of Ionic-Molecular Function of Cardiovascular Disease, Department of Cardiology, Tianjin Institute of Cardiology, Second Hospital of Tianjin Medical University, Tianjin, P.R. China; Kent and Medway Medical School, Canterbury, United Kingdom; School of Nursing and Health Studies, Hong Kong Metropolitan University, Hong Kong, P.R. China

Moi Hoon Yap

Visual Computing Lab, Manchester Metropolitan University, Manchester, North West, United Kingdom

Shaohua Kevin Zhou

Institute of Computing Technology, Chinese Academy of Science, Beijing, China

This page intentionally left blank

About the editors

Ayman S. El-Baz is a distinguished professor at the University of Louisville, Kentucky, United States and the University of Louisville at Alamein International University (UofL-AIU), New Alamein City, Egypt. Dr. El-Baz earned his BSc and MSc degrees in electrical engineering in 1997 and 2001, respectively. He earned his PhD in electrical engineering from the University of Louisville in 2006. Dr. El-Baz was named as a Fellow for Coulter, AIMBE, and NAI for his contributions to the field of biomedical translational research. Dr. El-Baz has almost two decades of hands-on experience in the fields of bio-imaging modeling and non-invasive computer-assisted diagnosis systems. He has authored or coauthored more than 700 technical articles (182 journals, 46 books, 97 book chapters, 253 refereed-conference papers, 214 abstracts, and 38 US patents and Disclosures).



Jasjit S. Suri, Ph.D., MBA, is an innovator, visionary, scientist, and an internationally known world leader in Biomedical Engineering and its Management. Dr. Suri received the Director General's Gold medal in 1980 and is a Fellow of (i) Institute of Electrical and Electronic Engineers, (ii) American Institute of Medical and Biological Engineering, (iii) American Society of Ultrasound in Medicine, (iv) Society of Vascular Medicine, and (v) Asia Pacific Vascular Society. He is also the recipient of Lifetime Achievement Award from Marquis. He is currently Chairman of AtheroPoint, Roseville, CA, USA, dedicated to imaging technologies for cardiovascular and stroke. He has won numerous awards, and has ~24,000 citations, co-authored 50 books, over 50 patents, and has an H-index of 75.



This page intentionally left blank

Acknowledgments

The completion of this book would not have been possible without the participation and assistance of so many people whose names may not all be enumerated. Their contributions are sincerely appreciated and gratefully acknowledged. However, the editors would like to express their deep appreciation and indebtedness particularly to Dr. Ali H. Mahmoud and Mohamed Shehata for their endless support.

This page intentionally left blank

Microscopy Cancer Cell Imaging in B-lineage Acute Lymphoblastic Leukemia

1

Anubha Gupta¹, Shiv Gehlot¹ and Ritu Gupta²

¹*Department of ECE, Indraprastha Institute of Information Technology-Delhi (IIIT-D), IIIT
Delhi, Delhi, India*

²*All India Institute of Medical Sciences (AIIMS), New Delhi, India*

1.1 Introduction

Acute lymphoblastic leukemia (ALL) is a type of white blood cancer, in which the B- and T-lymphocytes are affected. This cancer constitutes approximately 20% of pediatric malignancies [1]. At diagnosis, patients with acute leukemia may have a total of roughly 10^{12} malignant cells. The disease is considered to be in complete remission (patient is not showing any symptom of the disease) when fewer than 5% of the cells in bone marrow samples are morphologically identifiable blasts. However, these patients may still have as many as 10^{10} malignant cells. From that point until an overt clinical relapse, the level of leukemic cells in the body is mostly unknown, resulting in clinical management strategies that do not discriminate among patients by their residual disease levels. Thus, patients with 10^{10} leukemic cells are treated on the same regimen as those with much lower levels or, perhaps, with no leukemia.

Morphologically, the healthy progenitor cells and cancer blood cells present at low numbers appear similar under the microscope to the naked eye. Hence, if a patient's bone marrow is tested via microscopic examination, leukemia would be diagnosed in the progressed state when the number of white blood cells is observed to be exceptionally high in numbers. Thus, the disease is diagnosed not because the pathologist can identify the cancer cell but because of the medical knowledge that such a high number of a particular blood cell cannot be spotted in the microscopic slide of a healthy subject. This implies that, whether accidentally or otherwise, leukemia would be detected only in the advanced stages during routine testing. However, it is essential to make early disease diagnosis for better cure and improve the overall survival of the subjects suffering from cancer.

Similarly, for patients in clinical remission during cancer therapy, the number of cancer cells is generally below the conventional methods' detection limit. If left undetected and therefore untreated, this leads to a frank relapse of the disease. The advanced medical tests using flow cytometry are not utilized under routine

check-ups. They would be attempted only when the subject is undergoing cancer treatment, which may lead to a delay in the diagnosis. Moreover, they are costly and are not available widely in pathology laboratories or hospitals, particularly in rural areas. The costs involved in terms of infrastructure, reagents, highly skilled human resources, and time required preclude their widespread use in routine pathology testing.

1.2 Building a computer-assisted solution

Because the microscopic examination is readily available and cost-effective, conferring the ability to distinguish cancer cells from healthy cells to microscopic image processing evaluation will provide several benefits. First, the test can be included as part of routine clinical tests whenever a blood sample is collected. The test will become readily available to doctors wherever computer and microscope facilities exist. Second, the proposed method will eliminate the need for sophisticated high-end costly machines (e.g., flow cytometer), the requirement of expensive reagents and chemicals, and trained human resources to run those tests. In particular, such a device can serve as a boon for a rural society where hospitals and pathology labs generally run with a shortage of resources, including a skilled workforce. Thus, it is worthwhile to build computer-assisted diagnostic tools for blood disorders such as leukemia. To arrive at a conclusive decision on disease diagnosis and degree of progression, it is crucial to identify malignant cells and count the number of malignant vs healthy cells. Computer-assisted tools can be beneficial in automating the entire process of cell identification and counting. This will also be useful for objective evaluation of residual disease in leukemia wherein a large number of cells need to be analyzed in an objective manner for reliable diagnostic results. We started with the aim to build an image processing-based robust classification tool that minimizes the probability of miss and false alarm of disease detection.

The complete workflow of such a tool consists of the following steps:

1. Capture of images and preparation of the dataset,
2. Normalization of color stain to correct for abnormalities during the staining process,
3. Segmentation of cells of interest, and
4. Identification of cells as cancer or healthy cells.

Each of these four stages has its challenges that need to be addressed to build a final deliverable tool that can be deployed at a hospital to diagnose and monitor leukemia. In this chapter, we discuss the attempt to build an automated tool for B-acute lymphoblastic leukemia cancer and the generic steps and challenges encountered in the development of such tool.

1.3 Data preparation

1.3.1 Preparation of slide for microscopic imaging

Slide preparation is a sophisticated process that involves multiple stages. The following six steps are used for the slide preparation for any tissue:

- 1. Fixation:** Tissue fixation aims to preserve the sample in its natural state through the prevention of autolysis and putrefaction. Fixation is vital to avoid the introduction of artifacts in the samples that may affect the further analysis. Typically, chemical fixatives work by stabilizing the nucleic acids and proteins of the tissue. Some examples of the fixatives are formalin, potassium dichromate, and picric acid for solid tissues and alcohol based fixatives for blood and/or bone marrow smears.
- 2. Processing:** Tissue processing is used to replace water with a solidifying medium. This dehydration (water removal) is necessary to provide rigidity to the sample, enabling the thin section's slicing. At the same time, solidification must not be too severe to damage the tissue. Ethanol, acetone, methanol are some commonly used dehydrating agents.
- 3. Embedding:** Embedding is done to provide external support for the sectioning. In this process, the sample (tissue) is transferred to a mold containing a medium like wax or gelatine, which upon solidification, provides blocks used in sectioning.
- 4. Sectioning:** In sectioning, thin slices are obtained from an embedded sample using an instrument called a microtome. The thickness of the slices depends on the microscopy to be used for analysis. In light microscopy, 10 μm slices are obtained using a microtome mounted with a steel knife, whereas, for transmission electron microscopy, 50 nm slices are cut with an ultra-microtome having a diamond knife.
- 5. Staining:** Staining is used to highlight the different features of the tissues, which otherwise show unnoticeable variations. Some examples of the histology stains are Haematoxylin and Eosin (H&E), Giemsa Stain, Bielschowsky Stain, Mallory Trichrome, etc. H&E is the frequently used dye and contains H&E staining chemicals. Hematoxylin stains the acidic structure purple. Similarly, eosin is used to stain the base structure pink. For blood and bone marrow smears, the staining is done using Romanowski stains such as Wright's stain, Giemsa stain, etc.
- 6. Mounting:** To preserve and prepare the stained section for light microscopy, it is mounted on a clear glass slide and covered with a thin glass coverslip. A resin-based mounting medium is used to adhere to the coverslip to the slide. Finally, the section area in the case of histology and smear area for blood and bone marrow smears on the slide is covered with a coverslip using the mounting media.

Once the slide is prepared, its images are captured using a digital pathology scanner or a camera mounted on the microscope.

1.3.2 Capture of microscopic images from healthy and cancer subjects for B-acute lymphoblastic leukemia cancer

In this section, we discuss one of the recently released ALL dataset. A dataset of 118 subjects, 49 healthy and 69 patients diagnosed with B-lineage ALL (B-ALL), was prepared at Laboratory Oncology, All India Institute of Medical Sciences (AIIMS), New Delhi, India. A waiver for written informed consent is obtained from the Ethics Committee of AIIMS, New Delhi, on this dataset for research purposes. All the subject identifying information was removed entirely from the image dataset by the doctors at AIIMS before sharing it with the other researchers.

Microscopic images were captured from bone marrow aspirate slides of subjects. Slides were stained using Jenner-Giemsa stain for better visibility of B-type immature white blood cells, also called lymphoblasts, under the microscope. Images were captured in raw BMP format with a size of 2560x1920 pixels using the Nikon Eclipse-200 microscope equipped with a digital camera at 100x magnification. The subjects were randomly sampled. The normal data (healthy cell images) was collected from subjects who did not suffer from cancer and hence, the ground truth labels are 100% correct for this class. The malignant cell images were collected from the patients who were initially diagnosed with cancer and had a sizeable leukemic cell growth in their blood. The medical expert's domain knowledge is used to prepare the dataset because morphologically, the healthy cells and the malignant blasts appear the same under the microscope, as shown in Fig. 1.1A and B. Also, all the cells from the cancer patients' data would not be the cancer cells. Thus, there can be a label noise of low value in the cancer class. However, the label noise would be below 1%, as confirmed by the oncologist expert. Data were annotated, that is, the cells of interest were marked by three expert oncologists to identify B-type white blood blasts in the microscopic images.

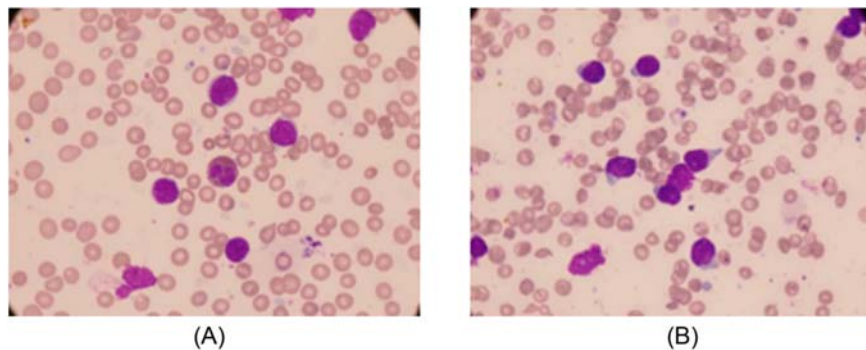


FIGURE 1.1

(A) Lymphoblasts (cancer cells) (B) Hematogones (healthy cells).

Different illumination settings were used to capture images from subject slides. The workflow involved in capturing and saving the images also varied to some extent. As this data was collected over three years and different members from the team contributed to the data collection, there is sufficient variability in the data that emulates real-life scenarios of data capture. In other words, the data collection procedure mimics the real-world data collection setting where the data comes from different sources and often multiple staff members are involved during the data collection. Overall, the expert oncologist has made sure that there is enough variability in the data by following the predesigned protocols. These involve:

- (1) Capturing images from different fields of view instead of focusing on one area of view of the slide.
- (2) Multiple slides per patient were made that provided more depth about the variability within the subject.

1.4 Normalization of color stain to correct for abnormalities during the staining process

Before imaging, microscopic slides are prepared manually using the staining chemicals and are, thus, prone to irregularities. As cell segmentation and classification may utilize color information, the performance of such tools is susceptible to color variations. This presents the need for color (stain) normalization of stained microscopic images for building any computer-assisted automated diagnostic tool. While preparing histopathology slides, the captured microscopic images exhibit color variations from batch to batch owing to the following reasons [2]:

- (1) **Illumination condition:** The first cause of color variation in microscopic images is illumination condition and camera type. This type of color variation is characterized by the product of camera response and spectral power distribution (SPD) of imaging light and requires correction. This is to note that in microscopic images, uneven illumination or vignetting, is not the case because a slide is tiny and is well-illuminated.
- (2) **Stain chemical:** Stain chemicals vary in composition by brands, by batches, and get affected over time due to chemical reactions. This variability in stain chemicals causes variations in the colors of stained tissues from image to image. This effect is captured in the stain's absorbance spectrum across different wavelengths or sensor channels.
- (3) **Stain quantity:** The time duration for which stain is left on the microscopic slide also causes color variations. If stain is left for a longer time, the quantity of stain absorbed is more and hence, is reflected as the depth of stain quantity at any pixel position.

Problems (1) and (3) listed earlier are related to the staining process, while (2) is related to the staining chemical. Owing to these staining problems, images collected from the slides of different subjects exhibit variations from batch to batch, as shown in Fig. 1.2. Thus, for building a robust tool, it is important to take care of the staining related errors.

Some of the widely used stain normalization methods are histogram equalization and color transfer methods. However, these are blind to histological information and lead to alteration of the same because they

1. ignore local color differences,
2. lead to smearing of histological components for overlapping PDFs of regions of interest (ROIs) in color spaces, and
3. may alter nucleus or cytoplasm boundaries and/or their textures.

Color deconvolution methods are the most promising methods that present a mathematical framework for stain color correction via singular value decomposition (SVD) and non-negative matrix factorization (NMF). However, existing SVD and NMF based methods replace the stain color basis of query image with that of the reference image instead of implementing basis transformation. These methods also visualize illumination and color variation as independent problems and do not entirely exploit the geometry of the underlying basis. Most of the existing techniques alter the reference image itself if treated as a query. Here, we discuss a recently proposed new method, namely GCTI-SN [2], that is a complete pipeline to address all the three causes of stain variations consisting of three stages [3]. The imaging process of the microscopic images is understood as below. While imaging, a specimen slide is exposed to incident light. Assuming a 3-sensor RGB [red (R), green (G), and blue (B) color sensors] camera, the intensity at a pixel p in the i th color sensor channel is given by [4]:

$$I(p, \lambda_i) = \int_{\lambda_i - \delta}^{\lambda_i + \delta} f_i(\lambda) E(\lambda) e^{-m_i(\lambda)d(p)} d\lambda \quad (1.1)$$

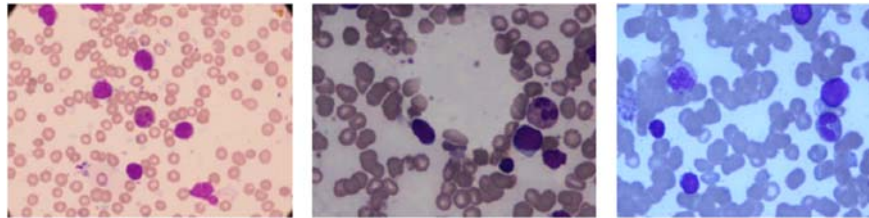


FIGURE 1.2

Color variations in microscopic images owing to staining process.

where $i=1, 2, 3$ correspond to R, G, and B channels, $f_i(\lambda) > 0$ for $\lambda \in (\lambda_i - \delta, \lambda_i + \delta)$ represents the i th sensor's response of camera within $\pm \delta$ of its color wavelength, $E(\lambda)$ denotes SPD of imaging light, $m_i(\lambda)$ denotes the characteristic absorbance of the stain in the i th sensor channel, and $d(p)$ denotes the stain depth or the quantity of stain bound at pixel position p . The stain chemical binds to the tissue of interest and absorbs colors of visible light spectrum according to its texture. The above equation can be simplified and the intensity at pixel p in the optical density (OD) space is defined as:

$$OD_i(p) = -\log \frac{I(p, \lambda_i)}{I^b(\lambda_i)} = m_i(\lambda_i)d(p), \quad (1.2)$$

where $I^b(\lambda_i)$ is a scalar quantity that denotes the background (BG) intensity in image or the intensity of unstained pixels, that is, with $d(p) = 0$. This equation is similar to the Beer-Lambert law and relates image intensity to stain's absorbance spectrum $m_i(\lambda)$ and the quantity of stain $d(p)$ present at that pixel. As both $I(p, \lambda_i)$ and $I^b(\lambda_i)$ are known for a microscopic image in each of the i th sensor channel, OD values for the image can be computed in all the three channels. Thus, at each pixel, we obtain a 3×1 vector of OD values. Stacking all pixels' OD values in a matrix, we obtain a $3 \times MN$ matrix I_{OD} representing the OD values of an $M \times N$ size RGB image.

The stain correction method requires the fixing of one image as the reference image. The input query images are stain normalized for the three errors listed earlier with reference to the reference image. The first stage carries out robust *illumination correction*. In the ideal scenario, unstained BG pixels in the image would be characterized by RGB value [1 1 1]. However, due to illumination variation, BG pixels have intensities different from [1 1 1]. Thus, (2) requires conversion from RGB to OD space, where BG pixel's intensity value is transformed to origin in the OD space. If illumination variation is not corrected appropriately, it leads to translation between the origins of the Cartesian frames of reference and query images in the OD space. This step requires a robust identification of the unstained BG in the query image and, thereafter, computation of (2) for every pixel.

In the second stage of color basis correction, GCTI-SN employed SVD of OD matrix into stain basis vector and stain quantity matrices as below:

$$I_{OD} = \Psi A \quad (1.3)$$

where I_{OD} is a $3 \times MN$ OD matrix, Ψ is a 3×3 stain basis matrix representing the characteristic absorbance of stains for each of the three channels, and A is a $3 \times MN$ matrix with each column storing the quantity of each of the staining chemical with both Ψ and A as unknowns. The GCTI-SN workflow consists of the following steps:

1. Finding stain basis vectors using SVD
2. Aligning the color basis frame of the query image to that of the reference image

3. Finding robust stain color vectors using the wedge finding method [5] for both query and reference images
4. Providing appropriate rotation to every pixel in the OD space that aligns the wedges of both query and reference images.

Finally, stain quantity correction is achieved via histogram normalization. The quantitative and qualitative results demonstrate the comparatively better performance of the GCTI-SN method vis-a-vis existing methods.

1.4.1 Quantitative results

We identified the ROI, that is, the nucleus of lymphoblasts in B-ALL images and compared the performance quantitatively in terms of mean square distance (MSD) of the stain color between the reference and the normalized query images over the ROI. This is realized by defining masks over ROIs shown as white circles in Fig. 1.3. On an average, each image mask covers at least three nuclei in that image. Fig. 1.4 shows the box plot, and Fig. 1.5 shows the qualitative results of the different methods.

1.5 Segmentation of cells of interest (in B-lineage ALL cancer)

Once the images are stain normalized, cells are required to be segmented out of the images. Cell segmentation methods can be broadly divided into the following

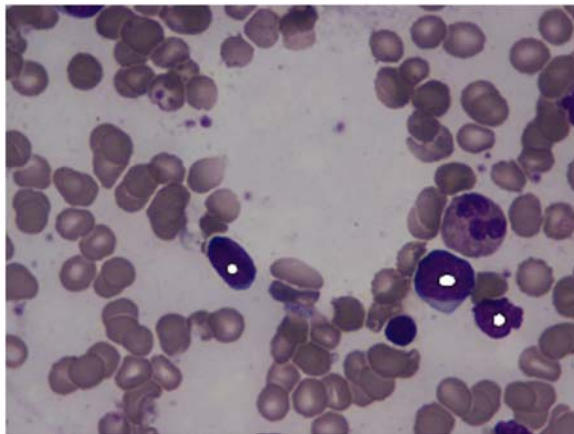
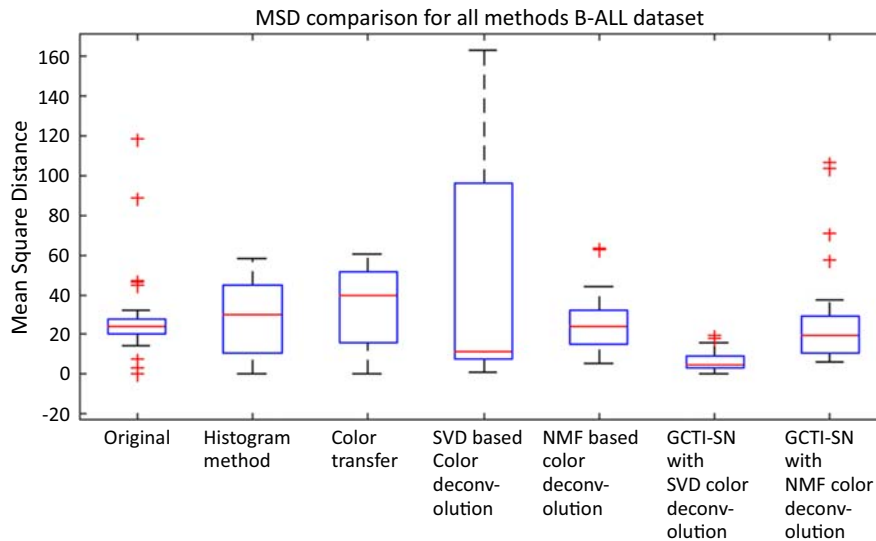


FIGURE 1.3

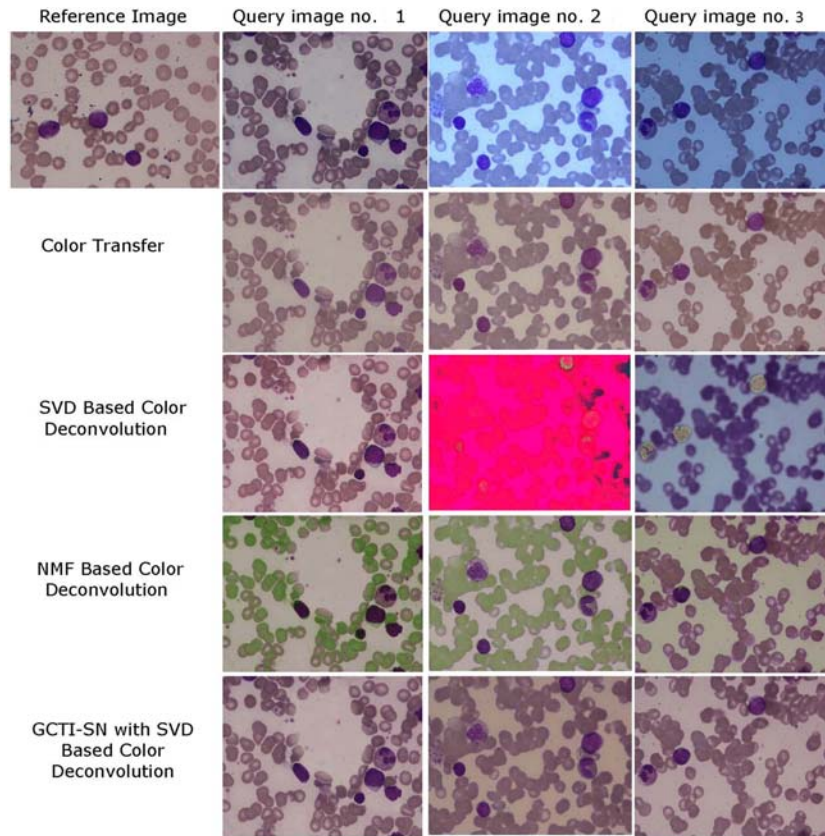
B-ALL Image with mask over nucleus to compute quantitative results in Mean Square Distance from the color of nucleus of the reference image.

**FIGURE 1.4**

Box Plot on MSD for B-All images on Jenner-Giemsa Stain (tested on 30 images) Methods used for comparison are: Histogram Method [6], Color Transfer Method [7], SVD-based Color Deconvolution Method [5], NMF-based Color Deconvolution Method [4], and GCTI-SN [2].

Taken from A Gupta et al., GCTI-SN: geometry-inspired chemical and tissue invariant stain normalization of microscopic medical images, Med. Image Anal. 65 (2020) 101788.

categories: intensity thresholding based, contour-based, region-based, and clustering-based methods. *Intensity thresholding* based segmentation is one of the simplest and fastest methods of image segmentation. However, it does not provide good segmentation results. *Active contour model* [8], popularly known as the snake model, works on deformable curves that change their shape according to the boundaries of the targeted object in the image. In this method, a set of internal and external forces define how snakes conform to an object boundary. These methods require an initial region of interest (ROI) as an input. As this ROI may vary from cell to cell, it cannot be fully automated. *Region-based segmentation* approaches generally look for connected components based on properties such as texture and brightness. Similar regions are combined, and the same procedure is repeated until the entire image is split into regions that belong to the same category. These approaches include seed-based region growing and merging methods [9]. *Image clustering* corresponds to segmentation via a grouping of similar pixels (based on some metric, say Euclidean distance on intensity) into a single cluster and correspondingly dividing it into multiple clusters. *k-means clustering and watershed* are some of the most often used algorithms in segmentation [10,11]. Machine *learning* techniques have also been employed for cell segmentation,

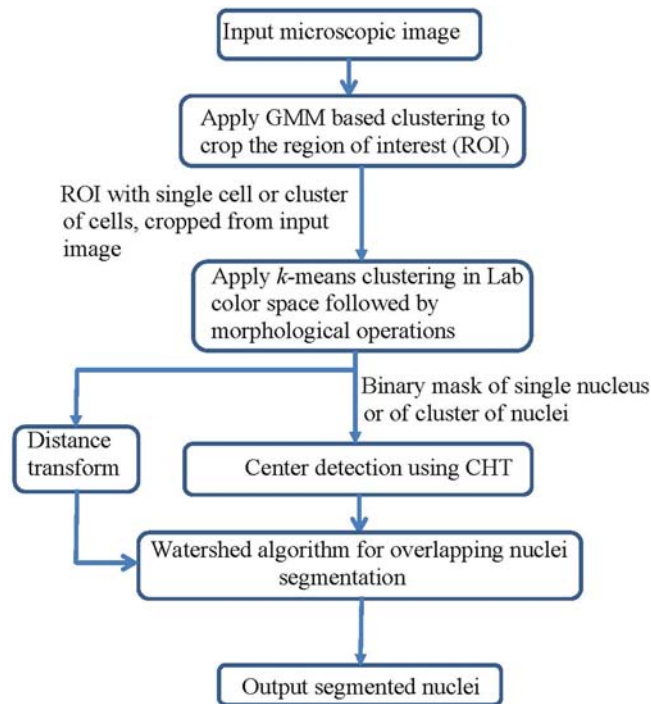
**FIGURE 1.5**

Stain Normalization-Qualitative Results of histogram equalization method, color transfer method, SVD and NMF based color deconvolution methods, and the GCTI-SN method [2] on three images of B-ALL stained with Jenner-Giemsa Stain.

wherein the hybrid watershed and support vector machine (SVM) classifier-based approaches have been used for cell nucleus segmentation from pap-smear images [12]. However, machine learning methods work on hand-crafted features, which can limit their performance.

B-ALL images require segmentation of the nucleus of B-lineage lymphoblasts. However, these cell nuclei appear isolated as well as in clusters. For classification, these cells need to be segregated out of the images and broken from clusters to individual cells. We implemented cell segmentation using three different methods in B-ALL stain normalized images:

1. Method-1: via fully-automated cell segmentation pipeline shown in Fig. 1.6
2. Method-2 via deep belief network (DBN) [13]

**FIGURE 1.6**

Fully automated cell segmentation pipeline.

3. Method-3: via a newly proposed method based on convolutional neural network (CNN).

1.5.1 Method-1 of cell segmentation using traditional image processing techniques

In method-1, we used supervised Gaussian mixture modeling (GMM) based clustering to locate possible ROIs, that is, the cell nuclei, in the original image. We cropped these ROIs from the original image for segmentation later. This step is very fast because it uses the apriori trained GMM model. Further, we used k -means in the Lab color space along with the morphological operations to extract a single or cluster of nuclei from the cropped ROIs extracted from the GMM clustering step. As k -means on the original image would have consumed significant time, GMM clustering added with k -means clustering provides robust and computationally fast segmentation methods. GMM stage helps with reducing the number of false positives that could otherwise get detected as nuclei with applying only

k-means stage. Next, we employed the combination of distance transform, circular Hough transform, and watershed algorithm to segment the overlapping nuclei.

1.5.2 Method-2 of cell segmentation using deep belief network

Deep learning methods provide another alternative to all these existing methods and are increasingly used in image segmentation. In method-2, we attempted a 4-layer deep belief NN [13]. A DBN consists of stacked layers of restricted Boltzmann machine. Again, the cell nuclei were identified in the Lab color space using K-means clustering. However, the cluster of nuclei was segmented innovatively by marking the joining ridge of two cell nuclei as one class label, foreground (cell nuclei) as the second class label and BG as the third class label. Next, the 4-layer DBN was trained, and the joining ridge pixels were identified. The pixels of these ridges were dropped to separate cell nuclei from clusters, leading to cell segmentation. This method worked better than the existing techniques that lead to the over-segmentation of some cells or fail to segment cells from the clusters [13].

1.5.3 Method-3 of cell segmentation using novel convolutional neural network architecture

Here, we propose a new method that utilizes a deeper CNN architecture for cell segmentation. Before presenting the explanation of this method, we provide a brief review of CNN architectures and of medical imaging segmentation methods using CNNs.

1.5.3.1 Brief review of convolutional neural network architectures

With deeper CNN architectures, segmentation is implemented via classification at the pixel level. However, the task is more complicated compared to image classification because the label of each pixel of the image is required to be predicted. The broader idea of deep learning architectures is that the NN can acquire the necessary features to carry out classification by the composition of functions implemented via different layers of the network. CNN architecture is particularly prevalent due to the requirement of a smaller number of parameters relative to a fully connected network. A CNN architecture consists of multiple convolutional layers. Each of them implements the convolution of the input image map with a small kernel (3x3, 5x5, or so on) followed by activation functions that incorporate non-linearity between different layers. Besides, there may or may not be pool layers between different CNN layers. These pool layers can upsample or down-sample the feature maps (output of any layer) depending upon the implementation. The last layers of a CNN network consist of fully connected layers followed by a softmax layer for classification similar to the traditional NN.

The training of CNNs involves the minimization of a targeted loss function that helps in extracting task-specific features. For classification at the image level, the input image is transformed into a feature vector after passing through a series of convolutional and pooling layers. This feature vector is then used to decide the label of the input image. However, segmentation is required to produce the output image of the same size as the input image, but with an output image containing each pixel label according to the object to which the pixel belongs. For example, say in an image, we would like to do segmentation of foreground and BG. This can be achieved via the identification and labeling of a foreground pixel with “one” and labeling a BG pixel with “zero.”

1.5.3.2 Semantic versus instance segmentation in medical imaging

In the context of medical imaging, another essential classification of the type of segmentation is semantic segmentation vs instance segmentation. Instance segmentation is generally more complex than semantic segmentation. In semantic segmentation, the aim is to map each pixel to the object it belongs to, whereas instance segmentation further requires identifying each instance of the object. For example, in cytology images (say microscopic images of blood cells), we might be interested in counting the number of cancer cells as a ratio of total cells. Hence, it is crucial to carry out cluster cell segmentation to identify each cell instance. Fig. 1.7 shows the difference between these two types of segmentations.

Moreover, we might build the cell classification tool that works on a cell image instead of cell clusters. For instance, instance segmentation is essential because semantic segmentation will treat a cluster as a single object, whereas instance segmentation will try to extract the individual cells from the clusters.

The first attempt for semantic segmentation using CNNs was made through a fully convolutional network (FCN) [14]. FCN consists of encoder–decoder type architecture, in which encoder is used for hierarchical downsampling of the image and decoder is used for upsampling the obtained feature maps, resulting in the segmented output mask of the same size as the input image. However, the decoder struggles to generate a smoother segmentation mask with the feature

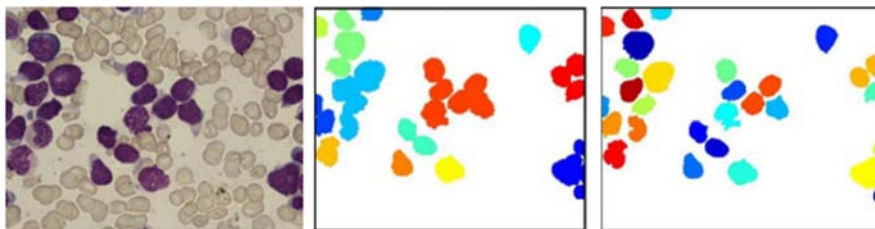


FIGURE 1.7

(A) Original image (B) semantic segmentation (each cluster of cells get one color label)

(C) instance segmentation (each cell gets individual color label).

maps obtained through repeated downsampling (by a factor of 32). To overcome this, skip connections from the previous layers were added that generate smoother segmentation masks (Fig. 1.8).

U-Net architecture [15] shown in Fig. 1.8 was designed for biomedical image segmentation and is one of the most popular architectures [16–20]. Owing to its good performance, U-Net became very popular in a short time. Its backbone architecture is being used actively by the research community. It employs encoder–decoder type architecture but with certain significant modifications. The decoder of the U-Net is deeper and is symmetrical to the encoder. The decoder upsamples the feature maps in the hierarchical fashion similar to downsampling in the encoder. The encoder downsamples the image in steps, where the number of feature maps is doubled and the size of the feature maps is halved with each step. A similar set up is used in the decoder, wherein the number of feature maps is halved and the size of the feature maps is doubled with each step. Finally, in the last layer of the decoder module, 1×1 conv filters are used to output the number of channels equal to the number of classes.

This arrangement gives this architecture a “U” shape and hence, it has the name of U-Net. Another important characteristic of U-Net is the concatenation of features from the encoder module to the decoder module. Each step of the decoder obtains concatenation of two sets of features- one set obtained for this step from the decoder and the other set from the corresponding step of the encoder. To match the spatial size of the concatenating feature maps, feature maps of the encoder are cropped to the same size as that of the decoder. However, this cropping step can be eliminated by using padding during the operation of convolution. The U-Net’s overall arrangement helps to capture complex structures in the data and generate smoother segmentation masks.

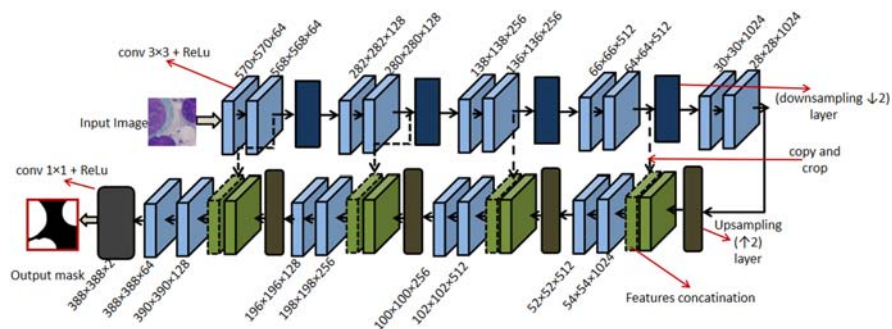


FIGURE 1.8

U-Net architecture [15].

Adapted from O. Ronneberger, P. Fischer, T. Brox, U-Net: convolutional networks for biomedical image segmentation, in: N. Navab, J. Hornegger, W. Wells, A. Frangi (Eds.), Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015, Lecture Notes in Computer Science, vol 9351, Springer, Cham.

The pixel-wise binary cross-entropy (BCE) loss function is used to train the U-Net. However, to achieve instance segmentation, the weighted loss function is used and boundary pixels are penalized more than results in segregated clusters. Generally, U-Net faces challenges in separating overlapped cells. Several approaches have been used to address this issue. For example, in [21], this problem is addressed by additionally predicting the contour of the ROI, while in [22] overlapping cell segmentation is formulated as a three-class classification: BG, foreground, and contour.

1.5.3.3 Method-3: novel proposed EDNiS-Net convolutional neural network for automated nuclei instance segmentation

On similar lines, we propose an encoder–decoder-based CNN for Nuclei-instance Segmentation (EDNiS-Net). The base module of the EDNiS-Net consists of three stacks of convolutional filters separated by non-linear functions that help with learning better discriminative functions, followed by a concatenation layer that promotes feature-reusability and provides a suitable weighting of essential features. Skip connections are introduced between the encoder and decoder modules to preserve the context information (Fig. 1.9). Also, to further enhance the instance segmentation, three output masks are predicted for each input image. These masks are foreground, BG, and edges.

An additional advantage of predicting the edge maps is that the scaled version of the edge maps can be subtracted from the foreground mask to separate the cells further. The architecture also utilizes batch normalization and dropout to check the stain variability and overfitting, respectively. The network is designed by utilizing several basic building blocks, named as base modules, to realize an encoder–decoder architecture, namely EDNiS-Net, as shown in Fig. 1.9.

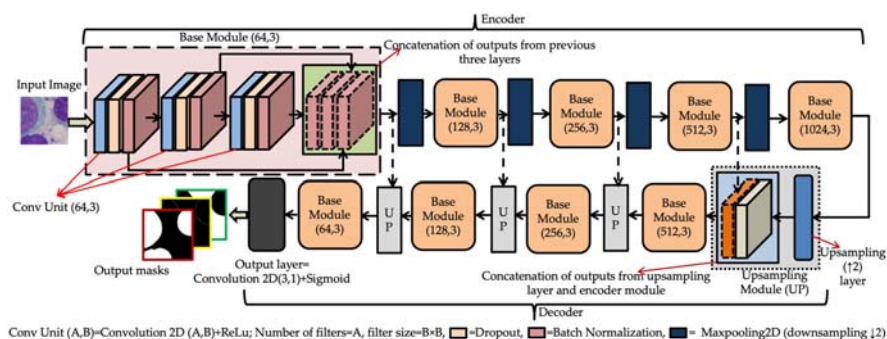


FIGURE 1.9

EDNiS-Net: Encoder-decoder-based convolutional neural network (CNN) for Nuclei-instance Segmentation.

1.5.3.3.1 Base module

The base module is designed to learn better discriminative functions and to have an efficient flow of features. The base module consists of three layers of $k \times 3 \times 3$ convolutional filters, where k is the number of filters. Each convolutional layer is followed by dropout and batch normalization layers. Batch normalization helps mitigate the effect of color variation of input images commonly observed in stained microscopic images. It also speeds up the training of the network. Dropout counters overfitting when the training data is less than the requirement as per the network's capacity, leading to satisfactory generalization even with less training data. Finally, the output of all three layers is concatenated and passed to the next layer. Concatenated features from all the layers allow reusability of features from the previous layers, wherein the network decides the importance of features for different labels. This strategy also provides shorter connections in the network that checks the vanishing gradient problem in deep architectures.

1.5.3.3.2 Encoder module

Encoder module consists of five base modules and four downsampling (maxpool) layers. The number of filters in the first base module is 64 and increases by a factor of two in the following base modules. Each base module, except the last, in the encoder, is followed by downsampling by two. Hence, the size of feature maps reduces progressively, while the number of filters increases. This arrangement helps to capture the intricate structure of the data.

1.5.3.3.3 Decoder module

The decoder module consists of base modules, upsampling modules, and an output layer. Upsampling module upsamples the incoming feature maps by a factor of two and concatenates the resultant features with the features of the same size from the encoder module. This direct flow of information from encoder to decoder helps in preserving the context information. A base module follows each upsampling module and thus, the size of feature maps increases progressively. After the final base module, the size of feature map is the same as the input image. Finally, the output layer has $c \times 1 \times 1$ filters, where c is the number of output channels set to three as we predict foreground (nuclei mask), BG, and edges.

1.5.3.3.4 Proposed loss function

For an input image \mathbf{I} with foreground \mathbf{M} , background \mathbf{B} , and edges \mathbf{E} , we predict $\hat{\mathbf{I}}$ with $\hat{\mathbf{M}}$, $\hat{\mathbf{B}}$, and $\hat{\mathbf{E}}$. Besides using the BCE loss and dice loss, we also introduce a distance learning term. Given \mathbf{M} , $\hat{\mathbf{M}}$ and $\hat{\mathbf{B}}$, we define:

$$d_{\mathbf{M}, \hat{\mathbf{M}}} = \mathbf{M} - \hat{\mathbf{M}}_F^2 \text{ and } d_{\mathbf{M}, \hat{\mathbf{B}}} = \mathbf{M} - \hat{\mathbf{B}}_F^2.$$

We minimize $d_{\mathbf{M}, \hat{\mathbf{M}}}$ and maximize $d_{\mathbf{M}, \hat{\mathbf{B}}}$ for better prediction of the segmentation mask and hence, define distance loss as:

$$\mathcal{L}_{dist} = \exp(d_{\mathbf{M}, \hat{\mathbf{M}}}) + \exp(-d_{\mathbf{M}, \hat{\mathbf{B}}})$$

We incorporated this term in the final loss function that is the weighted sum of BCE, dice loss, and distance loss as:

$$\mathcal{L} = w_1 \mathcal{L}_{BCE} + w_2 \mathcal{L}_{dist} + w_3 \mathcal{L}_{dice},$$

where BCE indicates binary cross entropy between input image \mathbf{I} and the predicted image $\hat{\mathbf{I}}$, and

$$\mathcal{L}_{dice} = 1 - 2 \times \frac{|\mathbf{I} \cap \mathbf{I}| + \epsilon}{|\mathbf{I}| + |\mathbf{I}| + \epsilon},$$

where ϵ is added to ensure non-zero denominator. Besides helping pixel classification, this loss also promotes distance learning that leads to better separation between the foreground and the BG.

1.5.3.3.5 Results and discussion

We report and compare results on two metrics: dice similarity coefficient (DSC or F1-score) and Intersection over Union (IoU or Jaccard index between the segmented and the ground truth masks) computed at the pixel level. We compare the quantitative and qualitative performance of the EDNiS-Net with U-Net trained using weighted combinations of BCE loss and dice loss, and U-Net trained with the proposed loss (U-Net +). [Table 1.1](#) shows the performance comparison in terms of DSC and IoU on our inhouse dataset of cell imaging on Multiple Myeloma, another type of blood cancer. For both the metrics, EDNiS-Net is leading the other two methods. U-Net + has better performance than U-Net, which proves the usefulness of the proposed loss function. This loss function promotes the BG-foreground separation leading to 0.6% and 1% performance gain in terms of DSC and IoU as compared to U-Net. Visualization of the segmented masks generated by U-Net + and EDNiS-Net is shown in [Fig. 1.10](#), where each instance of nucleus mask is assigned a different color, while the overlapping nuclei have the same color. EDNiS-Net is observed to separate majorly overlapped nuclei as well. It also has better precision and generates a smoother segmentation mask.

Table 1.1 Comparison of EDNiS-Net with U-Net and U-Net + on MM19 dataset.

Model	DSC	IoU
U-Net	0.933	0.885
U-Net +	0.939	0.895
EDNiS-Net	0.947	0.905

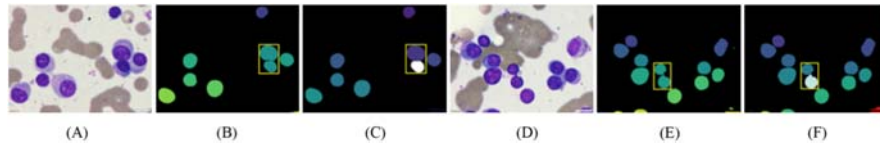


FIGURE 1.10

Predicted segmentation masks with different methods on sample images (A,D) of MM19 dataset. Each color denotes a different nucleus mask. In comparison to U-Net + (B,E), EDNiS-Net (C,F) is able to separate the overlapped nuclei efficiently.

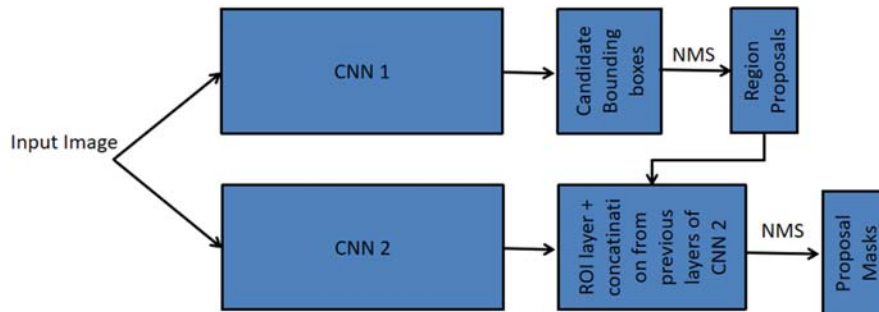


FIGURE 1.11

Cell segmentation with region-proposal based network [23].

Adapted from S.U. Akram et al., Cell segmentation proposal network for microscopy image analysis, in: Carneiro G. et al. (Eds.), Deep Learning and Data Labeling for Medical Applications, DLMIA 2016, LABELS 2016, Lecture Notes in Computer Science, vol 10008. Springer, Cham.

1.5.3.4 Region-proposal based convolutional neural network architectures

Another approach for nuclei/cell segmentation is based on object detection methods using region-proposal CNN architectures [24,25]. These methods predict axis-aligned bounding boxes representing the shapes of the objects of interest. Pixels inside bounding boxes are next classified to obtain refined instance segmentation. For example, Fig. 1.11 shows cell segmentation using a region proposal-based CNN network. In this architecture, the upper network produces the bounding boxes and the lower network uses these bounding boxes to generate the segmentation mask.

Although such networks have good performance in segregating a cluster, they perform relatively poor with reference to the accuracy of the segmented masks. Some authors have combined U-Net with detection networks for better cell segmentation performance [26,27]. Another related approach based on object detection is StarDist [28] that uses star-convex polygon instead of bounding boxes for representing the shape. For each pixel, distance to the boundary along predefined

radial directions is regressed along with the prediction of the pixel probability belonging to the object. Finally, a non-maximum suppression is performed to have the final segmentation mask. This approach seemed to work better than U-Net and Mask R-CNN [28].

1.6 Classification of cancer and healthy cells

Finally, the goal is to carry out the classification of cells for disease diagnosis and monitoring. Before the popularity of deep learning, SVM, NN, random forest, and naive Bayes trained with hand-crafted features were used as the popular machine learning methods to implement cell classification ALL cancer. For example, [29–39] used traditional machine learning methods of SVM, NN, Fuzzy C-means classifier, and K-nearest neighbor classifiers on the ALL-IDB dataset (<https://omictools.com/all-idb-tool>) and other datasets. The above classifiers used hand-crafted features, mainly consisting of color, shape, and texture features. As these features may not be the true descriptors of class discriminative characteristics, it may lead to non-optimal performance. Hence, an ideal solution would be to learn such discriminative features from the data itself. This solution is provided by the CNNs that consist of multiple two-dimensional filters arranged in a feed-forward manner.

A deeper architecture implements the composition of functions, allowing the network to learn a complex low-level and high-level features from the data. However, CNNs require a large training dataset that is generally not available, particularly, in the medical domain. This issue is resolved using transfer learning, wherein a pretrained CNN is fine-tuned with the given data. It is still a better approach to train a network from scratch, provided there is a required size dataset. A generic classification pipeline is shown in Fig. 1.12. The pipeline involves the extraction of the ROI (for example, a cell) using a segmentation algorithm (discussed in the previous section), which is subsequently identified as normal or cancer cell for diagnostic purposes. One of the significant limitations of the above

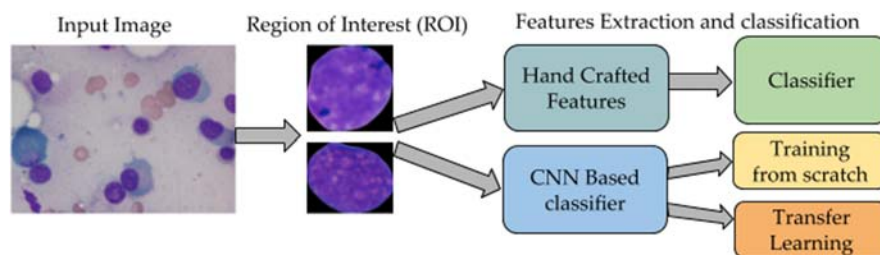


FIGURE 1.12

A generic classification pipeline.

works was the limited size of a dataset of fewer than 400 images. This can lead to overfitting on limited data that can potentially fail on prospective subjects' data. We addressed this issue by building a large-sized dataset for B-ALL cancer, as explained in the next section.

1.6.1 C-NMC 2019 challenge dataset

We prepared a dataset of 118 subjects, 49 healthy and 69 cancer subjects. We used our in-house developed method of stain normalization explained in [Section 1.4](#) followed by Method-1 of cell segmentation explained in [Section 1.5](#) on microscopic images. The data was appropriately curated. This dataset is publicly available at The Cancer Imaging Archive (TCIA) [\[41\]](#) and has been displayed according to the TCIA standard protocols [\[42\]](#). We also conducted the medical imaging challenge, namely, *Classification of Normal versus Malignant Cells in B-ALL White Blood Cancer Microscopic Images (C-NMC) 2019* [\[43\]](#), in the IEEE International Symposium on Biomedical Imaging (ISBI), 2019 conference held in Venice, Italy in April 2019.

Data is prepared at the subject-level and is split into the training set and testing set. The training set consists of 8491 cancer cell images collected from 60 cancer subjects and 4037 normal cell images collected from 41 healthy subjects, with a total of 1,2528 images [\[41\]](#). Test data contains a total of 2586 cell images, collected from 9 cancer subjects and 8 healthy subjects [\[41\]](#). None of the cell images of these subjects are used for training the classifier and hence, results of the test data are unbiased. As original cell images are of different sizes, all images are zero-padded such that the centroid of every cell is at the center. After zero-padding, all cell images are made to the size of 350×350 pixels. Overall, the dataset consists of three challenges:

- (1) The dataset has a class imbalance, that is, the number of cancer cell images is almost double the number of normal (healthy) cell images.
- (2) The dataset has label noise in cancer class. As explained above, the dataset is prepared with the ground truth obtained from the medical expert's domain knowledge.
- (3) It is required to capture subject variability that is crucial because the proposed tool should work robustly on the prospective subjects' data. This is also the reason that the data is prepared and made available at the subject level.

1.6.2 Classification on C-NMC 2019 dataset

The very first work carried out by us on the B-ALL dataset was in [\[44\]](#). In this work, we presented the stain deconvolutional layer-based CNN model to classify healthy and cancer cells by projecting the image data to OD space via stain deconvolution. However, a significant limitation of the approach was that the

train-test split was not done at the subject-level. Hence, the same subject's images could be present in both the training and the test data. This can cause the classifier to fail on the prospective (new unseen) subjects' data.

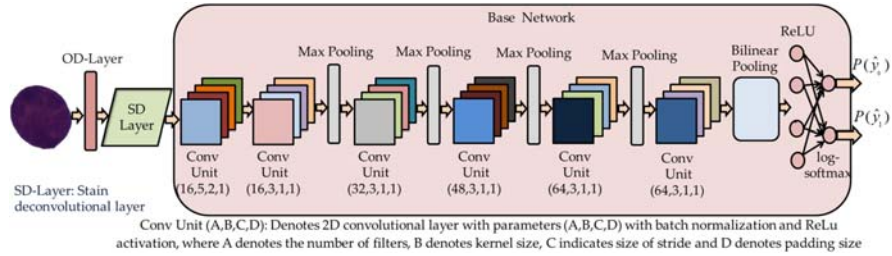
Learning from this, we again segmented and prepared the dataset at the subject level, which was subsequently deposited at TCIA and was named as C-NMC 2019 dataset. The dataset is prepared at the subject level so that all images of a given subject are present in either train, validation, or test set. In the 2019 challenge, several teams participated and have published their works on this dataset. The section below briefly discusses the top-performing methods of the challenge.

A neighborhood-correction algorithm in combination with fine-tuned ResNets was used in [45]. In this work, ResNets were used to generate the feature maps, which were then utilized for constructing the Fisher vectors. Finally, these Fisher vectors were exploited by a neighbors' majority-based label correction approach. This algorithm was used with an ensemble of ResNet50, ResNet101, and ResNet150, achieving a weighted-F1 score of 91.04%, leading to the first position in the challenge.

Authors in [46] also exploited a pretrained ResNeXt50, which was fine-tuned using a different learning rate for different layers. They utilized different augmentations during testing, resulting in the final weighted-F1 score of 88.9%. An ensembling of DenseNet121, SENet154, ResNet101, DeepTEN, VGG16, and InceptionV4 combined with pseudo-labeling approach was used in the DeepMen architecture in [47]. In DeepMen, a pretrained network is iteratively used to predict the labels of test data, which are then used to re-train the model, assuming them to be the ground truth. This approach achieved a final weighted-F1 score of 88.5%. Authors in [48] exploited PNASNet along with the fusion and voting approach, which resulted in a weighted-F1 score of 87.9%. Authors in [49] had split the majority class samples into two smaller subsets and combined them with the minority class samples to form two different training sets. These two sets were used to fine-tune two separate pretrained Inception ResNets. The resulting models were then jointly fine-tuned on the complete training set to make the final predictions. This method was able to achieve a weighted-F1 score of 87.6%. Other researchers used other similar methodologies based on Inception [50–53].

1.6.3 SDCT-AuxNet^θ CNN architecture for C-NMC 2019 dataset

Recently, we have introduced another architecture, namely, SDCT-AuxNet^θ [40]. This architecture has achieved, so far, the best performance compared to all the methods discussed earlier on the C-NMC dataset. The SDCT-AuxNet^θ utilizes SD-Net shown in Fig. 1.13 as the backbone architecture [54]. The SD-Net uses SD-layer [44] appended at the front of a compact CNN architecture. The SD-layer operates in the OD space (OD-space) and helps calculate stain-absorbed quantities based on the cell category. As the absorbed quantities comprise the texture of the tissue or of the cells of interest, these are better representatives of texture class features as compared to RGB pixel intensities. Thus, SD-layer helps in improving

**FIGURE 1.13**

Architecture of SD-Net for cell classification for B-ALL cancer diagnosis [40].

Taken from S. Gehlot, A. Gupta and R. Gupta, *SDCT-AuxNet⁰: DCT augmented stain deconvolutional CNN with auxiliary classifier for cancer diagnosis*, *Medical Image Anal.* 61 (2020) 101661.

classification performance. A detailed explanation of the SD-layer is provided in [44]. In [54], we modified this architecture, where the DCT-layer follows SD-layer. DCT-layer is inspired by the concept of flow cytometry, which is the conventional approach for cancer diagnosis. It is based on the spectroscopy or frequency domain analysis of the sample. DCT-layer, in combination with SD-layer, helps to boost the class discriminative features. DCT-layer projects the input image in the DCT domain and induces class-dependent sparsity.

Interestingly, DCT-layer induces more sparsity in the healthy cells than the cancer cells as observed in [40]. This difference in the sparsity helps enhance the performance of classification. Here, an input image is sequentially projected onto the OD-space that helps in calculating class-dependent stain quantities and then to the DCT-domain that induces class-dependent sparsity. As a next modification, SDCT-AuxNet⁰ is proposed that consists of an auxiliary classifier that is trained and tested with a novel strategy. SDCT-AuxNet⁰ has two classifiers: (1) SDCT-Net and (2) an auxiliary classifier. Also, this architecture uses two different transformations. While SDCT-Net uses bilinear pooling, auxiliary classifier uses spectral averaging (depth-wise averaging) of the feature maps to classify cell types.

Training of the SDCT-AuxNet⁰: The training of the SDCT-AuxNet⁰ is carried out in two steps. First, the SDCT-Net is trained in an end-to-end fashion. Once, the SDCT-Net is trained, the prediction on all the training data is made, and the correctly predicted samples are used to train the auxiliary classifier. The correctly predicted samples are passed through the feature extracting portion of the SDCT-Net, they are passed to the spectral averaging layer and then used to train the auxiliary classifier, as shown in Fig. 1.14.

Testing of the SDCT-AuxNet⁰: Similar to training, the testing of SDCT-AuxNet⁰ is also a two-step process, as shown in Fig. 1.15. Both the classifiers (SDCT-Net and auxiliary classifier) are used in a coupled fashion, and the coupling ratio is decided by a user-defined variable called confidence probability. First, for a given test sample, the prediction is made using the SDCT-Net and its confidence (probability) is observed. If the prediction probability is higher than

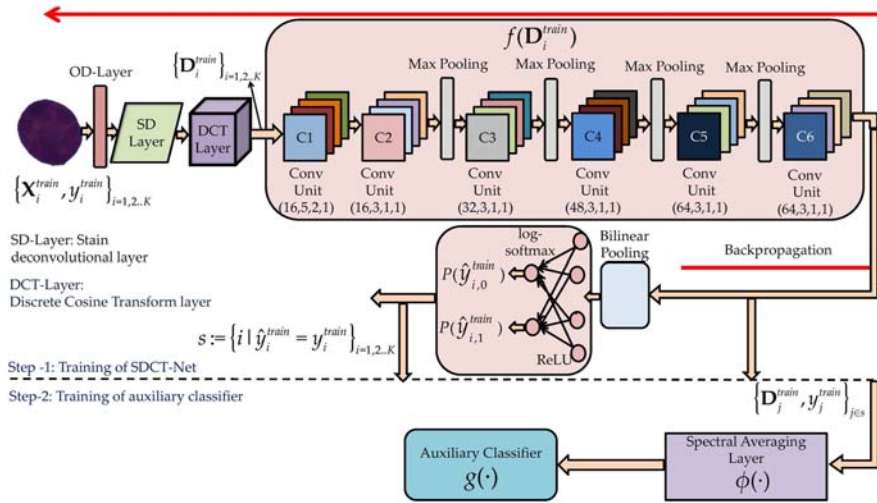


FIGURE 1.14

Training of SDCT-AuxNet⁰.

Taken from S. Gehlot, A. Gupta and R. Gupta, SDCT-AuxNet⁰: DCT augmented stain deconvolutional CNN with auxiliary classifier for cancer diagnosis, Medical Image Anal. 61 (2020) 101661.

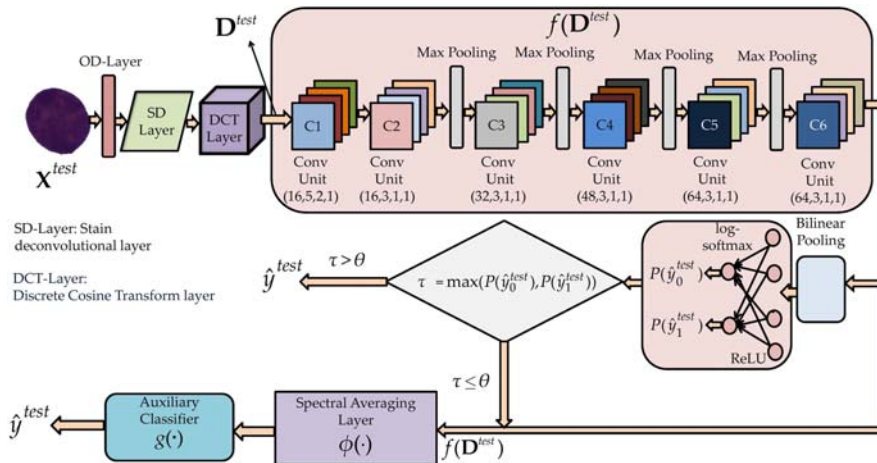


FIGURE 1.15

Testing of the SDCT-AuxNet⁰.

Taken from S. Gehlot, A. Gupta and R. Gupta, SDCT-AuxNet⁰: DCT augmented stain deconvolutional CNN with auxiliary classifier for cancer diagnosis, Medical Image Anal. 61 (2020) 101661.

the confidence probability, the prediction is considered to be final. However, if it is less than the confidence probability, the test sample is first passed to the feature extraction portion of the SDCT-Net and then to the spectral averaging. Finally, it is predicted using the auxiliary classifier. This prediction by the auxiliary classifier is considered to be the final decision. Thus, the confidence probability is deciding the contribution of both the classifiers. Also, SDCT-AuxNet⁰ is a generic architecture, with SDCT-Net being a subset of it with $\theta = 0$, while only the auxiliary classifier is active for $\theta = 1$ in the SDCT-AuxNet⁰ architecture. This architecture can achieve a weighted-F1 score of 94.8% that is highest on the C-NMC 2019 dataset.

1.7 Conclusions

In this chapter, we have addressed the problem of building an automated diagnostic tool for B-ALL cancer, a commonly occurring pediatric white blood cancer. We have explained the challenges associated with building the automated pipeline, the steps involved, and the efforts carried out in different stages as well as the current literature. In particular, we have explained the process of normalization, segmentation, and classification of cancer cells with respect to healthy cells with B-ALL as a prototype disease. We have described our efforts on the C-NMC 2019 B-ALL dataset, which is the largest curated dataset of cell images of B-ALL cancer. Deep learning architectures, results, and discussion on this dataset are provided in detail in this chapter.

References

- [1] M.B.B. Tyagi, V. Raina, Cancer incidences in urban Delhi-2001-05, *Asian Pac. J. Cancer Prev.* 10 (5) (2009) 799–806.
- [2] A. Gupta, et al., GCTI-SN: Geometry-inspired chemical and tissue invariant stain normalization of microscopic medical images, *Med. Image Anal.* 65 (2020) 101788.
- [3] R. Gupta, et al., Stain color normalization and segmentation of plasma cells in microscopic images as a prelude to development of computer assisted automated disease diagnostic tool in multiple myeloma, *Clin. Lymphoma, Myeloma Leukemia* 17 (1) (2017) e99.
- [4] Li, et al., A complete color normalization approach to histopathology images using color cues computed from saturation-weighted statistics, *IEEE Trans. Biomed. Eng.* 62 (7) (2015) 1862–1873.
- [5] M. Macenko et al., A method for normalizing histology slides for quantitative analysis, in: 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Boston, MA, 2009, pp. 1107–1110.
- [6] A.K. Jain, *Fundamentals of Digital Image Processing*, Prentice-Hall, Inc, 1989.
- [7] E. Reinhard, et al., Color transfer between images, *IEEE Computer Graph. Appl.* 21 (5) (2001) 34–41.

- [8] M. Kass, A. Witkin, D. Terzopoulos, Snakes: active contour models, *Int. J. Computer Vis.* 1 (1988) 321–331.
- [9] W. Gao, Y. Tang, X. Li, Segmentation of microscopic images for counting leukocytes, in: 2008 2nd International Conference on Bioinformatics and Biomedical Engineering, Shanghai, 2008, pp. 2609–2612.
- [10] J.M. Sharif et al., Red blood cell segmentation using masking and watershed algorithm: A preliminary study, in: 2012 International Conference on Biomedical Engineering (ICoBE), Penang, 2012, pp. 258–262.
- [11] M. Yan et al., K-means cluster algorithm based on color image enhancement for cell segmentation, in: 2012 5th International Conference on BioMedical Engineering and Informatics, Chongqing, 2012, pp. 295–299.
- [12] Orozco-Monteaudo, et al., Combined hierarchical watershed segmentation and svm classification for pap smear cell nucleus extraction, *Computación y. Sist.* 16 (2) (2012) 133–145.
- [13] R. Duggal et al., Overlapping Cell Nuclei Segmentation in Microscopic Images Using Deep Belief Networks, in: Proceedings of the Tenth Indian Conference on Computer Vision, Graphics and Image Processing (ICVGIP '16). Association for Computing Machinery, New York, Article 82, 2016, pp. 1–8.
- [14] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, 2015, pp. 3431–3440.
- [15] O. Ronneberger, P. Fischer, T. Brox, U-Net: convolutional networks for biomedical image segmentation, in: N. Navab, J. Hornegger, W. Wells, A. Frangi (Eds.), *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*, Lecture Notes in Computer Science, vol. 9351, Springer, Cham.
- [16] F. Milletari, N. Navab, S. Ahmad, V-Net: fully convolutional neural networks for volumetric medical image segmentation, in: 2016 Fourth International Conference on 3D Vision (3DV). Stanford, CA, 2016, pp. 565–571.
- [17] Ö. Çiçek et al., 3D U-Net: learning dense volumetric segmentation from sparse annotation, in: S. Ourselin, L. Joskowicz, M. Sabuncu, G. Unal, W. Wells (Eds.), *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2016*, Lecture Notes in Computer Science, vol. 9901. Springer, Cham.
- [18] V. Badrinarayanan, A. Kendall, R. Cipolla, SegNet: a deep convolutional encoder-decoder architecture for image segmentation, *IEEE Trans. Pattern Anal. Mach. Intell.* 39 (12) (2017) 2481–2495.
- [19] L. Xiaomeng, et al., H-DenseUNet: hybrid densely connected unet for liver and tumor segmentation from CT volumes, *IEEE Trans. Med. Imaging* 37 (12) (2018) 2663–2674.
- [20] Z. Zhou et al., UNet++: a nested U-net architecture for medical image segmentation, in: D. Stoyanov et al. (Eds.), *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, DLMIA 2018, ML-CDS 2018*, Lecture Notes in Computer Science, vol 11045, Springer, Cham.
- [21] Chen, et al., DCAN: deep contour-aware networks for object instance segmentation from histology images, *Med. Image Anal.* 36 (2017) 135–146.
- [22] N. Kumar, et al., A dataset and a technique for generalized nuclear segmentation for computational pathology, *IEEE Trans. Med. Imaging* 36 (7) (2017) 1550–1560.

- [23] S.U. Akram et al., Cell segmentation proposal network for microscopy image analysis, in: G. Carneiro et al. (Eds.), *Deep Learning and Data Labeling for Medical Applications, DLMIA 2016, LABELS 2016, Lecture Notes in Computer Science*, vol. 10008. Springer, Cham.
- [24] K. He et al., Mask R-CNN, 2017 IEEE International Conference on Computer Vision (ICCV), Venice, 2017, pp. 2980–2988.
- [25] J. W. Johnson, Automatic Nucleus Segmentation with Mask-RCNN, in: K. Arai, S. Kapoor (Eds.), *Advances in Computer Vision, CVC 2019, Advances in Intelligent Systems and Computing*, vol 944. Springer, Cham.
- [26] A. O. Vuola, S. U. Akram, J. Kannala, Mask-RCNN and U-Net ensembled for nuclei segmentation, in: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), Venice, Italy, 2019, pp. 208–212.
- [27] Z. Xu et al., US-Net for robust and efficient nuclei instance segmentation, in: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), Venice, Italy, 2019, pp. 44–47.
- [28] U. Schmidt et al. Cell detection with star-convex polygons, in: A. Frangi, J. Schnabel, C. Davatzikos, C. Alberola-López, G. Fichtinger (Eds.), *Medical Image Computing and Computer Assisted Intervention—MICCAI 2018, Lecture Notes in Computer Science*, vol. 11071. Springer, Cham.
- [29] G. Singh, G. Bathla, S. Kaur, Design of new architecture to detect leukaemia cancer from medical images, *Int. J. Appl. Eng. Res.* 11 (10) (2016) 7087–7094.
- [30] V. Singhal, P. Singh, Texture features for the detection of acute lymphoblastic leukaemia, in: S. Satapathy, A. Joshi, N. Modi, N. Pathak (Eds.), *Proceedings of International Conference on ICT for Sustainable Development, Advances in Intelligent Systems and Computing*, 409, Springer, Singapore, 2016.
- [31] P. Viswanathan, Fuzzy C means detection of leukemia based on morphological contour segmentation, *Procedia Computer Sci.* 58 (2015) 84–90.
- [32] M.M. Amin, et al., Recognition of acute lymphoblastic leukaemia cells in microscopic images using k-means clustering and support vector machine classifier, *J. Med. Signals Sens.* 5 (1) (2015) 49–58.
- [33] S.C. Neoh, et al., An intelligent decision support system for leukaemia diagnosis using microscopic blood images, *Sci. Rep.* 5 (2015) 14938.
- [34] R. Bhattacharjee, L.M. Saini, Robust technique for the detection of acute lymphoblastic leukemia, in: 2015 IEEE Power, Communication and Information Technology Conference (PCITC), 2015, pp. 657–662.
- [35] L. Putzu, G. Caocci, C. Di, Ruberto, Leucocyte classification for leukaemia detection using image processing techniques, *Artif. Intell. Med.* 62 (3) (2014) 179–191.
- [36] L. Putzu, C. Di Ruberto, White blood cells identification and classification from leukemic blood image, in: *Proceedings of the IWBBIO International Work-Conference on Bioinformatics and Biomedical Engineering*, 2013, pp. 99–106.
- [37] H.T. Madhloom, et al., A robust feature extraction and selection method for the recognition of lymphocytes vs acute lymphoblastic leukaemia, in: 2012 International Conference on Advanced Computer Science Applications and Technologies (ACSAT), Kuala Lumpur, 2012, pp. 330–335.
- [38] S. Mohapatra et al., Fuzzy based blood image segmentation for automated leukemia detection, in: 2011 International Conference on Devices and Communications (ICDeCom), Mesra, 2011, pp. 1–5.

- [39] S. Mohapatra, D. Patra, Automated cell nucleus segmentation and acute leukemia detection in blood microscopic images, in: 2010 IEEE International Conference on Systems in Medicine and Biology (ICSMB), 2010, pp. 49–54.
- [40] S. Gehlot, A. Gupta, R. Gupta, SDCT-AuxNet⁰: DCT augmented stain deconvolutional CNN with auxiliary classifier for cancer diagnosis, *Med. Image Analysis* 61 (2020) 101661.
- [41] A. Gupta, R. Gupta, ALL challenge dataset of ISBI 2019 [dataset], Cancer Imaging Archive (2019). Available from: <https://doi.org/10.7937/tcia.dc64i46r>.
- [42] K. Clark, et al., The cancer imaging archive (TCIA): maintaining and operating a public information repository, *J. Digital Imaging* 26 (2013) 1045–1057.
- [43] A. Gupta et al., Classification of normal vs malignant cells in B-ALL white blood cancer microscopic images, in: IEEE International Symposium on Biomedical Imaging (ISBI)-2019 challenges. <<https://competitions.codalab.org/competitions/20395>>.
- [44] R. Duggal et al., SD-Layer: Stain deconvolutional layer for CNNs in medical microscopic imaging, in: M. Descoteaux, L. Maier-Hein, A. Franz, P. Jannin, D. Collins, S. Duchesne (Eds.), *Medical Image Computing and Computer Assisted Intervention—MICCAI 2017*, Lecture Notes in Computer Science, vol 10435, Springer, Cham.
- [45] Y. Pan et al., Neighbourhood-correction algorithm for classification of normal and malignant cells, in: A. Gupta, R. Gupta (Eds.), *ISBI 2019 C-NMC Challenge: Classification in Cancer Cell Imaging*, Lecture Notes in Bioengineering, Springer, Singapore, 2019.
- [46] J. Prellberg, O. Kramer, Acute lymphoblastic leukaemia classification from microscopic images using convolutional neural network, in: A. Gupta, R. Gupta (Eds.), *ISBI 2019 C-NMC Challenge: Classification in Cancer Cell Imaging*, Lecture Notes in Bioengineering, Springer, Singapore, 2019.
- [47] F. Xiao et al., DeepMEN: Multi-model ensemble network for b-lymphoblast cell classification, in: A. Gupta, R. Gupta (Eds.), *ISBI 2019 C-NMC Challenge: Classification in Cancer Cell Imaging*, Lecture Notes in Bioengineering, Springer, Singapore, 2019.
- [48] T. Shi, et al., Ensemble convolutional neural networks for cell classification in microscopic images, in: A. Gupta, R. Gupta (Eds.), *ISBI 2019 C-NMC Challenge: Classification in Cancer Cell Imaging*, Lecture Notes in Bioengineering, Springer, Singapore, 2019.
- [49] Y. Liu, F. Long, Acute lymphoblastic leukaemia cells image analysis with deep bagging ensemble learning, in: A. Gupta, R. Gupta (Eds.), *ISBI 2019 C-NMC Challenge: Classification in Cancer Cell Imaging*, Lecture Notes in Bioengineering, Springer, Singapore, 2019.
- [50] E. Verma, V. Singh, ISBI challenge 2019: Convolution neural networks for b-all cell classification, in: A. Gupta, R. Gupta (Eds.), *ISBI 2019 C-NMC Challenge: Classification in Cancer Cell Imaging*, Lecture Notes in Bioengineering, Springer, Singapore, 2019.
- [51] X. Xie et al., Multi-streams and multi-features for cell classification, in A. Gupta, R. Gupta (Eds.), *ISBI 2019 C-NMC Challenge: Classification in Cancer Cell Imaging*, Lecture Notes in Bioengineering, Springer, Singapore.

- [52] S. Shah et al., Classification of normal and leukemic blast cells in b-all cancer using a combination of convolutional and recurrent neural networks, in A. Gupta, R. Gupta (Eds.), ISBI 2019 CNMC Challenge: Classification in Cancer Cell Imaging, Lecture Notes in Bioengineering, Springer, Singapore.
- [53] Y. Ding, Y. Yang, Y. Cui, Deep learning for classifying of white blood cancer, In A. Gupta, R. Gupta (Eds.), ISBI 2019 C-NMC Challenge: Classification in Cancer Cell Imaging, Lecture Notes in Bioengineering, Springer, Singapore.
- [54] S. Mourya et al., LeukoNet: DCT-based CNN architecture for the classification of normal vs Leukemic blasts in B-ALL Cancer, ArXiv abs/1810.07961, 2018.

Computational imaging applications in brain and breast cancer

2

Aimilia Gastounioti^{1,2,*}, Saima Rathore^{1,2,*}, Omid Haji Maghsoudi^{1,2,*}, Emily F. Conant², Despina Kontos^{1,2} and Spyridon Bakas^{1,2,3}

¹*Center for Biomedical Image Computing and Analytics, University of Pennsylvania, Philadelphia, PA, United States*

²*Department of Radiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States*

³*Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States*

2.1 Introduction

In the past decade, an increasing body of literature evolved surrounding the development of computational algorithms for computer-aided diagnosis of various diseases. The rapid development of advanced computational algorithms from the domain of machine learning, and particularly the reported performance of algorithms based on deep learning (DL), shows promise for application in the clinical environment to (1) assist clinicians with tedious daily tasks and allow them to focus more on complex or urgent patient management, (2) offer second reads or opinions on tasks that require specialized training, as well as (3) assist in the training and education of new clinical experts. However, the complexity of developing algorithms for these applications, as well as the somewhat limited validation of large and diverse datasets spanning diverse patient populations, has a direct effect in their reproducibility, which is reflected in the lack of adoption by clinical researchers and eventually, the clinical translation. This chapter offers an overview of the state-of-the-art computational techniques, with a particular focus on their application in the field of brain and breast cancer. We specifically focus on major applications that DL algorithms have shown promise in addressing clinical issues, as well as challenges and potential methods to improve the reproducibility of such algorithms in biomedical image analysis of brain and breast cancer patients. The included references should not be considered as an exhaustive

*These authors changed employment before the publication of this book chapter, but made contributions to it solely while employed by the University of Pennsylvania.

literature review, but as studies serving as examples for the points made in this chapter.

2.2 Building upon current clinical standards

2.2.1 Clinical standards

Successful treatment of cancer is challenged by the heterogeneous molecular, phenotypic, and radio-phenotypic nature of malignancies [1–6]. Especially, as most current diagnoses are based on direct tissue sampling, which often is only of one portion of a tumor, such samples can not capture the spectrum of cancer heterogeneity. Furthermore, the current clinical criteria for assessing treatment response, known as the Response Evaluation Criteria In Solid Tumors and the Response Assessment in Neuro-Oncology (also known as the Macdonald criteria), consider two-dimensional measurements of the major axes of a tumor on a subjectively chosen slice, representative of the tumor’s largest extent [7–9]. Although these criteria are widely used, they have limited reproducibility due to their intrinsic subjectivity [10]. Furthermore, the nature of the 2D measurement involved in these criteria would render them appropriate for tumors that grow in a consistent manner across the tumor boundaries, which contradicts the well-known nature of arbitrarily shaped tumors. Acknowledging the limitations of these criteria, and with the intention of improving tumor assessment, the diagnostic Imaging Reporting And Data Systems (IRADS) for (1) breast (i.e., BI-RADS) [11], (2) prostate (i.e., PI-RADS) [12], and (3) lung (i.e., LUNG-RADS) [13], have included imaging characteristics describing the apparent texture (also known as radiomic features) [14]. Such features have shown their promise on offering a macroscopic characterization of biological processes in the tumor microenvironment associated with clinical outcome and the underlying tumor molecular characteristics [15–22].

2.2.2 Tissue segmentation

To accurately assess a region of interest (ROI), for example, the tumor, the task of segmentation is essential, irrespective of whether it refers to the involuntary partitioning of an image that an expert radiologist conceptually performs to assess a clinical image and create a report for a patient, or to the actual delineation of the various structures appearing in a scan. The tedious and time-consuming manual processing steps involved in such a process, as well as the human rater variability that potentially impede not only the further analyses, but also their repeatability and reproducibility, assisted in making apparent the importance of automated approaches for the segmentation task, leading to various computer-aided segmentation methods. The majority of such methods reported in the literature are based on population-derived knowledge and its application to new data,

to *discriminate* ROIs within an image [23,24]. These are characterized as “discriminative methods” and examples beyond DL-based ones include support vector machines (SVMs), boosting, and random forests. The main limitation of these methods is that they are dependent on large, diverse, and well-annotated datasets, as well as in the way they handle ROIs that have not been previously seen/ included in the population data used for their training.

Although neural networks were introduced in the 1940s [25], these have been recently re-introduced with a stacking of multiple layers to process input data, hence providing the depth that attributes their naming “deep learning”. Convolutional neural networks (CNNs) typically include the convolution of the input data with several local kernels followed by a non-linear transformation on each layer, the output of which is sent (optionally down-sampled) on the following stacked layer. This processing increases the effectiveness of the analysis of the assessed input data, while produces inferences invariant to translation as they increase the field of view of each “neuron”. Minimization of a cost function (e.g., Kullback–Leibler divergence [26]) is included during their training and describes the difference between the ground truth labels provided during training and the final predicted labels.

The tumor segmentation problem is typically dealt as a voxel-wise classification problem, which results in spatial inconsistencies in the segmentation result, potentially due to high frequency intensity fluctuations in areas of increased imaging noise. Various strategies have been proposed to improve robustness to imaging noise including cascaded architectures [27], multi-scale image patches [23,27], and building CNNs with fully connected layers comprising convolutions with 12 or 13 kernels [28]. Integration of CNNs with Markov (MRF) or Conditional Random Fields (CRF), have also shown significant promise, implemented as a post-processing step, or as a formulation of neural networks, or even as recurrent neural networks (RNNs) such that both CNNs and CRF-RNNs can be trained with back-propagation algorithms end to end [23,29–31].

2.3 Deep learning applications in brain cancer

The increased use of magnetic resonance imaging (MRI) for brain tumor diagnosis has led to recent studies exploring the associations between MRI-based indices and clinical brain tumor diagnostic markers. Applications of DL in this field have been widely explored, examining the efficacy of MRI in developing diagnostic, prognostic, and predictive biomarkers [32].

2.3.1 Tumor grading

Brain tumors show heterogeneous characteristics, making it very hard to distinguish between different tumor types, such as primary gliomas and metastatic

tumors. From the clinical perspective, it is very important to provide a distinction between different tumor types for the treating physician to come up with an optimal treatment plan. Several computer-aided analytic tools have been proposed in the past with an aim to provide a more objective decision than human, and to possibly serve as a consistent and reproducible brain tumor diagnostic procedure. Ge et al. proposed a novel multi-stream deep CNN architecture for classifying gliomas into low-grade gliomas (LGGs) and high-grade gliomas (HGGs) [33]. Using sensor fusion method on T1, T2 and FLAIR images to enhance the performance through feature aggregation, the authors reported classification accuracy of 90.87% on their test data. Khawaldeh et al. [34] proposed a system for non-invasive detection of healthy brain scans, LGG, and HGG scans, using a modified version of AlexNet CNN. While training their model on whole-brain MRI of 130 subjects belonging to the three classes, they reported classification performance of 91.16%. Sajjad et al. [35] proposed an extensive data augmentation method fused with CNN for multi-grade classification of brain tumors (grade-I to grade-IV) using segmented brain tumor MRI. They used a pretrained VGG-19 CNN architecture for classification using transfer learning on a dataset of 121 subjects and achieved accuracies of 87.38% and 90.67% for data before and after augmentation, respectively.

Another set of DL-based methods dealt with the categorization of tumors into different tumor types. Fatih et al. combined CNN with neutrosophic set—expert maximum fuzzy-sure entropy (NS-CNN) for classification of brain tumors into benign and malignant categories [36]. They used the neutrosophic set—expert maximum fuzzy-sure method for tumor segmentation. These images were then fed to CNN to extract features, which were subsequently used within SVM classification framework. While carrying out the experimental evaluation of their method via five-fold cross-validation on 160 tumors (80 benign and 80 malignant), they achieved classification success rate of 95.62%. To distinguish between different brain tumor types, including glioma, meningioma, and pituitary tumors, Alqudah et al. used CNN under three different configurations: (1) on cropped lesions, (2) on uncropped lesions (i.e., the whole image), and (3) on the segmented lesions. Prospective evaluation of the trained model on a test set resulted in an accuracy of 98.93% for the cropped lesions, 99.00% for uncropped lesions, and 97.62% for the segmented lesions [37].

2.3.2 Survival analysis

Gliomas, in general, tend to show grim prognosis; therefore, prediction of overall survival at the initial presentation of the disease provides valuable information for surgical and treatment planning [38,39]. Prediction of overall survival of gliomas using DL approaches have gained considerable attention in the recent past [40]. In one of the earlier attempts to predict survival using brain MRI, Lao et al. used transfer learning method to calculate DL radiomic features and eventually predict overall survival in glioblastoma [41]. The authors extracted 1403 engineered

features and 98304 deep features from preoperative MRI data of 112 patients. Following the feature selection step, the authors constructed an imaging signature by using the least absolute shrinkage and selection operator Cox regression model. The proposed DL-based radiomic signature achieved a c-index of 0.710 (95%CI: 0.588–0.932) and hazard-ratio of 5.128 ($P < .001$, 95%CI: 2.029–12.960) in segregation of patients into prognostically different groups. In addition to MRI-based DL approaches, authors have also explored the effectiveness of machine learning in integrating genomic and clinical variables for assessment of survival in gliomas [42,43].

Among the recent studies, Nie et al. [44] proposed a two-layered learning approach to predict overall survival of HGGs. At the first layer, the authors applied DL on multimodal preoperative MRIs, including T1-post-contrast, resting-state functional MRI (rs-fMRI), and diffusion tensor imaging (DTI), to derive multiple metric maps. These maps included diffusivity maps, extracted from DTI, and fluctuation amplitude maps and functional connectivity maps, extracted from rs-fMRI. The authors applied a multi-channel architecture of 3D CNNs on the metric maps to extract high-level predictive features from each individual patch. At the second layer, these features along with the tumor-related features (tumor size and histopathological subtype) and demographics (age) were fed into an SVM to classify 56 HGGs into long and short survivors, reporting an accuracy of 90.66%. Han et al. [45] in a recent study, used 55 HGGs from their own institution and 128 HGGs from The Cancer Imaging Archive (TCIA). For each image, they calculated 348 engineered features and 8192 DL features, derived by using a pretrained CNN. Feature selection and Elastic Net-Cox modeling methods were applied to stratify patients into prognostically different groups. The survival analysis segregated the data into different prognostic groups with a log-rank test $P < .001$ and $P = .014$ in local and TCIA patients.

2.3.3 Radiogenomics

Radiogenomics is generally performed to understand how a particular genomic alteration impacts the imaging traits of the tumors [46]. Recently, there has been growing evidence that computational analysis can help in developing MRI-based surrogate indices of tumor molecular characteristics, without the need to perform expensive, advanced genetic testing. Three genetic markers, isocitrate dehydrogenase (IDH) mutation, 1p/19q co-deletion, and oxygen 6-methylguanine-DNA methyltransferase (MGMT) promoter methylation status are of significant interest in gliomas.

2.3.3.1 1p/19q

Recent studies on glioma based on TCIA datasets have uncovered the strong association of 1p/19q co-deletion with the patient outcomes, and several authors have developed non-invasive machine-learning-based methods to predict 1p/19q status [33,47]. Akkus et al. included preoperative T1-Gd and T2 images of a total of

159 LGGs (57 non-deleted and 102 co-deleted) for detection of 1p/19q status [47]. They assessed multiple configurations of a multi-scale CNN architecture, and found the best configuration to provide an accuracy of 87.7%, sensitivity of 93.3%, and specificity of 82.22%. Ge et al. proposed a novel multi-stream deep CNN architecture for detection of 1p/19q co-deletion status [33]. The authors employed sensor fusion method on T1, T2 and FLAIR images to enhance performance through feature aggregation, and used 2D brain slices coupled with 2D image augmentation to mitigate overfitting. Experiments using the proposed method showed 89.39% accuracy in classifying glioma patients with and without 1p/19q co-deletion.

2.3.3.2 *Isocitrate dehydrogenase*

IDH mutation is fundamentally important and serves as a prognostic indicator for gliomas. Non-invasive assessment of IDH mutation by computational imaging signatures could significantly influence therapeutic decision-making. Several MRI-based predictors of IDH have been developed using DL approaches in the recent past. Li et al. proposed a DL-based model, wherein they used a modified CNN architecture having six convolutional layers and a fully connected layer to segment tumors [48]. As an alternative to extracting radiomic features from segmented images, the features were derived by normalizing the output of final convolutional layer. Features discriminative of IDH status were calculated using paired t-test and F-score. The authors reported an AUC of 0.92 compared to 0.86 obtained using traditional machine learning for predicting IDH1 on a dataset of 151 LGGs. Recently, Liang et al. have proposed a novel multimodal 3D DenseNet model to predict *IDH* status using MRI data [49]. The evaluation of the method yielded an accuracy of 84.6% and an AUC of 85.7% on validation data from the International Brain Tumor Segmentation (BraTS) 2017 challenge [40,50–53]. Similarly, Chang et al. employed a residual CNN to predict the IDH mutation status on multi-center data of 406 subjects, including grade II-IV [54]. Their method achieved classification accuracies of 82.8%, 83.0%, and 85.7% on training, validation, and testing cohorts, respectively.

2.3.3.3 *6-methylguanine-DNA methyltransferase*

MGMT status has been increasingly used as a prognostic and predictive biomarker for gliomas [55]. MGMT constrains the repair process of temozolomide-induced therapeutic DNA damage in glioma patients, and thereby strongly correlates with the overall survival. Moreover, there is clinical evidence suggesting that MGMT has strong association with better treatment response to temozolomide [55], thereby making it very important to distinguish between MGMT methylated and unmethylated tumors prior to treatment. Levner et al. was one of the pioneering groups to develop DL-based radiogenomic approaches [56]. They proposed a method to predict MGMT status in de novo glioblastoma patients by extracting features from brain MRIs using S-transform based space-frequency texture analysis and utilizing the features within a neural network framework to

predict the methylation status of a glioblastoma. The author's group reported a classification success rate of 87.7% on 59 patients. Han et al. used bi-directional convolutional RNN to leverage the spatial aspects of 3D MRIs to predict MGMT status of gliomas [57]. The authors reported accuracies of 67% and 62% on validation and test data of glioblastoma patients acquired from TCIA. Korfiatis et al. evaluated the performance of three different residual CNN architectures (ResNet50, ResNet34 and ResNet18) to predict MGMT status on MRIs of 155 subjects [58]. Among the three chosen architectures, the ResNet50 outperformed the others, achieving an accuracy of 94.90%. The corresponding accuracies were 80.72%, and 76.75% on ResNet34 and ResNet18, respectively. Recently, Chang et al. have proved the effectiveness of CNN in detecting IDH1 mutation, MGMT methylation, and 1p/19q co-deletion status on a dataset of 256 subjects acquired from TCIA [59]. They achieved classification accuracies of 94%, 92%, and 83% for IDH1 mutation, 1p/19q co-deletion, and MGMT methylation status.

2.3.4 Pseudoprogression

The independent and added value of DL in developing predictive biomarkers for brain tumors are explored using surrogate measures for pseudoprogression (PSP) and distinguishing from true progression (TRP). The term PSP is defined as the intensity enhancement that brain tumor patients show in certain regions (in the vicinity of the resected tumor) in the months following the completion of chemoradiotherapy, and it initially progresses but usually stabilizes and may regress on follow-up MRI sequences. This mimics TRP but has different characteristics when compared with each other. It is caused by therapy-associated but not tumor growth-associated pathologic changes. It is becoming increasingly important to accurately distinguish PSP and TRP because therapeutic strategies need to be adapted by the treating physicians for each of these entities; for example, additional surgical resection may be pursued for TRP but not for PSP. An incorrect diagnosis of progression (i.e. PSP) may lead to unnecessary surgical resection or erroneous termination of a successful therapy. An increasing attention has been recently paid to this field of research. Jang et al. proposed a hybrid approach coupling a CNN algorithm to a conventional machine learning, long short-term memory (CNN-LSTM) approach, to distinguish between PSP and TRP in glioblastoma [60]. The evaluation of their method on multiinstitutional data of 78 patients yielded an AUC of 0.83. Building upon these successful past efforts of discriminating PSP from TRP, Akbari et al. [61] utilized MRI sequences of 63 glioblastoma patients comprising diffusion-weighted imaging, DSC-MRI and conventional imaging sequences (T1, T2, T2-FLAIR, T1-Gd) to distinguish between PSP and TRP. The authors adapted a CNN model pretrained on ImageNet LSVRC-2010 dataset for quantifying the imaging profile of lesion of interest on the follow-up MRI scans. The DL features were integrated with an SVM classifier; the corresponding accuracy in train-test setting was 87.50% for predicting PSP (AUC = 0.81) and 78.26% for TRP (AUC = 0.87).

2.4 Deep learning applications in breast cancer

Breast cancer is the most common cancer in women worldwide, with an estimated (approximately) 270,000 new cancer cases diagnosed in the United States (U.S.) in 2019 [62]. Breast cancer is also the primary cause of cancer-related deaths in women in developing countries where many cases present at advanced stages, and is the second most common cause of death in women in developed countries where screening is more common. Most developed health care systems have implemented population breast cancer screening programs based on evidence from randomized trials and real-world screening that it reduces breast cancer mortality. Breast cancer screening involves visual assessment of digital mammograms or digital breast tomosynthesis (DBT) images by readers to identify suspicious abnormalities that warrant further investigation or, recall from screening, to rule in or rule out breast cancer. As visual mammographic image interpretation is subjective, there may be both a high and variable false-positive and false-negative rates. In addition, as mammographic imaging is anatomic rather than a physiologic imaging modality using contrast (i.e., MRI), malignant lesions may be hidden or “masked” by normal breast tissue that is similar in optical density to some breast cancers.

Much of the effort to improve breast cancer screening outcomes has focused on intensifying screening, for example, double-reading instead of single-reading and more frequent or supplemental screening (with breast ultrasound or MRI), which entail increased resources and often comes at a cost of higher false-positive rates. Furthermore, personalized breast cancer screening regimens tailored to an individual’s breast cancer risk are increasingly being advocated. For example, it has been widely established that increased breast density not only masks the detection of some breast cancers but also is an independent risk factor for developing breast cancer [63], and the reporting of mammographic density is mandated in more than 70% of the U.S. to identify the women who may benefit from supplemental screening [64]. This complex landscape of breast cancer screening offers several opportunities for improvements via DL, with research currently focusing primarily on three aspects: (1) increasing accuracy in breast density evaluation [65–69] and breast cancer risk assessment [70–72], (2) improving performance in breast cancer diagnosis [73–76], and (3) enhancing efficacy in breast cancer clinical practice [77–79].

2.4.1 Increasing accuracy in breast cancer risk assessment

Gastouniotti et al. [71,80] were among the first studies to explore the potential of CNNs in breast cancer risk assessment in analyzing mammographic images. The authors proposed a hybrid computational approach that employs CNNs to optimally fuse parenchymal complexity measurements generated by conventional texture analysis into discriminative meta-features relevant for breast cancer risk

prediction. This study showed that CNNs can capture sparse, subtle, yet relevant interactions between localized patterns present in texture feature maps derived from mammographic images, thereby improving the breast cancer risk prediction performance of conventional parenchymal texture analysis (AUC = 0.90 vs AUC = 0.79, $P < .05$).

In more recent studies [70,72], CNNs have been used to train DL models fed directly with raw digital mammograms toward predicting the risk of future breast cancer development. Dembrower et al. [72] found that the DL risk score could more accurately help predict which women were at risk for future breast cancer compared with age-adjusted dense area (odds ratio: 1.6 and 1.3, respectively). In addition, the false-negative rate was lower for the deep neural network than for the age-adjusted dense area (31% vs 36%; $P = .006$), and the difference was most pronounced for women later diagnosed with more aggressive cancers. Yala et al. [70] showed that DL can also improve 5-year risk prediction when compared to an established breast cancer risk model that includes subjective breast density categories (Tyrer-Cuzick model, version 8 [TC]). When the DL model was used, 31% of women were identified as having high risk, compared with 18% when the TC model was used. Interestingly, the image-only DL model also outperformed TC and it provided accurate risk assessment when traditional risk factor information was unavailable, which can be particularly beneficial to patients who do not know their family history of breast or ovarian cancer.

2.4.2 Reproducible breast density assessment for improved breast cancer risk prediction

The most commonly used method to assess breast density in the clinical setting is the visual grading of breast density by the interpreting radiologist into one of four categories outlined by the American College of Radiology BI-RADS [81]. However, there is a well-established and large degree of inter- and intra-reader variability in the assignment of breast density, particularly among less-experienced readers [82]. In one of the earliest studies, Kallenberg et al. [68] introduced a fully-automated method based on learning a feature hierarchy from unlabeled data followed by a classifier to estimate percent density (PD). The results of this study showed a very strong positive correlation between DL-based PD scores and manual PD scores ($r = 0.85$), while in a case-control evaluation setting, the DL-based PD scores yielded an AUC of 0.59, which is competitive to reported AUCs in the literature on similar populations. In another study [69], the authors used a dataset of over 200,000 screening mammography exams, each containing at least four images corresponding to the standard four views used in screening mammography, to train a DL model for the task of BI-RADS breast density classification. The level of agreement between the trained DL classifier and the BI-RADS density classes in the data was found to be similar to that between the human experts and the BI-RADS density classes in the data, as well

as among the human experts themselves. Mohamed et al. [67] applied transfer learning to develop a DL approach based on the AlexNet architecture to classify digital mammography images into BI-RADS density classes. When applied to an unseen set of 1850 images, the proposed approach achieved an AUC of 0.94.

More recently, Lehman et al. [65] have presented another DL model based on the ResNet architecture for the same classification task. In an unseen test set of 8677 mammograms, the DL model showed good agreement in BI-RADS density classification ($k = 0.67$) with radiologists and with radiologists in consensus in a reader study set ($k = 0.78$). Lately, the same DL model was also implemented at a dedicated breast imaging practice, where it showed had a high clinical acceptance rate among both academic (94.9%) and community (90.7%) radiologists and reduced the proportion of mammograms assessed as dense from 47% to 41% ($P < .001$) [66].

2.4.3 Improving performance in breast cancer diagnosis

The Digital Mammography DREAM challenge [73], held between 2016 and 2017, encouraged the development of DL models to classify screening mammograms according to whether cancer was present or not. A large set of digital mammography images of over 640,000 images from approximately 80,000 women was used by several teams to train DL models, resulting in the development of many novel approaches to improve cancer detection. The most successful teams achieved a sensitivity of up to 87% compared to the radiologist sensitivity of 88% on the same dataset.

DL has also been used to compare breast cancer detection performance of radiologists reading mammographic examinations unaided versus supported by an artificial intelligence (AI) system [75]. Using screening digital mammographic examinations from 240 women and readings from 14 radiologists, the authors showed that radiologists improved their cancer detection at mammography when using an AI system for support (unaided, AUC = 0.87; supported by AI, AUC = 0.89; $P = .002$), without requiring additional reading time (unaided, 146 seconds; supported by AI, 149 seconds; $P = .15$). In another study, a commercial DL model for identifying breast cancer in screening mammograms was evaluated using two large datasets from the UK and the US [74]. The authors showed an absolute reduction of 5.7% and 1.2% in false positives, and 9.4% and 2.7% in false negatives (U.S. and UK, respectively). In an independent study of six radiologists, the same DL model also outperformed all the human readers (AUC was greater than the AUC for the average radiologist by an absolute margin of 11.5%). Moreover, in a simulation in which the DL model participated in the double-reading process that is used in the UK, the authors reported non-inferior performance as well as reduction of the workload of the second reader by 88%. Most recently, a retrospective multi-reader study has shown that a DL algorithm for detecting breast cancer can be used as an effective diagnostic support tool for radiologists in mammography interpretation [76]. Specifically, a DL model based

on the ResNet architecture showed an AUC of 0.94–0.97 in multiple validation datasets collected from five institutions in South Korea, the US, and the UK. It also showed significantly better performance than 14 radiologists in 320 independent mammograms, resulting in a significant improvement in radiologists' DL-aided diagnostic performance.

2.4.4 Enhancing efficacy in breast cancer clinical practice

DL has also been found to be able to reduce the workload of cases to be read by a radiologist by excluding exams with a very low likelihood of cancer. Using a total of 2652 digital mammography exams and interpretations by 101 radiologists, Rodriguez-Ruiz et al. [78] showed that there is potential to use DL to reduce the mammogram reading workload by 17%, while only excluding 1% of true-positive exams. Moreover, the exclusion of exams with the lowest likelihood of cancer in screening might not change radiologists' breast cancer detection performance. Another DL model evaluated in screening exams from over 7000 women was capable of significantly decreasing the number of mammograms that a radiologist needed to read by 34% in a theoretical diagnostic setting (15% cancer prevalence) and by nearly 91% for a setting with low prevalence, as would be expected in a screening facility [77]. Finally, the recently published reader study by Conant et al. [79] is the first to evaluate the potential of DL to enhance efficiency in DBT reading. By comparing the performance of 24 radiologists (13 of whom were breast subspecialists) reading 260 DBT examinations both with and without AI, this study showed that the concurrent use of an accurate DBT AI system can improve cancer detection accuracy while reducing the reading time by 57% (reading time 64.1 seconds without versus 30.4 seconds with AI).

2.5 Conclusion

In conclusion, we note that advanced computational algorithms, particularly those based on DL, hold a great promise for applications in brain and breast cancer, as demonstrated by the example studies cited in this chapter. Considering the current several reported successes and promise shown in the literature within breast and brain cancer applications, there is an associated mounting interest by both the research and the medical community in terms of both developing and potentially translating these algorithms in a clinical environment. However, the algorithmic complexity, as well as the somewhat limited size and diversity of datasets used for algorithmic validation, is hindering the clinical adoption. To facilitate this clinical translation, there is a clear need for large multiinstitutional datasets spanning diverse patient populations, with corresponding annotations while following consistent protocols for quality assurance. These efforts can thereby lead to benchmarking datasets used for comparative studies deciding on the relative

performance and ranking of algorithms, such as the BraTS dataset [40,50–53]. Currently, the paradigm for collecting knowledge from such datasets is based on collaborations pooling together multiinstitutional datasets. However, driven by the commonly accepted fact that this paradigm faces various legal, privacy, and data ownership concerns, multiple initiatives of alternatives paradigms based on distributed learning have recently been introduced and appreciated on example cases, showing algorithmic performance similar to the one when pooled datasets are used [83–85]. We look forward to this current research field and its potential impact on the next generation of computational algorithms for medical applications toward ultimately leading to a new era of computational precision diagnostics.

Acknowledgments

The work was partially supported by the Susan G. Komen for the Cure® Breast Cancer Foundation [PDF17479714] and by the National Institutes of Health (NCI: U01CA242871).

References

- [1] S.P. Nicolou, *Gauging Heterogeneity in Primary Versus Recurrent Glioblastoma*, Oxford University Press, 2015.
- [2] J.P. O'Connor, C.J. Rose, J.C. Waterton, R.A. Carano, G.J. Parker, A. Jackson, Imaging intratumor heterogeneity: role in therapy response, resistance, and clinical outcome, *Clin. Cancer Res.* 21 (2) (2015) 249–257.
- [3] R. Fisher, L. Pusztai, C. Swanton, Cancer heterogeneity: implications for targeted therapeutics, *Br. J. Cancer* 108 (3) (2013) 479–485.
- [4] D. Zardavas, A. Irrthum, C. Swanton, M. Piccart, Clinical management of breast cancer heterogeneity, *Nat. Rev. Clin. Oncol.* 12 (7) (2015) 381.
- [5] I. Dago-Jack, A.T. Shaw, Tumour heterogeneity and resistance to cancer therapies, *Nat. Rev. Clin. Oncol.* 15 (2) (2018) 81.
- [6] N. McGranahan, C. Swanton, Biological and therapeutic impact of intratumor heterogeneity in cancer evolution, *Cancer Cell* 27 (1) (2015) 15–26.
- [7] P. Therasse, S.G. Arbuck, E.A. Eisenhauer, J. Wanders, R.S. Kaplan, L. Rubinstein, et al., New guidelines to evaluate the response to treatment in solid tumors, *J. Natl Cancer Inst.* 92 (3) (2000) 205–216.
- [8] P.Y. Wen, D.R. Macdonald, D.A. Reardon, T.F. Cloughesy, A.G. Sorensen, E. Galanis, et al., Updated response assessment criteria for high-grade gliomas: response assessment in neuro-oncology working group, *J. Clin. Oncol.* 28 (11) (2010) 1963–1972.
- [9] D.R. Macdonald, T.L. Cascino, S.C. Schold Jr, J.G. Cairncross, Response criteria for phase II studies of supratentorial malignant glioma, *J. Clin. Oncol.* 8 (7) (1990) 1277–1280.
- [10] M. Vos, B. Uitdehaag, F. Barkhof, J. Heimans, H. Baayen, W. Boogerd, et al., Interobserver variability in the radiological assessment of response to chemotherapy in glioma, *Neurology.* 60 (5) (2003) 826–830.
- [11] E.S. Burnside, E.A. Sickles, L.W. Bassett, D.L. Rubin, C.H. Lee, D.M. Ikeda, et al., The ACR BI-RADS® experience: learning from history, *J. Am. Coll. Radiol.* 6 (12) (2009) 851–860.

- [12] J.G. Bomers, J.O. Barentsz, Standardization of multiparametric prostate MR imaging using PI-RADS, *BioMed. Res. Int.* 2014 (2014).
- [13] E.A. Kazerooni, M.R. Armstrong, J.K. Amorosa, D. Hernandez, L.A. Liebscher, H. Nath, et al., ACR CT accreditation program and the lung cancer screening program designation, *J. Am. Coll. Radiol.* 12 (1) (2015) 38–42.
- [14] A. Zwanenburg, M. Vallières, M.A. Abdalah, H.J. Aerts, V. Andrearczyk, A. Apte, et al., The Image Biomarker Standardization Initiative: standardized quantitative radiomics for high-throughput image-based phenotyping, *Radiology.* (2020) 191145.
- [15] C.C. Jaffe *Imaging and Genomics: Is There a Synergy?:* Radiological Society of North America, Inc.; 2012.
- [16] M. Carol Proud, Radiogenomics: the promise of personalized treatment in radiation oncology? *Clin. J. Oncol. Nurs.* 18 (2) (2014) 185.
- [17] B.S. Rosenstein, C.M. West, S.M. Bentzen, J. Alsner, C.N. Andreassen, D. Azria, et al., Radiogenomics: radiobiology enters the era of big data and team science, *Int. J. Radiat. Oncol. Biol. Phys.* 89 (4) (2014) 709–713.
- [18] A.M. Rutman, M.D. Kuo, Radiogenomics: creating a link between molecular diagnostics and diagnostic imaging, *Eur. J. Radiol.* 70 (2) (2009) 232–241.
- [19] H.J. Aerts, The potential of radiomic-based phenotyping in precision medicine: a review, *JAMA Oncol.* 2 (12) (2016) 1636–1642.
- [20] M.A. Mazurowski, J. Zhang, L.J. Grimm, S.C. Yoon, J.I. Silber, Radiogenomic analysis of breast cancer: luminal B molecular subtype is associated with enhancement dynamics at MR imaging, *Radiology.* 273 (2) (2014) 365–372.
- [21] B.J. Gill, D.J. Pisapia, H.R. Malone, H. Goldstein, L. Lei, A. Sonabend, et al., MRI-localized biopsies reveal subtype-specific differences in molecular and cellular composition at the margins of glioblastoma, *Proc. Natl. Acad. Sci.* 111 (34) (2014) 12550–12555.
- [22] H. Li, Y. Zhu, E.S. Burnside, K. Drukker, K.A. Hoadley, C. Fan, et al., MR imaging radiomics signatures for predicting the risk of breast cancer recurrence as given by research versions of MammaPrint, Oncotype DX, and PAM50 gene assays, *Radiology.* 281 (2) (2016) 382–391.
- [23] K. Kamnitsas, C. Ledig, V.F. Newcombe, J.P. Simpson, A.D. Kane, D.K. Menon, et al., Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation, *Med. Image Anal.* 36 (2017) 61–78.
- [24] V. Rao, M.S. Sarabi, A. Jaiswal, Brain tumor segmentation with deep learning, *Multimodal Brain Tumor Segmentation Challenge (BraTS)* (2015) 56–59.
- [25] W.S. McCulloch, W. Pitts, A logical calculus of the ideas immanent in nervous activity, *Bull. Math. Biophys* 5 (4) (1943) 115–133.
- [26] S. Kullback, R.A. Leibler, On information and sufficiency, *Ann. Math. Stat.* 22 (1) (1951) 79–86.
- [27] M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, et al., Brain tumor segmentation with deep neural networks, *Med. Image Anal.* 35 (2017) 18–31.
- [28] J. Long, E. Shelhamer, T. Darrell (Eds.), Fully convolutional networks for semantic segmentation, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015.
- [29] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A.L. Yuille, Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs, *IEEE Trans. Pattern Anal. Mach. Intell.* 40 (4) (2017) 834–848.

- [30] Z. Liu, X. Li, P. Luo, C.-C. Loy, X. Tang (Eds.), Semantic image segmentation via deep parsing network, in: Proceedings of the IEEE International Conference on Computer Vision, 2015.
- [31] S. Zheng, S. Jayasumana, B. Romera-Paredes, V. Vineet, Z. Su, D. Du et al. (Eds.), Conditional random fields as recurrent neural networks, in: Proceedings of the IEEE International Conference on Computer Vision, 2015.
- [32] M.M. Shaver, P.A. Kohanteb, C. Chiou, M.D. Bardis, C. Chantaduly, D. Bota, et al., Optimizing neuro-oncology imaging: a review of deep learning approaches for glioma imaging, *Cancers*. 11 (6) (2019) 829.
- [33] C. Ge, I.Y. Gu, A.S. Jakola, J. Yang, Deep learning and multi-sensor fusion for glioma classification using multistream 2D convolutional networks, in: Conference Proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society IEEE Engineering in Medicine and Biology Society Annual Conference. 2018 Jul;2018:5894–5897. PubMed PMID: 30441677. Epub 2018/11/18. eng.
- [34] S. Khawaldeh, U. Pervaiz, A. Rafiq, R.S. Alkhawaldeh, Noninvasive grading of glioma tumor using magnetic resonance imaging with convolutional neural networks, *Appl. Sci.* 8 (1) (2018) 27.
- [35] M. Sajjad, S. Khan, K. Muhammad, W. Wu, A. Ullah, S.W. Baik, Multi-grade brain tumor classification using deep CNN with extensive data augmentation, *J. Comput. Sci.* 30 (2019) 174–182.
- [36] F. Özyurt, E. Sert, E. Avci, E. Dogantekin, Brain tumor detection based on convolutional neural network with neutrosophic expert maximum fuzzy sure entropy, *Measurement* 147 (2019) 106830.
- [37] A.M. Alqudah, H. Alquraan, I.A. Qasmieh, A. Alqudah, W. Al-Sharu, Brain tumor classification using deep learning technique—a comparison between cropped, uncropped, and segmented lesion images with different sizes, *Int. J. Adv. Trends Comp. Sci. Eng.* 86 (2019) 3684–3691. Crossref Web. 2019:3684–91.
- [38] L. Macyszyn, H. Akbari, J.M. Pisapia, X. Da, M. Attiah, V. Pigrish, et al., Imaging patterns predict patient survival and molecular subtype in glioblastoma via machine learning techniques, *Neuro-oncology* 18 (3) (2015) 417–425.
- [39] S. Bakas, G. Shukla, H. Akbari, G. Erus, A. Sotiras, S. Rathore, et al. (Eds.), Integrative radiomic analysis for pre-surgical prognostic stratification of glioblastoma patients: from advanced to basic MRI protocols, in: *Medical Imaging 2020: Image-Guided Procedures, Robotic Interventions, and Modeling*: International Society for Optics and Photonics, 2020.
- [40] S. Bakas, M. Reyes, A. Jakab, S. Bauer, M. Rempfler, A. Crimi, et al. Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge. arXiv preprint arXiv:181102629. 2018.
- [41] J. Lao, Y. Chen, Z.C. Li, Q. Li, J. Zhang, J. Liu, et al., A deep learning-based radiomics model for prediction of survival in glioblastoma multiforme, *Sci. Rep.* 7 (1) (2017) 10353. PubMed PMID: 28871110. Pubmed Central PMCID: PMC5583361. Epub 2017/09/06. eng.
- [42] J. Hao, Y. Kim, T. Mallavarapu, J.H. Oh, M. Kang, Interpretable deep neural network for cancer survival analysis by integrating genomic and clinical data, *BMC Med. Genomics* 12 (10) (2019) 189.

- [43] G. Shukla, S. Bakas, S. Rathore, H. Akbari, A. Sotiras, C. Davatzikos, Radiomic features from multi-institutional glioblastoma MRI offer additive prognostic value to clinical and genomic markers: focus on TCGA-GBM collection, *Int. J. Radiat. Oncol. Biol. Phys.* 99 (2) (2017) E107–E108.
- [44] D. Nie, J. Lu, H. Zhang, E. Adeli, J. Wang, Z. Yu, et al., Multi-channel 3D deep feature learning for survival time prediction of brain tumor patients using multi-modal neuroimages, *Sci. Rep.* 9 (1) (2019) 1103.
- [45] W. Han, L. Qin, C. Bay, X. Chen, K.-H. Yu, N. Miskin, et al., Deep transfer learning and radiomics feature prediction of survival of patients with high-grade gliomas, *Am. J. Neuroradiol.* (2019).
- [46] A. Fathi Kazerooni, S. Bakas, H. Saligheh Rad, C. Davatzikos, Imaging signatures of glioblastoma molecular characteristics: a radiogenomics review, *J. Magn. Reson. Imaging* (2019).
- [47] Z. Akkus, I. Ali, J. Sedlar, J.P. Agrawal, I.F. Parney, C. Giannini, et al., Predicting Deletion of Chromosomal Arms 1p/19q in Low-Grade Gliomas from MR Images Using Machine Intelligence, *J. Digital Imaging* 30 (4) (2017) 469–476. PubMed PMID: 28600641. Pubmed Central PMCID: PMC5537096. Epub 2017/06/11. eng.
- [48] Z. Li, Y. Wang, J. Yu, Y. Guo, W. Cao, Deep Learning based Radiomics (DLR) and its usage in noninvasive IDH1 prediction for low grade glioma, *Sci. Rep.* 7 (1) (2017) 5467. PubMed PMID: 28710497. Pubmed Central PMCID: PMC5511238. Epub 2017/07/16. eng.
- [49] S. Liang, R. Zhang, D. Liang, T. Song, T. Ai, C. Xia, et al., Multimodal 3D DenseNet for IDH genotype prediction in gliomas, *Genes* 9 (8) (2018). PubMed PMID: 30061525. Pubmed Central PMCID: PMC6115744. Epub 2018/08/01. eng.
- [50] B.H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, et al., The multimodal brain tumor image segmentation benchmark (BRATS), *IEEE Trans. Med. Imaging* 34 (10) (2014) 1993–2024.
- [51] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. Kirby, et al., Segmentation labels and radiomic features for the pre-operative scans of the TCGA-LGG collection, *Cancer Imaging Archive* (2017) 286.
- [52] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. Kirby, et al., Segmentation labels and radiomic features for the pre-operative scans of the TCGA-GBM collection, *Cancer Imaging Archive* (2017).
- [53] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J.S. Kirby, et al., Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features, *Sci. Data* 4 (2017) 170117.
- [54] K. Chang, H.X. Bai, H. Zhou, C. Su, W.L. Bi, E. Agbodza, et al., Residual convolutional neural network for the determination of idh status in low- and high-grade gliomas from MR imaging, *Clin. Cancer Res.* 24 (5) (2018) 1073–1081. PubMed PMID: 29167275. Pubmed Central PMCID: PMC6051535. Epub 2017/11/24. eng.
- [55] M.E. Hegi, A.C. Diserens, T. Gorlia, M.F. Hamou, N. de Tribolet, M. Weller, et al., MGMT gene silencing and benefit from temozolomide in glioblastoma, *N. Engl. J. Med.* 352 (10) (2005) 997–1003. PubMed PMID: 15758010. Epub 2005/03/11. eng.
- [56] I. Levner, S. Drabycz, G. Roldan, P. De Robles, J.G. Cairncross, R. Mitchell, Predicting MGMT methylation status of glioblastomas from MRI texture, in: *Medical Image Computing and Computer-Assisted Intervention: MICCAI International Conference on*

- Medical Image Computing and Computer-Assisted Intervention, 2009;12(Pt 2):522–530. PubMed PMID: [20426152](#). Epub 2009/01/01. eng.
- [57] L. Han, M.R. Kamdar, MRI to MGMT: predicting methylation status in glioblastoma patients using convolutional recurrent neural networks, in: Pacific Symposium on Biocomputing Pacific Symposium on Biocomputing, 2018;23:331–342. PubMed PMID: [29218894](#). Pubmed Central PMCID: [PMC5728677](#). Epub 2017/12/09. eng.
- [58] P. Korfiatis, T.L. Kline, D.H. Lachance, I.F. Parney, J.C. Buckner, B.J. Erickson, Residual deep convolutional neural network predicts MGMT methylation status, *J. Digital Imaging* 30 (5) (2017) 622–628. PubMed PMID: [28785873](#). Pubmed Central PMCID: [PMC5603430](#). Epub 2017/08/09. eng.
- [59] P. Chang, J. Grinband, B.D. Weinberg, M. Bardis, M. Khy, G. Cadena, et al., Deep-learning convolutional neural networks accurately classify genetic mutations in gliomas, *AJNR Am. J. Neuroradiol.* 39 (7) (2018) 1201–1207. PubMed PMID: [29748206](#). Pubmed Central PMCID: [PMC6880932](#). Epub 2018/05/12. eng.
- [60] B.S. Jang, S.H. Jeon, I.H. Kim, I.A. Kim, Prediction of pseudoprogression versus progression using machine learning algorithm in glioblastoma, *Sci. Rep.* 8 (1) (2018) 12516. PubMed PMID: [30131513](#). Pubmed Central PMCID: [PMC6104063](#). Epub 2018/08/23. eng.
- [61] H. Akbari, S. Rathore, S. Bakas, M.P. Nasrallah, G. Shukla, E. Mamourian, et al., Histopathology-validated machine learning radiographic biomarker for noninvasive discrimination between true progression and pseudo-progression in glioblastoma, *Cancer* (2020).
- [62] Cancer facts and figures 2019. Atlanta, GA: American Cancer Society, 2019. <<https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/cancer-facts-figures-2019.html>> (accessed 3.01.20).
- [63] V.A. McCormack, I. dos Santos Silva, Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis, *Cancer Epidemiol. Biomarkers Prev.* 15 (6) (2006) 1159–1169. PubMed PMID: [16775176](#).
- [64] Are You Dense Advocacy. D.E.N.S.E. State Efforts. <<http://areyoudenseadvocacy.org/>> (accessed 8.01.20).
- [65] C.D. Lehman, A. Yala, T. Schuster, B. Dontchos, M. Bahl, K. Swanson, et al., Mammographic breast density assessment using deep learning: clinical implementation, *Radiology* 290 (1) (2018) 52–58.
- [66] B.N. Dontchos, A. Yala, R. Barzilay, J. Xiang, C.D. Lehman, External validation of a deep learning model for predicting mammographic breast density in routine clinical practice, *Acad. Radiol.* (2020).
- [67] A.A. Mohamed, W.A. Berg, H. Peng, Y. Luo, R.C. Jankowitz, S. Wu, A deep learning method for classifying mammographic breast density categories, *Med. Phys.* 45 (1) (2018) 314–321.
- [68] M. Kallenberg, K. Petersen, M. Nielsen, A. Ng, P. Diao, C. Igel, et al., Unsupervised deep learning applied to breast density segmentation and mammographic risk scoring, *IEEE Trans. Med. Imaging* 35 (5) (2016) 1322–1331.
- [69] N. Wu, K.J. Geras, Y. Shen, J. Su, S.G. Kim, E. Kim, et al., editors. Breast density classification with deep convolutional neural networks, in: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2018.

- [70] A. Yala, C. Lehman, T. Schuster, T. Portnoi, R. Barzilay, A deep learning mammography-based model for improved breast cancer risk prediction, *Radiology*. (2019) 182716.
- [71] A. Gastouniotti, A. Oustimov, M.-K. Hsieh, L. Pantalone, E.F. Conant, D. Kontos, Using convolutional neural networks for enhanced capture of breast parenchymal complexity patterns associated with breast cancer risk, *Acad. Radiol.* (2018).
- [72] K. Dombrower, Y. Liu, H. Azizpour, M. Eklund, K. Smith, P. Lindholm, et al., Comparison of a deep learning risk score and standard mammographic density score for breast cancer risk prediction, *Radiology*. (2019) 190872.
- [73] The digital mammography DREAM challenge. <<https://www.synapse.org/#!Synapse:syn4224222/wiki/401743>> (accessed 8.01.20).
- [74] S.M. McKinney, M. Sieniek, V. Godbole, J. Godwin, N. Antropova, H. Ashrafian, et al., International evaluation of an AI system for breast cancer screening, *Nature*. 577 (7788) (2020) 89–94.
- [75] A. Rodríguez-Ruiz, E. Krupinski, J.-J. Mordang, K. Schilling, S.H. Heywang-Köbrunner, I. Sechopoulos, et al., Detection of breast cancer with mammography: effect of an artificial intelligence support system, *Radiology*. 290 (2) (2018) 305–314.
- [76] H.-E. Kim, H.H. Kim, B.-K. Han, K.H. Kim, K. Han, H. Nam, et al., Changes in cancer detection and false-positive recall in mammography using artificial intelligence: a retrospective, multireader study, *Lancet Digital Health* 2 (3) (2020) e138–e148.
- [77] T. Kyono, F.J. Gilbert, M. van der Schaar, Improving workflow efficiency for mammography using machine learning, *J. Am. Coll. Radiol.* (2019).
- [78] A. Rodriguez-Ruiz, K. Lång, A. Gubern-Merida, J. Teuwen, M. Broeders, G. Gennaro, et al., Can we reduce the workload of mammographic screening by automatic identification of normal exams with artificial intelligence? A feasibility study, *Eur. Radiol.* (2019) 1–8.
- [79] E.F. Conant, A.Y. Toledano, S. Periaswamy, S.V. Fotin, J. Go, J.E. Boatsman, et al., Improving accuracy and efficiency with concurrent use of artificial intelligence for digital breast tomosynthesis, *Radiol. Artif. Intell.* 1 (4) (2019) e180096.
- [80] Convolutional neural network approach for enhanced capture of breast parenchymal complexity patterns associated with breast cancer risk, in: A. Oustimov, A. Gastouniotti, M.-K. Hsieh, L. Pantalone, E.F. Conant, D. Kontos (Eds.), *SPIE Medical Imaging, International Society for Optics and Photonics*, 2017.
- [81] C.J. D’Orsi, *ACR BI-RADS Atlas: Breast Imaging Reporting and Data System*, American College of Radiology, 2013.
- [82] B.L. Sprague, E.F. Conant, T. Onega, M.P. Garcia, E.F. Beaber, S.D. Herschorn, et al., Variation in mammographic breast density assessments among radiologists in clinical practice: a multicenter observational study, *Ann. Intern. Med.* 165 (7) (2016) 457–464.
- [83] N. Rieke, J. Hancox, W. Li, F. Milletari, H. Roth, S. Albarqouni, et al., The future of digital health with federated learning. *arXiv preprint arXiv:200308119*. 2020.
- [84] Multi-institutional deep learning modeling without sharing patient data: A feasibility study on brain tumor segmentation, in: M.J. Sheller, G.A. Reina, B. Edwards, J. Martin, S. Bakas (Eds.), *International MICCAI Brainlesion Workshop*, Springer, 2018.
- [85] K. Chang, N. Balachandar, C. Lam, D. Yi, J. Brown, A. Beers, et al., Distributed deep learning networks among institutions for medical imaging, *J. Am. Med. Inform. Assoc.* 25 (8) (2018) 945–954.

This page intentionally left blank

Deep neural networks and advanced computer vision algorithms in the early diagnosis of skin diseases

**Joanna Jaworek-Korjakowska¹, Moi Hoon Yap², Debotosh Bhattacharjee³,
Pawel Kleczek¹, Andrzej Brodzicki¹ and Marek Gorgon¹**

¹*Department of Automatic Control and Robotics, AGH University of Science and Technology,
Krakow, Lesser Poland, Poland*

²*Visual Computing Lab, Manchester Metropolitan University, Manchester, North West,
United Kingdom*

³*Department of Computer Science and Engineering, Jadavpur University, Kolkata, West Bengal,
India*

In this chapter, we summarize the state of the art of the computerized analysis of dermoscopy images with the use of deep neural networks and advanced machine learning algorithms. This chapter begins by discussing the medical and technical backgrounds of the research problem and then presents details of state-of-the-art solutions using deep neural networks to solve problems including image enhancement, segmentation, pattern recognition, and classification.

In this chapter, we not only showcase recent advances in computer vision and deep learning in dermoscopy image analysis but also explore future directions for this exciting subfield of medical image analysis.

3.1 Introduction and motivation for the early diagnosis of melanoma

Malignant melanoma is one of the most deadly forms of skin cancer and furthermore the most advancing cancer among many white-skinned populations. Incidence and mortality rates caused by cutaneous melanoma, which is the most aggressive kind of skin cancer, have significantly increased during the past few decades among Caucasian populations worldwide [1]. National health services report that melanoma is currently responsible for nearly 70% of all skin cancer-related deaths in the United States and in Australia [2,3].

Two of the most crucial factors suspected to cause melanoma are short but intense sun exposure and unhealthy living habits including smoking and high alcohol consumption. Owing to a constant depletion of ozone layer in stratosphere (resulting in higher exposure to UV radiation), unhealthy nutrition habits as well as poor diet, skin melanoma is likely to become one of the most common malignant neoplasms, with incidence rate even 2–10 times higher than nowadays in the future [1,4].

Moreover, no effective treatment of melanoma in advanced stages has been developed so far. However, early and small melanomas (in the in situ stage) are curable in about 99%. Only a simple excision is needed to close treatment phase (as discussed in Section 3.3) [4]. Therefore, the early diagnosis of melanoma and distinguishing melanoma from other types of skin melanocytic lesions have become extremely important issue (Fig. 3.1).

Based on these data, both education, prevention, early diagnosis, and sophisticated treatment of melanoma are the main goals of modern dermatology. There is a high demand for developing computer-aided diagnostic systems (CADs) facilitating the early detection of melanoma, which would lower its misdiagnosis rate. It has been observed that the accuracy of expert dermatologists in diagnosing melanoma is estimated to be 75%–84% which is partly caused by the subjectivity of the diagnostic judgments [5].

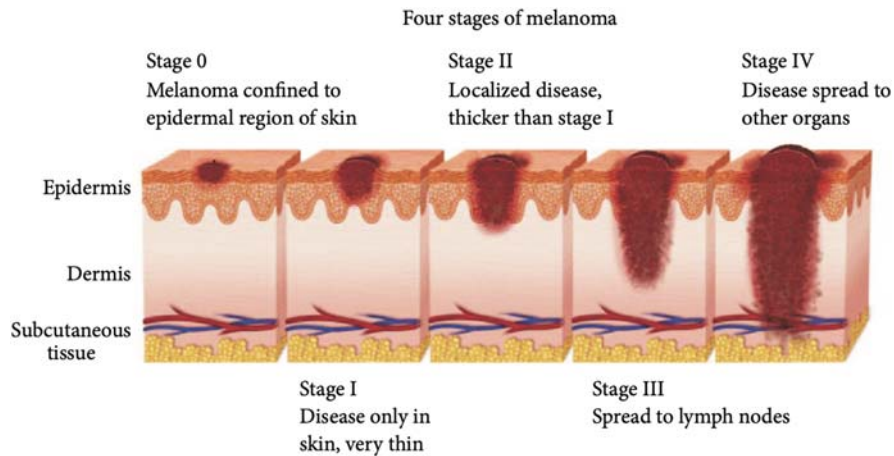
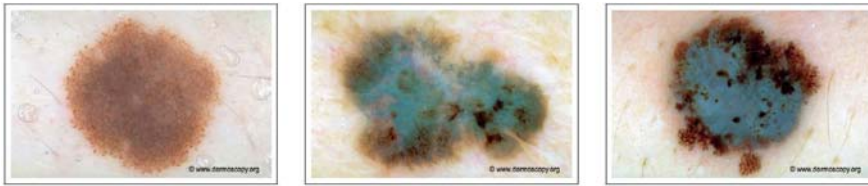


FIGURE 3.1

Four stages of melanocytic skin lesion. Stage 0 which is called melanoma in situ: only in epidermis which is the outer layer of the skin contains melanoma cells, In Stages I and II melanoma cells can be found in the layer directly under the epidermis with no sign that it has spread to lymph nodes or other parts of the body while stages III and IV indicate that the cells have spread to other organs.

**FIGURE 3.2**

Examples of dermoscopic skin mole images in different stages of advancement.

Furthermore, the analysis of dermoscopy images is an extremely difficult task due to the low contrast between the healthy background and skin lesion, variegated coloring inside of the region, irregular borders, and numerous artifacts (Fig. 3.2). There is, therefore, a need to develop automatic image analysis methods based on computer vision and artificial intelligence to enhance the diagnosis of skin melanocytic lesions, which will provide accurate, reliable, and reproducible results. During the last few years, a rapid development in the fields of digital dermatology and artificial intelligence has been observed worldwide. Methods of computer vision and machine learning allow not only for automation of different tasks in various fields, but also for obtaining unbiased and repeatable results. In this chapter we present the most important and groundbreaking works in areas including skin lesion segmentation, melanocytic lesion classification, local pattern segmentation, as well as hardware implementations, that have been published during the last few years.

To present the aforementioned research topics, we have organized the chapter as follows. [Section 3.1](#) Introduction and motivation to early diagnosis of melanoma presents the skin cancer awareness, the problem of melanoma misdiagnosis and motivation of the undertaken research. [Section 3.2](#) Artificial intelligence and computer vision in melanoma diagnosis introduces into the most important dermoscopy image analysis stages, discusses the deep learning methods and their subdivisions. [Section 3.3](#) Medical diagnostic procedures for screening of skin diseases presents the commonly used medical devices and diagnostic procedures. [Section 3.4](#) State-of-the-Art Survey on skin mole segmentation methods includes a detailed review of different representative studies in deep learning for skin lesion segmentation. [Section 3.5](#) Improved local and global patterns detection algorithms by deep learning algorithms presents a new approach to local pattern recognition based on convolutional U-Net architecture. [Section 3.6](#) Early classification of skin melanomas in dermoscopy presents a systematic review on the classification of melanocytic lesions based on common machine learning methods as well as deep learning solutions. [Section 3.7](#) How to speed up the classification process with FPGA? explains the importance of hardware implementation of the aforementioned algorithms, while [section 3.8](#) Challenges and future directions concludes the review.

3.2 Artificial intelligence and computer vision in melanoma diagnosis

Artificial intelligence has been explored since 1960s; however, in the past few years, it has acquired immense popularity. If this is the new era of the Industrial Revolution, AI is surely going to be one of its driving forces not to take us far in the world but also overtake us in many applications and well-known solutions. Artificial intelligence is a family of more or less sophisticated methods including machine learning, deep learning, general artificial intelligence, artificial narrow intelligence, among others [6]. Artificial intelligence-related concepts and terms increasingly take center stage in a variety of settings, achieving remarkable results in image, and signal processing tasks as well as computer-vision challenges, especially regarding medical aspects. Deep learning models which are advances neural networks in general are currently one of the most popular architectures used for unsolved tasks in areas like image segmentation and classification, as well as signal analysis and data reconstruction. Fig. 3.3 presents a brief taxonomy of deep learning models where most of the methods have been used to improve the dermoscopy image analysis process.

The most commonly used deep learning models, which are used in dermoscopy images analysis, contain basic and modified architectures like convolutional neural networks (CNN), recurrent neural networks (RNN), and autoencoders. A CNN consists of one or more convolutional layers followed by a pooling layer, which is responsible for reducing the spatial size of the convolved feature. The network is closed by one or more fully connected layers as in a standard multi-layer neural network. The architecture of a CNN is designed to take advantage of the 2D or even 3D structures of an input image. The effectiveness of CNNs has

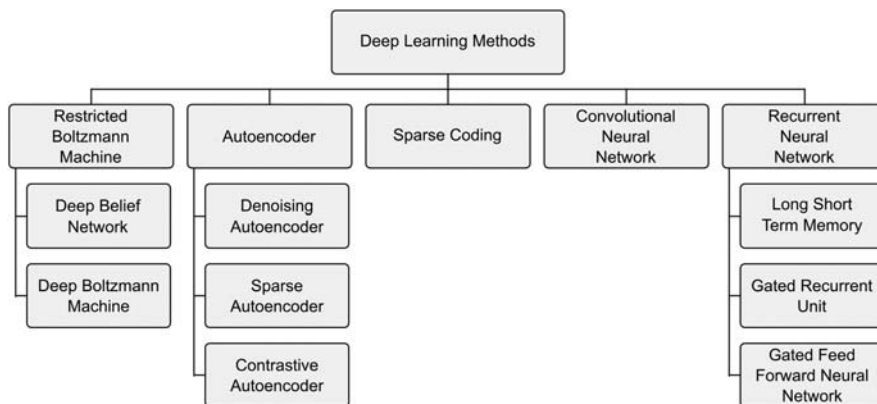


FIGURE 3.3

A brief taxonomy of deep learning models.

been proven in many tasks of computer vision due to their powerful feature representation, especially for segmentation and classification tasks. For visual datasets, like dermoscopy images, the first CNN layers might describe edges and dark local structures in the image, while high layers in the network refer to object parts and even the category of the object viewed (global features). As there are various CNN architectures a very promising approach is the concept of transfer learning which enables to use pretrained models (with already calculated weights) to extract feature representations for our new images. A RNN is a class of artificial neural networks which add an interesting twist to basic neural networks. The connections between nodes form a directed graph along a temporal sequence which allows it to exhibit temporal dynamic behavior. Mixed models CNN-RNN are successfully used for image classification tasks. Another group of deep neural networks are autoencoders that are composed of two, symmetrical deep-belief networks that typically have few layers representing the encoding part of the net, and the second set of layers that are responsible for the decoding half. Deep autoencoders are useful in image segmentation, information retrieval, modeling and data compression. A systematic review on the use of deep neural networks, including convolutional autoencoders, can be found in [7] and [8].

One of many areas of application for advanced methods of computer vision and artificial intelligence is a rapidly growing field of dermatology including dermoscopy image analysis and digital pathology. The development of computer-aided diagnosis (CAD) systems for automated diagnosis of melanoma is of high importance. The CAD systems for dermoscopy consists of following stages: (a) image preprocessing for image enhancement, black frame removal, smoothing of air bubbles, black hair detection and in-painting, (b) skin lesion segmentation responsible for correct detection of region of interest, (c) feature extraction containing shape analysis, color variegation, local and global patterns analysis, and (d) classification of skin lesion into two classes containing benign and malignant cases, or a step further into specific melanocytic lesion types. Methods which are most often used for the performance of each of the steps have been stated in Fig. 3.4.

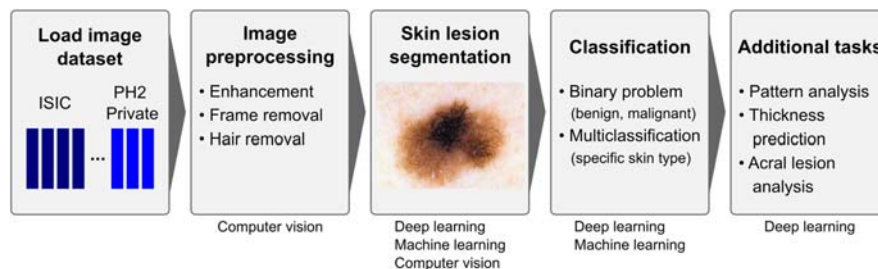


FIGURE 3.4

Illustration of CAD system for melanocytic lesion classification. Combination of individual stages with computer vision and artificial intelligence methods.

Table 3.1 summarizes the comparison of three main approaches: computer vision, conventional machine learning methods, and deep learning models in terms of strengths and weaknesses for dermoscopy image analysis. We have used characteristics including data preparation, data preprocessing, segmentation, feature representation method, generalization and classification [9].

Nowadays, the ISIC Archive and the PH2 dataset are the most employed skin lesion sources to benchmark deep-learning based tools. In Table 3.2 we have outlined the publicly available skin lesion datasets and provided additional information regarding the total amount of images and a short medical description.

Table 3.1 Comparison of computer vision, machine learning and deep learning methods for dermoscopy image analysis.

Characteristics	Computer vision	Machine learning	Deep learning
Data preparation and preprocessing	Computer vision is responsible for data preparation and preprocessing. Based on well-known methods noise and artifacts can be removed.	Machine learning requires data to be prepared and preprocessed to improve the calculation of feature vectors.	Data preprocessing and normalization is not compulsory in deep learning to obtain good results.
Feature representation method	Helps to compute feature vectors for the classification process.	Uses manually engineered feature vectors that are applications dependent.	Has the ability to determine features from input image and learn the most efficient patterns and relationships to boost recognition accuracy.
Generalization and classification	Enables data labeling and perform feature calculation.	Based on arbitrary feature selection only for labeled data. Feature selection is followed by dimensionality reduction which makes it hardly generalizable.	Helps to automatically capture spatial, temporal dependencies and scale invariant features from unlabeled raw sensor data.
Training time	Computational time depends on the algorithm advancement and data complexity.	Requires lesser training data, gives mostly lower accuracy, takes less time to train on CPU.	Requires large amount of data to avoid overfitting, provides high accuracy, longer training time and requires GPU.

Table 3.2 Research datasets for skin image analysis [10].

Dataset	Total images	Total melanomas	Description
ISIC Archive 2018/2019	13,786	1019	Biopsy confirmation has been performed for more than 50% of cases, while the ground-truth for other samples was follow-up, expert consensus, or confirmation by in-vivo confocal microscopy
PH2	200	40	Most of the melanomas have been confirmed by biopsy
SD-198	6584	198 disease state	Consumer grade photographs of skin
Derm7pt	2,000	—	Dermoscopic and clinical images of skin lesions have been classified based on the 7-point checklist criteria

3.3 Medical diagnostic procedures for screening of skin diseases

Based on the information provided earlier, accurate and early screening is a key to an early diagnosis of a disease. In this era of modern medical advancement, there is a development of sophisticated medical diagnosis modalities and techniques for early and timely screening of skin diseases. The medical imaging diagnostic modalities serve as necessary screening tools for dermatologists to quantify the grade of abnormality based on the characteristics of skin lesions (including among others asymmetry, color variation, local structures analysis) and plan the course of treatment. Medical diagnostic images related to screening of skin diseases can be broadly divided into two categories: dermoscopic (i.e., microscopic images) and non-dermoscopic (i.e., clinical images) [11,12]. A concise overview of various screening procedures commonly used for the early diagnosis of skin melanoma has been shortly discussed. Dermoscopy, also known as epiluminescence microscopy, is a non-invasive procedure that refers to the examination of skin lesions using skin surface microscopy [13]. It allows in-vivo evaluation of the microscopic structure and colors of the epidermis and the papillary dermis, which is not perceptible by the naked eye [4,14].

The procedure requires a dermatoscope with a high-quality magnifying lens with a light source that allows dermatologists to examine the lesion patterns minutely on the skin. The device allows capturing videos or images via suitable attachments like smartphones or cameras (Fig. 3.6).

Next one, high-frequency ultrasound, is a reliable and straightforward non-invasive procedure to examine skin lesions in the dermis, epidermis, and extended to the subcutaneous layer. In this procedure, the transducer probe emits high-frequency

sound waves, ranging between 7.5 and 100 MHz, for visualization of different layers of the skin at different depths [15]. The higher the frequency, the higher the image resolution and lower penetration depth. This procedure plays a vital role in the assessment of size, shape, depth, consistency, and vascularity of skin melanoma and basal cell carcinomas.

Skin biopsy is an invasive procedure where a small piece of skin is removed from the abnormal region for examination under the microscope. Histopathological examination, which is mostly performed for biopsy skin material, is considered as a “gold standard” for the diagnosis of skin-related diseases [16]. Depending on the tool used for the removal of the abnormal skin region and depth of spread of abnormality inside the skin surface, biopsies may be of four different kinds, namely: shave, saucerization, punch and excisional biopsy (see Fig. 3.5) [17,18].

Skin biopsy is generally performed to diagnose skin cancer, skin infections, or skin disorders like psoriasis [20]. Fig. 3.6 shows a skin tumor sample with visible blood vessels and its corresponding biopsy pathological slide, which is diagnosed as basal cell carcinoma.

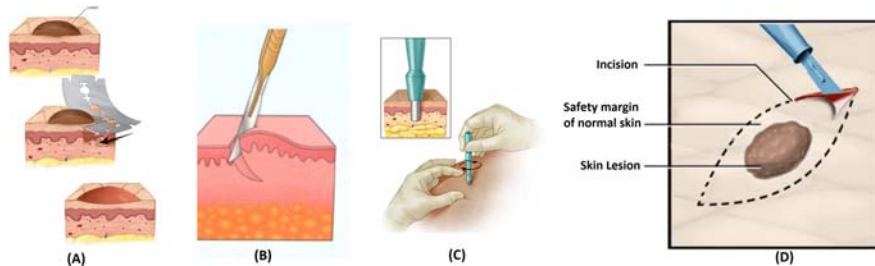


FIGURE 3.5

Skin biopsies: (A) shave biopsy, (B) saucerization biopsy, (C) punch biopsy, and (D) excisional biopsy [19].

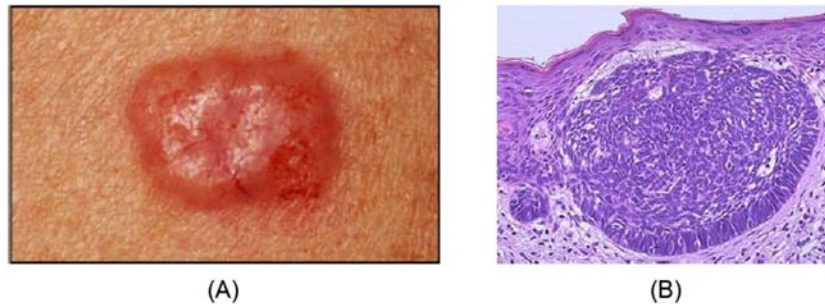


FIGURE 3.6

Skin tumor: (A) clinical view and (B) its corresponding biopsy slide [21].

Table 3.3 Comparative analysis of different diagnostic imaging modalities.

Imaging techniques	Diagnostic efficacy	Artifacts	Risk factors/ limitations	Cost of screening
Skin biopsy	Depends on the optimal choice of biopsy site and technique.	Hemorrhage, crush and split artifacts, fragmentation artifacts.	Bruising, bleeding, or soreness at the biopsy site.	Moderate (The exact cost depends on location and depth of skin lesions.) Meager
Dermoscopy	More accurate than visual inspection of a skin lesion if clinicians are aware of dermoscopic features.	Ruler markings, dark corner, ink markers, air bubbles and hairs.	Non-invasive procedure so no risk factors.	
High-frequency USG	Gives a high resolution image of skin structure and much more efficient than dermoscopy.	Hair follicle, bubbles in the gel, improper probe placement.	Cannot provide tissue diagnosis, and compression of the probe results in false thinning.	Low

In [Table 3.3](#) we have given a comparative analysis of different medical imaging diagnostic techniques discussed earlier in terms of diagnostic efficacy, artifacts, risk factors, and cost of screening.

As can be observed in [Table 3.3](#), each of the diagnostic imaging techniques has its pros and cons, but despite the development of a plethora of modern, sophisticated diagnostic modalities, dermatologists still prefer biopsy for early diagnosis of suspicious skin lesions. However, biopsy alone cannot be a reliable diagnostic procedure; hence, there is a need for subsequent other diagnosis procedures to reconfirm the spread of the disease. It may be noted that the earlier the disease is diagnosed, the more comfortable and effective is the treatment.

3.4 State-of-the-art survey on skin mole segmentation methods

Skin lesion segmentation is challenging but it is an important process in dermoscopic image processing due to several reasons: (1) Low contrast between the skin lesion (mole) and the healthy skin, variegated coloring and texture inside the lesion, irregular borders, and different artefacts.

Automatic segmentation of skin lesion is not a trivial task, this is due to most of the lesions has non-uniform coloring, and the surrounding skin is covered with

**FIGURE 3.7**

Examples of segmentation results and illustration of the non-uniform coloring and the surrounding of the skin lesions.

Table 3.4 The ISIC segmentation challenges and its top score.

Challenge dataset	Total number of images	Evaluation metric	Top score
ISIC 2016	1279	JSI	0.843
ISIC 2017	2750	JSI	0.765
ISIC 2018	3694	Thresholded JSI	0.802

the remaining parts after the preprocessing step, which make the process even harder to carry out (Fig. 3.7). Furthermore, it has to be accurate and efficient, because the subsequent steps including border irregularity analysis, asymmetry analysis, feature extraction, and classification crucially depend on it. Therefore, the segmentation algorithms are one of the most widely explored in the dermoscopy image analysis. Owing to the difficulties described earlier, skin lesion challenges based on the ISIC database are organized yearly [22]. These challenges evaluate the performance of submitted classifiers using Jaccard Similarity Index (JSI), which is widely known as Intersect over Union (IoU) in computer vision research. JSI is defined as the overlap of the automatic segmentation to the manually delineated ground truth binary mask:

$$JSI(A, M) = \frac{|A \cap M|}{|A \cup M|}$$

where $JSI(A, M) = 1$ represents the best result, a 100% overlap with the manual delineation.

The performance of winning algorithms in ISIC segmentation competitions are summarized in Table 3.4.

Before the era of deep learning, image processing methods and conventional machine learning were widely used in research. There are some previous works presented in depth review on lesion segmentation on dermoscopic images [23,24]. Korotkov et al. [24] and Pathan et al. [25] reported that the majority of algorithms were designed based on hand-crafted features in detection, classification and segmentation of skin lesions. These reviews focused on thresholding methods (histogram based or gray-scale), active contour (based on energy functions to find the

optimum border), region growing (recursively merging pixels of regions), and some methods were combined with color and texture classification to identify the skin lesions. Korotkov et al. [24] found the main barrier is the lack of dataset availability that makes it harder to assess the algorithms without bias. They concluded that there is a large discrepancy in previous works and the CAD systems were not ready for real-world applications.

There are limited work based on hand-crafted features after 2016. We found two non-deep learning approaches proposed by Ashour et al. [26,27]. Their approaches are based on a histogram-based clustering estimation algorithm to determine the required number of clusters using the neutrosophic c-means clustering method. They perform segmentation on ISIC 2016 dataset and neutrosophic k-means using genetic algorithm is used for skin lesion detection in dermoscopy images. To bridge the gap of the existing literature reviews, we are focusing on deep learning methods, which are the state of the art for segmentation algorithms. We refer the readers to previous review papers [24,25] for conventional methods.

The popularity of deep learning algorithms in skin lesions segmentation was driven by the International Skin Imaging Collaboration (ISIC). ISIC provides digital skin lesion image datasets with worldwide experts/dermatologists annotations and delineation of the skin lesion. This is the main driving force to encourage development of automated algorithms, particularly data-driven approaches, for the diagnosis of malignant skin lesions and other conditions. Additionally, ISIC organizes yearly skin lesion (classification or segmentation) challenges to encourage the use of datasets for to improve the performance of the CAD solutions and raise the awareness of skin cancer [22]. These datasets are now the benchmark datasets for computerized methods and many have used it to train the clinical aspects of skin.

ISIC 2016 Skin Lesion Analysis Toward Melanoma Detection Challenge dataset [28] (henceforth, ISIC 2016 dataset) consist of 1279 images, enabling the use of CNN in segmentation. Yuan et al. [29], the top score for ISIC 2016 segmentation challenge, leveraged 19-layer DCNN for an end-to-end fully automatic method for skin lesions segmentation. To improve the performance, they proposed to use a loss function based on Jaccard Distance. They experimented by using different parameter settings: input size, augmented strategies, loss function and optimization methods. Fivefold cross validation was used to fine-tuned the hyperparameters. In addition to ISBI 2016, they also evaluated their results on PH2 dataset. Overall, their proposed method outperformed the state-of-the-art methods. However, the method did not work well in images with low contrast.

On the similar ISIC 2016 segmentation challenge, Yu et al. [30] ranked second amongst 28 teams. Unlike Yuan et al. [29], they proposed very deep residual networks that consist of 50 layers for two-stage segmentation, followed by classification. According to their design principles, deeper networks produce more discriminative and richer features for recognition task. Although the result seems promising, it is very costly and inefficient due to two-stage framework and computational expensive deep networks.

In 2017, semantic segmentation based on a multi-stage fully convolutional networks (FCNs) for skin lesions segmentation was proposed by Bi et al. [31]. The multi-stage consists of: early stage of localized coarse appearance learning and later stage of detailed boundaries characteristics learning. Additionally, they enhanced the performance of the segmentation by deploying a parallel integration approach, which enabled the fusion of result. Similar to Yuan et al. [29], they also evaluated their algorithm on PH2 dataset [32] and ISIC 2016 dataset. They achieved 90.66% on PH2 but marginal improvement when compared to Team ExB in ISIC 2016 competition with 91.18%.

Other two-stage segmentation methods were proposed by Vesal et al. [33] and Goyal et al. [34]. For the first stage, they used Faster-RCNN. Then Vesal et al. [33] proposed a modified version of U-Net for skin lesion segmentation. Goyal et al. [34] implemented deep extreme method as second stage for segmentation.

The first attempt to perform multi-class segmentation was proposed by Goyal et al. [35]. They implemented semantic segmentation algorithm based on fully convolutional methods for multi-class segmentation and evaluated on ISBI challenge dataset 2017 [22] (henceforth, ISIC 2017 dataset). The multi-class segmentation distinguishes three classes, that is melanocytic nevus, melanoma and seborrheic keratosis.

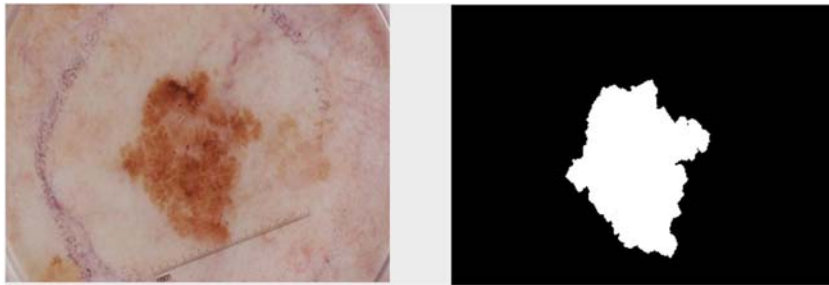
In pixel-wise dermoscopic skin lesions segmentation approach, Al-masni et al. [36] proposed a fully resolution convolutional network (FrCN) to learn full resolution features of each image pixels for skin segmentation. When evaluated on ISIC 2017 testing set, they achieved JSI of 77.11%. On the other hand, Soudani et al. [37] proposed the use of two deep learning classification models to recommend the most suitable lesions segmentation technique.

Recently, Goyal et al. [34] designed fully automated ensemble deep learning frameworks, which combine one of the best semantic segmentation method, that is DeeplabV3+ and one of the best instance segmentation algorithm, that is Mask R-CNN. The authors claimed Ensemble-L and Ensemble-A performed best in Sensitivity and Ensemble-S in specificity on ISIC 2017 dataset and PH2 testing dataset.

To compare the performance of segmentation, we compared the results on two publicly available skin lesion datasets, which are ISIC 2017 dataset [22] and PH2 dataset [32]. We are not able to demonstrate the results on ISIC Challenge 2018 [22] as they did not share the ground truth delineation of the images in testing set. ISIC 2017 segmentation dataset consists of 2750 images, with the split of 2000 images in training set, 150 in the validation set and 600 in the testing set. Fig. 3.8 is an example of ground truth delineation, where the boundary was loosely drawn by an expert. PH2 dataset has 200 images in which 160 images are nevus (atypical nevus and common nevus), and 40 images are of melanoma [32]. In this dataset, the ground truths delineation are precise represent the true boundaries of the skin lesion (high specificity), as shown in the Fig. 3.9. PH2 is used to test the robustness of the model trained by using the ISIC 2017 training set.

**FIGURE 3.8**

Example of ground truth delineation by experts. The binary mask illustrates an example of loosely drawn boundary in the dataset.

**FIGURE 3.9**

Example of ground truth delineation by experts. The binary mask illustrates an example of precisely drawn boundary in the dataset.

3.4.1 Comparison of the state of the art

Table 3.5 compares the performance of the state-of-the-art deep learning methods on ISIC 2017 testing dataset using the performance metrics used in the segmentation challenge, as presented in [36]. Overall, Yu et al. [30] achieved the best results, followed by Goyal et al. [34]. Other methods that achieved good results were SegNet [38] and classic U-Net [39].

Table 3.6 compares the performance of the state-of-the-art deep learning methods on PH2 dataset. The results are based on the training model built based on ISIC 2017 training set and test on the PH2 dataset. It is noted that Goyal et al. [34] achieved the overall best results.

Table 3.5 Comparisons of the state-of-the-art deep learning approaches on ISIC 2017 testing set.

Method	Accuracy	Dice	Jaccard index	Sensitivity	Specificity
Yuan et al.	0.934	0.849	0.765	0.825	0.975
SegNet	0.918	0.821	0.696	0.801	0.954
U-Net	0.901	0.763	0.616	0.672	0.972
Yu et al.	0.949	0.897	0.829	0.911	0.957
FrCN	0.940	0.870	0.771	0.854	0.967
Goyal et al.	0.941	0.871	0.793	0.899	0.950

Table 3.6 Comparisons of the state-of-the-art deep learning approaches on PH2 dataset.

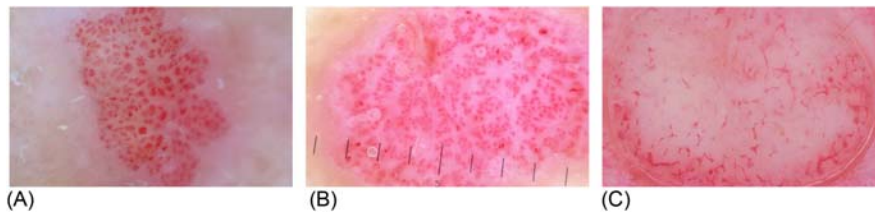
Method	Accuracy	Dice	JSI	Sensitivity	Specificity
FCN-16s	0.917	0.881	0.802	0.939	0.884
DeeplabV3 +	0.923	0.890	0.814	0.943	0.896
Mask R-CNN	0.937	0.904	0.830	0.969	0.897
Goyal et al.	0.938	0.907	0.839	0.932	0.929

3.4.2 Summary

The fully automated and end-to-end skin lesions segmentation can provide inference to aid in the decision of lesion diagnosis. Future challenge including designing robust algorithm that can work across different skin type and artefacts, including hair, hair follicles and coloring of surrounding skin. Other solution may incorporate different data augmentation techniques [40], which demonstrated promising results in enhancing the deep learning algorithms.

3.5 Improved local and global patterns detection algorithms by deep learning algorithms

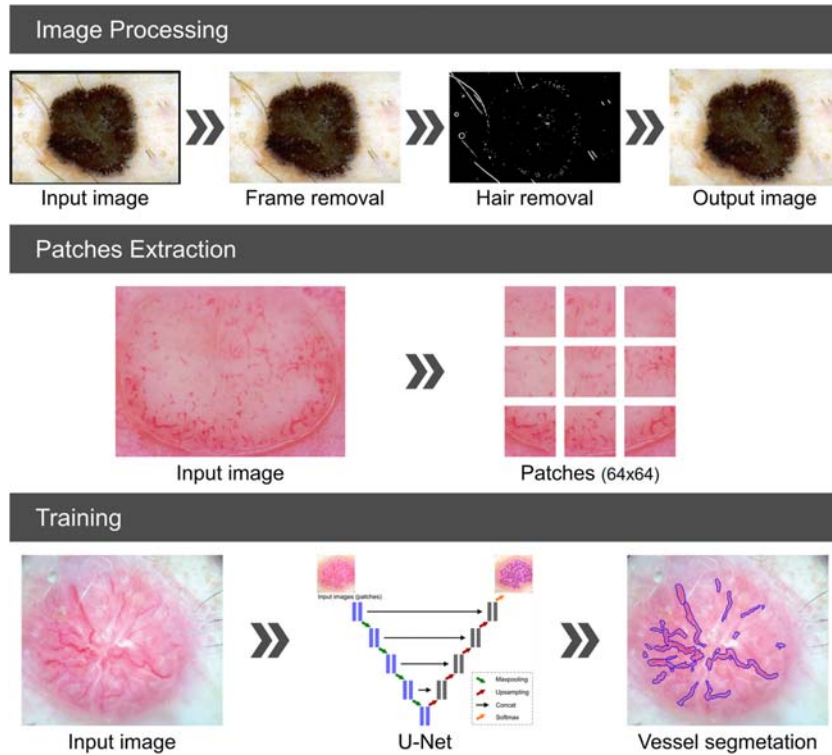
Dermoscopic features which are also called dermoscopic criteria are important statistical denominators while assessing the skin mole. Dermoscopic criteria consist of global patterns which can be observed in whole skin lesion while local features may appear only in parts of the mole. Corresponding global and local patterns are observed in particular skin lesion types. One of the most important local criteria is the atypical vascular pattern which has the second highest odds ratio of 7.42 and is among the three most important features that indicate the

**FIGURE 3.10**

Vascular pattern observed in melanocytic lesions as red dots and lines.

malignancy of the diagnosed skin mole. Atypical vascular pattern helps to differentiate between benign lesions and melanomas. Atypical vascular pattern has been introduced in 1994 by Fineberg and Rosen and described as linear-irregular or dotted vessels associated with other local structures (Fig. 3.10).

This pattern can be located in parallel or vertically to the skin's surface which will change in appearance as lines or dots and nodes, respectively. Researchers observe a relationship between the vascular pattern and observed tumor size and degree of malignancy. During the last few years only several research groups have proposed a solution for the detection of vascular structures in dermoscopic color images. As the vascular pattern detection and analysis tasks are among the most challenging in dermoscopy image processing Authors take advantage of DNN architectures including autoencoders and U-Nets. In work [41] Kharazmi et al. implemented a stack of sparse autoencoders for detection of cutaneous vessels based on feature calculation and selection from the raw data. The proposed framework achieved 95.4% detection accuracy for diverse vessel patterns. Betta et al. in work [42] described an approach for the identification of atypical vessels. Due to the difficulty in obtaining a relevant and reliable number of epiluminescence microscopy (ELM) images with vascular pattern, the training set has been augmented. Pixel classification has been performed in the HSL color space which determines Hue, Saturation, and Luminance. Components of the HSL color space have been calculated and a frequency histogram has been used for classification. As this approach is highly generalizable the Authors confirmed low specificity by misclassifying the area [23,43]. In 2014, Fabbrocini et al. described an algorithm based on texture analysis for vascular pattern detection [44]. The system performance has been tested on 200 dermoscopic images and scored 80% sensitivity, and 78% specificity, respectively. Texture descriptors including entropy, correlation, and inverse difference moment have been calculated based on the gray level co-occurrence matrix. The latest research in this area has been presented in [45] by Kharazmi et al. who has used independent component analysis to decompose the image into melanin and hemoglobin to differentiate between red areas and background. Using k-means clustering the area is divided into groups where the hemoglobin component can be easily separated from the surrounding and a vessel mask is generated as a result of global thresholding. The Authors achieved

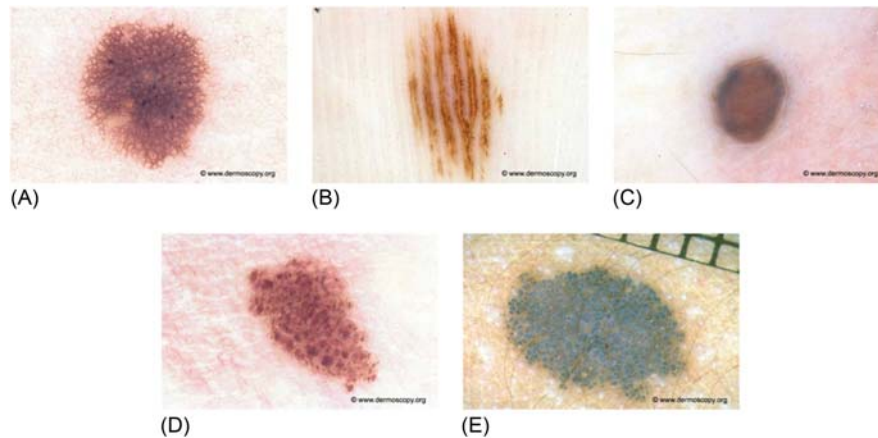
**FIGURE 3.11**

The system workflow of a CNN architecture described in [46] consists of (A) Image preprocessing stage which consists of black frame removal, hair inpainting, and further image normalization, (B) Patch extraction adapted to the network architecture based on a sliding window as well as data augmentation, (C) U-Net architecture implementation, hyperparameter specification, training and validation.

sensitivity and specificity of 90% and 86%, respectively for a dataset containing 500,000 manually segmented pixels provided by an expert. In research work [46] we have proposed a U-Net based solution where the proposed implementation has been divided into four main stages including preprocessing (image enhancement), patch extraction, training, and validation as illustrated in Fig. 3.11. The algorithm achieved an average DSC of 0.84, sensitivity 0.85, and specificity 0.81.

Another group of highly significant diagnostic parameters in determining malignancy of the skin lesion include the general appearance of global patterns of the lesion. The most commonly analyzed global patterns of skin lesions include (Fig. 3.12):

- Pigment network (reticular) pattern: appears mostly as thin brown lines over a light brown background. Most commonly observed global pattern in melanocytic lesions which mostly covers the entire skin lesion.

**FIGURE 3.12**

Global pattern observed in melanocytic skin moles: (A) pigment network, (B) parallel pattern, (C) homogeneous pattern, (D) globular pattern, and (E) cobblestone pattern.

- Globular pattern: occurs as round to oval structures variously sized and distributed throughout the skin lesion. Mostly light and dark brown as well as of gray-black coloration.
- Cobblestone pattern: very similar to the globular pattern but composed of larger and closely aggregated globules resembling a cobblestone.
- Homogeneous pattern: area without local features as well as other patterns mostly of brown, gray-blue, gray-black, or reddish-black pigmentation.
- Parallel pattern: it is found on the palms and soles due to the particular anatomy of these areas.

As the skin global pattern analysis is very challenging, there are only few works that address this problem and propose to classify the global pattern into five classes described earlier based on texture analysis. In Serrano et al. [47] used Markov random field (MRF) which is a graphical representation of a joint probability distribution. First, each image plane in $L^*a^*b^*$ color space is modeled as a MRF following a finite symmetric conditional model. Coupling of color components is taken into account by supposing that features of the MRF in the three color planes follow a multivariate normal distribution. The best classification rate is 86% on average over 100 tiles (sized 40 40).

In Sadeghi [48] uses the joint distribution of color in the $L^*a^*b^*$ color space to analyze the image texture. Comprehensive set of the state-of-the-art filter banks to model the texture by the joint probability distribution of filter responses. This distribution is represented by the frequency histogram of filter response cluster centers called textons. The implementation has been tested on 375 images and scored 86.8% accuracy.

3.6 Early classification of skin melanomas in dermoscopy

As the impact of machine learning in dermatology will increase in the following years, in this section we outline main approaches to skin lesions classification, quote most important latest methods, discuss general limitations regarding machine learning methods, and conclude this section with our perspectives about this field for the future (including applications of deep learning classifiers in teledermatology).

Before 2016, most studies (described in [Section 3.6.2](#)) followed the classical workflow of machine learning (preprocessing, segmentation, feature extraction, and classification) and focused on detecting a number of features describing criteria of (clinical) diagnostic algorithms (mentioned in [Section 3.6.1](#)) using hand-crafted features, extracted by means of traditional computer vision methods, to feed a single machine learning classifier. However, since 2016 [\[49\]](#), with recent advancements in the field of deep learning and the availability of large datasets of dermoscopic images required by deep learning models for reasonable training, the current trend is rather to directly classify the disease from the image using deep neural networks (as discussed in [Section 3.6.3](#)). In particular, CNNs have become the classifiers of choice due to their high accuracy. CNNs relieve the machine learning expert of the burden of manual “feature engineering” by automatically discovering high-level abstractions from low-level data [\[50\]](#).

Most of the research done so far in the area of automatic skin lesion diagnostics has concentrated on creating new methods for automatic differentiation between malignant melanoma and other skin lesions, that is, on the “benign vs malignant” classification, were restricted to lesions that were obviously melanocytic [\[51–53\]](#). However, as recently larger datasets became available, studies which extend this dichotomic classification to a trichotomic one (e.g., “melanomas vs seborrheic keratosis vs nevi” [\[54\]](#)) or even capable of differentiating into one of 10 or more individual clinical entities [\[55,56\]](#) started appearing in the literature.

3.6.1 Diagnostic algorithms

The first step in the diagnosis of a melanocytic lesion is a visual examination of the suspicious skin area, typically using a dermatoscope, by a dermatologist, who qualitatively and/or quantitatively assesses numerous individual ELM criteria to make the diagnosis. This so-called “pattern analysis” involves the assessment of two types of lesion’s dermoscopic patterns: both global (e.g., homogeneous, globular, starburst, etc.) and local (e.g., irregular streaks, pigment network, blue-whitish veil, etc.). To simplify diagnoses, following three rule-based diagnostic algorithms (based on the same criteria as pattern analysis) have been proposed. They become widely accepted by clinicians, and are nowadays used by dermatologists in common medical practice [\[57\]](#).

- ABCD rule refers to the four criteria found to be significant cofactors for diagnosing melanoma: Asymmetry (with respect to the latter three criteria),

Borders (i.e., sharp and abrupt changes), Colors (i.e., the presence of each of 6 “dermoscopic” colors), and Different structural components (e.g., pigment network, dots, globules, etc.). The likelihood of melanoma depends on adding up the scores for different features present in the lesion, each having its own score and factor.

- Seven-Point Checklist Method (7PCL) is based on the evaluation of two types of criteria, major and minor ones, each having a distinct score value. In total there are 7 criteria: three major (blue-whitish veil, atypical pigment network, and atypical vascular pattern), and four minor (irregular globules/dots, irregular blotches, irregular streaks, and regression structures). If the summed up score value of criteria present in the lesion exceeds a given threshold, the lesion is classified as melanoma.
- Menzies Method is a simplified algorithm based on the evaluation of both “positive” and “negative” indicators of melanoma. For a melanoma to be diagnosed, none of the two “negatives” (i.e., lesion’s symmetry and single color presence) should be found and at least 1 of 9 “positives” (e.g., blue-whitish veil) must be present.

Among the aforementioned methods, only pattern analysis is suitable for both melanocytic and non-melanocytic lesions. Fig. 3.13 shows the summary of all the methods described in this section. Additional information based on clinical covariates (e.g., age, gender, and lesion location), is also taken into account [58].

3.6.2 Approaches to detect the diagnostic criteria

Detailed information on the CAD systems for skin lesion diagnosis using this approach (i.e., focusing on at least partially emulating one of the diagnostic algorithms described in Section 3.6.1 using handcrafted features or on combining multiple pre-detected criteria in a single machine learning classifier model) can be found in review papers [23–25,53,59–62].

They typically use supervised machine learning techniques, such as decision trees and SVMs, and their sensitivity ranges between 80%–100% and specificity between 50%–95%. The drawbacks of this “traditional” approach are that it requires a vast application-specific expertise, particularly for feature extraction (a process which additionally is very time-consuming), and that errors and the loss of information in the first processing steps (e.g., poor segmentation) strongly affects feature extraction and, consequently, the classification quality as well.

Feature descriptors commonly used in the computerized analysis of dermoscopic images and related to ABCD criteria include [63]: (a) asymmetry: symmetry distance [64] and lesion’s centroid [65]; (b) border irregularity: Fourier feature [66], fractal geometry [67], area and perimeter [65,68,69], and irregularity index [70,71]; (c) color variegation: RGB statistical descriptors [65,72]; and (d) different structural components: pattern analysis [65,73,74], wavelet-based descriptors [75], texture descriptors [76], intensity distribution descriptors [77], and Haralick descriptors

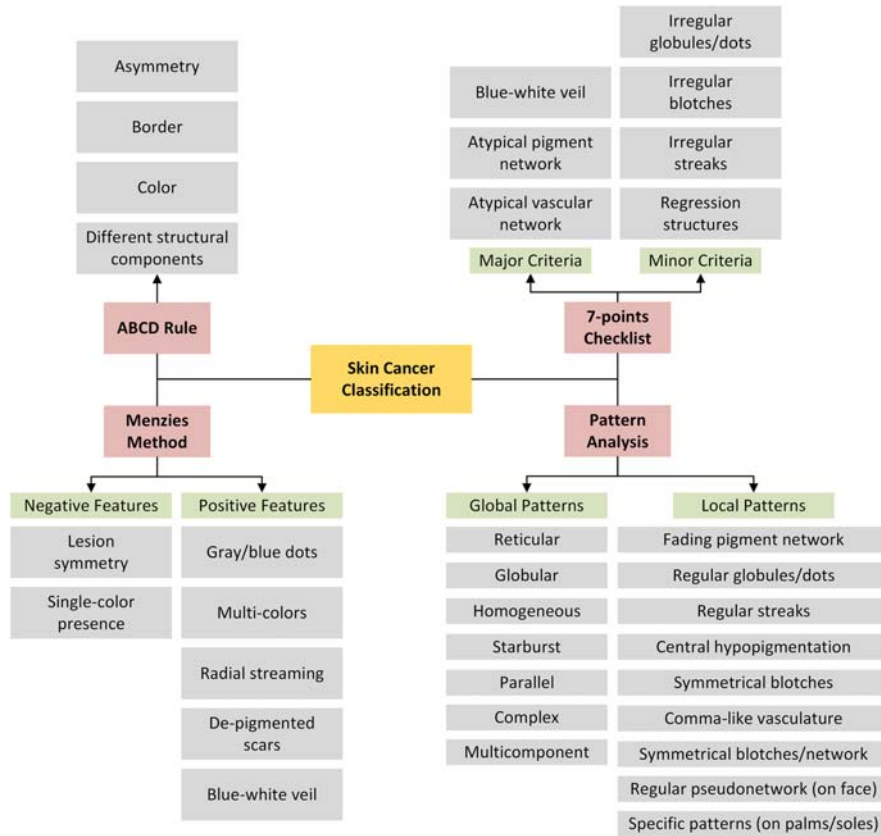


FIGURE 3.13 Summary of skin cancer classification methods used commonly by clinicians.

[78]. Some other methods focus on the detection of individual features such as streaks [79] and blue whitish-veil [80,81] or try to detect all features at once [44,82]. However, typically each pipeline (designed specifically for a particular feature) significantly increases complexity of the system and requires careful tuning of hyperparameters. For instance, irregular streaks detection involves obtaining precise lesion border detection to compute an “irregularity” index, describing how the lesion border differs from a straight line when the lesion is divided into segments.

3.6.3 Approaches to directly classify skin conditions

Within four recent years deep learning models, particularly CNNs, became a trend to deal with the task of directly classifying skin conditions due to their remarkable performance in this field—some presented models are competitive to or even

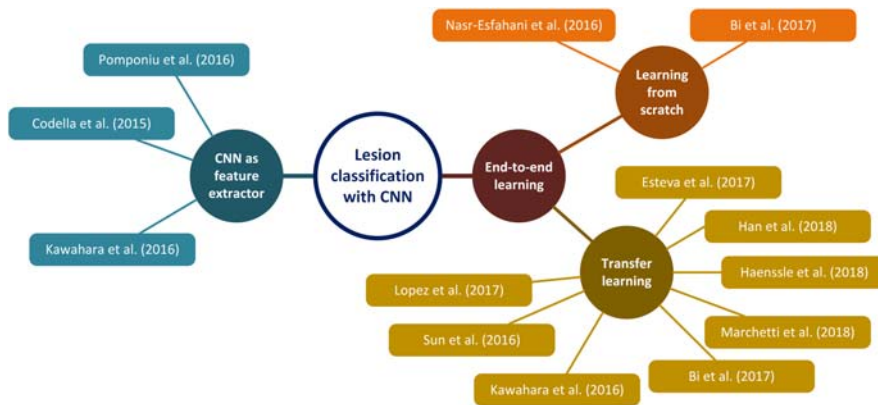


FIGURE 3.14

An overview of convolutional neural networks (CNNs) methods used for skin lesions classification and their corresponding categorization.

outperform the dermatologists [83,84]. As the speed of development in the area of deep learning for dermatology is tremendous, in this section we present only the most typical approaches and some of the most relevant studies (Fig. 3.14). Further discussion on the topic can be found in dedicated review articles [85,86].

3.6.3.1 Classifiers utilizing the convolutional neural networks as a feature extractor

A CNN can be used for classification by substituting the fully connected layers of a CNN (typically of the AlexNet [49,55,87], but other models—such as ResNets—were also used [30,49]), pretrained on a large dataset (such as ImageNet) with such classifier as for instance k-nearest-neighbor classifier (using cosine distance metrics) [87], support vector machine (SVM) [49], or a convolutional layer [55]. An interesting solution was proposed by Codella et al. [49], who showed that the use of deep features results in a better performance compared to classifiers that only used low-level handcrafted features (apart from the modified AlexNet outputs they also used low-level handcrafted features as well as features from sparse coding, a deep residual network, and a convolutional U-network). Similar results were obtained in [55] and [88].

3.6.3.2 Classifiers using end-to-end learning convolutional neural networks model training with transfer learning

As publicly available datasets of dermoscopic images are limited, a common strategy for skin lesion classification—known as transfer learning—is to use a CNN pretrained using large ImageNet [89] dataset (despite its non-medical image domain, the learned features have sufficiently high quality for lesion classification [90]) and adapt some of its weighting parameters to the actual classification

problem. Four notable works representing this trend, all using very similar approaches, are briefly described in the following sections.

Esteva et al. [91] presented a landmark publication, as it was the first study in the field of dermatology involving a CNN model trained with huge and diverse dataset, consisting of nearly 130,000 images (of which nearly 3400 were dermoscopic images representing over 2000 different skin lesions). The authors considered two binary classification problems: “keratinocyte carcinomas vs. benign seborrheic keratosis” and “malignant melanomas vs. benign nevi” (the latter was performed for both clinical and dermoscopic images). The model of choice was a GoogLeNet Inception v3, pretrained with ImageNet database and then fine-tuned to classify skin lesions using transfer learning. The novelty of this study was also the use of a tree-structured disease taxonomy which “leaves” are formed by the individual diseases and inner nodes group diseases that are visually and clinically similar. The CNN outputs a probability distribution with over 757 training classes (the probabilities of a coarser lesion class, i.e., an inner node at a higher level in the tree, are determined by summing up the probabilities of its child nodes). Authors demonstrated that a CNN trained for finer classes has a better performance than the one trained for the distinct classes that are of interest for the problem. The trained CNN achieved an AUC ROC of 0.96 for both melanomas and carcinomas, and of 0.94 for melanomas classified exclusively with dermoscopic images, which is a dermatologist-level diagnostic.

Haenssle et al. [83] adapted GoogLeNet Inception v3 model for the “melanoma vs. nevi” classification with transfer learning, but fine-tuned weights in all layers. The model achieved AUC ROC of 0.86. The publication included the largest number of dermatologists to date ($n = 58$) and was the first to evidently prove that additional clinical information improves both sensitivity and specificity of dermatologists.

The study by Han et al. [56] is particularly noteworthy for its scientific transparency as they have made their computer algorithm publicly available. The authors developed a fine-tuned ResNet model capable of classifying clinical images of lesions into one of 12 different classes representing individual diseases, which scored ROC AUCs of 0.96 for melanoma.

Sun et al. [92] experimented with CaffeNet and VGGNet models for the task of classifying lesions into one of 198 finely defined classes. The best average accuracy of 50.3% over all classes was achieved by a pretrained VGGNet optimized using transfer learning. A modified VGGNet was also used by Lopez et al. [93] for the “melanoma vs. benign” problem. The authors compared the accuracy of the same CNN architecture but trained in different modes: from scratch, pretrained with transfer learning and frozen layers, and pretrained with transfer learning and then fine-tuned; the last one achieved the highest accuracy of 81.3%.

3.6.3.3 Convolutional neural networks model training from scratch

Some approaches involve training the model “from scratch” [54,94] (these two methods used ResNet and a custom two-layer CNN, respectively). However, Menegola et al. [90] showed that fine-tuning a neural network pretrained only

over ImageNet, performed better than training a neural network from scratch (for fine-tuning they used the Argenziano et al. [4] dataset).

3.6.3.4 Ensembles of convolutional neural networks models

Another trend in the field of deep learning for dermoscopy is to use an ensemble of deep models instead of a single method (e.g., [95,96]) which helps increase effectiveness and reliability of predictions, particularly for ISIC dataset [97]. The works using this approach and listed in the following are particularly noteworthy.

Marchetti et al. [98] implemented five methods to feed a single classifier with all automated predictions from the 25 teams participating in the ISBI 2016 Challenge. The greedy fusion achieved highest performance among ensemble methods with a sensitivity of 58% and a specificity of 88%.

Bi et al. [54], who considered the classification of lesions into one of the three classes (as either melanoma, seborrheic keratosis, or nevus), did not train multiple CNNs for the same classification problem, but rather trained three ResNets for different problems (by fine-tuning a pretrained CNN): one for the original three-class problem and two binary 1-vs-all classifiers. The classifier scored an ROC AUC of 0.854 for melanomas and an average ROC AUC over all classes of 0.915.

Kawahara et al. [99] presented a unique architecture of a CNN ensemble: their CNN is composed of a number of modules each considering the same image but at a different resolution, and in an end layer their outputs are combined into a single layer (the weighting parameters are fully optimized by end-to-end learning). The model achieved mean accuracy of 79.5%.

3.7 Conclusions

One issue with the comparison of skin lesion classification methods is that the considered problem formulations of the individual works differ, sometimes only slightly. This occurs not only for the considered training classes and the used data, but also for the presented statistical quantities.

The size of the training dataset varied greatly among the presented methods, from less than 300 dermoscopic images [98] to nearly 130,000 clinical images [91] (in some cases data augmentation was used to increase the number of training samples [87]). Typically, the dataset of choice (at least for model fine-tuning) was ISIC Archive [28], used in [30,83,84,95].

Moreover, in addition to publicly accessible data archives some works use nonpublic archives of skin clinics for training and/or testing [83,91], or do not thoroughly present the applied training/testing procedure (e.g., do not disclose the exact size of the test set [83]), making results difficult or even impossible to reproduce and compare. Some results should be viewed critically as the test dataset was very limited [94], the algorithm was not tested with an independent test

dataset (only cross-validated on the training data) [87], or the test set diagnoses were not biopsy-proven [83].

The drawback of some methods is that they are semi-automatic (e.g., the region of interest for each skin lesion must be manually annotated [87]).

The vast majority of the DNN-based models use only images to output their diagnostics. However, dermatologists typically consider also clinical covariates (e.g., patient's age, sex, ethnicity, and anatomic location) as they are often correlated with certain skin conditions [58]. Approaches incorporating these information into the model achieved a diagnostic improvement of 4%–7% [83,100,101].

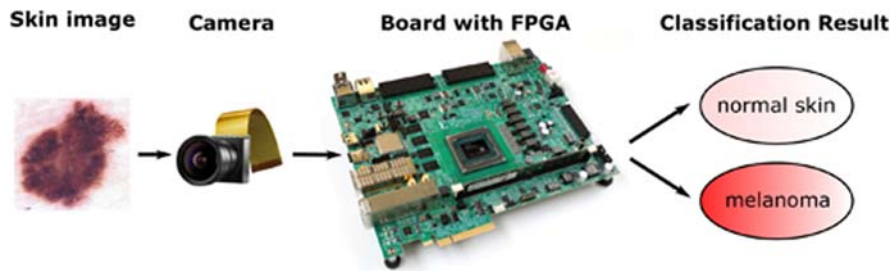
Finally, it is worth noting the recent work by Faes et al. [102], in which authors, healthcare professionals lacking any expertise with the DNN development, used the Google Cloud AutoML on the ISIC archive and reported a result comparable to other elementary classification tasks in this section.

3.8 How to speed up the classification process with field-programmable gate arrays?

In the light of the rapidly rising incidence rate early diagnosis of melanoma has become an important issue. There is a need to develop a computer-aided system to make the early detection of melanoma easier for non-experts and the general public. Without a doubt, in any CAD device the most important parameter is reliability and accuracy but it should also meet embedded system constraints such as real-time work, low resource utilization, power consumption, and cost [103].

One of the approaches is parallel computing. The main idea is to carry out some calculations or processes executions simultaneously. Large problems can often be divided into smaller ones, which can then be solved independently, at the same time. This can be done, for example, using GPGPUs (General-purpose computing on graphics processing units) or FPGAs (Field-Programmable Gate Arrays)—the latter being more portable and easily reconfigurable. An FPGA is a device composed of a set of programmable logic gates. In contrary to most integrated circuits, which are dedicated to one task, FPGA's hardware resources can be reconfigured to provide the desired functionality. Systems behavior is specified using Hardware Description Language (HDL) such as Verilog or VHDL. Modern FPGAs are fused together with processors or memory blocks on a complete system-on-chip boards. Furthermore, companies are offering tools for HDL code generation from high-level languages like C or C++ making development of hardware implementation easier for beginners. FPGAs have recently found many applications in image processing tasks.

Melanoma image analysis systems mostly consist of two parts mainly image processing with area segmentation and feature extraction followed by classification. Both of them can be easily parallelized as basic features are often calculated for each pixels based mostly on a small neighborhood.

**FIGURE 3.15**

Prototype of a system architecture for early melanoma diagnosis.

Dermoscopic images are complex and inhomogeneous. Moreover, in almost every image appear extraneous artifacts—skin lines, air bubbles, and hairs. The preprocessing step is, therefore, essential to improve the quality, as it reduces the artifacts and noise. The preprocessing step consists of the following: black frame removal, smoothing the air bubbles, and hair in-painting. These stages are not computationally demanding and can easily be implemented in hardware, using FPGA (Fig. 3.15). One of such ideas, applied to melanoma images, has been presented in [104]. Authors have implemented image enhancement including contrast and brightness change, conversion to gray level, invert transform and thresholding. Xilinx ISE Design Suite has been used to develop the system as it is dedicated to programming of FPGA reconfigurable circuits. It helps with the design, verification, and synthesis of programs written in HDLs (Verilog, VHDL).

However, we can easily implement the preprocessing and feature extraction stages on reconfigurable circuits, making them less time consuming, the most compute-intensive part remains classification. Sophisticated and complex classifiers need to be compared in terms of computational efficiency and the possibility of parallelization. The most extensive analysis in this area has been conducted by a research group from Auckland University of Technology in New Zealand, which is understandable, as Australia and New Zealand are countries with the highest melanoma incidence rate cases. In 2015 Afifi et al. examined five different machine learning classifiers specifically for malignant melanoma [105]. The SVM achieved the highest accuracy level among five different classifiers that have been tested, the others being Naive Bayes, Multi-Layer Perceptron, Nearest Neighbors and Random Forest. The authors accelerate the SVM classifier by implementing its time consuming computation part on FPGA as a coprocessor. They used Xilinx Zynq device and the High-Level Synthesis (HLS) design methodology [105]. One year later, in 2016, the same research group described a fully hardware SVM implementation, on the recent hybrid Zynq SoC, using the latest UltraFast HLS design methodology, and achieving better results than before [106]. In further research they examined different modifications to SVM algorithm, such as cascade or dynamic classifier [107,108]. With each modification,

they got increased computing speed, due to design optimization and cutting-edge technology. Their latest solution, presented in [103] retained 97% accuracy of software SVM-based application while achieving 36x acceleration factor.

In recent years, we have observed an increasing importance of real-time applications in different areas such as digital signal processing, wireless sensor networks, and healthcare. An idea of embedded real-time CAD system for medical purpose needs to be exploited. A low-cost handheld device for melanoma detection would be useful in the primary care. Leading companies, like Xilinx and Intel, manufacturing FPGA boards and software, have recently concentrated significantly on deep neural networks. They develop system-on-chip devices with dedicated AI-cores (e.g., Xilinx Versal), which main function is an acceleration of artificial intelligence algorithms. Although, to the best of our knowledge, there are currently no publications regarding the use of deep neural networks in melanoma, implemented on FPGA boards, such solutions begin to appear in other fields [109]. Based on the current trends in research and directions followed by private companies, we can expect that the next few years will bring novel ideas combining deep learning solutions with FPGA possibilities. We hope that this will ultimately lead to a further progress in melanoma classification and treatment.

3.9 Challenges and future directions

Currently, there are two main problems related to the application of deep learning in medicine, and particularly in dermatology—insufficient data publicly available, and the way most models output their predictions.

It is a challenging task to automatically detect skin cancer as skin lesions in the dermatology field exhibit huge variability. According to the recent advances reported for this task, the most successful machine learning technique addressing that problem is deep learning. However, large amount of data are necessary to apply deep learning. As it is a challenging task to collect medical data, particularly from skin cancer, even the rich ISIC archive is often not enough to fully exploit the potential of this methodology [30,56]. Not only is the total number of images too low, but also some classes of samples are underrepresented—some datasets (e.g., the one used by Liu et al. [101] or ISIC Archive [28]) contain mostly samples of skin lesions from light-skinned people, which contribute to the bias and prevents trained models to be successful in detecting skin cancer for dark-skinned people as well [110]. A CNN may learn to deal with the skin color only if it observes enough pictures of each type of skin during the training. Approaches dealing with these problems include data augmentation, transfer learning, weighted loss, and up/down-sampling [110,111]. Finally, datasets should include not only images but also clinical information (metadata) accompanying them, as they provide valuable information for classifiers [83,100,101] (especially when the model is trained on a sparse dataset). It is important to develop CAD systems integrating data from three main types of skin evaluations—clinical,

dermoscopic and histopathological—should large datasets of the respective types of images be publicly available.

The second problem is that most state-of-the-art predictive models yield the diagnosis only as the label that produced the highest probability, in certain cases also accompanied by a threshold or a ranking for dubious lesions [56,101]. Nevertheless, such results are not explainable by clinicians who, in fact, need to know also diagnostic clues which led to a particular classifier outcome (i.e., they would like to know the reasons for the selection of such a disease by the model) [112]. It should be the ultimate goal of each CAD system employed for skin diagnosis, even though being a challenging task. It has already been an issue at the early days of automatic dermoscopy image classification, as classifiers used mostly low-level features adopted mainly from the computer vision literature [52] which typically lacked particular clinical meaning (it was particularly visible in case of texture features). As then the “traditional” feature extraction literature moved toward high-level, clinically oriented features [113]. However, modern algorithms, such as deep neural networks, became increasingly complex. As the European Union’s recent General Data Protection Regulation [114] aims at enforcing interpretability of models at the legal level, we expect that the deep learning literature will soon eventually become more understandable and transparent—the first such solutions were already proposed [115].

3.10 Teledermatology

Telemedicine (also called e-health) is a rapidly developing field of medicine, providing access to medical knowledge regardless location and time [116]. The Ericsson mobile report [117] estimates the number of smartphones around the world at roughly 7.9 billion, and according to the American Food and Drug Administration roughly 500 millions smartphone users around the world use e-health mobile applications. Hence, embedding CAD systems in smartphones seems like a low-cost method of tackling the problem of early melanoma detection. The use of such mobile technologies for skin diagnosis (teledermatology) could help patients avoid dermatology clinic visits and decrease the amount of biopsies, while for inexperienced physicians they could provide a valuable training tool.

In 2017 Apple’s App Store alone offered more than 45 mobile applications related to mole diagnosis. The majority of them provided only information on melanoma, roughly half enable the user to perform self-examination (to take photos of their lesions and monitor alterations over time using basic visual comparison), while merely four applications performed automatic melanoma risk assessment or lesion classification based on image analysis [118].

The availability of machine learning models capable of accurately classifying images taken in “laboratory” conditions is only the first step toward making these solutions accessible by general users. However, the process of developing

such a technology is more complicated than just deploying such a model in a smartphone—some important technical and ethical aspects must be addressed.

The two technical issues are hardware-related. Some specific dermoscopic features of a lesion are distinctly visible only at sufficient magnification or in appropriate illumination conditions. Although the majority of high-quality modern smartphone cameras are equipped with a high-resolution sensor which pixel of sufficiently small size, which allows the user to capture photos of high-quality, they typically lack a source of white light which would uniformly illuminate an examined lesion and a quality optical zoom [118]. These deficiencies of bare smartphones could be overcome by using either a conventional dermatoscope with a mobile phone case or a smartphone dermatoscope (the latter currently having similar features as the former ones). Moreover, as such mobile applications will certainly be often used in remote places such as rural areas, where no internet connection is available, they would require the application to operate offline. Solutions using complex ensembles of deep learning models might be difficult to deploy on smartphones due to lack of sufficient amount of operational memory and processing power to produce the results within the reasonable time period. However, at least some of these difficulties had been overcome and the first DNN-based mobile solutions were already deployed [119].

The main ethical issues are user's data privacy protection (i.e., how the application secures and handles user data), which is often not respected by researchers/developers [120], and the fact that using untested and uncertified (by a board of experts) hypothetical CAD systems giving false negative results may delay user's treatment and consequently even lead them to death. Only the authors of SkinVision and Lūbax made public the results of clinical evaluation of their algorithms [121,122], the methods achieved the accuracy 81% and 91%, sensitivity 73% and 90%, and specificity 83% and 91%, respectively. and out of 4 applications for automatic diagnosis, only two were certified by authorities [118].

A systematic review on the topic of teledermatology (in the context of diagnosing skin lesions) can be found in [85,118].

References

- [1] C. Garbe, U. Leiter, Melanoma epidemiology and trends, *Clin. Dermatol.* 27 (1) (2009) 3–9.
- [2] Cancer Facts & Figures, <<http://www.cancer.org/research/cancerfactsstatistics/cancer-factsfigures2016/index>>. American Cancer Society, 2016 (accessed 11.06.16).
- [3] Australian Bureau of Statistics, 3303.0 Causes of death, Australia 2015. <<http://www.abs.gov.au/Causes-of-Death>>, 2016 (accessed 11.30.16).
- [4] G. Argenziano et al., *Interactive Atlas of Dermoscopy*. Edra Medical Publishing and New Media, 2000. ISBN: 978-8-88645-730-9.
- [5] G. Argenziano et al., Dermoscopy of pigmented skin lesions: results of a consensus meeting via the Internet. *J Am Acad Dermatol* 48 (5) (2003) 679–693. Available from: <https://doi.org/10.1067/mjd.2003.281>.

- [6] Understanding the Artificial Intelligence Taxonomy and its Ecosystem. <<https://analytic-sindiamag.com/understanding-the-artificial-intelligence-taxonomy-and-its-ecosystem/>>, 2018 (accessed 12.06.19).
- [7] Y. Bengio, A. Courville, P. Vincent, Representation learning: a review and new perspectives, *IEEE Trans. Pattern Anal. Mach. Intell.* 35 (8) (2013) 1798–1828. issn: 1939–3539. Available from: <https://doi.org/10.1109/TPAMI.2013.50>.
- [8] A. Shrestha, A. Mahmood, Review of deep learning algorithms and architectures, *IEEE Access.* 7 (2019) 53040–53065. Available from: <https://doi.org/10.1109/ACCESS.2019.2912200>. issn: 2169–3536.
- [9] H.F. Nweke, et al., Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: state of the art and research challenges, *Expert. Syst. Appl.* 105 (2018) 233–261.
- [10] P. Tschandl, C. Rosendahl, H. Kittler, The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions, *Sci. data* (2018).
- [11] M.H. Jafari, et al., Extraction of skin lesions from non-dermoscopic images for surgical excision of melanoma, *Int. J. Computer Assist. Radiology Surg.* 12.6 (2017) 1021–1030.
- [12] H. Kittler, et al., Diagnostic accuracy of dermoscopy, *Lancet. Oncol.* 3 (3) (2002) 159–165.
- [13] E. Errichetti, G. Stinco, Dermoscopy in general dermatology: a practical overview, *Dermatol. Ther.* (2016).
- [14] D. Lipsker, Clinical examination and differential diagnosis of skin lesions, 2013.
- [15] R.K. Mlosek, S. Malinowska, Ultrasound image of the skin, apparatus and imaging basics, *J. Ultrason.* (2013).
- [16] B. Werner, Skin biopsy and its histopathologic analysis: Why? What for? How? Part I, *An. Bras. Dermatol.* 84 (4) (2009) 391–395.
- [17] U. Nischal, K. Nischal, U. Khopkar, Techniques of skin biopsy and practical considerations, *J. Cutan. Aesthet. Surg.* 1 (2) (2008) 107.
- [18] Skin biopsies: what you should expect. <<https://www.yalemedicine.org/stories/skin-biopsy/>> (accessed 12.01.19).
- [19] Skin biopsy. <<https://www.mayoclinic.org/tests-procedures/skin-biopsy/about/pac-20384634>> (accessed 12.01.19).
- [20] Skin biopsy. <<https://medlineplus.gov/lab-tests/skin-biopsy/>> (accessed 12.02.19).
- [21] Lecture 33: Skin cancer, photoprotection, & sunscreen. <<https://www.memorangapp.com/flashcards/47237/Lecture+33%3A+Skin+Cancer%2C+Photoprotection%2C+%26+Sunscreen/>> (accessed 12.01.19).
- [22] N.C. Codella et al. Skin lesion analysis toward melanoma detection: a challenge at the 2017 international symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC), in: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), IEEE, 2018, pp. 168–172.
- [23] M.E. Celebi, W.V. Stoecker, R.H. Moss, Advances in skin cancer image analysis, *Comput. Med. Imaging Graph.* 35 (2) (2011) 83–186.
- [24] K. Korotkov, R. Garcia, Computerized analysis of pigmented skin lesions: a review, *Artif. Intell. Med.* 56 (2) (2012) 69–90.
- [25] S. Pathan, K.G. Prabhu, P. Siddalingaswamy, Techniques and algorithms for computer aided diagnosis of pigmented skin lesions: a review, *Biomed. Signal. Process. Control.* 39 (2018) 237–262.

- [26] A.S. Ashour, et al., A hybrid dermoscopy images segmentation approach based on neutrosophic clustering and histogram estimation, *Appl. Soft Comput.* 69 (2018) 426–434.
- [27] A.S. Ashour, et al., A novel optimized neutrosophic k-means using genetic algorithm for skin lesion detection in dermoscopy images, *Signal, Image Video Process.* 12 (7) (2018) 1311–1318.
- [28] D. Gutman et al. Skin lesion analysis toward melanoma detection: a challenge at the international symposium on biomedical imaging (ISBI) 2016, hosted by the international skin imaging collaboration (ISIC), arXiv preprint arXiv:1605.01397, 2016.
- [29] Y. Yuan, M. Chao, Y.-C. Lo, Automatic skin lesion segmentation using deep fully convolutional networks with Jaccard distance, *IEEE Trans. Med. Imaging* 36 (9) (2017) 1876–1886. Available from: <https://doi.org/10.1109/TMI.2017.2695227>. issn: 0278-0062.
- [30] L. Yu, et al., Automated melanoma recognition in dermoscopy images via very deep residual networks, *IEEE Trans. Med. Imaging* 36 (4) (2017) 994–1004.
- [31] L. Bi, et al., Dermoscopic image segmentation via multistage fully convolutional networks, *IEEE Trans. Biomed. Eng.* 64 (9) (2017) 2065–2074.
- [32] T. Mendonça et al. PH 2-A dermoscopic image database for research and benchmarking, in: Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE, IEEE, 2013, pp. 5437–5440.
- [33] S. Vesal et al. A multi-task framework for skin lesion detection and segmentation, in: OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis. Springer, 2018, pp. 285–293.
- [34] M. Goyal et al. Skin lesion boundary segmentation with fully automated deep extreme cut methods, in: Medical Imaging 2019: Biomedical Applications in Molecular, Structural, and Functional Imaging. Vol. 10953. International Society for Optics and Photonics, 2019, 109530Q.
- [35] M. Goyal, M.H. Yap, Multi-class semantic segmentation of skin lesions via fully convolutional networks. arXiv preprint arXiv:1711.10449, 2017.
- [36] M.A. Al-masni, et al., Skin lesion segmentation in dermoscopy images via deep full resolution convolutional networks, *Computer Methods Prog. Biomed.* 162 (2018) 221–231.
- [37] A. Soudani, W. Barhoumi, An image-based segmentation recommender using crowdsourcing and transfer learning for skin lesion extraction, *Expert. Syst. Appl.* 118 (2019) 400–410.
- [38] V. Badrinarayanan, A. Kendall, R. Cipolla, Segnet: a deep convolutional encoder-decoder architecture for image segmentation, *IEEE Trans. Pattern Anal. Mach. Intell.* 39 (12) (2017) 2481–2495.
- [39] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2015, pp. 234–241.
- [40] M. Goyal, M.H. Yap, Region of interest detection in dermoscopic images for natural data-augmentation. arXiv preprint arXiv:1807.10711, 2018.
- [41] P. Kharazmi, et al., A computer-aided decision support system for detection and localization of cutaneous vasculature in dermoscopy images via deep feature learning, *J. Med. Syst.* 42 (2) (2018). 33 Jan. issn: 1573-689X. Available from: <https://doi.org/10.1007/s10916-017-0885-2>.

- [42] G. Betta et al. Dermoscopic image-analysis system: estimation of atypical pigment network and atypical vascular pattern, in: IEEE International Workshop on Medical Measurement and Applications, 2006, MeMea 2006, Apr. 2006, pp. 63–67. Available from: <https://doi.org/10.1109/MEMEA.2006.1644462>.
- [43] M.E. Celebi, T. Mendonca, J.S. Marques, *Dermoscopy Image Analysis*, CRC Press, 2015.
- [44] G. Fabbrocini, et al., Automatic diagnosis of melanoma based on the 7-point checklist, in: J. Scharcanski, M.E. Celebi (Eds.), *Computer Vision Techniques for the Diagnosis of Skin Cancer*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2014, pp. 71–107. isbn: 978-3-642-39608-3. Available from: https://doi.org/10.1007/978-3-642-39608-3_4.
- [45] P. Kharazmi, et al., Automated detection and segmentation of vascular structures of skin lesions seen in dermoscopy, with an application to basal cell carcinoma classification, *IEEE J. Biomed. Health Inform.* 21 (6) (2017) 1675–1684. Available from: <https://doi.org/10.1109/JBHI.2016.2637342>. Nov. issn: 2168-2194.
- [46] J. Jaworek-Korjakowska, A deep learning approach to vascular structure segmentation in dermoscopy color images, *BioMed. Res. Int.* 2018 (2018).
- [47] C. Serrano, B. Acha, Pattern analysis of dermoscopic images based on Markov random fields, *Pattern Recognit.* 42 (2009) 1052–1057.
- [48] M. Sadeghi, et al., Global pattern analysis and classification of dermoscopic images using textons, *Med. Imaging* (2012).
- [49] N. Codella, et al., Deep learning, sparse coding, and SVM for melanoma recognition in dermoscopy images, *International Workshop on Machine Learning in Medical Imaging*, Springer, 2015, pp. 118–126.
- [50] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, *Nature* 521 (7553) (2015) 436–444.
- [51] J.W. Choi, et al., Differentiation of benign pigmented skin lesions with the aid of computer image analysis: a novel approach, *Ann. Dermatol.* 25 (3) (2013) 340–347. Available from: <https://doi.org/10.5021/ad.2013.25.3.340>.
- [52] M.E. Celebi, et al., A methodological approach to the classification of dermoscopy images, *Computerized Med. Imaging Graph.* 31 (6) (2007) 362–373. Available from: <https://doi.org/10.1016/j.compmedimag.2007.01.003>. issn: 08956111. url.
- [53] M.E. Celebi, N. Codella, A. Halpern, Dermoscopy image analysis: overview and future directions, *IEEE J. Biomed. Health Inform.* 23 (2) (2019) 474–478. Available from: <https://doi.org/10.1109/JBHI.2019.2895803>.
- [54] L. Bi et al., Automatic skin lesion analysis using large-scale dermoscopy images and deep residual networks, 2017. arXiv: 1703.04197 cs.CV.
- [55] J. Kawahara, A. BenTaieb, G. Hamarneh, Deep features to classify skin lesions, in: 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI), 2016, pp. 1397–1400. Available from: <https://doi.org/10.1109/ISBI.2016.7493528>.
- [56] S.S. Han, et al., Classification of the clinical images for benign and malignant cutaneous tumors using a deep learning algorithm, *J. Investig. Dermatol.* 138 (7) (2018) 1529–1538. Available from: <https://doi.org/10.1016/j.jid.2018.01.028>. issn: 0022-202X.
- [57] R.H. Johr, Dermoscopy: alternative melanocytic algorithms—the ABCD rule of dermatoscopy, menzies scoring method, and 7-point checklist, *Clin. Dermatol.* 20 (3) (2002) 240–247. Available from: [https://doi.org/10.1016/S0738-081X\(02\)00236-5](https://doi.org/10.1016/S0738-081X(02)00236-5). issn: 0738-081X.

- [58] K. Wolff, et al., *Fitzpatrick's Color Atlas and Synopsis of Clinical Dermatology*, seventh ed., McGraw-Hill Medical, New York, 2013.
- [59] M. Filho, Z. Ma, J.M. Tavares, A review of the quantification and classification of pigmented skin lesions: from dedicated to hand-held devices, *J. Med. Syst.* 39 (11) (2015) 1–12. Available from: <https://doi.org/10.1007/s10916-015-0354-8>. issn: 0148-5598.
- [60] R.B. Oliveira, et al., Computational methods for the image segmentation of pigmented skin lesions: a review, *Comput. Methods Prog. Biomed.* 131 (2016) 127–141. issn: 0169-2607. Available from: <https://doi.org/10.1016/j.cmpb.2016.03.032>. url. <http://www.sciencedirect.com/science/article/pii/S0169260716303418>.
- [61] R.B. Oliveira, et al., Computational methods for pigmented skin lesion classification in images: review and future trends, *Neural Comput. Appl.* 29 (3) (2018) 613–636 issn: 0941-0643. Available from: <https://doi.org/10.1007/s00521-016-2482-6>.
- [62] M.E. Celebi, et al., A State-of-the-art survey on lesion border detection in dermoscopy images, *Dermoscopy Image Anal.* (2015) 97–129.
- [63] J. Jaworek-Korjakowska, P. Kłeczek, Automatic classification of specific melanocytic lesions using artificial intelligence, *BioMed. Res. Int.* 2016 (2016) 1–17.
- [64] V. Ng, D. Cheung, Measuring asymmetries of skin lesions, in: *Proceedings of IEEE. Vol. 5. International Conference on Systems, Man, and Cybernetics*, IEEE Press, 1997, pp. 4211–4216.
- [65] Z. She, Y. Liu, A. Damatoa, Combination of features from skin pattern and ABCD analysis for lesion classification, *Skin. Res. Technol.* 13 (1) (2007) 25–33.
- [66] B. Kusumoputro, A. Ariyanto, Neural network diagnosis of malignant skin cancers using principal component analysis as a preprocessor, in: *Proceedings of IEEE. Vol. 1. World congress on computational intelligence*, IEEE Press, 1998, pp. 310–315.
- [67] E. Claridge, et al., Shape analysis for classification of malignant melanoma, *J. Biomed. Eng.* 14 (3) (1992) 229–234.
- [68] Y.I. Cheng, et al., Skin lesion classification using relative color features, *Skin. Res. Technol.* 14 (1) (2008) 53–64.
- [69] K. Cheung, *Image processing for skin cancer detection: malignant melanoma recognition*. PhD thesis. University of Toronto, 1997.
- [70] T. Lee, D. McLean, A.M. Stella, Irregularity index: a new border irregularity measure for cutaneous melanocytic lesions, *Med. Image Anal.* 7 (1) (2003) 47–64.
- [71] T. Lee, E. Claridge, Predictive power of irregular border shapes for malignant melanomas, *Skin. Res. Technol.* 11 (1) (2005) 1–8.
- [72] Y. Chang, et al., A systematic heuristic approach for feature selection for melanoma discrimination using clinical images, *Skin. Res. Technol.* 11 (3) (2005) 165–178.
- [73] A.J. Round, A.W.G. Duller, P.J. Fish, Lesion classification using skin patterning, *Skin. Res. Technol.* 6 (4) (2000) 183–192.
- [74] Z. She, P. Fish, Analysis of skin line pattern for lesion classification, *Skin. Res. Technol.* 9 (1) (2003) 73–80.
- [75] R. Walvick et al. Classification of melanoma using wavelet-transform-based optimal feature set, in: *Proceedings of SPIE. Vol. 5370. Medical Imaging: Image Processing*, 2004, pp. 944–951.
- [76] W.V. Stoecker, C.-S. Chiang, R.H. Moss, Texture in skin images: comparison of three methods to determine smoothness, *Computerized Med. Imaging Graph.* 16 (3) (1992) 179–190.

- [77] Melanoma Foundation of New Zealand, About melanoma—key information, 2014 <<http://www.melanoma.org.nz/About-Melanoma/Key-Information/>> (accessed 01.10.20).
- [78] S.V. Deshabhoina, et al., Melanoma and seborrheic keratosis differentiation using texture features, *Skin. Res. Technol.* 9 (4) (2003) 348–356.
- [79] H. Mirzaalian, T.K. Lee, G. Hamarneh, Learning features for streak detection in dermoscopic color images using localized radial flux of principal intensity curvature, 2012 IEEE Workshop Math. Methods Biomed. Image Anal. (2012) 97–101. Available from: <https://doi.org/10.1109/MMBIA.2012.6164758>.
- [80] A. Madooei, et al., Automatic detection of blue-white veil by discrete color matching in dermoscopy images, in: K. Mori, et al. (Eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2013, pp. 453–460. isbn: 978-3-642-40760-4.
- [81] J. Jaworek-Korjakowska, et al., Automatic detection of blue-whitish veil as the primary dermoscopic feature, in: L. Rutkowski, et al. (Eds.), *Artificial Intelligence and Soft Computing*, Springer International Publishing, 2017, pp. 649–657. isbn: 978-3-319-59063-9.
- [82] T. Wadhawan et al., Implementation of the 7-point checklist for melanoma detection on smart handheld devices, in: *Conference Proceedings of the IEEE Engineering in Medicine and Biology Society*, 2011, pp. 3180–3183. Available from: <https://doi.org/10.1109/IEMBS.2011.6090866>.
- [83] H. Haenssle, et al., Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists, *Ann. Oncol.* 29 (8) (2018) 1836–1842. Available from: <https://doi.org/10.1093/annonc/mdy166>. issn: 0923-7534.
- [84] T.J. Brinker, et al., Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task, *Eur. J. Cancer* 113 (2019) 47–54. Available from: <https://doi.org/10.1016/j.ejca.2019.04.001>. issn: 0959-8049.
- [85] A.G.C. Pacheco, R.A. Krohling, Recent advances in deep learning applied to skin cancer detection, 2019. arXiv: 1912.03280 eess.IV.
- [86] T.J. Brinker, et al., Skin cancer classification using convolutional neural networks: systematic review, *J. Med. Internet Res.* 20 (10) (2018) e11936. Available from: <https://doi.org/10.2196/11936>. issn: 1438-8871.
- [87] V. Pomponiu, H. Nejati, N.-. Cheung, Deepmole: deep neural networks for skin mole lesion classification, in: 2016 IEEE International Conference on Image Processing (ICIP), 2016, pp. 2623–2627. Available from: <https://doi.org/10.1109/ICIP.2016.7532834>.
- [88] L. Ballerini, et al., A color and texture based hierarchical K-NN approach to the classification of non-melanoma skin lesions, in: M.E. Celebi, G. Schaefer (Eds.), *Color Medical Image Analysis*, Springer Netherlands, Dordrecht, 2013, pp. 63–86. isbn: 978-94-007-5389-1. Available from: https://doi.org/10.1007/978-94-007-5389-1_4.
- [89] O. Russakovsky et al., Imagenet large scale visual recognition challenge, preprint arXiv:1409.0575 (2014).
- [90] A. Menegola et al., Knowledge transfer for melanoma screening with deep learning, in: 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017). IEEE, 2017. isbn: 9781509011728. Available from: <https://doi.org/10.1109/ISBI.2017.7950523>.
- [91] A. Esteva, et al., Dermatologist-level classification of skin cancer with deep neural networks, *Nature* 542 (7639) (2017) 115–118. Available from: <https://doi.org/10.1038/nature21056>. issn: 1476-4687.

- [92] X. Sun, et al., A benchmark for automatic visual classification of clinical skin disease images, in: B. Leibe, et al. (Eds.), *Computer Vision—ECCV 2016*, Springer International Publishing, Cham, 2016, pp. 206–222. isbn: 978-3-319-46466-4.
- [93] A. Romero Lopez et al., Skin lesion classification from dermoscopic images using deep learning techniques, in: 2017 13th IASTED International Conference on Biomedical Engineering (BioMed), 2017, pp. 49–54. Available from: <https://doi.org/10.2316/P.2017.852-053>.
- [94] E. Nasr-Esfahani et al. Melanoma detection by analysis of clinical images using convolutional neural network, in: 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2016, pp. 1373–1376. Available from: <https://doi.org/10.1109/EMBC.2016.7590963>.
- [95] N.C.F. Codella, et al., Deep learning ensembles for melanoma recognition in dermoscopy images, *IBM J. Res. Dev.* 61 (4/5) (2017). 5:1–5:15. issn: 0018-8646. Available from: <https://doi.org/10.1147/JRD.2017.2708299>.
- [96] N. Gessert et al., Skin lesion diagnosis using ensembles, unscaled multi-crop evaluation and loss weighting, 2018. arXiv: 1808.01694 cs.CV.
- [97] F. Perez, S. Avila, E. Valle, Solo or ensemble? Choosing a CNN architecture for melanoma classification, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2019.
- [98] M.A. Marchetti, et al., Results of the 2016 international skin imaging collaboration international symposium on biomedical imaging challenge: comparison of the accuracy of computer algorithms to dermatologists for the diagnosis of melanoma from dermoscopic images, *J. Am. Acad. Dermatol.* 78 (2) (2018) 270–277. Available from: <https://doi.org/10.1016/j.jaad.2017.08.016>. e1. issn: 0190-9622.
- [99] J. Kawahara, G. Hamarneh, Multi-resolution-tract CNN with hybrid pretrained and skin-lesion trained layers, in: L. Wang, et al. (Eds.), *Machine Learning in Medical Imaging*, Springer International Publishing, Cham, 2016, pp. 164–171. isbn: 978-3-319-47157-0.
- [100] A.G. Pacheco, R.A. Krohling, The impact of patient clinical information on automated skin cancer detection, *Computers Biol. Med.* 116 (2020) 103545. Available from: <https://doi.org/10.1016/j.compbimed.2019.103545>. issn: 0010-4825.
- [101] Y. Liu et al., A deep learning system for differential diagnosis of skin diseases, 2019. arXiv: 1909.05382 eess.IV.
- [102] L. Faes, et al., Automated deep learning design for medical image classification by health-care professionals with no coding experience: a feasibility study, *Lancet Digital Health* 1 (5) (2019) e232–e242. Available from: [https://doi.org/10.1016/S2589-7500\(19\)30108-6](https://doi.org/10.1016/S2589-7500(19)30108-6). issn: 2589-7500.
- [103] S. Afifi, H. Gholamhosseini, R. Sinha, A system on chip for melanoma detection using FPGA-based SVM classifier, *Microprocess. Microsyst.—Embedded Hardw. Des.* 65 (2019) 57–68.
- [104] I. Chiuchisan, O. Geman, An approach of FPGA technology in skin lesion detection, in: 2018 International Conference and Exposition on Electrical and Power Engineering (EPE), 2018, pp. 0175–0178.
- [105] S. Afifi, H. Gholamhosseini, R. Sinha, Hardware acceleration of SVM-based classifier for melanoma images, in: *PSIVT Workshops*, 2015.
- [106] S. Afifi, H. Gholamhosseini, R. Sinha, A low-cost FPGA-based SVM classifier for melanoma detection, in: 2016 IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES), 2016, pp. 631–636.

- [107] S. Afifi, H. Gholamhosseini, R. Sinha, SVM classifier on chip for melanoma detection. in: 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2017, pp. 270–274.
- [108] S. Afifi, H. Gholamhosseini, R. Sinha, Dynamic hardware system for cascade SVM classification of melanoma, *Neural Comput. Appl.* (2018) 1–12.
- [109] J. Qiu et al. Going deeper with embedded FPGA platform for convolutional neural network, in: *FPGA '16*, 2016.
- [110] C.N. Vasconcelos, B.N. Vasconcelos, Experiments using deep learning for dermoscopy image analysis, *Pattern Recognit. Lett.* (2017). Available from: <https://doi.org/10.1016/j.patrec.2017.11.005>. issn: 0167-8655.
- [111] F. Perez, et al., Data augmentation for skin lesion analysis, in: D. Stoyanov, et al. (Eds.), *OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis*, Springer International Publishing, Cham, 2018, pp. 303–311. isbn: 978-3-030-01201-4.
- [112] G.A. Zakhem, C.C. Motosko, R.S. Ho, How should artificial intelligence screen for skin cancer and deliver diagnostic predictions to patients, *JAMA Dermatol.* 154 (12) (2018) 1383–1384. Available from: <https://doi.org/10.1001/jamadermatol.2018.2714>. Dec. issn: 2168-6068.
- [113] C. Barata, M.E. Celebi, J.S. Marques, Development of a clinically oriented system for melanoma diagnosis, *Pattern Recognit.* 69 (2017) 270–285. Available from: <https://doi.org/10.1016/j.patcog.2017.04.023>. issn: 0031-3203.
- [114] B. Goodman, S. Flaxman, European Union regulations on algorithmic decision-making and a “Right to Explanation”, *AI Mag.* 38 (3) (2017) 50–57. Available from: <https://doi.org/10.1609/aimag.v38i3.2741>.
- [115] J. Kawahara, G. Hamarneh, Fully convolutional neural networks to detect clinical dermoscopic features, *IEEE J. Biomed. Health Inform.* 23 (2) (2019) 578–585. Available from: <https://doi.org/10.1109/jbhi.2018.2831680>.
- [116] E.M.T. Wurm, H.P. Soyer, A.C. Smith, *Telemedicine in Dermatology*, first ed., Springer-Verlag, Berlin Heidelberg, 2012. isbn: 3642208002.
- [117] E. Inc. Ericsson Mobility Report November 2019, Ericsson Inc., 2019. <<https://www.ericsson.com/en/mobility-report/reports/november-2019>> (accessed 28.01.20).
- [118] J. Jaworek-Korjakowska, P. Kleczek, eSkin: study on the smartphone application for early detection of malignant melanoma, *Wirel. Commun. Mob. Comput.* 2018 (2018). Available from: <https://doi.org/10.1155/2018/5767360>.
- [119] A. Ech-Cherif, M. Misbhauddin, M. Ech-Cherif, Deep neural network based mobile dermoscopy application for triaging skin cancer detection, in: 2019 2nd International Conference on Computer Applications Information Security (ICCAIS), 2019, pp. 1–6. doi:10.1109/CAIS.2019.8769517.
- [120] E. Chao, C.K. Meenan, L.K. Ferris, Smartphone-based applications for skin monitoring and melanoma detection, *Dermatol. Clin.* 35 (4) (2017). Available from: <https://doi.org/10.1016/j.det.2017.06.014>. Non-invasive technologies for the diagnosis and management of skin cancer, pp. 551–557. ISSN: 0733-8635.
- [121] T. Maier, et al., Accuracy of a smartphone application using fractal image analysis of pigmented moles compared to clinical diagnosis and histological result, *J. Eur. Acad. Dermatol. Venereol.* 29 (2014) 4. Available from: <https://doi.org/10.1111/jdv.12648>.
- [122] R.H. Chen, et al., Validation of a skin-lesion image-matching algorithm based on computer vision technology, *Telemed. e-Health* 22 (1) (2016) 45–50. Available from: <https://doi.org/10.1089/tmj.2014.0249>.

This page intentionally left blank

An accurate deep learning-based computer-aided diagnosis system for early diagnosis of prostate cancer

Islam R. Abdelmaksoud^{1,2}, Ahmed Shalaby¹, Mohammed Ghazal³,
Mohammed Elmogy², Ahmed AbouElfetouh², Ali Mahmoud¹ and Ayman S. El-Baz^{1,4}

¹University of Louisville, Louisville, KY, United States

²Faculty of Computers and Information, Mansoura University, Mansoura, Egypt

³Electrical, Computer and Biomedical Engineering Department, Abu Dhabi University, Abu Dhabi, United Arab Emirates

⁴University of Louisville at Alamein International University (UofL-AIU), New Alamein City, Egypt

4.1 Introduction

Prostate cancer has the second mortality rate after lung cancer among men in the US. About 14% of American men are subjected to prostate cancer during his lifetime. The chance of diagnosis of prostate cancer increases for older men. In 2020, the number of men who were expected to be diagnosed with prostate cancer in the US was about 191,930. The expected number of deaths in this year was about 33,330 [1]. The average age of the cases diagnosed with prostate cancer is 66 years. The 5-year survival rate is about 100% for local prostate cancer. This rate drops to 31% when the cancer spread to other organs of the body [2]. These numbers indicate the importance of early diagnosis of prostate cancer.

Currently, there are two initial techniques for diagnosing prostate cancer. The first one is digital rectal exam (DRE). In this exam, a physician checks the prostate gland with his gloved fingers through the rectum to find out any tumors. Although, this exam has a low cost and can be performed easily, DRE can detect only large tumors, and the adequacy of the physician who performs the exam is a key factor of its accuracy [3]. The second technique for diagnosing prostate cancer is prostate specific antigen (PSA) screening. This technique is a blood test that measures the concentration of PSA in the blood. Both healthy and cancerous prostates generate PSA in the blood. However, the generated amount of PSA increases in case of cancerous prostates and therefore can be used as indicator for prostate cancer. However, PSA screening has high false positive rates because the increased amounts of PSA in the blood may be caused by different reasons such

as, prostatitis and hyperplasia [4]. Although, PSA screening is an initial technique for diagnosing prostate cancer, its introduction led to reducing the mortality rate of prostate cancer by 20% [5]. The current objective of the diagnostic techniques of prostate cancer is to increase their accuracies of clinical significance tumors while decreasing the over-diagnosis of indolent tumors [6]. The accuracy of the two previously mentioned techniques of diagnosing prostate cancer is not high. Therefore, these techniques are employed to determine who should undergo biopsy. Biopsy, typically guided by ultrasound and recently by magnetic resonance imaging (MRI), is the definitive technique for diagnosing prostate cancer. However, biopsy has its own shortcomings. It is an invasive operation with high cost. Moreover, biopsy can miss more than 30% of malignant prostate tumors [7].

Recently, different MRI modalities have resulted in accurate diagnosis of prostate cancer without the drawbacks of invasive techniques [8,9]. The most common MRI modalities used by the current computer-aided diagnosis (CAD) systems for prostate cancer diagnosis are the following: T2-weighted, diffusion-weighted, dynamic contrast-enhanced (DCE) MRI, and magnetic resonance spectroscopy [10]. Diffusion-weighted imaging (DWI) is a functional MRI modality that measures the random motion of water molecules within the tissues to generate contrast in Mr images. An apparent diffusion coefficient (ADC) image eliminates the T2-weighting inherent to traditional DWI to show diffusion more specifically than traditional DWI. In the literature, a number of CAD systems have used DWI alone or in accompanied with other modalities to diagnose prostate cancer [11]. For example, Firjani et al. [12] proposed a K-nearest neighbor-based system for diagnosing prostate cancer from DWI. They used three intensity features from the prostate region to discriminate between benign and malignant lesions. Litjens et al. [13] proposed a support vector machine (SVM)-based multimodal system that employed T2-weighted MRI, DWI, and DCE-MRI. The used features of their system were appearance, anatomy, and pharmacokinetic parameters. Vos et al. [14] proposed a linear discriminant analysis-based multimodal system using the same modalities of the system by Litjens et al. [13]. The used features of their system were textural features in addition to a collection of ADC maps. Trigui et al. [15] proposed a multimodal system using the four common MRI modalities listed earlier. They evaluated the accuracy of two classification models which are SVM and random forest. Shoshana et al. [16] proposed a 3 Tesla (T) multimodal system to diagnose prostate cancer in both the transition and peripheral zones using radiomic features. The used radiomic features included co-occurrence, wavelet and kinetic features. A feature selection technique was employed to identify a small set of features with the highest accuracy. The best area under the curve achieved by their system was 0.71. Khalvati et al. [6] proposed a multimodal system that computed textural features from four MRI modalities. The best features with the best performance at each individual modality were chosen to form a final group of optimized features to be used for classification. The highest accuracy achieved was 88%. Yohannes et al. [17] compared the performance of convolutional neural networks (CNNs) that can learn the features automatically

with the performance of traditional techniques that use handcrafted features. They found that the use of CNNs improved the accuracy by 9%. They achieved a detection rate of 89% in case of CNN in comparison to 80% detection rate in case of traditional techniques. CNNs were used by Kwak et al. [18] to detect the aggressiveness of prostate cancer from histological images.

The main drawback of the most proposed DWI-based systems for diagnosing prostate cancer is that the classification is performed at a single b – value. The used b -value is not consistent between the different systems. However, the proposed system has two advantages. First, the accuracy of prostate cancer diagnosis is tested at multiple b -values as the used datasets were acquired at seven b – values. Second, the used datasets were acquired at two different magnetic field strengths which are 1.5 and 3 T. therefore, the reported accuracy is not biased to the magnetic field strength of the used scanner. The rest of this chapter is organized as follows: [Section 4.2](#) explains in details the proposed method. [Section 4.3](#) explains the conducted experiments and their results. [Section 4.4](#) summarizes the proposed system and list the various conclusions.

4.2 Methods

In this chapter, an automated system diagnosing prostate cancer using CNN is proposed. The DWI datasets used for developing that system were collected from 45 subjects (20 benign 25 malignant) at seven b -values (100, 200, . . . , 700 s/mm²). Each subject contains on average 25 cross-sectional DW images at each b -value. [Fig. 4.1](#) shows the three main steps of the proposed system. The first step is prostate segmentation. The purpose of this step is to separate the prostate region from the background to reduce the complexities and expedite the following steps by

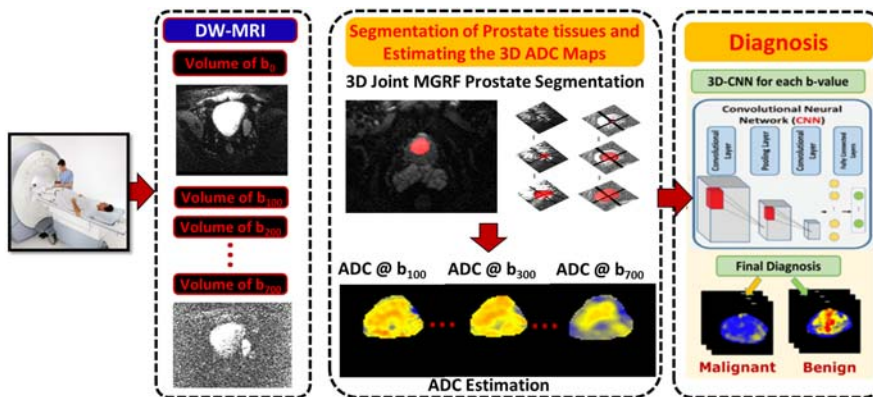


FIGURE 4.1

Three main steps of the proposed system.

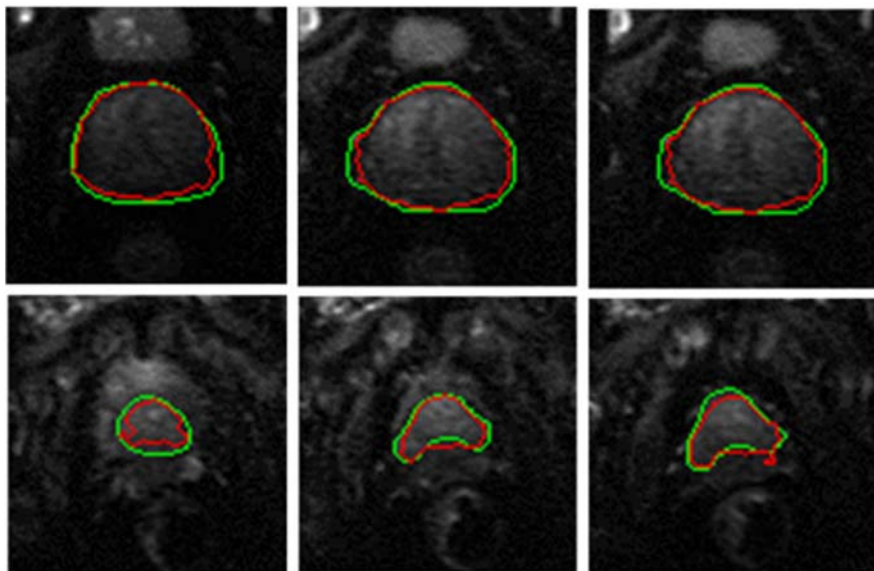


FIGURE 4.2

Segmentation results of two subjects. The green and red contours represent the ground truth and the model segmentations, respectively.

applying these steps on the region of interest. Prostate segmentation was performed using a level-set model that employed three features for enhanced accuracy. The employed features were intensity, shape prior, and spatial features. These three features were fused using a nonnegative matrix factorization (NMF) approach. More details about the segmentation model can be found in our previous work [19].

Fig. 4.2 shows the segmentation results of two subjects. The second step after prostate segmentation is to extract features from the segmented prostate regions that can discriminate between malignant and benign subjects. In the proposed system, ADC maps computed using the difference between two DW images are used as discriminating features. The third step of the proposed system is to feed the computed features of the second step to a classification model to identify the input subject as either malignant or benign. The used classification model is based on CNN that is trained using the ADC maps computed in the second step. The details of the processes of computing the ADC maps and training the CNN-based model are mentioned in the following subsections.

4.2.1 Feature Extraction

DWI is considered as one of the recent MRI modalities employed in the diagnosis of prostate cancer. The basic idea of DWI to generate contrast DW images is to

measure the freedom with which water molecules disseminate within tissues. The nature of this dissemination provides information about the cellularity of the tissues and the space inside and between the different cells of these tissues. This information can be used to recognize the presence and aggressiveness of cancer. DWI has some advantages over DCE-MRI. For example, the time required to generate DW images is short because DWI does not employ any sort of contrast materials. Although, these contrast materials result in high quality images, they cause harmful complications for patients, especially those who have kidney problems. DWI concentrates on the spread of water between cells to evaluate whether a certain region has abnormal restricted spread. This restricted spread of water is typically triggered by the decrease in the volume of space between cells. DWI uses the distinctions in the motion of water molecules within tissues to generate images. This motion has a random nature that is positively correlated with the signal loss of DWI. The signal loss is defined by [20]:

$$S_d \sim e^{-b \times ADC} \quad (4.1)$$

where b is a variable related to the magnitude and timing of gradient pulses. DWI scanners use gradient pulses instead of steady state gradients as gradient pulses have better sensitivity to the spread of fluids [21].

The signal loss defined in Eq. (4.1) determines the relationship between the intensities of the pixels of a diffusion image acquired at some b -value and the intensities of the pixels of a diffusion image acquired with the absence of diffusion ($b = 0 \text{ s/mm}^2$). The following equation define these intensities:

$$S_d = S_0 \times e^{-b \times ADC} \quad (4.2)$$

As the quality of DWI is typically not high [22], the quantitative maps computed from DWI, known as ADC maps, are used to diagnose prostate cancer. ADC maps can be used for this purpose because cancerous tissues have restricted diffusions and therefore lower ADC values in comparison to higher ADC values of healthy tissues. The ADC map at a certain b -value is computed by measuring the distinction between a diffusion image acquired at that b -value and the counterpart image acquired at b_0 ($b = 0 \text{ s/mm}^2$) using the following equation:

$$ADC = -\frac{\ln S_b - \ln S_0}{b} \quad (4.3)$$

In Eq. (4.3), the natural logarithm is applied to the results of dividing the pixels' intensities of the b image by the pixels' intensities of b_0 image to eliminate the T2-effect. Fig. 4.3 shows an example of two subjects (one benign and one malignant) where it represents the ADC color maps at the seven b -values for each subject. After the ADC maps are calculated, they are used to train a CNN-based classification model. Recent study shows that incorporating handcrafted features can enhance the performance of CNN [22]. To take the advantage of idea in our system, we decided to use ADC maps as inputs to CNN classifier instead of the DWI. This process is detailed in the following subsection and in Section 4.3.

The extracted ADC maps for two different Subjects

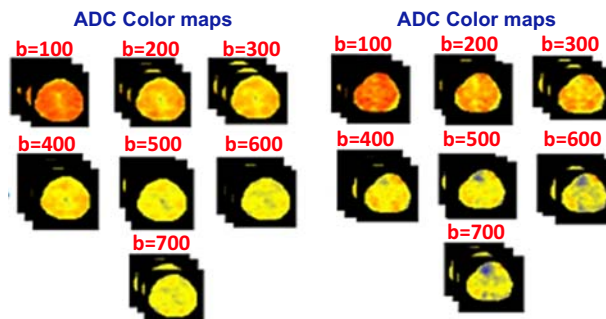


FIGURE 4.3

An example of the calculation of ADC maps for two subjects.

4.2.2 CNN-based classification

Deep learning improves the accuracy of classification over conventional neural networks using a higher number of network layers. Augmenting the number of layers in a network enables it to learn high levels of abstractions as the first layers learn primitive features while the following deep layers use these primitive features to learn abstract high level features [23]. Deep learning has resulted in many successes in multiple domains. For example, deep learning has been used as the domain of analyzing medical imaging to diagnose different kinds of diseases [23]. In the proposed system, one of the most dominant deep learning techniques (CNN) is used for early diagnosis of prostate cancer. When conventional classifiers are used, the good design of the handcrafted features that they employ is a key factor in their accuracies. The effect of handcrafted features vanishes in the case of CNNs as CNNs can learn the discriminating features automatically.

CNNs differ from conventional neural networks in three main aspects. The first difference is in the type of inputs. CNNs take as input images or volume of images, whereas conventional neural networks accept only vectors. The transformation of images to vectors to be suitable for conventional neural networks led to the loss of useful information, which in turn affects negatively their accuracies. The second difference is in weight sharing. The number of parameters of a CNN is much lower than the number of parameters of a similar conventional neural network as conventional networks do not support weight sharing. The third difference is in the types of supported layers. For example, CNNs support pooling layers which are not found in conventional neural networks. CNN is formed by concatenating the following types of layers.

Convolutional layers, pooling layers, and fully-connected layers are three main types of layers that form CNNs. Convolutional layers apply a set of filters on their inputs. Each filter results in a feature map. These feature maps are concatenated to produce the output volumes. A non-linearity function, such as

rectifying linear unit (ReLU), is applied pixelwise on these feature maps. ReLU is defined using this equation:

$$f(x) = \max(0, x) \quad (4.4)$$

The advantage of ReLU activation is that its convergence time is shorter than other famous activations like sigmoid. Pooling layers have the ability to decrease the width and height of feature maps. They achieve this decrease by representing a group of values in a certain window by a single value. This single value is typically either the average or the maximum value of these values. Therefore, the common types of pooling are either average pooling or maxpooling. Fully-connected layers are the type of layers that form conventional neural networks. Each neuron of a fully-connected layer has direct connection to all neurons of the next layer.

The CNN of the proposed system has a total of six convolutional layers. The first convolutional layer has 64 filter of size 7×7 . The second convolutional layer has 64 filters of size 3×3 . Each of the remaining four convolutional layers has 128 filters of size 3×3 . Three layers which are batch normalization, ReLU, and pooling layers follow each convolutional layer. All the used pooling layers are maxpooling except for the final pooling layer before the fully-connected layer. This final pooling layer is an average pooling layer. The advantages of using batch normalization layers is that they enable CNNs to employ high learning rates which lead to reducing the training time of the network [24]. The output classification probability is obtained by a softmax layer concatenated to the fully-connected layer.

4.3 Experimental results

Experiments were performed using DWI datasets acquired from 45 subjects at two different magnetic field strengths, which are 1.5 and 3 T. Two main experiments were performed to evaluate the accuracy of CNN and to compare its accuracy to the accuracy of SVM. SVM introduced the concept of hyperplanes with the maximum margins between the two output classes. To ensure that the solution found by SVM is unique, SVM use convex loss function [25]. The rationale behind comparing the proposed CNN to SVM is that SVMs have good performance for input data of high dimension which is the case of the input data to the proposed system.

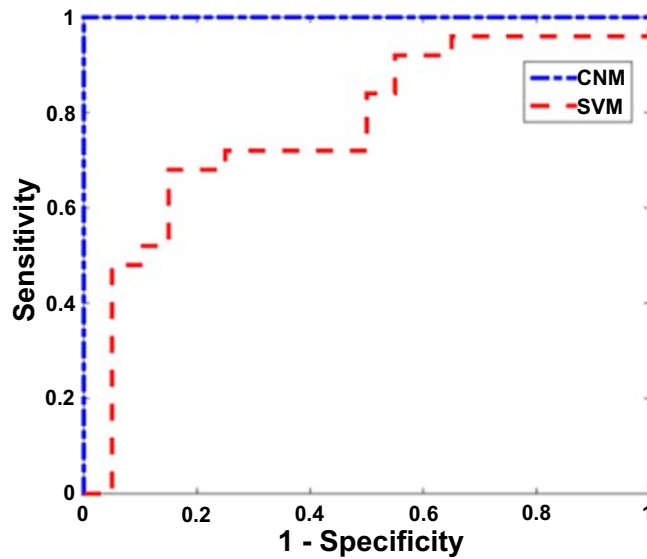
In the first experiment, the 45 ADC maps computed in the second step of our system at the first b-value ($b\text{-value} = 100 \text{ s/mm}^2$) is used to train the CNN in a “leave one subject out” (LOSO) scenario. A similar experiment is performed for each of the remaining b – values (200, 300, . . . , 600 s/mm^2). Table 4.1 lists the different performance measures at each b-value. The results in that table reflects the high performance of CNN.

Table 4.1 Performance measures of CNN.

b-Value (s/mm ²)	Accuracy (%)	Sensitivity (%)	Specificity (%)
100	95.6	100.0	90.0
200	91.1	100.0	80.0
300	97.8	100.0	95.0
400	91.1	100.0	80.0
500	97.8	100.0	95.0
600	91.1	100.0	80.0
700	91.1	100.0	80.0

Table 4.2 Performance measures of SVM.

b-Value (s/mm ²)	Accuracy (%)	Sensitivity	Specificity
100	71.1	72.0	70.0
200	60.0	68.0	50.0
300	57.8	68.0	45.0
400	62.2	68.0	55.0
500	53.3	56.0	50.0
600	55.6	56.0	55.0
700	60.0	68.0	50.0

**FIGURE 4.4**

ROC curve for CNN and SVM at b-value = 100 s/mm².

In the second experiment, the 45 computed ADC maps at the seven b-values is used to train SVM in an LOSO scenario. Table 4.2 lists the different performance measures of SVM at each b-value.

A quick analysis of the results in Tables 4.1 and 4.2 indicates that the accuracy of the CNN is high at the different b-values even if the DWI datasets are acquired at two magnetic field strengths (1.5 and 3 T). This means CNN can effectively classify data acquired with different acquisition systems. However, the use of datasets acquired by different scanners has negative effect on the accuracy of SVM.

Fig. 4.4 shows the receiver operating characteristic (ROC) curve for both the CNN and SVM at a single b-value ($b\text{-value} = 100 \text{ s/mm}^2$).

4.4 Conclusion

In this chapter, an automated system for diagnosing prostate cancer from DWI using CNN is proposed. The developed system consists of three main steps. First, prostate segmentation using a level-set model to delineate the region of interest (prostate). The evolution of the level set is guided by a set of information fused using NMF. Second, the ADC maps of the prostate region are computed at each b-value and used as inputs to the classifier of the final step. Finally, a CNN trained using these ADC maps to distinguish malignant subjects from benign ones. Although the used DWI datasets in the proposed system were acquired at two magnetic field strengths, the accuracy of CNN is high and is not affected negatively by this change in magnetic field strength.

In addition to the prostate cancer [26–28], this work could also be applied to various other applications in medical imaging, such as the kidney [29–56], the heart [57–73], the lung [74–121], the brain [122–143], the vascular system [144–153], the retina [154–158], the bladder [159] and injury prediction [160] as well as several non-medical applications [161–166].

References

- [1] American Cancer Society, Key statistics for prostate cancer. <http://www.cancer.org/cancer/prostate-cancer/about/key-statistics.html>.
- [2] National Cancer Institute, Cancer of the prostate—SEER stat fact sheets. Available. <https://seer.cancer.gov/statfacts/html/prost.html>.
- [3] K. Mistry, G. Cable, *Meta-analysis of prostate-specific antigen and digital rectal examination as screening tests for prostate carcinoma*, J. Am. Board. Family Pract. 6 (2) (2003) 95–101.
- [4] J. Hugosson, et al., *Mortality results from the goteborg randomised population-based prostate-cancer screening trial*, Lancet Oncol. 11 (8) (2010) 725–732.

- [5] F.H. Schroder, et al., Prostate-cancer mortality at 11 years of follow-up, *N. Engl. J. Med.* 366 (11) (2012) 981–990.
- [6] F. Khalvati, A. Wong, M.A. Haider, Automated prostate cancer detection via comprehensive multi-parametric magnetic resonance imaging texture feature models, *BMC Med. Imaging* 15 (1) (2015) 27.
- [7] A. Taira, et al., Performance of transperineal template-guided mapping biopsy in detecting prostate cancer in the initial and repeat biopsy setting, *Prostate Cancer Prostatic Dis.* 13 (1) (2010) 71.
- [8] J.L. Speight, M. Roach III, Advances in the treatment of localized prostate cancer: the role of anatomic and functional imaging in men managed with radiotherapy, *J. Clin. Oncol.* 25 (8) (2007) 987–995.
- [9] A. Thon, et al., Computer aided detection in prostate cancer diagnostics: a promising alternative to biopsy? a retrospective study from 104 lesions with histological ground truth, *PLoS One* 12 (10) (2017) e0185995.
- [10] S. Wang, et al., Computer aided-diagnosis of prostate cancer on multiparametric MRI: a technical review of current research, *BioMed. Res. Int.* (2014).
- [11] G. Lematre, et al., Computer-aided detection and diagnosis for prostate cancer based on mono and multi-parametric MRI: a review, *Computers Biol. Med.* 60 (2015) 8–31.
- [12] A. Firjani, et al., A diffusion-weighted imaging based diagnostic system for early detection of prostate cancer, *J. Biomed. Sci. Eng.* 6 (03) (2013) 346.
- [13] G. Litjens, et al., Automatic computer aided detection of abnormalities in multi-parametric prostate MRI79630T SPIE Medical Imaging, International Society for Optics and Photonics, 2011p. 79630T.
- [14] P. Vos, et al., Automatic computer-aided detection of prostate cancer based on multiparametric magnetic resonance image analysis, *Phys. Med. Biol.* 57 (6) (2012) 1527.
- [15] R. Trigui, et al., A classification approach to prostate cancer localization in 3T multi-parametric MRI, in: *Advanced Technologies for Signal and Image Processing (ATSIP), 2016 2nd International Conference on.* IEEE, 2016, pp. 113–118.
- [16] S.B. Ginsburg, et al., Radiomic features for prostate cancer detection on MRI differ between the transition and peripheral zones: preliminary findings from a multi-institutional study, *J. Magn. Reson. Imaging* 46 (1) (2017) 184–193.
- [17] Y.K. Tsehay, et al., Convolutional neural network based deep-learning architecture for prostate cancer detection on multiparametric magnetic resonance images 1013405 SPIE Medical Imaging, International Society for Optics and Photonics, 2017p. 1013405.
- [18] J.T. Kwak, S.M. Hewitt, Nuclear architecture analysis of prostate cancer via convolutional neural networks, *IEEE Access.* 5 (2017) 18526–18533.
- [19] P. McClure, et al., A novel NMF guided level-set for DWI prostate segmentation, *J. Computer Sci. Syst. Biol.* 7 (6) (2014) 209–216.
- [20] T.A. Huisman, Diffusion-weighted imaging: basic concepts and application in cerebral stroke and head trauma, *Eur. Radiol.* 13 (10) (2003) 2283–2297.
- [21] J. Hrabe, et al., Principles and limitations of NMR diffusion measurements, *J. Med. Phys./Assoc. Med. Phys. India* 32 (1) (2007) 34.
- [22] Y.J. Choi, et al., Functional MR imaging of prostate cancer, *Radiographics* 27 (1) (2007) 63–75.

- [23] H. Greenspan, et al., Guest editorial deep learning in medical imaging: overview and future promise of an exciting new technique, *IEEE Trans. Med. Imaging* 35 (5) (2016) 1153–1159.
- [24] S. Ioffe, C. Szegedy, Batch normalization: accelerating deep network training by reducing internal covariate shift, in: *International Conference on Machine Learning*, 2015, pp. 448–456.
- [25] E. Niaf, et al., Kernel-based learning from both qualitative and quantitative labels: application to prostate cancer diagnosis based on multiparametric MR imaging, *IEEE Trans. Image Pro-cessing* 23 (3) (2014) 979–991.
- [26] I. Reda, M. Ghazal, A. Shalaby, M. Elmogy, A. AbouEl-Fetouh, B.O. Ayinde et al., A novel adcs-based cnn classification system for precise diagnosis of prostate cancer, in: *2018 24th International Conference on Pattern Recognition (ICPR)*. IEEE, 2018, pp. 3923–3928.
- [27] I. Reda, A. Khalil, M. Elmogy, A. Abou El-Fetouh, A. Shalaby, M. Abou El-Ghar, et al., Deep learning role in early diagnosis of prostate cancer, *Technol. cancer Res. Treat.* 17 (2018) 1533034618775530.
- [28] I. Reda, B.O. Ayinde, M. Elmogy, A. Shalaby, M. El-Melegy, M.A. El-Ghar, et al., A new CNN-based system for early diagnosis of prostate cancer, in: *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. IEEE, 2018, pp. 207–210.
- [29] A.S. Chowdhury, R. Roy, S. Bose, F.K.A. Elnakib, A. El-Baz, Non-rigid biomedical image registration using graph cuts with a novel data term, in: *Proceedings of IEEE International Symposium on Biomedical Imaging: From Nano to Macro, (ISBI'12)*, Barcelona, Spain, 2–5 May 2012, pp. 446–449.
- [30] A. El-Baz, A.A. Farag, S.E. Yuksel, M.E. El-Ghar, T.A. Eldiasty, M.A. Ghoneim, *Application of deformable models for the detection of acute renal rejection, Deformable models*, Springer, New York, NY, 2007, pp. 293–333.
- [31] A. El-Baz, A. Farag, R. Fahmi, S. Yuksel, M.A. El-Ghar, T. Eldiasty, Image analysis of renal DCE MRI for the detection of acute renal rejection, in: *Proceedings of IAPR International Conference on Pattern Recognition (ICPR'06)*, Hong Kong, 2006, pp. 822–825.
- [32] A. El-Baz, A. Farag, R. Fahmi, S. Yuksel, W. Miller, M.A. El-Ghar, et al., A new CAD system for the evaluation of kidney diseases using DCE-MRI, in: *Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention, (MICCAI'08)*, Copenhagen, Denmark, 1–6 October 2006, pp. 446–453.
- [33] A. El-Baz, G. Gimel'farb, M.A. El-Ghar, A novel image analysis approach for accurate identification of acute renal rejection, in: *Proceedings of IEEE International Conference on Image Processing, (ICIP'08)*, San Diego, California, USA, 12–15 October 2008, pp. 1812–1815.
- [34] A. El-Baz, G. Gimel'farb, M.A. El-Ghar, Image analysis approach for identification of renal transplant rejection, in: *Proceedings of IAPR International Conference on Pattern Recognition, (ICPR'08)*, Tampa, Florida, USA, 2008, pp. 1–4.
- [35] A. El-Baz, G. Gimel'farb, M.A. El-Ghar, New motion correction models for automatic identification of renal transplant rejection, in: *Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention, (MICCAI'07)*, Brisbane, Australia, 29 October–2 November 2007, pp. 235–243.

- [36] A. Farag, A. El-Baz, S. Yuksel, M.A. El-Ghar, T. Eldiasty, A framework for the detection of acute rejection with dynamic contrast enhanced magnetic resonance imaging, in: *Proceedings of IEEE International Symposium on Biomedical Imaging: From Nano to Macro, (ISBI'06)*, Arlington, Virginia, USA, 2006, pp. 418–421.
- [37] F. Khalifa, G.M. Beache, M.A. El-Ghar, T. El-Diasty, G. Gimel'farb, M. Kong, et al., Dynamic contrast-enhanced MRI- based early detection of acute renal transplant rejection, *IEEE Trans. Med. Imaging* 32 (10) (2013) 1910–1927.
- [38] F. Khalifa, A. El-Baz, G. Gimel'farb, M.A. El-Ghar, Non-invasive image-based approach for early detection of acute renal rejection, in: *Proceedings of International Conference Medical Image Computing and Computer-Assisted Intervention, (MICCAI'10)*, Beijing, China, 20–24 September 2010, pp. 10–18.
- [39] F. Khalifa, A. El-Baz, G. Gimel'farb, R. Ouseph, M.A. El-Ghar, Shape-appearance guided level-set deformable model for image segmentation, in: *Proceedings of IAPR International Conference on Pattern Recognition, (ICPR'10)*, Istanbul, Turkey, 23–26 August 2010, pp. 4581–4584.
- [40] F. Khalifa, M.A. El-Ghar, B. Abdollahi, H. Frieboes, T. El-Diasty, A. El-Baz, A comprehensive non-invasive framework for automated evaluation of acute renal transplant rejection using DCE-MRI, *NMR Biomed.* 26 (11) (2013) 1460–1470.
- [41] F. Khalifa, M.A. El-Ghar, B. Abdollahi, H.B. Frieboes, T. El-Diasty, A. El-Baz, Dynamic contrast-enhanced MRI-based early detection of acute renal transplant rejection, in: *2014 Annual Scientific Meeting and Educational Course Brochure of the Society of Abdominal Radiology, (SAR'14)*, Boca Raton, Florida, 23–28 March 2014, p. CID: 1855912.
- [42] F. Khalifa, A. Elnakib, G.M. Beache, G. Gimel'farb, M.A. El-Ghar, G. Sokhadze, et al., 3D kidney segmentation from CT images using a level set approach guided by a novel stochastic speed function, in: *Proceedings of International Conference Medical Image Computing and Computer-Assisted Intervention, (MICCAI'11)*, Toronto, Canada, 18–22 September 2011, pp. 587–594.
- [43] F. Khalifa, G. Gimel'farb, M.A. El-Ghar, G. Sokhadze, S. Manning, P. McClure, et al., A new deformable model-based segmentation approach for accurate extraction of the kidney from abdominal CT images, in: *Proceedings of IEEE International Conference on Image Processing, (ICIP'11)*, Brussels, Belgium, 11–14 September 2011, pp. 3393–3396.
- [44] M. Mostapha, F. Khalifa, A. Alansary, A. Soliman, J. Suri, A. El-Baz, Computer-aided diagnosis systems for acute renal transplant rejection: challenges and methodologies, in: A. El-Baz, L. saba, J. Suri (Eds.), *Abdomen and Thoracic Imaging*, Springer, 2014, pp. 1–35.
- [45] M. Shehata, F. Khalifa, E. Hollis, A. Soliman, E. Hosseini-Asl, M.A. El-Ghar, et al., A new non-invasive approach for early classification of renal rejection types using diffusion-weighted MRI, in: *IEEE International Conference on Image Processing (ICIP)*, 2016. IEEE, 2016, pp. 136–140.
- [46] F. Khalifa, A. Soliman, A. Takieldeeen, M. Shehata, M. Mostapha, A. Shaffie, et al., Kidney segmentation from CT images using a 3D NMF-guided active contour model, in: *IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, 2016. IEEE, 2016, pp. 432–435.
- [47] M. Shehata, F. Khalifa, A. Soliman, A. Takieldeeen, M.A. El-Ghar, A. Shaffie, et al., 3d diffusion mri-based cad system for early diagnosis of acute renal rejection, in:

- Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium on. IEEE, 2016, pp. 1177–1180.
- [48] M. Shehata, F. Khalifa, A. Soliman, R. Alrefai, M.A. El-Ghar, A.C. Dwyer, et al., A level set-based framework for 3d kidney segmentation from diffusion mr images, in: IEEE International Conference on Image Processing (ICIP), 2015. IEEE, 2015, pp. 4441–4445.
- [49] M. Shehata, F. Khalifa, A. Soliman, M.A. El-Ghar, A.C. Dwyer, G. Gimel'farb, et al., A promising non- invasive cad system for kidney function assessment, in: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2016, pp. 613–621.
- [50] F. Khalifa, A. Soliman, A. Elmaghraby, G. Gimel'farb, A. El-Baz, 3d kidney segmentation from abdominal images using spatial-appearance models, *Comput. Math. Methods Med.* 2017 (2017) 1–10.
- [51] E. Hollis, M. Shehata, F. Khalifa, M.A. El-Ghar, T. El-Diasty, A. El-Baz, Towards non-invasive diagnostic techniques for early detection of acute renal transplant rejection: a review, *Egypt. J. Radiol. Nucl. Med.* 48 (1) (2016) 257–269.
- [52] M. Shehata, F. Khalifa, A. Soliman, M.A. El-Ghar, A.C. Dwyer, A. El-Baz, Assessment of renal transplant using image and clinical-based biomarkers, in: Proceedings of 13th Annual Scientific Meeting of American Society for Diagnostics and Interventional Nephrology (ASDIN'17), New Orleans, LA, USA, 10–12 February 2017.
- [53] M. Shehata, F. Khalifa, A. Soliman, M.A. El-Ghar, A.C. Dwyer, A. El-Baz, Early assessment of acute renal rejection, in: Proceedings of 12th Annual Scientific Meeting of American Society for Diagnostics and Interventional Nephrology (ASDIN'16), Pheonix, AZ, USA, 19–21 February 2016, 2017.
- [54] A. Eltanboly, M. Ghazal, H. Hajjdiab, A. Shalaby, A. Switala, A. Mahmoud, et al., Level sets-based image segmentation approach using statistical shape priors, *Appl. Math. Comput.* 340 (2019) 164–179.
- [55] M. Shehata, A. Mahmoud, A. Soliman, F. Khalifa, M. Ghazal, M.A. El-Ghar, et al., 3d kidney segmentation from abdominal diffusion mri using an appearance-guided deformable boundary, *PLoS One* 13 (7) (2018) e0200082.
- [56] H. Abdeltawab, M. Shehata, A. Shalaby, F. Khalifa, A. Mahmoud, M.A. El-Ghar, et al., A novel cnn-based cad system for early assessment of transplanted kidney dysfunction, *Sci. Rep.* 9 (1) (2019) 5948.
- [57] F. Khalifa, G. Beache, A. El-Baz, G. Gimel'farb, Deformable model guided by stochastic speed with application in cine images segmentation, in: Proceedings of IEEE International Conference on Image Processing, (ICIP'10), Hong Kong, 26–29 September 2010, pp. 1725–1728.
- [58] F. Khalifa, G.M. Beache, A. Elnakib, H. Sliman, G. Gimel'farb, K.C. Welch, et al., A new shape-based framework for the left ventricle wall segmentation from cardiac first-pass perfusion MRI, in: Proceedings of IEEE International Symposium on Biomedical Imaging: From Nano to Macro, (ISBI'13), San Francisco, CA, 7–11 April 2013, pp. 41–44.
- [59] F. Khalifa, G.M. Beache, A. Elnakib, H. Sliman, G. Gimel'farb, K.C. Welch, et al., A new nonrigid registration framework for improved visualization of transmural perfusion gradients on cardiac first-pass perfusion MRI, in: Proceedings of IEEE International Symposium on Biomedical Imaging: From Nano to Macro, (ISBI'12), Barcelona, Spain, 2–5 May 2012, pp. 828–831.

- [60] F. Khalifa, G.M. Beache, A. Firjani, K.C. Welch, G. Gimel'farb, A. El-Baz, A new nonrigid registration approach for motion correction of cardiac first-pass perfusion MRI, in: *Proceedings of IEEE International Conference on Image Processing, (ICIP'12)*, Lake Buena Vista, Florida, 30 September–3 October 2012, pp. 1665–1668.
- [61] F. Khalifa, G.M. Beache, G. Gimel'farb, A. El-Baz, A novel CAD system for analyzing cardiac first-pass MR images, in: *Proceedings of IAPR International Conference on Pattern Recognition (ICPR'12)*, Tsukuba Science City, Japan, 11–15 November 2012, pp. 77–80.
- [62] F. Khalifa, G.M. Beache, G. Gimel'farb, A. El-Baz, A novel approach for accurate estimation of left ventricle global indexes from short-axis cine MRI, in: *Proceedings of IEEE International Conference on Image Processing, (ICIP'11)*, Brussels, Belgium, 11–14 September 2011, pp. 2645–2649.
- [63] F. Khalifa, G.M. Beache, G. Gimel'farb, G.A. Giridharan, A. El-Baz, A new image-based framework for analyzing cine images, in: A. El-Baz, U.R. Acharya, M. Mirmedhdi, J.S. Suri (Eds.), *Handbook of Multi Modality State-of-the-Art Medical Image Segmentation and Registration Methodologies*, vol. 2, Springer, New York, 2011, pp. 69–98. ch. 3.
- [64] F. Khalifa, G.M. Beache, G. Gimel'farb, G.A. Giridharan, A. El-Baz, Accurate automatic analysis of cardiac cine images, *IEEE Trans. Biomed. Eng.* 59 (2) (2012) 445–455.
- [65] F. Khalifa, G.M. Beache, M. Nitzken, G. Gimel'farb, G.A. Giridharan, A. El-Baz, Automatic analysis of left ventricle wall thickness using short-axis cine CMR images, in: *Proceedings of IEEE International Symposium on Biomedical Imaging: From Nano to Macro, (ISBI'11)*, Chicago, Illinois, 30 March–2 April 2011, pp. 1306–1309.
- [66] M. Nitzken, G. Beache, A. Elnakib, F. Khalifa, G. Gimel'farb, A. El-Baz, Accurate modeling of tagged CMR 3D image appearance characteristics to improve cardiac cycle strain estimation, in: *Image Processing (ICIP), 2012 19th IEEE International Conference on*. Orlando, Florida, USA, IEEE, Sep. 2012, pp. 521–524.
- [67] M. Nitzken, G. Beache, A. Elnakib, F. Khalifa, G. Gimel'farb, A. El-Baz, Improving full-cardiac cycle strain estimation from tagged cmr by accurate modeling of 3D image appearance characteristics, in: *Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on*. Barcelona, Spain. IEEE, May 2012, pp. 462–465 (Selected for oral presentation).
- [68] M.J. Nitzken, A.S. El-Baz, G.M. Beache, Markov-gibbs random field model for improved full-cardiac cycle strain estimation from tagged cmr, *J. Cardiovasc. Magn. Reson.* 14 (1) (2012) 1–2.
- [69] H. Sliman, A. Elnakib, G. Beache, A. Elmaghraby, A. El-Baz, Assessment of myocardial function from cine cardiac MRI using a novel 4D tracking approach, *J. Comput. Sci. Syst. Biol.* 7 (2014) 169–173.
- [70] H. Sliman, A. Elnakib, G.M. Beache, A. Soliman, F. Khalifa, G. Gimel'farb, et al., A novel 4D PDE-based approach for accurate assessment of myocardium function using cine cardiac magnetic resonance images, in: *Proceedings of IEEE International Conference on Image Processing (ICIP'14)*, Paris, France, 27–30 October 2014, pp. 3537–3541.
- [71] H. Sliman, F. Khalifa, A. Elnakib, G.M. Beache, A. Elmaghraby, A. El-Baz, A new segmentation-based tracking framework for extracting the left ventricle cavity from

- cine cardiac MRI, in: Proceedings of IEEE International Conference on Image Processing, (ICIP'13), Melbourne, Australia, 15–18 September 2013, pp. 685–689.
- [72] H. Sliman, F. Khalifa, A. Elnakib, A. Soliman, G.M. Beache, A. Elmaghraby, et al., Myocardial borders segmentation from cine MR images using bi-directional coupled parametric deformable models, *Med. Phys.* 40 (9) (2013) 1–13.
- [73] H. Sliman, F. Khalifa, A. Elnakib, A. Soliman, G.M. Beache, G. Gimel'farb, et al., Accurate segmentation framework for the left ventricle wall from cardiac cine MRI, in: Proceedings of International Symposium on Computational Models for Life Science, (CMLS'13), vol. 1559, Sydney, Australia, 27–29 November 2013, pp. 287–296.
- [74] B. Abdollahi, A.C. Civelek, X.-F. Li, J. Suri, A. El-Baz, PET/CT nodule segmentation and diagnosis: a survey, in: L. Saba, J.S. Suri (Eds.), *Multi Detector CT Imaging*, Taylor, Francis, 2014, pp. 639–651. ch. 30.
- [75] B. Abdollahi, A. El-Baz, A.A. Amini, A multi-scale non-linear vessel enhancement technique, in: Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE. IEEE, 2011, pp. 3925–3929.
- [76] B. Abdollahi, A. Soliman, A. Civelek, X.-F. Li, G. Gimel'farb, A. El-Baz, A novel gaussian scale space-based joint MGRF framework for precise lung segmentation, in: Proceedings of IEEE International Conference on Image Processing, (ICIP'12). IEEE, 2012, pp. 2029–2032.
- [77] B. Abdollahi, A. Soliman, A. Civelek, X.-F. Li, G. Gimel'farb, A. El-Baz, A novel 3D joint MGRF framework for precise lung segmentation, in: *Machine Learning in Medical Imaging*. Springer, 2012, pp. 86–93.
- [78] A.M. Ali, A.S. El-Baz, A.A. Farag, A novel framework for accurate lung segmentation using graph cuts, in: Proceedings of IEEE International Symposium on Biomedical Imaging: From Nano to Macro, (ISBI'07). IEEE, 2007, pp. 908–911.
- [79] A. El-Baz, G.M. Beache, G. Gimel'farb, K. Suzuki, K. Okada, Lung imaging data analysis, *Int. J. Biomed. Imaging* 2013 (2013) 1–2.
- [80] A. El-Baz, G.M. Beache, G. Gimel'farb, K. Suzuki, K. Okada, A. Elnakib, Computer-aided diagnosis systems for lung cancer: challenges and methodologies, *Int. J. Biomed. Imaging* 2013 (2013) 1–46.
- [81] A. El-Baz, A. Elnakib, M. Abou El-Ghar, G. Gimel'farb, R. Falk, A. Farag, Automatic detection of 2D and 3D lung nodules in chest spiral CT scans, *Int. J. Biomed. Imaging* 2013 (2013) 1–11.
- [82] A. El-Baz, A.A. Farag, R. Falk, R. La Rocca, A unified approach for detection, visualization, and identification of lung abnormalities in chest spiral CT scans, *International Congress Series*, vol. 1256, Elsevier, 2003, pp. 998–1004.
- [83] A. El-Baz, A.A. Farag, R. Falk, R. La Rocca, Detection, visualization and identification of lung abnormalities in chest spiral CT scan: Phase-I, in: Proceedings of International conference on Biomedical Engineering, Cairo, Egypt, vol. 12, no. 1, 2002.
- [84] A. El-Baz, A. Farag, G. Gimel'farb, R. Falk, M.A. El-Ghar, T. Eldiasty, A framework for automatic segmentation of lung nodules from low dose chest CT scans, in: Proceedings of International Conference on Pattern Recognition, (ICPR'06), vol. 3. IEEE, 2006, pp. 611–614.
- [85] A. El-Baz, A. Farag, G. Gimel'farb, R. Falk, M.A. El-Ghar, A novel level set-based computer-aided detection system for automatic detection of lung nodules in low dose

- chest computed tomography scans, *Lung Imaging Comput. Aided Diagn.* 10 (2011) 221–238.
- [86] A. El-Baz, G. Gimel'farb, M. Abou El-Ghar, R. Falk, Appearance-based diagnostic system for early assessment of malignant lung nodules, in: *Proceedings of IEEE International Conference on Image Processing, (ICIP'12)*. IEEE, 2012, pp. 533–536.
- [87] A. El-Baz, G. Gimel'farb, R. Falk, A novel 3D framework for automatic lung segmentation from low dose CT images, in: A. El-Baz, J.S. Suri (Eds.), *Lung Imaging and Computer Aided Diagnosis*, Taylor Francis, 2011, pp. 1–16. ch. 1.
- [88] A. El-Baz, G. Gimel'farb, R. Falk, M. El-Ghar, Appearance analysis for diagnosing malignant lung nodules, in: *Proceedings of IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI'10)*. IEEE, 2010, pp. 193–196.
- [89] A. El-Baz, G. Gimel'farb, R. Falk, M.A. El-Ghar, A novel level set-based CAD system for automatic detection of lung nodules in low dose chest CT scans, in: A. El-Baz, J.S. Suri (Eds.), *Lung Imaging and Computer Aided Diagnosis*, vol. 1, Taylor, Francis, 2011, pp. 221–238. ch. 10.
- [90] A. El-Baz, G. Gimel'farb, R. Falk, M.A. El-Ghar, A new approach for automatic analysis of 3D low dose CT images for accurate monitoring the detected lung nodules, in: *Proceedings of International Conference on Pattern Recognition (ICPR'08)*. IEEE, 2008, pp. 1–4.
- [91] A. El-Baz, G. Gimel'farb, R. Falk, M.A. El-Ghar, A novel approach for automatic follow-up of detected lung nodules, in: *Proceedings of IEEE International Conference on Image Processing, (ICIP'07)*, vol. 5. IEEE, 2007, pp. V–501.
- [92] A. El-Baz, G. Gimel'farb, R. Falk, M.A. El-Ghar, A new CAD system for early diagnosis of detected lung nodules, in: *Image Processing, 2007. ICIP 2007. IEEE International Conference on*, vol. 2. IEEE, 2007, pp. II–461.
- [93] A. El-Baz, G. Gimel'farb, R. Falk, M.A. El-Ghar, H. Refaie, Promising results for early diagnosis of lung cancer, in: *Proceedings of IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI'08)*. IEEE, 2008, pp. 1151–1154.
- [94] A. El-Baz, G.L. Gimel'farb, R. Falk, M. Abou El-Ghar, T. Holland, T. Shaffer, A new stochastic framework for accurate lung segmentation, in: *Proceedings of Medical Image Computing and Computer-Assisted Intervention (MICCAI'08)*, 2008, pp. 322–330.
- [95] A. El-Baz, G.L. Gimel'farb, R. Falk, D. Heredis, M. Abou El-Ghar, A novel approach for accurate estimation of the growth rate of the detected lung nodules, in: *Proceedings of International Workshop on Pulmonary Image Analysis*, 2008, pp. 33–42.
- [96] A. El-Baz, G.L. Gimel'farb, R. Falk, T. Holland, T. Shaffer, A framework for unsupervised segmentation of lung tissues from low dose computed tomography images, in: *Proceedings of British Machine Vision, (BMVC'08)*, 2008, pp. 1–10.
- [97] A. El-Baz, G. Gimel'farb, R. Falk, M.A. El-Ghar, 3D MGRF-based appearance modeling for robust segmentation of pulmonary nodules in 3D LDCT chest images, in: *Lung Imaging and Computer Aided Diagnosis*. Chapter, 2011, ch. 3, pp. 51–63.
- [98] A. El-Baz, G. Gimel'farb, R. Falk, M.A. El-Ghar, Automatic analysis of 3D low dose CT images for early diagnosis of lung cancer, *Pattern Recognit.* 42 (6) (2009) 1041–1051.

- [99] A. El-Baz, G. Gimel'farb, R. Falk, M.A. El-Ghar, S. Rainey, D. Heredia, et al., Toward early diagnosis of lung cancer, in: *Proceedings of Medical Image Computing and Computer-Assisted Intervention, (MICCAI'09)*. Springer, 2009, pp. 682–689.
- [100] A. El-Baz, G. Gimel'farb, R. Falk, M.A. El-Ghar, J. Suri, Appearance analysis for the early assessment of detected lung nodules, in: *Lung Imaging and Computer Aided Diagnosis*. Chapter, 2011, ch. 17, pp. 395–404.
- [101] A. El-Baz, F. Khalifa, A. Elnakib, M. Nitzken, A. Soliman, P. McClure, et al., A novel approach for global lung registration using 3D Markov Gibbs appearance model, in: *Proceedings of International Conference Medical Image Computing and Computer-Assisted Intervention (MICCAI'12)*, Nice, France, 1–5 October 2012, pp. 114–121.
- [102] A. El-Baz, M. Nitzken, A. Elnakib, F. Khalifa, G. Gimel'farb, R. Falk, et al., 3D shape analysis for early diagnosis of malignant lung nodules, in: *Proceedings of International Conference Medical Image Computing and Computer-Assisted Intervention (MICCAI'11)*, Toronto, Canada, 18–22 September 2011, pp. 175–182.
- [103] A. El-Baz, M. Nitzken, G. Gimel'farb, E. Van Bogaert, R. Falk, M.A. El-Ghar, et al., Three-dimensional shape analysis using spherical harmonics for early assessment of detected lung nodules, *Lung Imaging Computer Aided Diagnosis*. (2011) 421–438. chapter.
- [104] A. El-Baz, M. Nitzken, F. Khalifa, A. Elnakib, G. Gimel'farb, R. Falk, et al., 3D shape analysis for early diagnosis of malignant lung nodules, in: *Proceedings of International Conference on Information Processing in Medical Imaging, (IPMI'11)*, Monastery Irsee, Germany (Bavaria), 3–8 July 2011, pp. 772–783.
- [105] A. El-Baz, M. Nitzken, E. Vanbogaert, G. Gimel'Farb, R. Falk, M. Abo El-Ghar, A novel shape-based diagnostic approach for early diagnosis of lung nodules, in: *Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on*. IEEE, 2011, pp. 137–140.
- [106] A. El-Baz, P. Sethu, G. Gimel'farb, F. Khalifa, A. Elnakib, R. Falk, et al., Elastic phantoms generated by microfluidics technology: validation of an imaged-based approach for accurate measurement of the growth rate of lung nodules, *Biotechnol. J.* 6 (2) (2011) 195–203.
- [107] A. El-Baz, P. Sethu, G. Gimel'farb, F. Khalifa, A. Elnakib, R. Falk, et al., A new validation approach for the growth rate measurement using elastic phantoms generated by state-of-the-art microfluidics technology, in: *Proceedings of IEEE International Conference on Image Processing, (ICIP'10)*, Hong Kong, 6–29 September 2010, pp. 4381–4383.
- [108] A. El-Baz, P. Sethu, G. Gimel'farb, F. Khalifa, A. Elnakib, R. Falk, et al., Validation of a new imaged-based approach for the accurate estimating of the growth rate of detected lung nodules using real CT images and elastic phantoms generated by state-of-the-art microfluidics technology, in: A. El-Baz, J.S. Suri (Eds.), *Handbook of Lung Imaging and Computer Aided Diagnosis*, vol. 1, Taylor & Francis, New York, 2011, pp. 405–420.
- [109] A. El-Baz, A. Soliman, P. McClure, G. Gimel'farb, M.A. El-Ghar, R. Falk, Early assessment of malignant lung nodules based on the spatial analysis of detected lung nodules, in: *Proceedings of IEEE International Symposium on Biomedical Imaging: From Nano to Macro, (ISBI'12)*. IEEE, 2012, pp. 1463–1466.

- [110] A. El-Baz, S.E. Yuksel, S. Elshazly, A.A. Farag, Non-rigid registration techniques for automatic follow-up of lung nodules, in: *Proceedings of Computer Assisted Radiology and Surgery, (CARS'05)*, vol. 1281. Elsevier, 2005, pp. 1115–1120.
- [111] [A.S. El-Baz, J.S. Suri, Lung Imaging and Computer Aided Diagnosis, CRC Press, 2011.](#)
- [112] A. Soliman, F. Khalifa, N. Dunlap, B. Wang, M. El-Ghar, A. El-Baz, An iso-surfaces based local deformation handling framework of lung tissues, in: *Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium on. IEEE, 2016*, pp. 1253–1259.
- [113] A. Soliman, F. Khalifa, A. Shaffie, N. Dunlap, B. Wang, A. Elmaghraby, et al., Detection of lung injury using 4D-CT chest images, in: *Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium on. IEEE, 2016*, pp. 1274–1277.
- [114] A. Soliman, F. Khalifa, A. Shaffie, N. Dunlap, B. Wang, A. Elmaghraby, et al., A comprehensive framework for early assessment of lung injury, in: *Image Processing (ICIP), 2017 IEEE International Conference on. IEEE, 2017*, pp. 3275–3279.
- [115] A. Shaffie, A. Soliman, M. Ghazal, F. Taher, N. Dunlap, B. Wang, et al., A new framework for incorporating appearance and shape features of lung nodules for precise diagnosis of lung cancer, in: *Image Processing (ICIP), 2017 IEEE International Conference on. IEEE, 2017*, pp. 1372–1376.
- [116] A. Soliman, F. Khalifa, A. Shaffie, N. Liu, N. Dunlap, B. Wang, et al., Image-based cad system for accurate identification of lung injury, in: *Image Processing (ICIP), 2016 IEEE International Conference on. IEEE, 2016*, pp. 121–125.
- [117] A. Soliman, A. Shaffie, M. Ghazal, G. Gimel'farb, R. Keynton, A. El-Baz, A novel cnn segmentation framework based on using new shape and appearance features, in: *2018 25th IEEE International Conference on Image Processing (ICIP). IEEE, 2018*, pp. 3488–3492.
- [118] A. Shaffie, A. Soliman, H.A. Khalifeh, M. Ghazal, F. Taher, R. Keynton, et al., On the integration of ct- derived features for accurate detection of lung cancer, in: *2018 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT). IEEE, 2018*, pp. 435–440.
- [119] A. Shaffie, A. Soliman, H.A. Khalifeh, M. Ghazal, F. Taher, A. Elmaghraby, et al., Radiomic-based framework for early diagnosis of lung cancer, in: *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019). IEEE, 2019*, pp. 1293–1297.
- [120] A. Shaffie, A. Soliman, M. Ghazal, F. Taher, N. Dunlap, B. Wang, et al., A novel autoencoder-based diagnostic system for early assessment of lung cancer, in: *2018 25th IEEE International Conference on Image Processing (ICIP). IEEE, 2018*, pp. 1393–1397.
- [121] [A. Shaffie, A. Soliman, L. Fraiwan, M. Ghazal, F. Taher, N. Dunlap, et al., A generalized deep learning-based diagnostic system for early diagnosis of various types of pulmonary nodules, Technol. Cancer Res. Treat. 17 \(2018\) 1533033818798800.](#)
- [122] Y. ElNakieb, A. Soliman, A. Mahmoud, O. Dekhil, A. Shalaby, M. Ghazal, et al., Autism spectrum disorder diagnosis framework using diffusion tensor imaging, in: *2019 IEEE International Conference on Imaging Systems and Techniques (IST). IEEE, 2019*, pp. 1–5.
- [123] R. Haweel, O. Dekhil, A. Shalaby, A. Mahmoud, M. Ghazal, R. Keynton, et al., A machine learning approach for grading autism severity levels using task-based

- functional mri, in: 2019 IEEE International Conference on Imaging Systems and Techniques (IST). IEEE, 2019, pp. 1–5.
- [124] O. Dekhil, M. Ali, R. Haweel, Y. Elnakib, M. Ghazal, H. Hajjdiab, et al., A comprehensive framework for differentiating autism spectrum disorder from neurotypicals by fusing structural mri and resting state functional mri, *Seminars in Pediatric Neurology*, Elsevier, 2020, p. 100805.
- [125] R. Haweel, O. Dekhil, A. Shalaby, A. Mahmoud, M. Ghazal, A. Khalil, et al., A novel framework for grading autism severity using task-based fmri, in: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI). IEEE, 2020, pp. 1404–1407.
- [126] B. Dombroski, M. Nitzken, A. Elnakib, F. Khalifa, A. El-Baz, M.F. Casanova, Cortical surface complexity in a population-based normative sample, *Transl. Neurosci.* 5 (1) (2014) 17–24.
- [127] A. El-Baz, M. Casanova, G. Gimel'farb, M. Mott, A. Switala, An MRI-based diagnostic framework for early diagnosis of dyslexia, *Int. J. Computer Assist. Radiol. Surg.* 3 (3–4) (2008) 181–189.
- [128] A. El-Baz, M. Casanova, G. Gimel'farb, M. Mott, A. Switala, E. Vanbogaert, et al., A new CAD system for early diagnosis of dyslexic brains, in: Proc. International Conference on Image Processing (ICIP'2008). IEEE, 2008, pp. 1820–1823.
- [129] A. El-Baz, M.F. Casanova, G. Gimel'farb, M. Mott, A.E. Switala, A new image analysis approach for automatic classification of autistic brains, in: Proc. IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI'2007). IEEE, 2007, pp. 352–355.
- [130] A. El-Baz, A. Elnakib, F. Khalifa, M.A. El-Ghar, P. McClure, A. Soliman, et al., Precise segmentation of 3-D magnetic resonance angiography, *IEEE Trans. Biomed. Eng.* 59 (7) (2012) 2019–2029.
- [131] A. El-Baz, A.A. Farag, G. Gimel'farb, S.G. Hushek, Automatic cerebrovascular segmentation by accurate probabilistic modeling of tof-mra images, *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2005*, Springer, 2005, pp. 34–42.
- [132] A. El-Baz, A. Farag, A. Elnakib, M.F. Casanova, G. Gimel'farb, A.E. Switala, et al., Accurate automated detection of autism related corpus callosum abnormalities, *J. Med. Syst.* 35 (5) (2011) 929–939.
- [133] A. El-Baz, A. Farag, G. Gimelfarb, Cerebrovascular segmentation by accurate probabilistic modeling of tof-mra images, *Image Analysis*, vol. 3540, Springer, 2005, pp. 1128–1137.
- [134] A. El-Baz, G. Gimel'farb, R. Falk, M.A. El-Ghar, V. Kumar, D. Heredia, A novel 3D joint Markov-gibbs model for extracting blood vessels from PC-mra images, *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2009*, vol. 5762, Springer, 2009, pp. 943–950.
- [135] A. Elnakib, A. El-Baz, M.F. Casanova, G. Gimel'farb, A.E. Switala, Image-based detection of corpus callosum variability for more accurate discrimination between dyslexic and normal brains, in: Proc. IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI'2010). IEEE, 2010, pp. 109–112.
- [136] A. Elnakib, M.F. Casanova, G. Gimel'farb, A.E. Switala, A. El-Baz, Autism diagnostics by centerline-based shape analysis of the corpus callosum, in: Proc. IEEE

- International Symposium on Biomedical Imaging: From Nano to Macro (ISBI'2011). IEEE, 2011, pp. 1843–1846.
- [137] A. Elnakib, M. Nitzken, M. Casanova, H. Park, G. Gimel'farb, A. El-Baz, Quantification of age-related brain cortex change using 3D shape analysis, in: Pattern Recognition (ICPR), 2012 21st International Conference on. IEEE, 2012, pp. 41–44.
- [138] M. Nitzken, M. Casanova, G. Gimel'farb, A. Elnakib, F. Khalifa, A. Switala, et al., 3D shape analysis of the brain cortex with application to dyslexia, in: Image Processing (ICIP), 2011 18th IEEE International Conference on. Brussels, Belgium: IEEE, Sep. 2011, pp. 2657–2660, (Selected for oral presentation. Oral acceptance rate is 10 percent and the overall acceptance rate is 35 percent).
- [139] F.E.-Z.A. El-Gamal, M.M. Elmogy, M. Ghazal, A. Atwan, G.N. Barnes, M.F. Casanova, et al., A novel cad system for local and global early diagnosis of alzheimer's disease based on pib-pet scans, in: 2017 IEEE International Conference on Image Processing (ICIP). IEEE, 2017, pp. 3270–3274.
- [140] M.M. Ismail, R.S. Keynton, M.M. Mostapha, A.H. ElTanboly, M.F. Casanova, G.L. Gimel'farb, et al., Studying autism spectrum disorder with structural and diffusion magnetic resonance imaging: a survey, *Front. Hum. Neurosci.* 10 (2016) 211.
- [141] A. Alansary, M. Ismail, A. Soliman, F. Khalifa, M. Nitzken, A. Elnakib, et al., Infant brain extraction in t1-weighted mr images using bet and refinement using lcdg and mgrf models, *IEEE J. Biomed. Health Inform.* 20 (3) (2016) 925–935.
- [142] E.H. Asl, M. Ghazal, A. Mahmoud, A. Aslantas, A. Shalaby, M. Casanova, et al., Alzheimer's disease diagnostics by a 3d deeply supervised adaptable convolutional network, *Front. Biosci. (Landmark Ed.)* 23 (2018) 584–596.
- [143] O. Dekhil, M. Ali, Y. El-Nakieb, A. Shalaby, A. Soliman, A. Switala, et al., A personalized autism diagnosis cad system using a fusion of structural mri and resting-state functional mri data, *Front. Psychiatry* 10 (2019) 392 [Online]. Available: <https://www.frontiersin.org/article/10.3389/fpsy.2019.00392>.
- [144] A. Mahmoud, A. El-Barkouky, H. Farag, J. Graham, A. Farag, A non-invasive method for measuring blood flow rate in superficial veins from a single thermal image, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2013, pp. 354–359.
- [145] A. El-Baz, A. Farag, G. Gimel'farb, M.A. El-Ghar, T. Eldiasty, Probabilistic modeling of blood vessels for segmenting mra images, in: 18th International Conference on Pattern Recognition (ICPR'06), vol. 3. IEEE, 2006, pp. 917–920.
- [146] A. El-Baz, A.A. Farag, G. Gimel'farb, M.A. El-Ghar, T. Eldiasty, A new adaptive probabilistic model of blood vessels for segmenting mra images, *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2006*, vol. 4191, Springer, 2006, pp. 799–806.
- [147] A. El-baz, A. Shalaby, F. Taher, M. El-Baz, M. Ghazal, M.A. El-Ghar, et al., Probabilistic modeling of blood vessels for segmenting magnetic resonance angiography images, *Med. Res. Arch.* 5 (3) (2017).
- [148] A.S. Chowdhury, A.K. Rudra, M. Sen, A. Elnakib, A. El-Baz, Cerebral white matter segmentation from MRI using probabilistic graph cuts and geometric shape priors, *ICIP (2010)* 3649–3652.

- [149] Y. Gebru, G. Giridharan, M. Ghazal, A. Mahmoud, A. Shalaby, A. El-Baz, Detection of cerebrovascular changes using magnetic resonance angiography, *Cardiovascular Imaging and Image Analysis*, CRC Press, 2018, pp. 1–22.
- [150] A. Mahmoud, A. Shalaby, F. Taher, M. El-Baz, J.S. Suri, A. El-Baz, Vascular tree segmentation from different image modalities, *Cardiovascular Imaging and Image Analysis*, CRC Press, 2018, pp. 43–70.
- [151] F. Taher, A. Mahmoud, A. Shalaby, A. El-Baz, A review on the cerebrovascular segmentation methods, in: 2018 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT). IEEE, 2018, pp. 359–364.
- [152] H. Kandil, A. Soliman, L. Fraiwan, A. Shalaby, A. Mahmoud, A. ElTanboly, et al., A novel mra framework based on integrated global and local analysis for accurate segmentation of the cerebral vascular system, in: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). IEEE, 2018, pp. 1365–1368.
- [153] F. Taher, A. Soliman, H. Kandil, A. Mahmoud, A. Shalaby, G. Gimel'farb, et al., Accurate segmentation of cerebrovasculature from tof-mra images using appearance descriptors, *IEEE Access*. (2020).
- [154] A.A. Sleman, A. Soliman, M. Ghazal, H. Sandhu, S. Schaal, A. Elmaghraby, et al., Retinal layers oct scans 3-d segmentation, in: 2019 IEEE International Conference on Imaging Systems and Techniques (IST). IEEE, 2019, pp. 1–6.
- [155] N. Eladawi, M. Elmogy, M. Ghazal, O. Helmy, A. Aboelfetouh, A. Riad, et al., Classification of retinal diseases based on oct images, *Front. Biosci. (Landmark Ed.)* 23 (2018) 247–264.
- [156] A. ElTanboly, M. Ismail, A. Shalaby, A. Switala, A. El-Baz, S. Schaal, et al., A computer-aided diagnostic system for detecting diabetic retinopathy in optical coherence tomography images, *Med. Phys.* 44 (3) (2017) 914–923.
- [157] H.S. Sandhu, A. El-Baz, J.M. Seddon, “Progress in automated deep learning for macular degeneration, *JAMA Ophthalmol.* (2018).
- [158] M. Ghazal, S.S. Ali, A.H. Mahmoud, A.M. Shalaby, A. El-Baz, Accurate detection of non-proliferative diabetic retinopathy in optical coherence tomography images using convolutional neural networks, *IEEE Access*. 8 (2020) 34. 387–34 397.
- [159] K. Hammouda, F. Khalifa, A. Soliman, H. Abdeltawab, M. Ghazal, M. Abou El-Ghar, et al., A 3D CNN with a learnable adaptive shape prior for accurate segmentation of bladder wall using mr images, in: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI). IEEE, 2020, pp. 935–938.
- [160] A. Naglah, F. Khalifa, A. Mahmoud, M. Ghazal, P. Jones, T. Murray, et al., Athlete-customized injury prediction using training load statistical records and machine learning, in: 2018 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT). IEEE, 2018, pp. 459–464.
- [161] A.H. Mahmoud, Utilizing radiation for smart robotic applications using visible, thermal, polarization images. Ph.D. dissertation, University of Louisville, 2014.
- [162] A. Mahmoud, A. El-Barkouky, J. Graham, A. Farag, Pedestrian detection using mixed partial derivative based his togram of oriented gradients, in: 2014 IEEE International Conference on Image Processing (ICIP). IEEE, 2014, pp. 2334–2337.
- [163] A. El-Barkouky, A. Mahmoud, J. Graham, A. Farag, An interactive educational drawing system using a humanoid robot and light polarization, in: 2013 IEEE International Conference on Image Processing. IEEE, 2013, pp. 3407–3411.

- [164] A.H. Mahmoud, M.T. El-Melegy, A.A. Farag, Direct method for shape recovery from polarization and shading, in: 2012 19th IEEE International Conference on Image Processing. IEEE, 2012, pp. 1769–1772.
- [165] M.A. Ghazal, A. Mahmoud, A. Aslantas, A. Soliman, A. Shalaby, J.A. Benediktsson, et al., Vegetation cover estimation using convolutional neural networks, *IEEE Access*. 7 (2019) 132. 563–132 576.
- [166] M. Ghazal, A. Mahmoud, A. Shalaby, A. El-Baz, Automated framework for accurate segmentation of leaf images for plant health assessment, *Environ. Monit. Assess.* 191 (8) (2019) 491.

Adaptive graph convolutional neural network and its biomedical applications

Junzhou Huang and Ruoyu Li

The University of Texas at Arlington, Arlington, TX, United States

5.1 Introduction

Although the Convolutional neural networks (CNNs) have been proven to be supremely successful on a wide range of machine learning problems [1], the classical CNNs come with a requirement of regularly shaped tensor as input [2]. For instance, images [1] and videos [3] are respectively modeled as 2-D and 3-D tensors when being fed to a traditional convolutional neural network. As opposite to this stereotype, in many real-world applications, it is more likely to encounter the data deployed on an irregular grid for instead and more generally in a non-Euclidean domain. Despite being treated as regularly shaped tensors, those data are intuitively more appropriate to be seen as graph-structured, which make it convenient to handle the varying neighborhood vertex connectivity as well as the non-Euclidean metrics. Under these circumstances, the stationarity and the compositionality, which empower convolutional operator to work on multi-dimensional grid, does no longer exist on graph-structured data. Consequently, it is necessary to reformulate convolutional operator to make it compatible with graphs.

However, extending convolutional operator from regularly shaped grids to irregular graphs is not trivial. For the simplicity of constructing convolutional kernel, most early-stage neural networks on graph assumed that input data are still low-dimensional. Because, in their setup, the convolver handled a subset of nodes grouped by node degree respectively [4,5]. And their convolutional kernel is over-localized and infeasible to learn hierarchical representations from complex graphs with unpredictable and flexible node connectivity, for example chemical molecules and social networks [6]. In some applications, for example classification of point cloud [7], the topological structure of graph is more informative than the node features or the edge features alone. Unfortunately, the existing graph convolution networks [8] cannot thoroughly exploit the geometric property of graph due to the difficulties of implementing a parameterized spatial kernel compatible with different scenario of node neighborhood. Similar difficulty is also

interfering the extension of graph convolutional networks to new applications, for example human activity recognition [9] and co-citation networks [10]. Besides, given the flexibility of topology structure of graph and the $O(N^2)$ scale of parameters to define node-pair connectivity, learning a topology-preserving spatial convolutional kernel for every unique graph data sample is impractical in real-world scenarios.

Inherited from classical CNNs, a shared convolutional kernel among training samples is one of the common assumptions. Consequently, to guarantee a unified dimensionality of input/output for all samples at consecutive layers, the input graph data have to be pruned, which is also a constraint of utilizing traditional networks on graph directly. However, this kind of preprocessing on graph-structured data could destroy the completeness of graph-oriented information. For instance, the coarsening of molecule is hard to be justified chemically, and it is likely that the coarsened graph has lost the key sub-structures that differentiate the molecule from others. In Fig. 5.1, removal of any carbon atom from the graph breaks the Benzene ring. It would be much better if the graph CNNs could accept and preserve the original graph structures of data. Beyond the spatial graph convolution applied to quasi-grid graphs, another kind of graph convolutional neural networks, who were derived from the graph Fourier transform [11], are promising to offer an elastic kernel for sophisticated graph-structured data. The article introduces spectral network in session 2 and 3.

Finally, in this article, the graph-structured data used in experiments either have an intrinsic graph structure (e.g. chemical molecules) or have one reconstructed through clustering of vertices (e.g. point-cloud). In existing graph networks [4,5,9], the initial graph structures are enforced to be stable during the training process. While, on the other hand, it is doubtful that the graph topology derived in unsupervised fashion happens to be optimal for each particular learning task. Although there were pioneering works who included supervised graph reconstructions with fully connected networks [5], the computational complexity of net weights constrains the initialization of graph only feasible to small graph. Furthermore, the graph topology that fits in one pre-trained network may not work well with another kind of graph neural network [12].

In this article, we introduce a novel spectral graph convolution neural network, compatible with graph data of diverse topological structures, for example the

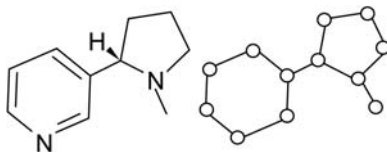


FIGURE 5.1

Example of molecular graph: nicotine (C₁₀H₁₄N₂, SMILES: CN1CCC[C@H]1c2cccnc2) and its intrinsic graph (omit hydrogen atoms).

organic molecules that consist of a different number of atoms. To deal with the fixed graph that may be stale, we choose to grant the network the capability of learning the supplement to graph topological structure. Therefore, different from a parametric kernel formed by a fixed Laplacian L [4–6], we parametrize the graph Laplacian itself. While, given the goal of preserving topology on each individual graph, we cannot learn the L as trainable parameters directly, let alone the unacceptable computational cost. For instead, we parameterize the distance metric between pairs of node feature vectors as an indirect learning of the self-organized structure of each graph sample. A reasonable assumption is that the distance metric parameters are shared by all samples that belong to the same type, for example molecular graphs. Consequently, each individual sample is able to train the network on a unique and adaptive graph Laplacian that preserves its uniqueness and infers any undiscovered task-related substructures. A customized L will lead to a customized kernel that combines neighbors’ features. It is interesting to question what exact graph that optimally empower a particular task. For instance, the chemical bonds, found via chemical experiments, naturally build a graph for any compound. However, it is never guaranteed that the convolver that works on intrinsic graph has extracted every meaningful feature. We introduced a so-called *residual graph* to unveil the substructures that the intrinsic graph does not present. Furthermore, to guarantee residual graphs to be the optimal supplement to the intrinsic graphs for the task, we learn the residual graph along with the rest of network.

To implement the learning of adaptive residual graphs, we are faced with two major problems: (1) how to efficiently construct residual graph during training; (2) how to preserve unique graph topology in the batch-wise training. A direct learning of L is of exponential complexity and with $O(N^2)$ parameters for a $R^{N \times d}$ graph sample. Allowing the topological structures preserved in M training samples means complexity of $O(MN^2)$, which is unscalable for large graphs. While, an indirect learning of graph structure based on the *Mahalanobis* distance metric [13] and transformation in feature manifold is able to reduce the complexity scale to $O(d^2)$ or even $O(d)$, assuming the metric parameters are shared by all graphs. More importantly, after this, the learning complexity becomes *independent* of graph size N . When execute the convolutional operator, both intrinsic graph and learned residual graph will be used in the kernel: $\hat{L} = L + L'$. Owing to that, the proposed layer is a spectral graph convolution (SGC) layer [5] but on the learned Laplacian, we name the new graph convolutional layer as SGC-LL. In terms of forward-pass of topological knowledge, the self-constructed residual graph Laplacian L' will be pass to next graph convolution layer as one of training inputs.

In classical CNNs, back-propagation generally updates kernel weights to adjust the relationship between neighboring nodes at each feature dimension individually. Then it sums up signals from all filters to construct hidden-layer activations [1]. To grant the graph convolutional network a similar capability, we applied a re-parameterization on the output feature using a linear transform

weight and bias. Finally, the $O(d^2)$ training parameters in the proposed graph convolution layer consist of two segments: the *Mahalanobis* distance metric parameters and the feature transform weight and bias. To facilitate the diverse input graph of varying number of nodes, we need to pad zeros to both feature and adjacency tensor. Therefore, we also modify existing graph pooling and gather layer to recover the original data from padded tensors before layer execution. Because the introduced graph network is capable of being adapted with respect to graph topology, and more importantly the graph being used is also adaptive toward learning task, we name the network as Adaptive Graph Convolutional Network (AGCN) [14] to highlight these valuable features.

AGCN, first introduced by us in [14], in which we verified the effectiveness of the new spectral convolution layer on multiple graph-structured datasets ranging from chemical molecules and 3D point cloud generated by LIDAR [15]. AGCN had demonstrated overwhelmingly better accuracy on both graph classification and graph attributes regression. Furthermore, we also explored the application of AGCN to more sophisticated applications, such as survival analysis [16]. During this exploration, we extended the theoretical work of AGCN via combining with graph attention network (GAT) [17]. The introduced DeepGraphSurv [18] delivered state-of-the-art accuracy on survival prediction on three benchmark tumor whole-slide image (WSI) datasets. Moreover, the end-to-end graph representation learning network is also promising for other large-scale image analysis tasks. The works were initially introduced in conference papers [14,18]. Some technical advantages of the AGCN architecture are summarized in the following:

1. Construct and learn unique graph Laplacian for each individual training graph sample. Preserve the completeness of original information, especially in terms of spatial topology.
2. Low computational expense in scale of $O(d^2)$, independent of vertex number N , making the network more attractive in training tasks on large-scale graphs.
3. Explainable model. Graph neural networks have an intrinsic advantage on interpretability. Discovering the hidden substructures on graph and the representations learned upon those substructures cross-validate the effectiveness of learned *residual graphs* from AGCN.
4. Residual graph is compatible with both spatial and spectral graph convolutional network. We chose SGC [5] as baseline and build SGC-LL on top, while the idea of residual graph and the learning of graph Laplacian via distance learning is trivial to be extended to other graph neural networks, for example GAT [17] and Differentiable Graph Module [19].

The rest of the article is organized as follows: (1) in Session 2, we introduce the origin and the evolutionary roadmap of graph convolutional network (GCN) and its success on molecular graph data and the recent massive applications to computer vision, drug discovery and natural language processing; (2) Session 3 provides audience an overview of mathematical formulations and analysis of introduced SGC-LL layer and what components AGCN comprises besides SGC-LL layers; (3) Experimental setup, results,

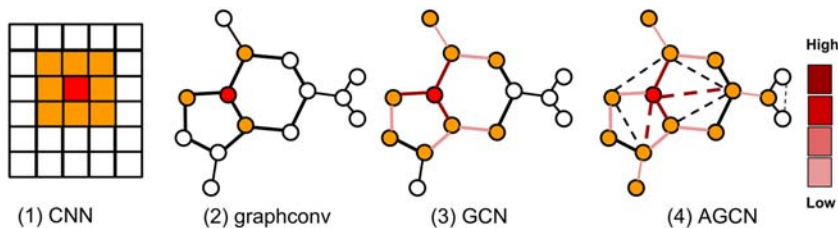


FIGURE 5.2

Red spot is the central node of convolution, the orange nodes represent receptive field. (1) classical CNN; (2) NeuralFP (spatial convolution) [20]; (3) ChevNet (spectral convolution) [21]; (4) AGCN based on learned residual graph (*dash lines*).

and discussions are presented in Session 4 with comparison with state-of-the-arts in GCN family; (4) Conclusion, future works, and acknowledgment are provided in Session 5 (Fig. 5.2).

5.2 Related work

5.2.1 Evolution of graph convolutional neural networks

As graph is a more common and generic topology of data in real-world scenarios, there is an increasing interest of generalizing CNNs to graph data since the significant success on computer vision starting in 2011. The majority of advances in this direction could be categorized as *spatial GCN* and *spectral GCN* according to the domain in which the convolution executed.

5.2.1.1 Spatial graph convolutional neural networks

The first trial of formulating an analogy of CNN on graph was accomplished by [4]. Particularly, the authors proposed a sparse kernel that aggregated the vertex features element-wisely from its neighbors. The finite-size kernel is nonparametric though, and the graph was self-constructed and constricted by data. [20] extended spatial kernel [4] to molecular graph and used dedicated weight matrices for the cluster of nodes of same degree. The drawback of [20] is that the node degree scenario should predefined and fixed, otherwise the network cannot be designed and initialized. And the spatial kernel [4,20] is also over-localized since the adjacency matrix form an undirected graph is merely able to represent the 1-hop connections, therefore, the kernel cannot assign weights to those peripheral nodes not directly linked to the central node. To relieve the constrain, the diffusion-convolutional network (DCNN) [22] builds K transform matrices to handle at most K -hops diffusion of node features on graph, allowing output $x' \in \mathcal{R}^{[N \times K \times d]}$. To tackle the over-localized kernel, DGCN [23] executed two parallel convolution networks on

two views of graph data balancing local and global consistency toward a semi-supervised problem.

Previous networks [20,22,23] more or less handled the challenge of diverse node degree, while none of them formulated a CNN-alike convolution on graph, and their convolutional layers were without loss of generality able to be approximated as the assembly of a series of fully connected layers. PTCHY-SAN [23] ignored the graph scale and only selected a fixed number of nodes with a fixed receptive field, following one of graph labeling procedures. The receptive field were picked from its direct neighborhood. And lastly the normalized neighborhood from the receptive field serve the final aggregation operation. [24] proposed LGCL, transforming graph data back to grid-like structure, selects a fixed number of neighbor nodes for each feature dimension, and then then applies a 1D-CNN on top. [25] further discussed the theoretical generalization of CNN from grid to manifold and finally graphs. GraphSAGE [26] argued that above transductive algorithms required the presence of nodes in training and cannot deal with graphs consisting of unseen nodes. However, [23,24] did not utilize the entire set of nodes, while, for some scenarios such as drug attribute prediction [27], a selective aggregation of nodes means a damage to local substructures, which have to stay intact to learn meaningful hidden representations.

5.2.1.2 Spectral graph convolutional neural networks

Another category of graph convolutional operator is defined and executed in Fourier domain. [4] first proposed to compute the graph convolution based on the convolution theorem and the eigen-decomposition of graph Laplacian matrix $L = U\Lambda U^T$. Then, the convolution is written as:

$$g_{\{\theta\}} \times x = Ug_{\{\theta\}}(\Lambda)U^T x \quad (5.1)$$

where U is the eigenvectors of the normalized graph Laplacian: $L = I - D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$. D is the degree matrix and A is the adjacency matrix. While, this trivial solution comes with massive computational cost from the eigen-decomposition. More importantly, the spectrum filtering may result in non-locality on spatial domain after applying inverse graph Fourier transform U . [5] attempted to tackle the spatial non-locality after filtering by the nonparametric spectral kernel [4] by parametrizing $g_{\{\theta\}}(\Lambda)$ in terms of $diag(W)$, whose parameters are $W \in \mathbb{R}^N$. Furthermore, another non-linear approximation of kernel $g_{\{\theta\}}(\Lambda)$ was proposed as:

$$g_{\{\theta\}}(\Lambda) = \sum_{k=0}^{K-1} \theta_k \Lambda^k \quad (5.2)$$

The kernel Eq. (5.2) mitigated the spatial non-locality by smoothing the spectrum filtering, while the computational cost is still high. [28] introduced a truncated Chebyshev expansion of kernel that comprises K items as $T_k(\Lambda)$. A recurring evaluation of $T_k(\Lambda)$ with $T_0(x) = 1, T_1(x) = x$ is formulated as: $2xT_{k-1}(x) - T_{k-2}(x)$. When convolving with feature x , the $O(N^2)$ complex multiplication with dense Fourier basis U is replaced with the multiplication with the

sparse \hat{L} , and it also saves the eigen-decomposition of \hat{L} . Consequently, the overall complexity is reduced to $O(K|\xi|) \ll O(N^2)$. $|\xi|$ is the count of non-zeros in \hat{L} and the number of graph edge. Both [28] and [29] relied on the approximation of spectral convolutional operator finalized as:

$$g_{\{\theta\}}(\Lambda) \times x \simeq \sum_{k=0}^{K-1} \theta_k T_k(\hat{L})x \quad (5.3)$$

\hat{L} is the re-normalized graph Laplacian defined as: $\hat{L} = \frac{2}{\lambda_{\{max\}}}L - I_N$, where

$\lambda_{\{max\}}$ is the maximum eigenvalue of L . While [29] further simplified the evaluation of Eq. (5.3) by setting $K = 1$ and $\lambda_{\{max\}} \simeq 2$ to alleviate the overfitting to local structures. Because local neighborhood might deliver a biased representation of graphs especially when the node degree distribution is skewed. Authors argued that a stack of multiple linearly approximated convolutions, that is $K = 1$, is also able to recover a similar multi-hop knowledge aggregation as $K > 1$. Given $K = 1$ and $\lambda_{\{max\}} = 2$, and the recurring evaluation equations, the Eq. (5.3) was further simplified in [29] as:

$$g_{\{\theta\}} \times x \simeq \theta_0 - \theta_1 D^{-\frac{1}{2}} A D^{-\frac{1}{2}} x \quad (5.4)$$

with layer-wide shared parameters $\{\theta_0, \theta_1\}$. An additional assumption $\theta = \theta_0 = -\theta_1$ brought us the linear approximation of single-layer spectral convolution as:

$$g_{\{\theta\}} \times x \simeq \theta \left(I_N + D^{-\frac{1}{2}} A D^{-\frac{1}{2}} \right) x \quad (5.5)$$

However, as expected, a repeated application of Eq. (5.5) leads to gradient exploding/vanishing. So, within such layer, a re-normalization of adjacency A is executed by $\hat{A} = A + I_N$ to control the eigenvalues of \hat{L} fall into range $[0, 2]$. The baseline of spectral GCN in the article is [28] with the K -rank Chebyshev polynomials as an approximation of kernel. The introduced AGCN is also founded on the formulations given in [28] and inspired by [29].

5.2.2 Neural network on molecular graph

Biological and chemical research are two fields in which GCN made significant progress in recent years. Given the nature of organic compounds as molecular graph, it is straightforward to formulate chemical compounds as graphs and to perform graph neural network (GNN) on top for representation learning and prediction tasks. Within those, the GCNs, with localized kernels and deep network architecture, have successfully driven big progress on problems, for example drug property prediction [30], drug discovery [31,32], reaction prediction [33,34]. The early success of GNN on molecular graph was the *NeuralFP* proposed in [20]. While, [20] had a strong assumption on node degree distribution, making the spatial neighborhood aggregation difficult to be generalized to graph of skewed degree distribution. And if designing a different kernel for different feature channel, the massive matrices multiplications are unscalable for

large graphs. Consider the real-world scenarios, we have to balance the representational capability and the computational cost. While, [20] groups nodes across graphs in batch according to node degree and lets nodes of same degree share transform parameters, on the direction against over-locality overfitting. But, it failed to perform a parametric elastic kernel to aggregate K -hop receptive field as what a CNN does.

Spectral GCNs balance the trade-off between the representational power and the computational cost by reducing the number of parameters and making the kernels shared by entire nodes. In [14], we introduced GCNs to molecular graph classification and drug property prediction tasks. Not doing any node truncation or edge pruning, we devised a novel spectral convolutional layer deployed on full graphs preserving original topology. Besides, we argued that the original graphs are not optimal for particular tasks. Beyond a kernel designed with fixed structure, the SGC-LL layer [14] makes graph structure trainable along with the rest of network. As opposite to the node ‘Selection, Assembly, Normalization’ procedure by [24] that prunes the graph to fit a pre-defined kernel, we let the kernel be adaptive to different graphs. More discussions on it in Session 3.

5.2.3 Attention on graph

Following the success of attention-only sequence-mapping networks, for example *Transformer* [35], and a variety of attention mechanism on natural language understanding tasks [36,37], the attention scheme was introduced to GNNs as Graph Attention Network (GAT) [17], leveraging a masked self-attentional layer that allows nodes specifying weights to different neighbors on graph. The mask applied to node selection is where the graph structure introduced, GAT computes attention coefficients $e_{\{i,j\}}$ where $A_{\{i,j\}} > 0$. And the neighborhood N_i are node i 's first-order neighbors. Similar to [29], GAT claimed that a stacking of first-order aggregation renders similar effects.

$$\alpha_{\{i,j\}} = \text{softmax}_j(e_{i,j}) = \exp(e_{\{i,j\}}) / \sum_{k \in N_i} \exp(e_{\{i,k\}}) \quad (5.6)$$

Eq. (5.6) gave the expression of a normalized attention coefficient of node i on its neighbor j . To learn the node-wise representations that best compute point-wise attentions, GAT first applied a linear transformation W on graph's node features, the formulation of single attentional layer:

$$\alpha_{\{i,j\}} = \frac{\exp(\text{LeakyReLU}(a^T [Wx_i || Wx_j]))}{\sum_{k \in N_i} \exp(\text{LeakyReLU}(a^T [Wx_i || Wx_k]))} \quad (5.7)$$

where $||$ represents a concatenation operation, and the weight vector $a \in R^{2d}$ need be learned. As output of the layer, once attention coefficients are learned, aggregated feature on node i :

$$x'_i = \sigma \left(\sum_{\{j \in N_i\}} \alpha_{\{i,j\}} Wx_j \right) \quad (5.8)$$

In which σ is the non-linear activation function. Also, similar to [35], GAT is able to extend as multi-head attention by applying K independent linear node feature transformation and attention vector parametrized as (W^k, a^k) . Then layer output is the concatenated feature: $x'_i = \parallel_{\{k\}} x_i^{(k)}$. In [18], we combined AGCN and GAT and introduced a novel graph attention network based on adaptive graph. With a jointly learned neighbors N_i and attention coefficient $\alpha_{\{i,j \in N_i\}}$, our new network is able to outperform others in sophisticated image understanding tasks, such as survival prediction, on benchmark datasets.

5.2.4 Neural network for survival analysis

Survival analysis [38] is a set of statistical inference models where the output is the elapsed time until a pre-defined event occurs. The event can be anything of interest, ranging from vehicle part failures to adverse drug reactions. Clinical trials are aimed to assess different treatment regimens with biological death as the primary event of interest to observe. An accurate estimate of survival probability provides invaluable information for clinical interventions.

The Cox proportional hazards model [39] is the most popular model in survival analysis. While, the classical Cox model and its early followers overly simplified the patient's survival probability as linear mapping from covariates. Recently, Katzman et-al have designed a fully connected network (FCN) called *DeepSurv* [40] to learn a nonlinear mapping of covariates to the representations in survival prediction. Although the neural networks [40,41] outperformed the linear Cox survival model, their networks cannot directly work on pathological images. Along with the success of convolutional neural networks (CNNs) on generic images, pathological image, as well as CT [42] and MRI [43], have become ideal data sources for training DL-based survival models. Among them, whole slide images are one of the most valuable data formats due to their massive multi-level pathological information on nidus and its surrounding tissues [44,45].

WSISA [46] was the first success of introducing whole slide pathological images (WSIs) as major data source to survival prediction. Because the data size of single whole slide pathological image is usually at gigabyte level, to have a cost-efficient algorithm, most of existing methods on WSIs, including [45,46], are based on a set of patches with reasonable size, like 128×128 , as inputs. Therefore, a patch sampling on WSIs is required before running the algorithm. However, *WSISA* model [46] comprises a series of CNNs, each of which was trained with a cluster of similar patch samples collected from all training WSIs, respectively. Therefore, the representations extracted by CNNs were over-localized for WSIs since their receptive field is constrained to be less than a patch's size equivalent to a physical area of 0.063 mm^2 . The pathological sections of nidus from patients contain more than the regions of interest (e.g., tumor cells), therefore, the representations drawn from random patches may not strongly correspond to the disease. Furthermore, it has been widely recognized that the topological properties of instances on pathological images are crucial for a wide range of

medical tasks, including cell subtype classification and cancer classification. While, *WSISA* is neither able to learn any topological representations from WSIs nor to construct feature maps upon given topological structures.

5.3 Method

In the session, we introduce the spectral convolutional network built on adaptive residual graph, that is AGCN, and the graph attentional layer on adaptive graph as well as the end-to-end whole-slide-image (WSI) based survival prediction that is DeepGraphSurv.

5.3.1 Spectral graph convolution-LL layer

As elaborated in Sessions 1 and 2, the existing GCNs are tackling two major challenges: (1) how to utilize the complete set of graph nodes and edges without making the model impossible to train; (2) how to balance the localized kernel and the global structures that generalize the network for high-level prediction tasks. To have a kernel that mixed both local and non-label features on graph, we insist to borrow the kernel defined on spectrum domain as [28] together with the K -rank Chebyshev expansion as approximation to control computational cost free of the graph size. Through tuning K , the layer has a configurable receptive field as classical CNNs. In practice, we set $K = 2$ in most cases. And we found it was also worth tuning K for best performance.

The root cause why previous GCNs cannot train entire graphs is that the diversity of graph structures and the skewed node degree distribution make it so infeasible to directly parameterize node neighborhood in any trivial formulation, given $O(2^{N^2})$ scale of possible node connectivity on graph of N nodes. To make graph trainable, we grant data the capability of self-construction of graph structure by training parameterized distance metric of nodes, so that the graph Laplacian itself becomes trainable and adaptive along with data. Given the learned distance metric weights M , we construct the residual graph, as supplement to the original graph, using node features x for each sample. Since the new spectral graph convolutional layer is executed on a learned graph Laplacian, we name it as *SGC-LL* layer with formulation as:

$$g_{\{\theta\}}(\Lambda) \times x \simeq \sigma \left(\sum_{k=0}^{K-1} \theta_k T_k(F(L, x, M))x \right) \quad (5.9)$$

In which $F(L, x, M)$ is the function to learn and update graph Laplacian for each graph sample. L is the free, original graph Laplacian. A trade-off coefficient λ_l is introduced in combination of original L and residual graph L_r in form of $L' = L + \lambda_l L_r$. And σ is the nonlinear activation (Fig. 5.3).

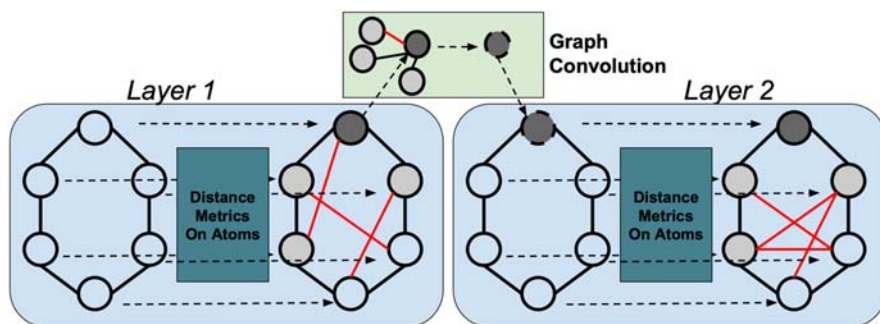


FIGURE 5.3

Illustration of execution of two consecutive SGC-LL layers. Red lines represent node-wise connections from residual graph. Convolution is running on (original + residual) graph.

5.3.1.1 Learning residual graph Laplacian

In real applications, some graph data come with their intrinsic graph topological structures, such as organic molecules. Molecules, when modeled as molecular graphs, have the atoms as graph nodes and the chemical bonds as graph edges. Importantly all the chemical bonds are justified by chemical experiment. However, some data, on which GCNs is about to perform, do not naturally show any graph structures, for example 3D point-cloud data. Under these circumstances, we will need to construct graphs before feeding to GCNs. Besides above extreme cases, it is mostly likely that the original graphs, either from domain knowledge or obtained via a graph initialization, failed to effectively unveil the hidden substructures among the remote nodes on original graphs. The finite receptive field and the computational cost prevent one GCN layer from aggregating nodes graph-widely, ignoring the constraints of local neighborhood defined by graph.

Use chemical compounds and property prediction as example, the initial graph given by SMILES sequence of compound does not disclose anything on toxicity directly. The useful representations of toxicity on the compound are supposed to be composed of those atoms, while not necessarily supported by the bonds from the original graph. If learning on original graphs, it may be difficult to learn the optimal representations. Therefore, we introduce a so-called *residual graph*, defined as a supplemental graph, patching the substructure missing on original graph that may assists the learning. The residual graphs have the identical node number as original graphs. To get rid of the curse of varying graph size, in SGC-LL layer, the distance metrics are the parameters to learn.

For graph structured data, the Euclidean distance is not necessarily the optimal metric to measure node-wise similarity. In articles of metrics learning, the algorithms were divided into supervised and unsupervised learning [47]. The optimal metric obtained in unsupervised fashion minimizes the intra-cluster distances and

also maximizes the inter-cluster distances. For labeled datasets, the optimal metric is the one that minimizes the learning loss. The generalized *Mahalanobis* distance between two nodes (x_i, x_j) is formulated as:

$$D_{\{i,j\}} = \sqrt{(x_i - x_j)^T M (x_i - x_j)} \quad (5.10)$$

where M is a symmetric positive semi-definite matrix, decomposed as $M = W_d W_d^T$. And the dense transform matrices $W_d \in \mathbb{R}^{d \times d^*}$, converting features to the manifold where the Euclidean distance still serve, is one of the training parameters of SGC-LL layer. If $M = I$, Eq. (5.10) reduces to the Euclidean distance. Then, we normalize the distances in Eq. (5.10) via Gaussian kernel:

$$G_{\{i,j\}}^D = \left(\frac{1}{2\pi \text{var}(D)} \right) \exp\left(-\frac{D_{\{i,j\}}^2}{2 \text{var}(D)} \right) \quad (5.11)$$

To make computations efficient, a sparse graph is required. Therefore, thresholding on G is to only keep the significant connections in constructed residual graph with adjacency matrix: $A_r = \text{thred}(G^D)$. Therefore, $A_r(W^d)$ is a differentiable function of distance metrics parameter W^d . In training process, the gradients back-propagated from training loss update the distance metrics W^d , and then, in next forward-pass, the residual graphs in batch will be reconstructed using the updated W^d . By doing metric learning before building residual graph, we are supposed to obtain the optimized metric parameters W^{d^*} , applying which onto the node features is able to form the substructures that better serve the fitting of graph node embedding.

5.3.1.2 Re-parameterization on feature transform

At a classical convolution layer, given the input feature as $H \times W \times d'$ and the dimensionality of output feature as $H' \times W' \times d''$, each dimension of output features is the sum of the feature maps, each of which is convolved by a kernel independently. Therefore, each layer has $d' \times d''$ kernels to learn. It means that the resulted features are not only built upon the neighbor vertices, but also depend on the rest of intra-vertex features from input. However, on graph convolution, it is not theoretically explainable to construct and learn a separate topological structure (graph) for each feature dimension. To construct the mapping of both intra- and inter-vertex features, at SGC-LL layer, we introduce a linear feature transform matrix and bias vector applied on the output features. Based on, the re-parameterization on output feature is formulated as:

$$y = g_{\{\theta\}}(\Lambda) \times x \simeq \sigma\left(W \left(\sum_{k=0}^{K-1} \theta_k T_k(F(L, x, M)) \right) x\right) + b \quad (5.12)$$

in which the linear transformation matrix $W \in \mathbb{R}^{d' \times d''}$ and the bias vector $b \in \mathbb{R}^{d''}$ are trained together with the distance metrics W_d . In total, at each SGC-LL layer, we have parameters up to the scale of $O(d)$, where $d = \max\{d', d^*, d''\}$, independent of graph scale and degree. At next SGC-LL layer, the spectral filters will be built in another feature domain with different metrics.

Algorithm 5.1: SGC-LL layer

Input: $\mathbf{X} = \{x_i\}$, $\mathbf{L} = \{L_i\}$; **Parameters:** W_d, W, b, λ_l

1. **for** j -the sample in batch $\{X, L\}$ **do**
 2. $A_r^i \leftarrow \text{Eq. (5.10), Eq. (5.11)}$
 3. $L_r^i = I_N - D_r^{\{-\frac{1}{2}\}} A_r^i D_r^{\{-\frac{1}{2}\}}$
 4. $L' = L + \lambda_l L_r^i$
 5. $Y^i \leftarrow \text{Eq. (5.3)}$; K -order Chebyshev approximation
-

In the [Algorithm 5.1](#), the composition of introduced SGC-LL layer is elaborated. Using iteration is for ease of narrative, and the [Eqs. \(5.10, 5.11\)](#) were not explicitly expressed in batch-mode. As to the implementation, the for-loop in the algorithm, with no loss of generality, could be replaced using the batch-wise tensor multiplication operator given by PyTorch.

5.3.2 Adaptive graph convolution network architecture

The introduced new graph network is named as Adaptive Graph Convolution Network (AGCN), because SGC-LL layers are designed to efficiently learn topological structure of residual graphs, adaptive to both training data and context of training task. Besides SGC-LL layer, AGCN also comprises *graph pooling* layer and *graph gather* layer [\[48\]](#).

Graph Pooling The graph pooling operation is conducted feature-wisely. For node feature vector x_i , at the i th vertex of graph, the pooling operator replaces the j th feature value, $x_i(j)$, with the maximum at the j -th feature among i th vertex and its neighbors vertices, $x_i(j) = \max_{\epsilon \in \{i, N_i\}} x_\epsilon(j)$, if the layer is a max-pooling layer. If the layer is for avg-pooling, $x_i(j) = \text{mean}_{\epsilon \in \{i, N_i\}} x_\epsilon(j)$. In AGCN, due to graph structure is adaptive and being updated along training progress, the neighborhood of i th vertex is as well changing w.r.t the update of graph adjacency A_r .

Graph Gather The graph gather layer sums up all the vertex features along the feature dimension as the final representation of graph data. The output tensor at gather layer for a batch of B graphs is of shape $(B \times d)$, where d is the feature cardinality of vertex representations. It will be used as input for a graph-level classification or regression. Without a graph gather layer, the AGCN is also able to be trained and used for vertex-wise inference tasks. Training is executed with given labels on vertices or in a weakly-supervised fashion by replying on graph-level label alone. The vertex-wise predictions include graph completion and many predictions on social networks.

Bilateral filter layer. The purpose of using a bilateral filter layer [\[49\]](#) in AGCN is to proactively prevent over-fitting, consider the data scales of graph data are not comparable to other machine learning problems, for example ImageNet. Residual graphs definitely push the model to a better fitting to training

tasks, while, at the risk of over-fitting. To mitigate over-fitting, we introduced a revised bilateral filtering layer to regularize the activation from SGC-LL layer by augmenting the spatial locality of updated graph Laplacian L' . We also introduced batch normalization layers to avoid gradient explosion or vanishing.

Network configuration. The AGCN consists of multiple consecutive layer combos, the core layer of which is the SGC-LL layer. A classic layer combo consists of one SGC-LL layer, one batch normalization layer as well as one graph max pooling layer. See Fig. 5.4 for illustration. Since a residual graph Laplacian is learned at SGC-LL layer. At the graph pooling layer that follows, the updated graph Laplacian L'_i , of sample i , will replace L_i when finding neighborhood N_i until next SGC-LL layer. As last convolutional layer transformed features, at next SGC-LL layer, the residual graphs have to be reconstructed from the scratch. While, the learned L' will become the “original” graph Laplacian L at following layers.

Padding. Because for data like organic compounds, local sub-structures are decisive on chemical properties, for example toxicity. For instance, aromatic hydrocarbon is usually strongly toxic. However, if the hydrogen (H) atom was replaced by methyl group (-CH₃), the toxicity of compound would be significantly reduced. Therefore, graph coarsening or feature dropping/averaging will damage the completeness of informative local substructures, resulting in wrong predictions. Therefore, when preparing data, we pad X and graph Laplacian L tensors to as large as that of the maximum node per graph in the dataset. Then, when used at layers, we remove the zeros padded to X, L and execute the calculations, for example Algorithm 1, on graphs of original size. By doing this, we maintain a unified batch shape required by deep learning frameworks, for example PyTorch, without pruning any decisive local structures on graphs. While, data padding and recovery lead to extra computational cost, which is linear to graph size N .

To make predictions toward particular tasks, a classifier or a regressor need to be added on top of output graph embeddings. We can either simply do logistic

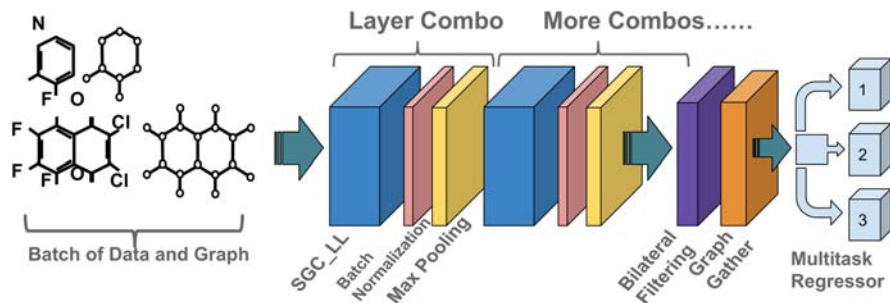


FIGURE 5.4

AGCN network. Directly feed it the graph-structured data and their original graphs.

regression or insert a fully connected layer before classifier. In experiments, we add one linear layer between last graph layer and the softmax. Besides, adding a linear layer make AGCN adaptable to a multi-task scenario, making the AGCN to deliver multiple predictions for different tasks in one-shot. We will present several experiments about using AGCN in multi-task learning.

5.3.3 Graph attention network on adaptive graph

As introduced in session 2.3, GAT [17] computes pair-wise attention coefficients between a node and its neighborhood using node features and learned transformation (W, a) shared across data. Given that the parameter scale of attentional layer is irrelevant to graph size and that GAT only consists of attentional layers, similar to AGCN, GAT is also able to accept graph data of diverse structures. However, attentional layers in GAT only aggregate features from attended neighbors, that is the first-order neighbors. Therefore, the representational capability of GAT is weaker than SGC-LL layers. Because at an SGC-LL layer there are up to K -hop neighbors included in kernel, equivalent to a receptive field of size $(2K + 1) \times (2K + 1)$ on grid. In practice, we set $K \geq 2$. While, GAT's attentional layers aggregate neighbors with an equivalent (3×3) kernel on grid.

In AGCN, if no graph gather layer added, the pooling layers and convolutional layers should not reduce graph size passed through the network, after removing the padded zeros, the output tensor should be of shape $(B \times N_i \times d)$, where B is batch size and N_i is the graph node count at i th sample and d is the feature dimensionality. If we set the output feature dimensionality at last SGC-LL as 1, then a stack of SGC-LL layers is also able to generate an attention mask on graph $G(V, E), \alpha \in R^{|V| \times 1}$. And then, different from the GAT aimed at learning graph embedding, the node-wise attention coefficients could be used in a *weighted* graph gathering:

$$y = \text{gather}(G) = \sum_{\{i \in V\}} \alpha_i x_i \quad (5.13)$$

where $x_i \in R^d$ is the i -th graph node embedding from AGCN for node-wise embedding and α_i is the attention coefficient from the attention net to be applied on every dimension of x_i . The output dimensionality at the weighted gather layer is $(B \times d)$, namely one d -dimensional vector for each graph. The node with a larger attention coefficient will be of more weight in the resulted final representation of graph.

As the attention coefficients $\alpha_i = \text{Attention}(G(V, E))$, are the direct output from the SGC-LL layer based attention net, when the final weighted embedding per graph y used in prediction, the gradients derived from loss will be propagated back to parameters in attention net. Consequently, the attention coefficients are updated along with node-wise embedding toward lowering the loss. The introduced joint end-to-end approach of attentional graph embedding learning is supposed to deliver better representation of graphs for a particular learning task. In

next section, we introduce one application of the attentional graph network on adaptive graph in a critical medical mission.

5.3.4 DeepGraphSurv framework

Medical image is a more direct observation compared to other formats of patient data toward an accurate survival time prediction. While, prior to CNNs, medical imaging analysis is based on handcrafted features, irrelevant to survival. On positive side, it has less chance of overfitting, but its accuracy and robustness are both unsatisfactory. CNNs are proved to be able to generate more comprehensive and generic representations of medical images. However, due to the tremendous data scale of whole slides pathological images (WSIs), no existing CNNs are able to accept the WSIs without down-sampling or cropping. State-of-the-arts of CNN based survival data models were all trained with sampled patches losing the informative global topological structures among patches, which is crucial for making decision for the entire WSI.

Graph is widely employed to represent topological structures among entities. However, modeling a WSI as graph is not straightforward. Cell-graph [50] is infeasible for tasks on WSIs due to the huge number of cells included and that many of them are possibly noisy nodes, that is isolated cells. To control the complexity of overall approach, the granularity of our model is set at patch level, for local substructure smaller than patches, we assumed that the CNN for patch feature extraction is able to represent them and include in patch embedding. To construct graphs for a WSI, patches become graph nodes, and the graph edges were to be built from the scratch. The extracted patch embeddings are, therefore, the original node features, when constructing node-wise connections using methods like clustering. Given a cluster, we set an edge appear on any two nodes belong to the same cluster.

Not all sampled patches will be used. For quality assurance purpose, we may have to dump some patches drawn from the marginal areas in which few cells are included. The extensive cleaning preprocessing was done via a visual check by professionals. Therefore, the cardinality of resulted patch samples per graph differs. Namely, the graphs that represent WSIs are of different number of nodes. In our experiment, vertex features are generated by the VGG-16 network pre-trained on ImageNet [51]. Due to the lack of patch labels, we cannot fine-tune the network on WSI patches. We will introduce how the proposed graph CNN model mitigates this deficiency in next session. The graph edges were initialized by thresholding the Euclidean distances between all patch pairs. The distances were calculated using the 128-dimensional node features that were first generated by a VGG-16 pre-trained network and then compressed by principal component analysis (PCA).

After graph is constructed, with minor change on output dimension, the SGC-LL layer is able to generate an attentional mask over graph, equivalent to the importance of each graph node in final graph representations. With a similar architecture

as AGCN proposed in [14], we created a graph attention network that comprises a stack of SGC-LL layers, parallel to the AGCN aimed at graph embedding learning, to learn node-wise attention coefficients. The output of the graph attentional mask is about to be applied in the final weighted gather layer, elaborated in Eq. (5.13).

As shown in Fig. 5.5, the end-to-end approach, named as DeepGraphSurv, first converts a WSI into a bag of patch sample represented as a 4096-dimensional embedding, then a graph was built using the compressed embeddings of nodes on cost-efficiency purpose, follow Eqs. (5.10) and (5.11). Different from previous deep learning-based survival models that simply act as feature extractor [46], DeepGraphSurv is able to directly generates predicted risks. We combine the regression of survival risk with the graph embedding learning on WSIs. The loss function is negative Cox log partial likelihood for censored survival data as shown in the following [39]:

$$L(R) = \sum_{i \in \{i: S_i = 1\}} (-R_i + \log \sum_{\{j \in \{j: T_j \geq T_i\}\}} \exp(R_j)) \quad (5.14)$$

S_i and T_i are respectively the censor status and the survival time measurement of i -th patient.

Eq. (5.14) is evaluated in batch-wise. R_i R_j are the risks given by DeepGraphSurv for sample i, j . A well-trained network is the one that gives a sample of longer survival time relatively less risk. During the training, the fine-tuned patch-wise embedding and the survival-related residual graph of WSIs are accessible at each SGC-LL layer of network, while the later layers usually provide more high-level topology-aware features about WSI. We also visualize the

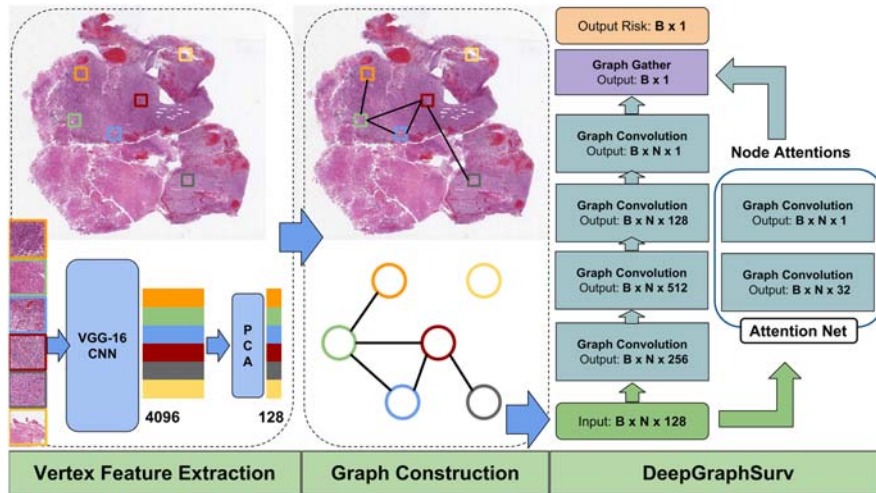


FIGURE 5.5

DeepGraphSurv: an WSI based end to end survival prediction workflow.

attention coefficients of graph node on the actual coordinates of corresponding patch on WSI in experiment session.

5.4 Experiment

5.4.1 Drug-property prediction

5.4.1.1 Baseline model

In experiment session, we compared the introduced AGCN with the state-of-the-art GNNs. [4], that constructs a spectral graph kernel by linear B-spline interpolation, is referred as *graphconv*. Neural fingerprint [20], referred as *NeuralFP*, is the graph neural network particularly designed for molecular graphs. It uses kernel constructed in spatial domain for each node cluster grouped by the same node degree. We refer to the spectral graph convolution equipped with a K -localized spectral filter as *GCN* [28], in which a Chebyshev approximation is applied for a fast evaluation of consecutive tensor multiplications. Using [28] as comparison is to demonstrate the additional knowledge representations empowered by the learned residual graphs. Graph attentional network (*GAT*) is also feasible for learning embeddings on molecular graphs, due to its definition that has no prerequisites on graph topology. As opposite, to enable the use of [4,28] for molecular graphs of different scales and structures, a graph pruning is inevitable, sacrificing performance. Besides, graph isomorphism network (*GIN*) [52] is introduced as a GNN that learns node embedding via MLP and the aggregated node and edge feature vectors. The graph-level representation of *GIN* is given by averaging the node embeddings. Those baseline models included for the experiment of molecular graph classification are more or less feasible with diverse graph inputs of varying node count and topological structure. Because the methods, for example *NeuralFP*, *GIN* and *GAT* execute the feature aggregation within the first-order neighbors, and parameterized transforms were built and initialized either for entire nodes or for node clusters [20]. Their over-localized, non-parametric kernels excluded most topological structure of molecular graphs from modeling. And that is the reason that AGCN outperformed the baselines on the drug classification benchmarks.

5.4.1.2 Dataset

The datasets used in this experiment are all about drug property prediction. Given labels obtained from the extensive lab or clinical experiment over a long list of organic compounds, it is possible to learn some patterns from the compounds related to certain predefined biochemical properties, such as toxicity and solubility. With the open-sourced cheminformatics software RDKit [53], it is straightforward to convert any compound to its corresponding graph representation, molecular graph, which consists of node list, edge list, node features and edge features. The node features extracted from molecular atoms: atomic number, atom

degree, formal charge, chiral tag, number of Hs, hybridization, the Boolean indicator on aromatic and the scaled atom mass. And the edge features include a 4-digit one-hot vector to represent bond type, the indicator on bond aromatic, as well as two Boolean features that determine if bond is conjugated or in-ring.

Downstream task datasets we utilized in experiment are 4 multi-task, binary drug classification dataset from MoleculeNet [54]. They are:

1. Tox21. Toxicity clinical data with labels on 12 tasks corresponding to different biological syndrome. Each label represents an observation toward one property of the compound.
2. Toxcast. Another toxicology measurement of drugs collected from the same initiative as Tox21, providing toxicology experimental results for a large library of compounds based on in vitro high-throughput screening. It offers 617 experiments on over 8K compounds.
3. SIDER. A dataset that contains marketed drugs and adverse drug reaction (ADR) with 27 group of organ classes [55].
4. ClinTox. Collected data on drugs approved by the FDA and the drugs that failed clinical trials for toxicity reasons. The dataset has two binary labels: (1) clinical trial toxicity pass or fail; (2) FDA approval or not [56].

5.4.1.3 Experimental result

The baseline models and AGCN were trained with the identical training dataset and tested over the same dataset. To measure the classification accuracy, RoC-AUC was chosen as the metric for comparison. Because for the datasets such as Tox21, there are 12 individual classification tasks, the numbers presented are averaged RoC-AUC of each task. To remove randomness, we also applied 4-fold cross-validation and averaged the numbers. From the results listed in Table 5.1, it is obvious that the AGCN outperformed other four baselines on 3 of 4 datasets. While, at SIDER dataset, GAT gave the best classification accuracy. Considering that GAT is constructed using complete original graph, therefore, there is chance that GAT outperformed AGCN on some tasks. Given that AGCN has more

Table 5.1 Class-average ROC-AUC on test dataset.

	Tox21	Toxcast	ClinTox	SIDER
No. graph	7831	8575	1478	1427
No. task	12	617	2	27
NeuralFP [20]	0.7341	0.6384	0.7469	0.5525
GraphSAGE [26]	0.7470	0.6335	0.5924	0.6040
GCN [28]	0.7481	0.6739	0.7573	0.5914
GAT [17]	0.7540	0.6460	0.5886	0.6090
GIN [52]	0.7480	0.6340	0.5804	0.5730
AGCN	0.8016	0.7033	0.7688	0.5921

parameters to train, it is likely to have under-fitting issue on a small dataset like SIDER [55]. Recently, researchers have also found that GNNs could benefit from a self-supervised pre-training before fine-tuning toward classification tasks [57]. Table 5.1 also includes RoC-AUC of GAT and GIN tested over the aforementioned 4 molecular graph datasets reported in [57]. It showed that, even with well-designed pretrain task, their performances were still worse than AGCN, who did not experience pretrains, on 3 of 4 tasks. And because of more parameters introduced, AGCN had more significant advantage on relatively larger datasets which the model was able to fit better.

5.4.2 DeepGraphSurv and survival prediction

GNNs are not only feasible for learning powerful representation from molecular graph for graph classification tasks, they have also shown promising results on computer vision mission [58,59]. The key part of applying GNNs onto computer vision domain is how to properly define the graph structures on multimedia data. Using video understanding as example, if video frames defined as graph node and graph edges represent correlations between key frames, then video classification will be naturally modeled as graph classification problem. And the graph node classification and attribute regression are equivalent to video segmentation and video caption respectively. In this experiment, we introduce WSI-based survival prediction and how a WIS is represented as graph. Then, DeepGraphSurv and the baselines survival models will be compared on extensive datasets.

5.4.2.1 Dataset

As to experimental benchmark dataset, we utilized the whole slide pathological images (WSIs) from a generic cancer patient dataset TCGA [60], which was originally released by The Cancer Genome Atlas project, whose research objective is to discover correlation between genetic errors in DNA and the occurrence of 33 cancer subtypes. We trained and evaluated the baseline models and the introduced DeepGraphSurv over the WSIs associated with two common cancer subtypes from the TCGA dataset: glioblastoma multiforme (GBM) and lung squamous cell carcinoma (LUSC). Besides, NLST (National Lung Screening Trials [61]) is another medical research that employed 53,454 heavy smokers, whose age 55 to 74 with at least 30-year smoking history, as the high-risk patient group for lung cancer survival modeling and analysis. We also committed an experiment over the WSIs data of NLST that consists of both squamous-cell carcinoma (SCC) and adenocarcinoma (ADC) patients and their lung tissue images to evaluate the performance of our model on survival prediction for patients of mixed tumor subtype. The numeric facts on the datasets in the experiments are listed in Table 5.2. Some patients have multiple WSIs collected and included in dataset, we executed models on all data and report the average prediction per patient. Because the constructed graphs over patches on WSI is data-specific, not patient-specific instead. Data quality in Table 5.2 is mostly related to image resolution. Average size of

Table 5.2 The statistics of whole slides pathological image (WSI) datasets.

Data source	Cancer subtype	No. patient	No. WSI	Quality	Avg. size
TCGA	LUSC	463	535	Medium	0.74 Gb
TCGA	GBM	365	491	Low	0.50 Gb
NLST	ADC & SCC	263	425	High	0.74 Gb

single WSI is to emphasize the challenges in experiments, because loading these WSIs of that size to memory is already difficult.

5.4.2.2 Baseline model

The baseline models included in survival prediction experiments are divided into two categories: classic methods and deep learning-based end-to-end methods. Classic methods, such as LASSO-Cox linear model [39], BoostCI [62] and a multi-task learning framework proposed for Survival Analysis, called MTLSA [63] are not able to directly output survival probability or survival time, and the regression is executed after extracting features from the raw data, no matter the data type is text or image. Therefore, the performances of classic methods largely depend on the quality of extracted hand-crafted features from raw data. Unfortunately, they were entirely not designed for WSI-based survival analysis, which requires extensive calculation. And no hand-crafted features are designed particularly for images at the scale of WSI. So, for a fair comparison, we first feed those baselines with the selected predefined features extracted on patches using CellProfiler [64], an open-source scientific software for cell image analysis. Patches on WSI data were randomly sampled, and the final features of WSI were the averaged ones over those from corresponding patches. Furthermore, we also feed the classic baselines with the WSI-level features generated by DeepGraphSurv using the same sampled patches to demonstrate the gain of performance brought by the end-to-end fine-tuned topology-aware global features from the new network.

Besides classic baselines relying on pre-calculated features, we compared DeepGraphSurv with the cutting-edge deep learning-based survival models on WSI. WSISA [46] is the first approach that directly works on WSI. While, training of WSISA is expensive and unscalable. It is required to train a CNN, for example VGG-16, for each cluster of patch samples. Therefore, WSISA neglected the wide-existing topological relationships among the patches, which are of great value to survival analysis. As opposite, Graph CNNs are built upon the topological structures and have recognized power of learning structured features on graph-structured data. To demonstrate the capability of graph feature learning over data with no intrinsic graph structure given, we add another GCN as baseline method. We concatenate the latest spectral GCN model [28], as feature extractor, with a Cox regression function who gave probability prediction. While, [28] executed convolution over a pre-defined, fixed spectrum kernel, lack of adaptiveness if the initial graph is not good.

5.4.2.3 Experimental result

As far as we know, DeepGraphSurv is the first survival model that utilizes graph-based attention scheme. As shown in Fig. 5.6, after 40 epochs of training, the regions that comprise the patches of high attentional coefficients have correctly highlighted the most of Regions of Interest (RoIs) corresponding to tumor-related cell clusters. The ground-truth of RoIs were annotated by experts. The embedding learned on those patches of high attentional coefficient will be of higher weights in the final representation of WSI [see Eq. (5.13)], and then if the higher coefficients predicted by attentional network geographically coincide with the RoIs related to tumor cells, the final graph representation will be consequently more tumor-oriented and more helpful in survival prediction.

The concordance probability (C-index) is the measurement of survival prediction. It is defined as the fraction of all pairs of patients whose predicted survival times/risks are correctly ordered as all censored patients that can be reasonably ordered. Formulating survival order as directed graph $G_i(D, \Xi)$ where the edge $\xi_{\{i,j\}}$ implies survival time $T_i < T_j$, then C-index of graph G_i , given the risk prediction $f(x)$, is defined as:

$$C(D, G_i, f(x)) = 1/|\xi| \sum_{\xi_{\{i,j\}}} 1_{f(x_i) < f(x_j)}$$

where $1_{f(x_i) < f(x_j)}$ is the indicator function formulated as: $1_{\{a < b\}} = 1$ if $a < b$, otherwise 0. $f(x)$ is the risk predicted by survival model for WSI x . When a patient has more than one WSI included in dataset, we average the predicted risks for this patient before calculating the C-index. A model who is able to more correctly order the censored patients by the predicted survival time is supposed to have higher C-index. The C-index result of the baseline models and the introduced DeepGraphSurv from three experiments are reported in Table 5.3.

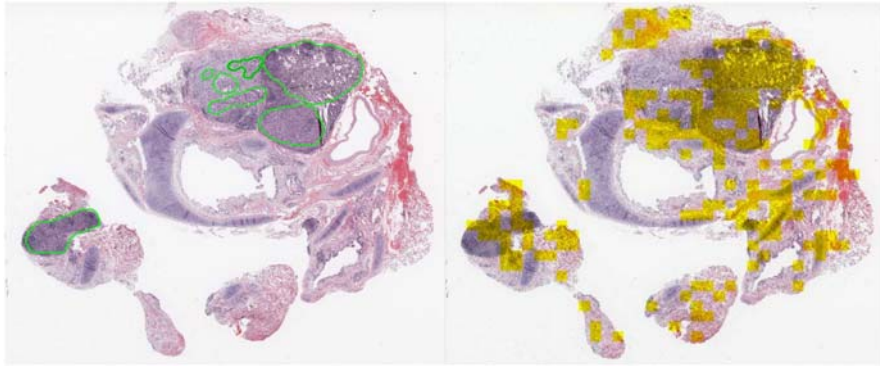


FIGURE 5.6

Predicted RoIs from DeepGraphSurv: Yellow colored regions in Right are the patches of high attentional coefficient; Green lines in Left drew the ground-truth RoIs.

Table 5.3 C-index on testing set for survival models.

Model	LUSC	GBM	NLST
LASSO-Cox	0.5280	0.5280	0.4738
LASSO-Cox + DeepGraphSurv feature	0.5663	0.5165	0.5663
BoostCI	0.5633	0.5543	0.5705
BoostCI + DeepGraphSurv feature	0.5800	0.5130	0.5716
EnCox	0.5216	0.5597	0.4883
EnCox + DeepGraphSurv feature	0.5740	0.5231	0.5742
RSF	0.5066	0.5570	0.5964
RSF + DeepGraphSurv feature	0.5492	0.5193	0.5491
MTLSA	0.5386	0.5787	0.6042
MTLSA + DeepGraphSurv feature	0.5247	0.5630	0.5573
WSISA	0.6380	0.5760	0.6539
GCN + Cox	0.6280	0.5901	0.6845
<i>DeepGraphSurv</i>	<i>0.6606</i>	<i>0.6215</i>	<i>0.7066</i>

Testing dataset was separated from the training set to avoid data leakage. The classic survival models, for example LASSO-Cox [39], failed to deliver compelling prediction accuracy on WSI datasets. Possible explanation may be: (1) sampled patches are only part of a WSI given the computational constraint; (2) the data quality of patches may vary case by case. Therefore, the features extracted from randomly sampled patches bring a noisy and biased representation of WSIs. Moreover, the hand-crafted features offered by CellProfiler are the generic descriptors of pathological images, not particularly designed for images at the scale of WSI. So, we believe it is the quality of hand-crafted features that limit the performance of classic survival models. After feeding the topology-aware graph embedding generated by DeepGraphSurv to classic baselines as input, the C-index showed a notable lift as large as 0.04 on average on NLST and LUSC datasets. And the outcome supported our argument that the features fine-tuned with survival labels are better representations of WSIs for survival prediction purpose. However, we also observe that, due to the lower image quality of GBM WSI data, only using the fine-tuned patch features cannot improve prediction accuracy. Because we see that WSISA [46], a CNN based patch feature extractor, cannot beat MTLISA [63], which still use hand-crafted features as input. This means that, given a low image quality, CNNs cannot always learn a better representation of image from scratch than heuristic features, even though survival censorship labels were used as penalty in the training of CNNs.

DeepGraphSurv generates predictions by encoding patch features with their topological structure via spectral graph convolutions. When the patch features are not discriminative, the topological structure among the patch instances will play a significant role in recognition of survival patterns of WSI. This explains the additional accuracy given by DeepGraphSurv compared to WSISA on GBM dataset,

which is of lower resolution. On other dataset of higher quality, DeepGraphSurv is able to deliver a larger margin of gain on C-index compared to the baselines who cannot learn anything from topological structures. The baseline *GCN + Cox* is the model in which we use the graph representation learned from GCN [28] with LASSO-Cox survival regression. Compared to this baseline, on all 3 datasets DeepGraphSurv still showed a lift, that comes from a better hidden representation disclosed on residual graphs built upon the learned distance metrics. Due to that WSI data do not have any intrinsic graph structures, the initial graphs constructed with the patch embeddings from a VGG-16 network are not guaranteed to be optimal in terms of learning graph representations. While, DeepGraphSurv makes graph structure trainable and adaptive in learning graph representations. Besides, the introduced end-to-end approach is able to directly generate risk prediction, the entire network including the survival regressor are optimized jointly.

5.5 Conclusion

In this article, we introduce a novel GCN with a new spectral graph convolution layer (SGC-LL) that works on trainable, adaptive graph. SGC-LL learns the residual graph Laplacian via learning the optimal distance metric and the feature transform that best serve model fitting. The AGCN is the first GCN that has no prerequisites on topological structures and graph size. By combining with a node-wise attention network trained over graph, AGCN also presented outstanding representation learning capability on sophisticated image understanding tasks, for example survival prediction on whole slide images (WSIs). Efficient discovering the survival-related sub-structures over patches sampled from a whole slide image is proved to be a promising solution to boost accuracy of survival prediction. Extensive multi-task graph classification experiments over various molecular graph data indicated that the AGCN outperformed the state-of-the-art graph network models in terms of accuracy. For survival prediction, we first modeled WSI as graph and introduced an end-to-end framework called DeepGraphSurv to initialize and learn a topological-aware representations for each WSI. Instead of unsupervised graph, DeepGraphSurv creatively utilized a survival-specific graph trained under supervision of survival labels. The effectiveness of introduced method has been verified by improved accuracy of risk ranking on multiple cancer patient datasets across carcinoma subtypes.

Regarding future work, pre-training strategy is a promising technique for biochemical applications in which ground-truth labels are expensive to collect. Owing to the limit of data scale, a deep network is easy to be under-fitting, and the data augmentation is difficult to be applied to molecular graph, which is derived from chemical formula, and a random manipulation of original graph may not generate any usable data. Therefore, pre-training is also supposed to help convergence of AGCN on small training data with a skewed class distribution.

References

- [1] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, *Advances in Neural Information Processing Systems* (2012) 1097–1105.
- [2] I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning*, MIT Press, 2016.
- [3] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, L. Fei-Fei, Large-scale video classification with convolutional neural networks, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1725–1732.
- [4] J. Bruna, W. Zaremba, A. Szlam, Y. LeCun, Spectral networks and locally connected networks on graphs, 2013. arXiv preprint arXiv:1312.6203.
- [5] M. Henaff, J. Bruna, Y. LeCun, Deep convolutional networks on graph-structured data, 2015. *arXiv:1506.05163*.
- [6] D.I. Shuman, S.K. Narang, P. Frossard, A. Ortega, P. Vandergheynst, The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains, *IEEE Signal Process. Mag.* 30 (3) (2013) 83–98.
- [7] R. Shapovalov, E. Velizhev, O. Barinova. Nonassociative markov networks for 3d point cloud classification, in: *Proceedings of the International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XXXVIII*, 2010, Part 3A.
- [8] Y. Zhang, M. Rabbat, A graph-CNN for 3D point cloud classification, in: *Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2018, pp. 6279–6283.
- [9] S. Yan, Y. Xiong, D. Lin, Spatial temporal graph convolutional networks for skeleton-based action recognition, in: *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.
- [10] B. Zhang, M. Al Hasan, November. Name disambiguation in anonymized graphs using network embedding, in: *Proceedings of the ACM on Conference on Information and Knowledge Management*, 2017, pp. 1239–1248.
- [11] A. Sandryhaila, J.M. Moura, Discrete signal processing on graphs: Graph Fourier transform, in: *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, IEEE, 2013, pp. 6167–6170.
- [12] F. Shoeleh, M. Asadpour, Graph based skill acquisition and transfer learning for continuous reinforcement learning domains, *Pattern Recognit. Lett.* 87 (2017) 104–116.
- [13] R. De Maesschalck, D. Jouan-Rimbaud, D.L. Massart, The mahalanobis distance, *Chemom. Intell. Lab. Syst.* 50 (1) (2000) 1–18.
- [14] R. Li, S. Wang, F. Zhu, J. Huang, Adaptive graph convolutional neural networks, in: *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.
- [15] B. Wu, A. Wan, X. Yue, K. Keutzer, Squeezeseg: convolutional neural nets with recurrent crf for real-time road-object segmentation from 3D lidar point cloud, in: *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2018, pp. 1887–1893.
- [16] T.H. Scheike, M.J. Zhang, Extensions and applications of the Cox-Aalen survival model, *Biometrics* 59 (4) (2003) 1036–1045.
- [17] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, Y. Bengio, Graph attention networks, 2017. *arXiv:1710.10903*.
- [18] R. Li, J. Yao, X. Zhu, Y. Li, J. Huang, Graph CNN for survival analysis on whole slide pathological images, in: *Proceedings of the International Conference on*

- Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2018, pp. 174–182.
- [19] A. Kazi, L. Cosmo, N. Navab, M. Bronstein, Differentiable Graph Module (DGM) graph convolutional networks, 2020. *arXiv:2002.04999*.
- [20] D.K. Duvenaud, D. Maclaurin, J. Iparraguirre, R. Bombarell, T. Hirzel, A. Aspuru-Guzik, et al., Convolutional networks on graphs for learning molecular fingerprints, *Advances in Neural Information Processing Systems (2015)* 2224–2232.
- [21] D.K. Hammond, P. Vandergheynst, R. Gribonval, Wavelets on graphs via spectral graph theory, *Appl. Comput. Harmon. Anal.* 30 (2) (2011) 129–150.
- [22] J. Atwood, D. Towsley, Diffusion-convolutional neural networks, *Advances in Neural Information Processing Systems (2016)* 1993–2001.
- [23] C. Zhuang, Q. Ma, April. Dual graph convolutional networks for graph-based semi-supervised classification, in: *Proceedings of the World Wide Web Conference*, 2018, pp. 499–508.
- [24] M. Niepert, M. Ahmed, K. Kutzkov. Learning convolutional neural networks for graphs, in: *Proceedings of the International Conference on Machine Learning*, 2016, June, pp. 2014–2023.
- [25] H. Gao, Z. Wang, S. Ji, Large-scale learnable graph convolutional networks, in: *Proceedings of the Twenty-Fourth ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2018, July, pp. 1416–1424.
- [26] F. Monti, D. Boscaini, J. Masci, E. Rodola, J. Svoboda, M.M. Bronstein. Geometric deep learning on graphs and manifolds using mixture model CNNs, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 5115–5124.
- [27] M.S. Attene-Ramos, N. Miller, R. Huang, S. Michael, M. Itkin, R.J. Kavlock, et al., The Tox21 robotic platform for the assessment of environmental chemicals—from vision to reality, *Drug Discov. Today* 18 (15–16) (2013) 716–723.
- [28] M. Defferrard, X. Bresson, P. Vandergheynst, Convolutional neural networks on graphs with fast localized spectral filtering, *Advances in Neural Information Processing Systems (2016)* 3844–3852.
- [29] T.N. Kipf, M. Welling. Semi-supervised classification with graph convolutional networks, 2016. *arXiv:1609.02907*.
- [30] K. Liu, X. Sun, L. Jia, J. Ma, H. Xing, J. Wu, et al., Chemi-Net: a molecular graph convolutional network for accurate drug property prediction, *Int. J. Mol. Sci.* 20 (14) (2019) 3389.
- [31] P.C. Rathi, R.F. Ludlow, M.L. Verdonk, Practical high-quality electrostatic potential surfaces for drug discovery using a graph-convolutional deep neural network, *J. Med. Chem.* (2019).
- [32] Y. Pathak, S. Laghuvarapu, S. Mehta, U.D. Priyakumar. Chemically interpretable graph interaction network for prediction of pharmacokinetic properties of drug-like molecules, in: *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 1, pp. 873–880, 2020, April.
- [33] K. Do, T. Tran, S. Venkatesh. Graph transformation policy network for chemical reaction prediction, in: *Proceedings of the Twenty-Fifth ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019, July, pp. 750–760.
- [34] S. Ishida, K. Terayama, R. Kojima, K. Takasu, Y. Okuno, Prediction and interpretable visualization of retrosynthetic reactions using graph convolutional networks, *J. Chem. Inf. Model.* 59 (12) (2019) 5026–5033.

- [35] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, et al., Attention is all you need, *Adv. Neural Inf. Process. Syst.* (2017) 5998–6008.
- [36] R. Paulus, C. Xiong, R. Socher. A deep reinforced model for abstractive summarization, 2017. *arXiv:1705.04304*.
- [37] A.P. Parikh, O. Täckström, D. Das, J. Uszkoreit. A decomposable attention model for natural language inference, 2016. *arXiv:1606.01933*.
- [38] R.G. Miller Jr, *Survival Analysis*, 66, John Wiley & Sons, 2011.
- [39] D.R. Cox, *Regression models and life-tables*, *J. R. Stat. Soc., B: Stat. Methodol.* 34 (2) (1972) 187–202.
- [40] J.L. Katzman, U. Shaham, A. Cloninger, J. Bates, T. Jiang, Y. Kluger, Deep survival: a deep cox proportional hazards network, *Stat 1050* (2) (2016).
- [41] V.S. Dave, M. Al Hasan, B. Zhang, C.K. Reddy, Predicting interval time for reciprocal link creation using survival analysis, *Social Netw. Anal. Min.* 8 (1) (2018) 16.
- [42] C. Yan, J. Yao, R. Li, Z. Xu, J. Huang, Weakly supervised deep learning for thoracic disease classification and localization on chest x-rays, in: *Proceedings of the ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, 2018, August, pp. 103–110.
- [43] R. Li, Y. Li, R. Fang, S. Zhang, H. Pan, J. Huang, Fast preconditioning for accelerated multi-contrast MRI reconstruction, in: *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, Cham, 2015, October, pp. 700–707.
- [44] R. Li, J. Huang. Fast regions-of-interest detection in whole slide histopathology images, in: *Proceedings of the International Workshop on Patch-based Techniques in Medical Imaging*. Springer, Cham, 2015, October, pp. 120–127.
- [45] X. Zhu, J. Yao, J. Huang, December. Deep convolutional neural network for survival analysis with pathological images, in: *Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2016, pp. 544–547.
- [46] X. Zhu, J. Yao, F. Zhu, J. Huang. WSISA: making survival prediction from whole slide histopathological images, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 7234–7242.
- [47] F. Wang, J. Sun, *Survey on distance metric learning and dimensionality reduction in data mining*, *Data Min. Knowl. Discov.* 29 (2) (2015) 534–564.
- [48] J. Gomes, B. Ramsundar, E.N. Feinberg, V.S. Pande. Atomic convolutional networks for predicting protein-ligand binding affinity, 2017. *arXiv:1703.10603*.
- [49] R. Gadde, V. Jampani, M. Kiefel, D. Kappler, P.V. Gehler. Superpixel convolutional networks using bilateral inceptions, in: *Proceedings of the European Conference on Computer Vision*, Springer, Cham, 2016, October, pp. 597–613.
- [50] C. Gunduz, B. Yener, S.H. Gultekin, The cell graphs of cancer, *Bioinformatics* 20 (Suppl. 1) (2004) i145–i151.
- [51] J. Deng, W. Dong, R. Socher, L.J. Li, K. Li, L. Fei-Fei. Imagenet: a large-scale hierarchical image database, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2009, June, pp. 248–255.
- [52] K. Xu, W. Hu, J. Leskovec, S. Jegelka, How powerful are graph neural networks? 2018. *arXiv:1810.00826*.
- [53] G. Landrum. RDKit: open-source cheminformatics, 2006.

- [54] Z. Wu, B. Ramsundar, E.N. Feinberg, J. Gomes, C. Geniesse, A.S. Pappu, et al., MoleculeNet: a benchmark for molecular machine learning, *Chem. Sci.* 9 (2) (2018) 513–530.
- [55] H. Altae-Tran, B. Ramsundar, A.S. Pappu, V. Pande, Low data drug discovery with one-shot learning, *ACS Cent. Sci.* 3 (4) (2017) 283–293.
- [56] K.M. Gayvert, N.S. Madhukar, O. Elemento, A data-driven approach to predicting successes and failures of clinical trials, *Cell Chem. Biol.* 23 (10) (2016) 1294–1301.
- [57] W. Hu, B. Liu, J. Gomes, M. Zitnik, P. Liang, V. Pande, et al. Strategies for pre-training graph neural networks, 2019, *arXiv:1905.12265*.
- [58] L. Liang, L. Jin, Y. Xu, Adaptive GNN for image analysis and editing, *Advances in Neural Information Processing Systems*, 2019, pp. 3643–3654.
- [59] J. Gao, C. Xu, CI-GNN: building a category-instance graph for zero-shot video classification, *IEEE Trans. Multimed.* (2020).
- [60] C. Kandath, M.D. McLellan, F. Vandin, K. Ye, B. Niu, C. Lu, et al., Mutational landscape and significance across 12 major cancer types, *Nature* 502 (7471) (2013) 333–339.
- [61] B.S. Kramer, C.D. Berg, D.R. Aberle, P.C. Prorok. Lung cancer screening with low-dose helical CT: results from the National Lung Screening Trial (NLST), 2011.
- [62] A. Mayr, M. Schmid, Boosting the concordance index for survival data—a unified framework to derive and evaluate biomarker combinations, *PLoS One* 9 (1) (2014) e84483.
- [63] Y. Li, J. Wang, J. Ye, C.K. Reddy. A multi-task learning formulation for survival analysis, in: *Proceedings of the Twenty-Second ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, August, pp. 1715–1724.
- [64] M.R. Lamprecht, D.M. Sabatini, A.E. Carpenter, CellProfiler™: free, versatile software for automated biological image analysis, *Biotechniques* 42 (1) (2007) 71–75.

Further reading

- J. Du, S. Zhang, G. Wu, J.M. Moura, S. Kar, Topology adaptive graph convolutional networks. *arXiv:1710.10370*, 2017.
- Q. Zhang, J. Chang, G. Meng, S. Xu, S. Xiang, C. Pan, Learning graph structure via graph convolutional networks, *Pattern Recogn.* 95 (2019) 308–318.
- W. Hamilton, Z. Ying, J. Leskovec, Inductive representation learning on large graphs, *Advances in Neural Information Processing Systems*, 2017, pp. 1024–1034.
- K. Xu, C. Li, Y. Tian, T. Sonobe, K.I. Kawarabayashi, S. Jegelka. Representation learning on graphs with jumping knowledge networks. *arXiv:1806.03536*, 2018.
- S.G. Rohrer, K. Baumann, Maximum unbiased validation (MUV) data sets for virtual screening based on PubChem bioactivity data, *J. Chem. Inf. Model.* 49 (2) (2009) 169–184.
- G. Subramanian, B. Ramsundar, V. Pande, R.A. Denny, Computational modeling of β -secretase 1 (BACE-1) inhibitors using ligand based approaches, *J. Chem. Inf. Model.* 56 (10) (2016) 1936–1949.
- I.F. Martins, A.L. Teixeira, L. Pinheiro, A.O. Falcao, A Bayesian approach to in silico blood-brain barrier penetration modeling, *J. Chem. Inf. Model.* 52 (6) (2012) 1686–1697.

Deep slice interpolation via marginal super-resolution, fusion, and refinement

Cheng Peng¹, Wei-An Lin¹, Haofu Liao², Rama Chellappa¹ and Shaohua Kevin Zhou³

¹*Electrical and Computer Engineering (ECE), University of Maryland, College Park, MD, United States*

²*Computer Science, University of Rochester, Rochester, NY, United States*

³*Institute of Computing Technology, Chinese Academy of Science, Beijing, China*

6.1 Introduction

Magnetic resonance imaging (MRI) is one of the prevailing gold standards for diagnostic purposes. It is not only noninvasive but also better at targeting different human tissues with specific contrasts that reveal the underlying anatomy. The main disadvantage of MRI compared to other medical imaging modalities (e.g. computed tomography, or CT) is its long acquisition time, which is governed by the duration of the frequency signals to be emitted by atoms and sampled by the machine. There has been a long history of studies on accelerating the MRI sampling process [1–4] by undersampling in the 2D k -space during acquisition; however, only a relatively small number of studies [5–8] are focused on interpolating between the sampled slices.

In practice, most MR volumes are taken anisotropically with a high resolution within slices and a sparse resolution between slices. For example, Fig. 6.1 shows a brain MR scan whose axial direction is sparsely sampled. As a result, image quality suffers when viewing from coronal and sagittal directions.

It is desirable to have a consistent resolution across all dimensions, both for visualization and for medical analysis tasks such as brain volume estimation.

Traditionally, slice interpolation has been done with two groups of methods: intensity-based and deformation-based methods. Linear and cubic spline interpolation methods are classic examples of intensity-based methods that directly perform the interpolation based on the intensity of the adjacent slices. Deformation-based methods estimate deformation fields between adjacent slices and then interpolate in-between pixels based on the estimated fields. However, these methods require that adjacent MR slices contain similar anatomical structures. That is, the structural change must be sufficiently small so

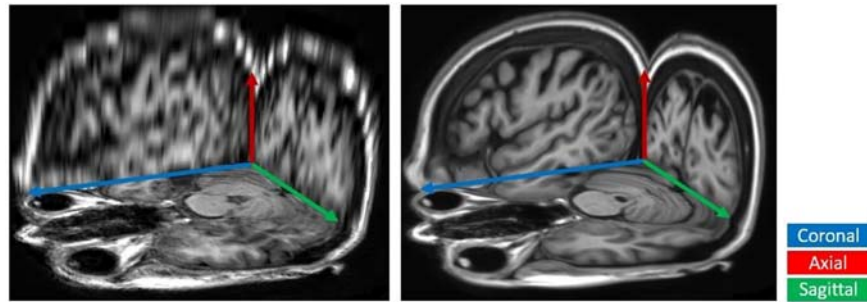


FIGURE 6.1

The axial, coronal, and sagittal views of an anisotropic MR volume are fitted to isotropic resolution through (*Left*) linear interpolation and (*Right*) our proposed slice interpolation method.

that a dense pixel correspondence can be established between adjacent slices. When the anatomical variation between slices is significant, more sophisticated modeling approach is needed.

Recently, deep convolutional neural networks (DCNNs) have been outperforming traditional approaches on medical image analysis due to their ability to model complex variations within data [4,9,10]. For slice interpolation, DCNNs can be applied to learn a mapping from an anisotropic MR to isotropic. However, directly addressing the task in 3D is challenging due to the high memory consumption of 3D networks. In this work, we break down the task of 3D slice interpolation into a sequence of 2D problems to produce anatomically consistent slice interpolations while being memory-feasible. Specifically, we propose a novel marginal super-resolution (MSR) to super-resolve isotropic views in the sagittal and coronal directions by a 2D CNN. The interpolation along the axial direction can be estimated by a fusion of the isotropic sagittal and coronal views. Finally, the interpolated slices are processed to recover more details via refinement.

Our main contributions can be summarized as follows:

1. We propose a novel MSR approach to break down the 3D slice interpolation problem into several 2D problems, which are more feasible in terms of GPU memory consumption and the amount of data available for training.
2. We propose a two-view fusion approach to incorporate the 3D anatomical structure. The interpolated slices after fusion achieve high structural consistency. The final refinement further recovers fine details.
3. We perform extensive evaluations on a large-scale MR dataset, and show that the proposed method outperforms all the competing CNN models, including 3D CNNs, in terms of quantitative measurement, visual quality, and brain matter segmentation.

6.2 Related work

6.2.1 Traditional slice interpolation methods

Early work on interpolating volumetric medical data dates back to 1992, when Goshtasby et al. [5] proposed to leverage the small and gradual anatomic differences between consecutive slices, and find correspondence between pixels by searching through small neighborhoods. A slew of methods were proposed in the subsequent years, focusing on finding more accurate deformation fields, including shape-based methods [6], morphology-based methods [7], registration-based methods [8], etc. Linear interpolation can be regarded as a special example, which essentially assumes no deformation between slices.

An important assumption of these methods is that adjacent slices contain similar anatomical structures, i.e., the changes in the structures have to be sufficiently small such that a dense correspondence can be found between two slices. This assumption largely limits the applicability of slice interpolation methods especially when slices are sparsely sampled. Furthermore, these methods did not utilize the information outside the two adjacent slices.

6.2.2 Learning-based super-resolution methods

Slice interpolation can be viewed as a special case of 3D super-resolution. Here, we review the literatures of 2D Single Image Super-Resolution (SISR), especially those approaches based on CNNs. Dong et al. [11] first proposed SRCNN, which learns a mapping that optimally transforms low-resolution (LR) images to high-resolution images. Many subsequent studies explored strategies to improve SISR such as using deeper architectures and weight-sharing [12–14]. However, these methods require bilinear upsampling as a preprocessing step, which drastically increases computational complexity [15]. To address this issue, Dong et al. [15] proposed to apply deconvolution layers for LR image to be directly upsampled to finer resolution. Furthermore, many studies have shown that residual learning provided better performance in SISR [16–18]. Specifically, Zhang et al. [18] incorporated both residual learning and dense blocks [19] and introduced Residual Dense Blocks (RDB) to allow for all layers of features to be seen directly by other layers, achieving state-of-the-art performance.

Generative Adversarial Networks (GAN) [20] have also been incorporated in SISR to improve the visual quality of the generated images. Ledig et al. pointed out that training SISR networks solely by L1 or L2 loss intrinsically leads to blurry estimations, and proposed SRGAN [17], which generated much sharper and realistic images compared to other approaches, despite having lower peak signal to noise ratios.

Though available computation capacity has been increasing, 3D CNNs are still limited by memory capacity due to a considerable increase in the size of network parameters and input data. A common compromise is to extract small patches from 3D volume to reduce the input size [21]; however, this also limits the

effective receptive field of the network. In practice, 3D CNNs are also limited by the amount of training data to ensure generalization.

6.3 Problem formulation and baseline convolutional neural networks approaches

Let $I(x, y, z) \in \mathbb{R}^{N \times N \times N}$ denote an isotropic MR volume. By convention, we refer the x axis as the ‘‘sagittal’’ axis, the y axis as the ‘‘coronal’’ axis, and the z axis as the ‘‘axial’’ axis. Accordingly, there are three types of slices:

1. the sagittal slice for a given x: $I^x(y, z) = I(x, y, z) \forall x$;
2. the coronal slice for a given y: $I^y(x, z) = I(x, y, z) \forall y$;
3. the axial slice for a given z: $I^z(x, y) = I(x, y, z) \forall z$.

We also define a slab of s slices, say along the x axis, as

$$\mathbb{I}^{x,s} = \left\{ I^{x+l}(y, z) \mid l = \frac{-(s-1)}{2}, \dots, 0, \dots, \frac{s-1}{2} \right\} \quad (6.1)$$

$\mathbb{I}^{y,s}$ and $\mathbb{I}^{z,s}$ are defined similarly. Without loss of generality, in this work we consider slice interpolation along the axial axis. From $I(x, y, z)$, the corresponding anisotropic MR volume is defined as

$$I_{\downarrow k}(x, y, z) = I(x, y, kz) \quad (6.2)$$

where k is the sparsity factor. The *goal* of slice interpolation is to find a transformation $T: \mathbb{R}^{N \times N \times \frac{N}{k}} \rightarrow \mathbb{R}^{N \times N \times N}$ that can optimally transform $I_{\downarrow k}(x, y, z)$ back to $I(x, y, z)$.

There are two possible baseline realizations of T using CNNs:

1. **2D CNN.** More in line with the traditional methods, a 2D CNN takes two adjacent slices $I_{\downarrow k}^z(x, y)$ and $I_{\downarrow k}^{z+1}(x, y)$ as inputs, and directly estimates the in-between missing slices. One major drawback of this approach is that a simple 2D CNN has limited capabilities of modeling the variations in highly anisotropic volumes.
2. **3D CNN.** A 3D CNN is learned as a mapping from the sparsely sampled volume $I_{\downarrow k}(x, y, z)$ to a fully sampled volume $I(x, y, z)$. This straightforward approach, however, suffers from training memory issue and insufficient training data.

Below, we present our proposed algorithm that retains the advantages of the baseline CNN models discussed above while mitigating their disadvantages.

6.4 The proposed algorithm

We propose to break down the 3D slice interpolation problem into a series of 2D tasks and interpolate the contextual information from all three anatomical views

to achieve structurally consistent reconstruction and improved memory efficiency. The two stages are as follows:

1. MSR, where we provide high-quality estimates of the interpolated slices by extrapolating context from sagittal and coronal axes.
2. Two-view Fusion and Refinement (TFR), where we fuse the estimations and further refine with information from the axial axis.

6.4.1 Marginal super-resolution

Fig. 6.2 demonstrates the pipeline of MSR. Given $I_{\downarrow k}(x, y, z)$, we view it as a sequence of 2D sagittal slices $I_{\downarrow k}^x(y, z)$ marginally from the sagittal axis. The same volume can also be treated as $I_{\downarrow k}^y(x, z)$ from the coronal axes. We make an observation that super-resolving $I_{\downarrow k}^x(y, z)$ to $I^x(y, z)$ and $I_{\downarrow k}^y(x, z)$ to $I^y(x, z)$ are equivalent to applying a sequence of 2D super-resolution along the x axis and y axis, respectively. Therefore, we apply a residual dense network (RDN) [18] \mathcal{M}_θ to upsample $I_{\downarrow k}^x(y, z)$ and $I_{\downarrow k}^y(x, z)$ as follows:

$$I_{sag}^x(y, z) = \mathcal{M}_\theta\left(\mathbb{I}_{\downarrow k}^{x,s}(y, z)\right), I_{cor}^y(x, z) = \mathcal{M}_\theta\left(\mathbb{I}_{\downarrow k}^{y,s}(x, z)\right) \quad (6.3)$$

Instead of super-resolving 2D slices independently, we propose to take a slab of s slices as input and estimate a single SR output. Using a larger s allows more context to be modeled. The MSR process is repeated for all x and y . Finally, the super-resolved slices can be reformatted as sagittally and coronally super-resolved volumes, $I_{sag}(x, y, z)$ and $I_{cor}(x, y, z)$, respectively. We apply the following L_1 loss to train the RDN:

$$L_{MSR} = \|\mathcal{M}_\theta\left(\mathbb{I}_{\downarrow k}^{x,s}\right) - I_{gt}^x\|_1 + \|\mathcal{M}_\theta\left(\mathbb{I}_{\downarrow k}^{y,s}\right) - I_{gt}^y\|_1 \quad (6.4)$$

Where $I_{gt}^x = I^x(y, z)$ and $I_{gt}^y = I^y(x, z)$ in the isotropic MR volume.

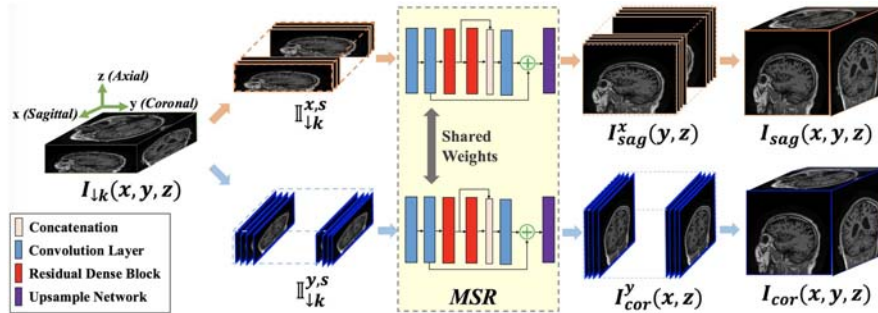


FIGURE 6.2

Marginal super-resolution pipeline.

From the axial perspective, $I_{sag}(x, y, z)$ and $I_{cor}(x, y, z)$ provide line-by-line estimations for the missing axial slices. However, since no constraint is enforced on the estimated axial slices, inconsistent interpolations lead to noticeable artifacts (See Section 6.5.4). We resolve this problem in the second TFR stage of the proposed pipeline.

6.4.2 Two-view fusion and refinement

The TFR stage is the counterpart of MSR which further improves the quality of slice interpolation by learning the structural variations along the axial direction.

As shown in Fig. 6.3, we first resample the sagittally and coronally super-resolved volumes $I_{sag}(x, y, z)$ and $I_{cor}(x, y, z)$ from the axial direction to obtain $I_{sag}^z(x, y)$ and $I_{cor}^z(x, y)$, respectively. A fusion network \mathcal{F}_ϕ takes the two slices as inputs and combines information from the two views. The objective function for training the fusion network is:

$$L_{fuse} = \|I_{fuse}^z(x, y) - I_{gt}^z\|_1 \quad (6.5)$$

Where $I_{fuse}^z(x, y) = \mathcal{F}_\phi(I_{sag}^z, I_{cor}^z)$ is the output of the fusion network, and $I_{gt}^z = I^z(x, y)$ in the isotropic MR volume. After training, the fusion network is applied to all the *interpolated* slices $\{I_{sag}^z | (z \bmod k) \neq 0\}$ and $\{I_{cor}^z | (z \bmod k) \neq 0\}$, yielding an MR volume $I_{fuse}(x, y, z)$.

After fusion, the interpolated slices already have visually pleasing qualities. Finally, to improve between-slice consistency along the axial axis, a refinement network \mathcal{R}_ψ takes a slab of $k + 1$ slices $\mathbb{I}_{fuse}^{z, k+1}$ as input and generates a consistent output slab $\mathbb{I}_{refine}^{z, k+1}$. The size is selected as $k + 1$ to make sure the refinement network has information from one or two observed slices. The pipeline is illustrated in Fig. 6.4. The loss function is given by:

$$L_{refine} = \|\mathbb{I}_{refine}^{z, k+1} - \mathbb{I}_{gt}^{z, k+1}\|_1 \quad (6.6)$$

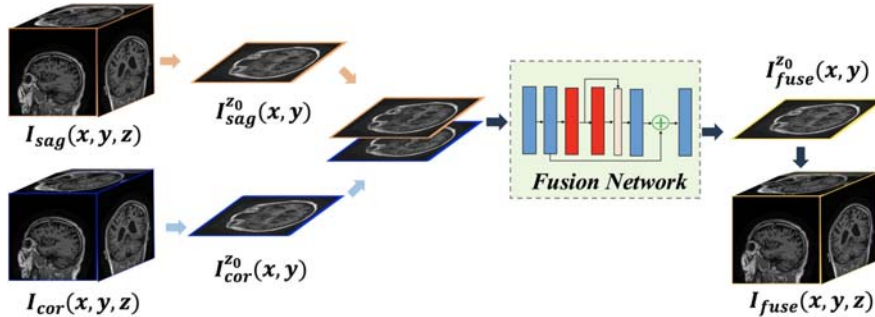


FIGURE 6.3

Two-view fusion pipeline.

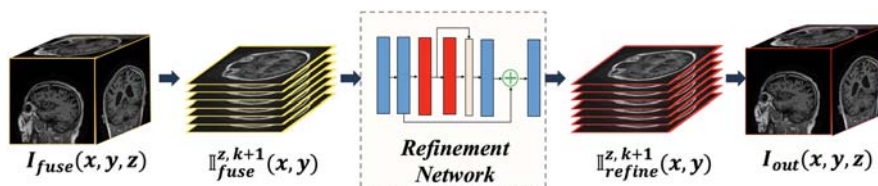


FIGURE 6.4

Refinement pipeline.

6.4.3 Comparison with baseline convolutional neural networks approaches

A 2D CNN estimates missing slices solely based on adjacent MR scans. In contrast, the proposed MSR and TFR take into account the full context from sagittal, coronal, and axial views, thus providing strong estimates of the in-between slices. A 3D CNN directly maps a sparsely sampled MR volume to a fully sampled MR volume. Due to memory limitation, a volume often needs to be divided into small patches during training, which limits the effective receptive field of 3D CNNs. In the proposed method, interpolation in 3D space is treated as a sequence of 2D operations, which ensures that the networks can be trained without relying on patches, thus allowing full contextual information to be captured. Furthermore, there are sufficient samples to train 2D CNNs, which mitigates the problem of overfitting issue that plagues 3D CNNs.

6.5 Experiments

6.5.1 Implementation details

We implement the proposed framework using PyTorch1. The RDN [18] architecture with two RDBs are used as the building unit for our networks. For Fusion, Refinement, and baseline 2D CNN models, where the inputs and outputs have the same image size, we replace the upsampling network in RDN with one convolutional layer. The input to the MSR network has $s = 3$. Note that due to memory constraint, 3D CNN only uses one RDB. We train the models with Adam optimization, with a momentum of 0.5 and a learning rate of 0.0001, until they reach convergence.

6.5.2 Dataset

We employ 120 T1 MR brain scans from the publicly available Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset. The MR scans are isotropically

Table 6.1 Quantitative evaluations for different slice interpolation approaches.

Sparsity	Method	PSNR(dB)	SSIM	DICE GM/WM	HD (90th pct.) GM/WM
4	LI	26.39	0.8317	0.7716/0.7296	3.607/7.965
	2D CNN	31.24	<i>0.9313</i>	<i>0.8813/0.8334</i>	3.176/12.36
	3D CNN	<i>31.34</i>	0.9292	0.8536/0.8265	<i>2.898/7.373</i>
	Ours	32.22	0.9441	0.9021/0.8593	2.494/6.240
8	LI	23.45	0.7165	0.6611/0.6105	4.487/10.59
	2D CNN	27.88	<i>0.8444</i>	<i>0.7783/0.7425</i>	<i>4.322/12.84</i>
	3D CNN	27.38	0.8390	0.7684/0.7468	4.583/9.017
	Ours	28.87	0.8808	0.8189/0.7828	3.960/8.127

For DICE and HD performance metrics, we present results on gray matter (GM)/white matter (WM) segmentation. The best results are in bold and the second best italic.

sampled at $1\text{ mm} \times 1\text{ mm} \times 1\text{ mm}$, and zero-padded to $256 \times 256 \times 256$ pixels, ending up with 30720 slices in each of sagittal, coronal, and axial directions.

We further down-sample the isotropic volumes by factors of $k = 4$ and $k = 8$, yielding $I_{\downarrow k}(x, y, z)$ of sizes $256 \times 256 \times 64$ and $256 \times 256 \times 32$, respectively. The data is split into training/validation/testing sets with 95/5/20 samples. Note that during test time, we only select slices that contain mostly brain tissues, the number of samples for each sparsity are presented in Table 6.1.

6.5.3 Evaluation metrics

We compare different slice interpolation approaches using two types of quantitative metrics. First, we use Peak Signal-to-Noise Ratio (PSNR) and Structured Similarity Index (SSIM) to measure low-level image quality. Second, we evaluate the quality of the interpolated slices through gray/white-matter segmentation. The segmentation network has a U-Net architecture, which is one of the winning models in MRBrainS challenge [22], and is trained on the OASIS dataset [23]. Dice Coefficient (DICE) and Hausdorff Distance (HD) between the segmentation maps of ground truth slices and generated slices are calculated. Due to the memory limitation of 3D CNN, we can at most super-resolve a limited region of $144 \times 144 \times 256$ pixels during evaluation. For fair comparisons, the evaluation metrics are calculated over the same region across all methods.

6.5.4 Visual comparisons

In Fig. 6.5, we present the observed slices $I_{\downarrow k}^z$ and $I_{\downarrow k}^{z+1}$ along with the interpolated slices produced by different methods. Specifically we demonstrate the second of three interpolated MR slices for 4x sparsity, and the third of seven interpolated slices for 8x sparsity. We highlight the region where the anatomical structures significantly change compared to the observed slices $I_{\downarrow k}^z$ and $I_{\downarrow k}^{z+1}$. We observe

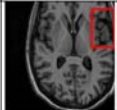
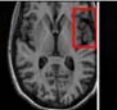
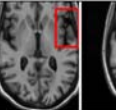
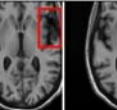
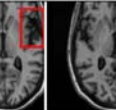
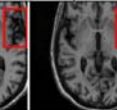

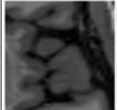
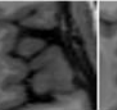
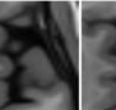
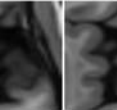
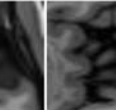
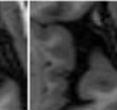

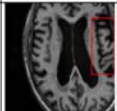
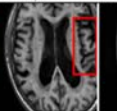
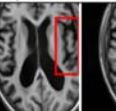
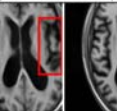
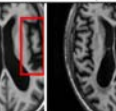
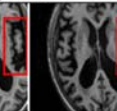

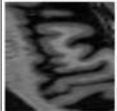
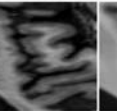
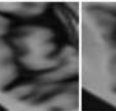
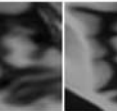
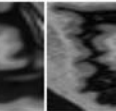
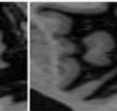

Sparsity	$I_{\downarrow k}^z$	LI	2D CNN	3D CNN	Ours	GT	$I_{\downarrow k}^{z+1}$
4							
							
27.37/0.8465 32.34/0.9441 32.72/0.9436 34.11/0.9607 PSNR(dB)/SSIM							
8							
							
25.51/0.7681 28.29/0.8205 29.51/0.8824 31.87/0.9249 PSNR(dB)/SSIM							

FIGURE 6.5

Visual comparisons of slice interpolation approaches. For $4 \times$ sparsity, the second of three interpolated MR slices is presented. For $8 \times$ sparsity, the third of seven interpolated slices is presented.

that although 2D CNN has comparable performance in terms of PSNR and SSIM, it tends to produce false anatomical structures in the zoomed regions. 3D CNN can resolve more accurate details. However, the improvement is quite limited, which we attribute to the fact that 3D CNN requires more training MR volumes to generalize and has smaller receptive field due to patch-based training. Our method benefits from the large receptive field of 2D CNN and two-view fusion, which not only produces sharper images, but also correctly estimates brain anatomy. The sharp and accurate estimation is crucial in clinical applications such as diagnosing Alzheimer’s Disease by brain volume estimation.

In Fig. 6.6, we demonstrate the advantage of the proposed method in brain matter segmentation. It is clear that although 2D and 3D CNN generates visually plausible interpolation as presented in Fig. 6.6, the brain matters are easily misclassified due to incorrect anatomical structures and blurred details.

6.5.5 Ablation study

In this section, based on 4x sparsity, we evaluate the effectiveness of each proposed components. The following settings are considered:

1. MSR_{sag}^n : Slice interpolation based on only sagittal view MSR. We consider number of input slices $n = 1, 3$.

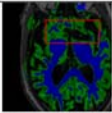
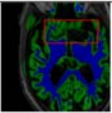
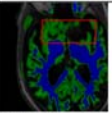
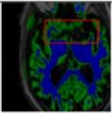
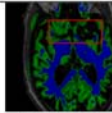
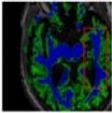
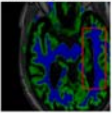
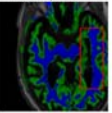
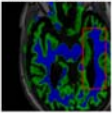
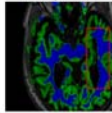
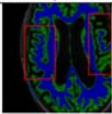
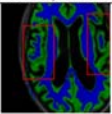
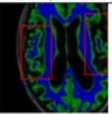
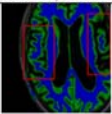
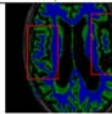
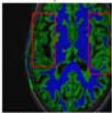
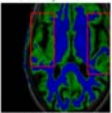
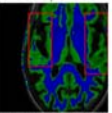
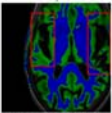
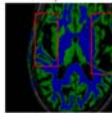
Sparsity	LI	2D CNN	3D CNN	Ours	GT
4	 0.6787/0.7972	 0.8143/0.8776	 0.8190/0.8714	 0.8664/0.9085	 GM/WM
	 0.6808/0.7161	 0.8103/0.8631	 0.7950/0.8606	 0.8598/0.9115	 GM/WM
8	 0.5139/0.7240	 0.6619/0.8224	 0.6878/0.8584	 0.7798/0.8853	 GM/WM
	 0.5910/0.6947	 0.6516/0.8021	 0.6507/0.8186	 0.7471/0.8540	 GM/WM

FIGURE 6.6

Visual comparison of gray matter (Green)/white matter (Blue) segmentation over different methods, with respective DICE scores listed under the images.

2. MSR_{cor}^n : Slice interpolation based on only coronal view MSR. We consider number of input slices $n = 1, 3$.
3. *Fused*: Slice interpolation with fusion network. Inputs to the network are MSR_{sag}^3 and MSR_{cor}^3 .
4. *Refined*: The proposed full pipeline.

From Table 6.2, it is clear that each proposed component improves the quality of slice interpolation. However, even without fusion and refinement, the axial slices interpolated by MSR_{sag}^3 and MSR_{cor}^3 are already better than the baseline 2D/3D CNNs.

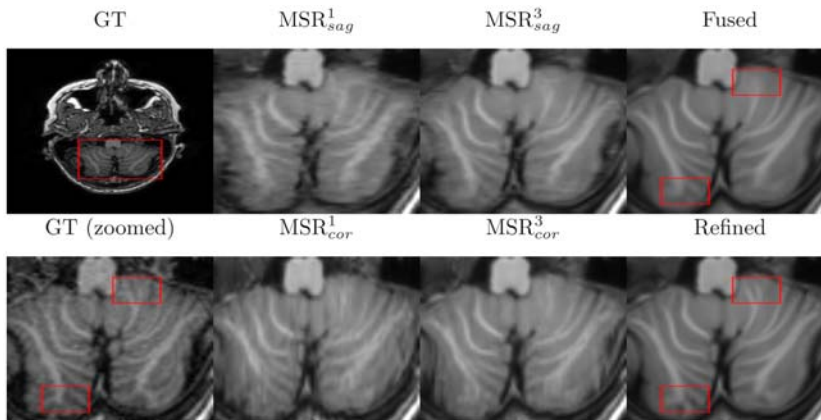
Visual comparisons are shown in Fig. 6.7, where we select a challenging slice with abundant anatomical details. From Fig. 6.7, it is clear that marginally super-resolving axial slices from coronal and sagittal views leads to noticeable horizontal (MSR_{sag}^3) and vertical (MSR_{cor}^3) artifacts. Furthermore, some small details are better resolved by MSR_{sag}^3 while others are better resolved by MSR_{cor}^3 . The fusion network combines the features from MSR_{sag}^3 and MSR_{cor}^3 , which effectively reduces inconsistency. With the additional axial information, the fused slice is then further improved by the refinement network.

In addition to L_1 loss, we also experiment on GAN loss at refinement stage. However, we find that GAN tends to generate fake anatomical details, which is undesirable in medical applications.

Table 6.2 Quantitative ablation study.

Stage	PSNR (dB)	SSIM
Baseline 2D CNN	31.24	0.9313
Baseline 3D CNN	31.34	0.9292
MSR_{sag}^1	30.28	0.9129
MSR_{cor}^1	30.56	0.9178
MSR_{sag}^3	31.43	0.9314
MSR_{cor}^3	31.61	0.9339
<i>Fused</i>	<i>32.02</i>	<i>0.9413</i>
<i>Refined</i>	32.22	0.9441

Baseline numbers are also included for comparison. The best results are in bold and the second best italic.

**FIGURE 6.7**

Visual comparison for the proposed components.

6.6 Conclusion

In this work, we proposed a multistage 2D CNN framework called deep slice interpolation. This framework allows us to recover missing slices with high quality, even when the distance between observed slices are sparsely sampled. We evaluated our approach on a large ADNI dataset, demonstrating that our method outperforms possible 2D/3D CNN baselines, both visually and quantitatively. Furthermore, we have illustrated that the MR slices estimated by the proposed method have superior segmentation accuracy. In the future, we plan to investigate the potential application of the proposed framework on real screening MRI, which often has a very low slice density.

References

- [1] S. Ravishankar, Y. Bresler, MR image reconstruction from highly undersampled k-space data by dictionary learning, *IEEE Trans. Med. Imaging* 30 (5) (2011) 1028–1041.
- [2] M. Lustig, D. Donoho, J.M. Pauly, Sparse mri: The application of compressed sensing for rapid mr imaging, *Magn. Resonan. Med.* 58 (6) (2007) 1182–1195.
- [3] S. Ma, W. Yin, Y. Zhang, A. Chakraborty, An efficient algorithm for compressed MR imaging using total variation and wavelets, in: 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2008), 24–26 June 2008, Anchorage, Alaska, USA, 2008.
- [4] J. Schlemper, J. Caballero, J.V. Hajnal, A.N. Price, D. Rueckert, A deep cascade of convolutional neural networks for dynamic MR image reconstruction, *IEEE Trans. Med. Imaging* 37 (2) (2018) 491–503.
- [5] A.A. Goshtasby, D.A. Turner, L.V. Ackerman, Matching of tomographic slices for interpolation, *IEEE Trans. Med. Imaging* 11 (4) (1992) 507–516.
- [6] G.J. Grevera, J.K. Udupa, Shape-based interpolation of multidimensional greylevel images, *IEEE Trans. Med. Imaging* 15 (6) (1996) 881–892.
- [7] T. Lee, W. Wang, Morphology-based three-dimensional interpolation, *IEEE Trans. Med. Imaging* 19 (7) (2000) 711–721.
- [8] G.P. Penney, J.A. Schnabel, D. Rueckert, M.A. Viergever, W.J. Niessen, Registration-based interpolation, *IEEE Trans. Med. Imaging* 23 (7) (2004) 922–926.
- [9] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015—18th International Conference Munich, Germany, 5–9 October 2015, Proceedings, Part III, 2015, pp. 234–241.
- [10] S. Liu, D. Xu, S.K. Zhou, T. Mertelmeier, J. Wicklein, A.K. Jerebko, et al., 3d anisotropic hybrid network: Transferring convolutional features from 2d images to 3d anisotropic volumes, *CoRR abs/1711.08580*, 2017.
- [11] C. Dong, C.C. Loy, K. He, X. Tang, Image super-resolution using deep convolutional networks, *CoRR abs/1501.00092*, 2015.
- [12] J. Kim, J.K. Lee, K.M. Lee, Accurate image super-resolution using very deep convolutional networks. *CoRR abs/1511.04587*, 2015.
- [13] K. Zhang, W. Zuo, S. Gu, L. Zhang, Learning deep CNN denoiser prior for image restoration, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, 21–26 July 2017, 2017, pp. 2808–2817.
- [14] J. Kim, J.K. Lee, K.M. Lee, Deeply-recursive convolutional network for image super-resolution, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, 27–30 June 2016, 2016, pp. 1637–1645.
- [15] C. Dong, C.C. Loy, X. Tang, Accelerating the super-resolution convolutional neural network. *CoRR abs/1608.00367*, 2016.
- [16] B. Lim, S. Son, H. Kim, S. Nah, K.M. Lee, Enhanced deep residual networks for single image super-resolution. *CoRR abs/1707.02921*, 2017.
- [17] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, et al.: Photo-realistic single image super-resolution using a generative adversarial network, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, 21–26 July 2017, 2017, pp. 105–114.

- [18] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, Y. Fu, Residual dense network for image super-resolution. CoRR abs/1802.08797, 2018.
- [19] G. Huang, Z. Liu, K.Q. Weinberger, Densely connected convolutional networks. CoRR abs/1608.06993, 2016.
- [20] I.J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, et al., Generative adversarial networks. CoRR abs/1406.2661, 2014.
- [21] Y. Chen, F. Shi, A.G. Christodoulou, Z. Zhou, Y. Xie, D. Li, Efficient and accurate MRI super-resolution using a generative adversarial network and 3d multilevel densely connected network. CoRR abs/1803.01417, 2018.
- [22] Z. Akkus, A. Galimzianova, A. Hoogi, D.L. Rubin, B.J. Erickson, Deep learning for brain MRI segmentation: State of the art and future directions, *J. Digital Imaging* 30 (4) (2017) 449–459.
- [23] D.S. Marcus, T.H. Wang, J. Parker, J.G. Csernansky, J.C. Morris, R.L. Buckner, Open access series of imaging studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults, *J. Cogn. Neurosci.* 19 (9) (2007) 1498–1507.

This page intentionally left blank

Explainable deep learning approach to predict chemotherapy effect on breast tumor's MRI

7

Mohammed El Adoui¹, Mohammed Amine Larhman², Stylianos Drisis³ and Mohammed Benjelloun⁴

¹Postdoctoral researcher, ILIA Department, Faculty of Engineering, University of Mons, Belgium

²VP of AI, Artificial Intelligence (AI) Department, Sol One, Bruges, Belgium

³Radiologist - Hospital Marie Curie, Charleroi, Belgium

⁴Professor, Faculty Of Engineering, ILIA Department, University of Mons, Belgium

7.1 Introduction

Patients with local breast tumors could systematically undergo to a neoadjuvant chemotherapy (NAC) before the surgery. The increase in the number of medical imaging examinations performed particularly for the diagnosis of breast cancer has facilitated the development of various techniques to support breast cancer monitoring. Among various imaging modalities, magnetic resonance imaging (MRI) is one of the most important tools in clinical diagnosis. In recent years, due to its excellent results, deep learning (DL) has been widely applied to the area of medical imaging. Indeed, this technique provides many sub-modalities including dynamic contrast-enhanced (DCE-MRI) and diffusion-weighted (DW-MRI) [1].

Deep convolutional neural networks (CNN) [2] produce multiple computer vision applications [3,4]. Various findings associated with healthcare and medical diagnosis like in lung lesion detection [5], breast cancer prognostic on mammograms image [6], cardiac anomalies detection [7], etc have been performed.

Intra-tumoral variations through the first chemotherapy have been regularly utilized as a guide to prognosticate the breast cancer response to NAC [8–11]. Accordingly, many image processing systems were introduced such as texture analysis [12] or voxel by voxel comparison [13,14]. However, lately, researchers have used many deep DL architectures to classify the intratumor response (responsive or nonresponsive) of breast cancer using DCE-MRI scans [15–17]. Recently, Richard Ha et al. [15] have evaluated a cohort of 3107 volumetric slices of 141 tumors on ten convolutional CNN layers. Authors in [18] proposed a method based on 4 principal treatments: pretreatment of mammograms scans,

features extraction by CNN, tumor detection, and ultimately, tumor labeling using YOLO architecture [19].

In the present research, we developed an original DL design to classify breast cancer response to NAC based on DCE-Mr images collected before and after the initial chemotherapy. In this study, the pathological complete response (pCR) based on biopsy analysis was employed as a ground truth. As the size of the tumor may remain unchanged during the first sessions of chemotherapy, it would be complicated with the techniques currently used to detect the rate of response of a breast tumor to chemotherapy. Nevertheless, some researchers report that in individual chemotherapy, intra-tumoral developments could occur inside the breast tumor [20–22]. Hence, we investigate a new deep neural network using increased DCE-MRI images obtained previously for primary chemotherapy (baseline) based on the biopsy ground to obtain a model for classifying breast cancer response to neoadjuvant therapy utilizing parallelly two DCE-MRI scans. We also explain the obtained results with a visual heatmap to facilitate the introduction of this model in clinical routine.

7.2 Materials and developed methods

7.2.1 Study population

We used in this research a clinical cohort including 42 subjects having local breast tumors provided by the Jules Bordet¹ Institute in Brussels, Belgium. Based on pCR, 14 patients were responding to the chemotherapy and 28 others were not responding. All 3D DCE-Mr scans obtained pre and postchemotherapy for each patient following the same protocol were used in this study. To reduce unnecessary data, from the DCE-Mr volumes, only slices including the tumor were cropped. Consequently, using the axial plane, 763 images were obtained before chemotherapy, and the 763 corresponding slices were obtained following the initial chemotherapy. We divided the cohort into 25% of validation data and 75% of training data while respecting that there is no overlap between these two datasets.

As described in our previous paper [23], patients with an early morphological response (EMR) are designed by the shrinking of their tumor's maximum diameter (Dmax) after the initial session of therapy. Early morphological nonresponders (EMNo-Res) design patients with an augmentation of maximum diameter after the primary session. Table 7.1 presents the clinical features of the used dataset.

7.2.2 Magnetic resonance imaging protocol

All MRI scans were realized at different time points using a Siemens 1.5 T [24,25]. The DCE-MRI acquisition favored six T1-weighted volumes obtained in

¹<https://bordet.be/en/>.

Table 7.1 Clinical features of the used cohort.

Clinical features	Total	EMR ^a	EMNo-Res ^b	pCR	Non-pCR
Number of patients with response status	42	28	14	14	28
Mean age	55	58	52	55	51
Mean tumor size (in mm)	34.8	32.1	38.3	28	38.2
Patients with positive lymph node	24	20	4	6	17
Patients with negative lymph node	18	8	10	8	11
Patients with ER + (positive estrogen receptors)	19	11	8	1	18
Patients with triple negative	10	8	2	6	4
Patients with HER2 +	13	9	4	7	6

^aEarly morphological response.

^bEarly morphological nonresponders.

the coronal view before and after injecting 0.3 mmol/kg of gadodiamide (Omniscan) [26], with a flip angle of 25°, TR/TE of 9.0/4.39 Ms.

As previously shown in our published articles [8,9,16], the primary MRI exam was performed within 7 days before the first chemotherapy session. The second exam was performed between 24–72 hours after the initial chemotherapy session.

7.2.3 Image preprocessing

We applied three preprocessing treatments before starting the training and the validation processes. In this study, it was required to select the tumor volume by applying a manual crop of the tumor volumes obtained before and after the chemotherapy with the radiologist validation. This was aimed to reduce any various unnecessary voxels of the breast volume. To exclude any artifacts on DCE-Mr volume, a bias-correction algorithm was performed.

The next preprocessing was to apply an affine image registration to align tumor volumes received before and following the primary chemotherapy. By implementing this registration, the first slice of the tumor acquired before chemotherapy matches its corresponding one acquired back to chemotherapy. Subsequently, a region growing segmentation was applied to the tumors to select only voxels matching with the tumor. This preprocessing measure will be significant to assist the deep CNN in getting a couple of 2D slices (Fig. 7.1).

7.2.4 Convolution neural network architecture development

Fig. 7.2 presents the architecture of the deep multiinput CNN structure that we designed, which is formed of two VGG-like [27] parts taking two cropped

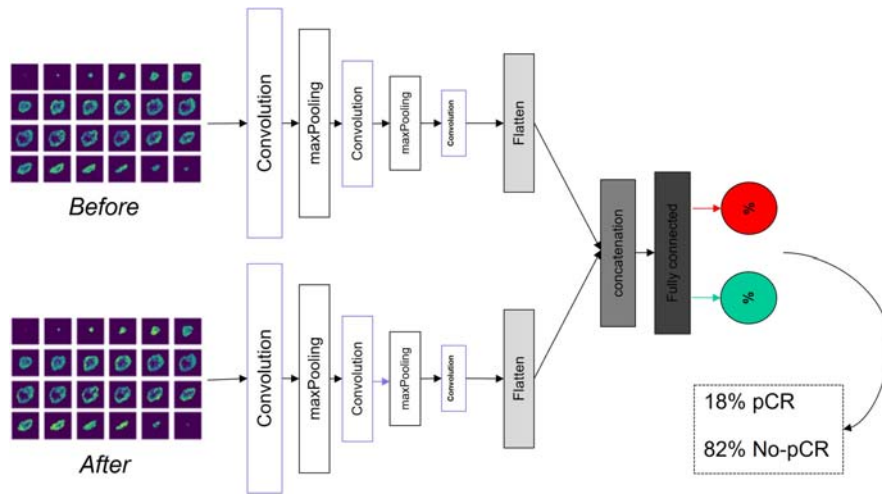


FIGURE 7.1

Proposed multiinput architecture to classifier pCR and no-pCR tumors.

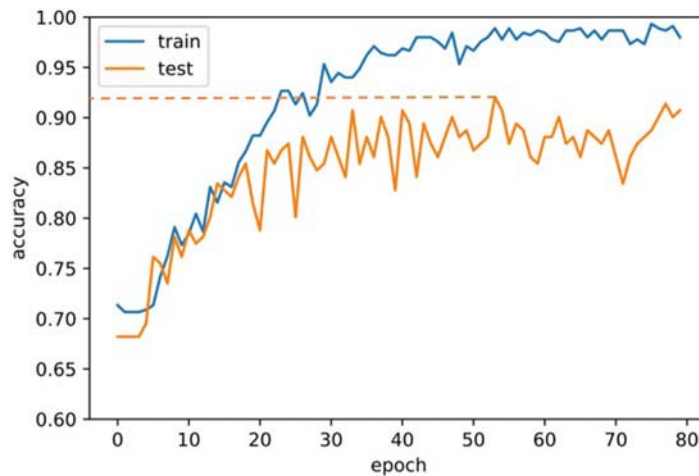


FIGURE 7.2

Accuracy of train and test process within 80 epochs.

volumes from axial DCE-Mr volumes. Each section holds 4 layers of convolution succeeded by a rectified linear unit (ReLU) activation function and Max-Pooling layer. To avoid any image down-sampling, the max-pooling layer was not used for the first block. 32 kernels were fixed for each layer of convolution inside the

second and the third blocks, and 64 kernels for the third and the fourth blocks. A Dropout operation was performed after every 2 layers of convolution, which help to generalize the trained model and bypass eventual overfitting effects. Then, the two branches after a fully connected layer of 512 hidden units were concatenated. At the final layer, the sigmoid function was applied to make the classification.

Glorot and Bengio's approach [28] were used for initializing the neural network weights. This approach was widely used to solve the vanishing or exploding gradient problems by performing a sophisticated initialization technique that ensures that all the layers learn at the same rate.

To perform the CNN training, data augmentation was implemented to the training cohort utilizing random linear methods of rotations, translations, horizontal flips, etc. for each training epoch, the data reproduction function was applied to create new samples. Therefore, more than 67500 unique samples were used to train the developed DL design.

The training was made by using Stochastic Gradient Descente (SGD) [29] inside 80 epochs. Based on fine-tuning, the used learning rate was fixed to 6×10^{-4} .

To compile the model, we applied categorical cross-entropy as loss function and standard accuracy metric based on the calculation of the mean accuracy rate across all predictions. Table 7.1 summarizes the employed and checked parameters.

7.3 Results

7.3.1 Quantitative results

Fig. 7.2 shows the accuracy during the 80 epochs. The greatest obtained accuracy is 92.72% using the test dataset. The area under the ROC curve (AUC) is 0.93. In our prior study [13], the AUC obtained by using the Parametric Response Map (PRM) was 0.89. Table 7.2 provides a comparison of accuracy and AUC calculated for different methods using the same dataset.

Table 7.2 Checked and used values of the training parameters.

Learning parameters	Checked values	Used value
Learning rate	0.5, 0.25, 0.1, 0.05, 0.0005	0.00005
Batch size	4, 10, 12, 16, 32	8
Adaptive learning rate	SGD, Adam, Adagrade.	SGD
Learning rate decay	Yes, No	Yes (1e-6)
Activation function	ReLU, Elu, Sigmoid	ReLU and sigmoid for the last layer
Dropout rate	0.25, 0.3, 0.5, 0.75	0.25 & 0.30 and 0.40 for the last layer

According to the radiologist, these findings are suitable for classifying the breast cancer response following the initial session of NAC. This could help radiologists to determine if a patient could stay receiving chemotherapy or not (Table 7.3).

7.3.2 Qualitative results

To reflect visually the most valuable features for classifying responsive and non-responsive patients, the Grad-Cam algorithm [30] was developed to be adapted to the proposed model. To evaluate the ability of the proposed model to classify pCR and no-pCR, we applied the Gradient class activation maps (grad-CAM) algorithm to no segmented tumors. As demonstrated in Fig. 7.3, for pCR patients, most tumor areas participating in the prediction refer to the internal tumor

Table 7.3 The obtained accuracy and AUC.

	Validation accuracy	AUC
Proposed model with DL	92.88%	0.93
PRM method without DL	–	0.89

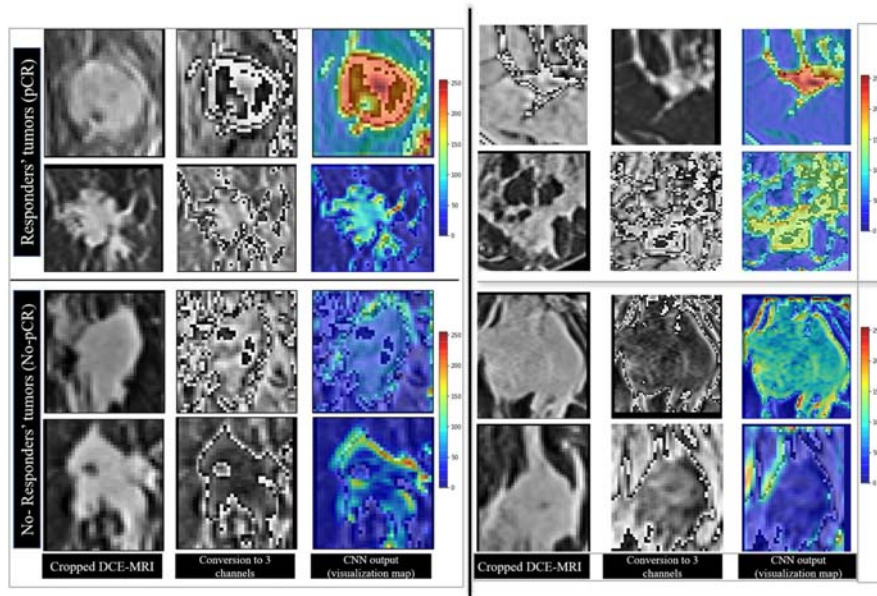


FIGURE 7.3

Randomly selected examples with grad-cam visualization.

volume. While, for nonresponsive patients, the neighboring zones and those positioned on the peripheral region of the tumor contribute more than the total tumor zone. These findings substantiate the critical position of the external tumor's regions to classify the response.

These results were interpreted by the collaborating radiologist. Indeed, this was explained by the fact that the most aggressive tumors try to protect themselves against the chemotherapy effect, by making their periphery and external areas stronger.

7.4 Discussion

In this chapter, we presented a new multiinput DL architecture for classifying and predicting pCR and no-pCR breast tumors.

The human could easily determine the rules for a given diagnosis and justify his decision. In the literature related to our research, it is not always explained how the decision was made by the DL model. To make this work more explainable and coherent, we added a qualitative validation of the results. This validation concerns the inspection of the degree of feature extraction by responsive and nonresponsive tumors. Indeed, it was necessary to explain and present information to radiologists in understandable and logical terms. Grad-CAM was employed to visually interpret the results of the developed deep CNN. When comparing Grad-CAM for pCR and non-pCR, we noted that CNN usually refines the internal tumor zone and highlights its neighboring zones. While for responsive patients, the deep CNN concentrates on all the internal tumor volume. Notwithstanding the great results shown in this chapter, the latter still has some weaknesses. Mostly, the size of the original cohort is still restricted. Indeed, the emphasis on response prediction of patients who underwent neoadjuvant therapy and the selection of pre and posttreatment MRI scans, as well as the ethical licenses administering the use of data limited the size of the cohort. Furthermore, the given approach does expect a sequence of data preprocessing, including the manual tumor volume selection and the affine registration.

7.5 Conclusion

We proposed in this chapter a new multiinput deep CNN model to classify the response to chemotherapy. Quantitative validations, including accuracy and AUC, were used to evaluate its performance. Using Grad-CAM to visualize tumor regions associated with the CNN classification is one of the most important contributions in this work. Indeed, this allows greater explicability than the approaches without a visualized decision-making part. The collected heatmaps could be feasibly utilized by radiologists to interpret CNN guidance concerning NAC.

In the future studies, we consider using the developed design for extra cohorts, which are still below annotation. We also plan to facilitate the tumor volume cropping and affine registration ideally over DL-based approaches. To implement a useful tool, we plan to integrate the developed DL model into a collective cloud web interface [31] to be used in a simple way by the radiologists, which will intend to produce a prospective validation process.

Aknowledgments

We extend our thanks to the radiology service of Jules Bordet in Brussels for helping us to use their dataset and evaluate the developed method. We would equally acknowledge Mr. Adriano Guttadauria for his support in managing the server machine of IT and the Artificial Intelligence unit used in this study.

References

- [1] Julie C. Dutoit, L.Verstraete Koenraad, Whole-body MRI, dynamic contrast-enhanced MRI, and diffusion-weighted imaging for the staging of multiple myeloma, *Skelet. Radiol.* 46 (6) (2017) 733–750.
- [2] SanaUllah Khan, Naveed Islam, Zahoor Jan, Ikram Ud Din, Joel J.P.C. Rodrigues, A novel deep learning-based framework for the detection and classification of breast cancer using transfer learning, *Pattern Recognit. Lett.* 125 (2019) 1–6.
- [3] Justin Ker, Lipo Wang, Jai Rao, Tchoyoson Lim, Deep learning applications in medical image analysis, *IEEE Access* 6 (2017) 9375–9389.
- [4] Sampo Kuutti, Richard Bowden, Yaochu Jin, Phil Barber, Saber Fallah, A survey of deep learning applications to autonomous vehicle control, *IEEE Trans. Intell. Transp. Syst.* (2020).
- [5] Antong Chen, Jennifer Saouaf, Bo Zhou, Randolph Crawford, Jianda Yuan, Junshui Ma, et al., A deep learning-facilitated radiomics solution for the prediction of lung lesion shrinkage in non-small cell lung cancer trials, *IEEE 17th Int. Symposium Biomed. Imaging (ISBI)* (2020) 678–682. IEEE, 2020.
- [6] Dooman Arefan, A.Mohamed Aly, A. Wendie, Margarita L. Berg, Zuley, Jules H. Sumkin, et al., Deep learning modeling using normal mammograms for predicting breast cancer risk, *Med. Phys.* 47 no. 1 (2020) 110–118.
- [7] G. Swapna, K.P. Soman, R. Vinayakumar, Automated detection of cardiac arrhythmia using deep learning techniques, *Procedia Comput. Sci.* 132 (2018) 1192–1201.
- [8] Adoui El, Stylianos Drisis Mohammed, Benjelloun Mohammed, Predict breast tumor response to chemotherapy using a 3D deep learning architecture applied to DCE-MRI data, *International Work-Conference on Bioinformatics and Biomedical Engineering*, Springer, Cham, 2019, pp. 33–40.
- [9] El Adoui, Mohamed Amine Larhman Mohammed, Drisis Stylianos, Benjelloun Mohammed, Deep Learning approach predicting breast tumor response to neoadjuvant treatment using DCE-MRI volumes acquired before and after chemotherapy, *Med. Imaging 2019: Comput. Diagnosis* 10950 (2019) 109502I. International Society for Optics and Photonics.

- [10] Francesco Schettini, Tomás Pascual, Benedetta Conte, Nuria Chic, Fara Brasó-Maristany, Patricia Galván, et al., HER2-enriched subtype and pathological complete response in HER2-positive breast cancer: a systematic review and *meta*-analysis, *Cancer Treat. Rev.* 84 (2020) 101965.
- [11] Richard Ha, Christine Chin, Jenika Karcich, Michael Z. Liu, Peter Chang, Simukayi Mutasa, et al., Prior to initiation of chemotherapy, can we predict breast tumor response? Deep learning convolutional neural networks approach using a breast MRI tumor dataset, *J. Digital Imaging* 32 (5) (2019) 693–701.
- [12] Na Eun, Daesung Lae, Eun Ju Kang, Jeong Seon Son, Ji. Hyun Park, Jeong-Ah Youk, et al., Texture analysis with 3.0-T MRI for association of response to neoadjuvant chemotherapy in breast cancer, *Radiology* 294 (1) (2020) 31–41.
- [13] Adoui El, Stylianos Drisis Mohammed, Benjelloun Mohammed, A PRM approach for early prediction of breast cancer response to chemotherapy based on registered MR images, *Int. J. Comput. Assist. Radiol. Surg.* 13 (8) (2018) 1233–1243.
- [14] Nariman Jahani, Eric Cohen, Meng-Kang Hsieh, P.Weinstein Susan, Pantalone Lauren, Hylton Nola, et al., prediction of treatment response to neoadjuvant chemotherapy for breast cancer via early changes in tumor heterogeneity captured by DCE-MRi registration, *Sci. Rep.* 9 (1) (2019) 1–12.
- [15] Richard Ha, Christine Chin, Jenika Karcich, Michael Z. Liu, Peter Chang, Simukayi Mutasa, et al., “Prior to initiation of chemotherapy, can we predict breast tumor response? Deep learning convolutional neural networks approach using a breast MRI tumor dataset, *J. Digital Imag.* 32 no. 5 (2019) 693–701.
- [16] Adoui El, Stylianos Drisis Mohammed, Benjelloun Mohammed, Multi-input deep learning architecture for predicting breast tumor response to chemotherapy using quantitative MR images, *Int. J. Comp. Assist. Radiol. Surg.* 15 (9) (2020) 1491–1500.
- [17] Michael Z. Liu, Simukayi Mutasa, Peter Chang, Maham Siddique, Sachin Jambawalikar, Ha Richard, A novel CNN algorithm for pathological complete response prediction using an I-SPY TRIAL breast MRI database, *Magnetic Reson. imaging* 73 (2020) 148–151.
- [18] Mohammed A. AL-MASNI, AL-ANTARI, A. Mugahed, Jeong-Min PARK, et al., Simultaneous detection and classification of breast masses in digital mammograms via a deep learning YOLO-based CAD system, *Comput. Methods Prog. Biomed.* 157 (2018) 85–94.
- [19] Joseph REDMON, Santosh DIVVALA, Ross GIRSHICK, et al., You only look once: Unified, real-time object detection, *Proc. IEEE Conf. Computer Vis. Pattern Recognit.* (2016) 779–788.
- [20] M. El Adoui, S. Drisis, M.A. Larhmam, M. Lemort, M. Benjelloun, Breast cancer heterogeneity analysis as index of response to treatment using MRI images: a review, *Imaging Med.* 9 (4) (2017) 109–119.
- [21] El Adoui, Mohammed, Stylianos Drisis, and Mohammed Benjelloun. Analyzing breast tumor heterogeneity to predict the response to chemotherapy using 3D Mr images registration. *Proceedings of the 2017 International Conference on Smart Digital Environment*, 56–63. 2017.
- [22] N. Cho, S.A. Im, I.A. Park, S.H. Lee, M. Li, W. Han, et al., Breast cancer: early prediction of response to neoadjuvant chemotherapy using parametric response maps for MR imaging, *Radiology* 272 (2) (2014) 385–396.

- [23] Stylianos Drisis, El. Adoui Mohammed, Flamen Patrick, Benjelloun Mohammed, Dewind Roland, Paesmans Mariane, et al., Early prediction of neoadjuvant treatment outcome in locally advanced breast cancer using parametric response mapping and radial heterogeneity from breast MRI, *J. Magn. Reson. Imaging* 51 (5) (2020) 1403–1411.
- [24] R.A. Little, H. Barjat, J.I. Hare, M. Jenner, Y. Watson, S. Cheung, et al., Evaluation of dynamic contrast-enhanced MRI biomarkers for stratified cancer medicine: how do permeability and perfusion vary between human tumours? *Magn. Reson. Imaging* 46 (2018) 98–105.
- [25] S. Drisis, T. Metens, M. Ignatiadis, K. Stathopoulos, S.-L. Chao, M. Lemort, Quantitative DCE-MRI for prediction of pathological complete response following neoadjuvant treatment for locally advanced breast cancer: the impact of breast cancer subtypes on the diagnostic accuracy, *Eur. Radiol.* 26 (5) (2016) 1474–1484.
- [26] W.S. Kerwin, X. Zhao, C. Yuan, T.S. Hatsukami, K.R. Maravilla, H.R. Underhill, et al., Contrast-enhanced MRI of carotid atherosclerosis: dependence on contrast agent, *J. Magn. Reson. Imaging* 30 (1) (2009) 35–40.
- [27] Hussam Qassim, Abhishek Verma, David Feinzimer, Compressed residual-VGG16 CNN model for big data places image recognition, 2018 IEEE 8th Annu. Comput. Commun. Workshop Conf. (CCWC) (2018) 169–175. IEEE.
- [28] Xavier et B.E.N.G.I.O. GLOROT, Yoshua, Understanding the difficulty of training deep feedforward neural networks, *Proc. Thirteen. Int. Conf. Artif. Intell. Stat.* (2010) 249–256.
- [29] Sixin Zhang, E.Choromanska Anna, Yann LeCun, Deep learning with elastic averaging SGD, *Adv. Neural Inf. Process. Syst.* (2015) 685–693.
- [30] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-cam: visual explanations from deep networks via gradient-based localization, *Proc. IEEE Int. Conf. Computer Vis.* (2017) 618–626.
- [31] Sidi Ahmed Mahmoudi, El. Adoui Mohammed, Amin Belarbi Mohammed, Amine Larhman Mohammed, Lecron Fabian, Cloud-based platform for computer vision applications, *Proc. 2017 Int. Conf. Smart Digital Environ.* (2017) 195–200.

Deep learning interpretability: measuring the relevance of clinical concepts in convolutional neural networks features

Mara Graziani, Vincent Andrearczyk and Henning Müller

*Institute of Information Systems, University of Applied Sciences Western Switzerland (HES-SO),
Sierre, Switzerland*

8.1 Introduction

A large variety of tasks is being solved by algorithmic systems implementing classical Machine learning (ML) and Deep learning (DL) techniques. DL, in particular, has emerged as a better-performing substitute for hand-crafted feature extraction [1], which is more traditional for ML in health applications. Not provided with any insights about the decision-making, end-users seem to report wobbly confidence in the DL decision process [2]. Some of the inherent risks that even a perfectly well performing DL model may hide are the codification of biases and the weak accountability of decision-making. The flawed system for pneumonia risk detection analyzed by Caruana et al. in [3] is an example. Despite its high performance, the model learned to assign a lower risk of death to cases of pneumonia with concurring asthma because of misleading correlations in the data. A correct diagnosis would have taken the opposite decision given the high risk of death with this preexisting condition. The misleading correlation (i.e. presence of asthma thus low risk of death from pneumonia) was rather a consequence of the effective care given to these patients by healthcare specialists that were promptly reacting to reduce the risk of death, consequently lowering the recorded risk for these patients. The misleading feature “presence of asthma” was captured because of model interpretability.

The perception of DL as a black-box that gives little insights about the final output is a limiting factor for the acceptance and consequent use of DL models by physicians [4]. The near-perfect accuracy of DL models may only be apparent for a few very specific tasks, dropping significantly in real-world practice [5]. This shows, as argued in [6], that the evaluation of DL models only on the basis of task performance is fundamentally incomplete. The need for interpretability in

the development of AI for health emerges as impellent for two main reasons. On the one hand, the interaction between physicians and AI is improved by interpretability methods. On the other hand, interpretability can be used as an alternative to the test performance to validate the model decision-making process. Interpretability can be used, for example, for pointing humans those subtle visual features in the image that make the diagnosis at the borderline between two choices, causing low inter-rater disagreement. In retinopathy [7] DL and physicians can interact to decipher borderline cases such as the detection of the plus disease in the retina of preborn babies. High performance in this task can make a considerable difference in saving babies from blindness. The performance of the combination of humans and DL were shown to be the highest in terms of the inter-rater agreement also for other application domains, for example cancer diagnosis [8]. Interpretability in the sense of explaining the rationale for AI decisions to its final users is therefore an important prerequisite for the application of AI to health-care, which will be further discussed in this chapter.

Most of the interpretable AI methods can be categorized according to a few factors, namely *global* versus *local* interpretability, *built-in* versus *posthoc* methods and *feature* versus *concept* attribution, described later in this chapter (in Section 8.2.2). The popular activation maps in [9], for example, generate explanations of the decision for a single input (i.e., local), without requiring to retrain the model parameters (i.e., posthoc), highlighting the most salient input pixels (i.e., feature attribution). Concept-based methods such as the Concept Activation Vectors (CAVs) proposed by Kim et al. [10] generate explanations in terms of arbitrary high-level concepts. This method is also posthoc and shows that clinically relevant features, that we refer to as *clinical concepts*, can be directly used to explain complex DL models. These explanations help the users of interpretable AI to think more systematically about the relevance of specific features within the AI model [11]. By including experiments on diabetic retinopathy, Kim et al. showed the applicability of CAVs to health-care, explaining DL decisions in terms of the presence or absence of a clinical feature. One limitation of CAVs is that they express a clinical feature only in terms of either its presence or absence, whereas continuous or categorical measures are more frequently used to describe clinical factors, for example, the size of a lesion. Regression Concept Vectors (RCVs) were proposed to extend CAVs to continuous and categorical clinical features [12]. Research on the applicability of RCVs shows that these explanations can fit the requirements of various medical tasks ranging from histopathology [12–14], to radiomics [15] and retinopathy [13,16]. By directly matching the semantics of the end-users, RCVs explain DL decisions in relation to well-known prognostic factors and clinical guidelines. This approach to “subject-centric” explanations (SCEs), as referred to in [17], shows promise for interactive explanations, learning about the model behavior from the outside.

The main focus of this chapter is the application of concept-based interpretability to measure the relevance of clinical concepts in DL decisions. Within the chapter, we clarify the terminology around AI interpretability, presenting an

in-depth analysis of the existing tools for health applications. The concept-attribution approach of RCVs is discussed for obtaining SCEs that relate to clinical concepts, fostering the interaction between physicians and DL models. The detection of plus disease in Retinopathy of Prematurity (ROP) cases is presented as the main application domain. The application to the ROP is relevant because the detection of plus disease is at the edge between two fundamentally different treatment planning strategies and causes large disagreement rates in the diagnoses. In Section 8.2, we review the literature concerning AI interpretability. Sections 8.3 and 8.4 present the methods and the experimental results, respectively. In Section 8.3.1, in particular, we introduce ROP and the task of plus disease detection. Section 8.5 presents insights and in-depth discussions on the analyses. The conclusions in Section 8.6 summarize the key points in this chapter and present a higher-level discussion on XAI research for computer-assisted diagnosis systems.

8.2 Related work on interpretable artificial intelligence

8.2.1 Motivations

The research field in AI interpretability has grown very quickly in the last four years (see Fig. 8.1 on the *left*). In the medical imaging domain, the number of publications per year concerning interpretable AI development also presents a marked increase (Fig. 8.1 on the *right*). The rising interest in interpreting DL models can be traced back to the evidence that the classic metrics of model performance (e.g., classification accuracy, loss) are not sufficient to describe the model's principles of inner functioning.

As Doshi Velez and Kim argue in [6], the need for interpretability stands out in problems that suffer from incompleteness in their formalization. Particularly in medical imaging, model performance in terms of the specificity and sensitivity of the predictions is evaluated for a precollected testing set for which experts have

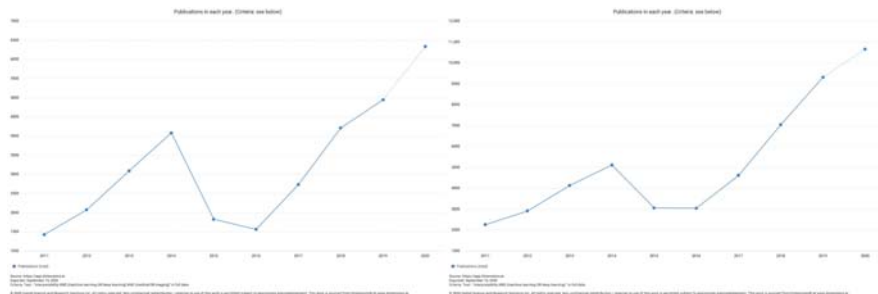


FIGURE 8.1

Trends of the research fields in interpretable AI (on the *left*) and interpretable AI for medical imaging (on the *right*). AI, Artificial intelligence.

agreed regarding a ground-truth diagnosis. The generalization to unseen data may not hold, causing significant drops in performance in real-world applications [5]. Depending on the application, the sole measurement of task performance may lead to incomplete model evaluations on various fronts. Some models may require fairness, for example, not encoding biases that would induce gender or racial discrimination in their decisions. Other models may require the robustness to adversarial attacks, for example biometrics and person identification. Model accountability (in the sense of taking responsibility for the decisions) may be another desideratum, for example in credit allowance or automated driving applications [18]. The motivation for interpretable AI development, therefore, directly stems from the application requirements. Interpretability in health-care applications aims at avoiding erroneous diagnosis since automatic predictions can be analyzed and interpreted before a final decision by an expert. In autonomous driving, interpretability mostly aims at demonstrating the causes for an accident (for insurance liability reasons, among others) once the mistake has already happened [18]. The EU's General Data Protection Regulation, in effect since May 2018, officialized the need for safety, fairness and explainability of AI deployment in the real world. The so-called "right to obtain an explanation" provides individuals with the right to inquire about the transparency, accountability and explainability of how their data were handled by the automated decisions. For example, if an automatic system was to deny a loan application, the denied person has the right to ask for an explanation regarding the decision in the form of "meaningful information about the logic of processing."

8.2.2 Related terminology

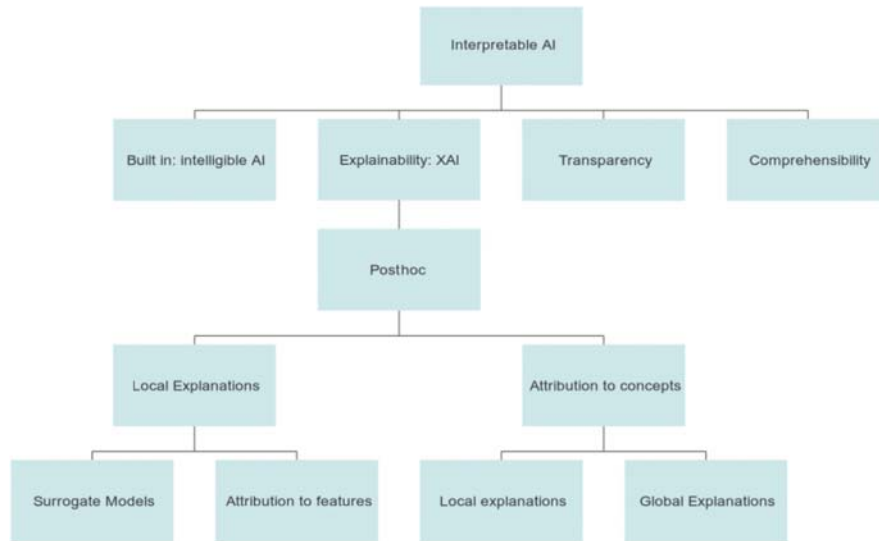
Differences in the specification of the interpretability objectives (e.g. for debugging, for explaining wrong decisions or as a way of proving model safety, fairness and accountability) inevitably lead to inconsistencies in the terminology related to interpretable AI. The words "interpretable," "explainable," "intelligible," "understandable," "transparent," and "comprehensible" have often been used interchangeably in the literature, causing confusion and different taxonomies [19–24]. A formal definition of *interpretability* was given by Doshi-Velez and Kim in [6] as that of "explaining or presenting in understandable terms to a human" the decisions of a ML system. The concept of "interpreting" is therefore inherently linked to that of "explaining" in this definition that we adopt. In the analysis of interpretability from the perspective of social sciences [19], Miller agrees with assigning the same meaning to explainable and interpretable, in the sense of "providing explanations" to humans. Interpretability as intended in [6], however, seems to correspond to what is meant as intelligible in the taxonomy presented in [20], namely the large set of possible actions and developments to obtain a system that is "clear enough to be understood" by humans. According to the scheme in [20], being interpretable or explainable also means being intelligible, but the opposite is not necessarily true. *Intelligible* AI, in practice, does not imply the generation

of explanations. This could be obtained, for example, by adding interpretability constraints on the model objective function being optimized [25], or by visualizing the network internal features [25]. The set of *explainable* models is thus a rather smaller subset than that of intelligible models, although interpretable and explainable are still being used to refer to the same purpose, which is “explaining” decisions to humans [6,19,20]. Transparency and comprehensibility also appear as terms related to interpretable AI [22]. The former generally refers to a set of descriptions that are most relevant to AI developers. *Transparency* is defined in [22] as the description of the structure, equations, parameter values and assumptions necessary to understand the inner model mechanisms. Lipton further divides this definition into model simulatability, decomposability of the parameters and algorithmic transparency [26]. Finally, the notion of model *comprehensibility* is described in [22] as the ability of the learning algorithm to generate meta-descriptions about its inner working mechanism that can be interpreted in natural language. The taxonomy relative to interpretable AI is nevertheless subject to continuous change and updates. Among these definitions proposed in the literature, two should be retained for this chapter, namely that of explainable AI (XAI) as a means of generating explanations for AI decisions and that of intelligible (or interpretable) AI as a wider set of tools also including methodologies that do not necessarily aim at generating explanations. In the context of XAI, Miller further clarifies the elusivity about the concept of providing explanations [19]. The sole association between model output and possible causes is not sufficient to provide a “good explanation”. These should be rather contextualized with the user’s needs, promoting the interaction and answering contrastive questions, that is, “why did the model output was class P instead of another class Q?” From these approaches to interpretability, the definitions of human-centric or Subject-Centric Explanations (SCEs) arose as a way of identifying XAI methods that are tailored to the user’s needs and the requirements of the application domain. SCEs aim at fostering the interaction between end-users and ML or DL models and are expressed in the ontology of the application domain.

8.2.3 Related work on explainable artificial intelligence

8.2.3.1 Explainable artificial intelligence for medical applications

Proving the safety and reliability of the model decision-making is an emerging challenge in the deployment of AI to the medical domain. Several techniques for interpretable AI development find a relevant application in the medical domain, providing interesting insights into various tasks [27–34]. A large part of the interpretable AI methods for medical applications generate visual explanations to provide justifications for the model predictions. Among these, saliency maps are the most frequently used to generate explanations in pathology [30–33], retinopathy [28], and radiology [27,34–36]. Numerous interpretable AI approaches can be categorized by the schema in Fig. 8.2 of which the elements are described in detail.

**FIGURE 8.2**

Categorization of interpretable artificial intelligence approaches.

This section reviews the main approaches and introduces the relevant technical terminology to categorize XAI methods. By focusing on explainability techniques, this section does not include inherently intelligible models (e.g. linear regression), models with built-in interpretability (e.g. a decision tree, also part of intelligible AI) and dataset exploration methods (e.g. dimensionality reduction techniques or the retrieval of influential instances [37]). We also exclude geometrical approaches such as Singular Vector Canonical Correlation Analysis [38].

Some of the technical terms used to distinguish most of the current approaches for obtaining interpretable AI were introduced by Lipton in [26]. In particular, Lipton distinguishes *local* versus *global* explanations and *built-in* versus *posthoc* methods. Local explanations refer to explanations that are only true for a single input. Global explanations, on the contrary, explain the model behavior for an entire set of inputs, for example all images of a single class in the dataset. Built-in methods, as explained by Lipton, introduce interpretability as one of the objectives of the model optimization function. These methods are included in the more general notion of intelligible AI. An example is that of inherently interpretable models, for example linear regression, where the linear increase of a feature value corresponds to a proportional increase in the model output. Posthoc methods are on the other hand methods that generate explanations without requiring the retraining of the model parameters with interpretability constraints. Finally, attribution methods generate explanations by identifying either the most relevant features, in feature attribution, or the most relevant concepts, in concept attribution, to the network decisions.

Referring back to Miller’s formalism of explainability [19], feature attribution methods answer to the question “What would the model output be if the *value of this input feature* was different?”. Concept attribution provides explanations in terms of high-level concepts that can match the semantics of the end-users. In the medical scenario, these can be directly clinically relevant concepts. Therefore, concept attribution answers the questions of the type “Why did the model output class P, and would it answer class Q if this *clinical attribute* was different?”. The clinical attribute may be a visual feature, for example, describing the size of a lesion. The shift between feature and concept attribution is mainly at the interpretation level. In some cases, for example in imaging applications, the values of individual features (e.g. the raw input pixels) appear rather incomprehensible to humans [10]. The aim of concept attribution, as further explained in Sections 8.2.3 and 8.3.2, is to generate explanations that can directly relate to the ontology of the receivers of the explanation.

The next sections present a review of several XAI methods. Despite being rather long, this review is not exhaustive of all the methods existing in the literature and does not include several interpretability approaches that do not generate explanations (such as intelligible models, transparent and comprehensible models). The review is organized as follows. In Section 8.2.3.2, we review the most common visualization and feature attribution methods. In Section 2.3.3 we introduce the related work to concept attribution. Finally, in Section 8.2.4 we review the evaluation of XAI methods in the literature.

8.2.3.2 Visualization methods and feature attribution

Visualization methods were proposed, at first, for interpreting the remarkable increase in performance given by the application of deep Convolutional Neural Networks (CNNs) to computer vision tasks. Visualizations were proposed to either visualize the learned features or to highlight the most salient input features. These methods evolved into the generation of explanations by feature attribution. The network output for a single input (local method) is explained by a subset of the input features. These input features identify, in the case of images, the most important pixels that are then displayed as a heatmap. In this section, we further clarify the literature related to visualization approaches and we report the formalization of feature attribution as proposed in [39].

A cornerstone of early interpretability development is the deconvolution paper by Zeiler and Fergus. Their approach is twofold. On the one hand, they visualize the filters learned by various CNNs (an idea already formulated in [40] by the Activation Maximization approach) through the inversion of the convolution operations, that is where the name “deconvolution” comes from. On the other hand, they generate saliency heatmaps by systematically occluding portions of the input image with a gray square [41]. Their occlusion sensitivity method monitors the output of the classifier to these input perturbations. Simonyan’s saliency maps further develop this idea by computing the relevance of individual pixels rather than entire input regions [42]. Each value at a location of Simonyan’s saliency

heatmap represents the derivative of the decision function with respect to the input pixel in that same location. Guided backpropagation upgrades this formulation, using the signals from the high layers as additional guidance to avoid the flow of negative gradients [43]. The Layerwise Relevance Propagation (LRP) in [44] further expands the idea of saliency in [42]. LRP decomposes the contribution of each pixel (also obtained by a derivative operation) at each layer in the CNN computation. The propagation of this relevance through the layers is then evaluated to obtain the LRP values that are visualized as a heatmap. An entirely different approach that also generates visual explanations, often called activation maps, is that of Class Activation Mapping (CAM) [9]. Individual CNN feature maps are used to obtain a heatmap of the network's attention before being spatially averaged and linearly combined to produce the network prediction. One limitation of CAM is that it can only be applied to CNNs with a global average pooling layer, rarely used in recent state-of-the-art architectures. Grad-CAM is proposed in [45] as a generalization of CAM that directly takes into account the cascade of gradients at each CNN layer. In this way, the activation maps can be obtained from a wider variety of CNN architectures, including those used for image captioning and query answering. When applied to classification tasks, CAM and Grad-CAM are equivalent up to a normalization constant that is proportional to the number of pixels in the feature maps [45]. The Grad-CAM framework does not generalize to multiple occurrences of same-class instances in the input image considering the gradients with respect to entire feature maps. This limitation is partially addressed by Grad-CAM++, which considers the gradients directly at the pixel level [46]. The prediction difference analysis method in [47], based on the framework in [48], proposes a probabilistic approach to generating heatmaps. The basic idea is that of estimating the relevance of a feature by measuring the change in the prediction when that specific feature is unknown. This change is obtained by evaluating the difference between the probability of the prediction when conditioning on the complete feature set and when conditioning on the feature set where that specific feature is removed [47]. The framework Local-Interpretable Model Agnostic Explanations (LIME) proposed by Ribeiro in [49], is a local posthoc XAI method that uses linear surrogate models to generate explanations for a single input image. To clarify eventual confusions, the adjective "Local" in LIME refers to the approximation of the DL decision function in the locality of an input sample. For instance, the linear surrogate model approximates the decisions of the DL model in a neighborhood of the input sample. Using LIME to explain CNNs is similar to using a sparse linear model to approximate the complex decision function of the CNN. The first step of the application of LIME to images consists of clustering pixels into superpixels (that are used as features) using color, texture and other types of local similarities. Randomly hiding some of the superpixels generates perturbations (called samples) of the original images which can be used to compute the relevance of each superpixel to the decision-making. Some common algorithms to extract superpixels are Simple Linear Iterative Clustering [50] and Felzenszwalb's graph-based image

segmentation [51]. DeepLIFT is proposed in [52] to generate attribution scores on the basis of the difference between the neuron activations and a “reference activation” that is computed, for example, by using a blurred version of the original input image. Finally, SHapley Additive exPlanations [53] are proposed as a framework that unifies some of the method formulations for explaining predictions, including LRP, LIME and DeepLIFT.

All of the methods presented in this section generate posthoc local explanations that attribute the CNN decisions to a set of input features. This approach is summarized by the formal definition by Sundararajan et al. of the framework of attribution to features [39], reported in the following.

Given an input image $X = (x_1, \dots, x_n) \in \mathbb{R}^n$ and a CNN with a decision function f mapping the input image to a class probability, the attribution vector is defined as:

$$A_f(X, X') = (a_1, \dots, a_n) \in \mathbb{R}^n$$

where each a_i explains the contribution of each pixel x_i to $f(X)$, $i = 1, \dots, n$

8.2.3.3 Concept attribution

Concept attribution aims at addressing a key difficulty in the generation of pixel-based explanations for CNNs, namely that humans understand high-level concepts more easily than the raw input pixel values or the internal CNN activation values [10]. This section reviews the related work on interpreting CNNs by using human-friendly concepts and presents the framework of attribution to concepts, as opposed to that of attribution to features.

The reference paper for generating explanations in terms of high-level concepts is testing with CAVs [10]. The interpretation of a ML or DL model, in the sense of generating posthoc explanations, is seen as a translation problem. The state of the model is defined by Kim et al. as a vector space E_m (e.g. the space of the CNN activations) [10]. The basis vectors in this space correspond to the input features and neural activations. Another vector space E_h is used to describe the space of high-level concepts and interactions understandable to humans, with bases vectors corresponding to the high-level concepts. Generating explanations means finding a function $g: E_m \rightarrow E_h$. The method in [10] proposes a way of obtaining the translation g . Given a concept of interest, they collect a set of example images representative of the concept. The CAV is then learned in the space of the activations of a CNN layer as a linear classification task that separates the set of examples with the concept from a set of random images (not containing the concept). The CAV for that concept, for instance, is the unit vector representing the linear classifier. In other words, the CAV models the presence or absence of a human-friendly concept and it is computed as the unit weight vector representing the linear classifier that separates images with the concept from those without the concept in the space of activations of a CNN layer. The performance of the linear classifier is indicative of how well the concept is learned in the network representation. The use of linear classifiers, that are inherently intelligible, is also addressed

as linear probing in [54]. The work on CAVs presents numerous extensions in the literature. The automatic extraction of visual concepts is proposed in [55] to obtain insights on the concepts learned automatically by CNNs despite not having explicit knowledge of all of them. Causal Concept Effect aims at establishing the causal effect of the presence of a concept in the input image [56]. The latest concept bottleneck models propose the training of DL models on images with annotations for both the ground-truth labels and the presence of concepts [57].

The RCVs in [12,13] extend CAVs to model not only the presence or absence of a concept, but also continuous-valued measures. These measures do not need necessarily additional annotations, as they can be directly computed on the images. The development of RCVs is particularly relevant to the medical domain since clinical concepts are expressed as observed measurements that do not fit in the binary formulation of CAVs, for example radiomic features [15], nuclei pleomorphism [12,13], vessel features of the retina [13,16].

Finally, concept attribution is defined in [13] for a set of Q concepts $\{c_i\}_{i=1}^Q$ as in the following:

A vector \mathbf{V}_{c_i} , being either the CAV or the RCV, represents a concept c_i in the activation space of a CNN layer l . The concept attribution vector $A_f(\Phi^l(\mathbf{X}), \{\mathbf{V}_{c_i}\}_{i=1}^Q) = (a_1, \dots, a_Q)$ represents with each a_i the relevance of the concept c_i to the CNN decision function $f(\mathbf{X})$ for an input \mathbf{X} .

Explanations obtained with concept attribution are defined locally around the input image \mathbf{X} . Being independent of the pixel locations the attribution values a_i can be agglomerated in multiple ways to obtain global explanations (valid for an entire class or an entire set of inputs). Some methods such as the TCAV and the Br scores are proposed within the works on CAVs and RCVs themselves. For this reason, the box of attribution to concepts in Fig. 8.2 leads to both local and global explanations.

8.2.4 Evaluation of explainable artificial intelligence methods

In the previous sections, we presented XAI methods that generate explanations in terms of visualizations or high-level concepts. Both approaches provide immediate feedback on the network internal state and can give insights on the criteria for decision-making. We discuss in this section the need for quantitative evaluation methods for XAI, motivated by the risk of confirmation bias if only a qualitative assessment is performed to evaluate XAI plausibility [58]. We then present a review of the studies proposing quantitative evaluations, arguing that some of the evaluation approaches do not generalize to all medical tasks. These considerations are relevant for both evaluating the existing XAI methods and developing new ones that can better suit the clinical needs.

XAI methods should highlight the relevant information behind the model's decision-making, while showing properties of robustness, implementation invariance, consistency, appropriateness and reliability [39]. Without these properties XAI methods would lose the user's trust as a meaningful way of assessing DL

decisions. Testing the reliability and trustworthiness of visual explanations only by visual inspection, however, is subject to the risk of confirmation bias. Confirmation bias is defined in cognitive psychology as the human tendency to attribute greater confidence to a hypothesis, even if false, when explanations are generated for it [lomb]. From a technical perspective, the quantitative evaluation of XAI methods is complicated because of the lack of ground-truth. We do not know, in fact, what input features are important to a model. Remarkable work in the literature focuses on developing evaluation methods, particularly on evaluating the consistency of saliency maps. By the term consistency, we refer to a series of desired invariances and dependencies that XAI outcomes should present. This includes, for example, implementation and input invariance and the dependency on the model parameters. Implementation and input invariance are addressed by the works in [39,59]. The same explanations, according to Sundararajan [39], should be generated for functionally equivalent networks, namely models with different architectures but reporting the same outputs for the same inputs. XAI outcomes should be invariant to constant shifts in the input data, although some of them show sensitivity are easily fooled by simple changes in the background color [59]. The dependency on the model parameters is evaluated in [60] by a series of randomization tests. The similarity of the explanations is compared when the learned model parameters are reinitialized to random values layer by layer in a cascading way, and completely, namely by resetting all the parameters to random values. The outcomes of the randomization tests show that XAI methods are inconsistent and perhaps assign the wrong attribution values to the wrong features [bim]. These results further stress the need for a solid evaluation of XAI outcomes that goes beyond simple visual inspection.

Concerning XAI development for medical applications, the deployment to the clinical setting further expands the desiderata for these methods. In the first place, explanations should be targeted at helping physicians with decision-making, without requiring extra expertise in the theoretical aspects of AI systems [59]. This goes in favor of the human-centric or SCE approaches that consider the user's need in the development of interpretable AI. In addition to this, Tokenaboni expands the list of desirable properties for XAI methods for clinical application. The explanations should be evaluated according to a series of factors. The first of these factors is the appropriateness of the explanation to the clinical domain. Clinically irrelevant, inconsistent and unnecessary explanations do not support physicians and should be given a lower priority. Explanations that cannot be translated into action (may this mean asking for additional analyses, for the confirmation or the modification of preexistent choices) should also be avoided, as they do not help with the clinical workflow. Finally, Tokenaboni adds the invariance to shifts in the XAI implementation parameters as a further evaluation of the consistency. Within the medical context, a few works assess the trustworthiness and reliability of XAI visualizations [33,60]. By using lesion contours annotations Arun et al. assess four points of saliency methods, namely their utility for localization tasks, their sensitivity to the randomization of the parameter weights, their

repeatability and reproducibility [60]. The instability of XAI visualization methods applied to emerges from this study on chest X-rays. It is important to point out that evaluating XAI methods on the basis of their localization performance as in [60], however, may not generalize to all clinical tasks. The work in [33], for example, discusses how this approach would easily fail in the context of histopathology images. This is mainly due to the fact that in these images there is not a clear central subject on the foreground but rather a structural disposition of many instances (e.g. connective, adipose, or epithelium cells) at several scales. It could be sufficient for the CNN to focus on one or a few instances, thus causing low evaluations of the localization capability of the CNN.

The existing studies on the evaluation of XAI methods show, finally, that the rigorous evaluation of these studies still necessitates sustained research that keeps into account the application domain [59].

8.3 Methods

8.3.1 Retinopathy of prematurity

8.3.1.1 *Relevant background*

We present in this section the applicative domain of the works presented in this chapter, namely the classification of plus disease in ROP. ROP affects premature babies born before 31 weeks of gestation and weighing less than 1.3 kg. This disease of the eye causes the abnormal growth of the blood vessels in the retina to more than 14,000 premature infants per year only in the U.S. If prompt action is not taken, the aggressivity of ROP may remain stable or advance further. Medical treatment is required by around 10% of the babies to avoid the degeneration of ROP. To analyze ROP an indirect ophthalmoscope is used to visually inspect the retina. With the digitalization of medical images, special cameras are used to take high-resolution pictures of the retina. These pictures are analyzed by multiple experts and can be used to track disease evolution over time, as shown in Fig. 8.3. The ROP diagnosis consists of identifying the affected zones of the retina, staging the disease on a scale from 0 to 5 (in Fig. 8.4). The risk if

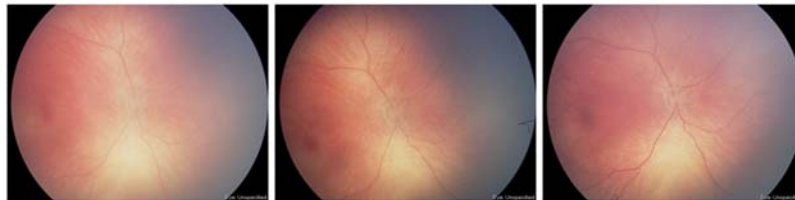


FIGURE 8.3

ROP progression from grade 0 to 2 in the right eye of the same patient. No presence of plus is noticeable.

ROP advances is that of blindness due to the total detachment of the retina (as in the picture on the right in Fig. 8.4).

The plus disease is a condition that may cooccur to ROP, illustrated in Fig. 8.4. In retinas affected by plus, the blood vessels appear enlarged and twisted and preannounce the worsening of the disease. The prompt detection of this condition is necessary for preventing the exacerbation of ROP and retinal detachment. Preplus indicates an intermediate stage where the severity of the eventual vascular abnormalities is not yet sufficient to define the presence of plus, but it is remarkable enough to plan earlier intervention. The presence of preplus or plus is assessed on the basis of the coexistence of clinical factors such as increased venous dilation and arterial tortuosity. The distinction between the two diseases is very subtle, and it is often a reason for strong disagreement among experts (Fig. 8.5).

8.3.1.2 Dataset for the experiments

The dataset for the experiments consists of 4800 de-identified posterior retinal images from a private dataset. The images were obtained by a commercially available camera, namely RetCam by Natus Medical Incorporated (in Pleasanton, CA). A total of 3024 images was used for training the network, consisting of 1084 images without plus, 1074 images with signs of Preplus and 1080 images containing plus. The testing set comprises 100 images, including 54 normal, 31 preplus disease, and 15 plus disease images. The assignment of the images to

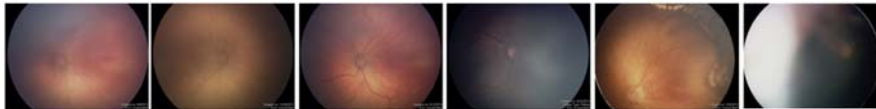


FIGURE 8.4

Example of ROP grades from 0 to 5 in multiple patients. plus is not noticeable in any of the images, except from the presence of preplus in the image for grade 4.



FIGURE 8.5

ROP progression on the same patient eye from the absence of plus disease, to preplus and presence of plus. The ROP grade is 2 in the first 2 pictures on the left and shifts to 3 in the picture on the right. ROP, Retinopathy of prematurity.

each category is obtained by the majority voting from three expert assessments. The high class imbalance between plus and normal cases is a consequence of the low prevalence of the ROP disease (only 3%).

8.3.1.3 Task and classification model

The task is the automatic ternary classification of images into normal, preplus and plus classes.

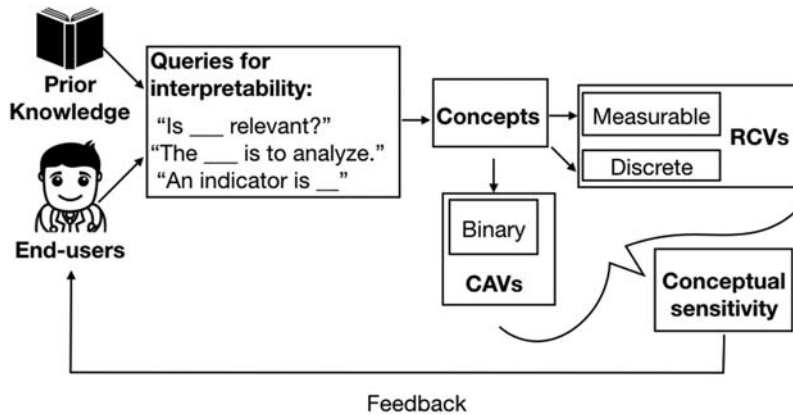
We reproduce the preprocessing pipeline from Brown et al. [61]. The retinal vasculature is segmented by a U-Net [62]. The network assigns, for instance, a probability to each pixel for being part of a vessel. This segmentation step bypasses the domain shifts due to variations in terms of pigmentation, illumination, and non-vascular pathology. A resizing operation is performed to uniformize the image size to 224×224 pixels, the input size of the CNN used for the classification. An Inception-V1 network [63], pretrained on ImageNet, is then finetuned on the ROP dataset to classify the images as normal, preplus or plus. We train the CNN with stochastic gradient descent and a categorical cross-entropy loss for 100 epochs. The learning rate is maintained constant to $1e - 4$. Data augmentation is applied with right-angle rotations and horizontal and vertical flipping. The hyperparameters are tuned by five-fold cross-validation as in Brown et al. [61].

8.3.2 Concept attribution with regression concept vectors

8.3.2.1 Identification of the concepts

The starting point of attribution to concepts is the identification of clinical concepts that should be used for generating explanations. In this section we will describe the workflow for the selection of clinical concepts represented in Fig. 8.6, starting from the collection of information to the final definition of a list of concepts and how to measure them. In the framework defined in [13], we identify mainly two sources of information that can drive the selection of clinical concepts, namely the prior knowledge on the domain and the consultation with domain experts. Fig. 8.6 shows these two as the starting point of the workflow for the selection of clinical concepts. Although for the experiments in this chapter we directly interacted with ophthalmologists to define the clinical concepts, we clarify in this section both approaches for completeness. We present the approaches from a high-level perspective so that they could be used also in other applications. The details on the interaction process and the type of questions that led to the identification of the concepts are reported in Section 8.4.2.1.

We describe in the following some of the sources of prior knowledge (represented as the starting point of the workflow in the box of the top left of Fig. 8.6) that can lead to the identification of clinical concepts. Prior knowledge can be represented in multiple ways, for example, collections of existing guidelines, as reports of previous studies or as annotated data. We discuss each of these in

**FIGURE 8.6**

Workflow for the selection of clinical concepts.

Reproduced from M. Graziani et al., *Concept attribution: explaining CNN decisions to physicians*, *Comput.*

Biol. Med. 123 (2020) 103865.

detail. Existing guidelines that are followed for human decision-making constitute an important resource, being based on several years of studies and joint efforts toward identifying decisive factors. Some examples of these are the well-established Nottingham grading (NGH) for breast cancer or the Gleason score in prostate cancer grading. The guidelines specify a list of factors that should be assessed by the pathologists to determine the tumor grade. For example, abnormalities in the appearance of nuclei and cells is one of the criteria in the NGH. In addition to the guidelines, the combination of handcrafted visual features and ML has been studied for several years before transitioning to DL. The handcrafted feature extraction driven by expert knowledge in the domain drives, in some cases, the extraction of powerful features with prognostic relevance [64–66]. Written reports, besides, justify the decision-making by describing the image content and the main causes that led to the diagnosis. Information to identify clinical concepts can be collected from all of these sources, namely the grading guidelines, handcrafted features and written reports. The selection of the clinical concepts performed in this way is particularly useful to verify that domain-knowledge is reflected in the layer activations of the network.

As the box on the bottom left of Fig. 8.6 suggests, the end users of the DL algorithm, in our case the physicians (ophthalmologists), can contribute to the selection of clinical concepts. As [13] suggests, this is rather an interactive process where the list of concepts is refined over multiple iterations until the explanations satisfy the users' inquiries on the model's decision-making. The direct interaction is useful to understand the expectations of the physicians on the DL decision-making. Physicians may be interested in validating that the model

decisions are in line with the guidelines of clinical practice as supposed in [12]. A question of interest, in this case, could be “Is the nuclei shape relevant to the automatic classification as tumor?” Confounding factors can be specified to make sure that irrelevant features are not used to make the classification, for example, by stating “changes in color appearance do not influence the classification.”

The next step in the workflow is to understand whether the factors identified by prior knowledge and end-users can translate to questions about the model decision-making. The question “is nuclei size relevant to the classification?” for example could be posed to translate the attention to the nuclei size and shape in the NGH into a relevant question about the automated decision-making. In this case, “nuclei size” is identifiable as a concept for the analysis. For an additional example, let us suppose that from the interaction with experts it emerged that they suspect that the watermarks at the bottom of the images may influence the model’s attention. The sentence “the influence of the presence of watermarks is to investigate” therefore translate into a relevant question that should be answered by concept attribution: “is the presence of watermarks a relevant factor to the decision?” The concept “presence of watermarks” is therefore added to the list of potential concepts. Note that this is a confounding factor with no clinical relevance and the outcome of the concept attribution analysis should show that this is not a relevant concept. If otherwise, this may highlight a bias in the decision-making requiring further analyses of the model and, if needed, the retraining of the parameters.

It is important to notice at this point that the concepts do not necessarily need to be specified in terms of the input features or the training data. Additional concepts can be defined using new data with annotations or from the metadata. Some concepts can be specific to the type of data being analyzed, as undefined for some data types. RGB color measures, for instance, are undefined for single-channel image modalities, for example computed tomography scans. Besides, to generate the explanations the list of concepts does not need to be perfect, and it will not likely be exhaustive of all possible concepts.

8.3.2.2 Computing the regression concept vector

In this section, we formalize the computation of RCVs as described in [13].

The output of the CNN internal layer is used to find the RCV for that layer. This procedure is posthoc and does not require the training of the parameters. The space of the activations of layer l , $\Phi^l(\mathbf{X})$, is considered. We extract $\Phi^l(\mathbf{X})$ for $\mathbf{X} \in X$ where X is the training dataset, a subset of it, or an additional dataset describing the concepts. For each image $\mathbf{X} \in X$ we have access to, or we can compute, a value of the clinical concept for which we are seeking the RCV. We represent this operation of accessing or evaluating the value of the clinical concept by $c(\mathbf{X})$. Given one image representing tumor cells in a tissue slide, the average number of pixels in the segmentation of the nuclei regions can represent a value for

the concept “nuclei size”. We seek the linear regression that can model the value of the concept $c(\mathbf{X})$ for each $\mathbf{X} \in X$ as in the following:

$$c(\mathbf{X}) = \mathbf{V}_c \cdot \Phi^l(\mathbf{X}) + \text{error}$$

The RCV for the concept c is \mathbf{V}_c . The RCV components can be found by applying linear least squares (LLS) estimation to X_{concepts} . Fig. 8.7 illustrates the approach for a 2-dimensional space.

If l is a dense layer of width p , \mathbf{V}_c is a p -dimensional vector in the space of its activations. If l is a convolutional layer the output of $\Phi^l(\mathbf{X})$ has spatial and channel dimensions (height, width, channels) represented as $w \times h \times p$. The simplest way of solving LLS in this space is to flatten $\Phi^l(\mathbf{X})$ to a one-dimensional array of whp elements as in [12,54]. This operation is widely discussed in [13], where better approaches are also proposed. Unrolling the convolutional maps may cause the explosion of the dimensionality of whp . The flattening operation, besides, breaks the natural 2D structure of the representation of convolutional feature maps, assigning neighboring features to independent dimensions. A spatial aggregation, i.e. global pooling, along the (height, width) of each feature map is a solution to this shortcoming, generating a representation of $\Phi^l(\mathbf{X})$ as a one-dimensional array of p elements. This solution, only briefly mentioned in [54] and tested in [13] improves the quality of the regression fit. A further solution proposed in [13] is adding a regularization term to the optimization:

$$\mathbf{V}_c^{\text{ridge}} = \arg \min_{\mathbf{V}_c} (\|c(\mathbf{X}) - \mathbf{V}_c \Phi^l(\mathbf{X})\|_2^2 + \lambda \|\mathbf{V}_c\|_2^2)$$

As opposed to CAVs, RCVs allow expressing the influence of the concept in terms of increasing values rather than its sole presence. For this reason, they are more suited to medical applications, which often consider continuous measures.

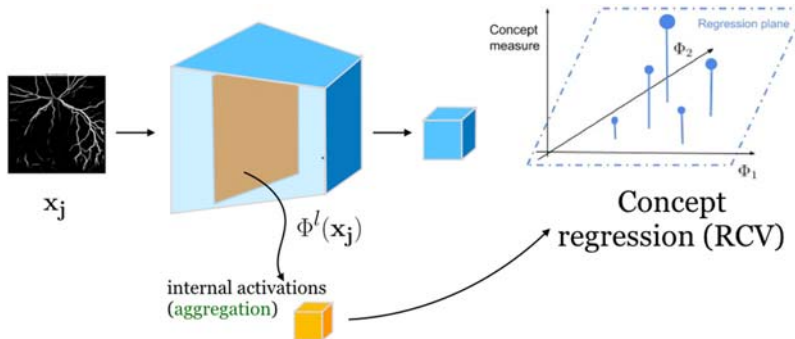


FIGURE 8.7

In this two-dimensional example, the RCV is the direction represented by the regression plane. In higher dimensions, unwanted pixel dependencies are removed by an aggregating operation of the internal layer representation.

The RCV represents the direction of the strongest increase of the concept measures for the concept c and it is normalized to obtain a unit vector \mathbf{V}_c .

8.3.2.3 Generating local explanations by conceptual sensitivity

In this section, we summarize how local explanations can be generated for a single input by a derivative operation [13].

The conceptual sensitivity proposed in [10] constitutes the way of generating explanations for a single input image x in terms of a concept c , and it represents the impact of changes in the concept value $c(x)$ to the network output. It is defined for CAVs, hence for binary concepts, in [10]. The same formula can be applied to RCVs for categorical and continuous concepts. In the following paragraphs, we report the definition of conceptual sensitivity for binary and multitask classification.

For a binary classification task, the conceptual sensitivity $S_c^1(\mathbf{X}) \in \mathbb{R}$ is defined as the directional derivative of the network output $f(\mathbf{X})$ over the CAV or the RCV direction \mathbf{V}_c , computed as a scalar product:

$$S_c^1(X) = \mathbf{V}_c \cdot \frac{\partial f(X)}{\partial \Phi^l(X)}$$

$S_c^1(X)$ represents the network responsiveness to changes in the input along the direction of the increasing values of the concept measures. The sign of $S_c^1(X)$ represents the direction of change, while its magnitude represents the rate of change. When moving along the RCV direction, the output $f(\mathbf{X})$ may either increase (positive conceptual sensitivity), decrease (negative conceptual sensitivity) or remain unchanged (conceptual sensitivity equals zero). In a binary classification network with a single neuron in the decision layer, the decision function is a logistic regression over the activations of the penultimate layer. A positive value of the sensitivity to a concept can be interpreted as an increase of $p(y=1|\mathbf{X})$ when the representation $\Phi^l(X)$ is moved toward the direction of the increasing values of the concept. Negative conceptual sensitivity can be interpreted as an increase in $p(y=0|\mathbf{X})$ when the same shift in the representation is applied. Conceptual sensitivities scores are informative about the concept influence on the decision for the single input image.

The derivation of the scores for multiclass classification tasks is straightforward.

Given the class label k , we consider the corresponding k th neuron in layer L . The neuron activation before softmax, $\Phi^{L,k}(\mathbf{X})$, is a vector of real numbers representing the raw prediction values. These values are then squashed by the softmax into a probability distribution, namely the probability of the label k to be assigned to the input data point \mathbf{X} . The conceptual sensitivity score for class k is computed as:

$$S_c^{l,k}(X) = \mathbf{V}_c \cdot \frac{\partial \Phi^{L,k}(X)}{\partial \Phi^l(X)}$$

The sensitivity scores can be computed for each class k , thus obtaining a vector of K elements. Large absolute values of the conceptual sensitivity for a single

class correspond to a strong impact in the decision function when the activations are shifted along the direction of the RCV. The derivative of the decision function can be obtained by stopping gradient backpropagation at the l th layer of the network.

8.3.2.4 Agglomerating scores for global explanations

In this section, we report two ways in the literature of agglomerating the concept sensitivity scores to obtain global explanations of model behavior for an entire set of input data, for example for a full class. These two ways are, for instance, the TCAV score proposed in [10] and the Br score in [13]. These ways of agglomerating score take into consideration different aspects and can be seen as complementary. Alternative scores can explore additional characteristics of the conceptual sensitivity, for example the ratio between positive and negative sensitivities or the largest variation. The UBS score in [15] proposes a layer-agnostic score that allows the comparison of the concept sensitivities across all layers in a CNN. Note that one score is computed for each concept analyzed. If we were to consider three concepts in our analysis, for example, three TCAV scores would be computed, namely one agglomerating all the conceptual sensitivity values obtained for the first concept, one agglomerating all of those for the second and finally one agglomerating the values for the third.

The TCAV score is defined in [10] as the fraction of k -class inputs for which the activation vector of layer l is positively influenced by the concept c

$$TCAV = \frac{|\{\mathbf{X} \in X_k : S_c^{l,k}(\mathbf{X}) > 0\}|}{|X_k|}$$

where $X_k \subset X_{task}$ is the set of inputs with label k . The TCAV score is bounded between zero and one. If no images are influencing the decision with a positive gradient, TCAV is zero. In the original paper, TCAV is only defined for CAVs [10], but its application is the same for RCVs.

Bidirectional relevance (Br) scores are proposed for medical tasks with two class labels symbolizing either the presence or the absence of a condition, for example tumor in [13]:

$$Br = R^2 \times \left(\frac{\bar{\mu}}{\hat{\sigma}} \right)$$

The coefficient of determination $R^2 \leq 1$ measures if the concept vector is representative of the concept. The coefficient of variation $\hat{\sigma}/\bar{\mu}$ is the standard deviation of the conceptual sensitivities over their average. This score is large when the RCV models correctly the concept values, i.e. R^2 is 1, and when the conceptual sensitivity values are consistent for all input samples, lying closely around their sample mean. Br explodes to infinite if $\hat{\sigma} = 0$. A normalization per layer is applied to scores for multiple concepts such that the highest magnitude is equal to 1. This scaling permits the comparison of the scores among concepts.

8.4 Experiments and results

8.4.1 Network performance on the retinopathy of prematurity task

We report in this section the results of the DL classification experiment on the ROP dataset.

The first step before classification is the vessel segmentation by a U-Net model, for which we report the output of some segmentations, compared to the raw images in Fig. 8.8.

The mean area under the ROC curve was computed on the validation sets, across five cross-validation splits. We obtain 0.94 (standard deviation 0.01) for the diagnosis of normal (i.e. binary classification normal versus preplus/plus) and 0.98 (0.01) for the diagnosis of plus disease (i.e. binary classification plus versus normal/preplus). The classification on the test set of the best model (based on cross-validation) achieves 91% accuracy on the 100 test images, sensitivity of 93% and specificity of 94%. The model loss and accuracy over training of one split of the cross-validation are reported in Fig. 8.9.

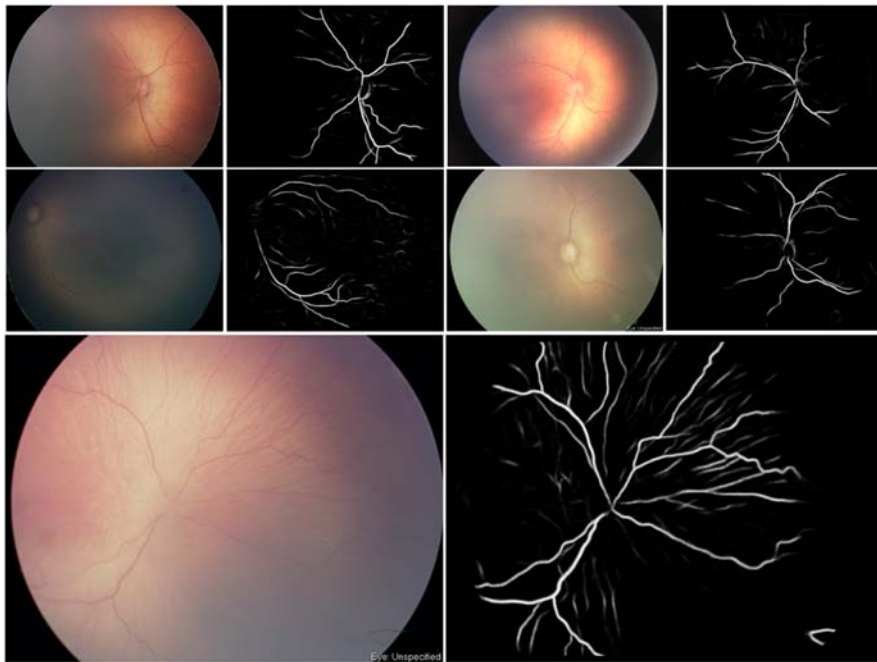


FIGURE 8.8

Raw input images and outputs of the vessel segmentation from the U-Net model.

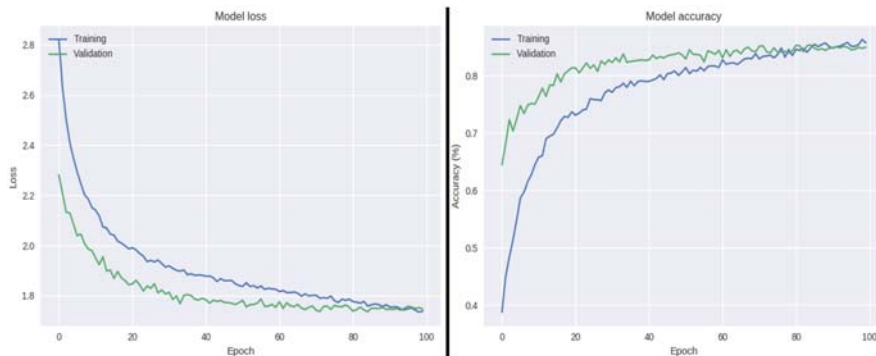


FIGURE 8.9

Model loss (on the left) and accuracy (on the right) at each training epoch on the training and validation sets.

Image credits: James M. Brown.

8.4.2 Results of concept attribution

8.4.2.1 Identification of the concepts

In the following sections, we describe the explainability results obtained on the ROP classification task using the methods described in [Section 8.3.2](#). In this section, in particular, we report in detail the process that led to the identification of the clinical concepts used for the research described in [\[16\]](#). As mentioned in [Section 8.3.2.1](#), the definition of the concepts was mostly driven by the interaction with the physicians (ophthalmologists).

This interaction focused on clarifying the visual factors that are taken into consideration to diagnose plus disease in ROP. This was made by asking them to sort the images in terms of the degree of aggressiveness of ROP and in particular of the plus disease. In [Fig. 8.10](#), we show some images for the three classes, namely normal, preplus and plus. Insights about these images were discussed with the physicians. To explain the visual differences in the images, they sketched out one of the patterns showing the exacerbation of the disease in [Fig. 8.11](#), namely the tortuosity of the retinal vessels.

Vessel tortuosity is also present in the literature of ROP as an important pattern for the detection of plus. Mathematical models, for example, describe the vessel appearance in terms of the Cumulative Tortuosity Index (CTI) [\[67,68\]](#). The aim of the feature modeling in the literature is to describe the appearance of the vessels in the whole retinal sample by computing standard statistics of the mathematical model representing the pervessel features. These features can then be used with standard ML algorithms for automatic ROP diagnosis. The features are divided into three groups according to their computation: point-based, segment-based or tree-based. An example of a point-based feature is the average point diameter. It describes the width of the vessels, in the direction normal to the blood flow, at each location in the image. Segment-based features are computed

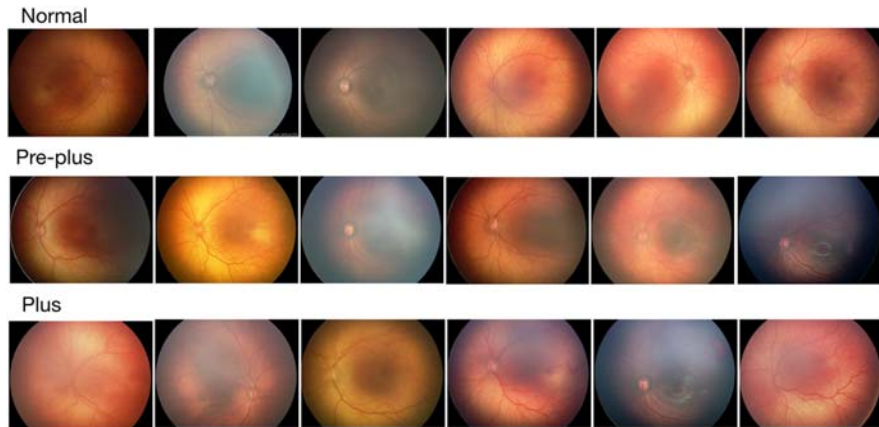
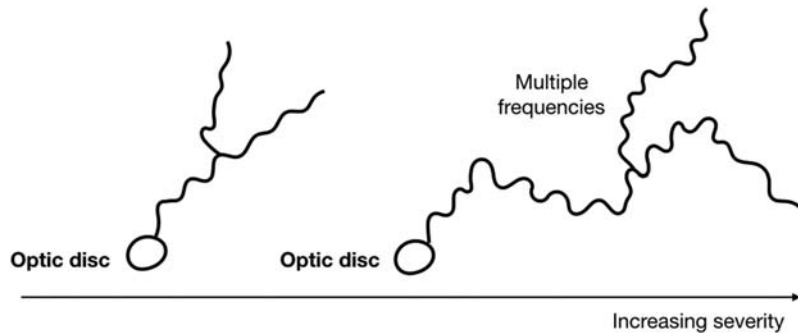
**FIGURE 8.10**

Image examples for each class, namely normal, preplus and plus.

**FIGURE 8.11**

Sketch from the interaction with ophthalmologists during the identification of clinical concepts for the explainability analysis.

on segments of the vessel trace. For example, average segment diameter is obtained by dividing the number of pixels in the vessel by the vessel curve length. Finally, the distance to the center of the optic disk is an example of a tree-based feature. This feature represents, for each vessel segment, the distance between the ending point of the vessel and the disk center. More details and exhaustive comparison of the feature types can be found in [42].

Following this interaction with the physicians, we considered the handcrafted feature design used for ML approaches in ROP applications [67,68]. We computed 11 feature types following the well-established and validated approach in [68]. The pool of the extracted features was separated into two clusters to

differentiate the signal originating from normal and abnormal vessels. The normal and abnormal clusters were then fit into a Gaussian Mixture Model (GMM) and the means, variances and the mixing component were used as GMM statistics. For each of the 11 types of features, we extracted eight standard statistics (minimum, second minimum, maximum, second maximum, mean, median and second and third moments) and the five GMM statistics, obtaining a total of $11 \times (8 + 5) = 143$ handcrafted features. These features were extracted from the automated vessel segmentations obtained by the U-Net model described in 4.1. We trained 100 random forest classifiers on random train-test splits with replacement to rank the features in terms of importance according to their Gini coefficient. The median of the cumulative tortuosity index appeared in the top 5 for all 100 models, confirming the selection of tortuosity as a relevant concept. The Kernel Density Estimation of the top 10 features is shown in Fig. 8.12. We sorted the image samples in the training set by increasing values of the 10 retained feature types and we sampled some of the images to create the visualizations in Figs. 8.13–8.15. We presented these images to the physicians to collect feedback on which extracted features aligned the most with clinically relevant aspects according to them.

From the inspection of the sortings in Figs. 8.13–8.15, the features of curvature “curvature mean” and “curvature median” appeared informative about the class differences, while the utility of the “Average point diameter mean” was not clear to the experts. By this interaction with the physicians, we refined the list of concepts to six measures covering a wide set of clinically interpretable features, including the notion of tortuosity discussed in Fig. 8.11. Features with a frequency of appearance lower than 10% in the ranking were discarded. The retained measures are described in Table 8.1.

Fig. 8.16 shows examples of the vessel segmentations retrieved from the training data according to their minimum and maximum values of the mean and median

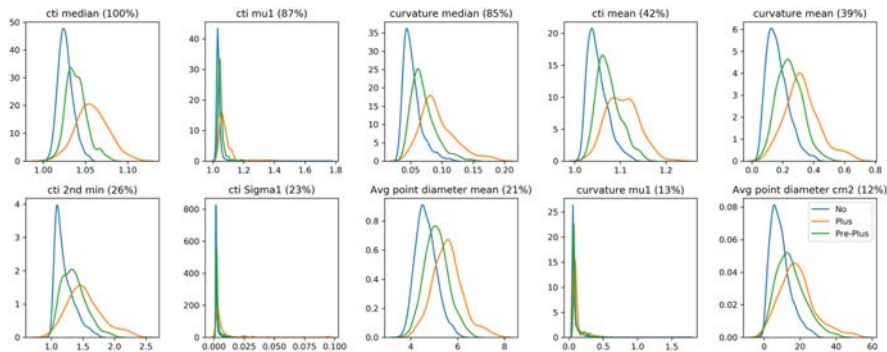
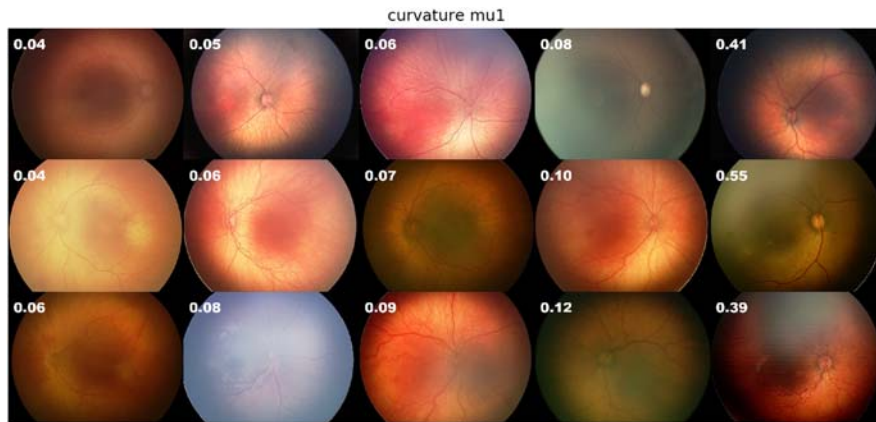
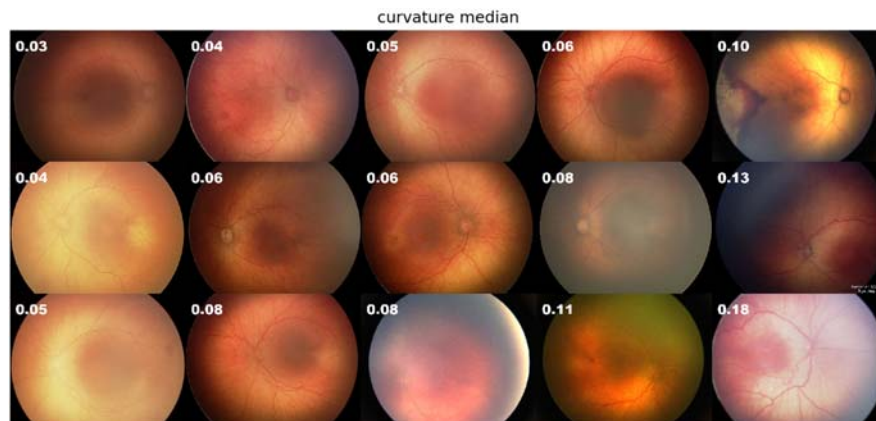


FIGURE 8.12

Kernel Density Estimation of the handcrafted feature values for the three classes (normal, plus and pre-plus).

**FIGURE 8.13**

Images in the training dataset sorted for increasing values of the feature “mean curvature” as defined in [68].

**FIGURE 8.14**

Images in the training dataset sorted for increasing values of the feature “median curvature” as defined in [68].

statistics computed for the features in Table 8.1. The RCVs in the next section will find a direction in the activation space of the CNN layers that represents the change from the minimum to the maximum values of these features.

The analysis in this section led to the central research question in [16], namely whether the concept-based explanations of concept attributions can be used to establish a link between the handcrafted features and the deep features.

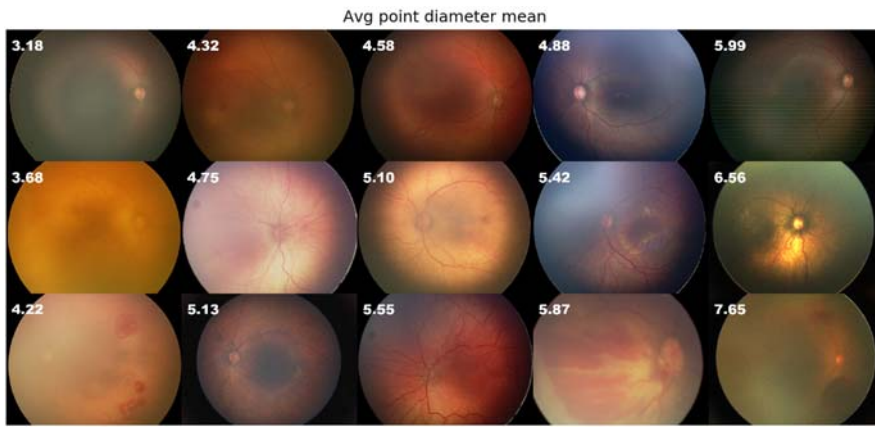


FIGURE 8.15

Images in the training dataset sorted for increasing values of the feature “Average point diameter mean” as defined in [68].

Table 8.1 Handcrafted feature description and clinical interpretation. $k(s)$ describes the rate of changing velocity between points with respect to the rate of changing the curve length between points. L_c and L_x denote respectively curve and chord length. W_n denotes the width of the vessel in the normal direction.

Feature	Name abbreviation	Description	Clinical interpretation
curvature	CURV	$k(s)$	Rate of direction change
Avg Segment Diameter	ASD	$\#pixels/L_c(x)$	Global dilation
Avg Point Diameter	APD	$W_n(x)$	Absolute dilation
Cumulative Tortuosity Index	CTI	$cti(x) = L_c(x)/L_x(x)$	Curving, curling, twisting rate

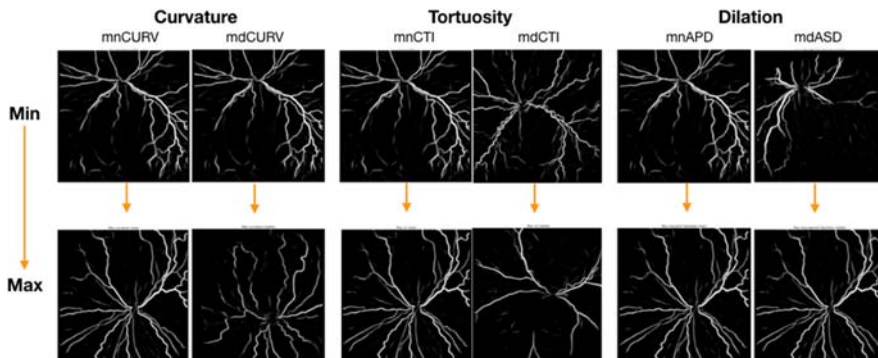


FIGURE 8.16

Examples of the vessel segmentations according to their values of the handcrafted features. The top row shows the masks retrieved from the training data having the lowest value of the feature. The bottom row shows the masks with the largest value of the feature. The mn- and md- prefixes stand respectively for mean and median.

8.4.2.2 Computation of the regression concept vectors

Since the preplus disease represents a natural progression from normal to plus disease, we compute the RCVs on the set of training images for normal and plus. This was not done in [16], where the RCVs were computed separately on the two input classes. The R^2 for multiple layers of the network are reported in Table 8.2, evaluating the presence of the concepts at multiple layers in the CNN, as explained in Section 8.3.2.2. Two pooling strategies for aggregating the feature maps before computing the RCVs are compared, for which we illustrate the

Table 8.2 Coefficient of determination R^2 for the ROP concepts. The pooling strategy is indicated on the top left of each block. The labels of the other columns refer to the layers of Inception-V1. Higher values of R^2 reflect the stronger presence of the concept. Results partially replicated from our study in [13].

max pool	conv1	Mixed3b	Mixed4b	Mixed4c	Mixed5c
<i>medianCTI</i> R^2	0.59	0.66	0.64	0.63	0.67
<i>meanCTI</i> R^2	0.49	0.56	0.50	0.47	0.56
<i>medianCURV</i> R^2	0.65	0.72	0.69	0.67	0.71
<i>meanCURV</i> R^2	0.65	0.70	0.61	0.57	0.72
<i>medianASD</i> R^2	0.55	0.66	0.58	0.56	0.64
<i>medianAPD</i> R^2	0.69	0.76	0.69	0.66	0.76
avg pool	conv1	Mixed3b	Mixed4b	Mixed4c	Mixed5c
<i>medianCTI</i> R^2	0.68	0.75	0.70	0.72	0.72
<i>meanCTI</i> R^2	0.56	0.63	0.54	0.55	0.56
<i>medianCURV</i> R^2	0.62	0.73	0.75	0.76	0.71
<i>meanCURV</i> R^2	0.65	0.74	0.68	0.69	0.71
<i>medianASD</i> R^2	0.69	0.74	0.67	0.67	0.64
<i>meanAPD</i> R^2	0.72	0.80	0.76	0.77	0.76

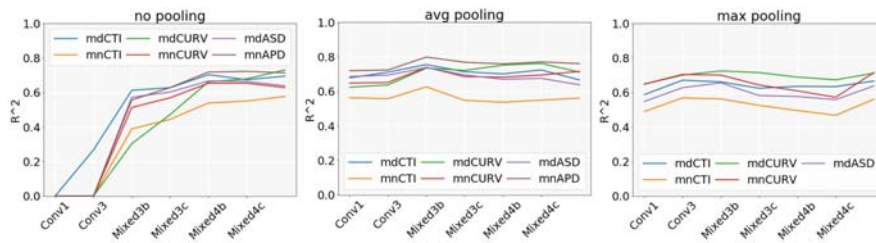


FIGURE 8.17

Comparison of the regression of concepts of curvature (mdCURV and mnCURV), dilation (mdASD, mnAPD) and tortuosity (mdCTI and mnCTI) in ROP images of class normal and plus. Note that the “md” and “mn” prefixes stand respectively for median and mean.

Results replicated from our study in M. Graziani et al., Concept attribution: explaining CNN decisions to physicians, Comput. Biol. Med. 123 (2020) 103865.

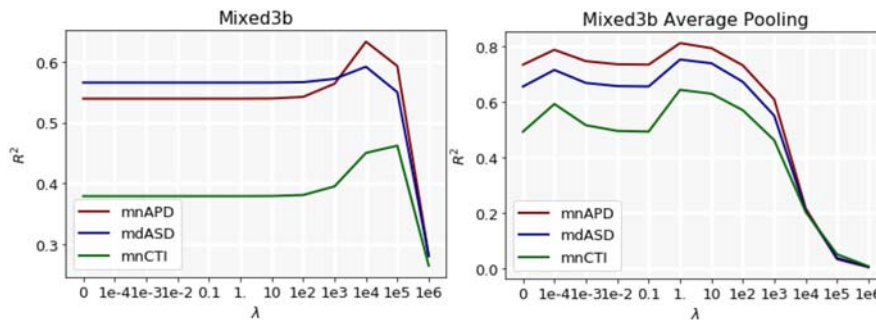


FIGURE 8.18

Impact of the parameter λ (strength of the regularization) on the ridge regression with (on the *left*) and without (on the *right*) global average pooling for the ROP concepts. The pooling operation reduces the need for regularization and leads to higher values of R^2 . A subset of ROP concepts is shown, representing dilation (mdASD and mnAPD) and tortuosity (mnCTI).

Results replicated from our study in M. Graziani et al., Concept attribution: explaining CNN decisions to physicians, Comput. Biol. Med. 123 (2020) 103865.

differences in Fig. 8.17. The results for the regularized regression are compared against multiple values of the regularization penalty in Fig. 8.18.

8.4.2.3 Evaluation of the conceptual sensitivities

In this section, we present two examples of the conceptual sensitivities as local explanations, as introduced in Section 8.3.2.3.

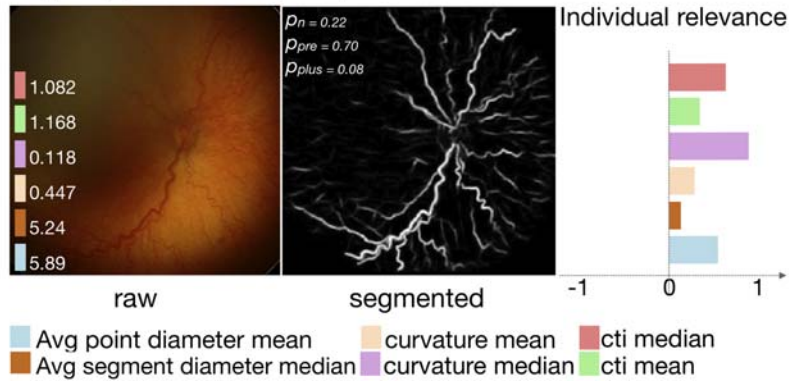
GT: *plus*; prediction: *pre-plus*

FIGURE 8.19

Conceptual sensitivities for a misclassification of a plus image as a preplus. The original values of six concept measures are displayed on top of the raw input image on the left. The network probabilities for the three classes, normal, preplus and plus are reported as p_n , p_{pre} and p_{plus} .

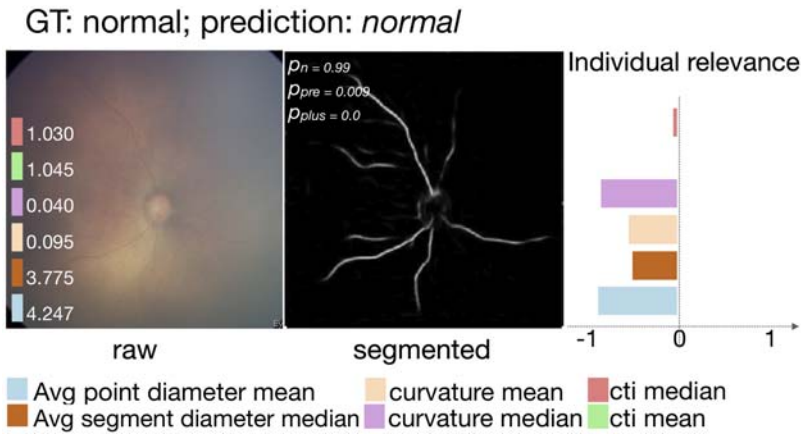
Image reproduced from our work in M. Graziani, et al. Improved interpretability for computer-aided severity assessment of retinopathy of prematurity, in: Medical Imaging 2019: Computer-Aided Diagnosis. Vol. 10950. International Society for Optics and Photonics, 2019.

In Fig. 8.19, we show the sensitivities for a misclassified image. The original values of the handcrafted features (which were used as concept measures) are reported on the left of the image. The network probability of each class is shown on top of the segmentation as p_n , p_{pre} and p_{plus} . The analysis highlights the fact that higher values of curvature and tortuosity would increase the prediction probability of the plus class. Similarly, Fig. 8.20 presents the conceptual sensitivities for a correctly classified image.

8.4.2.4 Global explanations with Br

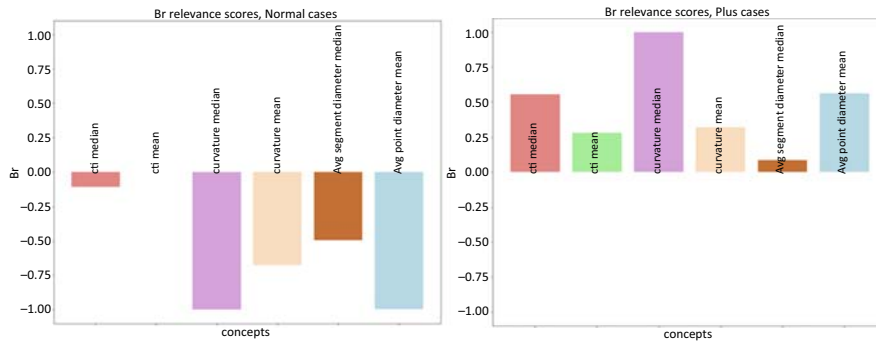
The global explanations, as explained in Section 8.3.2.4, are summarized for inputs of the normal and plus classes in Fig. 8.21.

From the global explanations, curvature median appears as the most relevant concept to detect plus images with $Br = 1.0$. Avg point diameter mean is, on the other side, the most important concept for the detection of normal cases with $Br = -0.99$. The negative score shows that an increase of the value for this clinical concept would correspond to a decrease in the network output, hence a shift toward the prediction of the normal class. Avg point diameter mean and CTI median appear as equally important with $Br = 0.56$ for the detection of plus.

**FIGURE 8.20**

Conceptual sensitivities for a correct classification of an image of the normal class. The original values of six concept measures are displayed on top of the raw input image on the left. The network probabilities for the three classes, normal, preplus and plus are reported as p_n , p_{pre} and p_{plus} .

Image reproduced from our work in M. Graziani, et al. Improved interpretability for computer-aided severity assessment of retinopathy of prematurity, in: Medical Imaging 2019: Computer-Aided Diagnosis. Vol. 10950. International Society for Optics and Photonics, 2019.

**FIGURE 8.21**

Global Br scores on the testing set for normal and plus images. Positive scores represent a shift toward the prediction of the normal class (*left*) or plus class (*right*) when the concept measure increases. Negative scores represent a shift toward these same classes when the concept measure decreases.

Figure reproduced from our work in M. Graziani, et al. Improved interpretability for computer-aided severity assessment of retinopathy of prematurity, in: Medical Imaging 2019: Computer-Aided Diagnosis. Vol. 10950. International Society for Optics and Photonics, 2019.

8.5 Discussion of the results

The presence of plus disease has a relevant impact on the treatment planning for ROP. Its diagnosis is, however, highly subjective, being mostly based on the identification of vessel dilation and tortuosity. The performance for the diagnosis of plus disease of the fully automated system in [Section 8.4.1](#) compares, if not exceeds, that of ROP experts [69]. This result aims at showing that the use of DL models can introduce objectivity in the assessment of ROP severity, supporting physicians with difficult decisions such as establishing the presence of plus disease. This result shows a high potential of improving the clinical outcomes from the integration of DL and experts, similar to the results in other medical applications [7,8].

Interpreting the model predictions is a necessary step to validate the model's decision-making. The proposed approach to ROP classification is particular since the images classified by the Inception-V1 are not continuous multichannel inputs, like natural images, they are binary masks of vessel segmentations. Feature attribution methods to obtain visualizations may therefore not provide sufficient insights on the decision-making. Since the inputs are binary masks, this application is also challenging for concept attribution. Despite their versatility in many other applications [12–14,70], basic visual features such as image intensity and texture cannot be extracted from the binary masks of the vessels. The concept selection had to be defined on purpose for this task. The interaction with the ophthalmologists was essential to the formulation of the clinical concepts. The tortuosity measures of CTI emerge as relevant from the exchange with the physicians in [Fig. 8.11](#). This result is also in agreement with the analysis of the Gini coefficients, with CTI appearing in the top five for all the training repetitions of the random forests model classification. The vessel curvature is another interesting feature, according to the Gini coefficients. The visual differences between images with increasing values of vessel curvature features (i.e. curvature mean and median) suggest the relation between increased curvature and the presence of plus. The selection of clinical concepts has both upsides and downsides. On the upside, arbitrary concepts can be used to formulate explanations that directly address specific answers about the application. They do not need large annotated datasets, since the concepts can be computed automatically on the images, for example the handcrafted vessel features. The downside of the arbitrary choice is that the selection of the concepts itself is a delicate process, as seen in [Section 8.4.2.1](#). The selection needs multiple iterations and it requires the participation of experts, to find the clinically relevant visual patterns, and developers, to implement the modeling of such patterns.

The computation of the RCVs (in [Section 8.4.2.2](#)) is straightforward after selecting clinical concepts. The results compare the RCVs obtained by applying the improvements proposed in [13] to obtain more stable vectors, including appropriate feature map pooling and regression regularization. The average pooling of

the features leads to the most stable vectors, with the regularization only leading to small improvements. It is important to note that we compute R^2 and relevance scores on unseen test data, and not on the training data as generally done in statistics. The rationale behind this is to check whether the correlation learned in the input features can generalize and is robust enough to be predictive of unseen data. This is informative on the robustness and the reliability of the explanations. The risk of CAVs of capturing spurious correlation is therefore reduced in this application of RCVs by this evaluation on test data. Yet, more research is still needed to clarify the causal link between the presence of the concept and the decision [56].

The insights given by the local (in Section 8.4.2.3) and global explanations (in Section 8.4.2.4) represent a first attempt in bridging the gap between handcrafted visual features used for plus disease detection in classic ML approaches and the data-driven learning of features that is automated in CNNs. The scores reflect the clinical expectation that emerged from the interaction with the physicians, reporting high relevance for curvature and tortuosity in the diagnosis of plus disease. The relevance of average diameter mean as a discriminant factor for normal images is yet to be investigated. The visualizations in Figs. 8.19 and 8.20 propose a possible way of integrating the local explanations as a tool to assist the diagnosis, showing the conceptual sensitivities, the original and segmented images and the raw values of the hand-crafted features.

8.6 Conclusions

This chapter has covered important topics in the quest for interpretable AI in the medical domain, presenting an explainability approach with an application on ROP. The often unclear terminology has led to confusion and multiple taxonomies for interpretability [19–24]. By reviewing these works, we identified in section 2.2.2 the terms for which most of the taxonomies agree in the definition. Importantly, we clarified the use of interpretability, explainability, and intelligibility. Interpretability and explainability can be used interchangeably for referring to the generation of explanations for the model decisions. Intelligibility refers to a wider group of methods that includes inherently interpretable models [3] and the introduction of interpretability as an additional model objective [25].

We presented the framework of concept attribution as opposed to the visualization techniques that are wide-spreading in the medical community. As discussed in Section 2.4, visualization methods may lead to unstable explanations that do not inspire reliability [33,60]. Concept attribution comes as an alternative approach to visualizations that can provide further insights on the network decision-making, both at the global and the local level. Being post hoc, it does not need the retraining of the parameters, and it can thus be applied to any network. If a more performant and accurate model needs to be developed for ROP, concept attribution could be applied to the updated model.

The use of clinical concepts to explain the decisions may foster the comparison between the explanations of the models used in multiple institutions. This is in line with future developments of AI for healthcare, with federated learning approaches also promoting the exchange of information [71]. Clinical concepts, moreover, generate explanations that are at a higher level of abstraction than heat-maps. This makes the comparison of network behavior independent from the input images used to generate the explanations. By selecting concepts that match preexisting guidelines, explanations can help the physicians for verifying if the same set of values of principles is followed by the model decision-making. New hypotheses on the learned clinical concepts can be tested, also to verify that the network does not contain biases. For example, CAVs and RCVs could be used to inspect if the watermarks and text annotations, often present in medical images, affect the classifications.

From a more global perspective, explaining the automated decision-making of AI is a task at the frontier of two research worlds: the clinical and the developmental. Explanations should be generated with a human-centric approach, considering the requirements of the receivers of the explanations. For this reason, domain experts and DL developers should join forces to develop methods that can make the automated choices less intimidating and more understandable for physicians, while at the same time more stable and reliable from the development perspective.

Acknowledgments

This work was possible because of the following projects, part of the European Union's Horizon 2020 research and innovation program: PROCESS (grant agreement No 777533), AI4MEDIA (grant agreement No 825619) and EXAMODE (grant agreement No 825292). We thank Jayashree Kalpathy-Cramer and Michael F. Chiang for the important directives on the research and the explanations on ROP, James Brown for providing the deepROP code, the trained model weights, and some of the images in this chapter. We also thank Veysi Yildiz for providing the feature vectors representing the clinical concepts.

References

- [1] L. Nanni, S. Ghidoni, S. Brahmam, *Handcrafted vs. non-handcrafted features for computer vision classification*, *Pattern Recognit.* 71 (2017) 158–172.
- [2] J. Zhou, F. Chen, *DecisionMind: revealing human cognition states in data analytics-driven decision making with a multimodal interface*, *J. Multimodal User Interfaces* 12 (2) (2018) 67–76.
- [3] R. Caruana et al., *Intelligible models for healthcare: predicting pneumonia risk and hospital 30-day readmission*, in: *Proceedings of the Twenty-first ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2015.

- [4] A.J. London, Artificial intelligence and black-box medical decisions: accuracy versus explainability, *Hastings Center Report* 49 (1) (2019) 15–21.
- [5] S. Yune, et al. Real-world performance of deep-learning-based automated detection system for intracranial hemorrhage, in: *Proceedings of the SIIM Conference on Machine Intelligence in Medical Imaging*, San Francisco, 2018.
- [6] F. Doshi-Velez, B. Kim, Toward a rigorous science of interpretable machine learning. *arXiv:1702.08608*, 2017.
- [7] J.M. Brown, J.P. Campbell, A. Beers, K. Chang, S. Ostmo, R.V.P. Chan, et al., Imaging and informatics in retinopathy of prematurity (i-ROP) research consortium. automated diagnosis of plus disease in retinopathy of prematurity using deep convolutional neural networks, *JAMA Ophthalmol.* 136 (7) (2018) 803–810. Available from: <https://doi.org/10.1001/jamaophthalmol.2018.1934>. PMID: 29801159; PMCID: PMC6136045.
- [8] D. Wang et al., Deep learning for identifying metastatic breast cancer. *arXiv:1606.05718*, 2016.
- [9] B. Zhou et al., Learning deep features for discriminative localization, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016.
- [10] B. Kim et al., Interpretability beyond feature attribution: quantitative testing with concept activation vectors (TCAV), in: *Proceedings of the International Conference on Machine Learning*, PMLR, 2018.
- [11] C. J. Cai et al. Human-centered tools for coping with imperfect algorithms during medical decision-making, in: *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 2019.
- [12] M. Graziani, V. Andrearczyk, H. Müller, Regression concept vectors for bidirectional explanations in histopathology. *Understanding and Interpreting Machine Learning in Medical Image Computing Applications*, Springer, Cham, 2018, pp. 124–132.
- [13] M. Graziani, et al., Concept attribution: explaining CNN decisions to physicians, *Comput. Biol. Med.* 123 (2020) 103865.
- [14] M. Graziani, et al., Interpretable CNN pruning for preserving scale-covariant features in medical imaging. *Interpretable and Annotation-Efficient Learning for Medical Image Computing*, Springer, Cham, 2020, pp. 23–32.
- [15] H. Yeche, J. Harrison, T. Berthier, UBS: a dimension-agnostic metric for concept vector interpretability applied to radiomics. *Interpretability of Machine Intelligence in Medical Image Computing and Multimodal Learning for Clinical Decision Support*, Springer, Cham, 2019, pp. 12–20.
- [16] M. Graziani et al. Improved interpretability for computer-aided severity assessment of retinopathy of prematurity, in: *Proceedings of the Medical Imaging 2019: Computer-Aided Diagnosis*. Vol. 10950. International Society for Optics and Photonics, 2019.
- [17] L. Edwards, M. Veale, Slave to the algorithm: Why a right to an explanation is probably not the remedy you are looking for, *Duke L. Tech. Rev.* 16 (2017) 18.
- [18] L.M. Cysneiros, M. Raffi, J.C. Sampaio do Prado Leite, Software transparency as a key requirement for self-driving cars, in: *Proceedings of the IEEE Twenty-sixth International Requirements Engineering Conference (RE)*, Banff, AB, 2018, pp. 382–387. Available from: <https://doi.org/10.1109/RE.2018.00-21>.
- [19] T. Miller, Explanation in artificial intelligence: Insights from the social sciences, *Artif. Intell.* 267 (2019) 1–38.

- [20] M.-A. Clinciu, H. Hastie, A survey of explainable AI terminology, in: Proceedings of the First Workshop on Interactive Natural Language Technology for Explainable Artificial Intelligence (NL4XAI 2019), 2019.
- [21] M. Chromik, M. Schuessler, A taxonomy for human subject evaluation of black-box explanations in XAI, ExSS-ATEC@ IUI, 2020.
- [22] A.B. Arrieta, et al., Explainable Artificial Intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI, *Inf. Fusion* 58 (2020) 82–115.
- [23] A. Adadi, M. Berrada, Peeking inside the black-box: a survey on explainable artificial intelligence (XAI), *IEEE Access*, 2018, pp. 1–1. <https://doi.org/10.1109/ACCESS.2018.2870052>.
- [24] E. Tjoa, C. Guan, A survey on explainable artificial intelligence (XAI): toward medical XAI. *ArXiv abs/1907.07374*, 2019, n. pag.
- [25] D. Bertsimas, A. King, R. Mazumder, Best subset selection via a modern optimization lens, *Ann. Statist.* (2016) 813–852.
- [26] Z.C. Lipton, The mythos of model interpretability, *Queue* 16 (3) (2018) 31–57.
- [27] X. Wang, et al. Chestx-ray8: hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017.
- [28] R. Gargeya, T. Leng, Automated identification of diabetic retinopathy using deep learning, *Ophthalmology* 124 (7) (2017) 962–969.
- [29] C. González-Gonzalo, et al. Improving weakly-supervised lesion localization with iterative saliency map refinement, 2018.
- [30] Y. Huang, A.C.S. Chung, Evidence localization for pathology images using weakly supervised learning, in: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, Cham, 2019.
- [31] B. Korbar, et al. Looking under the hood: deep neural network visualization to interpret whole-slide image analysis outcomes for colorectal polyps, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2017.
- [32] Y. Xu, et al., Large scale tissue histopathology image classification, segmentation, and visualization via deep convolutional activation features, *BMC Bioinform.* 18 (1) (2017) 1–17.
- [33] M. Graziani, et al. Evaluation and comparison of CNN visual explanations for histopathology, 2020.
- [34] M. Reyes, et al., On the interpretability of artificial intelligence in radiology: challenges and opportunities, *Radiol.: Artif. Intell.* 2 (3) (2020) e190043.
- [35] S. Pereira, et al., Automatic brain tumor grading from MRI data using convolutional neural networks and quality assessment. *Understanding and Interpreting Machine Learning in Medical Image Computing Applications*, Springer, Cham, 2018, pp. 106–114.
- [36] A. Hosny, et al., Deep learning for lung cancer prognostication: a retrospective multi-cohort radiomics study, *PLoS Med.* 15 (11) (2018) e1002711.
- [37] P. Koh, Wei, P Liang, Understanding black-box predictions via influence functions. *arXiv:1703.04730*, 2017.
- [38] Raghu, M, et al. SVCCA: singular vector canonical correlation analysis for deep learning dynamics and interpretability. *Adv. Neural Inf. Process. Syst.* 2017.

- [39] M. Sundararajan, A. Taly, Q. Yan, Axiomatic attribution for deep networks. *arXiv:1703.01365*, 2017.
- [40] D. Erhan, et al., Visualizing higher-layer features of a deep network, *Univ.Montreal 1341 (3) (2009) 1*.
- [41] M D. Zeiler, R. Fergus, Visualizing and understanding convolutional networks, in: *Proceedings of the European Conference on Computer Vision*, Springer, Cham, 2014.
- [42] K. Simonyan, A. Vedaldi, A. Zisserman Deep inside convolutional networks: visualising image classification models and saliency maps, 2014.
- [43] J. T. Springenberg, et al., Striving for simplicity: the all convolutional net. *arXiv:1412.6806*, 2014.
- [44] S. Bach, et al., On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation, *PLoS One 10 (7) (2015) e0130140*.
- [45] R. R. Selvaraju, et al. Grad-cam: visual explanations from deep networks via gradient-based localization, in: *Proceedings of the IEEE International Conference on Computer Vision*, 2017.
- [46] A. Chattopadhyay, et al. Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks, in: *Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2018.
- [47] L M. Zintgraf, et al. Visualizing deep neural network decisions: prediction difference analysis, *ICLR*, 2017.
- [48] R.-Š. Marko, I. Kononenko, Explaining classifications for individual instances, *IEEE Trans. Knowl. Data Eng.* 20 (5) (2008) 589–600.
- [49] M. Ribeiro, S Tulio, Singh, C Guestrin, Why should I trust you? Explaining the predictions of any classifier, in: *Proceedings of the Twenty-Second ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016.
- [50] R. Achanta, et al., SLIC superpixels compared to state-of-the-art superpixel methods, *IEEE Trans. Pattern Anal. Mach. Intell.* 34 (11) (2012) 2274–2282.
- [51] P.F. Felzenszwalb, P.H. Daniel, Efficient graph-based image segmentation, *Int. J. Comp. Vision* 59 (2) (2004) 167–181.
- [52] A. Shrikumar, P. Greenside, A. Kundaje, Learning important features through propagating activation differences. *arXiv:1704.02685*, 2017.
- [53] S.M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, *Adv. Neural Inf. Process. Syst.* (2017).
- [54] G. Alain, Y. Bengio, Understanding intermediate layers using linear classifier probes. *arXiv:1610.01644*, 2016.
- [55] A. Ghorbani, et al., Toward automatic concept-based explanations, *Adv. Neural Inf. Process. Syst.* (2019).
- [56] Y. Goyal, et al. Explaining classifiers with Causal Concept Effect (CaCE). *arXiv:1907.07165*, 2019.
- [57] P.W. Koh, et al. Concept bottleneck models. *arXiv:2007.04612*, 2020.
- [58] M. Yang, B. Kim, Benchmarking attribution methods with relative feature importance, *arXiv-1907*, 2019.
- [59] S. Tonekaboni, et al. What clinicians want: contextualizing explainable machine learning for clinical end use. *arXiv:1905.05134*, 2019.
- [60] N. Arun, et al. Assessing the (un) trustworthiness of saliency maps for localizing abnormalities in medical imaging. *arXiv:2008.02766*, 2020.

- [61] J. M. Brown, et al. Fully automated disease severity assessment and treatment monitoring in retinopathy of prematurity using deep learning, in: Proceedings of the Medical Imaging 2018: Imaging Informatics for Healthcare, Research, and Applications. Vol. 10579. International Society for Optics and Photonics, 2018.
- [62] O. Ronneberger, P. Fischer, T. Brox U-net: Convolutional networks for biomedical image segmentation, in: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, Cham, 2015.
- [63] C. Szegedy, et al. Going deeper with convolutions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015.
- [64] J. Whitney, et al., Quantitative nuclear histomorphometry predicts oncotype DX risk categories for early stage ER + breast cancer, *BMC Cancer* 18 (1) (2018) 610.
- [65] X. Wang, et al. Computer extracted features of cancer nuclei from H&E stained tissues of tumor predicts response to nivolumab in non-small cell lung cancer, 2018, pp. 12061–12061.
- [66] G. Lee, et al. Cell orientation entropy (CORE): predicting biochemical recurrence from prostate cancer tissue microarrays, in: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Berlin, Heidelberg, 2013.
- [67] W.E. Hart, et al., Measurement and classification of retinal vascular tortuosity, *Int. J. Med. Inform.* 53 (2–3) (1999) 239–252.
- [68] E. Ataer-Cansizoglu, et al., Computer-based image analysis for plus disease diagnosis in retinopathy of prematurity: performance of the “i-ROP” system and image features associated with expert diagnosis, *Trans. Vision Sci. Technol.* 4 (6) (2015) 5. -5.
- [69] J.M. Brown, et al., Automated diagnosis of plus disease in retinopathy of prematurity using deep convolutional neural networks, *JAMA Ophthalmol.* 136 (7) (2018) 803–810.
- [70] M. Graziani, H. Muller, V. Andrearczyk Interpreting intentionally flawed models with linear probes, in: Proceedings of the IEEE International Conference on Computer Vision Workshops, 2019.
- [71] K. Chang, et al., Distributed deep learning networks among institutions for medical imaging, *J. Am. Med. Inform. Assoc.* 25 (8) (2018) 945–954.

Computational lung sound classification: a review

9

Truc Nguyen and Franz Pernkopf

Signal Processing and Speech Communication Lab., Graz University of Technology, Graz, Austria

9.1 Introduction

Respiratory diseases represent an immense worldwide health burden. According to the World Health Organization (WHO), the “big five” respiratory diseases, which include asthma, chronic obstructive pulmonary disease (COPD), acute lower respiratory tract infections, lung cancer and tuberculosis, cause severe illness and mortality of more than 7 million people each year worldwide [1]. Currently, COVID-19 has caused globally more than 158 million infections and 3,296,000 deaths [2]. On March 11, 2020, the WHO officially announced that COVID-19 has reached global pandemic status. Furthermore, according to [3], the “big five” lung diseases, except lung cancer, have increased during COVID-19 epidemics. These respiratory diseases are characterized by highly similar symptoms, that is the adventitious breathing, which could be a confounding factor during diagnosis [4]. Owing to their severe consequences, particularly in the case of COVID-19, an early and accurate diagnosis of these types of diseases has become crucial.

Most diseases related to an obstructed or restricted respiratory system can be characterized from the sounds generated while breathing [5]. If there are physiological changes in the lungs due to a disease, it can cause changes in the lung sound pattern [6]. Therefore, lung sounds can be considered as relevant indicators of respiratory health [7,8]. The lung sounds are classified as normal and adventitious. Normal respiratory sounds are heard when no respiratory disorders exist. Adventitious lung sounds (ALSs) are superimposed with the normal respiratory sounds [9]. Depending on the duration, ALSs can be categorized into two main types: continuous and discontinuous. Continuous adventitious sounds (CASs) are usually longer than 250 ms of duration. There are different kinds of CASs corresponding to different frequency ranges such as wheezes with a pitch higher than 400 Hz, rhonchi with a maximal pitch at 200 Hz, stridor with a pitch higher than 500 Hz and squawk’s pitch around 200–300 Hz. In contrast, discontinuous adventitious sounds (DASs) have a duration shorter than 25 ms. They can be classified as fine or coarse depending on their duration and frequency. Fine crackles have a short duration of 5 ms and high frequency with a pitch around

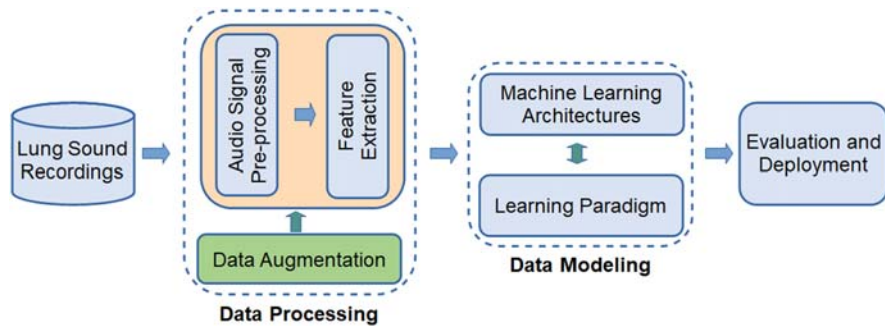
650 Hz, whereas coarse crackles have a longer duration that is 15 ms and a lower frequency of around 350 Hz. Beside that, pleural rub is around 15 ms duration and its pitch is often lower than 350 Hz [5].

Auscultation is a traditionally important method to diagnose pulmonary diseases and abnormalities using a stethoscope. This process mainly relies on the physician [10]. In the last decades, computational methods that is computational lung sound analysis (CLSA) have been developed for automated detection and classification of ALSs. In particular, digital recording, signal processing techniques, and machine learning algorithms are used to analyze the signals. They potentially overcome the conventional method's limitations such as subjectiveness and offer advantages for medical diagnosis [5,11]. Furthermore, they are carefully evaluated in real-life scenarios and can be used as portable easy-to-use devices without the necessity of expert interaction.

Over the years, the popularity of deep learning [12] has introduced some noticeable changes to the classical pattern recognition framework in general and for CLSA systems in particular. The processing steps of conventional machine learning methods involve preprocessing/transformation of the lung sound signal, feature extraction, and classification [13]. Most research following this paradigm has focused on enhancing the robustness of each of these steps. Feature extraction is used to extract informative feature vectors to represent patterns in specific ways relevant to the task. This is described in the literature as handcrafted or handmade features. Its main objective is to create features that place patterns belonging to the same class close to each other in the feature space, while simultaneously maximizing their distance to the other classes. However, with deep learning, the handcrafted features are replaced by 2D representations, for example, time–frequency representations such as spectrograms or mel frequency cepstral coefficients (MFCCs) of the audio signals.

In CLSA, there are two popular classification tasks, namely (1) ALS classification and (2) respiratory disease classification (RDC). ALS classification (ALSC) is a task recognizing either abnormal events (i.e., either crackles or wheezes or others) or normal and abnormal sounds that is respiratory cycles or recordings including ALS events [14–29]. While for RDC, several categories have been considered for example, binary classification (health and pathological) [10,30,31], ternary chronic classification (healthy, chronic and nonchronic diseases) [27,31–36] or multiple-class classification of distinct pathologies [37–44]. The systems have been evaluated on several nonpublic datasets such as the Tromsø 7 study [45] or the multichannel lung sound data [46] and public datasets such as R.A.L.E. [47], the ICBHI 2017 dataset [8], the Abdullah University Hospital 2020 dataset [48] or the largest open-access lung sound database, namely the HF_Lung_V1 [49].

In this chapter, we summarize and categorize systematically algorithms for LSC based on processing steps illustrated in Fig. 9.1. Data processing techniques for audio signal preprocessing, feature extraction, and data augmentation are detailed in Sections 9.2. Section 9.3 discusses neural network architectures and

**FIGURE 9.1**

Processing steps of a typical lung sound classification system.

other learning paradigms. Furthermore, lung sound datasets used in the lung sound classification tasks are introduced in [Section 9.4](#). [Section 9.5](#) discusses recent advances and open challenges of LSC and concludes the chapter.

As a complementary read to this chapter, Pramono et al. [5] published a systematic review of automatic adventitious respiratory sound analysis between 1938 and 2016 using handcrafted features for conventional machine learning such as k-nearest neighbor classifiers, support vector machines, logistic regression, probabilistic classification, etc. In addition, Nguyen et al. [36] also categorized algorithms based on conventional machine learning and deep learning approaches for both ALSC and RDC using the ICBHI 2017 dataset. Methods for evaluating LSC algorithms are not further addressed in this chapter.

9.2 Data processing

Lung sound signals are usually represented as discrete-time samples. The recording properties of the lung sound strongly depend on the recording devices. In addition to single signal (channel) preprocessing techniques, a multichannel lung sound processing has been proposed, where the multiple signals are recorded simultaneously by the recording device [28]. In this section, we overview (1) preprocessing techniques of audio signals and (2) the most relevant features used in LSC systems.

9.2.1 Audio signal preprocessing

As the audio signals of respiratory sounds are often recorded using different equipment under verified conditions, they needed to be preprocessed to have the same fundamental characteristics. There are relevant preprocessing techniques such as (1) signal splitting, (2) noise filtering, (3) resampling, (4) amplitude

scaling between -1 and 1 and (5) segment splitting and (6) padding as shown in Fig. 9.2.

9.2.1.1 Signal splitting

There are three popular analysis levels of lung sounds in LSC systems, namely event-wise, respiratory cycle-wise, and recording-wise. These levels correspond to the tasks of adventitious event detection [14–19], ALS classification [20–29], and RDC [10,27,30–43], respectively. Lung sound datasets, which are used for evaluation of the LSC systems, provide annotations for recordings at different analysis levels. They can be used to split lung sound recordings into adventitious events or respiratory cycles and their corresponding labels.

9.2.1.2 Noise filtering

Filtering is applied to remove all nonrelevant information such as heart sounds, background noise, and characteristics of the recording device. Commonly, band-pass filters are used in both conventional LSC systems [19,38,39] and deep learning-based LSC systems [30,37,42].

9.2.1.3 Resampling

Lung sound signals have been recorded with a wide range of sampling frequencies [8]. As adventitious lung signals are typically presented within a frequency range of up to 2 kHz, the sampling rate is at 4 kHz or more. For further processing, the same sampling frequency for the data is advocated.

9.2.1.4 Amplitude scaling

In conventional LSC systems, amplitude scaling is usually performed on each signal so that all samples have values between -1 and 1. This compensates for the intensity difference of data collected from different sources while preserving the important statistical parameters of the respiratory sounds.

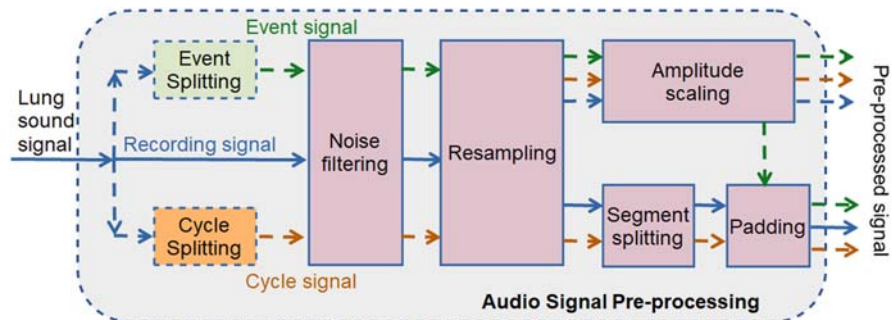


FIGURE 9.2

Various preprocessing steps for different signal analysis levels.

9.2.1.5 Segment splitting

In deep learning-based systems of adventitious LSC [29,36,50–53] and RDC [51], audio signals are split with/without overlap into segments with a fixed length.

9.2.1.6 Padding

Padding of audio signals is mostly used for deep learning-based LSC systems using convolutional neural networks (CNNs) [29,50–53]. They often require the same portion of data extracted from the preprocessed audio signals. Usually, there is a wide variation in length of the respiratory cycles and recordings. Hence, audio signals are often split with/without overlap into segments of a fixed length and it is necessary to complete partial segments. There are various padding methods available such as zero padding [52–54], sample padding [29,35,51] or smart padding [50]. According to [29,50,51], the LSC systems using sample padding and smart padding outperform that using zero padding with the same system settings. In addition, smart padding is also considered to augment data [50] which is discussed in detail in Section 9.2.3.

It is noticeable that some deep learning-based LSC systems, which take advantage of high-level feature extraction and normalization of the deep neural network architectures, ignore noise filtering and amplitude scaling [32–36]. Furthermore, there are several systems performing splitting and padding or resizing of 2D signal representations [32–35,55], instead of processing the 1D audio signals.

9.2.2 Feature extraction

Feature extraction is an important step for practical algorithms in pattern recognition and classification. It enhances the information for classification. For LSC systems, there are two popular types of feature extraction (1) features for conventional classifiers and (2) time-frequency representations for deep learning. Common features of each category are summarized in Tables 9.1 and 9.2.

9.2.2.1 Features for conventional classifiers

According to a systematic review [5], automatic ALS detection and classification systems between 1938 and 2016 were mostly introduced within the classical pattern recognition framework and relied on the hand-crafted features. The hand-crafted features encode the information after removing nonrelevant information in the preprocessing step. A compact set of features (feature vectors) are presented as input to the classifier. The hand-crafted feature set usually consists of features in the time domain, in the frequency domain that is cepstral features, spectral features (i.e., spectral centroid, brightness, entropy, kurtosis, skewness, roll-off, flux); melodic features (i.e., pitch, chromagrams); MFCCs, linear prediction coefficients (LPCs), linear prediction cepstral coefficients (LPCCs) or features of discrete or continuous wavelet transform (DWT/CWT). They have been used and further

Table 9.1 Summary of classical features. Several columns within the column “features” indicate a sequential decomposition.

Type	Features		References	
			Before 2019	From 2019
Spectral	Spectral centroid, spectral spread, spectral skewness, spectral kurtosis, spectral entropy, spectral flatness, spectral roughness, spectral irregularity, spectral flux, spectral flux inc, spectral halfwave, spectral median, spectral brightness, spectral roll-off, power spectrum, deformation, spectrum mean, median frequency		[5,15,30]	[14,19,44,56,57]
	Power spectral density		[5,58]	
	Mel spectrogram	Features on Local binary pattern (LBP)	[20,25]	
Melodic	Pitch, pitch smoothing, inharmonicity, inharmonicity smoothing, voicing, voicing smoothing		[5,30]	[14,19]
MFCCs	MFCCs		[5,10,22,25,26,31,58]	[14,19]
	Delta-MFCCs			
Linear prediction coefficients (LPCs)	LPCs		[5]	[59]
LPCCs	LPCC-based features		[5]	[23]
Empirical mode decomposition (EMD) variants	Intrinsic mode functions (IMFs)	Gray level difference Matrix (GLDM)	Gradient second moment, gradient contrast, mean gradient, inverse different moment, gradient entropy	[60]
	IMFs	Spectral features, their statistics, MFCCs		[5]
	IMFs	Gammatone cepstral coefficients (GTCCs)		[62]
	IMFs	2D, 3D phase space representations	Ellipse area, interquartile range	[63]
Hilbert–Huang Transform	IMFs		[21]	[64]

Time domain	Zero-crossing rate Variance, range, sum of moving averaging Amplitude and statistical features Shannon entropy, energy entropy	[5,15,31,44]	[14,61]
Wavelet transform (WT)	Wavelet coefficients, entropy Spectral features, their statistics, MFCCs	[5,31,58,65]	[14,66–68] [61,67]

Table 9.2 Summary of time-frequency features. Several columns within the column “features” indicate a sequential decomposition of the methods.

Type	Features	References	
STFT	Spectrogram	[10,16,19,28,38,39,41,49,69–72]	
	Spectrogram	Color image	[73–76]
		Gray image	[38]
	Mel spectrogram	[19,29,32,33,37,51,53]	
	Mel spectrogram image	[36,77]	
	Gammatonegram	[32–35]	
	Gammatonegram image	[55]	
	MFCCs	[16,26,27,32,42,49,52,53,56,78–80]	
	STFT coefficients	[52]	
	STFT-based chromagram	[53]	
Scalogram/scalogram image	[35,69,81,82]		
Wavelet transform			
Constant-Q transform (CQT)	CQT-based chromagram	[32,33,53]	
	Chroma Energy Normalized (CENS) chromagram	[53]	
	CQT cepstral coefficient	[83]	
S-transform	Spectrogram	[18]	
EMD	Scalogram of CWT	[43]	
Frequency domain	Energy summation (EM)	[49]	
	Pitch, median frequency and bandwidth	[17]	

explored in many recent LSC studies. Table 9.1 summarizes common hand-crafted features including references to literature.

Feature selection and/or transformation algorithms such as the principal component analysis [58,61], linear Wilcoxon Rank Sum statistical test [14,31], minimum redundancy maximum relevance algorithm [19], random forests [24], and auto-encoders [61] are used to determine the significance of the extracted features and to remove redundant features. This helps to enhance the performance in terms of accuracy and computational cost of the classifiers.

It is noticeable that in conventional LSC, the hand-crafted features are extracted for short time frames obtained by windowing the audio signals. To exploit temporal information of consecutive frames of an ALS event, statistical moments (i.e. mean, standard deviation, coefficient of variation, skewness, kurtosis) are further calculated from the distributions of the short-term features [24]. The long-term features are usually necessary for a respiratory-wise and recording-wise classification system. The short time analysis is also the basis for recurrent networks (RNNs) in deep learning-based LSC systems.

9.2.2.2 Time-frequency representations for deep learning

Similar to image and audio signal processing applications, deep learning techniques have their own added characteristics suited for health informatics [84] in general and for lung sound classifications in particular. Deep neural network (DNN) architectures such as CNN and RNN variations implicitly learn a representation which can be considered as automatic feature extraction. In particular, CNNs and various combinations of CNNs are equipped with feature learning layers such as convolutional layers (see Section 9.3.1.2). These layers often play the role of a high-level feature extractor of 2D feature maps (i.e. visual representation of time—frequency maps).

There are different 2D representations such as the spectrogram of short-time Fourier transform (STFT), constant-Q transform (CQT) and Stockwell transform (S-transform) (known for local spectral phase properties) or scalogram of the wavelet transform. The time–frequency representations are able not only to capture both fine-grained temporal and spectral information but also present a much wider time context of the recordings. Mel and gammatone filter banks, which were designed to mimic the human auditory system, are often used for filtering to obtain a mel spectrogram, gammatonegram or MFCCs. Furthermore, filter banks dramatically reduce the number of features of each time frame. In addition to these transformations, empirical mode decomposition (EMD) has been used to decompose audio signals into intrinsic mode functions (IMFs), which are further processed to obtain visual representations.

Often combinations of visual features are exploited in LSC. They can be concatenated to construct a 2D matrix [49,52], but they can also be considered as individual visual representation inputs for multiinput DNNs [51,69] or ensembles of single-input DNNs [32,34,35].

We summarize different time–frequency features of recent LSC system in Table 9.2.

9.2.3 Data augmentation

Lung sound datasets are relatively limited in either the number of patients or recordings. Furthermore, an imbalance between categories is common. These dataset challenges cause an under/over-estimation in performance and a limited generalization ability of the LSC system. To prevent this situation, data augmentation is a popular and efficient solution in state-of-the-art LSC systems to approach the category imbalance and increase the diversity of the training data set. The simplest approaches randomly select a sample and modify it by for example, adding noise. There are two main directions of data augmentation techniques: (1) time domain and (2) time–frequency domain data augmentation.

9.2.3.1 Time domain

The approach generates new training data samples (instances) in time domain. In particular, different signal transformations are performed on audio signals such as

time stretching, pitch or random shifting, variations of volume or speed, etc. In addition, smart padding or a concatenation-based augmentation technique [50] for audio signals has been used.

9.2.3.2 Time–frequency domain

Data augmentation techniques are performed for time–frequency representations. A popular technique used from the field of speech recognition is vocal tract length perturbation. Furthermore, most of the augmentation techniques for images in the field of computer vision are used such as mixup augmentation [85], horizontal/vertical flipping, random cropping or color transformations. In addition, generative adversarial network variants [78,81], which synthesizes novel data instances from spectrograms, have been introduced.

A combination of several augmentation techniques has often been used in recent studies. The most popular data augmentation techniques are summarized in Table 9.3.

Table 9.3 Summary of data augmentation techniques.

Input type	Techniques	References
Audio signals	Pitch shift	[38,50,55,71,77,78]
	Random shifting	[36,42,50,51,53,77]
	Volume adjusting	[36,50,53,71]
	Noise addition	[36,42,50,51,55,71,77]
	Speed variation	[50,51,53,55,71,77]
	Time stretching	[29,36,38,42,71,78]
	Cropping/masking	[25,53,55,71]
	Background noise	[78]
	Dynamic range compression	[38,78]
	Smart padding	[50]
Time–frequency representation	Vocal tract length perturbation	[25,29,36,51]
	Mixup augmentation	[32,33,59,71]
	Horizontal/vertical flipping	[36]
	Random cropping	[25]
	SpecAugment techniques	[78]
	Color mapping for RGB spectrogram image	[43]
	Generative adversarial network variants	[41,78,81]

9.3 Data modeling

This section summarizes common machine learning approaches and learning paradigms used for LSC systems.

9.3.1 Machine learning

There are two main directions: (1) conventional classifiers for hand-crafted features and (2) deep neural networks.

9.3.1.1 Conventional classifiers

The conventional classifiers are usually used in combination with classical features in LSC systems. They are summarized in [Table 9.4](#).

9.3.1.2 Deep learning architectures

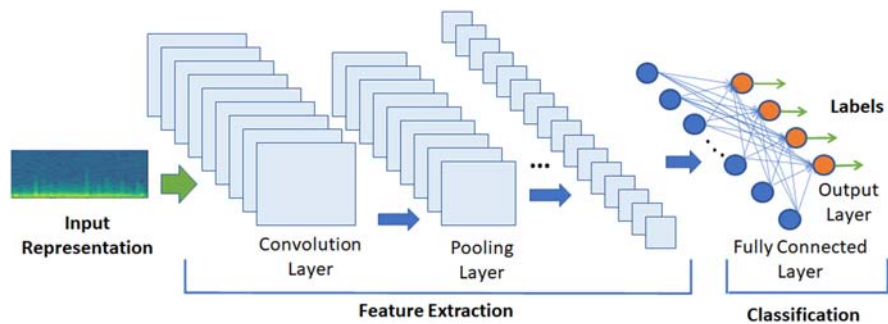
Deep learning architectures often use 2D feature representation. They perform both high-level feature extraction and classification. For LSC several different architectures are used.

9.3.1.2.1 Convolutional neural networks

Similar to acoustic scene classification and acoustic event detection [86], classification tasks for lung sounds and respiratory diseases are mostly approached by 2D CNN-based network architectures. [Fig. 9.3](#) shows a typical LSC system using a CNN architecture. The CNN is able to collect information for longer lung sound

Table 9.4 Summary of conventional classifiers.

Classifier	References
Random forest	[5,24,26,40,61]
SVMs	[5,10,15,19,20,23–26,44,57,59,61,66–68,83]
Multilayer perceptrons	[5,20,21,23,24,60,61,67]
k-NNs	[5,15,20,25,44,58,59,61,62,67]
GMMs	[5,22,25]
HMMs	[5,22]
Decision trees and variations	[15,19,23,26,30,44,58,59]
Bayes rule-based classifiers	[23,58]
Logistic regression	[5,26,59]
Linear discriminant analysis	[5,17,19,44,59,74]
Ensemble classifiers	[63]
Self organized map	[5,61]
K-mean clustering	[61]

**FIGURE 9.3**

A typical convolutional neural networks architecture for lung sound classification.

excerpts such as respiratory cycles containing several ALS events. CNNs are used in two common ways as follows.

1. *Standard CNN architectures:* CNNs with various architectures originally proposed for ImageNet applications are exploited such as VGGs, ResNets, AlexNets, InceptionNets, GoogLeNets or Mobilenets. They are often utilized in transfer learning-based LSC systems.
2. *Custom CNN architectures:* Several self-defined 2D CNNs are proposed for LSC. Furthermore, parallel architectures such as multiinput CNNs [19,51] and parallel-pooling CNN architectures [58] have been introduced to make efficiently use of more information from various feature representations. Beside 2D CNNs, 1D CNNs are used in a few studies [65,87].

9.3.1.2.2 Recurrent networks

Some LSC systems have used RNN variations such as gated recurrent units (GRUs), long short-term memories (LSTMs) or their bidirectional extensions BiGRUs and BiLSTMs. RNNs can account for tasks related to detection of ALS events and classification of ALSs and respiratory diseases.

9.3.1.2.3 Hybrid systems

Hybrid CNN models have been proposed, which combine CNN models with either RNN variants or special mechanisms such as mixture of expert (MoE) layers. In addition, hybrid LSC systems have been developed as combination of conventional LSC and deep learning. For instance, classical features are fed into deep multilayer perceptrons [24,25]. Furthermore, conventional classifiers are used for learned representations from CNN architectures [25,73].

Common deep learning architectures are summarized in Table 9.5.

In addition, some deep learning-based LSC systems are focusing on low computational costs by using or proposing small model architectures such as

Table 9.5 Summary of deep learning architectures.

Type	References		
CNNs	<i>Standard CNN architectures</i>		
	VGGs	[29,37,52,65,72,76,77]	
	AlexNets	[52,53,55,72,76,81]	
	ResNets	[36,55,70,75,81]	
	InceptionNets	[35,55,75]	
	MobileNets	[72,75,77]	
	GoogLeNets	[55,81]	
RNNs	<i>Custom CNN architectures</i>		
	1D CNNs	[65,87]	
	CNNs	[10,25,29,33,34,38,41,43,50,53,56,71,73,76,80,82]	
	Multi-Input (MI)— CNNs	[19,39,51]	
	Noise Mask—RNN variations	[26,78,79]	
	RNNs	[18,26,27,78]	
	GRUs/BiGRUs	[16,27,42,48,52]	
	LSTMs/BiLSTMs	[27,48,52]	
	Hybrid systems	CNN—RNNs	[32]
		CNN—LSTMs/ CNN—Bi LSTMs	[9,48,51,70,87] [28,37,48]
CNN—GRUs/CNN— Bi GRUs		[32,33] [34,79]	
CNN—MoE			
Autoencoder			
Deep multilayer perceptrons		[24,25] [25,73]	
CNN—SVMs			

lightweight CNNs, MobileNets, depth-wise convolutional layers in custom CNN models [52] or weight quantization [77].

9.3.2 Learning paradigm

9.3.2.1 Transfer learning

Transfer learning is an efficient solution for performance improvement in state-of-the-art LSC systems, as these systems can save training time and utilize effectively knowledge from source domains. In particular, it allows the improvement of models through transferring knowledge from all or parts of a CNN model, which is pretrained on a source domain dataset. Generally, there are two popular approaches of transfer learning. (1) The knowledge from the representation part

of the pretrained model is not modified during post-training. This means that the weights are frozen and transfer learning works as feature extractor [51,73]. (2) All or parts of the pretrained model are fine-tuned on the target domain of lung sound data [36,50,73,74].

State-of-the-art LSC studies using transfer learning are summarized in Table 9.6. ResNets are a common architecture, and often ImageNet is used as source domain for model pretraining.

In addition to common transfer learning, cotuning and stochastic normalization have been recently exploited in [36]. This approach makes use of all parts of the pretrained model that is the representation part and output layers, and uses a cotuning paradigm [88] during training.

Beside transfer learning, knowledge distillation has also been used. Particularly, a student—teacher scheme is applied for RDC [33]. A teacher with specific CNN architecture is trained as usual. Afterwards, the teacher’s embedding is distilled to the student to assist in the student’s learning process.

9.3.2.2 Postprocessing

Many LSC systems process data at basic units of either frames or segments in all analysis modes illustrated in Fig. 9.4. Thus, it is necessary to perform a postprocessing step to fuse results of each frame/segment for the event, respiratory cycle or full recording. Popular methods for LSC systems are majority voting [16,35] or averaging of prediction results [32].

Some state-of-the-art LSC systems use multiple models as ensembles. The prediction outputs of each classifier can be combined to enhance accuracy [29,32–35,44,58,74]. A snapshot ensemble [29] builds an ensemble of multiple models at moderate additional training cost. It is an effective solution to deal with training multiple models. However, ensembles are often infeasible in practical applications due to their high computational requirements.

Table 9.6 Summary of transfer learning for lung sound classification with both the feature extraction and fine-tuning approach.

CNN architectures	Source domain	References
ResNets	ImageNet ICBHI 2017	[36,50,75,76] [36]
VGGs	ImageNets Audio Set	[65,76] [37]
Inception-V3	ImageNets	[75]
MobileNet-V2	ImageNets	[75,77]
CNNs	ICBHI 2017	[51]

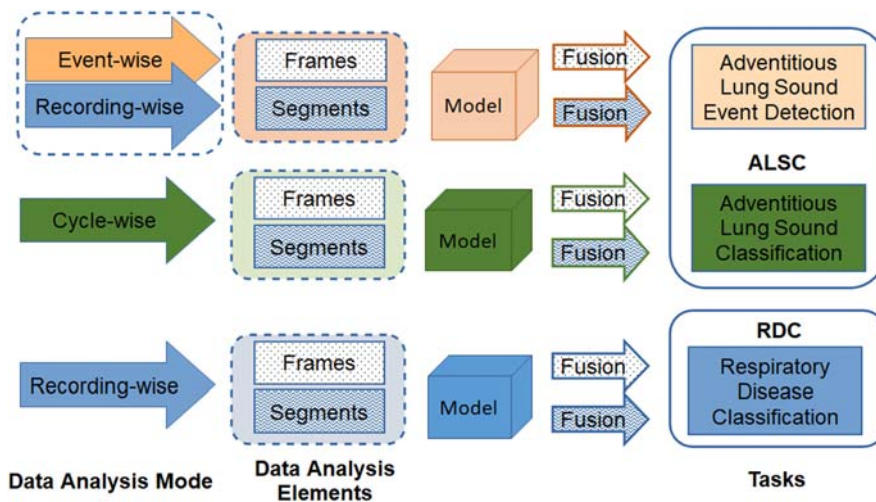


FIGURE 9.4

Block diagram of data processing for different lung sound classification tasks and corresponding data analysis modes.

9.4 Recent public lung sound datasets

Thirteen publicly available lung sound datasets were listed in [8]. There are three prominent datasets of lung sounds, which have been mainly used in recent research. They are described in the following section.

9.4.1 ICBHI 2017 dataset

The ICBHI Scientific Challenge dataset (ICBHI 2017) [8] is freely available for research. The dataset consists of 920 annotated audio samples from 126 subjects corresponding to patient pathological conditions that is healthy and seven distinct disease categories (Pneumonia, Bronchiectasis, COPD, upper respiratory tract infection, lower respiratory tract infection, Bronchiolitis, Asthma). The audio samples were recorded using different stethoscopes that is AKGC417L, Meditron, Litt3200 and LittC2SE. The recording duration ranges from 10 to 90 seconds and the sampling rate ranges from 4000 to 44100 Hz. Each recording is composed of a certain number of breathing cycles with corresponding annotations of the beginning and the end, and the presence/absence of crackles and/or wheezes. The annotations of the database support to split audio recordings into respiratory cycles. The cycle duration ranges from 0.2 to 16 seconds and the average cycle duration is 2.7 seconds. The database includes 6898 different respiratory cycles with 3642 normal cycles, 1864 crackles, 886 wheezes, and 506 cycles containing of both crackles and wheezes.

9.4.2 The Abdullah University Hospital 2020 dataset

The dataset introduced by the King Abdullah University Hospital in 2020 contains lung sounds acquired by an electronic stethoscope placed on various regions of the chest wall [48]. The recording was performed using the 3MLittmann Electronic Stethoscope model 3200. The dataset includes respiratory sounds from 120 subjects (35 healthy and 77 patients). It contains one recording per subject. The duration of each recording ranges from 5 to 30 seconds, which is sufficient to cover at least one respiratory cycle. Each recording was replicated three times corresponding to various frequency filters that emphasize certain body sounds. The dataset consisted of a total of 308 lung sound recordings, each is of 5 seconds duration. The dataset can be used to detect seven pulmonary diseases from lung sounds that is asthma, heart failure, pneumonia, bronchitis, pleural effusion, lung fibrosis and COPD as well as normal breathing sounds. Furthermore, it can be used to identify the correct type of lung sound such as normal, crepitations, wheezes, crackles, bronchial, wheezes and crackles or bronchial and crackles.

9.4.3 HF_Lung_V1 dataset

HF_Lung_V1 is an open access lung sound database used for developing automated inhalation, exhalation and adventitious sound detection algorithms working at event or recording level. It is the largest lung sound dataset introduced in 2021. The dataset comprises 9765 audio files of lung sounds (duration of 15 seconds each), 34,095 inhalation labels, 18,349 exhalation labels, 13,883 CAS labels (8457 wheeze labels, 686 stridor labels, and 4740 rhonchi labels), and 15,606 DAS labels (all crackles). The dataset includes the start and end time of all the events such as inhalation, exhalation, wheeze (, stridor, rhonchus, and DAS events in the recordings). It has to be noted that the labels in a single label file is made only by one annotator. The anotators were trained to have good agreement, but the labeling is still not perfect [49].

9.5 Conclusion

There are recent advances in deep learning, which help to improve performance in state-of-the-art LSC as follows. (1) Robust deep neural network architectures are commonly exploited to automatically extract high-level features. (2) Transfer learning is a promising technique in accounting for limitations of data quantity in lung sound datasets through exploiting knowledge from datasets in the same or different fields. (3) Data augmentation techniques are increasingly applied in LSC to increase diversity and number of data as well as balancing the classes of the lung sound dataset. It helps in dealing with overfitting of DNN architectures.

This chapter presents a comprehensive review on recent studies for computational LSC. There are two important classification tasks: the classification of adventitious lung sounds and respiratory diseases. We summarized and categorized as a structural review the topics from data processing such as audio signal processing, feature extraction, and data augmentation to data modeling such as neural network architectures and learning paradigms. Besides the recent advances, one main challenge arises when LSC models are deployed in real-world diagnosis systems. The clinical setting introduces many challenges such as noise, bad signal quality, and artifacts. Deep neural networks combined with transfer learning and data augmentation have been promising avenues for these challenges.

References

- [1] Forum of International Respiratory Societies, European Respiratory Society. The global impact of respiratory disease, 2017.
- [2] WHO <<https://covid19.who.int/>> (accessed 11.05.21).
- [3] M.T. Barbosa, M. Morais-Almeida, C.S. Sousa, J. Bousquet, The “big five” lung diseases in covid-19 pandemic—a google trends analysis, *Pulmonology* 27 (1) (2021) 71.
- [4] N. Chen, M. Zhou, X. Dong, J. Qu, F. Gong, Y. Han, et al., Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in wuhan, china: a descriptive study, *Lancet* 395 (10223) (2020) 507–513.
- [5] R.X.A. Pramono, S. Bowyer, E. Rodriguez-Villegas, Automatic adventitious respiratory sound analysis: a systematic review, *PLoS One* 12 (5) (2017). Available from: <https://doi.org/10.1371/journal.pone.0177926>.
- [6] H. Pasterkamp, S.S. Kraman, G.R. Wodicka, Respiratory sounds advances beyond the stethoscope, *Am. J. Respir. Crit. Care Med.* 156 (3) (1997) 974–987. Available from: <https://doi.org/10.1164/ajrccm.156.3.9701115>.
- [7] M. Sarkar, I. Madabhavi, N. Niranjana, M. Dogra, Auscultation of the respiratory system, *Ann Thorac. Med.* 10 (3) (2015) 158.
- [8] B.M. Rocha, D. Filos, L. Mendes, I. Vogiatzis, E. Perantoni, E. Kaimakamis, et al., A respiratory sound database for the development of automated classification, *Precision Medicine Powered by pHealth and Connected Health*, Springer, 2018, pp. 33–37.
- [9] C. Jácome, J.C. Aviles-Solis, A.M. Uhre, H. Pasterkamp, H. Melbye, Adventitious and normal lung sounds in the general population: comparison of standardized and spontaneous breathing, *Respir. Care*, 63 (11) (2018) 1379–1387. Available from: <https://doi.org/10.4187/respcare.06121>.
- [10] M. Aykanat, O. Kılıc, B. Kurt, S. Saryal, Classification of lung sounds using convolutional neural networks, *EURASIP J. Image Video Process.* 2017 (1) (2017) 65.
- [11] A. Gurung, C.G. Scraftford, J.M. Tielsch, O.S. Levine, W. Checkley, Computerized lung sound analysis as diagnostic aid for the detection of abnormal lung sounds: a systematic review and meta-analysis, *Respir. Med.* 105 (9) (2011) 1396–1403.
- [12] L. Deng, D. Yu, Deep learning: methods and applications, *Found. Trends Signal Process.* 7 (3) (2014) 197–387. Available from: <https://doi.org/10.1561/20000000039>.

- [13] R.O. Duda, P.E. Hart, D.G. Stork, *Pattern Classification and Scene Analysis*, second ed., Wiley Interscience, 1995.
- [14] R.X.A. Pramono, S.A. Imtiaz, E. Rodriguez-Villegas, Evaluation of features for classification of wheezes and normal respiratory sounds, *PLoS One* 14 (3) (2019) e0213659. Available from: <https://doi.org/10.1371/journal.pone.0213659>. PMID: 30861052. PMCID: PMC6414007.
- [15] M. Grønnesby, J.C.A. Solis, E. Holsbø, H. Melbye, L.A. Bongo, Feature extraction for machine learning based crackle detection in lung sounds from a health survey, *arXiv:1706.00005*, 2017.
- [16] E. Messner, M. Fediuk, P. Swatek, S. Scheidl, F. Smolle-Jüttner, H. Olschewski, et al., Crackle and breathing phase detection in lung sounds with deep bidirectional gated recurrent neural networks, in: *Proceedings of EMBC, IEEE*, 2018, pp. 356–359.
- [17] S.H. Li, B.S. Lin, C.H. Tsai, C.T. Yang, B.S. Lin, Design of wearable breathing sound monitoring system for real-time wheeze detection, *Sensors* (2017) 171.
- [18] H. Chen, X. Yuan, Z. Pei, M. Li, J. Li, Triple-classification of respiratory sounds using optimized s-transform and deep residual networks, *IEEE Access* 7 (2019) 32845–32852.
- [19] B.M. Rocha, D. Pessoa, A. Marques, P. Carvalho, R.P. Paiva, Automatic classification of adventitious respiratory sounds: a (un) solved problem, *Sensors* 21 (2021) 57. Available from: <https://doi.org/10.3390/s21010057>.
- [20] N. Sengupta, M. Sahidullah, G. Saha, Lung sound classification using local binary pattern, *arXiv:1710.01703*, 2017.
- [21] Y. X. Liu, Y. Yang, Y. H. Chen, Lung sound classification based on hilbert-huang transform features and multilayer perceptron network, in: *Proceedings of the Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), IEEE*, 2017, pp. 765–768.
- [22] N. Jakovljevic, T. Loncar-Turukalo, Hidden markov model based respiratory sound classification, in: *Proceedings of the International Conference on Biomedical and Health Informatics, Springer*, 2017, pp. 39–43.
- [23] H. Mukherjee, P. Sreerama, A. Dhar, et al., Automatic lung health screening using respiratory sounds, *J. Med. Syst.* 45 (19) (2021). Available from: <https://doi.org/10.1007/s10916-020-01681-9>.
- [24] A. Monaco, N. Amoroso, L. Bellantuono, E. Pantaleo, S. Tangaro, R. Bellotti, Multi-time-scale features for accurate respiratory sound classification, *Appl. Sci.* 10 (23) (2020) 8606.
- [25] D. Bardou, K. Zhang, S.M. Ahmad, Lung sounds classification using convolutional neural networks, *Artif. Intell. Med.* 88 (2018) 58–69.
- [26] K. Kochetov, E. Putin, M. Balashov, A. Filchenkov, A. Shalyto, Noise masking recurrent neural network for respiratory sound classification, in: *Proceedings of the International Conference on Artificial Neural Networks. Springer*, 2018, pp. 208–217.
- [27] D. Perna, A. Tagarelli, Deep auscultation: predicting respiratory anomalies and diseases via recurrent neural networks, in: *Proceedings of the IEEE Thirty-second International Symposium on Computer-Based Medical Systems (CBMS), IEEE*, 2019, pp. 50–55.
- [28] E. Messner, M. Fediuk, P. Swatek, S. Scheidl, F.M. Smolle-Jüttner, H. Olschewski, et al., Multi-channel lung sound classification with convolutional recurrent neural networks, *Comput. Biol. Med.* 122 (2020) 103831. Available from: <https://doi.org/10.1016/j.compbiomed.2020.103831>. ISSN 0010–4825.

- [29] T. Nguyen, F. Pernkopf, Lung sound classification using snapshot ensemble of convolutional neural networks, in: Proceedings of the Forty-Second Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), IEEE, 2020, pp. 760–763.
- [30] G. Chambres, P. Hanna, M. Desainte-Catherine, Automatic detection of patient with respiratory diseases using lung sound analysis, in: Proceedings of the International Conference on Content-Based Multimedia Indexing (CBMI), IEEE, 2018, pp. 1–6.
- [31] X.H. Kok, S.A.S. Imtiaz, E. Rodriguez-Villegas, A novel method for automatic identification of respiratory disease from acoustic recordings, in: Proceedings of the Forty-first Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2019, pp. 2589–2592.
- [32] L. Pham, I. McLoughlin, H. Phan, M. Tran, T. Nguyen, R. Palaniappan, Robust deep learning framework for predicting respiratory anomalies and diseases, in: Proceedings of the Forty-Second Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE, 2020, pp. 164–167.
- [33] L.D. Pham, H. Phan, R. Palaniappan, A. Mertins, I. McLoughlin, Cnn-moe based framework for classification of respiratory anomalies and lung disease detection, *IEEE J. Biomed. Health Inform.* (2021).
- [34] D. Ngo, L. Pham, A. Nguyen, B. Phan, K. Tran, T. Nguyen, Deep learning framework applied for predicting anomaly of respiratory sounds, in: Proceedings of the International Symposium on Electrical and Electronics Engineering (ISEE), IEEE, 2021.
- [35] L.D. Pham, H. Phan, R. King, A. Mertins, I. McLoughlin, Inception-based network and multi-spectrogram ensemble applied for predicting respiratory anomalies and lung diseases, in: Proceedings of the Forty-Third Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), IEEE, 2021.
- [36] T. Nguyen, F. Pernkopf, Lung sound classification using co-tuning and stochastic normalization, *IEEE Transaction on Biomedical Engineering* (2022). Available from: <https://doi.org/10.1109/TBME.2022.3156293>.
- [37] L. Shi, K. Du, C. Zhang, H. Ma, W. Yan, Lung sound recognition algorithm based on VGGish-BiGRU, *IEEE Access* 7 (2019) 139438–139449. Available from: <https://doi.org/10.1109/ACCESS.2019.2943492>.
- [38] Z. Tariq, S.K. Shah, Y. Lee, Lung disease classification using deep convolutional neural network, in: Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2019, pp. 732–735. Available from: <https://doi.org/10.1109/BIBM47256.2019.8983071>.
- [39] Z. Tariq, S. K. Shah, Y. Lee, Multimodal lung disease classification using deep convolutional neural network, in: Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2020, pp. 2530–2537. Available from: <https://doi.org/10.1109/BIBM49941.2020.9313208>.
- [40] L. Wu, L. Li, Investigating into segmentation methods for diagnosis of respiratory diseases using adventitious respiratory sounds, in: Proceedings of the Forty-Second Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), 2020, pp. 768–771. Available from: <https://doi.org/10.1109/EMBC44109.2020.9175783>.
- [41] M. García-Ordás, J. Benítez Andrades, I. García, C. Benavides, H. Alaiz Moreton, Detecting respiratory pathologies using convolutional neural networks and variational autoencoders for unbalancing data, *Sensors* 20 (2020).

- [42] V. Basu, S. Rana, Respiratory diseases recognition through respiratory sound with the help of deep neural network, in: Proceedings of the Fourth International Conference on Computational Intelligence and Networks (CINE), 2020, pp. 1–6.
- [43] S.B. Shuvo, S.N. Ali, S.I. Swapnil, T. Hasan, M.I.H. Bhuiyan, A lightweight CNN model for detecting respiratory diseases from lung auscultation sounds using EMD-CWT-based hybrid scalogram, *IEEE J. Biomed. Health Inform.* Available from: <https://doi.org/10.1109/JBHI.2020.3048006>.
- [44] L. Fraiwan, O. Hassanin, M. Fraiwan, B. Khassawneh, A.M. Ibnian, M. Alkhodari, Automatic identification of respiratory diseases from stethoscopic lung sound signals using ensemble classifiers, *Biocybern. Biomed. Eng.* 41 (1) (2021) 1–14. Available from: <https://doi.org/10.1016/j.bbe.2020.11.003>.
- [45] J.C. Aviles-Solis, S. Vanbelle, P.A. Halvorsen, N. Francis, J.W.L. Cals, E.A. Andreeva, et al., International perception of lung sounds: a comparison of classification across some European borders, *BMJ Open Respir Res.* 4 (1) (2017) e000250. Available from: <https://doi.org/10.1136/bmjresp-2017-000250>. PMID: 29435344. PMCID: PMC5759712.
- [46] E. Messner, M. Hagmueller, P. Swatek, F. Pernkopf, A robust multichannel lung sound recording device, *Biodevices* (2016) 34–39.
- [47] Dataset, RALE: a computer-assisted instructional package, *Respir. Care* 35 (1990) 1006.
- [48] M. Fraiwan, L. Fraiwan, B. Khassawneh, A. Ibnian, A dataset of lung sounds recorded from the chest wall using an electronic stethoscope, *Data Brief* 35 (2021) 106913.
- [49] F.S. Hsu, S.R. Huang, C.W. Huang, C.J. Huang, Y.R. Cheng, C.C. Chen, et al., Benchmarking of eight recurrent neural network variants for breath phase and adventitious sound detection on a self-developed open-access lung sound database—HF_Lung_V1, *PLoS One* 16 (7) (2021) e0254134. Available from: <https://doi.org/10.1371/journal.pone.0254134>.
- [50] S. Gairola, F. Tom, N. Kwatra, M. Jain, Respirenet: a deep neural network for accurately detecting abnormal lung sounds in limited data setting, in: Proceedings of Forty-Third Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), IEEE (2021).
- [51] T. Nguyen, F. Pernkopf, Crackle detection in lung sounds using transfer learning and multi-input convolutional neural networks, in: Proceedings of the Forty-Third Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), IEEE, 2021.
- [52] S.Y. Jung, C.H. Liao, Y.S. Wu, S.M. Yuan, C.T. Sun, Efficiently classifying lung sounds through depth-wise separable CNN models with fused STFT and MFCC features, *Diagnostics* 11 (2021) 732. Available from: <https://doi.org/10.3390/diagnostics11040732>.
- [53] A. Srivastava, S. Jain, R. Miranda, S. Patil, S. Pandya, K. Kotecha, Deep learning based respiratory sound analysis for detection of chronic obstructive pulmonary disease, *PeerJ Comput. Sci.* 7 (2021) e369. Available from: <https://doi.org/10.7717/peerj-cs.369>. PMID: 33817019. PMCID: PMC7959628.
- [54] V.I. Quandt, E.R. Pacola, S.F. Pichorim, H.R. Gamba, M.A. Sovierzoski, Pulmonary crackle characterization: approaches in the use of discrete wavelet transform regarding border effect, mother-wavelet selection, and subband reduction, *Res. Biomed. Eng.* 31 (2) (2015) 148–159. Available from: <https://doi.org/10.1590/2446-4740.0639>.
- [55] S. Gupta, M. Agrawal, D. Deepak, Gammatonegram based triple classification of lung sounds using deep convolutional neural network with transfer learning, *Biomed.*

- Signal Process. Control 70 (2021) 102947. Available from: <https://doi.org/10.1016/j.bspc.2021.102947>.
- [56] R. Hazra, S. Majhi, Detecting respiratory diseases from recorded lung sounds by 2D CNN, in: Proceedings of the Fifth International Conference on Computing, Communication and Security (ICCCS), 2020, pp. 1–6. Available from: <https://doi.org/10.1109/ICCCS49678.2020.9277101>.
- [57] S. Z. H. Naqvi, M. Arooj, S. Aziz, M. U. Khan, M. A. Choudhary and M. N. ul Hassan, Spectral analysis of lungs sounds for classification of asthma and pneumonia wheezing, in: Proceedings of the International Conference on Electrical, Communication, and Computer Engineering (ICECCE), 2020, pp. 1–6. Available from: <https://doi.org/10.1109/ICECCE49384.2020.9179417>.
- [58] S. Ulukaya, G. Serbes, Y.P. Kahya, Overcomplete discrete wavelet transform based respiratory sound discrimination with feature and decision level fusion, Biomed. Signal Process. Control 38 (2017) 322–336. Available from: <https://doi.org/10.1016/j.bspc.2017.06.018>. ISSN 1746–8094.
- [59] N.S. Haider, B.K. Singh, R. Periyasamy, et al., Respiratory sound based classification of chronic obstructive pulmonary disease: a risk stratification approach in machine learning paradigm, J. Med. Syst. 43 (2019) 255. Available from: <https://doi.org/10.1007/s10916-019-1388-0>.
- [60] S. Hadiyoso, A. Rizal, Empirical mode decomposition and grey level difference for lung sound classification, Trait. du Signal 38 (1) (2021) 175–179. Available from: <https://doi.org/10.18280/ts.380118>.
- [61] A. Elsetrønning, A. Rasheed, J. Bekker, O. San, On the effectiveness of signal decomposition, feature extraction and selection on lung sound classification, arXiv:2012.11759, 2020.
- [62] S.Z.H. Naqvi, M.A. Choudhary, Z. Tariq, A. Waseem, Automated detection and classification of multichannel lungs signals using EMD, in: Proceedings of the International Conference on Electrical, Communication, and Computer Engineering (ICECCE), 2020, pp. 1–6. Available from: <https://doi.org/10.1109/ICECCE49384.2020.9179244>.
- [63] S.I. Khan, R.B. Pachori, Automated classification of lung sound signals based on empirical mode decomposition, Expert Syst. Appl. 184 (2021) 115456. Available from: <https://doi.org/10.1016/j.eswa.2021.115456>. ISSN 0957–4174.
- [64] G. Altan, Y. Kutlu, N. Allahverdi, Deep learning on computerized analysis of chronic obstructive pulmonary disease, IEEE J. Biomed. Health Inform. 24 (5) (2020) 1344–1350. Available from: <https://doi.org/10.1109/JBHI.2019.2931395>.
- [65] N. Baghel, V. Nangia, M.K. Dutta, ALSD-Net: automatic lung sounds diagnosis network from pulmonary signals, Neural Comput. Appl. (2021). Available from: <https://doi.org/10.1007/s00521-021-06302-1>.
- [66] A. H. Falah, J. Jondri, Lung sounds classification using stacked autoencoder and support vector machine, in: Proceedings of the Seventh International Conference on Information and Communication Technology (ICoICT), 2019, pp. 1–5, doi: 10.1109/IcoICT.2019.8835278.
- [67] F. Meng, Y. Shi, N. Wang, M. Cai, Z. Luo, Detection of respiratory sounds based on wavelet coefficients and machine learning, EEE Access 8 (2020) 155710–155720. Available from: <https://doi.org/10.1109/ACCESS.2020.3016748>.
- [68] A. Yadav, M. K. Dutta, J. Prinosil, Machine learning based automatic classification of respiratory signals using wavelet transform, in: Proceedings of the Forty-Third

- International Conference on Telecommunications and Signal Processing (TSP), 2020, pp. 545–549, doi: 10.1109/TSP49548.2020.9163565.
- [69] K. Minami, H. Lu, H. Kim, S. Mabu, Y. Hirano, S. Kido, Automatic classification of large-scale respiratory sound dataset based on convolutional neural network, in: Proceedings of the Nineteenth International Conference on Control, Automation and Systems (ICCAS), 2019, pp. 804–807. Available from: <https://doi.org/10.23919/ICCAS47443.2019.8971689>.
- [70] C.H. Hsiao, et al., Breathing sound segmentation and detection using transfer learning techniques on an attention-based encoder-decoder architecture, IEEE Engineering in Medicine & Biology Society (EMBC) 2020, 754–759. Available from: <https://doi.org/10.1109/EMBC44109.2020.9176226>.
- [71] Y. Ma, X. Xu, Y. Li, Lungrn + nl: an improved adventitious lung sound classification using non-local block resnet neural network with mixup data augmentation, in: Proceedings of the Interspeech, 2020, pp. 2902–2906.
- [72] Y.S. Wu, C.H. Liao, S.M. Yuan, Automatic auscultation classification of abnormal lung sounds in critical patients through deep learning models, in: Proceedings of the Third IEEE International Conference on Knowledge Innovation and Invention (ICKII), 2020, pp. 9–11. Available from: <https://doi.org/10.1109/ICKII50300.2020.9318880>.
- [73] F. Demir, A. Sengur, V. Bajaj, Convolutional neural networks based efficient approach for classification of lung diseases, *Health Inform. Sci. Syst.* 8 (1) (2020) 1–8.
- [74] F. Demir, A.M. Ismael, A. Sengur, Classification of lung sounds with cnn model using parallel pooling structure, *IEEE Access* 8 (2020) 105376–105383.
- [75] V. Vaityshyn, H. Porieva, A. Makarenkova, Pre-trained convolutional neural networks for the lung sounds classification, in: Proceedings of the IEEE Thirty-Ninth International Conference on Electronics and Nanotechnology (ELNANO), 2019, pp. 522–525. Available from: <https://doi.org/10.1109/ELNANO.2019.8783850>.
- [76] R. Shethwala, S. Pathar, T. Patel, P. Barot, Transfer learning aided classification of lung sounds-wheezes and crackles, in: Proceedings of the Fifth International Conference on Computing Methodologies and Communication (ICCMC), 2021, pp. 1260–1266. Available from: <https://doi.org/10.1109/ICCMC51019.2021.9418310>.
- [77] J. Acharya, A. Basu, Deep neural network for respiratory sound classification in wearable devices enabled by patient specific model tuning, *IEEE Trans. Biomed. Circuits Syst.* 14 (3) (2020) 535–544.
- [78] K. Kochetov, A. Filchenkov, Generative adversarial networks for respiratory sound augmentation, in: Proceedings of the International Conference on Control, Robotics and Intelligent System, 27–29 October 2020, Xiamen, China. ACM, New York, 6 pp. <<https://doi.org/10.1145/3437802.3437821>>
- [79] A. Manzoor, Q. Pan, H.J. Khan, S. Siddeeq, H.M.A. Bhatti, M.A. Wedagu, Analysis and detection of lung sounds anomalies based on NMA-RNN, in: Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2020, pp. 2498–2504. Available from: <https://doi.org/10.1109/BIBM49941.2020.9313197>.
- [80] S. Jayalakshmy, B. L. Priya, N. Kavya, CNN based Categorization of respiratory sounds using spectral descriptors, in: Proceedings of the International Conference on Communication, Computing and Industry 4.0 (C2I4), 2020, pp. 1–5. Available from: <https://doi.org/10.1109/C2I451079.2020.9368933>.
- [81] S. Jayalakshmy, L. Priya, G.F. Sudha, Chapter 7—Synthesis of respiratory signals using conditional generative adversarial networks from scalogram representation,

- Generative Adversarial Networks for Image-to-Image Translation, first ed. Elsevier Top of Form, 2021, pp. 161–183.
- [82] Y. Ma, X. Xu, Q. Yu, Y. Zhang, Y. Li, J. Zhao, et al., Lungbrn: a smart digital stethoscope for detecting respiratory disease using biresnet deep learning algorithm, in: Proceedings of the IEEE Biomedical Circuits and Systems Conference (BioCAS), IEEE, 2019, pp. 1–4.
 - [83] S. Jayalakshmy, S. Rithika, S. Rajasri, Average power based classification of respiratory sounds using SVM classifier, in: Proceedings of the IEEE International Conference on System, Computation, Automation and Networking (ICSCAN), 2019, pp. 1–5, doi: 10.1109/ICSCAN.2019.8878752.
 - [84] T.M. Navamani, Chapter 7—Efficient deep learning approaches for health informatics, in: A.K. Sangaiah (Ed.), Deep Learning and Parallel Computing Environment for Bioengineering Systems, Academic Press, 2019, pp. 123–137. ISBN 9780128167182. Available from: <https://doi.org/10.1016/B978-0-12-816718-2.00014-2>.
 - [85] H. Zhang, M. Cisse, Y. N. Dauphin, D. Lopez-Paz, Mixup: beyond empirical risk minimization, ICLR, 2018.
 - [86] J. Abeßer, A review of deep learning based methods for acoustic scene classification, Appl. Sci. 10 (6) (2020) 2020. Available from: <https://doi.org/10.3390/app10062020>.
 - [87] M. Fraiwan, L. Fraiwan, M. Alkhodari, et al., Recognition of pulmonary diseases from lung sounds using convolutional neural networks and long short-term memory, J. Ambient Intell. Human Comput. (2021). Available from: <https://doi.org/10.1007/s12652-021-03184-y>.
 - [88] K. You, Z. Kou, M. Long, J. Wang, Co-tuning for transfer learning, Adv. Neural Inf. Process. Syst. 33 (2020).

This page intentionally left blank

Clinical applications of machine learning in heart failure

10

Xinmu Li^{1,*}, Sharen Lee^{2,*}, George Bazoukis³, Gary Tse^{1,4,5} and Tong Liu¹

¹*Tianjin Key Laboratory of Ionic-Molecular Function of Cardiovascular Disease, Department of Cardiology, Tianjin Institute of Cardiology, Second Hospital of Tianjin Medical University, Tianjin, P.R. China*

²*Cardiovascular Analytics Group, Hong Kong, China-UK Collaboration*

³*University of Nicosia, Medical School, Nicosia, Cyprus*

⁴*Kent and Medway Medical School, Canterbury, United Kingdom*

⁵*School of Nursing and Health Studies, Hong Kong Metropolitan University, Hong Kong, P.R. China*

10.1 Introduction

Heart failure (HF) is a heterogeneous clinical syndrome caused by structural or functional abnormalities leading to a reduction in cardiac output that is inadequate to meet the needs of peripheral tissues. Currently, there are an estimated 64.3 million HF patients worldwide, with a disease prevalence of 1%–2% in the general adult population of developed countries [1]. HF classification is most often based on the left ventricular ejection fraction (LVEF), traditionally divided into HF with reduced ejection fraction (HFrEF) and HF with preserved ejection fraction (HFpEF) [2]. Over the recent years, the concept of HF with mid-range ejection fraction (HFmrEF) has also been introduced [3], an intermediate category with its own phenotype [4]. Furthermore, HF can be stratified into acute and chronic based on disease onset. Chronic HF has a progressively deteriorating clinical course, with episodes of acute decompensation that results in acute respiratory distress that commonly requires hospitalization. As a result, the repeated admission and the irreversible progression lead to a heavy burden on the healthcare system and a leading global cause of death.

Early diagnosis is of critical importance in managing chronic HF to prolong patient survival and improve disease prognosis. While history and physical examination are essential to identify fluid overload and respiratory distress, these signs and symptoms are largely non-specific. Thus, investigations are important to confirm the diagnosis, evaluate the severity and determine the underlying cause. Typical investigations include electrocardiogram (ECG), chest X-ray, biomarkers such as pro-brain natriuretic peptide (pro-BNP), and echocardiogram.

* Joint first authors.

Over the recent years, the introduction of artificial intelligence into medicine has been revolutionizing the diagnosis and management of diseases. Artificial intelligence uses “big data,” a high volume and high variety set of data, to derive previously inaccessible insights through novel processing methods [5]. In the case of HF, the use of artificial intelligence in the analysis of investigation results can yield further prognostic implications [6]. Neural network analysis, a form of machine learning, is particularly useful in cardiovascular medicine by capturing electrophysiological and hemodynamic findings in different investigations, hence enabling the identification of novel phenotypes and the development of triage algorithms [7]. However, drawbacks such as difficulty to interpret given its “black box” nature and its lack of standardization should be noted. In HF, artificial neural network analysis is instrumental given its multifactorial disease nature in disease development and progression. In this chapter, the use of artificial intelligence and neural network in the diagnosis, management and prevention of HF is discussed in detail.

10.2 Diagnosis

10.2.1 Automatic diagnosis, classification, and phenotyping of heart failure

For the characterization of HF, machine learning-driven methodologies can be broadly divided into diagnosis, identification, and classification/phenotyping. First, diagnosis involves distinguishing patients with HF from those without HF. A neural network and fuzzy logic-based technique successfully distinguished both groups based on data from laboratory investigation results, patient history, physical examination, electrocardiography, vectorcardiography, echocardiography, and therapies [8,9]. Neural network analysis allows for patient classification based on these parameters both alone and in combination. For example, a recent study demonstrated that the deep learning approach combining a support vector machine (SVM)/k-nearest neighbor (k-NN) algorithm with long short time memory enables the differentiation between normal sinus rhythm, arrhythmic and congestive HF ECG signals with an accuracy up to 96.8% [10]. Similarly, a study analyzing over 100,000 frontal chest X rays from over 46,000 patients was able to achieve an area-under-the-receiver-operator-characteristic-curve (AUC) of 0.82 in the diagnosis of HF simply based on the radiological findings. Besides improving the interpretability of the model, the researchers developed Generative Visual Rationales, a visual output that illustrates the classifying features, to allow for the differentiation between correctly trained, overfitted and under fitted models [11].

Another study based on Bayesian classifier and linear discriminant analysis using heart rate variability (HRV) measures distinguished these patients with a sensitivity and specificity of 81.8% and 98.1% [12]. Similarly using HRV data, other techniques based on non-equilibrium decision-tree-based SVM [13] and

sparse auto-encoder-based deep learning algorithm [14] were utilized, achieving accuracies of 89.8% and 72.4%, respectively. Using regularized logistic regression, neural network models, k-NN, SVM, multiple perceptrons (MLP), AUCs between 0.73 and 0.78 were achieved [15]. Numerous other studies have also used HRV-based datasets for diagnosis [16–21]. More recent work has explored the use of left ventricular long-axis myocardial velocity patterns to distinguish HFpEF from healthy controls in their initial analyses using unsupervised machine learning, which was independently evaluated using hypertensives and breathless subjects [22]. Also based on myocardial velocity traces, the same group of authors used multiple kernel learning to differentiate cardiac and non-cardiac causes of breathlessness [23]. Based on spatiotemporal variation in left ventricular deformation, another study used principal-component analysis and distance-weighted k-NN to identify subjects with HFpEF with 81% accuracy. More recent work has used additional data modalities such as urinary proteomics to further improve discrimination [24,25]. A study investigating the level of agreement concerning the HF diagnosis to identify HFrEF, HFmrEF and HFpEF, between HF specialists and AI-Clinical Decision Support System (AI-CDSS). This was prospectively tested for its diagnostic performance of AI-CDSS in consecutive patients presenting with dyspnea to the outpatient clinic [26].

Second, additional efforts have been focused on using machine learning techniques in the identification of HF patients from a pool of hospitalized patients or identification of patients with similar characteristics. For example, automated identification of patients was made possible from terms such as “chf, hf, nyha, failure, congestive, and lasix” from electronic health records [27]. Algorithms based on machine-learning approach using unstructured notes, algorithm 5: a machine-learning approach using structured and unstructured data significantly outperformed a logistic regression-based model [27]. The same group subsequently proposed automated identification of HF hospitalization was achieved based on demographics, laboratory results, vital signs, problem-list diagnoses, and HF-related medications, using similar machine learning techniques [28]. Based on the entire vocabulary of covariates in all medical notes (random non-HF notes vs random HF notes), natural language processing was compared to predictive modeling [29]. The natural language processing achieved better sensitivity (100% vs 56%) but a lower positive predictive value (38% vs 82%), with comparable specificity (98% vs 96%).

The application of neural network analysis further improves the model predictive performance through accounting for temporal relations between events. A classical example of such a model is the REverse Time AttentIoN model (RETAIN), a recurrent neural network (RNN) model that can predict the onset of a specific disease based on medical and clinical events over a series of medical visits. The transferability of the RETAIN model to the prediction of HF has been tested on a large cohort of over 400 hospitals, where an AUC of 82% has been achieved [30]. Another observational study based on electronic health records has applied a RNN model to identify incident HF cases from more than 30,000

patients. The AUC for the RNN model is significantly higher than the best performing baseline machine-learning model (0.883 vs 0.834) under an 18-month observation window [31].

Third, for classification, advances in machine learning techniques extend from simplistic categorization based on LVEF alone for the utilization of several data modalities that include demographics, clinical examination, laboratory exams, medical history, electrocardiographic data, echocardiographic data, and HRV [32–35]. The incorporation of neural networks encourages the inclusion of imaging findings in the classification of HF. For example, a study achieved an overall accuracy of 84% in the differentiation of HF and non-HF patients, with around 80% accuracy in the classification of HF New York Heart Association severity through the use of a neural network classifier on creatine kinase metabolic parameters from 31 phosphorus cardiovascular magnetic resonance spectrometry [36]. Another study was able to differentiate between types of cardiomyopathy among HF patients based on an artificial neural network-based model on speckle-tracking echocardiography, achieving a sensitivity of 89% and specificity of 76% [37].

10.2.2 Detection of heart failure-associated arrhythmia

Besides the diagnosis of HF, the early detection of HF-associated arrhythmia is critical for reducing acute decompensations. Electrophysiological alterations are common under the cardiac structural remodeling and functional deterioration in HF. By applying artificial analysis to ECG, the diagnostic and prognostic accuracy can be raised by reducing human error and the identification of subtle electrophysiological abnormalities.

Atrial fibrillation (AF) is the most common arrhythmia found in HF patients. The paroxysmal onset of arrhythmias, for example, AF with a fast ventricular rate, can precipitate the acute decompensation of chronic HF. Multimodality machine learning models integrating clinical and echocardiographic parameters could predict new-onset AF [38]. For example, a study incorporating a convolutional neural network into an ECG-based prediction model can identify patients with AF, despite being under sinus rhythm, up to a sensitivity of 82.3% and specificity of 83.3% using 649,931 ECGs from over 180,000 patients [39]. Integrating artificial neural networks into automatic ECG analysis may allow for an accessible, inexpensive and accurate point-of-care early detection of AF and HF before the classical ECG manifestations, hence allowing for early intervention before the occurrence of fast AF or associated thromboembolic events.

Recognizing the importance of early detection, commercial artificial intelligence-driven wearables/handheld devices have been developed to alert the patient and allow cheap and accessible continuous ECG monitoring in an outpatient setting [40,41]. Although a standard 12-lead ECG should evaluate conduction and repolarization parameters, improvement in ECG detection driven by artificial intelligence has raised the quality of these commercial tracings, which

has shown to be non-inferior to standard ECG for the screening of sinus rhythm and AF [42–44]. The use of multi-objective optimization neural networks for compressing ECG data can reduce background noise and avoid storage of redundant data, allowing ease of commercial application [45]. By identifying the increased frequency of AF episodes, or the presence of fast AF, early intervention can take place to prevent further cardiac deterioration. In addition to AF, preliminary studies have reported that HF itself can also be identified through single-lead ECGs with AUC of 0.874 and 0.929 in internal and external validation, respectively, through developing a convolutional neural network model using over 39,000 ECGs from approximately 17,000 patients [46]. The availability of continuous ECG monitoring also upholds great research potential for studying the inter-relationship between AF and HF, since existing studies usually focus on paroxysmal AF or AF under an inpatient setting.

10.3 Management

10.3.1 Prognostic prediction

Given the chronic deteriorating disease course of HF, the prediction of mortality is of critical importance in decision-making. Although many predictors and risk scores have been identified, most only achieve modest success when applied to other populations beyond the cohort from which the models were developed [47–49]. While the subpar predictive performance is due to multifactorial causes, including dependence on non-universally available variables, the heterogeneity of the HF population and epidemiological variations, the failure to account for the multi-dimensional inter-predictor relationship is one of the most important causes [50].

Recently, the integration of machine learning approaches, which can capture the interactions between different variables, has significantly improved the predictive performance of risk stratification models. Studies have shown that machine learning approaches improve the predictive performance of benchmark logistic or Cox regression models in predicting mortality in HF [51–53]. A study reported that the addition of multiple plasma biomarkers to the Meta-Analysis Global Group in Chronic Heart Failure Risk Score (MAGGIC) risk score, a well-established risk score for mortality in HF based on over 39,000 patients from 30 studies, was able to raise the AUC to greater than 0.70 [54,55]. In addition, machine learning models allow the further stratification of HF patients into different phenotypic subgroups with distinct clinical characteristics and outcomes [56,57]. In a study based on the Swedish Heart Failure Registry of more than 44,000 patients, a random survival forest (RSF) model with cluster analysis identified four phenotypic HF subgroups with marked one-year survival difference (AUC = 0.83) [58]. An individualized management approach with better-tailored

treatment regimens, enabled by machine-learning driven models, may enhance the effectiveness of current HF therapies.

However, the HF data from electronic health records are often imbalanced with indistinct features that affect the accuracy of machine learning models. A recent study applied a convolutional neural network with feature rearrangement to tackle the problem of imbalance and transform raw data into extractable features, which achieved superior predictive performance for HF in-hospital, 30-day and 1-year mortality compared to traditional machine learning techniques such as RSF, MLP and SVM [59].

On the contrary, while the machine learning approach can also improve the prediction for hospital readmissions under HF, the predictive performance remains modest at best. A recent study using data from the Telemonitoring to Improve Heart Failure Outcomes trial applied different machine learning models to improve the performance of the traditional logistic regression model in the prediction of 30- and 180-day all-cause and HF-specific readmissions [60]. However, the AUC of the machine-learning model remains to be below 0.7. Another large-scale study testing the performance of machine learning models predicting 30-day all-cause readmission, based on the American Heart Association Get With the Guidelines Heart Failure registry with over 200,000 patients, found the AUC of the validation sets to range between 0.607 and 0.624 [61]. Unfortunately, the use of neural network models brought little improvement to the prediction of HF readmission. In an observational study based on electronic health records of more than 270,000 patients, the best performing model, which combined a RNN with conditional random fields, only achieved an AUC of 0.642 [62]. Although the use of neural network analysis has limited improvement upon the prediction findings, proposals have been made to improve the clinical interpretability of these models. An attention-based neural network model for HF readmission employed an attention mechanism, which marks essential features with attention signal, to improve the interpretability of the model. Although the AUC of the model is only 0.691, the study demonstrates that the addition of patient-specific attention weights can help to overcome the “black box” nature of artificial intelligence-driven models [63].

The modest predictive performance may be due to the significant training errors that are present in the models in the first place. First, hospital readmissions are multifactorial, including social health determinants that are not commonly recognized as independent predictors [64]. Moreover, the models assume the linear relationship between the predictors and readmission. Additionally, the temporal variation in the predictors was not accounted for in the evaluation of their predictiveness. Recent studies have demonstrated that glycemic and lipid variability are predictive for cardiovascular complications and mortality in patients with diabetes mellitus [65–68]. Future studies integrating the temporal dynamicity of biomarkers may improve the predictive performance of machine learning models. Overall, external validation with large, diverse populations is needed for the clinical application of machine learning-based models in HF.

The application of neural network analysis is particularly critical in incorporating medical imaging findings in the prognostic evaluation of HF patients. Artificial neural network models have been applied to echocardiographic data as early as 1995, where it has been used to predict the one-year mortality among 57 male and 38 female patients with HF [69]. In the current day and age, neural network analysis, combined with artificial intelligence-driven novel imaging interpretation techniques, can help identify prognostic indicators and algorithms. A study applied artificial neural networks into the interpretation of 123 iodine meta-iodobenzylguanidine single-photon emission computer tomography demonstrated that a planar global washout of $>30\%$ with a reduction in LVEF $>10\%$ has a sensitivity of 100% and specificity of 50% in the prediction of cardiac events in HF patients [70]. Moreover, the application of a convolutional neural network improved the image quality of cardiac magnetic resonance imaging (cMRI) given the undersampled radial cine images under a short reconstruction time by applying the spatio-temporal associations across different time frames [71]. As a result of the improved imaging resolution, in addition to novel analysis techniques, the further application of artificial neural networks on cMRI interpretation raises the segmentational accuracy of ventricles, septum and apex to over 92%. Automated pipelines have been developed to reduce the time-consuming and operator-dependent process of manual myocardial border delineation, where convolutional neural networks evaluated their predictive performances [72]. By advancing the interpretation of imaging, greater diagnostic and prognostic values can be extracted [73].

Furthermore, neural network analysis enhances the prognostic value of investigations through the retainment and utilization of data details that are lost in conventional analysis. Temporal data from the cardiopulmonary exercise test, the gold standard for cardiorespiratory fitness quantification among HF patients, are currently simplified into summary indices for ease of interpretation. A study comparing the prognostic performance of a feedforward neural network model, traditional multivariate regression models and conventional summary indices have found the neural network model to have the best performance (AUC = 0.842), in addition to being significantly more accurate than the commonly used cardiopulmonary exercise test risk score (AUC = 0.759) [74]. Superior predictive performance was reported by another cohort study of around 2000 patients on the prediction of cardiovascular mortality in HF using cardiopulmonary exercise testing, where an artificial neural network model was compared to the predictive performance of conventional Cox and logistic regression (AUC: 0.72 vs 0.69 vs 0.70) [75]. Similarly, a recent study applying a video-based, beta-to-beat, end-to-end deep learning approach for the estimation of LVEF from echocardiography was able to classify HF_{rEF} with an AUC of 0.97 from the training cohort and an AUC of 0.96 from the validation cohort [76]. In this case, the loss of data detail from echocardiography is due to human interpretation. The video-based neural network can improve prediction accuracy by minimizing the loss of details. The findings demonstrate that the prognostic value of investigations can be maximized

through the application of neural network analysis by minimizing the loss of details from the simplification of data.

10.3.2 Development of therapy

Neural network models are essential to the development of critical devices used to manage HF. For example, cardiac delay predictions under cardiac resynchronization therapy (CRT), an atrial-synchronized biventricular pacing device that aims to target the systolic dyssynchrony between ventricles, thus improve the cardiac output, can be improved through the use of an analog spiking neural network to assist adaptive CRT devices in its provision of optimal heartbeats [77]. Furthermore, left ventricular assist devices (LVAD) serve as mechanical circulatory support by pumping from the left ventricle to the aorta. It is a critical device to bridge HF patients to a heart transplant, but it is prone to inducing ischemia, ventricular suction and pulmonary congestion in patients due to its constant pumping speed. A recent study has developed a real-time convolutional neural network that allows for the estimation of preload based on LVAD flow, therefore enable for a sensorless physiological control system that responses to patients' hemodynamic changes and avoid the problems mentioned above without the need for additional sensors to measure the LVAD flow or pressure [78].

10.3.3 Optimal patient selection for specific therapies or recommendation of optimal therapy

Furthermore, machine learning can help to identify responders for specific therapies, thus improve the specificity of HF management. A recent retrospective population study has derived a machine-learning-driven predictive model for one-year all-cause mortality based on a 20-years follow-up of around 27,000 patients. By intervening on the actionable variables identified through the model, 8% of the 2844 patients who were predicted to die within a year were saved through prescriptions of specific treatments, including beta-blockers, hydralazine, angiotensin receptor antagonist, CRT and more [79]. Another study explored a deep neural network model that uses multidimensional echocardiographic data to identify distinct HFpEF subgroups with poorer prognosis and responsiveness toward spironolactone. Patients in the high-risk phenogroup had a higher event-free survival rate with spironolactone therapy in addition to higher rates of HF hospitalization and cardiovascular mortality, suggesting that this group of patients are more likely to be responsive to spironolactone [80].

As mentioned previously, CRT is an important device used to improve the LVEF of selected HF patients. However, a substantial number of patients selected under the current guidelines remained treatment non-responsive [81,82]. Recent randomized controlled trials have demonstrated that machine learning models integrating clinical and imaging parameters can better select the optimal patients

for CRT, ultimately reduce all-cause mortality and acute exacerbation of HF [83,84]. The machine-learning-driven improvement in the capability to stratify the patient suitability for CRT is attributed to capturing latent relationships between known prognostic predictors [85]. Another study predicted all-cause mortality or HF hospitalization within 12 months after CRT for patients in the Comparison of Medical Therapy, Pacing, and Defibrillation in Heart Failure (COMPANION) trial was able to stratify patients into quartiles based on their risk of adverse events using an RSF model combining QRS duration and bundle branch block morphology, with a statistically significant 8-fold survival difference between the highest and lowest risk group. More importantly, the model's predictive performance is better than stratification based on QRS duration or bundle branch block morphology alone [86]. However, the results from the direct application of the neural network model to raw ECG waveforms to predictive models were not promising, likely due to the presence of excessive noise and a lack of structured data [87]. With better accessibility of these risk calculators, better shared decision making between physicians and patients for CRT selection can be achieved.

For recommendation of optimal therapy, a multidimensional patient similarity assessment technique was described that leverages multiple types of information from electronic health records and predicts a medication plan for each new patient based on prior knowledge and data from similar patients [88]. This divided patients into different groups based on hierarchical clustering algorithms. and achieved an AUC of 0.74 for predicting HF therapy response.

10.4 Prevention

While research attention is focused on the secondary and tertiary prevention of HF, primary prevention is in fact, the most cost-effective way to reduce the global disease burden of HF. The key to primary prevention is the early identification of the population at risk, thus enabling earlier intervention to reduce the occurrence of HF ultimately.

Additionally, machine learning approach can be applied to the general public to identify subjects with potential or undiagnosed HF. Recently, machine learning approaches have been applied to population-based electronic health records to identify high-risk individuals [89]. Besides conventional methods such as the RSF model, RNNs have been applied to capture the temporal information, which improves the predictive performance [15,90]. A study identified asymptomatic left ventricular dysfunction by combining 12-lead ECG and echocardiogram data in a convolutional neural network model with more than 52,000 patients to an accuracy of 85.7%. The ability to diagnose asymptomatic left ventricular dysfunction, a prediagnostic HF disease state, demonstrates the HF-preventing potential of neural network analysis by identifying high-risk patients and enabling early

intervention [91]. Similarly, another study also applied a convolutional neural network model of the left ventricular posterior wall positioning for the detection of left ventricular hypertrophy, another potential prediagnostic HF cardiac change. The relative error between the model findings and the hospital measurements is less than 15%, which is less than the threshold of 20% for an acceptability repeatability error in clinical practice [92]. However, predictions do not equate to diagnosis, particularly in a heterogeneous disorder like HF. Clinicians should interpret the findings based on the clinical context and target risk factors present. From a public health perspective, the prediction of prediagnostic HF highlights the areas that warrant attention from the government to reduce the burden of HF.

10.5 Conclusion

In conclusion, machine learning with neural network models is critical for advancing HF diagnosis, management, and prevention. The incorporation of neural network models improves diagnostic accuracy, optimizes triage algorithms, and offers opportunities for high-risk identification in the heart failure population. The benefits of neural network analysis are particularly applicable to HF given its multifactorial nature. The latent features between clinical, ECG, echocardiographic, and other investigation results are identified by neural network models, thus allowing for the improved prediction performance against traditional machine learning approaches. Despite the advantages of neural network modeling, the disadvantages including difficulty in interpretation with potential limited clinical transferability and applicability. Researchers have addressed these areas of weakness through techniques such as attention-weighting and feature transformation to increase the interpretability of findings from the “black box.” In the future, studies to improve the predictive performance for HF-related adverse events, such as hospital readmissions, are needed to evaluate better patients’ quality of life and the disease burden on the public health system.

References

- [1] A. Groenewegen, F.H. Rutten, A. Mosterd, A.W. Hoes, *Epidemiology of heart failure*, *Eur. J. Heart Fail.* 22 (8) (2020) 1342–1356.
- [2] Y. Sun, N. Wang, X. Li, Y. Zhang, J. Yang, G. Tse, et al., *Predictive value of H2 FPEF score in patients with heart failure with preserved ejection fraction*, *ESC. Heart Fail.* (2021).
- [3] I. Lakhani, K.S.K. Leung, G. Tse, A.P.W. Lee, *Novel mechanisms in heart failure with preserved, midrange, and reduced ejection fraction*, *Front. Physiol.* 10 (2019) 874.

- [4] X. Zhang, Y. Sun, Y. Zhang, F. Chen, S. Zhang, H. He, et al., Heart failure with midrange ejection fraction: prior left ventricular ejection fraction and prognosis, *Front. Cardiovasc. Med.* 8 (794) (2021).
- [5] P. Bachtiger, C.M. Plymen, P.A. Pabari, J.P. Howard, Z.I. Whinnett, F. Opoku, et al., Artificial intelligence, data sensors and interconnectivity: future opportunities for heart failure, *Card. Fail. Rev.* 6 (2020) e11.
- [6] G. Bazoukis, S. Stavrakis, J. Zhou, S.C. Bollepalli, G. Tse, Q. Zhang, et al., Machine learning versus conventional clinical methods in guiding management of heart failure patients—a systematic review, *Heart Fail. Rev.* 26 (1) (2021) 23–34.
- [7] C. Krittanawong, K.W. Johnson, R.S. Rosenson, Z. Wang, M. Aydar, U. Baber, et al., Deep learning for cardiovascular medicine: a practical primer, *Eur. Heart J.* 40 (25) (2019) 2058–2073.
- [8] C.O. Akinyokun, O.U. Obot, F.-M.E. Uzoka (Eds.), *Application of Neuro-Fuzzy Technology in Medical Diagnosis: Case Study of Heart Failure*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.
- [9] W.E. Sanders Jr., T. Burton, A. Khosousi, S. Ramchandani, Machine learning: at the heart of failure diagnosis, *Curr. Opin. Cardiol.* 36 (2) (2021).
- [10] A. Cinar, S.A. Tuncer, Classification of normal sinus rhythm, abnormal arrhythmia and congestive heart failure ECG signals using LSTM and hybrid CNN-SVM deep neural networks, *Comput. Methods Biomech. Biomed. Engin* 24 (2) (2021) 203–214.
- [11] J.C.Y. Seah, J.S.N. Tang, A. Kitchen, F. Gaillard, A.F. Dixon, Chest radiographs in congestive heart failure: visualizing neural network learning, *Radiology* 290 (2) (2019) 514–522.
- [12] M.H. Asyali (Eds.), Discrimination power of long-term heart rate variability measures, in: *Proceedings of the Twenty-Fifth Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE Cat No03CH37439)*, 17–21 September 2003, 2003.
- [13] W. Chen, L. Zheng, K. Li, Q. Wang, G. Liu, Q. Jiang, A novel and effective method for congestive heart failure detection and quantification using dynamic heart rate variability measurement, *PLoS One* 11 (11) (2016) e0165304.
- [14] W. Chen, G. Liu, S. Su, Q. Jiang, H. Nguyen (Eds.), A CHF detection method based on deep learning with RR intervals, in: *Proceedings of the Thirty-Ninth Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 11–15, 2017.
- [15] E. Choi, A. Schuetz, W.F. Stewart, J. Sun, Using recurrent neural network models for early detection of heart failure onset, *J. Am. Med. Inf. Assoc.* 24 (2) (2017) 361–370.
- [16] Y. Isler, M. Kuntalp, Combining classical HRV indices with wavelet entropy measures improves to performance in diagnosing congestive heart failure, *Computers Biol. Med.* 37 (10) (2007) 1502–1510.
- [17] G. Liu, L. Wang, Q. Wang, G. Zhou, Y. Wang, Q. Jiang, A new approach to detect congestive heart failure using short-term heart rate variability measures, *PLoS One* 9 (4) (2014) e93399.
- [18] R. Mahajan, T. Viangteeravat, O. Akbilgic, Improved detection of congestive heart failure via probabilistic symbolic pattern recognition and heart rate variability metrics, *Int. J. Med. Inform.* 108 (2017) 55–63.

- [19] P. Melillo, R. Fusco, M. Sansone, M. Bracale, L. Pecchia, Discrimination power of long-term heart rate variability measures for chronic heart failure detection, *Med. Biol. Eng. Comput.* 49 (1) (2011) 67–74.
- [20] A. Narin, Y. Isler, M. Ozer, Investigating the performance improvement of HRV Indices in CHF using feature selection methods based on backward elimination and statistical significance, *Computers Biol. Med.* 45 (2014) 72–79.
- [21] L. Pecchia, P. Melillo, M. Sansone, M. Bracale, Discrimination power of short-term heart rate variability measures for CHF assessment, *IEEE Trans. Inf. Technol. Biomed.* 15 (1) (2011) 40–46.
- [22] S. Sanchez-Martinez, N. Duchateau, T. Erdei, G. Kunszt, S. Aakhus, A. Degiovanni, et al., Machine learning analysis of left ventricular function to characterize heart failure with preserved ejection fraction, *Circ. Cardiovasc. Imaging* 11 (4) (2018) e007138.
- [23] S. Sanchez-Martinez, N. Duchateau, T. Erdei, A.G. Fraser, B.H. Bijmens, G. Piella, Characterization of myocardial motion patterns by unsupervised multiple kernel learning, *Med. Image Anal.* 35 (2017) 70–82.
- [24] Z.Y. Zhang, S. Ravassa, E. Nkuipou-Kenfack, W.Y. Yang, S.M. Kerr, T. Koeck, et al., Novel urinary peptidomic classifier predicts incident heart failure, *J. Am. Heart Assoc.* 6 (8) (2017).
- [25] K. Rossing, H.S. Bosselmann, F. Gustafsson, Z.Y. Zhang, Y.M. Gu, T. Kuznetsova, et al., Urinary proteomics pilot study for biomarker discovery and diagnosis in heart failure with reduced ejection fraction, *PLoS One* 11 (6) (2016) e0157167.
- [26] D.-J. Choi, J.J. Park, T. Ali, S. Lee, Artificial intelligence for the diagnosis of heart failure, *NPJ Digital Med.* 3 (1) (2020) 54.
- [27] S. Blecker, S.D. Katz, L.I. Horwitz, G. Kuperman, H. Park, A. Gold, et al., Comparison of approaches for heart failure case identification from electronic health record data, *JAMA Cardiol.* 1 (9) (2016) 1014–1020.
- [28] S. Blecker, D. Sontag, L.I. Horwitz, G. Kuperman, H. Park, A. Reventovich, et al., Early identification of patients with acute decompensated heart failure, *J. Card. Fail.* 24 (6) (2018) 357–362.
- [29] S. Pakhomov, S.A. Weston, S.J. Jacobsen, C.G. Chute, R. Meverden, V.L. Roger, Electronic medical records for clinical research: application to the identification of heart failure, *Am. J. Manag. Care* 13 (6 Part 1) (2007) 281–288.
- [30] L. Rasmy, Y. Wu, N. Wang, X. Geng, W.J. Zheng, F. Wang, et al., A study of generalizability of recurrent neural network-based predictive models for heart failure onset risk using a large and heterogeneous EHR data set, *J. Biomed. Inf.* 84 (2018) 11–16.
- [31] B. Zinman, C. Wanner, J.M. Lachin, D. Fitchett, E. Bluhmki, S. Hantel, et al., Empagliflozin, cardiovascular outcomes, and mortality in type 2 diabetes, *N. Engl. J. Med.* 373 (22) (2015) 2117–2128.
- [32] A. Alonso-Betanzos, V. Bolon-Canedo, G.R. Heyndrickx, P.L. Kerkhof, Exploring guidelines for classification of major heart failure subtypes by using machine learning, *Clin. Med. Insights Cardiol.* 9 (Suppl. 1) (2015) 57–71.
- [33] P.C. Austin, J.V. Tu, J.E. Ho, D. Levy, D.S. Lee, Using methods from the data-mining and machine-learning literature for disease classification and prediction: a case study examining classification of heart failure subtypes, *J. Clin. Epidemiol.* 66 (4) (2013) 398–407.

- [34] Y. Isler, Discrimination of systolic and diastolic dysfunctions using multi-layer perceptron in heart rate variability analysis, *Comput. Biol. Med.* 76 (2016) 113–119.
- [35] S.J. Shah, D.H. Katz, S. Selvaraj, M.A. Burke, C.W. Yancy, M. Gheorghade, et al., Phenomapping for novel classification of heart failure with preserved ejection fraction, *Circulation* 131 (3) (2015) 269–279.
- [36] M. Solaiyappan, R.G. Weiss, P.A. Bottomley, Neural-network classification of cardiac disease from (31)P cardiovascular magnetic resonance spectroscopy measures of creatine kinase energy metabolism, *J. Cardiovasc. Magn. Reson.* 21 (1) (2019) 49.
- [37] J.L. Walsh, W.A. AlJaroudi, N. Lamaa, O.K. Abou Hassan, K. Jalkh, I.H. Elhajj, et al., A speckle-tracking strain-based artificial neural network model to differentiate cardiomyopathy type, *Scand. Cardiovasc. J.* 54 (2) (2020) 92–99.
- [38] G. Tse, J. Zhou, S.W.D. Woo, C.H. Ko, R.W.C. Lai, T. Liu, et al., Multi-modality machine learning approach for risk stratification in heart failure with left ventricular ejection fraction ≤ 45 , *ESC. Heart Fail.* (2020).
- [39] Z.I. Attia, P.A. Noseworthy, F. Lopez-Jimenez, S.J. Asirvatham, A.J. Deshmukh, B. J. Gersh, et al., An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction, *Lancet* 394 (10201) (2019) 861–867.
- [40] L. Desteghe, Z. Raymaekers, M. Lutin, J. Vijgen, D. Dilling-Boer, P. Koopman, et al., Performance of handheld electrocardiogram devices to detect atrial fibrillation in a cardiology and geriatric ward setting, *Europace* 19 (1) (2017) 29–39.
- [41] A.N. Koshy, J.K. Sajeev, N. Nerlekar, A.J. Brown, K. Rajakariar, M. Zureik, et al., Smart watches for heart rate assessment in atrial arrhythmias, *Int. J. Cardiol.* 266 (2018) 124–127.
- [42] G. Hindricks, T. Potpara, N. Dagres, E. Arbelo, J.J. Bax, C. Blomstrom-Lundqvist, et al., 2020 ESC Guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS): the task force for the diagnosis and management of atrial fibrillation of the European society of Cardiology (ESC) Developed with the special contribution of the European Heart Rhythm Association (EHRA) of the ESC, *Eur. Heart J.* 42 (5) (2021) 373–498.
- [43] A.D. William, M. Kanbour, T. Callahan, M. Bhargava, N. Varma, J. Rickard, et al., Assessing the accuracy of an automated atrial fibrillation detection algorithm using smartphone technology: the iREAD Study, *Heart Rhythm* 15 (10) (2018) 1561–1565.
- [44] J.P.J. Halcox, K. Wareham, A. Cardew, M. Gilmore, J.P. Barry, C. Phillips, et al., Assessment of remote heart rhythm sampling using the alivecor heart monitor to screen for atrial fibrillation: the REHEARSE-AF study, *Circulation* 136 (19) (2017) 1784–1794.
- [45] B. Zhang, J. Zhao, X. Chen, J. Wu, ECG data compression using a neural network model based on multi-objective optimization, *PLoS One* 12 (10) (2017) e0182500.
- [46] J. Cho, B. Lee, J.M. Kwon, Y. Lee, H. Park, B.H. Oh, et al., Artificial intelligence algorithm for screening heart failure with reduced ejection fraction using electrocardiography, *ASAIO J.* 67 (3) (2021) 314–321.
- [47] D.E. Lanfear, W.C. Levy, J. Stehlik, J.D. Estep, J.G. Rogers, K.B. Shah, et al., Accuracy of seattle heart failure model and heartmate II risk score in non-inotrope-dependent advanced heart failure patients: insights from the ROADMAP study (risk

- assessment and comparative effectiveness of left ventricular assist device and medical management in ambulatory heart failure patients), *Circ. Heart Fail.* 10 (5) (2017).
- [48] L.A. Allen, D.D. Matlock, S.M. Shetterly, S. Xu, W.C. Levy, L.B. Portalupi, et al., Use of risk models to predict death in the next year among individual ambulatory patients with heart failure, *JAMA Cardiol.* 2 (4) (2017) 435–441.
- [49] M. Canepa, C. Fonseca, O. Chioncel, C. Laroche, M.G. Crespo-Leiro, A.J.S. Coats, et al., Performance of prognostic risk scores in chronic heart failure patients enrolled in the European Society of Cardiology Heart Failure Long-Term Registry, *JACC Heart Fail.* 6 (6) (2018) 452–462.
- [50] E.D. Adler, A.A. Voors, L. Klein, F. Macheret, O.O. Braun, M.A. Urey, et al., Improving risk prediction in heart failure using machine learning, *Eur. J. Heart Fail.* 22 (1) (2020) 139–147.
- [51] A. Negassa, S. Ahmed, R. Zolty, S.R. Patel, Prediction model using machine learning for mortality in patients with heart failure, *Am. J. Cardiol.* (2021).
- [52] C. Ju, J. Zhou, S. Lee, M.S. Tan, T. Liu, G. Bazoukis, et al., Derivation of an electronic frailty index for predicting short-term mortality in heart failure: a machine learning approach, *ESC. Heart Fail.* (2021).
- [53] S. König, V. Pellissier, S. Hohenstein, A. Bernal, L. Ueberham, A. Meier-Hellmann, et al., Machine learning algorithms for claims data-based prediction of in-hospital mortality in patients with heart failure, *ESC. Heart Fail.* (2021).
- [54] J.A. Chirinos, A. Orlenko, L. Zhao, M.D. Basso, M.E. Cvijic, Z. Li, et al., Multiple plasma biomarkers for risk stratification in patients with heart failure and preserved ejection fraction, *J. Am. Coll. Cardiol.* 75 (11) (2020) 1281–1295.
- [55] S.J. Pocock, C.A. Ariti, J.J. McMurray, A. Maggioni, L. Kober, I.B. Squire, et al., Predicting survival in heart failure: a risk score based on 39 372 patients from 30 studies, *Eur. Heart J.* 34 (19) (2013) 1404–1413.
- [56] M.W. Segar, K.V. Patel, C. Ayers, M. Basit, W.H.W. Tang, D. Willett, et al., Phenomapping of patients with heart failure with preserved ejection fraction using machine learning-based unsupervised cluster analysis, *Eur. J. Heart Fail.* 22 (1) (2020) 148–158.
- [57] A.K. Hedman, C. Hage, A. Sharma, M.J. Brosnan, L. Buckbinder, L.M. Gan, et al., Identification of novel pheno-groups in heart failure with preserved ejection fraction using machine learning, *Heart* 106 (5) (2020) 342–349.
- [58] T. Ahmad, L.H. Lund, P. Rao, R. Ghosh, P. Warier, B. Vaccaro, et al., Machine learning methods improve prognostication, identify clinically distinct phenotypes, and detect heterogeneity in response to therapy in a large cohort of heart failure patients, *J. Am. Heart Assoc.* 7 (8) (2018).
- [59] Z. Wang, Y. Zhu, D. Li, Y. Yin, J. Zhang, Feature rearrangement based deep learning system for predicting heart failure mortality, *Comput. Methods Prog. Biomed.* 191 (2020) 105383.
- [60] B.J. Mortazavi, N.S. Downing, E.M. Bucholz, K. Dharmarajan, A. Manhapra, S.X. Li, et al., Analysis of machine learning techniques for heart failure readmissions, *Circ. Cardiovasc. Qual. Outcomes* 9 (6) (2016) 629–640.
- [61] J.D. Frizzell, L. Liang, P.J. Schulte, C.W. Yancy, P.A. Heidenreich, A.F. Hernandez, et al., Prediction of 30-day all-cause readmissions in patients hospitalized for heart failure: comparison of machine learning and other statistical approaches, *JAMA Cardiol.* 2 (2) (2017) 204–209.

- [62] A. Allam, M. Nagy, G. Thoma, M. Krauthammer, Neural networks versus logistic regression for 30 days all-cause readmission prediction, *Sci. Rep.* 9 (1) (2019) 9277.
- [63] P. Chen, W. Dong, J. Wang, X. Lu, U. Kaymak, Z. Huang, Interpretable clinical prediction via attention-based neural network, *BMC Med. Inf. Decis. Mak.* 20 (Suppl. 3) (2020) 131.
- [64] P.S. Keenan, S.L. Normand, Z. Lin, E.E. Drye, K.R. Bhat, J.S. Ross, et al., An administrative claims measure suitable for profiling hospital performance on the basis of 30-day all-cause readmission rates among patients with heart failure, *Circ. Cardiovasc. Qual. Outcomes* 1 (1) (2008) 29–37.
- [65] S. Lee, T. Liu, J. Zhou, Q. Zhang, W.T. Wong, G. Tse, Predictions of diabetes complications and mortality using hba1c variability: a 10-year observational cohort study, *Acta Diabetol.* (2020).
- [66] S. Lee, J. Zhou, W.T. Wong, T. Liu, W.K.K. Wu, I.C.K. Wong, et al., Glycemic and lipid variability for predicting complications and mortality in diabetes mellitus using machine learning, *BMC Endocr. Disord.* 21 (1) (2021) 94.
- [67] C.R.L. Cardoso, N.C. Leite, C.B.M. Moram, G.F. Salles, Long-term visit-to-visit glycemic variability as predictor of micro- and macrovascular complications in patients with type 2 diabetes: the Rio de Janeiro Type 2 Diabetes Cohort Study, *Cardiovasc. Diabetol.* 17 (1) (2018) 33.
- [68] S. Yokota, H. Tanaka, Y. Mochizuki, F. Soga, K. Yamashita, Y. Tanaka, et al., Association of glycemic variability with left ventricular diastolic function in type 2 diabetes mellitus, *Cardiovasc. Diabetol.* 18 (1) (2019) 166.
- [69] J. Ortiz, C.G. Ghefter, C.E. Silva, R.M. Sabbatini, One-year mortality prognosis in heart failure: a neural network approach based on echocardiographic data, *J. Am. Coll. Cardiol.* 26 (7) (1995) 1586–1593.
- [70] G. Currie, B. Iqbal, H. Kiat, Intelligent imaging: radiomics and artificial neural networks in heart failure, *J. Med. Imaging Radiat. Sci.* 50 (4) (2019) 571–574.
- [71] H. El-Rewaidy, A.S. Fahmy, F. Pashakhanloo, X. Cai, S. Kucukseymen, I. Csecs, et al., Multi-domain convolutional neural network (MD-CNN) for radial reconstruction of dynamic cardiac MRI, *Magn. Reson. Med.* 85 (3) (2021) 1195–1208.
- [72] N.A. Farrag, A. Lochbihler, J.A. White, E. Ukwatta, Evaluation of fully automated myocardial segmentation techniques in native and contrast-enhanced T1-mapping cardiovascular magnetic resonance images using fully convolutional neural networks, *Med. Phys.* 48 (1) (2021) 215–226.
- [73] D. Liu, Z. Jia, M. Jin, Q. Liu, Z. Liao, J. Zhong, et al., Cardiac magnetic resonance image segmentation based on convolutional neural network, *Comput. Methods Prog. Biomed.* 197 (2020) 105755.
- [74] J. Hearn, H.J. Ross, B. Mueller, C.P. Fan, E. Crowdy, J. Duhamel, et al., Neural networks for prognostication of patients with heart failure, *Circ. Heart Fail.* 11 (8) (2018) e005193.
- [75] J. Myers, C.R. de Souza, A. Borghi-Silva, M. Guazzi, P. Chase, D. Bensimhon, et al., A neural network approach to predicting outcomes in heart failure using cardiopulmonary exercise testing, *Int. J. Cardiol.* 171 (2) (2014) 265–269.
- [76] D. Ouyang, B. He, A. Ghorbani, N. Yuan, J. Ebinger, C.P. Langlotz, et al., Video-based AI for beat-to-beat assessment of cardiac function, *Nature* 580 (7802) (2020) 252–256.

- [77] Q. Sun, F. Schwartz, J. Michel, Y. Herve, R. Dalmolin, Implementation study of an analog spiking neural network for assisting cardiac delay prediction in a cardiac resynchronization therapy device, *IEEE Trans. Neural Netw.* 22 (6) (2011) 858–869.
- [78] M. Fetanat, M. Stevens, C. Hayward, N.H. Lovell, A sensorless control system for an implantable heart pump using a real-time deep convolutional neural network, *IEEE Trans. Biomed. Eng.* (2021). PP.
- [79] L. Jing, A.E. Ulloa Cerna, C.W. Good, N.M. Sauers, G. Schneider, D.N. Hartzel, et al., A Machine learning approach to management of heart failure populations, *JACC Heart Fail.* 8 (7) (2020) 578–587.
- [80] A. Pandey, N. Kagiya, N. Yanamala, M.W. Segar, J.S. Cho, M. Tokodi, et al., Deep-learning models for the echocardiographic assessment of diastolic dysfunction, *JACC Cardiovasc. Imaging* (2021).
- [81] W.T. Abraham, W.G. Fisher, A.L. Smith, D.B. Delurgio, A.R. Leon, E. Loh, et al., Cardiac resynchronization in chronic heart failure, *N. Engl. J. Med.* 346 (24) (2002) 1845–1853.
- [82] J.B. Young, W.T. Abraham, A.L. Smith, A.R. Leon, R. Lieberman, B. Wilkoff, et al., Combined cardiac resynchronization and implantable cardioversion defibrillation in advanced chronic heart failure: the MIRACLE ICD trial, *JAMA* 289 (20) (2003) 2685–2694.
- [83] M. Cikes, S. Sanchez-Martinez, B. Claggett, N. Duchateau, G. Piella, C. Butakoff, et al., Machine learning-based phenogrouping in heart failure to identify responders to cardiac resynchronization therapy, *Eur. J. Heart Fail.* 21 (1) (2019) 74–85.
- [84] M. Tokodi, W.R. Schwertner, A. Kovacs, Z. Toser, L. Staub, A. Sarkany, et al., Machine learning-based mortality prediction of patients undergoing cardiac resynchronization therapy: the SEMMELWEIS-CRT score, *Eur. Heart J.* 41 (18) (2020) 1747–1756.
- [85] A.K. Feeny, J. Rickard, D. Patel, S. Toro, K.M. Trulock, C.J. Park, et al., Machine learning prediction of response to cardiac resynchronization therapy: improvement versus current guidelines, *Circ. Arrhythm. Electrophysiol.* 12 (7) (2019) e007316.
- [86] M.M. Kalscheur, R.T. Kipp, M.C. Tattersall, C. Mei, K.A. Buhr, D.L. DeMets, et al., Machine learning algorithm predicts cardiac resynchronization therapy outcomes: lessons from the COMPANION trial, *Circ. Arrhythm. Electrophysiol.* 11 (1) (2018) e005499.
- [87] C. Cai, A.P. Tafti, C. Ngufor, P. Zhang, P. Xiao, M. Dai, et al., Using ensemble of machine learning methods to predict outcomes of cardiac resynchronization, *J. Cardiovasc. Electrophysiol.* (2021).
- [88] M. Panahiazar, V. Taslimitehrani, N.L. Pereira, J. Pathak, Using EHRs for heart failure therapy recommendation using multidimensional patient similarity analytics, *Stud. Health Technol. Inform.* 210 (2015) 369–373.
- [89] K. Ng, S.R. Steinhubl, C. de Filippi, S. Dey, W.F. Stewart, Early detection of heart failure using electronic health records: practical implications for time before diagnosis, data diversity, data quantity, and data density, *Circ. Cardiovasc. Qual. Outcomes* 9 (6) (2016) 649–658.

- [90] R. Chen, W.F. Stewart, J. Sun, K. Ng, X. Yan, Recurrent neural networks for early detection of heart failure from longitudinal electronic health record data: implications for temporal modeling with respect to time before diagnosis, data density, data quantity, and data type, *Circ. Cardiovasc. Qual. Outcomes* 12 (10) (2019) e005114.
- [91] Z.I. Attia, S. Kapa, F. Lopez-Jimenez, P.M. McKie, D.J. Ladewig, G. Satam, et al., Screening for cardiac contractile dysfunction using an artificial intelligence-enabled electrocardiogram, *Nat. Med.* 25 (1) (2019) 70–74.
- [92] Z. Jian, X. Wang, J. Zhang, X. Wang, Y. Deng, Diagnosis of left ventricular hypertrophy using convolutional neural network, *BMC Med. Inf. Decis. Mak.* 20 (1) (2020) 243.

This page intentionally left blank

Role of artificial intelligence and radiomics in diagnosing renal tumors: a survey

11

Mohamed Shehata¹, Ahmed Elmahdy², Ahmed Alksas¹, Rasha Abouelkheir², Ali Mahmoud¹, Mohamed Abou El-Ghar², Mohammed Ghazal³ and Ayman S. El-Baz^{1,4}

¹University of Louisville, Louisville, KY, United States

²Radiology Department, Urology and Nephrology Center, University of Mansoura, Mansoura, Egypt

³Electrical, Computer and Biomedical Engineering Department, Abu Dhabi University, Abu Dhabi, United Arab Emirates

⁴University of Louisville at Alamein International University (UofL-AIU), New Alamein City, Egypt

11.1 Introduction

Renal cell carcinomas (RCCs) are the most common and aggressive renal cancer (represents around 70% of all renal cancers). The World Health Organization states that the most common RCC subtypes are clear cell RCC (ccRCCs), and non ccRCCs including papillary RCC (paRCCs) and chromophobe RCC (chRCCs), responsible for about 70%, 15%, and 5% of all RCCs, respectively. This RCCs taxonomy is of immense importance as each subtype has its own prognosis [1]. On the other hand, angiomyolipomas (AMLs) and oncocytomas are benign renal tumors that can be easily misclassified as RCCs using conventional diagnostic techniques such as physical examination and/or visual qualifications, particularly if these AMLs do not have sufficient fat content. Consequently, this might lead to the unneeded surgical intervention of such benign tumors. In addition, approximately 15%–20% of RCC also turn out to be AML at surgery time [2]. Early assessment of RCC is essential to provide the proper management plan. Biopsy remains the gold standard; however, it is unfavorable due to its invasiveness, high cost, and adverse events such as bleeding and infection and it takes around a week for reporting. To account for these limitations, we develop a two-stage classification computer-aided diagnostic (CAD) system that has the ability to differentiate benign from malignant renal tumors and classify its subtypes using contrast-enhanced computed tomography (CE-CT) [3]. The aim of this study is to review the recent diagnostic applications of AI and ML in renal tumors. More details are discussed in the following sections.

11.2 Basic background

11.2.1 Deep learning

Deep learning (DL) is the most important ML tool in the general imaging and computer vision domains [4]. It utilizes neural networks with several layers (>20) [5]. It has the following five categories in radiology: classification, object detection, segmentation, image processing, and natural language processing “NLP.” The use of DL for clinical applications is an issue of great interest. Before using DL algorithms, it is important that they are assessed using various test datasets, and also that the training datasets are evaluated to reveal imbalances in the data distribution [6]. It is believed that DL will be involved in more and more in the field of diagnostic imaging, especially MRI. DL algorithms could conduct mundane tasks in the future, leaving radiologists to focus on intellectually demanding challenges [7].

11.2.2 Machine learning

Machine learning (ML) is regarded as an AI branch as it can extract significant patterns from instances. Computers will carry out well-defined and repetitive tasks continuously and indefatigably. Machines can be trained and even perform tasks accurately, which were considered as too complicated for machines, suggesting that the algorithms of ML are possibly helpful components of decision support and CAD systems [5]. These CAD systems are further divided into groups: computer-aided diagnosis (CADx) and computer-aided detection (CADE). The CADE systems are created to assist the radiologist in “detecting and locating the abnormal area” in images, while the CADx systems are developed to detect and further differentiate malignant from benign tissues [8]. CADE recognizes doubtful features present on the images to reduce the false-negative readings. As presently utilized, the radiologist first evaluates the image after that he activates this system and re-assess the areas which are marked by CADE prior to making the final report [9]. Representation learning is an ML subtype where no “hand-crafted” features are given; instead, the computer algorithm is taught the features needed to classify the given data. In general, the addition of data enhances the performance. systems depending on representation learning may provide higher quality performance than the conventional ML systems which integrate “hand-crafted” features [10]. Thus, the chief ML techniques objective is to form a model that can be utilized to carry out estimation, prediction, classification, or any other function. However, the commonest task is to classify the item of data into one of many predefined categories [11]. ML is categorized according to the nature of the data labeling into supervised, unsupervised, and reinforcement learning. In supervised learning, the labels of data are given to the algorithm in the training phase. In unsupervised learning, no labels of data are provided to the algorithm of learning [12].

11.2.3 Radiomics

Radiomics describes a variety of techniques to obtain quantitative features from images to enhance the accuracy of image interpretation. It transforms image data to feature space that facilitates subsequent data analysis to enhance decision support. It facilitates the extraction of complex structure that, while present in the image, is not apparent to the naked eye. Recent advances in artificial intelligence have accelerated the application of radiomics clinical practice [13]. The process starts with choosing and designing a standard imaging protocol to reduce the variabilities in image parameters and enhance the study's comparability and reproducibility [14]. Mackin and colleagues documented the same variability level in the radiomics features values estimated on the images of (CT) derived from dissimilar CT scanners as that of these radiomics features variability detected in the CT images of cases suffering from nonsmall cell pulmonary cancers. In MRI, image acquisition standardization may be a higher challenge due to the affection by several factors involving the contrast agents, the parameters of sequence, and hardware [15].

11.3 Steps of artificial intelligence-based diagnostic systems

11.3.1 Image acquisition

The imaging scanners provide raw data volumes that need to be processed to be utilizable in the medical fields. It is important to choose the most appropriate reconstruction algorithm for every single case as this will directly affect image quality and the ability to detect and analyze atypical imaging features. The reconstructed images are reserved in a big database (public database), to facilitate gaining of more data as well as providing easy access for various clinics [16,17].

11.3.2 Image segmentation

Following image acquisition, reduction down to the important parts that are named "volumes of interest" must be done [18]. Instead of segmenting the images manually, an automatic segmentation should be utilized. A probable solution is semiautomatic and automatic algorithms of segmentation. Before applying on a large scale, an algorithm has to meet the following criteria: 1st, it has to be reproducible, which means when it is utilized on similar data the results will not alter. The second factor is consistency. The algorithm should solve the problem and carry out the task without making anything that is not significant. It is essential that it is able to recognize the affected area in the dissimilar scans. It is also required to be perfect. It reveals the affected area in the most accurate way

probable. Only with perfect data, perfect outcomes can be reached. The outcomes should be finished as rapidly as probable [17].

11.3.3 Feature extraction and qualifications

These are divided into five groups: functions of the pixel intensity distribution (histogram), measures of the shape and size of defined image subregions, statistics of the relationship between intensity values of different pixels [e.g., gray-level co-occurrence matrix (GLCM), neighborhood gray-tone difference matrix (NGTDM), size zone matrix (SZM), and run-length matrix (RLM)], filter response, and fractal geometry [19]. Because of its extensive diversity, feature reduction is typically required to eliminate redundant information. Many dissimilar features require to be assessed with chosen algorithms to speed up this process. In Addition, features that are nonreproducible and unstable should be removed because features have decreased-fidelity will be liable to cause unrepeatable models and false findings [20].

11.3.4 Diagnostic analysis

Prior to the definite analysis, the molecular and clinical data are required to be incorporated. This has a large effect on the abstract from the analysis. There are various ways to finish the analysis of data. First, the dissimilar features are put side by side to each other to realize whether they have anything in common and to detect what it means in the condition of occurrence simultaneously. Another way is unsupervised or supervised analysis. The unsupervised analysis reviews the data included and has the ability to express them in graphs. The supervised analysis utilizes a result variable that can produce models of prediction [21].

11.4 Texture analysis

Texture analysis indicates the quantitative evaluation of the tumor heterogeneity by analyzing the relationship and distribution of voxel gray levels in every single image [22]. It has been effectively utilized in several fields, and it has been applied in (CT) and (MRI) as a CADx tool. Actually, this type of quantitative analysis is an advantage of CAD as it is unnoticeable by the human eye [23].

11.4.1 Principles

As Gillis et al. declared, “images are more than pictures, they are data.” Radiomics is the transformation of digital images into mineable, high-dimensional data and can be utilized to extract quantitative features on the basis of shape intensity, texture, and volume features [24]. A heterogeneous ill-defined

mass is more liable to be malignant than a homogeneous well-defined mass. On the other hand, the heterogeneity of lesions is not easy to calculate and may show inter-observer dissimilarity. Radiomic TA permits the evaluation of the pixel intensities' spatial interrelations and can be utilized to calculate the heterogeneity of lesion [25].

11.4.2 Statistical techniques

Statistical techniques are most frequently utilized to explain the spatial relationship of the gray-level value in the image [26]. Many are found in commercial image processing software or are straightforward to implement in-house. This includes first and second-order statics [27].

11.4.2.1 First-order statics

This depends only on the intensity histogram of the image, or a subregion thereof, such as mean intensity, standard deviation (SD), percentage of pixels within a particular range of values, entropy (irregularity), skewness (asymmetry), and kurtosis (tail weight). It is not independent of pixel location and does not confer any information on the spatial interrelation between gray values [27].

11.4.2.2 Second-order statics

Examples derived from the GLCM include second-order entropy, homogeneity, energy, correlation, and dissimilarity. They can be derived utilizing an RLM, which analyzes texture in a particular direction. Higher-order statistics of cliques of ≥ 3 pixels, like coarseness, busyness, and, contrast can be measured utilizing NGTDM. They show the benefit of assessing pixel values in context of a pixel neighborhood, taking the relationship with nearby pixels into account [27] (Table 11.1).

11.4.3 Model-based methods

These represent an image as a realization of a sophisticated mathematical model (such as fractal or stochastic). Image data are used to estimate the model parameters, which comprise the radiomic quantities used for subsequent image analysis. A disadvantage of the model-based approach is the computational complexity involved in the parameter estimation [28].

11.4.4 Transform methods

The texture of the image may be more readily analyzed when the data are transformed to a different space, as in a frequency-based or the scale space representation. These methods often employ the Fourier, Gabor, or wavelet transform. The wavelet transform is the most widely used because the choice of wavelet basis allows it to be adapted to the problem in question [28].

Table 11.1 Spectrum of statistical-based first-order and higher-order texture features.

Texture feature	Level/order	Description	Examples
Pixel intensity histogram	First order	A graph where the x-axis is pixel gray level and y-axis is frequency of occurrence.	Mean, standard deviation, skewness, or kurtosis of intensity distribution; optimum threshold; first-order entropy; mean of positive pixels (MPP).
Run-length matrix	Second order	Number of consecutive pixels (in a given direction) with the same intensity.	Run nonuniformity (of length or of intensity); emphasis on long or short runs.
Gray-level cooccurrence matrix	Second order	Two dimensional histogram of pixel pairs with a given spatial relationship.	Contrast, correlation, homogeneity, second-order entropy.
Advanced metrics	Higher order	Encode the relationships between any number of pixels in any configuration.	Potentially hundreds: autoregressive model parameters, Haar wavelet energy, geometric descriptors, neighborhood gray-tone differences.

11.4.5 Texture parameters

The most commonly used texture parameters are from six main categories:

1. Histogram of pixel values (first-order statistics)
2. Histogram of image gradient values (first-order statistics)
3. GLRLM (second-order statistical)
4. GLCM (second-order statistical)
5. Auto-regressive model (model-based)
6. Wavelet decomposition (transform-based)

11.4.5.1 Filtration-histogram method

This method includes an initial filtration to define image features of a specified size, followed by histogram analysis. The histograms of the pixel values in the filtered and unfiltered images are quantified using standard descriptors, specifically: mean, SD, skewness, and kurtosis [29] (Table 11.2).

11.4.5.2 Postprocessing software

It can be carried out retrospectively on images were taken in the similar contrast enhancement phase with the same technique; generally, no prospective acquisition is required. CT texture analysis (CTTA) can be carried out on either one-section or volumetric datasets to evaluate the heterogeneity of the tumor [30] (Fig. 11.1).

Table 11.2 Definitions of histogram parameters [29].

Parameter	Definition
Mean	The average pixel intensity within the region of interest
Standard deviation	Width of pixel intensity distribution
Skewness	Asymmetry of the histogram. Negative skew indicates tail on the left side of the histogram is heavier than on the right side. Positive skew indicates the reverse. Symmetric distributions have zero skew.
Kurtosis excess	Measure of histogram tail weight. Positive kurtosis indicates greater likelihood of extreme values compared to a Gaussian (normal) distribution. Negative kurtosis indicates reduced likelihood of extreme values.

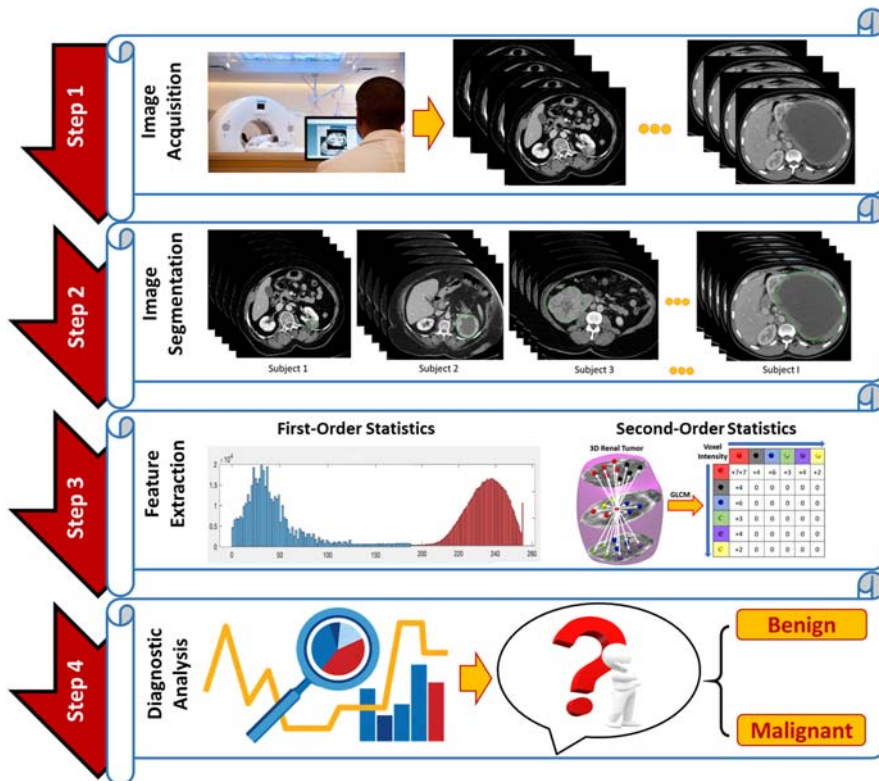


FIGURE 11.1

Steps of machine learning and texture analysis of renal tumors.

11.5 Clinical applications of artificial intelligence and radiomics

11.5.1 Benign versus malignant renal tumors

Fat-containing AML is the only benign solid mass confidently diagnosed by conventional imaging. The rest of the solid renal masses are assumed malignant by radiologists and referred for surgical resection. However, this leads to unnecessary resection of benign lesions in 13%–16% of patients [31]. Recent advances in AI and ML have proposed new methods of image analysis to help to solve this issue. CTTA determines the degree of lesion heterogeneity through analysis of signal intensity (SI) in each pixel in a predetermined region of interest (ROI). CTTA was able to correctly recognize (ccRCCs), (paRCCs), oncocytomas, and cysts in (91%, 100%, 89%, and 100%) of cases respectively, using a random forest (RF) model [32]. However, a recent study showed that a combination of RF models with shape metrics had higher accuracy [33]. On the other hand, differences in entropy are helpful in the differentiation between RCC and fat-poor AML (fp-AML) as well as between (chRCC) and oncocytoma which is very challengeable on conventional imaging [34]. Data derived from second-order statistics helped to differentiate Wilms tumor from clear cell sarcoma and rhabdoid tumor [35]. Value of statistical relational learning, specifically, RFGB (relational functional gradient boosting) is a hopeful CDS tool for the diagnosis of kidney masses because it learns models which are efficient and understandable. It facilitates the creation of diagnostic systems which are able to make decisions supported by explanations that are well understood by radiologists and urologists [36]. One recent study [37] reported higher accuracy in discrimination benign from malignant renal masses by freezing numbers of layers (before the mixed six layers) during transfer learning, and utilizing the ROI and RBR data sets. Another study reported that the histogram analysis metric SD in the pre-contrast images showed the best results in discrimination of chRCC from oncocytoma [38]. Radiomic ML and TA proved to have a promising role in recognition of lipid poor AML and discriminating it from atypical RCC and perform better than expert radiologists (80% as compared to 50%–71%) [39], yet some researchers do not recommend its application in clinical practice owing to the presence of a small number of pRCC that are marked as benign lesions using TA [40].

11.5.2 Renal cell carcinoma versus angiomyolipoma

Fp-AML could be distinguished from ccRCC with 80% accurateness based on TA of quadripasic contrast-enhanced CT (CECT). It is noteworthy that assessment of precontrast images alone or any individual phase of postcontrast phases read to similar high accuracy [41,42]. Other features of fp-AML included lack of pseudo-capsule, angular interphase (good delineation between mass and renal

parenchyma), lesser entropy (more uniform texture as compared to RCC) [42,43]. The CT-based radiomics nomogram presents encouraging predictive efficiency for distinguishing AML. Wvf from homogenous-ccRCC may help in tailoring accurate management [44]. Some authors suggested that ML-based models utilizing data derived from precontrast images can discriminate fp-AML from RCC (AUC = 0.9), eliminating the need of CE-CT [41,45]. On the other hand, newly introduced point shear elastography (pSWE) could benefit from the statistical models of ML in recognition of pf-AML. One recent study found that assessment of nonlinear support vector machine (SVM) was more accurate than median shear wave velocity in discriminating RCC from AML utilizing combined data from the lesion, cortex, and renal medulla [46]. TA and ML algorithms were also applied to MRI images, as researchers found high accuracy (AUC 0.89) of a MPP at SSF 2 on DWI b500 in differentiating fp-AML from RCC [47].

11.5.3 Renal cell carcinoma versus oncocytoma

Oncocytoma has similar imaging and histological features with chRCC [48]. Several studies tried to distinguish oncocytoma from other types of RCC using several parameters such as tensor Flow™ Inception [49], skewness and kurtosis [50], TA with predefined five classifiers [51], histogram with Fourier analysis [52], while other researchers found that mean at SSF 0 on DWI b1000 was the best parameter in MRI to discriminate oncocytoma [47].

11.5.4 Renal cell carcinoma versus renal cyst

The CTTA of heterogeneity of mass is effective in distinguishing renal cysts from RCC displaying low attenuation values in CT. It could be carried out on a one slice utilizing one measurement of ROI [53].

11.5.5 Subtyping of renal cell carcinoma

RCC has many histopathological subtypes. The most famous are clear cell, papillary, and chRCC. The differentiation between these subtypes is of clinical importance as it affects clinical outcomes and the rate of tumor recurrence after therapy [53–55]. TA could substitute invasive biopsy procedures in distinguishing papillary from non-ccRCC. One study found that low (SD, entropy, MPP) and high kurtosis are characteristics of non-ccRCC and attributed that to more heterogeneous texture present in ccRCC. On the other hand, they highlighted the value of MPP and skewness as tissue classifiers that could differentiate between pRCC and chRCC [55,56]. Regarding Similarly, researchers found that certain texture parameters could be used as markers

differentiating ccRCC from other less aggressive subtypes with high accuracy (AUC > 0.8) [57].

11.5.6 Grading of renal cell carcinoma

A huge number of radiomics signatures (RSs) are derived from multidetector CT images; these RSs are further processed by five predictive models designed for differential diagnosis of low-grade versus high-grade ccRCC [58,59]. Researchers found that data collected from triphasic CT is of greater accuracy as compared to those extracted from single phase acquisition [60]. Entropy is increased in ccRCC. High entropy is related to high-grade renal tumors as it represents heterogeneity and therefore can be used to discriminate Fuhrman I/II (low grade) from Fuhrman III/IV (high grade). It can be regarded as a supplementary marker while deciding the therapy aggressiveness [58,61–63]. Other useful ML architectures in radiomics include k-nearest neighbor classifiers, multilayer perceptrons, logistic regression, SVM, and RF [63,64]. The combined use of conventional imaging features and Ct radiomics is helpful in the detection of coagulative necrosis which is considered a sign of poor prognosis in ccRCC [65].

11.5.7 Staging of renal cell carcinoma

High values of TA-derived CT radiomics including histogram analysis (especially kurtosis and skewness), entropy, Fourier analysis, and Gray-level difference method are well correlated with high-stage ccRCC. They are also useful in preoperative recognition of adherent perinephric Fat related to low-stage tumors [66,67]. Another interesting application of CT radiomics is that it could be used to predict genetic mutations such as Bap1 [68].

11.5.8 Characterization of small renal mass

Small RCC < 4 cm could be indistinguishable from fpAML and represents a diagnostic challenge in conventional imaging. Several researchers proposed TA radiomics to discriminate these lesions. One group found that combined volumetric histogram analysis and reduced-FOV DWI are valuable in the characterization of small RCC and differentiate from benign lesions [69]. In agreement with convolutional neural networks, they concluded that the utilization of images taken in the corticomedullary phases yielded the best results [69]. Feng and colleagues used a SVM with recursive feature elimination to pick 11 features of TA to differentiate small RCC from fpAML [70]. Other groups found that TA features (entropy, histogram analysis, mean of the positive pixels) are helpful in the assessment of small RCC in the same way like large RCC with added diagnostic and prognostic potentials [71,72].

11.6 Merits and limitations

11.6.1 Merits

Improvement in diagnosis of renal tumor depends on improvement in DL and future research that focuses on the building more comprehensive medical databases that are not limited to research centers only and on new advances concerned with AI techniques. The use of improved algorithms could be achieved by mobile devices or through cloud services. The presence of specialized AI-based software which can provide image-guided, real-time, intraoperative decisions of resection of renal masses is one of the greatest achievements in the field of uro-imaging [3].

11.6.2 Limitations

Although texture analysis has proven useful in a wide variety of research applications, it still faces challenges to its adoption in a clinical setting. The major limitation is the there is no single platform or technique to be followed, actually, there is great variability of applications and derived parameters reaching up to hundreds and each research center develops its own protocol. Also, there is variability regarding image acquisition and reconstruction parameters, methods used for image segmentation, preprocessing techniques (e.g., noise reduction, contrast enhancement), and quantity and quality of texture feature outputs (statistical, model-based, etc.) [73–75]. Another limitation is that small lesions have fewer count statistics that may affect the results. On the other hand, tissue heterogeneity could be affected by parameters of the CT acquisition protocol that alter attenuation or pixel relationships. However, research groups reported that the first-order texture features are less prone to changes when using variable techniques unlike mean attenuation [76].

11.7 Future directions

TA has been described and used in several scientific articles, but it needs to be applied to a larger population sample before a certain protocol could be established prior to its introduction of everyday practice. Another issue is that it requires extra-human effort on workstations especially directed to the segmentation of a large number of images with extra fear of human errors. So, future directions should be aimed to reach automated techniques that require no human effort and provide TA results in the same setting with conventional images [77–79]. Automated segmentation and TA will also facilitate the creation of large databases, which enables more dedicated research studies, more collaboration between different institutes which paves the way to a generation of standardized techniques and analysis methods. An increasingly exciting the aspect of medical

imaging is artificial intelligence. Till now there are fears in the radiology community that AI and CAD are designed to replace radiologists, nevertheless, the advances in computer science are growing every day and gaining more territories as helpful and more solid tools in image-based decision making [77–95].

11.8 Conclusion

TA is an effective tool for distinguishing between benign and malignant renal masses, characterization of their pathological subtypes and grades, staging of RCC, and differentiation of small RCC. TA assesses variables not able to be detected by the naked human eye, so it acts as the eye of AI. On the other hand, human experience and diagnostic intuition have still a crucial role in future AI developments to ensure the proper performance of these systems and to recognize instances where AI might fail, leading to undesired consequences.

This work could also be applied to various other applications in medical imaging, such as the prostate [96–100], the kidney [101–119], the heart [120–137], the lung [138–187], the brain [188–210], the vascular system [211–221], the retina [222–231], the bladder [232–236], the liver [237,238], head and neck [239–241], and injury prediction [242] as well as several nonmedical applications [243–249].

References

- [1] G. Low, G. Huang, W. Fu, Z. Moloo, S. Girgis, Review of renal cell carcinoma and its common subtypes in radiology, *World J. Radiol.* 8 (5) (2016) 484.
- [2] T.J. van Oostenbrugge, J.J. Fütterer, P.F. Mulders, Diagnostic imaging for solid renal tumors: a pictorial review, *Kidney Cancer* 2 (2) (2018) 79–93.
- [3] R. Suarez-Ibarrola, S. Hein, G. Reis, C. Gratzke, A. Miernik, Current and future applications of machine and deep learning in urology: a review of the literature on urolithiasis, renal cell carcinoma, and bladder and prostate cancer, *World J. Urol.* 38 (10) (2020) 2329–2347.
- [4] H. Greenspan, B. Van Ginneken, R.M. Summers, Guest editorial deep learning in medical imaging: overview and future promise of an exciting new technique, *IEEE Trans. Med. Imaging* 35 (5) (2016) 1153–1159.
- [5] B.J. Erickson, P. Korfiatis, Z. Akkus, T.L. Kline, Machine learning for medical imaging, *Radiographics* 37 (2) (2017) 505–515.
- [6] D. Ueda, A. Shimazaki, Y. Miki, Technical and clinical overview of deep learning in radiology, *Jpn. J. Radiol.* 37 (1) (2019) 15–33.
- [7] M.A. Mazurowski, M. Buda, A. Saha, M.R. Bashir, Deep learning in radiology: an overview of the concepts and a survey of the state of the art with focus on MRI, *J. Magn. Reson. Imaging* 49 (4) (2019) 939–954.

- [8] A. Jalalian, S.B. Mashohor, H.R. Mahmud, M.I.B. Saripan, A.R.B. Ramli, B. Karasfi, Computer-aided detection/diagnosis of breast cancer in mammography and ultrasound: a review, *Clin. Imaging* 37 (3) (2013) 420–426.
- [9] R.A. Castellino, Computer aided detection (cad): an overview, *Cancer Imaging* 5 (1) (2005) 17.
- [10] A. Tang, R. Tam, A. Cadrin-Chênevert, W. Guest, J. Chong, J. Barfett, et al., Canadian association of radiologists white paper on artificial intelligence in radiology, *Can. Assoc. Radiol. J.* 69 (2) (2018) 120–135.
- [11] K. Kourou, T.P. Exarchos, K.P. Exarchos, M.V. Karamouzis, D.I. Fotiadis, Machine learning applications in cancer prognosis and prediction, *Comput. Struct. Biotechnol. J.* 13 (2015) 8–17.
- [12] G. Choy, O. Khalilzadeh, M. Michalski, S. Do, A.E. Samir, O.S. Pinykh, et al., Current applications and future impact of machine learning in radiology, *Radiology* 288 (2) (2018) 318–328.
- [13] H.J. Aerts, Data science in radiology: a path forward, *Clin. Cancer Res.* 24 (3) (2018) 532–534.
- [14] S.M. Ayyad, M. Shehata, A. Shalaby, A. El-Ghar, M. Ghazal, M. El-Melegy, et al., Role of ai and histopathological images in detecting prostate cancer: a survey, *Sensors* 21 (8) (2021) 2586.
- [15] D. Mackin, X. Fave, L. Zhang, D. Fried, J. Yang, B. Taylor, et al., Measuring ct scanner variability of radiomics features, *Invest. Radiol.* 50 (11) (2015) 757.
- [16] V. Kumar, Y. Gu, S. Basu, A. Berglund, S.A. Eschrich, M.B. Schabath, et al., Radiomics: the process and the challenges, *Magn. Reson. Imaging* 30 (9) (2012) 1234–1248.
- [17] S. Sanduleanu, H.C. Woodruff, E.E. De Jong, J.E. Van Timmeren, A. Jochems, L. Dubois, et al., Tracking tumor biology with radiomics: a systematic review utilizing a radiomics quality score, *Radiat. Oncol.* 127 (3) (2018) 349–360.
- [18] C. Parmar, E. Rios Velazquez, R. Leijenaar, M. Jermoumi, S. Carvalho, R.H. Mak, et al., Robust radiomics feature quantification using semiautomatic volumetric segmentation, *PLoS One* 9 (7) (2014) e102107.
- [19] G. Thibault, J. Angulo, F. Meyer, Advanced statistical matrices for texture characterization: application to cell classification, *IEEE Trans. Biomed. Eng.* 61 (3) (2013) 630–637.
- [20] I. Tunali, L.O. Hall, S. Napel, D. Cherezov, A. Guvenis, R.J. Gillies, et al., Stability and reproducibility of computed tomography radiomic features extracted from peritumoral regions of lung cancer lesions, *Med. Phys.* 46 (11) (2019) 5075–5085.
- [21] P. Lambin, R.T. Leijenaar, T.M. Deist, J. Peerlings, E.E. De Jong, J. Van Timmeren, et al., Radiomics: the bridge between medical imaging and personalized medicine, *Nat. Rev. Clin. Oncol.* 14 (12) (2017) 749–762.
- [22] B. Ganeshan, K.A. Miles, Quantifying tumour heterogeneity with CT, *Cancer Imaging* 13 (1) (2013) 140.
- [23] R. Thomas, L. Qin, F. Alessandrino, S.P. Sahu, P.J. Guerra, K.M. Krajewski, et al., A review of the principles of texture analysis and its role in imaging of genitourinary neoplasms, *Abdom. Radiol.* 44 (7) (2019) 2501–2510.
- [24] R.J. Gillies, P.E. Kinahan, H. Hricak, Radiomics: images are more than pictures, they are data, *Radiology* 278 (2) (2016) 563–577.
- [25] R.M. Summers, Texture analysis in radiology: does the emperor have no clothes? *Abdom. Radiol.* 42 (2) (2017) 342–345.

- [26] M.G. Lubner, A.D. Smith, K. Sandrasegaran, D.V. Sahani, P.J. Pickhardt, ct texture analysis: definitions, applications, biologic correlates, and challenges, *Radiographics* 37 (5) (2017) 1483–1503.
- [27] U. Bashir, M.M. Siddique, E. Mclean, V. Goh, G.J. Cook, Imaging heterogeneity in lung cancer: techniques, applications, and challenges, *Am. J. Roentgenol.* 207 (3) (2016) 534–543.
- [28] G. Castellano, L. Bonilha, L. Li, F. Cendes, Texture analysis of medical images, *Clin. Radiol.* 59 (12) (2004) 1061–1069.
- [29] K.A. Miles, B. Ganeshan, M.P. Hayball, CT texture analysis using the filtration-histogram method: what do the measurements mean? *Cancer Imaging* 13 (3) (2013) 400.
- [30] F. Ng, R. Kozarski, B. Ganeshan, V. Goh, Assessment of tumor heterogeneity by ct texture analysis: can the largest cross-sectional area be used as an alternative to whole tumor analysis? *Eur. J. Radiol.* 82 (2) (2013) 342–348.
- [31] F.U. Kay, I. Pedrosa, Imaging of solid renal masses, *Radiol. Clin.* 55 (2) (2017) 243–258.
- [32] S.P. Raman, Y. Chen, J.L. Schroeder, P. Huang, E.K. Fishman, CT texture analysis of renal masses: pilot study using random forest classification for prediction of pathology, *Acad. Radiol.* 21 (12) (2014) 1587–1596.
- [33] F.Y. Yap, B.A. Varghese, S.Y. Cen, D.H. Hwang, X. Lei, B. Desai, et al., Shape and texture-based radiomics signature on CT effectively discriminates benign from malignant renal masses, *Eur. Radiol.* 31 (2) (2021) 1011–1021.
- [34] K. Sasaguri, N. Takahashi, M. Takeuchi, R.E. Carter, B.C. Leibovich, A. Kawashima, Differentiation of benign from metastatic adrenal masses in patients with renal cell carcinoma on contrast-enhanced CT, *Am. J. Roentgenol.* 207 (5) (2016) 1031–1038.
- [35] H.J. Shin, J.Y. Kwak, E. Lee, M.-J. Lee, H. Yoon, K. Han, et al., Texture analysis to differentiate malignant renal tumors in children using gray-scale ultrasonography images, *Ultrasound Med. Biol.* 45 (8) (2019) 2205–2212.
- [36] G. Kunapuli, B.A. Varghese, P. Ganapathy, B. Desai, S. Cen, M. Aron, et al., A decision-support tool for renal mass classification, *J. Digital Imaging* 31 (6) (2018) 929–939.
- [37] L. Zhou, Z. Zhang, Y.-C. Chen, Z.-Y. Zhao, X.-D. Yin, H.-B. Jiang, A deep learning-based radiomics model for differentiating benign and malignant renal tumors, *Transl. Oncol.* 12 (2) (2019) 292–300.
- [38] B.A. Varghese, F. Chen, D.H. Hwang, S.Y. Cen, B. Desai, I.S. Gill, et al., Differentiation of predominantly solid enhancing lipid-poor renal cell masses by use of contrast-enhanced CT: evaluating the role of texture in tumor subtyping, *Am. J. Roentgenol.* 211 (6) (2018) W288–W296.
- [39] X.-Y. Sun, Q.-X. Feng, X. Xu, J. Zhang, F.-P. Zhu, Y.-H. Yang, et al., Radiologic-radiomic machine learning models for differentiation of benign and malignant solid renal masses: comparison with expert-level radiologists, *Am. J. Roentgenol.* 214 (1) (2020) W44–W54.
- [40] C. Erdim, A.H. Yardimci, C.T. Bektas, B. Kocak, S.B. Koca, H. Demir, et al., Prediction of benign and malignant solid renal masses: machine learning-based ct texture analysis, *Acad. Radiol.* 27 (10) (2020) 1422–1429.
- [41] M.-W. You, N. Kim, H. Choi, The value of quantitative ct texture analysis in differentiation of angiomyolipoma without visible fat from clear cell renal cell carcinoma on four-phase contrast-enhanced CT images, *Clin. Radiol.* 74 (7) (2019) 547–554.

- [42] Y. Ma, F. Cao, X. Xu, W. Ma, Can whole-tumor radiomics-based ct analysis better differentiate fat-poor angiomyolipoma from clear cell renal cell carcinoma: compared with conventional ct analysis? *Abdom. Radiol.* 45 (8) (2020) 2500–2507.
- [43] M. Shehata, A. Alksas, R.T. Abouelkheir, A. Elmahdy, A. Shaffie, A. Soliman, et al., A new computer-aided diagnostic (CAD) system for precise identification of renal tumors, in: *Proceedings of the Eighteenth International Symposium on Biomedical Imaging (ISBI)*, IEEE, 2021, pp. 1378–1381.
- [44] P. Nie, G. Yang, Z. Wang, L. Yan, W. Miao, D. Hao, et al., A ct-based radiomics nomogram for differentiation of renal angiomyolipoma without visible fat from homogeneous clear cell renal cell carcinoma, *Eur. Radiol.* 30 (2) (2020) 1274–1284.
- [45] R. Yang, J. Wu, L. Sun, S. Lai, Y. Xu, X. Liu, et al., Radiomics of small renal masses on multiphasic ct: accuracy of machine learning–based classification models for the differentiation of renal cell carcinoma and angiomyolipoma without visible fat, *Eur. Radiol.* 30 (2) (2020) 1254–1263.
- [46] H. Sagreiya, A. Akhbardeh, D. Li, R. Sigrist, B.I. Chung, G.A. Sonn, et al., Point shear wave elastography using machine learning to differentiate renal cell carcinoma and angiomyolipoma, *Ultrasound Med. Biol.* 45 (8) (2019) 1944–1954.
- [47] A. Razik, A. Goyal, R. Sharma, D. Kandasamy, A. Seth, P. Das, et al., MR texture analysis in differentiating renal cell carcinoma from lipid-poor angiomyolipoma and oncocytoma, *Br. J. Radiol.* 93 (1114) (2020). 20200569.
- [48] K. Sasaguri, N. Takahashi, Ct and mr imaging for solid renal mass characterization, *Eur. J. Radiol.* 99 (2018) 40–54.
- [49] H. Coy, K. Hsieh, W. Wu, M.B. Nagarajan, J.R. Young, M.L. Douek, et al., Deep learning and radiomics: the utility of google tensorflow inception in classifying clear cell renal cell carcinoma and oncocytoma on multiphasic ct, *Abdom. Radiol.* 44 (6) (2019) 2009–2020.
- [50] H. Yu, J. Scalera, M. Khalid, A.-S. Touret, N. Bloch, B. Li, et al., Texture analysis as a radiomic marker for differentiating renal tumors, *Abdom. Radiol.* 42 (10) (2017) 2470–2478.
- [51] Y. Li, X. Huang, Y. Xia, L. Long, Value of radiomics in differential diagnosis of chromophobe renal cell carcinoma and renal oncocytoma, *Abdom. Radiol.* 45 (10) (2020) 3193–3201.
- [52] B.A. Varghese, F. Chen, D.H. Hwang, S.Y. Cen, I.S. Gill, V.A. Duddalwar, Differentiating solid, non-macroscopic fat containing, enhancing renal masses using fast fourier transform analysis of multiphase ct, *Br. J. Radiol.* 91 (1089) (2018) 20170789.
- [53] N.Y. Kim, M.G. Lubner, J.T. Nystrom, J.F. Swietlik, E.J. Abel, T.C. Havighurst, et al., Utility of ct texture analysis in differentiating low-attenuation renal cell carcinoma from cysts: a bi-institutional retrospective study, *Am. J. Roentgenol.* 213 (6) (2019) 1259–1266.
- [54] C.L. Vendrami, Y.S. Velichko, F.H. Miller, A. Chatterjee, C.P. Villavicencio, V. Yaghmai, et al., Differentiation of papillary renal cell carcinoma subtypes on MRI: qualitative and texture analysis, *Am. J. Roentgenol.* 211 (6) (2018) 1234–1245.
- [55] G.-M.-Y. Zhang, B. Shi, H.-D. Xue, B. Ganeshan, H. Sun, Z.-Y. Jin, Can quantitative ct texture analysis be used to differentiate subtypes of renal cell carcinoma? *Clin. Radiol.* 74 (4) (2019) 287–294.
- [56] M.G. Lubner, N. Stabo, E.J. Abel, A.M. Del Rio, P.J. Pickhardt, CT textural analysis of large primary renal cell carcinomas: pretreatment tumor heterogeneity correlates with histologic findings and clinical outcomes, *Am. J. Roentgenol.* 207 (1) (2016) 96–105.

- [57] A. Goyal, A. Razik, D. Kandasamy, A. Seth, P. Das, B. Ganeshan, et al., Role of mr texture analysis in histological subtyping and grading of renal cell carcinoma: a preliminary study, *Abdom. Radiol.* 44 (10) (2019) 3336–3349.
- [58] J. Shu, Y. Tang, J. Cui, R. Yang, X. Meng, Z. Cai, et al., Clear cell renal cell carcinoma: Ct-based radiomics features for the prediction of fuhrman grade, *Eur. J. Radiol.* 109 (2018) 8–12.
- [59] X. He, Y. Wei, H. Zhang, T. Zhang, F. Yuan, Z. Huang, et al., Grading of clear cell renal cell carcinomas by using machine learning based on artificial neural networks and radiomic signatures extracted from multidetector computed tomography images, *Acad. Radiol.* 27 (2) (2020) 157–168.
- [60] F. Lin, E.-M. Cui, Y. Lei, L.-P. Luo, CT-based machine learning model to predict the fuhrman nuclear grade of clear cell renal cell carcinoma, *Abdom. Radiol.* 44 (7) (2019) 2528–2534.
- [61] Y. Deng, E. Soule, A. Samuel, S. Shah, E. Cui, M. Asare-Sawiri, et al., Ct texture analysis in the differentiation of major renal cell carcinoma subtypes and correlation with Fuhrman grade, *Eur. Radiol.* 29 (12) (2019) 6922–6929.
- [62] Z. Feng, Q. Shen, Y. Li, Z. Hu, CT texture analysis: a potential tool for predicting the fuhrman grade of clear-cell renal carcinoma, *Cancer Imaging* 19 (1) (2019) 1–7.
- [63] J. Shu, D. Wen, Y. Xi, Y. Xia, Z. Cai, W. Xu, et al., Clear cell renal cell carcinoma: Machine learning-based computed tomography radiomics analysis for the prediction of WHO/ISUP grade, *Eur. J. Radiol.* 121 (2019) 108738.
- [64] M. Nazari, I. Shiri, G. Hajianfar, N. Oveisi, H. Abdollahi, M.R. Deevband, et al., Noninvasive fuhrman grading of clear cell renal cell carcinoma using computed tomography radiomic features and machine learning, *La. Radiol. Med.* 125 (8) (2020) 754–762.
- [65] K. Xu, L. Liu, W. Li, X. Sun, T. Shen, F. Pan, et al., CT-based radiomics signature for preoperative prediction of coagulative necrosis in clear cell renal cell carcinoma, *Korean J. Radiol.* 21 (6) (2020) 670.
- [66] Z.-E. Khene, K. Bensalah, A. Largent, S. Shariat, G. Verhoest, B. Peyronnet, et al., Role of quantitative computed tomography texture analysis in the prediction of adherent perinephric fat, *World J. Urol.* 36 (10) (2018) 1635–1642.
- [67] T.S. Gill, B.A. Varghese, D.H. Hwang, S.Y. Cen, M. Aron, M. Aron, et al., Juxtatumoral perinephric fat analysis in clear cell renal cell carcinoma, *Abdom. Radiol.* 44 (4) (2019) 1470–1480.
- [68] A. Li, W. Xing, H. Li, Y. Hu, D. Hu, Z. Li, et al., Subtype differentiation of small (<4 cm) solid renal mass using volumetric histogram analysis of dwi at 3-t mri, *Am. J. Roentgenol.* 211 (3) (2018) 614–623.
- [69] T. Tanaka, Y. Huang, Y. Marukawa, Y. Tsuboi, Y. Masaoka, K. Kojima, et al., Differentiation of small (<4 cm) renal masses on multiphase contrast-enhanced ct by deep learning, *Am. J. Roentgenol.* 214 (3) (2020) 605–612.
- [70] Z. Feng, P. Rong, P. Cao, Q. Zhou, W. Zhu, Z. Yan, et al., Machine learning-based quantitative texture analysis of ct images of small renal masses: differentiation of angiomyolipoma without visible fat from renal cell carcinoma, *Eur. Radiol.* 28 (4) (2018) 1625–1633.
- [71] U.N. Hoang, S. Mojdeh Mirmomen, O. Meirelles, J. Yao, M. Merino, A. Metwalli, et al., Assessment of multiphasic contrast-enhanced mr textures in differentiating small renal mass subtypes, *Abdom. Radiol.* 43 (12) (2018) 3400–3409.

- [72] A.T. Scrima, M.G. Lubner, E.J. Abel, T.C. Havighurst, D.D. Shapiro, W. Huang, et al., Texture analysis of small renal cell carcinomas at mdct for predicting relevant histologic and protein biomarkers, *Abdom. Radiol.* 44 (6) (2019) 1999–2008.
- [73] M.G. Lubner, Radiomics and artificial intelligence for renal mass characterization, *Radiol. Clin.* 58 (5) (2020) 995–1008.
- [74] L. Lu, R.C. Ehmke, L.H. Schwartz, B. Zhao, Assessing agreement between radiomic features computed for multiple ct imaging settings, *PLoS One* 11 (12) (2016) e0166550.
- [75] M. Shafiq-ul Hassan, G.G. Zhang, K. Latifi, G. Ullah, D.C. Hunt, Y. Balagurunathan, et al., Intrinsic dependencies of ct radiomic features on voxel size and number of gray levels, *Med. Phys.* 44 (3) (2017) 1050–1062.
- [76] A. Chalkidou, M.J. O’Doherty, P.K. Marsden, False discovery rates in pet and ct studies with texture features: a systematic review, *PLoS One* 10 (5) (2015) e0124165.
- [77] A. Bhandari, M. Ibrahim, C. Sharma, R. Liang, S. Gustafson, M. Prior, Ct-based radiomics for differentiating renal tumours: a systematic review, *Abdom. Radiol.* 46 (5) (2021) 2052–2063.
- [78] M. Recht, R.N. Bryan, Artificial intelligence: threat or boon to radiologists? *J. Am. Coll. Radiol.* 14 (11) (2017) 1476–1480.
- [79] G. Wang, M. Kalra, C.G. Orton, Machine learning will transform radiology significantly within the next 5 years, *Med. Phys.* 44 (6) (2017) 2041–2044.
- [80] C. Parmar, P. Grossmann, J. Bussink, P. Lambin, H.J. Aerts, Machine learning methods for quantitative radiomic biomarkers, *Sci. Rep.* 5 (1) (2015) 1–11.
- [81] M. Shehata, F. Khalifa, A. Soliman, R. Alrefai, M. Abou El-Ghar, A.C. Dwyer, et al., A novel framework for automatic segmentation of kidney from dw-MRI, in: *Proceedings of the IEEE Twelfth International Symposium on Biomedical Imaging (ISBI), IEEE, 2015*, pp. 951–954.
- [82] M. Shehata, F. Khalifa, A. Soliman, A.T. Eldeen, M. Abou El-Ghar, T. El-Diasty, et al., An appearance-guided deformable model for 4d kidney segmentation using diffusion mri, *Biomedical Image Segmentation, CRC Press, 2016*, pp. 291–312.
- [83] F. Khalifa, M. Shehata, A. Soliman, M. Abou El-Ghar, T. El-Diasty, A.C. Dwyer, et al., A generalized mri-based cad system for functional assessment of renal transplant, in: *Proceedings of the Fourteenth International Symposium on Biomedical Imaging (ISBI 2017), IEEE, 2017*, pp. 758–761.
- [84] M. Shehata, M. Abou El-Ghar, T. Eldiasty, A. El-Baz, An integrated cad system of dwi-mri and laboratory biomarkers in diagnosis of kidney transplant dysfunction, in: *Proceedings of the European Congress of Radiology (ECR 2018), Austria Center Vienna, Bruno-Kreisky-Platz, Vol. 11220, 2018*.
- [85] H. Abdeltawab, M. Shehata, A. Shalaby, S. Mesbah, M. El-Baz, M. Ghazal, et al., A new 3D CNN-based cad system for early detection of acute renal transplant rejection, in: *Proceedings of the Twenty-fourth International Conference on Pattern Recognition (ICPR), IEEE Computer Society, 2018*, pp. 3898–3903.
- [86] M. Shehata, M. Ghazal, G. Beache, M. Abou El-Ghar, A. Dwyer, H. Hajjdiab, et al., Role of integrating diffusion mr image-markers with clinical-biomarkers for early assessment of renal transplants, in: *Proceedings of the Twenty-fifth IEEE International Conference on Image Processing (ICIP), 2018*, pp. 146–150.
- [87] M. Shehata, M. Ghazal, F. Khalifa, M. Abou El-Ghar, A. Khalil, A.C. Dwyer, et al., A novel cad system for detecting acute rejection of renal allografts based on integrating imaging-markers and laboratory biomarkers, in: *Proceedings of the IEEE*

- International Conference on Imaging Systems and Techniques (IST), IEEE, 2018, pp. 1–6.
- [88] M. Shehata, F. Taher, M. Ghazal, A. Mahmoud, G. Beache, M. Abou El-Ghar, et al., Early assessment of acute renal rejection post-transplantation: a combined imaging and clinical biomarkers protocol, in: Proceedings of the IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), IEEE, 2018, pp. 297–302.
- [89] M. Shehata, M. Abou El-Ghar, T. Eldiasty, A. El-Baz, in: Integrating clinical with diffusion image markers as a noninvasive alternative to renal biopsy, in: Proceedings of the European Congress of Radiology (ECR 2019), Austria Center Vienna, Bruno-Kreisky-Platz, vol. 11220, 2019.
- [90] M. Shehata, A. Shalaby, M. Ghazal, M. Abou El-Ghar, M. Badawy, G. Beache, et al., Early assessment of renal transplants using bold-MRI: Promising results, in: Proceedings of the IEEE International Conference on Image Processing (ICIP), IEEE, 2019, pp. 1395–1399.
- [91] M. Shehata, A. Shalaby, A. Mahmoud, M. Ghazal, H. Hajjdiab, M.A. Badawy, et al., *Towards Big Data in Acute Renal Rejection*, Chapman and Hall/CRC, 2019.
- [92] M. Shehata, M. Ghazal, H.A. Khalifeh, A. Khalil, A. Shalaby, A.C. Dwyer, et al., A deep learning based cad system for renal allograft assessment: diffusion, bold, and clinical biomarkers, in: Proceedings of the IEEE International Conference on Image Processing (ICIP), IEEE, 2020, pp. 355–359.
- [93] M. Shehata, H. Abdeltawab, M. Ghazal, A. Khalil, S. Shaker, A. Shalaby, et al., Accurate identification of renal transplant rejection: convolutional neural networks and diffusion MRI, *State of the Art in Neural Networks and their Applications*, Elsevier, 2021, pp. 91–115.
- [94] M. Shehata, F. Taher, M. Ghazal, S. Shaker, M. Abou El-Ghar, M. Badawy, et al., Early identification of acute rejection for renal allografts: a machine learning approach, *State of the Art in Neural Networks and their Applications*, Elsevier, 2021, pp. 197–218.
- [95] M. Shehata, F. Khalifa, A. Soliman, S. Shaker, A. Shalaby, M. El-Baz, et al., Early classification of renal rejection types: A deep learning approach, *Machine Learning in Medicine*, CRC Press, 2021, pp. 257–280.
- [96] I. Reda, M. Ghazal, A. Shalaby, M. Elmogy, A. AbouEl-Fetouh, B.O. Ayinde, et al., A novel adcs-based cnn classification system for precise diagnosis of prostate cancer, in: Proceedings of the Twenty-fourth International Conference on Pattern Recognition (ICPR), IEEE, 2018, pp. 3923–3928.
- [97] I. Reda, A. Khalil, M. Elmogy, A. Abou El-Fetouh, A. Shalaby, M. Abou El-Ghar, et al., Deep learning role in early diagnosis of prostate cancer, *Technol. Cancer Res. Treat.* 17 (2018). 1533034618775530.
- [98] I. Reda, B.O. Ayinde, M. Elmogy, A. Shalaby, M. El-Melegy, M.A. El-Ghar, et al., A new cnn-based system for early diagnosis of prostate cancer, in: Proceedings of the IEEE Fifteenth International Symposium on Biomedical Imaging (ISBI 2018), IEEE, 2018, pp. 207–210.
- [99] S.M. Ayyad, M.A. Badawy, M. Shehata, A. Alksas, A. Mahmoud, M. Abou El-Ghar, et al., A new framework for precise identification of prostatic adenocarcinoma, *Sensors* 22 (5) (2022). Available from: <https://www.mdpi.com/1424-8220/22/5/1848>.

- [100] K. Hammouda, F. Khalifa, M. El-Melegy, M. Ghazal, H.E. Darwish, M.A. El-Ghar, et al., A deep learning pipeline for grade groups classification using digitized prostate biopsy specimens, *Sensors* 21 (20) (2021) 6708.
- [101] M. Shehata, A. Shalaby, A.E. Switala, M. El-Baz, M. Ghazal, L. Fraiwan, et al., A multimodal computer-aided diagnostic system for precise identification of renal allograft rejection: preliminary results, *Med. Phys.* 47 (6) (2020) 2427–2440.
- [102] M. Shehata, F. Khalifa, A. Soliman, M. Ghazal, F. Taher, M. Abou El-Ghar, et al., Computer-aided diagnostic system for early detection of acute renal transplant rejection using diffusion-weighted MRI, *IEEE Trans. Biomed. Eng.* 66 (2) (2018) 539–552.
- [103] E. Hollis, M. Shehata, M. Abou El-Ghar, M. Ghazal, T. El-Diasty, M. Merchant, et al., Statistical analysis of adcs and clinical biomarkers in detecting acute renal transplant rejection, *Br. J. Radiol.* 90 (1080) (2017). 20170125.
- [104] M. Shehata, A. Alksas, R.T. Abouelkheir, A. Elmahdy, A. Shaffie, A. Soliman, et al., A comprehensive computer-assisted diagnosis system for early assessment of renal cancer tumors, *Sensors* 21 (14) (2021) 4928.
- [105] F. Khalifa, G.M. Beache, M.A. El-Ghar, T. El-Diasty, G. Gimel'farb, M. Kong, et al., Dynamic contrast-enhanced MRI based early detection of acute renal transplant rejection, *IEEE Trans. Med. Imaging* 32 (10) (2013) 1910–1927.
- [106] F. Khalifa, M.A. El-Ghar, B. Abdollahi, H. Frieboes, T. El-Diasty, A. El-Baz, A comprehensive non-invasive framework for automated evaluation of acute renal transplant rejection using DCE-MRI, *NMR Biomed.* 26 (11) (2013) 1460–1470.
- [107] F. Khalifa, A. Elnakib, G.M. Beache, G. Gimel'farb, M.A. El-Ghar, G. Sokhadze, et al., 3D kidney segmentation from CT images using a level set approach guided by a novel stochastic speed function, in: *Proceedings of International Conference Medical Image Computing and Computer-Assisted Intervention (MICCAI'11)*, Toronto, Canada, 18–22 September 2011, pp. 587–594.
- [108] M. Shehata, F. Khalifa, E. Hollis, A. Soliman, E. Hosseini-Asl, M.A. El-Ghar, et al., A new non-invasive approach for early classification of renal rejection types using diffusion-weighted MRI, in: *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, 2016, IEEE, 2016, pp. 136–140.
- [109] F. Khalifa, A. Soliman, A. Takieldeem, M. Shehata, M. Mostapha, A. Shaffie, et al., Kidney segmentation from CT images using a 3D NMF-guided active contour model, in: *Proceedings of the IEEE Thirteenth International Symposium on Biomedical Imaging (ISBI)*, 2016, IEEE, 2016, pp. 432–435.
- [110] M. Shehata, F. Khalifa, A. Soliman, A. Takieldeem, M.A. El-Ghar, A. Shaffie, et al., 3d diffusion mri-based cad system for early diagnosis of acute renal rejection, in: *Proceedings of the Biomedical Imaging (ISBI), 2016 IEEE Thirteenth International Symposium on*, IEEE, 2016, pp. 1177–1180.
- [111] M. Shehata, F. Khalifa, A. Soliman, R. Alrefai, M.A. El-Ghar, A.C. Dwyer, et al., A level set-based framework for 3D kidney segmentation from diffusion mr images, in: *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, 2015, IEEE, 2015, pp. 4441–4445.
- [112] M. Shehata, F. Khalifa, A. Soliman, M.A. El-Ghar, A.C. Dwyer, G. Gimel'farb, et al., A promising noninvasive cad system for kidney function assessment, in: *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2016, pp. 613–621.

- [113] F. Khalifa, A. Soliman, A. Elmaghraby, G. Gimel'farb, A. El-Baz, 3d kidney segmentation from abdominal images using spatial-appearance models, *Comput. Math. Methods Med.* 2017 (2017) 1–10.
- [114] E. Hollis, M. Shehata, F. Khalifa, M.A. El-Ghar, T. El-Diasty, A. El-Baz, Towards non-invasive diagnostic techniques for early detection of acute renal transplant rejection: A review, *Egypt. J. Radiol. Nucl. Med.* 48 (1) (2016) 257–269.
- [115] M. Shehata, F. Khalifa, A. Soliman, M.A. El-Ghar, A.C. Dwyer, A. El-Baz, Assessment of renal transplant using image and clinical-based biomarkers, in: *Proceedings of Thirteenth Annual Scientific Meeting of American Society for Diagnostics and Interventional Nephrology (ASDIN'17)*, New Orleans, LA, USA, 10–12 February 2017, 2017.
- [116] M. Shehata, F. Khalifa, A. Soliman, M.A. El-Ghar, A.C. Dwyer, A. El-Baz, Early assessment of acute renal rejection, in: *Proceedings of Twelfth Annual Scientific Meeting of American Society for Diagnostics and Interventional Nephrology (ASDIN'16)*, Phoenix, AZ, USA, 19–21 February 2016, 2016.
- [117] A. Eltanboly, M. Ghazal, H. Hajjdiab, A. Shalaby, A. Switala, A. Mahmoud, et al., Level sets-based image segmentation approach using statistical shape priors, *Appl. Math. Comput.* 340 (2019) 164–179.
- [118] M. Shehata, A. Mahmoud, A. Soliman, F. Khalifa, M. Ghazal, M.A. El-Ghar, et al., 3d kidney segmentation from abdominal diffusion MRI using an appearance-guided deformable boundary, *PLoS One* 13 (7) (2018) e0200082.
- [119] H. Abdeltawab, M. Shehata, A. Shalaby, F. Khalifa, A. Mahmoud, M.A. El-Ghar, et al., A novel cnn-based cad system for early assessment of transplanted kidney dysfunction, *Sci. Rep.* 9 (1) (2019) 5948.
- [120] K. Hammouda, F. Khalifa, H. Abdeltawab, A. Elnakib, G. Giridharan, M. Zhu, et al., A new framework for performing cardiac strain analysis from cine mri imaging in mice, *Sci. Rep.* 10 (1) (2020) 1–15.
- [121] H. Abdeltawab, F. Khalifa, K. Hammouda, J.M. Miller, M.M. Meki, Q. Ou, et al., Artificial intelligence based framework to quantify the cardiomyocyte structural integrity in heart slices, *Cardiovascular Eng. Technol.* (2021) 1–11.
- [122] F. Khalifa, G.M. Beache, A. Elnakib, H. Sliman, G. Gimel'farb, K.C. Welch, et al., A new shape-based framework for the left ventricle wall segmentation from cardiac first-pass perfusion MRI, in: *Proceedings of IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI'13)*, San Francisco, CA, 7–11 April 2013, pp. 41–44.
- [123] F. Khalifa, G.M. Beache, A. Elnakib, H. Sliman, G. Gimel'farb, K.C. Welch, et al., A new nonrigid registration framework for improved visualization of transmural perfusion gradients on cardiac first-pass perfusion MRI, in: *Proceedings of IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI'12)*, Barcelona, Spain, 2–5 May 2012, pp. 828–831.
- [124] F. Khalifa, G.M. Beache, A. Firjani, K.C. Welch, G. Gimel'farb, A. El-Baz, A new nonrigid registration approach for motion correction of cardiac first-pass perfusion MRI, in: *Proceedings of IEEE International Conference on Image Processing (ICIP'12)*, Lake Buena Vista, Florida, 30 September–3 October 2012, pp. 1665–1668.
- [125] F. Khalifa, G.M. Beache, G. Gimel'farb, A. El-Baz, A novel CAD system for analyzing cardiac first-pass MR images, in: *Proceedings of IAPR International*

- Conference on Pattern Recognition (ICPR' 12), Tsukuba Science City, Japan, 11–15 November 2012, pp. 77–80.
- [126] F. Khalifa, G.M. Beache, G. Gimel'farb, A. El-Baz, A novel approach for accurate estimation of left ventricle global indexes from short-axis cine MRI, in: Proceedings of IEEE International Conference on Image Processing (ICIP'11), Brussels, Belgium, 11–14 September 2011, pp. 2645–2649.
- [127] F. Khalifa, G.M. Beache, G. Gimel'farb, G.A. Giridharan, A. El-Baz, A new image-based framework for analyzing cine images, in: A. El-Baz, U.R. Acharya, M. Mirmedhdi, J.S. Suri (Eds.), *Handbook of Multi-Modality State-of-the-Art Medical Image Segmentation and Registration Methodologies*, Vol. 2, Springer, New York, 2011, pp. 69–98.
- [128] F. Khalifa, G.M. Beache, G. Gimel'farb, G.A. Giridharan, A. El-Baz, Accurate automatic analysis of cardiac cine images, *IEEE Trans. Biomed. Eng.* 59 (2) (2012) 445–455.
- [129] F. Khalifa, G.M. Beache, M. Nitzken, G. Gimel'farb, G.A. Giridharan, A. El-Baz, Automatic analysis of left ventricle wall thickness using short-axis cine CMR images, in: Proceedings of IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI'11), Chicago, Illinois, 30 March–2 April 2011, pp. 1306–1309.
- [130] M. Nitzken, G. Beache, A. Elnakib, F. Khalifa, G. Gimel'farb, A. El-Baz, Accurate modeling of tagged CMR 3D image appearance characteristics to improve cardiac cycle strain estimation, in: Proceedings of the Nineteenth IEEE International Conference on Image Processing (ICIP), Orlando, Florida, USA, IEEE, Sep. 2012, pp. 521–524.
- [131] M. Nitzken, G. Beache, A. Elnakib, F. Khalifa, G. Gimel'farb, A. El-Baz, Improving full-cardiac cycle strain estimation from tagged CMR by accurate modeling of 3D image appearance characteristics, in: Proceedings of the International Symposium on Biomedical Imaging (ISBI), 2012 Ninth IEEE. Barcelona, Spain, IEEE, May 2012, pp. 462–465.
- [132] M.J. Nitzken, A.S. El-Baz, G.M. Beache, Markov-gibbs random field model for improved full-cardiac cycle strain estimation from tagged cmr, *J. Cardiovasc. Magn. Reson.* 14 (1) (2012) 1–2.
- [133] H. Sliman, A. Elnakib, G. Beache, A. Elmaghraby, A. El-Baz, Assessment of myocardial function from cine cardiac MRI using a novel 4D tracking approach, *J. Comput. Sci. Syst. Biol.* 7 (2014) 169–173.
- [134] H. Sliman, A. Elnakib, G.M. Beache, A. Soliman, F. Khalifa, G. Gimel'farb, et al., A novel 4D PDE-based approach for accurate assessment of myocardium function using cine cardiac magnetic resonance images, in: Proceedings of IEEE International Conference on Image Processing (ICIP'14), Paris, France, 27–30 October 2014, pp. 3537–3541.
- [135] H. Sliman, F. Khalifa, A. Elnakib, G.M. Beache, A. Elmaghraby, A. El-Baz, A new segmentation-based tracking framework for extracting the left ventricle cavity from cine cardiac MRI, in: Proceedings of IEEE International Conference on Image Processing (ICIP'13), Melbourne, Australia, 15–18 September 2013, pp. 685–689.
- [136] H. Sliman, F. Khalifa, A. Elnakib, A. Soliman, G.M. Beache, A. Elmaghraby, et al., Myocardial borders segmentation from cine MR images using bi-directional coupled parametric deformable models, *Med. Phys.* 40 (9) (2013) 1–13.

- [137] H. Sliman, F. Khalifa, A. Elnakib, A. Soliman, G.M. Beache, G. Gimel'farb, et al., Accurate segmentation framework for the left ventricle wall from cardiac cine MRI, in: Proceedings of International Symposium on Computational Models for Life Science (CMLS'13), vol. 1559, Sydney, Australia, 27–29 November 2013, pp. 287–296.
- [138] A. Sharafeldein, M. Elsharkawy, N.S. Alghamdi, A. Soliman, A. El-Baz, Precise segmentation of covid-19 infected lung from ct images based on adaptive first-order appearance model with morphological/anatomical constraints, *Sensors* 21 (16) (2021) 5482.
- [139] M. Elsharkawy, A. Sharafeldein, F. Taher, A. Shalaby, A. Soliman, A. Mahmoud, et al., Early assessment of lung function in coronavirus patients using invariant markers from chest x-rays images, *Sci. Rep.* 11 (1) (2021) 1–11.
- [140] B. Abdollahi, A.C. Civelek, X.-F. Li, J. Suri, A. El-Baz, PET/CT nodule segmentation and diagnosis: A survey, in: L. Saba, J.S. Suri (Eds.), *Multi Detector CT Imaging*, Taylor, Francis, 2014, pp. 639–651.
- [141] B. Abdollahi, A. El-Baz, A.A. Amini, A multi-scale non-linear vessel enhancement technique, in: Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE. 2011, pp. 3925–3929.
- [142] B. Abdollahi, A. Soliman, A. Civelek, X.-F. Li, G. Gimel'farb, A. El-Baz, A novel gaussian scale space-based joint MGRF framework for precise lung segmentation, in: Proceedings of IEEE International Conference on Image Processing (ICIP'12), IEEE, 2012, pp. 2029–2032.
- [143] B. Abdollahi, A. Soliman, A. Civelek, X.-F. Li, G. Gimel'farb, A. El-Baz, A novel 3D joint MGRF framework for precise lung segmentation, *Machine Learning in Medical Imaging*, Springer, 2012, pp. 86–93.
- [144] A.M. Ali, A.S. El-Baz, A.A. Farag, A novel framework for accurate lung segmentation using graph cuts, in: Proceedings of IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI'07), 2007, pp. 908–911.
- [145] A. El-Baz, G.M. Beache, G. Gimel'farb, K. Suzuki, K. Okada, Lung imaging data analysis, *Int. J. Biomed. Imaging* 2013 (2013) 1–2.
- [146] A. El-Baz, G.M. Beache, G. Gimel'farb, K. Suzuki, K. Okada, A. Elnakib, Computer-aided diagnosis systems for lung cancer: challenges and methodologies, *Int. J. Biomed. Imaging* 2013 (2013) 1–46.
- [147] A. El-Baz, A. Elnakib, M. Abou El-Ghar, G. Gimel'farb, R. Falk, A. Farag, Automatic detection of 2D and 3D lung nodules in chest spiral CT scans, *Int. J. Biomed. Imaging* 2013 (2013) 1–11.
- [148] A. El-Baz, A.A. Farag, R. Falk, R. La Rocca, A unified approach for detection, visualization, and identification of lung abnormalities in chest spiral CT scans, *International Congress Series*, Vol. 1256, Elsevier, 2003, pp. 998–1004.
- [149] A. El-Baz, A.A. Farag, R. Falk, R. La Rocca, Detection, visualization and identification of lung abnormalities in chest spiral CT scan: phase-lin Proc. Int. Conf. Biomed. Engineering, Cairo, Egypt. 12 (1) (2002).
- [150] A. El-Baz, A. Farag, G. Gimel'farb, R. Falk, M.A. El-Ghar, T. Eldiasty, A framework for automatic segmentation of lung nodules from low dose chest CT scans, *Proc. Int. Conf. Pattern Recognition, (ICPR'06)* 3 (2006) 611–614. IEEE.
- [151] A. El-Baz, A. Farag, G. Gimel'farb, R. Falk, M.A. El-Ghar, A novel level set-based computer-aided detection system for automatic detection of lung nodules in low dose chest computed tomography scans, *Lung Imaging Computer Aided Diagn.* 10 (2011) 221–238.

- [152] A. El-Baz, G. Gimel'farb, M. Abou El-Ghar, R. Falk, Appearance-based diagnostic system for early assessment of malignant lung nodules, in: Proceedings of IEEE International Conference on Image Processing (ICIP'12), IEEE, 2012, pp. 533–536.
- [153] A. El-Baz, G. Gimel'farb, R. Falk, A novel 3D framework for automatic lung segmentation from low dose CT images, in: A. El-Baz, J.S. Suri (Eds.), *Lung Imaging and Computer Aided Diagnosis*, Taylor, Francis, 2011, pp. 1–16.
- [154] A. El-Baz, G. Gimel'farb, R. Falk, M. El-Ghar, Appearance analysis for diagnosing malignant lung nodules, in: Proceedings of IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI'10), IEEE, 2010, pp. 193–196.
- [155] A. El-Baz, G. Gimel'farb, R. Falk, M.A. El-Ghar, A novel level set-based CAD system for automatic detection of lung nodules in low dose chest CT scans, in: A. El-Baz, J.S. Suri (Eds.), *Lung Imaging and Computer Aided Diagnosis*, Vol. 1, Taylor, Francis, 2011, pp. 221–238.
- [156] A. El-Baz, G. Gimel'farb, R. Falk, M.A. El-Ghar, A new approach for automatic analysis of 3D low dose CT images for accurate monitoring the detected lung nodules, in: Proceedings of International Conference on Pattern Recognition (ICPR'08), IEEE, 2008, pp. 1–4.
- [157] A. El-Baz, G. Gimel'farb, R. Falk, M.A. El-Ghar, A novel approach for automatic follow-up of detected lung nodules, in: Proceedings of IEEE International Conference on Image Processing (ICIP'07), vol. 5, IEEE, 2007, pp. V–501.
- [158] A. El-Baz, G. Gimel'farb, R. Falk, M.A. El-Ghar, A new CAD system for early diagnosis of detected lung nodules, in: IEEE International Conference on Image Processing, 2007, ICIP 2007, vol. 2. IEEE, 2007, pp. II–461.
- [159] A. El-Baz, G. Gimel'farb, R. Falk, M.A. El-Ghar, H. Refaie, Promising results for early diagnosis of lung cancer, in: Proceedings of IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI'08), IEEE, 2008, pp. 1151–1154.
- [160] A. El-Baz, G.L. Gimel'farb, R. Falk, M. Abou El-Ghar, T. Holland, T. Shaffer, A new stochastic framework for accurate lung segmentation, in: Proceedings of Medical Image Computing and Computer-Assisted Intervention (MICCAI'08), 2008, pp. 322–330.
- [161] A. El-Baz, G.L. Gimel'farb, R. Falk, D. Heredis, M. Abou El-Ghar, A novel approach for accurate estimation of the growth rate of the detected lung nodules, in: Proceedings of International Workshop on Pulmonary Image Analysis, 2008, pp. 33–42.
- [162] A. El-Baz, G.L. Gimel'farb, R. Falk, T. Holland, T. Shaffer, A framework for unsupervised segmentation of lung tissues from low dose computed tomography images, in: Proceedings of British Machine Vision (BMVC'08), 2008, pp. 1–10.
- [163] A. El-Baz, G. Gimel'farb, R. Falk, M.A. El-Ghar, 3D MGRF-based appearance modeling for robust segmentation of pulmonary nodules in 3D LDCT chest images, in: *Lung Imaging and Computer Aided Diagnosis*, chapter, 2011, ch. 3, pp. 51–63.
- [164] A. El-Baz, G. Gimel'farb, R. Falk, M.A. El-Ghar, Automatic analysis of 3D low dose CT images for early diagnosis of lung cancer, *Pattern Recognit.* 42 (6) (2009) 1041–1051.
- [165] A. El-Baz, G. Gimel'farb, R. Falk, M.A. El-Ghar, S. Rainey, D. Heredia, et al., Toward early diagnosis of lung cancer, in: Proceedings of Medical Image Computing and Computer-Assisted Intervention (MICCAI'09), Springer, 2009, pp. 682–689.

- [166] A. El-Baz, G. Gimel'farb, R. Falk, M.A. El-Ghar, J. Suri, Appearance analysis for the early assessment of detected lung nodules, in: Lung Imaging and Computer Aided Diagnosis, chapter, 2011, ch. 17, pp. 395–404.
- [167] A. El-Baz, F. Khalifa, A. Elnakib, M. Nitzken, A. Soliman, P. McClure, et al., A novel approach for global lung registration using 3D Markov Gibbs appearance model, in: Proceedings of International Conference Medical Image Computing and Computer-Assisted Intervention (MICCAI' 12), Nice, France, 1–5 October 2012, pp. 114–121.
- [168] A. El-Baz, M. Nitzken, A. Elnakib, F. Khalifa, G. Gimel'farb, R. Falk, et al., 3D shape analysis for early diagnosis of malignant lung nodules, in: Proceedings of International Conference Medical Image Computing and Computer-Assisted Intervention (MICCAI' 11), Toronto, Canada, 18–22 September 2011, pp. 175–182.
- [169] A. El-Baz, M. Nitzken, G. Gimel'farb, E. Van Bogaert, R. Falk, M.A. El-Ghar, et al., Three-dimensional shape analysis using spherical harmonics for early assessment of detected lung nodules, in: Lung Imaging and Computer Aided Diagnosis, Chapter, 2011, ch. 19, pp. 421–438.
- [170] A. El-Baz, M. Nitzken, F. Khalifa, A. Elnakib, G. Gimel'farb, R. Falk, et al., 3D shape analysis for early diagnosis of malignant lung nodules, in: Proceedings of International Conference on Information Processing in Medical Imaging (IPMI'11), Monastery Irsee, Germany (Bavaria), 3–8 July 2011, pp. 772–783.
- [171] A. El-Baz, M. Nitzken, E. Vanbogaert, G. Gimel'Farb, R. Falk, M. Abo El-Ghar, A novel shape-based diagnostic approach for early diagnosis of lung nodules, in: Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on, IEEE, 2011, pp. 137–140.
- [172] A. El-Baz, P. Sethu, G. Gimel'farb, F. Khalifa, A. Elnakib, R. Falk, et al., Elastic phantoms generated by microfluidics technology: validation of an imaged-based approach for accurate measurement of the growth rate of lung nodules, *Biotechnol. J.* 6 (2) (2011) 195–203.
- [173] A. El-Baz, P. Sethu, G. Gimel'farb, F. Khalifa, A. Elnakib, R. Falk, et al., A new validation approach for the growth rate measurement using elastic phantoms generated by state-of-the-art microfluidics technology, in: Proceedings of IEEE International Conference on Image Processing (ICIP'10), Hong Kong, 26–29 September 2010, pp. 4381–4383.
- [174] A. El-Baz, P. Sethu, G. Gimel'farb, F. Khalifa, A. Elnakib, R. Falk, et al., Validation of a new imaged-based approach for the accurate estimating of the growth rate of detected lung nodules using real CT images and elastic phantoms generated by state-of-the-art microfluidics technology, in: A. El-Baz, J.S. Suri (Eds.), *Handbook of Lung Imaging and Computer Aided Diagnosis*, Vol. 1, Taylor & Francis, New York, 2011, pp. 405–420.
- [175] A. El-Baz, A. Soliman, P. McClure, G. Gimel'farb, M.A. El-Ghar, R. Falk, Early assessment of malignant lung nodules based on the spatial analysis of detected lung nodules, in: Proceedings of IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI' 12), IEEE, 2012, pp. 1463–1466.
- [176] A. El-Baz, S.E. Yuksel, S. Elshazly, A.A. Farag, Non-rigid registration techniques for automatic follow-up of lung nodules, *Proceedings of Computer Assisted Radiology and Surgery*, (CARS'05), Vol. 1281, Elsevier, 2005, pp. 1115–1120.

- [177] A.S. El-Baz, J.S. Suri, *Lung Imaging and Computer Aided Diagnosis*, CRC Press, 2011.
- [178] A. Soliman, F. Khalifa, N. Dunlap, B. Wang, M. El-Ghar, A. El-Baz, An iso-surfaces based local deformation handling framework of lung tissues, in: *Biomedical Imaging (ISBI), 2016 IEEE Thirteenth International Symposium on*, IEEE, 2016, pp. 1253–1259.
- [179] A. Soliman, F. Khalifa, A. Shaffie, N. Dunlap, B. Wang, A. Elmaghraby, et al., Detection of lung injury using 4D-CT chest images, in: *Biomedical Imaging (ISBI), 2016 IEEE Thirteenth International Symposium on*, IEEE, 2016, pp. 1274–1277.
- [180] A. Soliman, F. Khalifa, A. Shaffie, N. Dunlap, B. Wang, A. Elmaghraby, et al., A comprehensive framework for early assessment of lung injury, in: *IEEE International Conference on Image Processing (ICIP)*, IEEE, 2017, pp. 3275–3279.
- [181] A. Shaffie, A. Soliman, M. Ghazal, F. Taher, N. Dunlap, B. Wang, et al., A new framework for incorporating appearance and shape features of lung nodules for precise diagnosis of lung cancer, in: *IEEE International Conference on Image Processing (ICIP)*, IEEE, 2017, pp. 1372–1376.
- [182] A. Soliman, F. Khalifa, A. Shaffie, N. Liu, N. Dunlap, B. Wang, et al., Image-based cad system for accurate identification of lung injury, in: *IEEE International Conference on Image Processing (ICIP)*, IEEE, 2016, pp. 121–125.
- [183] A. Soliman, A. Shaffie, M. Ghazal, G. Gimel'farb, R. Keynton, A. El-Baz, A novel cnn segmentation framework based on using new shape and appearance features, in: *25th IEEE International Conference on Image Processing (ICIP)*, IEEE, 2018, pp. 3488–3492.
- [184] A. Shaffie, A. Soliman, H.A. Khalifeh, M. Ghazal, F. Taher, R. Keynton, et al., On the integration of ct derived features for accurate detection of lung cancer, in: *IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, IEEE, 2018, pp. 435–440.
- [185] A. Shaffie, A. Soliman, H.A. Khalifeh, M. Ghazal, F. Taher, A. Elmaghraby, et al., Radiomic-based framework for early diagnosis of lung cancer, in: *IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, IEEE, 2019, pp. 1293–1297.
- [186] A. Shaffie, A. Soliman, M. Ghazal, F. Taher, N. Dunlap, B. Wang, et al., A novel autoencoder-based diagnostic system for early assessment of lung cancer, in: *25th IEEE International Conference on Image Processing (ICIP)*, IEEE, 2018, pp. 1393–1397.
- [187] A. Shaffie, A. Soliman, L. Fraiwan, M. Ghazal, F. Taher, N. Dunlap, et al., A generalized deep learning-based diagnostic system for early diagnosis of various types of pulmonary nodules, *Technol. Cancer Res. Treat.* 17 (2018). 1533033818798800.
- [188] A.A.K. Abdel Razek, A. Alksas, M. Shehata, A. AbdelKhalek, K. Abdel Baky, A. El-Baz, et al., Clinical applications of artificial intelligence and radiomics in neuro-oncology imaging, *Insights into Imaging* 12 (1) (2021) 1–17.
- [189] Y. ElNakieb, M.T. Ali, O. Dekhil, M.E. Khalefa, A. Soliman, A. Shalaby, et al., Towards accurate personalized autism diagnosis using different imaging modalities: SMRI, FMRI, and DTI, in: *IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, IEEE, 2018, pp. 447–452.
- [190] Y. ElNakieb, A. Soliman, A. Mahmoud, O. Dekhil, A. Shalaby, M. Ghazal, et al., Autism spectrum disorder diagnosis framework using diffusion tensor imaging, in: *IEEE International Conference on Imaging Systems and Techniques (IST)*, IEEE, 2019, pp. 1–5.

- [191] R. Haweel, O. Dekhil, A. Shalaby, A. Mahmoud, M. Ghazal, R. Keynton, et al., A machine learning approach for grading autism severity levels using task-based functional MRI, in: IEEE International Conference on Imaging Systems and Techniques (IST), IEEE, 2019, pp. 1–5.
- [192] O. Dekhil, M. Ali, R. Haweel, Y. Elnakib, M. Ghazal, H. Hajjdiab, et al., A comprehensive framework for differentiating autism spectrum disorder from neurotypicals by fusing structural mri and resting state functional mri, *Seminars in Pediatric Neurology*, Elsevier, 2020, p. 100805.
- [193] R. Haweel, O. Dekhil, A. Shalaby, A. Mahmoud, M. Ghazal, A. Khalil, et al., A novel framework for grading autism severity using task-based FMRI, in: IEEE 17th International Symposium on Biomedical Imaging (ISBI), IEEE, 2020, pp. 1404–1407.
- [194] A. El-Baz, A. Elnakib, F. Khalifa, M.A. El-Ghar, P. McClure, A. Soliman, et al., Precise segmentation of 3-D magnetic resonance angiography, *IEEE Trans. Biomed. Eng.* 59 (7) (2012) 2019–2029.
- [195] A. El-Baz, A. Farag, A. Elnakib, M.F. Casanova, G. Gimel'farb, A.E. Switala, et al., Accurate automated detection of autism related corpus callosum abnormalities, *J. Med. Syst.* 35 (5) (2011) 929–939.
- [196] A. El-Baz, G. Gimel'farb, R. Falk, M.A. El-Ghar, V. Kumar, D. Heredia, A novel 3D joint Markov-gibbs model for extracting blood vessels from PC–MRA images, *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2009*, Vol. 5762, Springer, 2009, pp. 943–950.
- [197] A. Elnakib, A. El-Baz, M.F. Casanova, G. Gimel'farb, A.E. Switala, Image-based detection of corpus callosum variability for more accurate discrimination between dyslexic and normal brains, in: Proc. IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI'2010), IEEE, 2010, pp. 109–112.
- [198] A. Elnakib, M.F. Casanova, G. Gimel'farb, A.E. Switala, A. El-Baz, Autism diagnostics by centerline-based shape analysis of the corpus callosum, in: Proc. IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI' 2011), IEEE, 2011, pp. 1843–1846.
- [199] A. Elnakib, M. Nitzken, M. Casanova, H. Park, G. Gimel'farb, A. El-Baz, Quantification of age-related brain cortex change using 3D shape analysis, in: Pattern Recognition (ICPR), 2012 21st International Conference on. IEEE, 2012, pp. 41–44.
- [200] M. Nitzken, M. Casanova, G. Gimel'farb, A. Elnakib, F. Khalifa, A. Switala, et al., 3D shape analysis of the brain cortex with application to dyslexia, in: Image Processing (ICIP), 2011 18th IEEE International Conference on. Brussels, Belgium: IEEE, Sep. 2011, pp. 2657–2660.
- [201] F.E.-Z.A. El-Gamal, M.M. Elmogy, M. Ghazal, A. Atwan, G.N. Barnes, M.F. Casanova, et al., A novel cad system for local and global early diagnosis of alzheimer's disease based on pib-pet scans, in: 2017 IEEE International Conference on Image Processing (ICIP), IEEE, 2017, pp. 3270–3274.
- [202] M.M. Ismail, R.S. Keynton, M.M. Mostapha, A.H. ElTanboly, M.F. Casanova, G.L. Gimel'farb, et al., Studying autism spectrum disorder with structural and diffusion magnetic resonance imaging: a survey, *Front. Hum. Neurosci.* 10 (2016) 211.
- [203] A. Alansary, M. Ismail, A. Soliman, F. Khalifa, M. Nitzken, A. Elnakib, et al., Infant brain extraction in T1-weighted MR images using BET and refinement using LCDG and MGRF models, *IEEE J. Biomed. Health Inform.* 20 (3) (2016) 925–935.

- [204] E.H. Asl, M. Ghazal, A. Mahmoud, A. Aslantas, A. Shalaby, M. Casanova, et al., Alzheimer's disease diagnostics by a 3d deeply supervised adaptable convolutional network, *Front. Biosci. (Landmark Ed.)* 23 (2018) 584–596.
- [205] O. Dekhil, M. Ali, Y. El-Nakieb, A. Shalaby, A. Soliman, A. Switala, et al., A personalized autism diagnosis cad system using a fusion of structural mri and resting-state functional mri data, *Front. Psychiatry* 10 (392) (2019). Available from: <https://www.frontiersin.org/article/10.3389/fpsy.2019.00392>.
- [206] O. Dekhil, A. Shalaby, A. Soliman, A. Mahmoud, M. Kong, G. Barnes, et al., Identifying brain areas correlated with ados raw scores by studying altered dynamic functional connectivity patterns, *Med. Image Anal.* 68 (2021) 101899.
- [207] Y.A. Elnakieb, M.T. Ali, A. Soliman, A.H. Mahmoud, A.M. Shalaby, N.S. Alghamdi, et al., Computer aided autism diagnosis using diffusion tensor imaging, *IEEE Access* 8 (2020) 191. 298–191 308.
- [208] M.T. Ali, Y.A. Elnakieb, A. Shalaby, A. Mahmoud, A. Switala, M. Ghazal, et al., Autism classification using SMRI: a recursive features selection based on sampling from multi-level high dimensional spaces, in: 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI), IEEE, 2021, pp. 267–270.
- [209] M.T. Ali, Y. ElNakieb, A. Elnakib, A. Shalaby, A. Mahmoud, M. Ghazal, et al., The role of structure mri in diagnosing autism, *Diagnostics* 12 (1) (2022) 165.
- [210] Y. ElNakieb, M.T. Ali, A. Elnakib, A. Shalaby, A. Soliman, A. Mahmoud, et al., The role of diffusion tensor mr imaging (dti) of the brain in diagnosing autism spectrum disorder: Promising results, *Sensors* 21 (24) (2021) 8171.
- [211] A. Mahmoud, A. El-Barkouky, H. Farag, J. Graham, A. Farag, A non-invasive method for measuring blood flow rate in superficial veins from a single thermal image, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2013, pp. 354–359.
- [212] N. Elsaid, A. Saied, H. Kandil, A. Soliman, F. Taher, M. Hadi, et al., Impact of stress and hypertension on the cerebrovasculature, *Front. Bioscience-Landmark* 26 (12) (2021) 1643.
- [213] F. Taher, H. Kandil, Y. Gebru, A. Mahmoud, A. Shalaby, S. El-Mashad, et al., A novel mra-based framework for segmenting the cerebrovascular system and correlating cerebral vascular changes to mean arterial pressure, *Appl. Sci.* 11 (9) (2021) 4022.
- [214] H. Kandil, A. Soliman, F. Taher, M. Ghazal, A. Khalil, G. Giridharan, et al., A novel computer aided diagnosis system for the early detection of hypertension based on cerebrovascular alterations, *NeuroImage Clin.* 25 (2020) 102107.
- [215] H. Kandil, A. Soliman, M. Ghazal, A. Mahmoud, A. Shalaby, R. Keynton, et al., A novel framework for early detection of hypertension using magnetic resonance angiography, *Sci. Rep.* 9 (1) (2019) 1–12.
- [216] Y. Gebru, G. Giridharan, M. Ghazal, A. Mahmoud, A. Shalaby, A. El-Baz, *Detection of cerebrovascular changes using magnetic resonance angiography, Cardiovascular Imaging and Image Analysis*, CRC Press, 2018, pp. 1–22.
- [217] A. Mahmoud, A. Shalaby, F. Taher, M. El-Baz, J.S. Suri, A. El-Baz, *Vascular tree segmentation from different image modalities, Cardiovascular Imaging and Image Analysis*, CRC Press, 2018, pp. 43–70.
- [218] F. Taher, A. Mahmoud, A. Shalaby, A. El-Baz, A review on the cerebrovascular segmentation methods, in: *IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, IEEE, 2018, pp. 359–364.

- [219] H. Kandil, A. Soliman, L. Fraiwan, A. Shalaby, A. Mahmoud, A. ElTanboly, et al., A novel MRA framework based on integrated global and local analysis for accurate segmentation of the cerebral vascular system, in: IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), IEEE, 2018, pp. 1365–1368.
- [220] F. Taher, A. Soliman, H. Kandil, A. Mahmoud, A. Shalaby, G. Gimel'farb, et al., Accurate segmentation of cerebrovasculature from tof-mra images using appearance descriptors, IEEE Access, 2020.
- [221] F. Taher, A. Soliman, H. Kandil, A. Mahmoud, A. Shalaby, G. Gimel'farb, et al., Precise cerebrovascular segmentation, in: IEEE International Conference on Image Processing (ICIP), IEEE, 2020, pp. 394–397.
- [222] M. Elsharkawy, A. Sharafeldeen, A. Soliman, F. Khalifa, M. Ghazal, E. El-Daydamony, et al., A novel computer-aided diagnostic system for early detection of diabetic retinopathy using 3D-oct higher-order spatial appearance model, Diagnostics 12 (2) (2022) 461.
- [223] M. Elsharkawy, M. Elrazzaz, M. Ghazal, M. Alhalabi, A. Soliman, A. Mahmoud, et al., Role of optical coherence tomography imaging in predicting progression of age-related macular disease: a survey, Diagnostics 11 (12) (2021) 2313.
- [224] H.S. Sandhu, M. Elmogy, A.T. Sharafeldeen, M. Elsharkawy, N. El-Adawy, A. Eltanboly, et al., Automated diagnosis of diabetic retinopathy using clinical biomarkers, optical coherence tomography, and optical coherence tomography angiography, Am. J. Ophthalmol. 216 (2020) 201–206.
- [225] A. Sharafeldeen, M. Elsharkawy, F. Khalifa, A. Soliman, M. Ghazal, M. AlHalabi, et al., Precise higher-order reflectivity and morphology models for early diagnosis of diabetic retinopathy using oct images, Sci. Rep. 11 (1) (2021) 1–16.
- [226] A.A. Sleman, A. Soliman, M. Elsharkawy, G. Giridharan, M. Ghazal, H. Sandhu, et al., A novel 3d segmentation approach for extracting retinal layers from optical coherence tomography images, Med. Phys. 48 (4) (2021) 1584–1595.
- [227] A.A. Sleman, A. Soliman, M. Ghazal, H. Sandhu, S. Schaal, A. Elmaghraby, et al., Retinal layers oct scans 3-D segmentation, in: IEEE International Conference on Imaging Systems and Techniques (IST), IEEE, 2019, pp. 1–6.
- [228] N. Eladawi, M. Elmogy, M. Ghazal, O. Helmy, A. Aboelfetouh, A. Riad, et al., Classification of retinal diseases based on oct images, Front. Biosci. (Landmark Ed.) 23 (2018) 247–264.
- [229] A. ElTanboly, M. Ismail, A. Shalaby, A. Switala, A. El-Baz, S. Schaal, et al., A computer-aided diagnostic system for detecting diabetic retinopathy in optical coherence tomography images, Med. Phys. 44 (3) (2017) 914–923.
- [230] H.S. Sandhu, A. El-Baz, J.M. Seddon, Progress in automated deep learning for macular degeneration, JAMA Ophthalmol. (2018).
- [231] M. Ghazal, S.S. Ali, A.H. Mahmoud, A.M. Shalaby, A. El-Baz, Accurate detection of non-proliferative diabetic retinopathy in optical coherence tomography images using convolutional neural networks, IEEE Access 8 (2020) 34. 387–34 397.
- [232] K. Hammouda, F. Khalifa, A. Soliman, M. Ghazal, M. Abou El-Ghar, A. Haddad, et al., A cnn-based framework for bladder wall segmentation using mri, in: Fifth International Conference on Advances in Biomedical Engineering (ICABME), IEEE, 2019, pp. 1–4.

- [233] K. Hammouda, F. Khalifa, A. Soliman, M. Ghazal, M. Abou El-Ghar, A. Haddad, et al., A deep learning-based approach for accurate segmentation of bladder wall using mr images, in: IEEE International Conference on Imaging Systems and Techniques (IST), IEEE, 2019, pp. 1–6.
- [234] K. Hammouda, F. Khalifa, A. Soliman, H. Abdeltawab, M. Ghazal, M. Abou El-Ghar, et al., A 3d cnn with a learnable adaptive shape prior for accurate segmentation of bladder wall using MR images, in: IEEE 17th International Symposium on Biomedical Imaging (ISBI), IEEE, 2020, pp. 935–938.
- [235] K. Hammouda, F. Khalifa, A. Soliman, M. Ghazal, M. Abou El-Ghar, M. Badawy, et al., A multiparametric mri-based cad system for accurate diagnosis of bladder cancer staging, *Comput. Med. Imaging Graph.* 90 (2021) 101911.
- [236] K. Hammouda, F. Khalifa, A. Soliman, M. Ghazal, M. Abou El-Ghar, M. Badawy, et al., A CAD system for accurate diagnosis of bladder cancer staging using a multiparametric MRI, in: IEEE 18th International Symposium on Biomedical Imaging (ISBI), IEEE, 2021, pp. 1718–1721.
- [237] A. Alksas, M. Shehata, G.A. Saleh, A. Shaffie, A. Soliman, M. Ghazal, et al., A novel computer-aided diagnostic system for early assessment of hepatocellular carcinoma, in: 25th International Conference on Pattern Recognition (ICPR), IEEE, 2021, pp. 10 375–10 382.
- [238] A. Alksas, M. Shehata, G.A. Saleh, A. Shaffie, A. Soliman, M. Ghazal, et al., A novel computer-aided diagnostic system for accurate detection and grading of liver tumors, *Sci. Rep.* 11 (1) (2021) 1–18.
- [239] A.A.K.A. Razek, R. Khaled, E. Helmy, A. Naglah, A. AbdelKhalek, A. El-Baz, Artificial intelligence and deep learning of head and neck cancer, *Magnetic Reson. Imaging Clin.* 30 (1) (2022) 81–94.
- [240] A. Sharafeldeen, M. Elsharkawy, R. Khaled, A. Shaffie, F. Khalifa, A. Soliman, et al., Texture and shape analysis of diffusion-weighted imaging for thyroid nodules classification using machine learning, *Med. Phys.* (2021).
- [241] A. Naglah, F. Khalifa, R. Khaled, A.A.K. Abdel Razek, M. Ghazal, G. Giridharan, et al., Novel mri-based cad system for early detection of thyroid cancer using multi-input CNN, *Sensors* 21 (11) (2021) 3878.
- [242] A. Naglah, F. Khalifa, A. Mahmoud, M. Ghazal, P. Jones, T. Murray, et al., Athlete-customized injury prediction using training load statistical records and machine learning, in: IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), IEEE, 2018, pp. 459–464.
- [243] A.H. Mahmoud, Utilizing radiation for smart robotic applications using visible, thermal, and polarization images, (Ph.D. dissertation), University of Louisville, 2014.
- [244] A. Mahmoud, A. El-Barkouky, J. Graham, A. Farag, Pedestrian detection using mixed partial derivative based histogram of oriented gradients, in: IEEE International Conference on Image Processing (ICIP), IEEE, 2014, pp. 2334–2337.
- [245] A. El-Barkouky, A. Mahmoud, J. Graham, A. Farag, An interactive educational drawing system using a humanoid robot and light polarization, in: IEEE International Conference on Image Processing, IEEE, 2013, pp. 3407–3411.
- [246] A.H. Mahmoud, M.T. El-Melegy, A.A. Farag, Direct method for shape recovery from polarization and shading, in: 19th IEEE International Conference on Image Processing, IEEE, 2012, pp. 1769–1772.

- [247] M.A. Ghazal, A. Mahmoud, A. Aslantas, A. Soliman, A. Shalaby, J.A. Benediktsson, et al., Vegetation cover estimation using convolutional neural networks, *IEEE Access* 7 (2019) 132. 563–132 576.
- [248] M. Ghazal, A. Mahmoud, A. Shalaby, A. El-Baz, Automated framework for accurate segmentation of leaf images for plant health assessment, *Environ. Monit. Assess* 191 (8) (2019) 491.
- [249] M. Ghazal, A. Mahmoud, A. Shalaby, S. Shaker, A. Khelifi, A. El-Baz, Precise statistical approach for leaf segmentation, in: *IEEE International Conference on Image Processing (ICIP)*, IEEE, 2020, pp. 2985–2989.

A review of texture-centric diagnostic models for thyroid cancer using convolutional neural networks and visualized texture patterns

12

Ahmed Naglah¹, Fahmi Khalifa¹, Reem Khaled², Ahmed Abdel Khalek Abdel Razeq², Mohammed Ghazal³, Guruprasad Giridharan¹, Ali Mahmoud¹ and Ayman S. El-Baz^{1,4}

¹*University of Louisville, Louisville, KY, United States*

²*Department of Diagnostic Radiology, Faculty of Medicine, Mansoura University, Mansoura, Egypt*

³*Electrical, Computer and Biomedical Engineering Department, Abu Dhabi University, Abu Dhabi, United Arab Emirates*

⁴*University of Louisville at Alamein International University (UofL-AIU), New Alamein City, Egypt*

12.1 Introduction

In the United States, around 52,890 new cases of thyroid cancer and around 2180 deaths were estimated in 2020 according to the American Cancer Society's recent figures [1]. The rate of thyroid nodules' development is nearly 5% in women and 1% in men in terms of prevalence [2]. 7%–15% of thyroid nodule cases develop into malignant tumors (cancer), and this percentage depends on different factors such as age, sex, radiation exposure history, and family history [2]. We can categorize thyroid cancer into three main categories: differentiated thyroid cancer (DTC), medullary thyroid cancer, and anaplastic thyroid cancer. DTC has a share of 90% among thyroid cancer. DTC includes two main subcategories: papillary thyroid carcinoma (PTC) and follicular thyroid carcinoma (FTC), where PTC accounts for the majority of the DTC cases [2].

Different procedures are required as part of the diagnostic criteria of thyroid cancer, and those procedures include physical examination, blood test, ultrasound (US) imaging, magnetic resonance imaging (MRI), biopsy, and surgical operation. As detection of relatively small nodules becomes achievable over time due to the current advances in medical imaging, diagnosing cancer (nodule malignancy) and early stratification of nodules are still challenging and still require examination of biopsy-extracted thyroid tissue [2]. Surgical extraction using biopsy is still the gold standard of clinical evaluation. However, this invasive procedure is costly and may introduce a false negative error depending on the biopsy technique and

the size of the nodule being aspirated [3–6]. Biopsy is usually performed either by fine-needle aspiration or by surgical excision of the nodule.

Different approaches for thyroid cancer diagnosis use noninvasive methods, and those approaches have been addressed by several researchers and studies to enhance the accuracy of thyroid cancer detection and stratification [7–10]. Examination and analysis of medical images are widely used within this domain, and the field of artificial intelligence (AI) contributes to this by enriching image analysis and detecting new biomarkers in medical data including medical images. The type of images used as feed to the AI algorithms can affect the accuracy and efficiency of the desired computer-aided diagnosis (CAD) system. US imaging is widely used as a first-line assessment of potential thyroid nodules [2], and certain markers of thyroid nodules in US imaging can be associated with an increased risk of malignancy. Yet, the appearance of those markers in US images is operator-dependent, and also many features in US images need to be considered simultaneously during the evaluation to provide sufficient malignancy diagnostic accuracy [2]. Those factors can introduce several effects and limitations to the AI-based systems that are designed to use US images for thyroid nodule stratification [7–9]. Other than US, MR imaging modalities have also been used in the research community. For example, T1-weighted and T2-weighted MRIs were used to perform thyroid nodule classification [10]. Certain MRI modalities can help differentiate between different molecular components in the tissue. For instance, fats appear brighter in T1-weighted MRI scans [11], while fluids appear brighter in T2-weighted MRI scans. Studying T2-weighted MRI scans can help in the modeling of fluid patterns in tissue structures [12]. On top of that, diffusion-weighted MRI (DWI) can model the diffusivity of fluids in the tissue by evaluating the different constraints that limit the fluid diffusion in different directions [13,14]. Therefore, DWI can generally simulate the fluid patterns (statics and dynamics) in the tissue, and these patterns can be presented by measuring the apparent diffusion coefficient (ADC) from the raw MRI scans (this will be illustrated in more details in this chapter).

The cell proliferation process within the tissue of malignant thyroid nodules can have significant effect on the patterns of the extracellular matrix in the thyroid tissue. Research studies suggest that statistical analysis between ADC value and T2-weighted images can classify between malignant and benign nodules [15–17], which indicates the role of ADC in modeling diffusivity within the tissue. Statistical analysis can be useful in studying the overall effect of cancer on diffusivity. However, the propagation of cancer across the tissue can be associated with other effects on the spatial patterns. The cell structure of the thyroid tissue takes the form of colloid follicles surrounded by cuboidal epithelial cells. This structure enables the different molecular exchange processes that occur in the thyroid gland to achieve the desired functionality. Cancer propagation can disturb this structure in different ways according to the subtype of cancer. For instance, PTC can be associated with distinctive change in the shape and size of nucleus, while FTC can have further changes on the follicle structure level. Studying radiomics, and in more specific way features in DWI, can have significant contribution in understanding the

differences between malignant and benign tissue, and also in understanding how can we differentiate between thyroid cancer subtypes.

AI plays an important role in analyzing medical images and in detecting biomarkers and features in DWIs. AI has different approaches to address problems according to the application, the type of data, and the prior information that we have on the data. Statistical machine learning methods can fit more in application with hand-crafted features or with low number of features [18,19], while neural networks and deep learning technologies can fit in application where we need to perform automatic detection of features from the input data [20–26]. Inspired by the biological neural network, artificial neural network (ANN) is a network of neurons that models the relationship between the input features (which represents the independent variables) with the output (which represents the dependent variable). Depending on the type of data we have, the type of ANN can be chosen and the network can be designed [27,28]. Convolutional neural network (CNN) is a subtype of ANN where we have convolutional layers that learn the suitable filters that can be used on the input data. A filter is a function that linearly combines adjacent input features. Accordingly, CNN can fit the application where we have data that have dependent features such as images, videos, and audio. As cancer propagation can have significant effects on spatial patterns within the tissue, CNN can be used to automatically detect voxel dependence in MRI scans.

In this chapter, we review two texture-based diagnostic models for thyroid cancer presented in our previous work [20,29]. We demonstrate a methodology for visualizing deep-learning generated texture patterns in both 2D and 3D. The two models act as automation approaches for CAD of thyroid nodules instead of the traditional approach, which is based on statistical analysis (see Fig. 12.1). In the first model, we

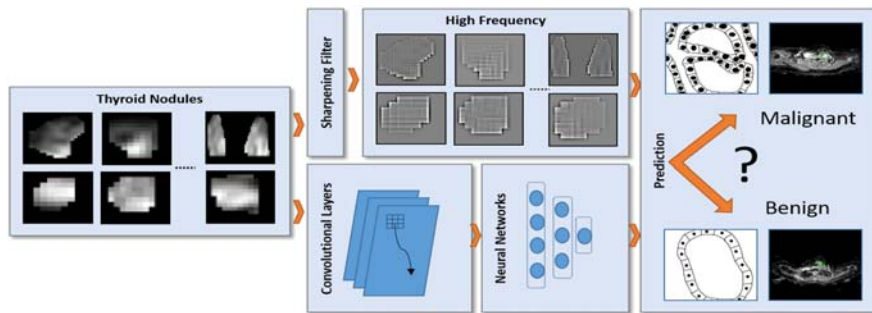


FIGURE 12.1

Illustrative diagram of the high-level approach used in the proposed models. Instead of using hand-crafted features that capture certain parameters from the MRI intensity or MRI high-frequency domains, we deploy CNN-based models that can learn texture patterns in each MRI modality. Capturing texture patterns in diffusion-weighted MRIs can help mapping structural differences between malignant and benign tumors to predict cancer as a way of bridging the gap between radiomics and microscopic domains.

explore the possibility to use 2D texture patterns in DWI to diagnose thyroid cancer, while in the second model we extend this to 3D to explore the possible enhancement in terms of accuracy and interpretability. In [Section 12.2](#), we review the material used to develop the presented models, and the protocols used for collecting data. In [Section 12.3](#), we review our exploratory work on the collected data using basic statistical analysis. In [Sections 12.4](#) and [12.5](#), we review the detailed methodology and consideration of the 2D and 3D texture models respectively. In [Section 12.7](#), we review the results obtained from the experiments performed to study the presented models. In [Section 12.8](#), we discuss the obtained results and the contribution of those models toward bridging the gap between radiomics and microscopic domains. [Section 12.9](#) is the conclusion of our study as a conclusion of the chapter.

12.2 Materials and collection protocols

12.2.1 Study participants and raw data collection

Data were collected in those studies from 49 patients with pathologically proven thyroid nodules. The age range was 25–70 years. Imaging of the thyroid gland was performed at Mansoura University, Egypt with a 1.5 T Ingenia MR scanner (Philips Medical Systems, Best, Netherlands) using a head/neck circular polarization surface coil. All participants were fully informed about the aims of the study and provided their informed consent. The inclusion criteria for the study were untreated patients with thyroid nodules whose malignancy status was unclear from ultrasound examination. Patients underwent thyroid core biopsy or surgery after MR imaging. Histopathologic diagnoses were provided by an experienced cytologist or pathologist. In total, there are 17 malignant nodules in 17 patients and 40 benign nodules in 32 patients included in our study.

DWI volumes that employ a multislice, single-shot, spin-echo, echo-planar imaging sequence with TR = 10,000 ms, TE = 108 ms, and 125 kHz bandwidth were collected and stored. Axial diffusion-weighted slices over the region of interests were 5 mm thickness with an inter-slice gap of 1 mm, 25 cm or 30 cm FOV, and 256×256 acquisition matrix. For DWI, a diffusion gradient was applied during the scanning process with b-values of $b = 500 \text{ s/mm}^2$, $b = 1000 \text{ s/mm}^2$, and $b = 1500 \text{ s/mm}^2$. T2-weighted images are measured using b-value of $b = 0 \text{ s/mm}^2$.

12.2.2 Nodule segmentation and apparent diffusion coefficient calculations

Different stages were implemented on the collected MR scans to prepare the dataset for the training model (see [Fig. 12.2](#)). Segmentation of thyroid nodules was implemented manually throughout our work. An experienced radiologist processed the segmentation of each nodule as it appeared in each T2-weighted slice

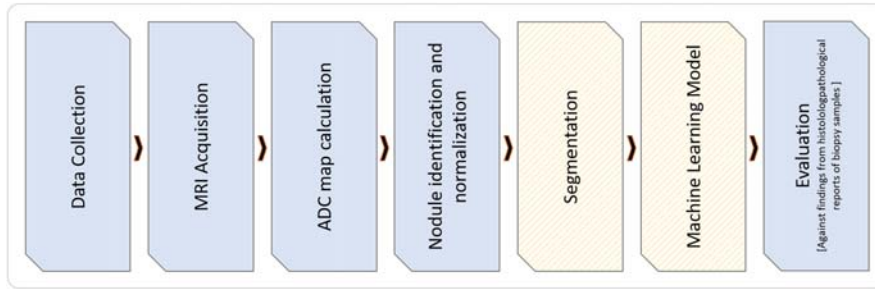


FIGURE 12.2

Schematic showing the pipeline used to train the proposed model. MRI data are collected from human subject cohort for developing and validating the model. Preprocessing steps such as segmentation and ADC maps were implemented to prepare inputs for the two-branch CNN. The objective of the proposed system was to learn the texture patterns in diffusion-weighted MRI images and correlate them with pathological results (as a ground truth for the diagnosis of nodules).

($b = 0 \text{ s/mm}^2$) and in each DWI slice. DWI scans were measured in the same session and using the same resolution, number of slices, and inter-slice gap. Accordingly, no registration was applied to align the different b-values. The produced manual segmentations were stored as binary images. Each binary image produced from DWI slice with $b = 0 \text{ s/mm}^2$ was reused during processing phases on the corresponding slice at all other b-values, and also was reused for the corresponding slice at ADC500, ADC1000, and ADC1500. We segmented each nodule in both T2-weighted images and ADC maps using a square-bounding box. We fixed the dimensions of the spatial domain by resizing extracted box into unified $48 \times 48 \times 20$ volumes by adding zero-padding channels. We then normalized the voxel-intensity in that volume to be in 0–1 range. Each segmented nodule was provided for the network model on a black background and padding. ADC maps were calculated at each non-zero b-value (500, 1000, and 1500) by combining the diffusion images at the corresponding b-value with the image at $b = 0 \text{ s/mm}^2$, and then we substituted, at the voxel level, this into the Stejskal–Tanner equation [30]. The generated images of this process are referred to as ADC500, ADC1000, and ADC1500. Since DWI as an absolute value usually does not reflect direct biological activity, the relative differences between DW-MRI at different b-values were used instead (i.e., ADC) to model the diffusivity in the tissue. Usually, a b-value of 0 is taken as reference, and therefore we computed three ADC values that correspond to 3 b-values of 500, 1000, and 1500 referenced to a b-value of 0.

For 2D model, in each slice, we calculate the size of the nodules' cross-section that appear in it. Then, we use the images of the slice that has higher nodules' footprint. On the other hand, regarding the 3D model, we use the whole volume as input for the 3D CNN used in that model (this CNN architecture will be illustrated later in this chapter). Having two models that address the same

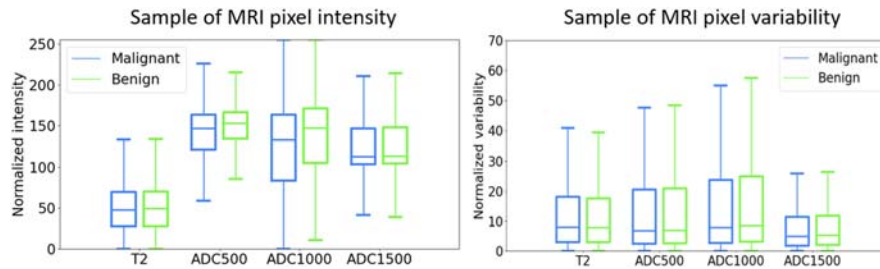


FIGURE 12.3

Example from the results obtained by the preliminary results that use statistical analysis between the malignant and benign groups. Box Plots represent the intensity level in the normal intensity MRI images (*left*) and the high-frequency version of the MRI images (*right*).

problem from different side usually helps in validating the results and providing enhanced interpretations.

12.3 Statistical analysis

As a preliminary analysis, we investigated if the intensity variations of the pixels can differentiate between malignant and benign groups (see Fig. 12.1). To evaluate that, we deployed a statistical analysis method to determine the differences between the two groups of patients as observed in each of the T2-weighted images and the ADC maps. Our analysis indicated a significant heterogeneity in intensity variation between T2-weighted and ADC maps, which supports the idea of feeding the T2-weighted images and the ADC maps each to a separate input layer of the CNN to enable learning independent textures in each branch and to enhance the accuracy of the desired system. Box plot representations are illustrated in Fig. 12.3. A high-pass filter using a 3×3 Laplacian filter invariant to 45 degrees rotations is used to model local intensity variations [31]. Welch two-sample t -test was implemented to evaluate the statistical significance and to evaluate the metrics presented in the results section. Voxels were classified into benign and malignant groups (35,625 and 15,764 pixels, respectively). Our choice of Welch two-sample t -test is supported by the high number of samples. R was used (statistical libraries in R) to generate the results.

12.4 2D texture model

In our previous studies [15,16], we propose two models that form novel CNN-based CAD systems that integrate T2-weighted images and ADC maps using a multiinput CNN network for thyroid nodules detection and classification,

see Figs. 12.4 and 12.5. In contrast to using different modalities of MRI with-including ADC maps as in [10], our model uses ADC maps as representative of cell density in the tissues [32]. Our model can enable searching for cancer biomarkers given cancer is associated with high rates of cell proliferation in many stages. With close relation to a study that uses multiparametric MRI radiomics as a predictive model [20,33], we use a CNN-based architecture instead of using hand-crafted features. We utilize a methodology of independent automatic feature extraction (or convolutions) of ADC and DWI before combining them within the dense fully connected layer. This approach boosts the feasibility to detect deep texture features from each modality without losing CNN automatic searching capability. The proposed system combines multiple ADC maps, i.e. from different gradient coefficients as a

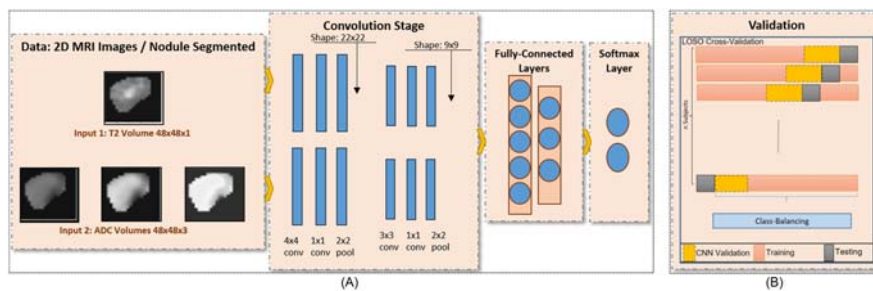


FIGURE 12.4

(A) Schematic of the proposed CNN-based diagnostic model (2D version) that presents the design and the layers of the multiinput 2D CNN deep learning framework. (B) Illustration for the cross-validation methodology used in our experiments.

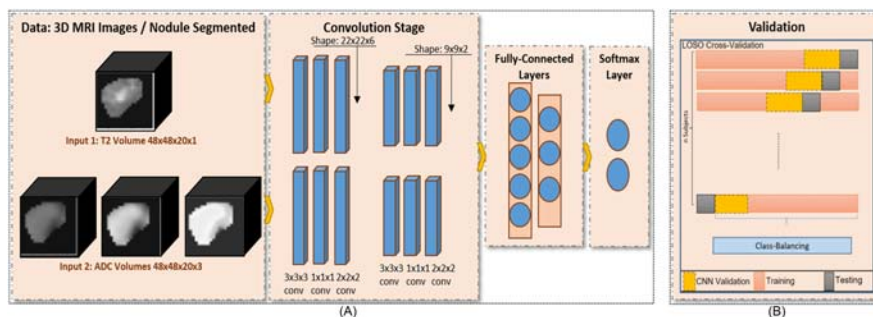


FIGURE 12.5

(A) Schematic of the proposed CNN-based diagnostic model (3D version) that presents the design and the layers of the multiinput 3D CNN deep-learning framework. (B) Illustration for the cross-validation methodology used in our experiments.

configurable parameter in the MRI scanner. As a multichannel input, the combination of all modalities is used by our CNN model to achieve enhanced learning of texture for more accurate diagnosis.

In this subsection specifically, we present the first model, which is based on 2D analysis of DWI using CNNs. The proposed architecture, shown in Fig. 12.4, consists of two branches. The two branches are identical in terms of the number of layers. The kernels generated by our architecture are influenced by combining T2 images and ADC images of the training samples during the learning process of the neural network. The convolution layers are composed of 4×4 conv (with 32 filters and 4×4 kernel size), 1×1 conv layer (with 16 filters and 1×1 kernel size), pooling layer (2×2 pool size, maximum value pooling). Each branch has two convolution blocks, and each branch feeds the dense fully connected layers (2 layers). Fully connected layer use ReLU activation function [34], and the output layer uses sigmoid activation function [35]. The total number of parameters in our proposed network is 45,589 parameters.

We handle unbalanced classes by weighted mean-square error loss function during the back-propagation in the network. Adam's optimization method is used to increment the parameters of the network during training phase [36]. We keep the learning rate and other parameters constant while examining all different scenarios.

12.5 3D texture model

Similar to the previous section, we extend the model to the 3D domain to perform accurate learning of the features. The proposed architecture, shown in Fig. 12.5, consists of two identical branches in the structure. The convolution layers were constructed from $3 \times 3 \times 3$ 3Dconv (with 32 filters and $3 \times 3 \times 3$ kernel size), $1 \times 1 \times 1$ 3Dconv (with 16 filters and $1 \times 1 \times 1$ kernel size), pooling block ($2 \times 2 \times 1$ pool size, maximum value pooling). Each branch had two convolution blocks before being concatenated into the fully connected layers—dense layers (2 layers). Those layers were one hidden layer of 10 neurons with ReLU activation function [34] and one output layer of 1 neuron with sigmoid activation function [35]. Total number of 127,829 parameters is encapsulated in our model.

12.6 Texture analysis

In this step, we demonstrate the method that we deployed to extract the learnt features out of the trained CNN model. For instance, we present case of 3D CNN. Generally, we extract the kernels of each of the T2-weighted images and the ADC maps after the last epoch of the training phase. As shown in Fig. 12.6A, those kernels are projected from the 3D to 2D form by averaging the 3 depth channels. Hierarchical agglomerative clustering is used for clustering the

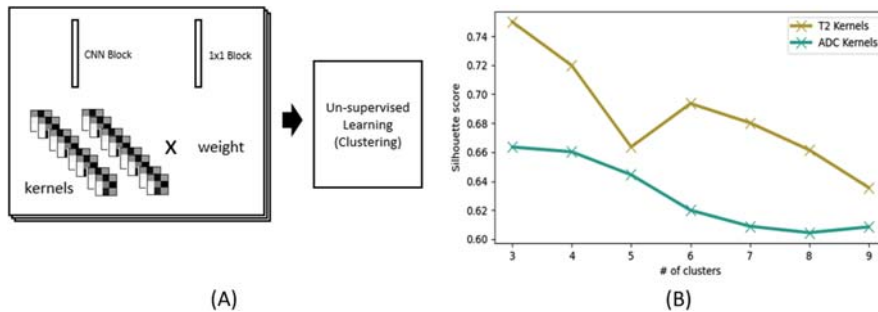


FIGURE 12.6

(A) Illustrative diagram for the extraction of the kernels from the weights of each layer, and the analysis of those kernels using nonsupervised clustering technique (hierarchical agglomerative clustering) to analyze the patterns seen in T2-weighted MRI images and ADC maps. (B) Silhouette score versus number of clusters in the clustering algorithm, as evaluation metric of the clustering.

produced 2D kernels [37]. We use Silhouette score to evaluate the produced clusters [38], see Fig. 12.6B.

12.7 Results

12.7.1 Statistical results

Following the results obtained by analyzing local intensity variations in each of the T2-weighted images and the ADC maps, a significant difference in the mean of those variations exists between benign and malignant groups. Table 12.1 shows the figures obtained from the Welch two-sample t -test. A significant difference between groups can be observed with $P < .05$. The obtained t value and the 95% confidence interval (CI) are shown in Table 12.1. In our analysis, we normalize the CI values with respect to the standard deviation (SD) of the benign group. We can see from the sign of CI that the malignant group has uplifted mean observed in T2-weighted images while the benign group has a uplifted mean in ADC maps. It can be implied that having independent convolution filters of T2-weighted images from those of ADC maps improve the learning capability of the CAD system. We can note that the convolution filters are mapped to the conv kernels in our proposed CNN architecture. The resulting box plots is illustrated in Fig. 12.3.

12.7.2 Diagnostic accuracy of 2D model

Leave-one-out cross-validation is used for the evaluation of the proposed system. During the training stage, prediction accuracy was found to converge by 100 epochs.

Table 12.1 Statistical analysis results using *t*-test on the variations of pixel-level intensity between both malignant and benign groups.

Modality	Statistical analysis quantitative results		
	CI Δ mean	<i>t</i>	<i>P</i>
T2	−4% to −1%	−2.28	.023
ADC500	5%–9%	7.87	<.001
ADC1000	26%–34%	14.87	<.001
ADC1500	4%–8%	6.12	<.001

Accordingly, we keep the network parameters fixed for the ablation study and when compared with other techniques. We use the following metrics for system evaluation in both 2D and 3D models: accuracy, sensitivity, specificity and dice coefficient.

12.7.2.1 Ablation study

We test the performance of the proposed model under different scenarios and combinations using an ablation study. First, we use a single-input CNN with structure similar to the proposed architecture in Fig. 12.4. Then, the performance of each modality is evaluated. We also study and benchmark our proposed system to other fusion and feature engineering methods. The first one is probability voting between the T2 and the ADC prediction that are generated from the single-input networks. The second method uses single-input CNN while having the T2 image and ADC image as channels (depth channels) to the input. However, the later architecture will limit the use of 1×1 conv blocks to eliminate any distortions within the used images.

The results and benchmarking of our proposed method, including all alternative scenarios indicated in the ablation study (i.e. single-input CNN), are shown in Table 12.2. The proposed method has achieved the highest performance compared to other CNN alterations.

12.7.2.2 Comparison with hand-crafted-based techniques

Given that the proposed system automatically extracts visual features from MRI imaging using convolution layers, we compared it to other ML methods that use hand-crafted features to evaluate the operation of our system. In that assessment, we use three types of hand-crafted features. The first one uses the statistical profile of image intensity. This profile is summarized using five parameters (mean, SD, entropy, skewness, kurtosis). Our aim here is to summarize and quantify the overall parameters of the whole image by presenting it using those statistical parameters. The first impression by the physician and healthcare providers can be modeled by this statistical profile while examining the MRI scan.

The second one uses a filter bank of nine kernels to assess intensity variations between neighbor voxels. The filter bank is required to capture edges and

Table 12.2 Quantitative results comparing the different scenarios including input data shape, ablation study scenarios and other models used for comparison purposes.

Input shape	Model	Evaluation metrics			Dice coefficient
		Accuracy	Sensitivity	Specificity	
2D	Single-input CNN (base-images + ADC)	0.82	0.72	0.87	0.74
2D	Single-input CNN (base-images only)	0.84	0.74	0.90	0.78
2D	Single-input CNN (ADC only)	0.82	0.70	0.90	0.76
2D	Two-CNN voting (base-images + ADC)	0.86	0.78	0.90	0.80
2D	Multiinput CNN (proposed 2D method)	0.88	0.82	0.91	0.82
2D	Hand-crafted features with DT classifier	0.70	0.70	0.70	0.35
2D	Hand-crafted features with NB classifier	0.77	0.70	0.80	0.65
2D	Hand-crafted features with RF classifier	0.77	0.77	0.77	0.59
2D	Hand-crafted features with SVM classifier	0.57	0.60	0.64	0.22
2D	Multiinput CNN (Proposed Method)	0.88	0.82	0.91	0.82
3D	Hand-crafted features with DT classifier	0.70	0.66	0.70	0.57
3D	Hand-crafted features with classifier	0.76	0.73	0.77	0.63
3D	Hand-crafted features with classifier	0.77	0.67	0.77	0.53
3D	Hand-crafted features with classifier	0.56	0.40	0.73	0.48
3D	Proposed multiinput CNN	0.87	0.69	0.97	0.79
3D	Convolutional neural networks AlexNet	0.61	0.53	0.66	0.49
3D	Convolutional Neural Networks	0.49	1.00	0.22	0.58
3D	Multiinput CNN (proposed 3D method)	0.87	0.69	0.97	0.79
3D	Single-input CNN (T2-Weighted only)	0.76	0.56	0.87	0.62

(Continued)

Table 12.2 Quantitative results comparing the different scenarios including input data shape, ablation study scenarios and other models used for comparison purposes. *Continued*

Input shape	Model	Evaluation metrics			Dice coefficient
		Accuracy	Sensitivity	Specificity	
3D	Single-Input CNN (ADC only)	0.72	0.63	0.77	0.61
3D	Two-CNN voting (base-images + ADC)	0.83	0.63	0.93	0.71
3D	Multiinput CNN (Proposed Method)	0.87	0.69	0.97	0.79

intensity variations in four orientations, lines (1D impulsive variation) in four orientations, and the point patterns (1D impulsive variation). The domains used to map those features are horizontal, vertical and diagonal orientations (two orientations). The third one uses morphological features of the nodules as a solid shape. Nodule size, aspect ratio, convex hull and bounding-box summarization of the tumor are used as morphological features. Also, the spherical harmonics of the nodule is used [39] by calculating infinite set of harmonics defined on spherical coordinate/domain. We can define the level of nonhomogeneity of the surface by the degree of the spherical harmonics, and we can have reflection toward the ability to classify between malignant and benign cases.

After collecting and extracting features, different classifiers are deployed for prediction: decision tree [40], Naive Bayes (NB) [41], random forest (RF) [42] and support vector machine (SVM) [43]. Table 12.2 summarizes the results obtained from each classifier. As clearly shown, the proposed CNN-based system outperforms all compared classifiers.

Moreover, we conducted further evaluation of the system robustness using the receiver operating characteristics (ROC) analysis curve. The ROC curve shows the relation between the false positive rate and the true positive rate when we adjust the threshold used for decision. Fig. 12.7 shows ROCs of the proposed multiinput CNN framework with respect to the other frameworks discussed earlier. The area under the curve (AUC) is higher for the proposed system with respect to other methods and other classifier models.

12.7.3 Diagnostic accuracy of 3D model

Fig. 12.5 shows the overall proposed framework, and the results obtained by our framework are summarized in Table 12.2. As shown, the proposed multiinput CNN system outperforms all other classifiers and methods. Our proposed CAD system resulted in the best performance when compared to machine learning

models that use statistical classifiers and are based on hand-crafted features. The proposed system obtained an AUC of 0.85 in comparison to 0.59 when using linear SVM classifier, see Fig. 12.8. Additionally, it obtained an accuracy, sensitivity, and specificity of 0.87, 0.69, and 0.97, respectively, in comparison to an 0.77, 0.67 and 0.77 when using RF classifier, which specifically achieved the best accuracy among the pool of classifiers used with hand-crafted features. Table 12.2 shows that using automatic feature selection produced by CNN helps in achieving better accuracy for prediction. Fig. 12.7A,B shows the training with respect to validation accuracy and loss function trend during the model training.

We also compared our model to other state-of-the-art neural networks. What is unique in our system is that it has relatively low number of layers compared to the compared models. It achieved an AUC of 0.85 in comparison to 0.67 and

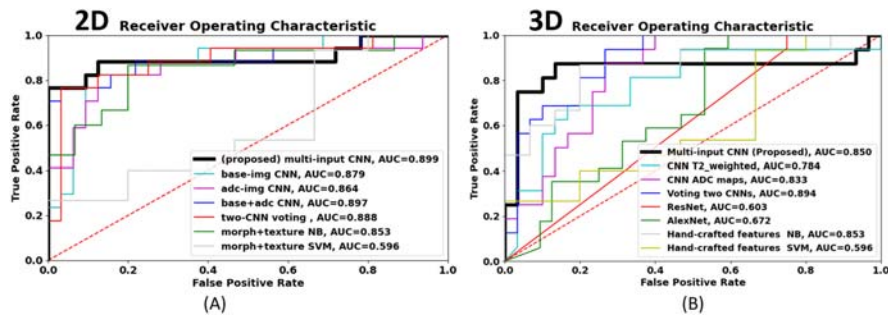


FIGURE 12.7

Receiver operating characteristic plots (ROCs) of the CNN models compared to other models. AUC is the area under the curve. *DT*, Decision tree; *RF*, random forest; *NB*, Naive Bayes; *SVM*, support vector machine. (A) Results from 2D model. (B) Results from 3D model.

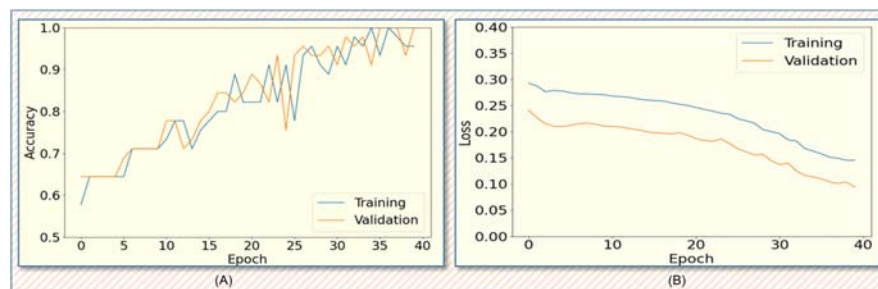


FIGURE 12.8

(A) Training versus validation accuracy plots versus the number of epochs during training phase. (B) Training versus validation loss function plots with the number of epochs during training phase, both at one of the training folds.

0.60 achieved using AlexNet and ResNet 18, respectively. Also, it resulted in accuracy of 0.87, sensitivity of 0.69 and specificity of 0.97. The accuracy, sensitivity and specificity using AlexNet were 0.61, 0.53, and 0.66, respectively, and those achieved using ResNet18 are 0.49, 1.00, and 0.22, respecting order. The achieved results show that using lower number of CNN layers can achieve better diagnostic accuracy, and this shows the advantage of the proposed method in comparison to other CNN-based techniques.

12.7.4 Texture pattern visualization

We have clustered the convolution kernels (filters) extracted and produced by the CNN after training phase, see Fig. 12.6A, and the clustering process was repeated for multiple times each with different number of expected clusters $k = 2, 4, 5, \dots, 9$. Fig. 12.6B shows the accuracy of the generated clusters by the use of Silhouette score as evaluation metric. Clusters produced from the T2-weighted kernels (green curve) achieved better accuracy in comparison to ADC kernels (blue curve). Moreover, $k = 3$ obtained the highest score in both T2-weighted and ADC images. Fig. 12.9A and B show the visualization of the generated clusters of T2-weighted and ADC kernels, respectively. The iterations (with the corresponding number of clusters, or k) are indicated on the y-axis. The generated clusters of the corresponding iteration is included in each row, and the cluster index inside each row is shown on the x-axis. Each cluster is represented by the mean of its member kernels, and then each mean is regulated from 0 to 1. We present a gray-scale visualization of each normalized mean (at each row-column position) using a 3×3 board image, as that 0–1 reflects a white–black gradient.

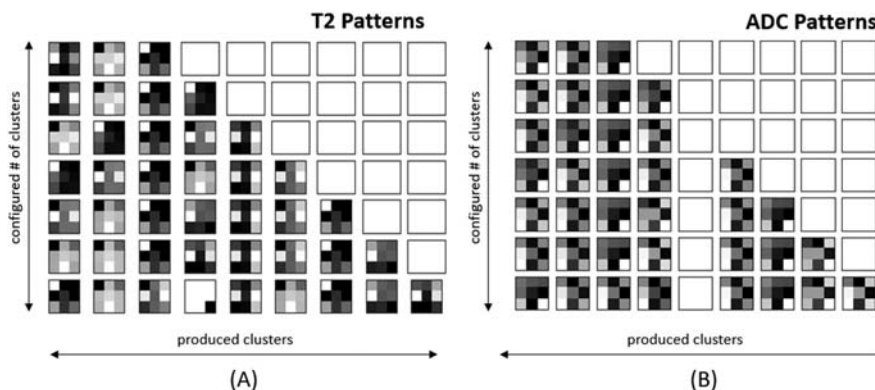


FIGURE 12.9

(A) Visualization of the texture obtained from T2-weighted images using CNN model. (B) Visualization of the texture obtained from ADC map images using CNN model. We can see a degree of heterogeneity according to this visualization that can help in distinguishing between malignant and benign thyroid nodules.

12.8 Discussion

In this chapter, we reviewed texture-based models that can be used as new CAD system to distinguish between malignant and benign thyroid nodules. Multiinput CNN can detect texture patterns from each input independently. We learn features from two modalities T2 and ADC maps using two-branch CNN. The first branch of our models learns the fluids patterns in the thyroid tissue by assessing texture patterns in T2-weighted MRI images. The second one learns the dynamics of tissue fluids by learning the texture patterns in ADC maps. For validation, we apply leave-one-out cross-validation on multimodal data collected from 49 patients that have pathologically proven nodules. We assessed the classification accuracy obtained from our system in comparison to other machine learning and deep learning approaches. We also use other hand-crafted features for our comparison. It can be seen from the experimental results that our system excels other models.

Preliminary studies were conducted to assess the possible benefits of integrating multiple MRI modalities as separate inputs of the proposed network, and that study shows heterogeneity in the intensity variability of malignant and benign samples. We used Welch two-sample *t*-test to assess the significant difference in mean variation across all modalities between the different groups/diagnosis (Table 12.1). We can see an opposite sign in the difference in mean between the two groups in T2-weighted images when compared to the corresponding difference in ADC maps (Table 12.1). This can also suggest that having independent features in each input can facilitate finding more fitted features.

We compared the results obtained by our system to other ML methods that use hand-crafted features, to evaluate the performance of our system in comparison to other popular models. In that comparison, we used three forms of hand-crafted features. The first one uses the statistical profile of image intensity. Five features (mean, SD, entropy, skewness, and kurtosis) are used to evaluate the statistical profile. This form aims to summarize the whole image by that profile. The first impression by the physician while examining the MRI scan can be mapped by this profile.

As a comment on the results, for instance from the 3D proposed system, the worst performance is obtained by linear SVM, which suggests a lack of a linear margin between classes. On the other hand, NB classifier shows the possibility of having a good distinguishable statistical distribution of the hand-crafted features between benign and malignant nodules. To benchmark our system, Fig. 12.7 shows ROCs of the proposed multiinput CNN system in comparison to the other models. Fig. 12.7A shows the 2D ROC curve, while Fig. 12.7B shows 3D figures. As per shown, the area under curve (AUC) of the proposed framework is higher in comparison with all methods, which shows higher accuracy. Fig. 12.8A,B shows the training with respect to validation accuracy and loss function trend during the model training. Hand-crafted features are not successful in providing a good diagnostic model, and this shows having multiinput CNNs with paired features can enhance the diagnostic accuracy of the CAD system.

To enrich that method, an ablation study has been conducted to evaluate the proposed method. The study shows that the proposed combination method using multiinput CNN outperformed single-input frameworks. For instance in the 3D ablation study, a single-input CNN with the same structure was built and evaluated. We implemented four scenarios. Scenarios 1 and 2 use T2-weighted images, and ADC maps, respectively. Scenario 3 uses a scheme of probability voting between the prediction of scenarios 1 and 2. Scenario 4 uses a single input that combines T2-weighted images and ADC maps in the input channels. Results obtained from all scenarios are represented in [Table 12.2](#). Using a multiinput CNN improves the diagnostic accuracy. The two-CNN voting scenario a low accuracy, sensitivity and dice coefficient compared to the proposed method, but, as a good point, they showed high specificity. The 3D ablation study suggests that having independent features for each input can enhance the detection performance of the CAD system.

The main focus of that work was to investigate the ability to extract automatic texture features associated with thyroid cancer by combining the texture in two input CNN with two independent branches. We designed the proposed networks to minimize the number of layers, which reflects the parameters in our model, to study the texture patterns that are linked to the anatomical features in the tissue. The proposed architecture supports fast computational. This can also enable further integration and combination with different MRI scanner devices to reflect the visual features automatically extracted from MRI images.

To assess and represent the texture patterns that might be used as a way to distinguish between benign and malignant tumor, we implemented an approach to extract and cluster the learned features from the trained CNN, and those patterns can be extracted from each MRI modality in a separate way. [Fig. 12.9A](#) and [B](#) show the obtained feature visualization in T2-weighted images and ADC map images respectively. We can see from the obtained visualization that there is heterogeneity in patterns between different MRI modalities. That can support the use of our models for thyroid nodule classification in contrast to hand-crafted features.

Despite the promising results, there are certain limitations that need to be worked on before moving with further clinical trials. One of these limitations are the number of samples, and this can affect the generalization of the pattern generated by the available data. Our model can be tested and validated on other cohorts with a higher number of subjects to assess the homogeneity of texture across different cohorts. We can also collect more samples to comprehensively cover thyroid cancer spectrum.

12.9 Conclusion

In summary, this chapter shows that extracting texture patterns using deep learning from MRI can improve the diagnostic performance in a noninvasive way and can help in performing accurate prediction, diagnosis, and stratification of thyroid cancer. Studying DWI and developing a diagnostic model that is texture-centric

can help in modeling the biological activities that can indicate medical conditions and the aggressiveness of those conditions. Having a method for visualization and interpretation of the learned texture patterns can support further analysis and further assessment from the medical side. Our models can be reapplied to other cohorts, and they can be adapted to perform classification of the types of thyroid cancer, and perhaps possible extension to other medical conditions. Staging of thyroid cancer can be assessed as a part of our future plan. We can also add other modalities to the model to study the heterogeneity and the dynamics of MRI texture patterns in a more comprehensive way. We can also adapt our model to study the texture patterns of thyroid tissues in other bio-images such as US. However, limited capability of modeling thyroid cancer can be observed in US in comparison to MRI, but having models that integrate US and MRI can contribute to the current development in the current field of precise and personalized medicine. We can also expand the data collection phase to include multiple samples from each subject at different time points. We can study the correlation between DWI patterns and the patterns of the cell proliferation as a biological process, which can be associated with the different stages of thyroid cancer.

Noninvasive CAD systems usually do not have suitable capabilities for interpretation, and thus they can have limited usability and low reliability. By having models that can capture accurate texture patterns in bio-images, and specifically DWI, can help in bridging the gap between radiomics and microscopic domain by introducing texture patterns that can be mapped to certain microscopic structures. Microscopic structures and features can be more indicative and illustrative for marking the biological state of the tissue. Accordingly, developing models similar to the proposed models in this chapter can contribute to enhanced noninvasive CAD paradigm toward precision medicine.

The reviewed work could also be applied to various other applications in medical imaging, such as the prostate [44–46], the kidney [47–74], the heart [75–92], the lung [93–140], the brain [141–167], the vascular system [168–178], the retina [179–183], the bladder [24–26,184,185] and injury prediction [18], as well as several nonmedical applications [186–192].

References

- [1] American Cancer Society, *Cancer Facts and Figures 2020*, American Cancer Society, Atlanta, GA, USA, 2020.
- [2] B.R. Haugen, E.K. Alexander, K.C. Bible, et al., 2015 American Thyroid Association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: the American Thyroid Association guidelines task force on thyroid nodules and differentiated thyroid cancer, *Thyroid* 26 (1) (2016) 1–133.
- [3] I. López Rojo, A. Gómez Valdazo, J. Gómez Ramirez, Current use of molecular profiling for indeterminate thyroid nodules, *Cir. Esp.* 96 (7) (2018) 395–400. Available from: <https://doi.org/10.1016/j.ciresp.2018.04.007>.

- [4] L.C. Pescatori, P. Torcia, L. Nicosia, et al., Which needle in the treatment of thyroid nodules? *Gland. Surg.* 7 (2) (2018) 111.
- [5] L.F. Alexander, N.J. Patel, et al., Thyroid ultrasound: diffuse and nodular disease, *Radiol. Clin.* (2020).
- [6] R. Mistry, C. Hillyar, et al., Ultrasound classification of thyroid nodules: a systematic review, *Cureus* 12 (3) (2020).
- [7] A.A. Ardakani, A. Gharbali, A. Mohammadi, Classification of benign and malignant thyroid nodules using wavelet texture analysis of sonograms, *J. Ultrasound Med.* 34 (11) (2015) 1983–1989.
- [8] F. Verburg, C. Reiners, Sonographic diagnosis of thyroid cancer with support of AI, *Nat. Rev. Endocrinol.* 15 (6) (2019) 319–321.
- [9] F. Ouyang, B. Guo, L. Ouyang, et al., Comparison between linear and nonlinear machine-learning algorithms for the classification of thyroid nodules, *Eur. J. Radiol.* 113 (2019) 251–257.
- [10] R. Zhang, Q. Liu, H. Cui, et al., Thyroid classification via new multi-channel feature association and learning from multi-modality MRI images, in: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), 2018, pp. 277–280.
- [11] B.H. Wokke, et al., Comparison of dixon and T1-weighted MR methods to assess the degree of fat infiltration in duchenne muscular dystrophy patients, *J. Magn. Reson. Imaging* 38 (3) (2013) 619–624.
- [12] S. Gupta, M. Soellinger, P. Boesiger, D. Poulikakos, V. Kurtcuoglu, Three-dimensional computational modeling of subject-specific cerebrospinal fluid flow in the subarachnoid space, *J. Biomech. Eng.* 131 (2) (2009).
- [13] H. Seo, J. Choi, C. Oh, Y. Han, H. Park, Isotropic diffusion weighting for measurement of a high-resolution apparent diffusion coefficient map using a single radial scan in MRI, *Phys. Med. Biol.* 59 (20) (2014) 6289–6303. Available from: <https://doi.org/10.1088/0031-9155/59/20/6289>.
- [14] D.-M. Koh, D.J. Collins, Diffusion-weighted MRI in the body: applications and challenges in oncology, *Am. J. Roentgenol.* 188 (6) (2007) 1622–1635.
- [15] Y. Hao, C. Pan, W. Chen, et al., Differentiation between malignant and benign thyroid nodules and stratification of papillary thyroid cancer with aggressive histological features: Whole-lesion diffusion-weighted imaging histogram analysis, *J. Magn. Reson. Imaging* 44 (6) (2016) 1546–1555.
- [16] A.M. Brown, S. Nagala, M.A. McLean, et al., Multi-institutional validation of a novel textural analysis tool for preoperative stratification of suspected thyroid tumors on diffusion-weighted MRI, *Magn. Reson. Med.* 75 (4) (2016) 1708–1716.
- [17] S. Schob, H.J. Meyer, J. Dieckow, et al., Histogram analysis of diffusion weighted imaging at 3T is useful for prediction of lymphatic metastatic spread, proliferative activity, and cellularity in thyroid cancer, *Int. J. Mol. Sci.* 18 (4) (2017) 821.
- [18] A. Naglah, et al., Athlete-customized injury prediction using training load statistical records and machine learning, in: 2018 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), 2018, pp. 459–464.
- [19] A. Naglah, A. ElDesouky, M. ElHelw, Real-time scale-adaptive compressive tracking using two classification stages, in: 2015 IEEE International Conference on Robotics and Biomimetics (ROBIO), 2015, pp. 363–367.

- [20] A. Naglah, F. Khalifa, R. Khaled, A. El-Baz, et al., Thyroid cancer computer-aided diagnosis system using MRI-based multi-input CNN model, in: 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI), 2021, pp. 1691–1694.
- [21] R.F. Hoey, et al., Bladder and bowel responses to lumbosacral epidural stimulation in uninjured and transected anesthetized rats, *Sci. Rep.* 11 (1) (2021) 1–21.
- [22] O. Dekhil, A. Naglah, M. Shaban, M. Ghazal, F. Taher, A. Elbaz, Deep learning based method for computer aided diagnosis of diabetic retinopathy, in: 2019 IEEE International Conference on Imaging Systems and Techniques (IST), 2019, pp. 1–4.
- [23] K. Hammouda, et al., A 3D CNN with a learnable adaptive shape prior for accurate segmentation of bladder wall using MR images, in: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), 2020, pp. 935–938.
- [24] K. Hammouda, et al., A CAD system for accurate diagnosis of bladder cancer staging using a multiparametric MRI, in: 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI), 2021, pp. 1718–1721.
- [25] K. Hammouda, et al., A multiparametric MRI-based CAD system for accurate diagnosis of bladder cancer staging, *Comput. Med. Imaging Graph.* 90 (2021) 101911.
- [26] K. Hammouda et al., A CNN-based framework for bladder wall segmentation using MRI, in: 2019 Fifth International Conference on Advances in Biomedical Engineering (ICABME), 2019, pp. 1–4.
- [27] A. Naglah, et al. Computer-aided diagnosis of acute myocardial infarction using time-dependent plasma metabolites, in: 2019 IEEE International Conference on Imaging Systems and Techniques (IST), 2019, pp. 1–5.
- [28] A. Naglah, et al. Ensemble learning of blood metabolic biomarkers: A novel cad system of heart failure, in: 2019 Fifth International Conference on Advances in Biomedical Engineering (ICABME), 2019, pp. 1–4.
- [29] A. Naglah, et al., Novel MRI-based CAD system for early detection of thyroid cancer using multi-input CNN, *Sensors* 21 (11) (2021) 3878.
- [30] E.O. Stejskal, J.E. Tanner, Spin diffusion measurements: spin echoes in the presence of a time-dependent field gradient, *J. Chem. Phys.* 42 (1) (1965) 288–292.
- [31] S.S. Negi, Y.S. Bhandari, A hybrid approach to image enhancement using contrast stretching on image sharpening and the analysis of various cases arising using histogram, in: International Conference on Recent Advances and Innovations in Engineering (ICRAIE-2014), 2014, pp. 1–6.
- [32] A. Surov, N. Garnov, Proving of a mathematical model of cell calculation based on apparent diffusion coefficient, *Transl. Oncol.* 10 (5) (2017) 828–830.
- [33] H. Wang, B. Song, N. Ye, et al., Machine learning-based multiparametric MRI radiomics for predicting the aggressiveness of papillary thyroid carcinoma, *Eur. J. Radiol.* 122 (2020) 108755.
- [34] A.F. Agarap, Deep learning using rectified linear units (ReLU), *arXiv Prepr. arXiv1803.08375*, 2018.
- [35] D.J. Finney, *Probit Analysis: a Statistical Treatment of the Sigmoid Response Curve*, Cambridge university press, Cambridge, 1952.
- [36] D.P. Kingma, J. Ba, Adam: a method for stochastic optimization, *arXiv Prepr. arXiv1412.6980*, 2014.
- [37] F. Murtagh, P. Legendre, Ward's hierarchical agglomerative clustering method: which algorithms implement Ward's criterion? *J. Classif.* 31 (3) (2014) 274–295.

- [38] T. Thinsungnoena, N. Kaoungkub, P. Durongdumronchaib, K. Kerdprasob, N. Kerdprasob, The clustering validity with silhouette and sum of squared errors, *Learning* 3 (7) (2015).
- [39] C. Müller, *Spherical Harmonics*, 17, Springer, 2006.
- [40] M.A. Friedl, C.E. Brodley, Decision tree classification of land cover from remotely sensed data, *Remote. Sens. Environ.* 61 (3) (1997) 399–409.
- [41] A. McCallum, K. Nigam, et al., A comparison of event models for naive bayes text classification in AAI-98 Workshop Learn. Text. Categorization 752 (1) (1998) 41–48.
- [42] A. Liaw, M. Wiener, et al., Classification and regression by randomForest, *R. N.* 2 (3) (2002) 18–22.
- [43] J.A.K. Suykens, J. Vandewalle, Least squares support vector machine classifiers, *Neural Process. Lett.* 9 (3) (1999) 293–300.
- [44] I. Reda, M. Ghazal, A. Shalaby, M. Elmogy, A. AbouEl-Fetouh, B.O. Ayinde, et al., A novel ADCS-based CNN classification system for precise diagnosis of prostate cancer, in: 2018 24th International Conference on Pattern Recognition (ICPR). IEEE, 2018, pp. 3923–3928.
- [45] I. Reda, A. Khalil, M. Elmogy, A. Abou El-Fetouh, A. Shalaby, M. Abou El-Ghar, et al., Deep learning role in early diagnosis of prostate cancer, *Technol. Cancer Res. Treat.* 17 (2018) 1533034618775530.
- [46] I. Reda, B.O. Ayinde, M. Elmogy, A. Shalaby, M. El-Melegy, M.A. El-Ghar, et al., A new CNN-based system for early diagnosis of prostate cancer, in: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). IEEE, 2018, pp. 207–210.
- [47] A.S. Chowdhury, R. Roy, S. Bose, F.K.A. Elnakib, A. El-Baz, Non-rigid biomedical image registration using graph cuts with a novel data term, in: Proceedings of IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI'12), Barcelona, Spain, 2–5 May, 2012, pp. 446–449.
- [48] A. El-Baz, A.A. Farag, S.E. Yuksel, M.E. El-Ghar, T.A. Eldiasty, M.A. Ghoneim, *Application of deformable models for the detection of acute renal rejection, Deformable Models*, Springer, New York, 2007, pp. 293–333.
- [49] A. El-Baz, A. Farag, R. Fahmi, S. Yuksel, M.A. El-Ghar, T. Eldiasty, Image analysis of renal DCE MRI for the detection of acute renal rejection, in: Proceedings of IAPR International Conference on Pattern Recognition (ICPR'06), Hong Kong, 2006, pp. 822–825.
- [50] A. El-Baz, A. Farag, R. Fahmi, S. Yuksel, W. Miller, M.A. El-Ghar, et al., A new CAD system for the evaluation of kidney diseases using DCE-MRI, in: Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI'08), Copenhagen, Denmark, 1–6 October, 2006, pp. 446–453.
- [51] A. El-Baz, G. Gimel'farb, M.A. El-Ghar, A novel image analysis approach for accurate identification of acute renal rejection, in: Proceedings of IEEE International Conference on Image Processing (ICIP'08), San Diego, California, USA, 12–15 October, 2008, pp. 1812–1815.
- [52] A. El-Baz, G. Gimel'farb, M.A. El-Ghar, Image analysis approach for identification of renal transplant rejection, in: Proceedings of IAPR International Conference on Pattern Recognition, (ICPR'08), Tampa, FL, USA, 2008, pp. 1–4.

- [53] A. El-Baz, G. Gimel'farb, M.A. El-Ghar, New motion correction models for automatic identification of renal transplant rejection, in: Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention, (MICCAI'07), Brisbane, Australia, 29 October–2 November, 2007, pp. 235–243.
- [54] A. Farag, A. El-Baz, S. Yuksel, M.A. El-Ghar, T. Eldiasty, A framework for the detection of acute rejection with Dynamic Contrast Enhanced Magnetic Resonance Imaging, in: Proceedings of IEEE International Symposium on Biomedical Imaging: From Nano to Macro, (ISBI'06), Arlington, Virginia, USA, 2006, pp. 418–421.
- [55] F. Khalifa, G.M. Beache, M.A. El-Ghar, T. El-Diasty, G. Gimel'farb, M. Kong, et al., Dynamic contrast-enhanced MRI- based early detection of acute renal transplant rejection, *IEEE Trans. Med. Imaging* 32 (10) (2013) 1910–1927.
- [56] F. Khalifa, A. El-Baz, G. Gimel'farb, M.A. El-Ghar, Non-invasive image-based approach for early detection of acute renal rejection, in: Proceedings of International Conference Medical Image Computing and Computer-Assisted Intervention (MICCAI'10), Beijing, China, 20–24 September 2010, pp. 10–18.
- [57] F. Khalifa, A. El-Baz, G. Gimel'farb, R. Ouseph, M.A. El-Ghar, Shape-appearance guided level-set deformable model for image segmentation, in: Proceedings of IAPR International Conference on Pattern Recognition, (ICPR'10), Istanbul, Turkey, 23–26 August, 2010, pp. 4581–4584.
- [58] F. Khalifa, M.A. El-Ghar, B. Abdollahi, H. Frieboes, T. El-Diasty, A. El-Baz, A comprehensive non-invasive framework for automated evaluation of acute renal transplant rejection using DCE-MRI, *NMR Biomedicine* 26 (11) (2013) 1460–1470.
- [59] F. Khalifa, M.A. El-Ghar, B. Abdollahi, H.B. Frieboes, T. El-Diasty, A. El-Baz, Dynamic contrast-enhanced MRI-based early detection of acute renal transplant rejection, in: 2014 Annual Scientific Meeting and Educational Course Brochure of the Society of Abdominal Radiology, (SAR'14), Boca Raton, FL, 23–28 March, 2014, p. CID: 1855912.
- [60] F. Khalifa, A. Elnakib, G.M. Beache, G. Gimel'farb, M.A. El-Ghar, G. Sokhadze, et al., 3D kidney segmentation from CT images using a level set approach guided by a novel stochastic speed function, in: Proceedings of International Conference Medical Image Computing and Computer-Assisted Intervention, (MICCAI'11), Toronto, Canada, 18–22 September, 2011, pp. 587–594.
- [61] F. Khalifa, G. Gimel'farb, M.A. El-Ghar, G. Sokhadze, S. Manning, P. McClure, et al., A new deformable model-based segmentation approach for accurate extraction of the kidney from abdominal CT images, in: Proceedings of IEEE International Conference on Image Processing, (ICIP'11), Brussels, Belgium, 11–14 September, 2011, pp. 3393–3396.
- [62] M. Mostapha, F. Khalifa, A. Alansary, A. Soliman, J. Suri, A. El-Baz, Computer-aided diagnosis systems for acute renal transplant rejection: challenges and methodologies, in: A. El-Baz, L.J. Suri (Eds.), *Abdomen and Thoracic Imaging*, Springer, 2014, pp. 1–35.
- [63] M. Shehata, F. Khalifa, E. Hollis, A. Soliman, E. Hosseini-Asl, M.A. El-Ghar, et al., A new non-invasive approach for early classification of renal rejection types using diffusion-weighted MRI, in: IEEE International Conference on Image Processing (ICIP), 2016. IEEE, 2016, pp. 136–140.
- [64] F. Khalifa, A. Soliman, A. Takieldean, M. Shehata, M. Mostapha, A. Shaffie, et al., Kidney segmentation from CT images using a 3D NMF-guided active contour model, in: IEEE 13th International Symposium on Biomedical Imaging (ISBI), 2016. IEEE, 2016, pp. 432–435.

- [65] M. Shehata, F. Khalifa, A. Soliman, A. Takieldeem, M.A. El-Ghar, A. Shaffie, et al., 3D diffusion MRI-based cad system for early diagnosis of acute renal rejection, in: Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium on. IEEE, 2016, pp. 1177–1180.
- [66] M. Shehata, F. Khalifa, A. Soliman, R. Alrefai, M.A. El-Ghar, A.C. Dwyer, et al., A level set-based framework for 3D kidney segmentation from diffusion MR images, in: IEEE International Conference on Image Processing (ICIP), 2015. IEEE, 2015, pp. 4441–4445.
- [67] M. Shehata, F. Khalifa, A. Soliman, M.A. El-Ghar, A.C. Dwyer, G. Gimel'farb, et al., A promising non- invasive cad system for kidney function assessment, in: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2016, pp. 613–621.
- [68] F. Khalifa, A. Soliman, A. Elmaghraby, G. Gimel'farb, A. El-Baz, 3D kidney segmentation from abdominal images using spatial-appearance models, *Computat. Math. Methods Med.* 2017 (2017) 1–10.
- [69] E. Hollis, M. Shehata, F. Khalifa, M.A. El-Ghar, T. El-Diasty, A. El-Baz, Toward non-invasive diagnostic techniques for early detection of acute renal transplant rejection: a review, *Egypt. J. Radiology Nucl. Med.* 48 (1) (2016) 257–269.
- [70] M. Shehata, F. Khalifa, A. Soliman, M.A. El-Ghar, A.C. Dwyer, A. El-Baz, Assessment of renal transplant using image and clinical-based biomarkers, in: Proceedings of 13th Annual Scientific Meeting of American Society for Diagnostics and Interventional Nephrology (ASDIN'17), New Orleans, LA, USA, 10–12 February 2017, 2017.
- [71] M. Shehata, F. Khalifa, A. Soliman, M.A. El-Ghar, A.C. Dwyer, A. El-Baz, Early assessment of acute renal rejection, in: Proceedings of 12th Annual Scientific Meeting of American Society for Diagnostics and Interventional Nephrology (ASDIN'16), Phoenix, AZ, USA, 19–21 February 2016, 2017.
- [72] A. Eltanboly, M. Ghazal, H. Hajjdiab, A. Shalaby, A. Switala, A. Mahmoud, et al., Level sets-based image segmentation approach using statistical shape priors, *Appl. Math. Comput.* 340 (2019) 164–179.
- [73] M. Shehata, A. Mahmoud, A. Soliman, F. Khalifa, M. Ghazal, M.A. El-Ghar, et al., 3d kidney segmentation from abdominal diffusion mri using an appearance-guided deformable boundary, *PLoS One* 13 (7) (2018) e0200082.
- [74] H. Abdeltawab, M. Shehata, A. Shalaby, F. Khalifa, A. Mahmoud, M.A. El-Ghar, et al., A novel cnn-based cad system for early assessment of transplanted kidney dysfunction, *Sci. Rep.* 9 (1) (2019) 5948.
- [75] K. Hammouda, F. Khalifa, H. Abdeltawab, A. Elnakib, G. Giridharan, M. Zhu, et al., A new framework for performing cardiac strain analysis from CINE MRI imaging in mice, *Sci. Rep.* 10 (1) (2020) 1–15.
- [76] F. Khalifa, G. Beache, A. El-Baz, G. Gimel'farb, Deformable model guided by stochastic speed with application in cine images segmentation, in: Proceedings of IEEE International Conference on Image Processing, (ICIP'10), Hong Kong, 26–29 September 2010, pp. 1725–1728.
- [77] F. Khalifa, G.M. Beache, A. Elnakib, H. Sliman, G. Gimel'farb, K.C. Welch, et al., A new shape-based framework for the left ventricle wall segmentation from cardiac first-pass perfusion MRI, in: Proceedings of IEEE International Symposium on Biomedical Imaging: From Nano to Macro, (ISBI'13), San Francisco, CA, 7–11 April 2013, pp. 41–44.

- [78] F. Khalifa, G.M. Beache, A. Elnakib, H. Sliman, G. Gimel'farb, K.C. Welch, et al., A new nonrigid registration framework for improved visualization of transmural perfusion gradients on cardiac first-pass perfusion MRI, in: *Proceedings of IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI'12)*, Barcelona, Spain, May 2–5, 2012, pp. 828–831.
- [79] F. Khalifa, G.M. Beache, A. Firjani, K.C. Welch, G. Gimel'farb, A. El-Baz, A new nonrigid registration approach for motion correction of cardiac first-pass perfusion MRI, in: *Proceedings of IEEE International Conference on Image Processing, (ICIP'12)*, Lake Buena Vista, FL, 30 September–3 October, 2012, pp. 1665–1668.
- [80] F. Khalifa, G.M. Beache, G. Gimel'farb, A. El-Baz, A novel CAD system for analyzing cardiac first-pass MR images, in: *Proceedings of IAPR International Conference on Pattern Recognition (ICPR'12)*, Tsukuba Science City, Japan, 11–15 November, 2012, pp. 77–80.
- [81] F. Khalifa, G.M. Beache, G. Gimel'farb, A. El-Baz, A novel approach for accurate estimation of left ventricle global indexes from short-axis cine MRI, in: *Proceedings of IEEE International Conference on Image Processing, (ICIP'11)*, Brussels, Belgium, 11–14 September, 2011, pp. 2645–2649.
- [82] F. Khalifa, G.M. Beache, G. Gimel'farb, G.A. Giridharan, A. El-Baz, A new image-based framework for analyzing cine images, in: A. El-Baz, U.R. Acharya, M. Mirmedhdi, J.S. Suri (Eds.), *Handbook of Multi Modality State-of-the-Art Medical Image Segmentation and Registration Methodologies*, 2, Springer, New York, 2011, pp. 69–98.
- [83] F. Khalifa, G.M. Beache, G. Gimel'farb, G.A. Giridharan, A. El-Baz, Accurate automatic analysis of cardiac cine images, *IEEE Trans. Biomed. Eng.* 59 (2) (2012) 445–455.
- [84] F. Khalifa, G.M. Beache, M. Nitzken, G. Gimel'farb, G.A. Giridharan, A. El-Baz, Automatic analysis of left ventricle wall thickness using short-axis cine CMR images, in: *Proceedings of IEEE International Symposium on Biomedical Imaging: From Nano to Macro, (ISBI'11)*, Chicago, IL, 30 March–2 April, 2011, pp. 1306–1309.
- [85] M. Nitzken, G. Beache, A. Elnakib, F. Khalifa, G. Gimel'farb, A. El-Baz, Accurate modeling of tagged CMR 3D image appearance characteristics to improve cardiac cycle strain estimation, in: *Image Processing (ICIP), 2012 19th IEEE International Conference on*, Orlando, FL, USA: IEEE, Sep. 2012, pp. 521–524.
- [86] M. Nitzken, G. Beache, A. Elnakib, F. Khalifa, G. Gimel'farb, A. El-Baz, Improving full-cardiac cycle strain estimation from tagged CMR by accurate modeling of 3D image appearance characteristics, in: *Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on*, Barcelona, Spain: IEEE, May, 2012, pp. 462–465, (Selected for oral presentation).
- [87] M.J. Nitzken, A.S. El-Baz, G.M. Beache, Markov-gibbs random field model for improved full-cardiac cycle strain estimation from tagged cmr, *J. Cardiovasc. Magn. Reson.* 14 (1) (2012) 1–2.
- [88] H. Sliman, A. Elnakib, G. Beache, A. Elmaghaby, A. El-Baz, Assessment of myocardial function from cine cardiac MRI using a novel 4D tracking approach, *J. Comput. Sci. Syst. Biol.* 7 (2014) 169–173.
- [89] H. Sliman, A. Elnakib, G.M. Beache, A. Soliman, F. Khalifa, G. Gimel'farb, et al., A novel 4D PDE-based approach for accurate assessment of myocardium function using cine cardiac magnetic resonance images, in: *Proceedings of IEEE International Conference on Image Processing (ICIP'14)*, Paris, France, 27–30 October, 2014, pp. 3537–3541.

- [90] H. Sliman, F. Khalifa, A. Elnakib, G.M. Beache, A. Elmaghraby, A. El-Baz, A new segmentation-based tracking framework for extracting the left ventricle cavity from cine cardiac MRI, in: *Proceedings of IEEE International Conference on Image Processing, (ICIP'13)*, Melbourne, Australia, 15–18 September 2013, pp. 685–689.
- [91] H. Sliman, F. Khalifa, A. Elnakib, A. Soliman, G.M. Beache, A. Elmaghraby, et al., Myocardial borders segmentation from cine MR images using bi-directional coupled parametric deformable models, *Med. Phys.* 40 (9) (2013) 1–13.
- [92] H. Sliman, F. Khalifa, A. Elnakib, A. Soliman, G.M. Beache, G. Gimel'farb, et al., Accurate segmentation framework for the left ventricle wall from cardiac cine MRI, in: *Proceedings of International Symposium on Computational Models for Life Science, (CMLS'13)*, 1559, Sydney, Australia, 27–29 November, 2013, pp. 287–296.
- [93] B. Abdollahi, A.C. Civelek, X.-F. Li, J. Suri, A. El-Baz, PET/CT nodule segmentation and diagnosis: a survey, in: L. Saba, J.S. Suri (Eds.), *Multi Detector CT Imaging*, Taylor, Francis, 2014, pp. 639–651.
- [94] B. Abdollahi, A. El-Baz, A.A. Amini, A multi-scale non-linear vessel enhancement technique, in: *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE. IEEE*, 2011, pp. 3925–3929.
- [95] B. Abdollahi, A. Soliman, A. Civelek, X.-F. Li, G. Gimel'farb, A. El-Baz, A novel Gaussian scale space-based joint MGRF framework for precise lung segmentation, in: *Proceedings of IEEE International Conference on Image Processing (ICIP'12)*. IEEE, 2012, pp. 2029–2032.
- [96] B. Abdollahi, A. Soliman, A. Civelek, X.-F. Li, G. Gimel'farb, A. El-Baz, A novel 3D joint MGRF framework for precise lung segmentation, *Machine Learning in Medical Imaging*, Springer, 2012, pp. 86–93.
- [97] A.M. Ali, A.S. El-Baz, A.A. Farag, A novel framework for accurate lung segmentation using graph cuts, in: *Proceedings of IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI'07)*. IEEE, 2007, pp. 908–911.
- [98] A. El-Baz, G.M. Beache, G. Gimel'farb, K. Suzuki, K. Okada, Lung imaging data analysis, *Int. J. Biomed. Imaging* 2013 (2013) 1–2.
- [99] A. El-Baz, G.M. Beache, G. Gimel'farb, K. Suzuki, K. Okada, A. Elnakib, et al., Computer-aided diagnosis systems for lung cancer: challenges and methodologies, *Int. J. Biomed. Imaging* 2013 (2013) 1–46.
- [100] A. El-Baz, A. Elnakib, M. Abou El-Ghar, G. Gimel'farb, R. Falk, A. Farag, Automatic detection of 2D and 3D lung nodules in chest spiral CT scans, *Int. J. Biomed. Imaging* 2013 (2013) 1–11.
- [101] A. El-Baz, A.A. Farag, R. Falk, R. La Rocca, A unified approach for detection, visualization, and identification of lung abnormalities in chest spiral CT scans, in: *International Congress Series*, 1256. Elsevier, 2003, pp. 998–1004.
- [102] A. El-Baz, A.A. Farag, R. Falk, R. La Rocca, Detection, visualization and identification of lung abnormalities in chest spiral CT scan: Phase-I *Proceedings of International Conference on Biomedical Engineering Cairo, Egypt*. 12 (1) (2002).
- [103] A. El-Baz, A. Farag, G. Gimel'farb, R. Falk, M.A. El-Ghar, T. Eldiasty, A framework for automatic segmentation of lung nodules from low dose chest CT scans, in: *Proceedings of International Conference on Pattern Recognition, (ICPR'06)*, 3. IEEE, 2006, pp. 611–614.
- [104] A. El-Baz, A. Farag, G. Gimel'farb, R. Falk, M.A. El-Ghar, A novel level set-based computer-aided detection system for automatic detection of lung nodules in low

- dose chest computed tomography scans, *Lung Imaging Computer Aided Diagnosis* 10 (2011) 221–238.
- [105] A. El-Baz, G. Gimel'farb, M. Abou, et al., Appearance-based diagnostic system for early assessment of malignant lung nodules, in: *Proceedings of IEEE International Conference on Image Processing, (ICIP'12)*. IEEE, 2012, pp. 533–536.
- [106] A. El-Baz, G. Gimel'farb, R. Falk, A novel 3D framework for automatic lung segmentation from low dose CT images, in: A. El-Baz, J.S. Suri (Eds.), *Lung Imaging and Computer Aided Diagnosis*, Taylor, Francis, 2011, pp. 1–16.
- [107] A. El-Baz, G. Gimel'farb, R. Falk, M. El-Ghar, Appearance analysis for diagnosing malignant lung nodules, in: *Proceedings of IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI'10)*. IEEE, 2010, pp. 193–196.
- [108] A. El-Baz, G. Gimel'farb, R. Falk, M.A. El-Ghar, A novel level set-based CAD system for automatic detection of lung nodules in low dose chest CT scans, in: A. El-Baz, J.S. Suri (Eds.), *Lung Imaging and Computer Aided Diagnosis*, 1, Taylor, Francis, 2011, pp. 221–238.
- [109] A. El-Baz, G. Gimel'farb, R. Falk, M.A. El-Ghar, A new approach for automatic analysis of 3D low dose CT images for accurate monitoring the detected lung nodules, in: *Proceedings of International Conference on Pattern Recognition, (ICPR'08)*. IEEE, 2008, pp. 1–4.
- [110] A. El-Baz, G. Gimel'farb, R. Falk, M.A. El-Ghar, A novel approach for automatic follow-up of detected lung nodules, in: *Proceedings of IEEE International Conference on Image Processing, (ICIP'07)*, 5. IEEE, 2007, pp. V–501.
- [111] A. El-Baz, G. Gimel'farb, R. Falk, M.A. El-Ghar, A new CAD system for early diagnosis of detected lung nodules, in: *Image Processing, 2007. ICIP 2007. IEEE International Conference on*, 2. IEEE, 2007, pp. II–461.
- [112] A. El-Baz, G. Gimel'farb, R. Falk, M.A. El-Ghar, H. Refaie, Promising results for early diagnosis of lung cancer, in: *Proceedings of IEEE International Symposium on Biomedical Imaging: From Nano to Macro, (ISBI'08)*. IEEE, 2008, pp. 1151–1154.
- [113] A. El-Baz, G.L. Gimel'farb, R. Falk, M. Abou El-Ghar, T. Holland, T. Shaffer, A new stochastic framework for accurate lung segmentation, in: *Proceedings of Medical Image Computing and Computer-Assisted Intervention (MICCAI'08)*, 2008, pp. 322–330.
- [114] A. El-Baz, G.L. Gimel'farb, R. Falk, D. Heredis, M. Abou El-Ghar, A novel approach for accurate estimation of the growth rate of the detected lung nodules, in: *Proceedings of International Workshop on Pulmonary Image Analysis*, 2008, pp. 33–42.
- [115] A. El-Baz, G.L. Gimel'farb, R. Falk, T. Holland, T. Shaffer, A framework for unsupervised segmentation of lung tissues from low dose computed tomography images, in: *Proceedings of British Machine Vision, (BMVC'08)*, 2008, pp. 1–10.
- [116] A. El-Baz, G. Gimel'farb, R. Falk, M.A. El-Ghar, 3D MGRF-based appearance modeling for robust segmentation of pulmonary nodules in 3D LDCT chest images, in: *Lung Imaging and Computer Aided Diagnosis*. Chapter 3, 2011, pp. 51–63.
- [117] A. El-Baz, G. Gimel'farb, R. Falk, M.A. El-Ghar, Automatic analysis of 3D low dose CT images for early diagnosis of lung cancer, *Pattern Recognit.* 42 (6) (2009) 1041–1051.
- [118] A. El-Baz, G. Gimel'farb, R. Falk, M.A. El-Ghar, S. Rainey, D. Heredia, et al., Toward early diagnosis of lung cancer, in: *Proceedings of Medical Image Computing and Computer-Assisted Intervention, (MICCAI'09)*. Springer, 2009, pp. 682–689.

- [119] A. El-Baz, G. Gimel'farb, R. Falk, M.A. El-Ghar, J. Suri, Appearance analysis for the early assessment of detected lung nodules, in: *Lung Imaging and Computer Aided Diagnosis*. Chapter 17, 2011, pp. 395–404.
- [120] A. El-Baz, F. Khalifa, A. Elnakib, M. Nitzken, A. Soliman, P. McClure, et al., A novel approach for global lung registration using 3D Markov Gibbs appearance model, in: *Proceedings of International Conference Medical Image Computing and Computer-Assisted Intervention, (MICCAI'12)*, Nice, France, 1–5 October 2012, pp. 114–121.
- [121] A. El-Baz, M. Nitzken, A. Elnakib, F. Khalifa, G. Gimel'farb, R. Falk, et al., 3D shape analysis for early diagnosis of malignant lung nodules, in: *Proceedings of International Conference Medical Image Computing and Computer-Assisted Intervention, (MICCAI'11)*, Toronto, Canada, 18–22 September 2011, pp. 175–182.
- [122] A. El-Baz, M. Nitzken, G. Gimel'farb, E. Van Bogaert, R. Falk, M.A. El-Ghar, et al., Three-dimensional shape analysis using spherical harmonics for early assessment of detected lung nodules, in: *Lung Imaging and Computer Aided Diagnosis*. chapter, 2011, ch. 19, pp. 421–438.
- [123] A. El-Baz, M. Nitzken, F. Khalifa, A. Elnakib, G. Gimel'farb, R. Falk, et al., 3D shape analysis for early diagnosis of malignant lung nodules, in: *Proceedings of International Conference on Information Processing in Medical Imaging, (IPMI'11)*, Monastery Irsee, Germany (Bavaria), 3–8 July, 2011, pp. 772–783.
- [124] A. El-Baz, M. Nitzken, E. Vanbogaert, G. Gimel'Farb, R. Falk, M. Abo El-Ghar, A novel shape-based diagnostic approach for early diagnosis of lung nodules, in: *Biomedical Imaging: From Nano to Macro*, 2011 IEEE International Symposium on. IEEE, 2011, pp. 137–140.
- [125] [A. El-Baz, P. Sethu, G. Gimel'farb, F. Khalifa, A. Elnakib, R. Falk, et al., Elastic phantoms generated by microfluidics technology: validation of an imaged-based approach for accurate measurement of the growth rate of lung nodules, *Biotechnol. J.* 6 \(2\) \(2011\) 195–203.](#)
- [126] A. El-Baz, P. Sethu, G. Gimel'farb, F. Khalifa, A. Elnakib, R. Falk, et al., A new validation approach for the growth rate measurement using elastic phantoms generated by state-of-the-art microfluidics technology, in: *Proceedings of IEEE International Conference on Image Processing, (ICIP'10)*, Hong Kong, 26–29 September, 2010, pp. 4381–4383.
- [127] [A. El-Baz, P. Sethu, G. Gimel'farb, F. Khalifa, A. Elnakib, R. Falk, et al., Validation of a new imaged-based approach for the accurate estimating of the growth rate of detected lung nodules using real CT images and elastic phantoms generated by state-of-the-art microfluidics technology, in: A. El-Baz, J.S. Suri \(Eds.\), *Handbook of Lung Imaging and Computer Aided Diagnosis*, 1, Taylor & Francis, New York, 2011, pp. 405–420.](#)
- [128] A. El-Baz, A. Soliman, P. McClure, G. Gimel'farb, M.A. El-Ghar, R. Falk, Early assessment of malignant lung nodules based on the spatial analysis of detected lung nodules, in: *Proceedings of IEEE International Symposium on Biomedical Imaging: From Nano to Macro, (ISBI'12)*. IEEE, 2012, pp. 1463–1466.
- [129] A. El-Baz, S.E. Yuksel, S. Elshazly, A.A. Farag, Non-rigid registration techniques for automatic follow-up of lung nodules, in: *Proceedings of Computer Assisted Radiology and Surgery, (CARS'05)*, 1281. Elsevier, 2005, pp. 1115–1120.
- [130] [A.S. El-Baz, J.S. Suri, *Lung Imaging and Computer Aided Diagnosis*, CRC Press, 2011.](#)

- [131] A. Soliman, F. Khalifa, N. Dunlap, B. Wang, M. El-Ghar, A. El-Baz An iso-surfaces based local deformation handling framework of lung tissues, in: Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium on. IEEE, 2016, pp. 1253–1259.
- [132] A. Soliman, F. Khalifa, A. Shaffie, N. Dunlap, B. Wang, A. Elmaghraby, et al., Detection of lung injury using 4D-CT chest images, in: Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium on. IEEE, 2016, pp. 1274–1277.
- [133] A. Soliman, F. Khalifa, A. Shaffie, N. Dunlap, B. Wang, A. Elmaghraby, et al., A comprehensive framework for early assessment of lung injury, in: Image Processing (ICIP), 2017 IEEE International Conference on. IEEE, 2017, pp. 3275–3279.
- [134] A. Shaffie, A. Soliman, M. Ghazal, F. Taher, N. Dunlap, B. Wang, et al., A new framework for incorporating appearance and shape features of lung nodules for precise diagnosis of lung cancer, in: Image Processing (ICIP), 2017 IEEE International Conference on. IEEE, 2017, pp. 1372–1376.
- [135] A. Soliman, F. Khalifa, A. Shaffie, N. Liu, N. Dunlap, B. Wang, et al., Image-based cad system for accurate identification of lung injury, in: Image Processing (ICIP), 2016 IEEE International Conference on. IEEE, 2016, pp. 121–125.
- [136] A. Soliman, A. Shaffie, M. Ghazal, G. Gimel'farb, R. Keynton, A. El-Baz, A novel CNN segmentation framework based on using new shape and appearance features, in: 2018 25th IEEE International Conference on Image Processing (ICIP). IEEE, 2018, pp. 3488–3492.
- [137] A. Shaffie, A. Soliman, H.A. Khalifeh, M. Ghazal, F. Taher, R. Keynton, et al., On the integration of CT-derived features for accurate detection of lung cancer, in: 2018 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT). IEEE, 2018, pp. 435–440.
- [138] A. Shaffie, A. Soliman, H.A. Khalifeh, M. Ghazal, F. Taher, A. Elmaghraby, et al., Radiomic-based framework for early diagnosis of lung cancer, in: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019). IEEE, 2019, pp. 1293–1297.
- [139] A. Shaffie, A. Soliman, M. Ghazal, F. Taher, N. Dunlap, B. Wang, et al., A novel autoencoder-based diagnostic system for early assessment of lung cancer, in: 2018 25th IEEE International Conference on Image Processing (ICIP). IEEE, 2018, pp. 1393–1397.
- [140] A. Shaffie, A. Soliman, L. Fraiwan, M. Ghazal, F. Taher, N. Dunlap, et al., A generalized deep learning-based diagnostic system for early diagnosis of various types of pulmonary nodules, *Technol. Cancer Res. Treat.* 17 (2018) 1533033818798800.
- [141] Y. ElNakieb, M.T. Ali, O. Dekhil, M.E. Khalefa, A. Soliman, A. Shalaby, et al., Toward accurate personalized autism diagnosis using different imaging modalities: SMRI, FMRI, and DTI, in: 2018 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT). IEEE, 2018, pp. 447–452.
- [142] Y. ElNakieb, A. Soliman, A. Mahmoud, O. Dekhil, A. Shalaby, M. Ghazal, et al., Autism spectrum disorder diagnosis framework using diffusion tensor imaging, in: 2019 IEEE International Conference on Imaging Systems and Techniques (IST). IEEE, 2019, pp. 1–5.
- [143] R. Haweel, O. Dekhil, A. Shalaby, A. Mahmoud, M. Ghazal, R. Keynton, et al., A machine learning approach for grading autism severity levels using task-based functional MRI, in: 2019 IEEE International Conference on Imaging Systems and Techniques (IST). IEEE, 2019, pp. 1–5.

- [144] O. Dekhil, M. Ali, R. Haweel, Y. Elnakib, M. Ghazal, H. Hajjdiab, et al., A comprehensive framework for differentiating autism spectrum disorder from neurotypicals by fusing structural MRI and resting state functional MRI, *Seminars in Pediatric Neurology*, Elsevier, 2020, p. 100805.
- [145] R. Haweel, O. Dekhil, A. Shalaby, A. Mahmoud, M. Ghazal, A. Khalil, et al., A novel framework for grading autism severity using task-based FMRI, in: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI). IEEE, 2020, pp. 1404–1407.
- [146] B. Dombroski, M. Nitzken, A. Elnakib, F. Khalifa, A. El-Baz, M.F. Casanova, Cortical surface complexity in a population-based normative sample, *Transl. Neurosci.* 5 (1) (2014) 17–24.
- [147] A. El-Baz, M. Casanova, G. Gimel'farb, M. Mott, A. Switala, An MRI-based diagnostic framework for early diagnosis of dyslexia, *Int. J. Computer Assist. Radiology Surg.* 3 (3-4) (2008) 181–189.
- [148] A. El-Baz, M. Casanova, G. Gimel'farb, M. Mott, A. Switala, E. Vanbogaert, et al., A new CAD system for early diagnosis of dyslexic brains, in: Proc. International Conference on Image Processing (ICIP'2008). IEEE, 2008, pp. 1820–1823.
- [149] A. El-Baz, M.F. Casanova, G. Gimel'farb, M. Mott, A.E. Switala, A new image analysis approach for automatic classification of autistic brains, in: Proc. IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI'2007). IEEE, 2007, pp. 352–355.
- [150] A. El-Baz, A. Elnakib, F. Khalifa, M.A. El-Ghar, P. McClure, A. Soliman, G. Gimel'farb, Precise segmentation of 3-D magnetic resonance angiography, *IEEE Trans. Biomed. Eng.* 59 (7) (2012) 2019–2029.
- [151] A. El-Baz, A.A. Farag, G. Gimel'farb, S.G. Hushek, Automatic cerebrovascular segmentation by accurate probabilistic modeling of Tof-MRA images, *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2005*, Springer, 2005, pp. 34–42.
- [152] A. El-Baz, A. Farag, A. Elnakib, M.F. Casanova, G. Gimel'farb, A.E. Switala, et al., Accurate automated detection of autism related corpus callosum abnormalities, *J. Med. Syst.* 35 (5) (2011) 929–939.
- [153] M. Ali, Y. Elnakieb, A. Shalaby, A. Mahmoud, A. Switala, M. Ghazal, et al., Autism classification using MRI: a recursive features selection based on sampling from multi-level high dimensional spaces, in: Proceedings of International Symposium on Biomedical Imaging (ISBI'21), pp. 267–270, 2021.
- [154] A. El-Baz, G. Gimel'farb, R. Falk, M.A. El-Ghar, V. Kumar, D. Heredia, A novel 3D joint Markov-Gibbs model for extracting blood vessels from PC–mra images, *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2009*, 5762, Springer, 2009, pp. 943–950.
- [155] A. Elnakib, A. El-Baz, M.F. Casanova, G. Gimel'farb, A.E. Switala, Image-based detection of corpus callosum variability for more accurate discrimination between dyslexic and normal brains, in: Proc. IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI'2010). IEEE, 2010, pp. 109–112.
- [156] A. Elnakib, M.F. Casanova, G. Gimel'farb, A.E. Switala, A. El-Baz, Autism diagnostics by centerline-based shape analysis of the corpus callosum, in: Proc. IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI'2011). IEEE, 2011, pp. 1843–1846.

- [157] A. Elnakib, M. Nitzken, M. Casanova, H. Park, G. Gimel'farb, A. El-Baz, Quantification of age-related brain cortex change using 3D shape analysis, in: *Pattern Recognition (ICPR), 2012 21st International Conference on*. IEEE, 2012, pp. 41–44.
- [158] M. Nitzken, M. Casanova, G. Gimel'farb, A. Elnakib, F. Khalifa, A. Switala, et al., 3D shape analysis of the brain cortex with application to dyslexia, in: *Image Processing (ICIP), 2011 18th IEEE International Conference on*. Brussels, Belgium: IEEE, Sep. 2011, pp. 2657–2660, (Selected for oral presentation. Oral acceptance rate is 10 percent and the overall acceptance rate is 35 percent).
- [159] F.E.-Z.A. El-Gamal, M.M. Elmogy, M. Ghazal, A. Atwan, G.N. Barnes, M.F. Casanova, et al., A novel cad system for local and global early diagnosis of Alzheimer's disease based on PIB-PET scans, in: *2017 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2017, pp. 3270–3274.
- [160] M.M. Ismail, R.S. Keynton, M.M. Mostapha, A.H. ElTanboly, M.F. Casanova, G.L. Gimel'farb, et al., Studying autism spectrum disorder with structural and diffusion magnetic resonance imaging: a survey, *Front. Hum. Neurosci.* 10 (2016) 211.
- [161] A. Alansary, M. Ismail, A. Soliman, F. Khalifa, M. Nitzken, A. Elnakib, et al., Infant brain extraction in t1-weighted MR images using bet and refinement using LCDG and MGRF models, *IEEE J. Biomed. health Inform.* 20 (3) (2016) 925–935.
- [162] E.H. Asl, M. Ghazal, A. Mahmoud, A. Aslantas, A. Shalaby, M. Casanova, et al., Alzheimer's disease diagnostics by a 3d deeply supervised adaptable convolutional network, *Front. Biosci. (Landmark Ed.)* 23 (2018) 584–596.
- [163] O. Dekhil, M. Ali, Y. El-Nakieb, A. Shalaby, A. Soliman, A. Switala, et al., A personalized autism diagnosis cad system using a fusion of structural mri and resting-state functional mri data, *Front. Psychiatry* 10 (2019) 392. Available from: <https://www.frontiersin.org/article/10.3389/fpsy.2019.00392>.
- [164] O. Dekhil, A. Shalaby, A. Soliman, A. Mahmoud, M. Kong, G. Barnes, et al., Identifying brain areas correlated with ADOS raw scores by studying altered dynamic functional connectivity patterns, *Med. Image Anal.* 68 (2021) 101899.
- [165] Y.A. Elnakieb, M.T. Ali, A. Soliman, A.H. Mahmoud, A.M. Shalaby, N.S. Alghamdi, et al., Computer aided autism diagnosis using diffusion tensor imaging, *IEEE Access.* 8 (2020) 191. 298–191 308.
- [166] M.T. Ali, Y. ElNakieb, A. Elnakib, A. Shalaby, A. Mahmoud, M. Ghazal, et al., The role of structure MRI in diagnosing autism, *Diagnostics* 12 (1) (2022) 165.
- [167] Y. ElNakieb, M.T. Ali, A. Elnakib, A. Shalaby, A. Soliman, A. Mahmoud, et al., The role of diffusion tensor MR imaging (DTI) of the brain in diagnosing autism spectrum disorder: promising results, *Sensors* 21 (24) (2021) 8171.
- [168] A. Mahmoud, A. El-Barkouky, H. Farag, J. Graham, A. Farag, A non-invasive method for measuring blood flow rate in superficial veins from a single thermal image, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2013*, pp. 354–359.
- [169] A. El-Baz, A. Farag, G. Gimel'farb, M.A. El-Ghar, T. Eldiasty, Probabilistic modeling of blood vessels for segmenting MRA images. In *L 18th International Conference on Pattern Recognition (ICPR'06)*, 3. IEEE, 2006, pp. 917–920.
- [170] A. El-Baz, A.A. Farag, G. Gimel'farb, M.A. El-Ghar, T. Eldiasty, A new adaptive probabilistic model of blood vessels for segmenting MRA images, *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2006*, 4191, Springer, 2006, pp. 799–806.

- [171] A. El-baz, A. Shalaby, F. Taher, M. El-Baz, M. Ghazal, M.A. El-Ghar, et al., Probabilistic modeling of blood vessels for segmenting magnetic resonance angiography images in *Med. Res. Arch.* 5 (3) (2017).
- [172] A.S. Chowdhury, A.K. Rudra, M. Sen, A. Elnakib, A. El-Baz, Cerebral white matter segmentation from MRI using probabilistic graph cuts and geometric shape priors, in: *ICIP*, 2010, pp. 3649–3652.
- [173] Y. Gebru, G. Giridharan, M. Ghazal, A. Mahmoud, A. Shalaby, A. El-Baz, Detection of cerebrovascular changes using magnetic resonance angiography, *Cardiovascular Imaging and Image Analysis*, CRC Press, 2018, pp. 1–22.
- [174] A. Mahmoud, A. Shalaby, F. Taher, M. El-Baz, J.S. Suri, A. El-Baz, Vascular tree segmentation from different image modalities, *Cardiovascular Imaging and Image Analysis*, CRC Press, 2018, pp. 43–70.
- [175] F. Taher, A. Mahmoud, A. Shalaby, A. El-Baz, A review on the cerebrovascular segmentation methods, in: *2018 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*. IEEE, 2018, pp. 359–364.
- [176] H. Kandil, A. Soliman, L. Fraiwan, A. Shalaby, A. Mahmoud, A. ElTanboly, et al., A novel MRA framework based on integrated global and local analysis for accurate segmentation of the cerebral vascular system, in: *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. IEEE, 2018, pp. 1365–1368.
- [177] F. Taher, A. Soliman, H. Kandil, A. Mahmoud, A. Shalaby, G. Gimel'farb, et al., Accurate segmentation of cerebrovasculature from Tof-MRA images using appearance descriptors. *IEEE Access*, 2020.
- [178] F. Taher, A. Soliman, H. Kandil, A. Mahmoud, A. Shalaby, G. Gimel'farb, et al., Precise cerebrovascular segmentation, in: *2020 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2020, pp. 394–397.
- [179] A.A. Sleman, A. Soliman, M. Ghazal, H. Sandhu, S. Schaal, A. Elmaghraby, et al., Retinal layers OCT scans 3-D segmentation, in: *2019 IEEE International Conference on Imaging Systems and Techniques (IST)*. IEEE, 2019, pp. 1–6.
- [180] N. Eladawi, M. Elmogy, M. Ghazal, O. Helmy, A. Aboelfetouh, A. Riad, et al., Classification of retinal diseases based on oct images, *Front. Biosci. (Landmark Ed.)* 23 (2018) 247–264.
- [181] A. ElTanboly, M. Ismail, A. Shalaby, A. Switala, A. El-Baz, S. Schaal, et al., A computer-aided diagnostic system for detecting diabetic retinopathy in optical coherence tomography images, *Med. Phys.* 44 (3) (2017) 914–923.
- [182] H.S. Sandhu, A. El-Baz, J.M. Seddon, Progress in automated deep learning for macular degeneration, *JAMA Ophthalmol.* (2018).
- [183] M. Ghazal, S.S. Ali, A.H. Mahmoud, A.M. Shalaby, A. El-Baz, Accurate detection of non-proliferative diabetic retinopathy in optical coherence tomography images using convolutional neural networks, *IEEE Access.* 8 (2020) 34. 387–34 397.
- [184] K. Hammouda, F. Khalifa, A. Soliman, M. Ghazal, M. Abou El-Ghar, A. Haddad, et al., A deep learning-based approach for accurate segmentation of bladder wall using MR images, in: *2019 IEEE International Conference on Imaging Systems and Techniques (IST)*. IEEE, 2019, pp. 1–6.
- [185] K. Hammouda, F. Khalifa, A. Soliman, H. Abdeltawab, M. Ghazal, M. Abou El-Ghar, et al., A 3D CNN with a learnable adaptive shape prior for accurate segmentation of bladder wall using MR images, in: *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2020, pp. 935–938.

- [186] A.H. Mahmoud, Utilizing Radiation for Smart Robotic Applications Using Visible, Thermal, and Polarization Images, (Ph.D. dissertation), University of Louisville, 2014.
- [187] A. Mahmoud, A. El-Barkouky, J. Graham, A. Farag, Pedestrian detection using mixed partial derivative based histogram of oriented gradients, in: 2014 IEEE International Conference on Image Processing (ICIP). IEEE, 2014, pp. 2334–2337.
- [188] A. El-Barkouky, A. Mahmoud, J. Graham, A. Farag, An interactive educational drawing system using a humanoid robot and light polarization, in: 2013 IEEE International Conference on Image Processing. IEEE, 2013, pp. 3407–3411.
- [189] A.H. Mahmoud, M.T. El-Melegy, A.A. Farag, Direct method for shape recovery from polarization and shading, in: 2012 19th IEEE International Conference on Image Processing. IEEE, 2012, pp. 1769–1772.
- [190] M.A. Ghazal, A. Mahmoud, A. Aslantas, A. Soliman, A. Shalaby, J.A. Benediktsson, et al., Vegetation cover estimation using convolutional neural networks, *IEEE Access*. 7 (2019) 132. 563–132 576.
- [191] M. Ghazal, A. Mahmoud, A. Shalaby, A. El-Baz, Automated framework for accurate segmentation of leaf images for plant health assessment, *Environ. Monit. Assess.* 191 (8) (2019) 491.
- [192] M. Ghazal, A. Mahmoud, A. Shalaby, S. Shaker, A. Khelifi, A. El-Baz, Precise statistical approach for leaf segmentation, in: 2020 IEEE International Conference on Image Processing (ICIP). IEEE, 2020, pp. 2985–2989.

This page intentionally left blank

Index

Note: Page numbers followed by “f” and “t” refer to figures and tables, respectively.

A

Abdullah University Hospital 2020 dataset, 208
Ablation study, 141–142
Activation Maximization approach, 163–165
Active contour model, 8–10
Acute lymphoblastic leukemia (ALL)
 in clinical remission, 1–2
 computer-assisted solution
 cancer and healthy cells, 19–24
 data preparation, 3–5
 normalization of color stain, 5–8
 segmentation of cells of interest, 8–19
 healthy progenitor cells and cancer blood cells, 1
Adaptive Graph Convolution Network (AGCN), 107–108
 bilateral filter layer, 117–118
 graph gather, 117
 graph pooling, 117
 network configuration, 118
 padding, 118
Adaptive residual graphs, 107
Adenocarcinoma (ADC), 124–125
Advanced computational algorithms, 39–40
Adventitious lung sounds (ALSs), 193–194
Affine registration, 154
AI-Clinical Decision Support System (AI-CDSS), 218–219
ALS classification, 194
Alzheimer’s Disease Neuroimaging Initiative (ADNI) dataset, 139–140
American Cancer Society, 265
American Heart Association Get With the Guidelines Heart Failure registry, 222
Anaplastic thyroid cancer, 265
Apparent diffusion coefficient (ADC), 84–85, 266
Area under curve (AUC), 151, 276, 279
Artificial intelligence (AI), 38–39, 49, 266
 benign *versus* malignant renal tumors, 242
 characterization of small renal mass, 244
 concept attribution, 165–166
 diagnostic analysis, 238
 evaluation of, 166–168
 feature extraction and qualifications, 238
 image acquisition, 237
 image segmentation, 237–238
 for medical applications, 161–163
 motivations, 159–160

 renal cell carcinoma
 vs. angiomyolipoma, 242–243
 grading of, 244
 merits and limitations, 245
 vs. oncocytoma, 243
 vs. renal cyst, 243
 staging of, 244
 subtyping of, 243–244
 visualization methods and feature attribution, 163–165
Artificial neural network models, 223
Atrial fibrillation (AF), 220
Attention on graph, 112–113
Audio signal preprocessing, 194–195
Autoencoders, 50–51
Automated decision-making, 188
Automated detection, 194
Automatic image analysis methods, 49

B

Background (BG) intensity, 6–7
Baseline convolutional neural networks
 approaches, 136, 139
Base module, 16
Beer-Lambert law, 6–7
“Benign vs malignant” classification, 64
Binary cross-entropy (BCE) loss function, 15
B-lineage lymphoblasts, 10–11
Boltzmann machine, 12
Boolean indicator, 122–123
b-value, 85

C

Cancer propagation, 266–267
Cardiac magnetic resonance imaging (cMRI), 223
Cardiac resynchronization therapy (CRT), 224
Causal concept effect, 165–166
Cell proliferation process, 266–267
Cell segmentation, 5–6
Class activation mapping (CAM), 163–165
C-NMC 2019 challenge dataset, 20–24
Cobblestone pattern, 63
Color deconvolution methods, 6–7
Comparison of Medical Therapy, Pacing, and Defibrillation in Heart Failure (COMPANION) trial, 224–225
Computational lung sound analysis (CLSA), 194

- Computational methods, 194
 - Computation capacity, 135–136
 - Computer-aided diagnosis (CAD) system, 48, 51, 235, 266
 - Computer-aided segmentation methods, 30–31
 - Computer vision, in melanoma diagnosis, 50–52
 - Concept activation vectors (CAVs), 158
 - Concept-based methods, 158
 - Conceptual sensitivities, 183–184
 - Concordance probability (C-index), 126
 - Conditional Random Fields (CRF), 31
 - Confirmation bias, 166–167
 - Continuous adventitious sounds (CASs), 193–194
 - Conventional classifiers, 204
 - Conventional machine learning, 194–195
 - Conventional neural networks, 88
 - Convolutional kernel, 105–106
 - Convolutional layers, 88–89
 - Convolutional neural networks (CNNs), 12–13, 31, 50–51, 67–69, 67*f*, 84–85, 105, 267
 - CNN-based classification, 88–89
 - custom CNN architectures, 204
 - Convolutional U-Net architecture, 49
 - COVID-19, 193
 - Cox proportional hazards model, 113
 - Cumulative Tortuosity Index (CTI), 177–178
 - Current computer-aided diagnosis (CAD) systems, 84–85
- D**
- Data augmentation, 128, 151, 194–195
 - time domain, 201–202
 - time–frequency domain, 202
 - Data compression, 50–51
 - Data preparation, 52
 - Data preprocessing, 52
 - Data processing, 194–195
 - amplitude scaling, 196
 - audio signal preprocessing, 195–197
 - noise filtering, 196
 - padding, 197
 - resampling, 196
 - segment splitting, 197
 - Decision tree, 276
 - Decoder module, 16
 - Deep autoencoders, 50–51
 - Deep convolutional neural networks (DCNNs), 134, 147
 - DeepGraphSurv framework, 120–122
 - baseline model, 125
 - dataset, 124–125
 - experimental result, 126–128
 - Deep learning (DL), 12, 29, 147, 157, 236
 - in brain cancer
 - pseudoprogression, 35
 - radiogenomics, 33–35
 - survival analysis, 32–33
 - tumor grading, 31–32
 - in breast cancer
 - accuracy in assessment, 36–37
 - breast density in the clinical setting, 37–38
 - clinical practice, 39
 - diagnosis, 38–39
 - local and global patterns detection algorithms, 60–63
 - Deep slice interpolation, 143
 - Deformation-based methods, 133–134
 - Dermoscopic images, 71
 - Dermoscopy, 51, 53
 - Diagnostic accuracy, of 2D model
 - ablation study, 274
 - hand-crafted-based techniques, 274–276
 - Diagnostic analysis, 238
 - Dice Coefficient (DICE), 140
 - Differentiated thyroid cancer (DTC), 265
 - Diffusion coefficient calculations, 268–270
 - Diffusion-weighted imaging (DWI), 84–85, 266
 - feature extraction, 86–87
 - Digital breast tomosynthesis (DBT), 36
 - Digital pathology, 51
 - Digital recording, 194
 - Digital rectal exam (DRE), 83–84
 - Discontinuous adventitious sounds (DASs), 193–194
 - Dominant deep learning techniques, 88
 - Drug-property prediction
 - baseline model, 122
 - dataset, 122–123
 - experimental result, 123–124
 - 2D Single Image Super-Resolution (SISR), 135
 - 2D texture model, 270–272
 - 3D texture model, 272
 - Dynamic contrast-enhanced (DCE-MRI), 147
- E**
- Early morphological nonresponders (EMNo-Res)
 - design, 148
 - Early morphological response (EMR), 148
 - Embedding, 3
 - Encoder–decoder-based CNN for Nuclei-instance Segmentation (EDNiS-Net), 15
 - Encoder module, 16
 - Epiluminescence microscopy (ELM) images, 61–62
 - Ericsson mobile report, 73

Evaluation metrics, 140
 Explainable AI (XAI), 160–161
 Explainable model, 108

F

Feature attribution, 163–165
 Feature descriptors, 65–66
 Feature extraction, 194–195

- features for conventional classifiers, 197–200
- time-frequency representations for deep learning, 201

 Feature representation method, 52
 Field-programmable gate arrays, 70–72
 Follicular thyroid carcinoma (FTC), 265
 Fully convolutional network (FCN), 13–14
 Fully resolution convolutional network (FrCN), 58

G

Gaussian Mixture Model (GMM), 178–179
 GCTI-SN workflow, 7–8
 Generalization and classification, 52
 Generative visual rationales, 218
 Gini coefficient, 178–179
 Glioblastoma multiforme (GBM), 124–125
 Global Br scores, 184–185
 Global patterns analysis, 51
 Globular pattern, 63
 Grad-Cam algorithm, 152–153
 Gradient class activation maps (grad-CAM) algorithm, 152–153
 Graph attention network (GAT), 108, 119–120
 Graph convolutional neural networks

- spatial graph convolutional neural networks, 109–110
- spectral graph convolutional neural networks, 110–111

 Graph isomorphism network (GIN), 122
 Graph-structured data, 106

H

Hand-rafterd features, 279
 Hardware Description Language (HDL), 70
 Hardware implementations, 49
 Hausdorff Distance (HD), 140
 HF_Lung_V1 dataset, 208
 HF New York Heart Association severity, 220
 HF with mid-range ejection fraction (HFmrEF), 217
 HF with preserved ejection fraction (HFpEF), 217
 HF with reduced ejection fraction (HFrEF), 217
 High-grade gliomas (HGGs), 31–32

Homogeneous pattern, 63
 Human decision-making, 170–171
 Hybrid systems, 204–205

I

ICBHI 2017 dataset, 207
 Illumination condition, 5
 Illumination correction, 7
 Image clustering, 8–10
 Image preprocessing, 149
 Imaging Reporting And Data Systems (IRADS), 30
 Individualized management approach, 221–222
 Information retrieval, 50–51
 Initial chemotherapy, 148
 Intel, 72
 Intelligibility, 187
 Intensity thresholding based segmentation, 8–10
 International Brain Tumor Segmentation (BraTS), 34
 Interpretability, 157–158
 Intersect over Union (IoU), 55–56
 Intra-tumoral developments, 148
 Intra-tumoral variations, 147–148
 ISIC 2016 Skin Lesion Analysis Toward Melanoma Detection Challenge dataset, 57
 ISIC 2017 training set, 59
 Isocitrate dehydrogenase, 34

J

Jaccard Similarity Index (JSI), 55–56
 Jenner-Giemsa Stain, 9*f*

K

$k \times 3 \times 3$ convolutional filters, 16
 Kernel Density Estimation, 179*f*
 K-means clustering, 8–10, 12
 k-nearest neighbor classifiers, 19, 84–85, 195
 Knowledge distillation, 206

L

LASSO-Cox linear model, 125
 Layerwise Relevance Propagation (LRP), 163–165
 Learning-based super-resolution methods, 135–136
 Learning paradigm

- postprocessing, 206
- transfer learning, 205–206

 Leave one subject out (LOSO), 89

Left ventricular assist devices (LVAD) serve, 224
 leukemic cells, 1
 Linear interpolation, 135
 Linear least squares (LLS) estimation, 173
 Local-Interpretable Model Agnostic Explanations (LIME), 163–165
 Logistic regression, 195
 Low-grade gliomas (LGGs), 31–32
 Lung squamous cell carcinoma (LUSC), 124–125

M

Machine learning (ML), 8–10, 157, 194, 220, 236
 conventional classifiers, 203
 deep learning architectures, 203–205
 in heart failure
 automatic diagnosis, classification, and phenotyping of heart failure, 218–220
 detection of HF-associated arrhythmia, 220–221
 development of therapy, 224
 optimal therapy, 224–225
 prevention, 225–226
 prognostic prediction, 221–224
 Magnetic field strengths, 91
 Magnetic resonance imaging (MRI), 83–84, 133, 147–149, 265–266
 Mahalanobis distance metric parameters, 107–108
 Malignancy diagnostic accuracy, 266
 Marginal super-resolution (MSR), 134, 137–138
 Markov random field (MRF), 63
 Mean square distance (MSD), 8
 Medical diagnostic procedures, 53–55
 Medullary thyroid cancer, 265
 Melanocytic lesion classification, 49
 Melanoma
 artificial intelligence and computer vision in diagnosis of, 50–52
 in dermoscopy, 64–69
 early diagnosis of, 47–49
 medical diagnostic procedures for, 53–55
 skin mole segmentation methods, 55–60
 Menzies Method, 65
 Meta-Analysis Global Group in Chronic Heart Failure Risk Score (MAGGIC) risk score, 221–222
 6-methylguanine-DNA methyltransferase, 34–35
 Mixture of expert (MoE) layers, 204
 Model-based methods, 239
 Molecular graph, neural network on, 111–112
 Mounting, 3
 MTLSA, 125
 Multiclass classification tasks, 174
 Multi-task learning framework, 125

N

Naive Bayes (NB), 276
 Network probability, 184
 Neural network analysis, 218
 on molecular graph, 111–112
 for survival analysis, 113–114
 Node-wise attention network, 128
 Nodule segmentation, 268–270
 Noninvasive CAD systems, 281
 Non-negative matrix factorization (NMF), 6–7
 Nottingham grading (NGH), 170–171

O

OASIS dataset, 140

P

Papillary thyroid carcinoma (PTC), 265
 Parallel pattern, 63
 Parametric Response Map (PRM), 151
 Pathological complete response (pCR), 148
 Peak Signal-to-Noise Ratio (PSNR), 140
 PH2 dataset, 58
 Pigment network (reticular) pattern, 62
 Pixel-wise dermoscopic skin lesions segmentation approach, 58
 1p/19q, 33–34
 Prediction difference analysis method, 163–165
 Probabilistic classification, 195
 Proposed loss function, 16–17
 Prostate cancer
 diffusion-weighted imaging (DWI)
 CNN-based classification, 88–89
 experimental results, 89–91
 feature extraction, 86–87
 Prostate segmentation, 85–86
 Prostate specific antigen (PSA) screening, 83–84
 Pseudoprogression, 35
 Public lung sound datasets
 Abdullah University Hospital 2020 dataset, 208
 HF_Lung_V1 dataset, 208
 ICBHI 2017 dataset, 207
 PyTorch1, 139

Q

Quantitative validations, 153

R

Radiomics, 237
 Random forest (RF), 276
 Randomization tests, 166–167
 Random survival forest (RSF) model, 221–222
 Raw data collection, 268
 Receiver operating characteristic (ROC) curve, 91, 276
 Rectifying linear unit (ReLU), 88–89
 Recurrent neural networks (RNN), 31, 50–51, 204, 219–220
 Region-based segmentation approaches, 8–10
 Region of interest (ROI), 8–10, 30–31, 126
 Registration-based methods, 135
 Regression concept vectors, 158
 agglomerating scores, 175
 computation, 172–174, 182–183
 conceptual sensitivity, 174–175
 identification of, 170–172, 177–181
 Renal cell carcinomas (RCCs), 235
 vs. angiomyolipoma, 242–243
 grading of, 244
 merits and limitations, 245
 vs. oncocytoma, 243
 vs. renal cyst, 243
 staging of, 244
 subtyping of, 243–244
 Research community, 14
 Residual Dense Blocks (RDB), 135
 Residual graph, 106–108
 Retinopathy of prematurity
 background, 168–169
 network performance on, 176
 task and classification model, 170
 Retinopathy of Prematurity (ROP), 158–159
 REverse Time Attention model (RETAIN), 219–220

S

Saliency methods, 167–168
 Sectioning, 3
 Segmentation, 19–20, 52
 Self-constructed residual graph Laplacian, 107
 Semantic segmentation vs instance segmentation, 13–15
 Seven-Point Checklist Method (7PCL), 65
 Shape-based methods, 135
 Signal processing techniques, 194
 Simonyan’s saliency maps, 163–165
 Simple Linear Iterative Clustering, 163–165
 Singular value decomposition (SVD), 6–7
 Skin biopsy, 54, 54f
 Skin lesion segmentation, 49, 55

Slice interpolation, 133–134
 Spatial aggregation, 173
 Spectral graph convolution (SGC) layer, 107, 128
 learning residual graph Laplacian, 115–116
 re-parameterization on feature transform, 113–114
 Squamous-cell carcinoma (SCC), 124–125
 Stain chemical, 5
 Stain correction method, 7
 Staining, 3
 Stain normalization methods, 6
 Stain quantity, 5
 Standard CNN architectures, 204
 Statistical analysis, 266–267, 270
 Statistical results, 273
 Statistical techniques
 first-order statics, 239
 second-order statics, 239
 Stochastic Gradient Descent (SGD), 151
 Structured Similarity Index (SSIM), 140
 Subject-centric explanations (SCEs), 158, 160–161
 Support vector machine (SVM), 8–10, 30–31, 84–85, 90r, 195, 276
 Surgical extraction, 265–266
 Survival analysis, 32–33, 113–114

T

Tele dermatology, 73–74
 Telemonitoring to Improve Heart Failure Outcomes trial, 222
 3 Tesla (T) multimodal system, 84–85
 Texture analysis, 147–148, 272–273
 principles, 238–239
 Texture parameters
 filtration-histogram method, 240
 postprocessing software, 240–241
 Texture pattern visualization, 278
 The Cancer Imaging Archive (TCIA), 20
 Thyroid cancer, 265
 Time–frequency representations, 194
 Tissue fixation, 3
 Tissue processing, 3
 Tissue segmentation, 30–31
 Traditional slice interpolation methods, 135
 Transfer learning, 32–33, 205–206
 Transform methods, 239
 Transparency, 160–161
 Tumor volume cropping, 154
 Two-view fusion and refinement, 138

U

Ultrasound (US) imaging, 265–266
 U-Net model, 14, 61–62, 176

V

Vessel segmentations, 179–180, 181*f*
Visual comparisons, 140–141
Visualization methods, 163–165

W

Welch two-sample *t*-test, 279

Whole slide images (WSIs), 108, 113–114,
124–125, 125*t*, 128

World Health Organization (WHO), 193, 235
WSISA, 113–114, 125

X

XAI visualization methods, 167–168
Xilinx, 72

STATE OF THE ART IN NEURAL NETWORKS AND THEIR APPLICATIONS

VOLUME 2

Examines the latest applications of neural networks in the medical field

Edited by

Ayman S. El-Baz and Jasjit S. Suri

KEY FEATURES

- Includes applications of neural networks, AI, machine learning, and deep learning techniques to a variety of imaging technologies
- Includes in-depth technical coverage of computer-aided diagnosis (CAD) for several diseases and disorders, including leukemia, brain and breast cancer, skin diseases, prostate cancer, and heart failure
- Written to help engineers, computer scientists, researchers, and clinicians understand the technology and applications of artificial neural networks

State of the Art in Neural Networks and Their Applications, Volume 2 presents the latest advances in artificial neural networks and their applications across a wide range of clinical diagnoses. Advances in the role of machine learning, artificial intelligence, deep learning, cognitive image processing, and suitable data analytics useful for clinical diagnosis and research applications are covered, including relevant case studies. The application of neural networks, artificial intelligence, and machine learning methods in biomedical image analysis has resulted in the development of CAD systems that aim toward the automatic early detection of several severe diseases.

ABOUT THE EDITORS

Ayman S. El-Baz is a distinguished professor at the University of Louisville, Kentucky, United States and the University of Louisville at Alamein International University (UofL-AIU), New Alamein City, Egypt. Dr. El-Baz was named as a fellow for Coulter, AIMBE, and NAI for his contributions to the field of biomedical translational research. Dr. El-Baz has almost two decades of hands-on experience in the fields of bioimaging modeling and noninvasive computer-assisted diagnosis systems. He has authored or coauthored more than 700 technical articles.

Jasjit S. Suri is an innovator, scientist, visionary, industrialist, and an internationally known world leader in biomedical engineering. Dr. Suri has spent over 25 years in the field of biomedical engineering/devices and its management. Dr. Suri was crowned with President's Gold Medal in 1980 and made a fellow of the IEEE, AIMBE, AIUM, SVM, and APVS for his outstanding contributions. He was awarded the Marquis Lifetime Achievement Award for his outstanding contributions to medical imaging. His main interests are in artificial intelligence in healthcare.



ACADEMIC PRESS

An imprint of Elsevier

elsevier.com/books-and-journals

ISBN 978-0-12-819872-8



9 780128 198728