

Medical Imaging

Artificial Intelligence, Image
Recognition, and Machine
Learning Techniques



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Medical Imaging

Artificial Intelligence, Image Recognition, and Machine Learning Techniques

K.C. Santosh, Sameer Antani,
D.S. Guru, and Nilanjan Dey



CRC Press

Taylor & Francis Group

Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

MATLAB® and Simulink® are trademarks of the MathWorks, Inc. and are used with permission. The MathWorks does not warrant the accuracy of the text or exercises in this book. This book's use or discussion of MATLAB® and Simulink® software or related products does not constitute endorsement or sponsorship by the MathWorks of a particular pedagogical approach or particular use of the MATLAB® and Simulink® software.

CRC Press
Taylor & Francis Group
52 Vanderbilt Avenue,
New York, NY 10017

© 2020 by Taylor & Francis Group, LLC
CRC Press is an imprint of Taylor & Francis Group, an Informa business

No claim to original U.S. Government works

Printed on acid-free paper

International Standard Book Number-13: 978-0-367-13961-2 (Hardback)

This book contains information obtained from authentic and highly regarded sources. Reasonable efforts have been made to publish reliable data and information, but the author and publisher cannot assume responsibility for the validity of all materials or the consequences of their use. The authors and publishers have attempted to trace the copyright holders of all material reproduced in this publication and apologize to copyright holders if permission to publish in this form has not been obtained. If any copyright material has not been acknowledged please write and let us know so we may rectify in any future reprint.

Except as permitted under U.S. Copyright Law, no part of this book may be reprinted, reproduced, transmitted, or utilized in any form by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying, microfilming, and recording, or in any information storage or retrieval system, without written permission from the publishers.

For permission to photocopy or use material electronically from this work, please access www.copyright.com (<http://www.copyright.com/>) or contact the Copyright Clearance Center, Inc. (CCC), 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400. CCC is a not-for-profit organization that provides licenses and registration for a variety of users. For organizations that have been granted a photocopy license by the CCC, a separate system of payment has been arranged.

Trademark Notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Library of Congress Cataloging-in-Publication Data

Names: Santosh, K. C., editor.

Title: Medical imaging : artificial intelligence, image recognition, and machine learning techniques / [edited by] KC Santosh, Sameer Antani, DS Guru, Nilanjan Dey.

Description: Boca Raton : Taylor & Francis, a CRC title, part of the Taylor & Francis imprint, a member of the Taylor & Francis Group, the academic division of T&F Informa, plc, 2020. | Includes bibliographical references and index.

Identifiers: LCCN 2019015642 | ISBN 9780367139612 (hardback : acid-free paper) | ISBN 9780429029417 (e-book)

Subjects: LCSH: Diagnostic imaging.

Classification: LCC RC78.7D53 M43215 2020 | DDC 616.07/54--dc23

LC record available at <https://lcn.loc.gov/2019015642>

Visit the Taylor & Francis Web site at
<http://www.taylorandfrancis.com>

and the CRC Press Web site at
<http://www.crcpress.com>

Contents

Preface.....	vii
Editors.....	xi
1. A Novel Stacked Model Ensemble for Improved TB Detection in Chest Radiographs.....	1
<i>Sivaramakrishnan Rajaraman, Sema Candemir, Zhiyun Xue, Philip Alderson, George Thoma, and Sameer Antani</i>	
2. The Role of Artificial Intelligence (AI) in Medical Imaging: General Radiologic and Urologic Applications.....	27
<i>Diboro Kanabolo and Mohan S. Gundeti</i>	
3. Early Detection of Epileptic Seizures Based on Scalp EEG Signals	43
<i>Abhishek Agrawal, Lalit Garg, Eliazar Elisha Audu, Ram Bilas Pachori, and Justin H.G. Dauwels</i>	
4. Fractal Analysis in Histology Classification of Non-Small Cell Lung Cancer	63
<i>Ravindra Patil, Geetha M., Srinidhi Bhat, Dinesh M.S., Leonard Wee, and Andre Dekker</i>	
5. Multi-Feature-Based Classification of Osteoarthritis in Knee Joint X-Ray Images	75
<i>Ravindra S. Hegadi, Dattatray N. Navale, Trupti D. Pawar, and Darshan D. Ruikar</i>	
6. Detection and Classification of Non-Proliferative Diabetic Retinopathy Lesions	97
<i>Ramesh R. Manza, Bharti W. Gawali, Pravin Yannawar, and K.C. Santosh</i>	
7. Segmentation and Analysis of CT Images for Bone Fracture Detection and Labeling.....	131
<i>Darshan D. Ruikar, K.C. Santosh, and Ravindra S. Hegadi</i>	
8. A Systematic Review of 3D Imaging in Biomedical Applications.....	155
<i>Darshan D. Ruikar, Dattatray D. Sawat, K.C. Santosh, and Ravindra S. Hegadi</i>	

9. A Review on the Evolution of Comprehensive Information for Digital Sliding of Pathology and Medical Image Segmentation	183
<i>M. Ravi and Ravindra S. Hegadi</i>	
10. Pathological Medical Image Segmentation: A Quick Review Based on Parametric Techniques	207
<i>M. Ravi and Ravindra S. Hegadi</i>	
Index	235

Preface

This book aims to provide advanced or up-to-date techniques in medical imaging through the use of artificial intelligence (AI), image recognition (IR), and machine learning (ML) algorithms/techniques. An image or a picture is worth a thousand words; which means that image recognition can play a vital role in medical imaging and diagnostics, for instance. The data/information in the form of image, i.e. a set of pixels, can be learned via AI, IR, and ML, since it is impossible to employ experts for big data. The book covers several different topics, such as tuberculosis (TB) detection; radiologic and urologic applications; epileptic seizures detection; histology classification of non-small cell lung cancer; osteoarthritis classification (using knee joint X-ray); non-proliferative diabetic retinopathy lesions classification; fractured bone detection and labeling (using CT images); usefulness of 3D imaging (a quick review); and pathological medical imaging and segmentation.

In Chapter 1, authors discuss the stacked generalization of models for TB detection in chest radiographs. TB is an airborne infection and a common cause of death related to antimicrobial resistance. In resource-constrained settings, where there is a significant lack of expertise in interpreting radiology images, there is a need of image-analysis-based computer-aided diagnosis (CADx) tools. Such tools have gained significance because they offer a promise to alleviate the human burden in screening in countries that lack adequate radiology resources. Very specifically, authors reported the use of convolutional neural networks (CNN), a class of deep learning (DL) models. We observed that such tools deliver promising results on visual recognition tasks with end-to-end feature extraction and classification. Besides, ensemble learning (EL) methods combine multiple models to offer promising predictions because they allow the blending of intelligence from different learning algorithms.

In Chapter 2, the authors provide a thorough idea on how artificial intelligence (AI) tools can help in medical imaging, using radiologic and urologic applications. The authors are convinced of the fact that diagnostic errors account for approximately 10% of patient deaths, and between 6 and 17% of adverse events occurring during hospitalization. At a total of ~20 million radiology errors per year, and 30,000 practicing radiologists, this averages to just under 700 errors per practicing radiologist. Errors in diagnosis have been associated with clinical reasoning, including: intelligence, knowledge, age, affect, experience, physical state (fatigue), and gender (male predilection for risk taking). These factors, and the limited access to radiologic specialists for up to 2/3 of the world, encourage a more urgent role for the use of AI in medical imaging, a huge focus of which is machine learning. Also, operator dependency in radiologic procedures, particularly sonography, has

led researchers to develop automated image interpretation techniques similar to those used for histopathology. AI now allows for the connection of image analysis with diagnostic outcome, in real time. AI has the potential to assist with care, teaching, and diagnosis of illness. According to the market research firm Tractica, the market for virtual digital assistants worldwide will reach \$16 billion by 2021. The term “machine learning” as it applies to *radiomics* is used to describe high throughput extraction of quantitative imaging features with the intent of creating minable databases from radiological images.

In Chapter 3, the authors discuss the early detection of epileptic seizures, which is based on scalp electroencephalography (EEG) signals. Their research aims to realize a seizure detector using the empirical mode decomposition (EMD) algorithm and a machine learning–based classifier that is robust enough for practical applications. They have conducted exhaustive tests on EEG data of 24 pediatric patients who suffered from intractable seizures. Their tool may serve as a potential avenue for real-time seizure detection.

In Chapter 4, authors reported the usefulness of fractals in histology classification of non-small cell lung cancer (NSCLC). This type of cancer accounts for 85% of all the lung cancers. Noninvasive identification of the histology of NSCLC aids in determining the appropriate treatment approaches. In this study, the authors observed the usage of radiomics with application of fractals for decision-making: histology classification of NSCLC using lung CT images. Again, their study suggests that fractals can play a vital role in radiomics, providing information about the gross tumor volume (GTV) structure, and also helping characterizing the tumor.

In Chapter 5, the authors explain the use of multiple features to classify osteoarthritis (OA) in knee joint X-ray images. OA is a commonly occurring disease in the joints of the knee, hip, and hands. It results in a loss of cartilage. Affected patients will experience severe pain, stiffness, and a grating sensation during the movement of the joints. The authors reported that the use of several different features, such as edge, curvature, and textures, could improve the performance in classification (normal and/or abnormal), where conventional machine learning classifiers are used.

In Chapter 6, the authors explained how non-proliferative diabetic retinopathy (NPDR) lesions could be detected and classified. Diabetic retinopathy began with a leakage of blood or fluid from the retinal blood vessels and it damages the retina. NPDR is an early stage of diabetic retinopathy and it is categorized by three stages: mild, moderate, and severe. These were tested, and the authors achieved a classification accuracy of 94% using artificial neural network.

In Chapter 7, the authors explain the use of image segmentation so that image region labeling becomes easier. In their study, they discuss bone fracture detection and labeling in computed tomography (CT) images. CT images are a crucial resource for assessing the severity and prognosis

of bone injuries caused by trauma or accident. Similarly, fracture detection is a very challenging task. In their work, the authors developed a computer-aided diagnosis (CAD) system, which not only precisely extracts and assigns unique labels to each fractured piece by considering patient-specific bone anatomy, but also effectively removes unwanted artifacts (like flesh) surrounded by bone tissues. In their tests (real patient-specific CT images), they have reported the maximum possible accuracy of 95%.

In Chapter 8, the authors provide a systematic review on 3D imaging in biomedical applications. The volume visualization or 3D imaging field is vibrant and one of the fastest growing fields in scientific visualization. It is focused on creating high-quality 3D images from acquired volumetric datasets to gain insights into underlying data. In this work, the authors primarily review detailed information about the state-of-the-art volume visualization techniques majorly applied in the biomedical field. Besides, they provide commonly used tools and libraries that are employed for volume visualization. Further, several applications are discussed.

In Chapter 9, the authors discuss the evolution of the digital sliding of pathology in medical imaging. In general, they point out the evolution in the digitalization of pathological slides and explain the advantages of pathology practices in the prediction of diseases, in minimizing efforts, and in clarifying disease information via diagnosis. For example, examining tiny tissue uncovers data that could empower the pathologist to render an accurate analysis and provide help with treatments.

In Chapter 10, the authors provide a quick review on pathological medical segmentation, which is based on parametric techniques. In their study, several different segmentation techniques are considered. Further, authors point out the comparison among the techniques (publicly available) and help readers find an appropriate one.

MATLAB® is a registered trademark of The MathWorks, Inc. For product information, please contact:

The MathWorks, Inc.
3 Apple Hill Drive
Natick, MA 01760-2098 USA
Tel: 508 647 7000
Fax: 508-647-7001
E-mail: info@mathworks.com
Web: www.mathworks.com



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Editors

K.C. Santosh (Senior member, IEEE) is an assistant professor and graduate program director for the Department of Computer Science, University of South Dakota (USD). Also, Dr. Santosh is an associate professor (visiting) for the School of Computing and IT, Taylor's University. Before joining the USD, Dr. Santosh worked as a research fellow at the U.S. National Library of Medicine (NLM), National Institutes of Health (NIH). He worked as a post-doctoral research scientist at the LORIA research centre, Université de Lorraine in direct collaboration with ITESOFT, France (industrial partner). He received his PhD diploma in computer science from INRIA – Université de Lorraine (France), and his MS in computer science from Thammasat University (Thailand). Dr. Santosh has demonstrated expertise in pattern recognition, image processing, computer vision, artificial intelligence, and machine learning with various applications in medical image analysis, graphics recognition, document information content exploitation, and biometrics. He has published more than 120 peer-reviewed research articles and two authored books (Springer); he has also edited several books (Springer, Elsevier, and CRC Press), journal issues (Springer), and conference proceedings (Springer). Dr. Santosh serves as an associate editor of the *International Journal of Machine Learning and Cybernetics* (Springer). For more information, please visit: <http://kc-santosh.org>.

Sameer Antani is a staff scientist and (acting) chief of the communications engineering branch and the computer science branch, respectively, at the Lister Hill National Center for Biomedical Communications, an intramural R&D division of the National Library of Medicine (NLM) part of the National Institutes of Health (NIH) in Bethesda, Maryland. He is a versatile senior researcher, leading several scientific and technical research explorations to advancing the role of computational sciences and engineering in biomedical research, education, and clinical care. His research applies and studies methods for explaining the behavior of machine intelligence methods in automated decision support in biomedical applications. For this, he draws on his expertise in biomedical image informatics, automatic medical image interpretation, information retrieval, computer vision, and related topics in computer science and engineering technology. His contributions include automated screening for high burden diseases such as (i) Tuberculosis (TB) in HIV positive patients using digital chest X-ray image analysis; (ii) cervical cancer in women using analysis of acetowhitened images of the cervix, and whole slide images of liquid-based Pap smears and histopathology images from cervical biopsies; and (iii) automated methods for detecting malaria parasites in microscopic images of thick and thin blood smears. Other contributions include functional MRI (fMRI) simulation for brain research and similarity retrieval; analysis of ophthalmic fundus images for glaucoma, and the OPEN-i® biomedical image retrieval system, which provides

text and visual search capability to retrieve over 3.7 million images from about 1.2 million of NLM's PubMed Central® Open-Access articles and other image datasets. Dr. Antani is a senior member of the International Society of Photonics and Optics (SPIE), the Institute of Electrical and Electronics Engineers (IEEE). He serves as the vice chair for computational medicine on the IEEE Computer Society's Technical Committee on Computational Life Sciences (TCCLS), and on the editorial boards of *Heliyon* and *Data*.

Dr. Antani is a senior member of the International Society of Photonics and Optics (SPIE), the Institute of Electrical and Electronics Engineers (IEEE), and the IEEE Computer Society. He serves as the vice chair for computational medicine on the IEEE Technical Committee on Computational Life Sciences (TCCLS), and as an associate editor for the *IEEE Journal of Biomedical and Health Informatics*

D.S. Guru is a professor at the Department of studies in computer science. He is known for his contributions to the fields of image processing and pattern recognition. He is a recipient of the BOYSCAST Fellowship awarded by the Department of science and technology, Govt of India, and also of the Award for Research Publications by the Department of science and technology, Karnataka Government. He has been recognized as the best ethical teacher in higher learning by Rotary North Mysore. He has supervised more than 15 PhD students and is currently supervising many more. He holds rank positions in both bachelors and masters education. He earned his doctorate from the University of Mysore and did his post-doctoral work at PRIP Lab, Michigan State University. He has been a reviewer for international journals by Elsevier Science, Springer, and IEEE Transactions. He has chaired and delivered lectures at many international conferences and workshops. He is a co-author for three textbooks, co-editor of three proceedings, and author of many research articles both in peer-reviewed journals and in proceedings.

Nilanjan Dey is an assistant professor in Department of Information Technology at Techno India College of Technology, Kolkata, India. He is a visiting fellow of the University of Reading, UK. He was an honorary Visiting Scientist at Global Biomedical Technologies Inc., CA (2012–2015). He was awarded his PhD from Jadavpur University in 2015. He is the editor-in-chief of *International Journal of Ambient Computing and Intelligence* and associate editor of *IEEE Access* and *International Journal of Information Technology*. He is the series co-editor of *Springer Tracts in Nature-Inspired Computing*, series co-editor of *Advances in Ubiquitous Sensing Applications for Healthcare*, series editor of *Computational Intelligence in Engineering Problem Solving and Intelligent Signal Processing and Data Analysis*. His main research interests include medical imaging, machine learning, computer-aided diagnosis, data mining, etc. He is the Indian Ambassador for the International Federation for Information Processing (IFIP)—Young ICT Group. He has been awarded as one among the top ten most-published academics in the field of computer science in India (2015–2017).

1

A Novel Stacked Model Ensemble for Improved TB Detection in Chest Radiographs

Sivaramakrishnan Rajaraman, Sema Candemir, Zhiyun Xue,
Philip Alderson, George Thoma, and Sameer Antani

CONTENTS

1.1	Introduction.....	1
1.2	Materials and Methods.....	7
1.2.1	Data Collection and Preprocessing.....	7
1.2.2	Proposal 1—Feature Extraction Using Local/Global Feature Descriptors and Classification Using SVM.....	8
1.2.3	Proposal 2—Feature Extraction and Classification Using a Customized CNN.....	10
1.2.4	Proposal 3—Feature Extraction Using Pre-Trained CNNs and Classification Using SVM.....	11
1.2.5	Proposal 4—Constructing Stacked Model Ensembles.....	12
1.3	Results and Discussion.....	14
1.4	Conclusion and Future Work.....	21
	Acknowledgments.....	22
	Conflict of Interest.....	22
	References.....	22

1.1 Introduction

Tuberculosis (TB) is an infectious disease caused by a rod-shaped bacterium called *Mycobacterium tuberculosis*. According to the 2018 World Health Organization (WHO) report, there were an estimated 10 million new TB cases, but only 6.4 million (64%) were reported for treatment [1]. Countries including India, China, Pakistan, South Africa, and Nigeria accounted for more than 60% of the people suffering from the infection. A chest X-ray (CXR), also called chest film or chest radiograph, is the most common imaging modality used to diagnose conditions affecting the chest and its contents [2, 3]. CXR diagnosis has revolutionized the field of TB diagnostics and is extremely useful in establishing a plausible diagnosis of the infection.

Clinicians initiate treatment for the infection based on their judgment of these radiology reports. Posterior-anterior (PA) and lateral CXR projections are routinely examined to diagnose the conditions and provide diagnostic evidence [4]. Figure 1.1 (a)–(e) shows some instances of abnormal and normal CXRs.

With significant advancements in digital imaging technology there is an increase in the use of CXRs for TB screening. However, there is a lack of expertise in interpreting radiology images, especially in TB endemic regions, which adversely impacts screening efficacy [5], an ever-growing backlog and increased opportunity for disease spread. Also, studies show that there is a high degree of variability in the intra-reader and inter-reader agreement during the process of scoring CXRs [6]. Thus, current research is focused on developing cost-effective, computer-aided diagnosis (CADx) systems that can assist radiologists in interpreting CXRs and improve the quality of diagnostic imaging [7]. These systems are highly competent in reducing intra-reader/inter-reader variability and detection errors [8–11]. There are several prior approaches using traditional image analysis and machine learning (e.g. support vector machine [SVM]) that are valuable for providing background on CADx tools for CXR analysis [12–15]. The reader is referred to these as background. They are promoted as a convenient tool to be used in systematic screening and triaging algorithms due to the increased availability of digital radiography, which presents numerous benefits over conventional radiography, including enhanced image quality, safety, and reduced operating expenses [16]. CADx tools have gained immense significance; the appropriate use and advancing of these systems could improve detection accuracy and alleviate the human burden in screening. Earlier CADx studies were based on image segmentation and textural feature extraction with grey-level co-occurrence matrix [17]. A CADx system for TB detection was proposed by Van Ginneken et al. [18], who used multi-scale feature banks for feature extraction and a weighted nearest-neighbor classifier for classification of TB-positive and normal cases. The study demonstrated area under a curve (AUC) values of 0.986 and 0.82 on two private CXR datasets. A technique based on pixel-level textural abnormality detection was proposed by Hogeweg et al. [19] to obtain AUC values between 0.67 and 0.86. However, a

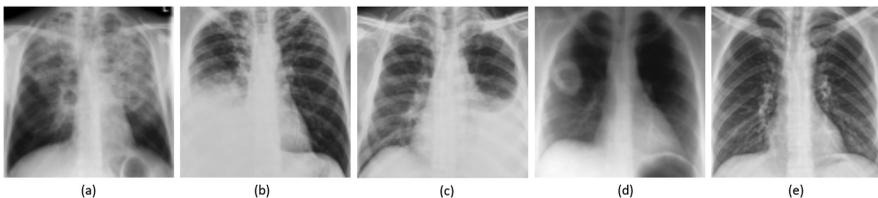


FIGURE 1.1

CXRs: (a) hyper-lucent cystic lesions in the upper lobes, (b) right pleural effusion, (c) left pleural effusion, (d) cavitary lung lesion in the right lung, and (e) normal lung.

comparative study of the proposed methods was hampered due to unavailability of public CXR datasets. Jaeger et al. [20] made available the public CXR datasets for TB detection, followed by Chauhan et al. [21], who helped to evaluate the proposed techniques on public datasets. Melendez et al. proposed the multiple instance learning methods for TB detection, which used moments of pixel intensities as features to be classified by an SVM classifier [22]. The authors obtained AUC between 0.86 and 0.91 by evaluating on three private CXR datasets. Jaeger et al. [5] proposed a combination of standard computer vision algorithms for extracting features from chest radiographs. The study segmented the region of interest (ROI) constituting the lungs, and extracted the features using a combination of algorithms that included a histogram of oriented gradients (HOG), local binary patterns (LBP), Tamura feature descriptors, and other algorithms. A binary classifier was trained on these extracted features to classify normal and TB-positive cases. CADx software based on machine learning (ML) approaches using a combination of textural and morphological features is also commercially available. This includes CAD4TB, a CADx software from the Image Analysis Group, Nijmegen, Netherlands that reported AUC ranging from 0.71 to 0.84 in a sequence of studies performed in detecting pulmonary abnormalities [23]. Another study achieved AUC of 0.87 to 0.90 by using an SVM classifier to classify pulmonary TB from the normal instances using texture and shape features [24]. However, the performance of textural features was found to be inconsistent across the imaging modalities. These features performed well as long as they were able to correlate with the disease, but delivered sub-optimal performance in instances when there was an overlapping of anatomical sites and images having complex appearances [25]. Feature descriptors such as bag-of-words (BOW) were also used in discriminating normal from pathological chest radiographs [26]. The method involves representing an image using a bag of visual words, constructed from a vocabulary of features extracted by local/global feature descriptors. A majority of CADx studies used handcrafted features that demand expertise in analyzing the images and account for variability in the morphology and texture of the ROI. On the other hand, deep learning (DL) models learn hierarchical layer-wise representation to model data at more and more abstract representations. These models are also known as hierarchical ML models that use a cascade of layers of non-linear processing units for end-to-end feature extraction and classification [27]. Convolutional neural networks (CNN), a class of DL models, have gained immense research prominence in tasks related to image classification, detection, and localization, as they deliver promising results without the need for manual feature selection [28]. Unlike kernel-based algorithms such as SVMs, DL models exhibit improved performance with an increasing number of training samples and computational resources [29].

Medical images contain visual representations of the interior of the body that aid in clinical analysis and medical intervention [30]. These images are specific to the internal structures of the body and have less in common

with natural images. Under these circumstances, a customized CNN, specifically trained on the underlying biomedical imagery, could learn “task-specific” features to aid in improved accuracy. The parameters of a custom model could be optimized for improvement in performance. The learned features and salient network activations could be visualized to understand the strategy the model adapts to learn these task-specific features [31]. However, the performance improvement of customized CNNs comes at the cost of huge amounts of labeled data, which are difficult to obtain, particularly in biomedical applications. Transfer Learning (TL) methods are commonly used to relieve issues with data inadequacy where DL models are pre-trained on large-scale datasets [32]. These pre-trained models could be used either as an initialization for visual recognition tasks or as feature extractors from the underlying data [33]. There are several pre-trained CNNs available, including AlexNet [34], VGGNet [35], GoogLeNet [36], ResNet [37], etc., which transfer knowledge gained from learning a comprehensive feature set from the large-scale datasets to the underlying task and serve as feature extractors in an extensive range of visual recognition applications, outperforming the handcrafted features [38]. Study of the literature reveals the use of pre-trained CNNs in detecting pleural effusion and cardiomegaly in chest radiographs [39]. The performance of the pre-trained models was compared to that of the classifiers operated on handcrafted features, including LBP [40], GIST [41], and PiCo descriptors [42]. It was observed that the combination of pre-trained CNN and PiCo features gave the best performance with an AUC of 0.89 and 0.93 for cardiomegaly and right pleural effusion respectively. The first application of CNNs to TB detection was proposed by Hwang et al. [43], who customized the architecture of AlexNet and trained on a private CXR dataset. The results obtained with random weight initializations were not promising; however, the model performed better, with an accuracy of 0.77 and an AUC of 0.82, when trained with pre-trained weights. The authors also evaluated the performance with publicly available Montgomery and Shenzhen datasets [20] to obtain an accuracy of 0.674 and 0.837, respectively. The application of CNN toward TB detection was demonstrated in another study that used a custom CNN model, a variant of the AlexNet framework, trained on a private CXR dataset of approximately 10,000 images [43]. The study achieved unsatisfactory results with the custom model with random weight initializations. However, with pre-trained CNNs, better results were obtained on the Montgomery and Shenzhen CXR datasets, achieving AUC of 0.884 and 0.926 respectively. In another study, the authors assessed the accuracy and stability in findings of DL models toward detecting abnormalities on frontal CXRs. The de-identified radiographs were processed with the Qure AI tool [44]. The scores were generated, recorded, and compared with the standard of reference (SOR) established with expert radiologists’ assessment toward detecting hilar prominence, pulmonary opacity, pleural effusion, and enlarged cardiac silhouette. It was observed that the value of

AUC ranged between 0.837–0.929 and 0.693–0.923 for DL and test radiologists, respectively. DL models had the lowest AUC of 0.758 toward assessing changes in pulmonary opacities. No statistical difference was observed between DL and SOR for these abnormalities. In a recent study, pre-trained CNNs were used as feature extractors for TB detection [45]. The study presents three different proposals for applying the pre-trained CNNs to extract features from the CXRs toward improving the accuracy of TB detection. The study used publicly available CXR datasets [20] and demonstrated superior performance to state-of-the-art methods.

The pioneering work on ensemble learning (EL) proved that multiple, diverse, and accurate base-learners could asymptotically build a strong-learner [46]. The generation of model ensembles could be categorized into homogeneous and heterogeneous methods. In the homogeneous method, base-learners use the same learning algorithm with different settings for training data and learning parameters. Some examples include the Bagging [47] and Boosting methods [48]. Heterogeneous methods generate base-learners with diverse learning algorithms. Different fusing strategies are used to combine the decision made by the base-learners. Majority voting [49] is a commonly used fusing method in which base-learners vote for a specific class and the predicted class collects the majority of votes. Simple average and weighted average methods [50] are also used. Stacking, otherwise called stacked generalization, is an optimal fusing technique that highlights each base-learner when it performs best and discredits it when it delivers sub-optimal performances [51]. Figure 1.2 gives a pictorial representation of this concept. This method introduces the concept of a second-level meta-learner that optimizes the combination of individual base-learners.

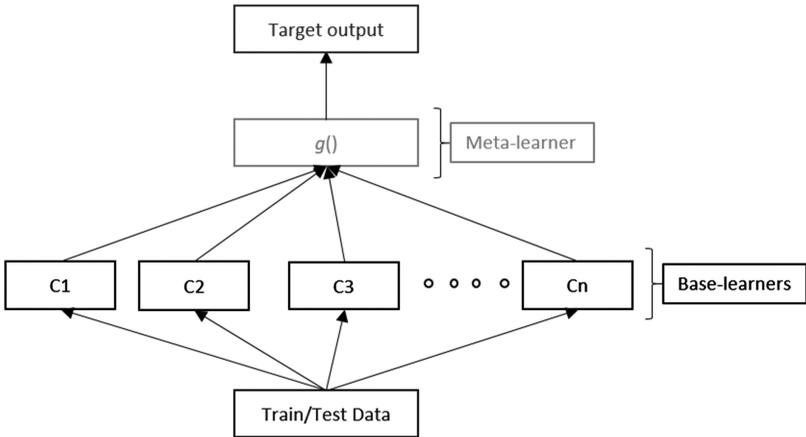


FIGURE 1.2
The concept of stacking.

Base-learners are diverse. They make different errors in the data and are accurate in different regions of the feature space. Individual base-learners are represented by $C = \{C_1, C_2, \dots, C_n\}$ where “n” is the number of base-learners. The meta-learner $g()$ learns how the individual base-learners make errors, and estimates and corrects their biases. Since both DL and EL have their inherent advantages in constructing non-linear decision-making functions, the combination of the two could efficiently handle tasks related to analyzing and understanding the underlying data. However, the literature is sparse in the use of ensemble models for TB detection. In the only study performed by Lakhani and Sundaram [52], the authors evaluated the efficacy of an ensemble of deep CNNs toward TB detection in chest radiographs. The study used publicly available CXR datasets, pre-trained, and untrained CNNs and performed ensembles on the best-performing models. It revealed that the ensemble of pre-trained CNNs along with data augmentation performed better with a sensitivity of 97.3% in comparison to other models.

This study aims to evaluate the performance of a stacked model ensemble that optimally combines the predictions using handcrafted feature descriptors and customized and pre-trained CNNs, and presents four different proposals toward improving the accuracy of TB detection from PA CXR images. In the first proposal, we use local and global feature descriptors including GIST, HOG, and SURF to extract the features from chest radiographs. The extracted features are used to train an SVM classifier to classify the normal and abnormal CXRs. In the second proposal, we evaluate the performance of a customized CNN-based DL model that learns task-specific features toward classifying TB-positive and healthy controls. The proposed model is optimized for its architecture and hyper-parameters by performing Bayesian optimization in the process of minimizing the classification error. In the third proposal, we use four different pre-trained CNN models to extract features from the chest radiographs, and an SVM classifier is trained on these features to detect TB manifestations. In the fourth and final proposal, we perform a stacked ensemble of models from different proposals to evaluate their performance on disease detection. The contributions of this work are as follows: (a) comparative analysis of the performance of local/global feature descriptors including GIST, HOG, and SURF toward classifying TB-positive and healthy chest radiographs, (b) proposing customized CNN-based DL models, optimized for their architecture and hyper-parameters toward learning task-specific features, (c) visualizing the learned features and salient network activations in the customized model to understand the learning dynamics, (d) comparing the performance of pre-trained DL models as feature extractors for the underlying task, and (e) evaluating the performance of stacked model ensembles toward the task of improving the accuracy of TB detection. This chapter is organized as follows: Section 1.2 elaborates on the materials and methods, Section 1.3 discusses the results, and Section 1.4 concludes the chapter.

1.2 Materials and Methods

1.2.1 Data Collection and Preprocessing

This study is evaluated on four CXR datasets that include the two publicly available datasets from Montgomery County, Maryland, and Shenzhen, China, maintained by the National Library of Medicine (NLM) [20]. The CXRs in the Montgomery collection have pixel resolutions of either 4892×4020 or 4020×4892 . The CXRs in the Shenzhen collection have resolutions of approximately 3000×3000 pixels. The Montgomery collection has 58 TB-positive cases and 80 healthy controls. The Shenzhen dataset has a total of 662 CXRs, which include 336 TB-positive CXRs and 326 healthy controls. Ground truth information for these datasets is available in the form of clinical findings, roughly annotating the abnormal locations in the CXR images. The acquisition and sharing of these datasets are exempted from National Institutes of Health (NIH) IRB review (#5357). The third dataset is from India, acquired by the National Institute of Tuberculosis and Respiratory Diseases, New Delhi, and made available by the authors [21]. This dataset contains two subsets of CXR collections from different X-ray machines and a balanced distribution of CXR images for TB-positive and normal cases. The CXRs in the India collection have resolutions ranging from 1024×1024 to 2480×2480 pixels. GT labels are available as global annotations for the normal and abnormal classes. TB manifestations in this dataset are obvious and distributed throughout the lungs. The fourth dataset is a private collection of CXRs obtained from Kenya under a retrospective study agreement with Moi Teaching and Referral Hospital, Eldoret, Kenya and with the assistance of Indiana University School of Medicine and Academic Model Providing Access to Healthcare (AMPATH), a Kenyan NGO. This dataset contains 238 abnormal CXRs and 729 healthy controls. Disease labels are made available as lung zone-based clinical findings from expert radiologists. The CXRs in the Kenya collection have resolutions of either 2004×2432 or 1932×2348 pixels.

The datasets include PA CXRs that contain regions other than the lungs which are irrelevant for lung TB detection. To alleviate issues due to models learning features that are irrelevant to detecting lung TB and demonstrate sub-optimal performance, the lung region constituting the ROI is segmented by a method that uses anatomical atlases with non-rigid registration [24]. This segmentation method follows a content-based image retrieval approach to identify the training examples that bear resemblance to the patient CXR by using Bhattacharyya similarity measure and partial Radon transform. The patient-specific anatomical lung shape model is created using SIFT-flow [53] for registering the training masks for the patient CXRs. The refined lung boundaries are extracted using graph-cut optimization and customized energy function [54]. An instance of a CXR with the detected lung region and cropped lung area using the proposed method is shown in Figure 1.3.

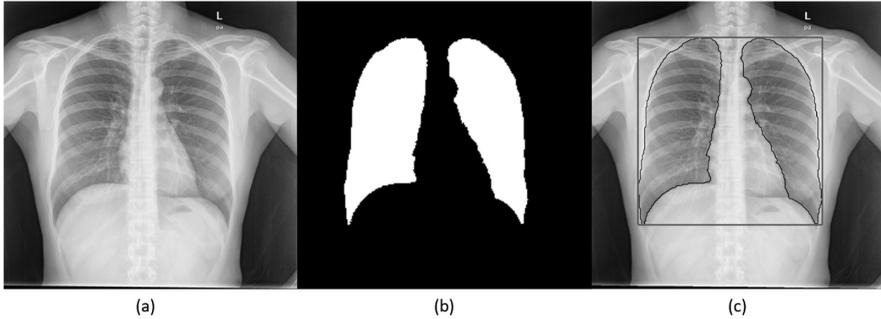


FIGURE 1.3

Lung ROI segmentation: (a) CXR, (b) computed lung mask, (c) Segmented ROI.

After lung segmentation, the resulting image is cropped to the size of a bounding box that contains all the lung pixels. The resultant images are enhanced for contrast by applying contrast limited adaptive histogram equalization (CLAHE).

1.2.2 Proposal 1—Feature Extraction Using Local/Global Feature Descriptors and Classification Using SVM

In the first proposal (P1), we evaluated the performance of global descriptors including GIST and HOG and local descriptors including SURF toward identifying TB manifestations. Since pre-trained CNNs demand down-sampling of the underlying data to fit the specific requirements of the input layer, a lot of potentially viable information pertaining to the signs of TB infection may be lost. The best way to overcome this issue is to use the local/global feature descriptors to extract discriminative information from the entire CXR image without the need for rigorous down-sampling. The GIST feature descriptor summarizes the information pertaining to the gradients, orientations, and scales for different regions of a given image to provide a robust description of the image. The process results in image filtering into low-level features including intensity, color, motion, and orientation at multiple scales in space. GIST captures these features toward identifying the salient image locations that significantly differ from those of the neighbors. Given an input image, the GIST descriptor convolves the image with 32 Gabor filters [55] at four different scales and eight different orientations, producing a total of 32 feature maps with the same size as that of the input image. The process results in the computation of the direction of low- and high-frequency repetitive gradients for a given image. Each of these feature maps is then divided into 16 regions with a 4×4 square grid and the feature values are averaged within each sub-region. The averaged values from the 16 sub-regions are concatenated for the 32 different feature maps, resulting in a total of 512 GIST descriptors for a given image. Across the datasets, eight orientations per scale and four blocks are used in this study.

HOG feature descriptors were introduced by Dalal and Triggs [56], and are used in computer vision applications for object-detection tasks for the purpose of counting the gradient orientation occurrences in localized image regions. HOG measures the first-order image gradient pooled in overlapping orientation bins, and gives a compressed and encoded version of an image. It counts the occurrences of different gradient orientations and maintains geometric invariance and photometric transformations. The process involves computing the gradients, creating cell histograms, and generating and normalizing the descriptor blocks. Given an image, HOG computes the gradient orientations and plots the histogram of these orientations, giving the probability of the existence of a given gradient with a specific orientation in a given path. The features are extracted over small blocks in a repetitive fashion to preserve information pertaining to the local structures and the block-wise features are finally concatenated into a feature vector. HOG descriptors are computed on a dense grid of uniformly spaced cells and use overlapping local contrast normalization to aid in improved accuracy. An increase in the cell size helps to capture spatial information on a large scale. A block comprises a number of cells, and a reduced block size helps to capture the significance of local pixels and suppress changes in illumination. In this proposal, the number of cells overlapping between adjacent blocks is chosen to be half the block size to ensure adequate normalization of contrast. The cell size is varied and the results are visualized across the datasets to observe the degree of variation in the amount of shape information encoded in the feature vector. We also visualized the effect of a reduced block size in the process of capturing the significance of local pixels and suppressing changes due to illumination variations. We empirically evaluated the values for the cell size parameter, number of bins, and block size that gave the best accuracy and set them to [32 32], 9, and [2 2] respectively.

BOW is a technique adapted from the world of information retrieval to computer vision applications. Contrary to text, images do not contain words. So, this method creates a bag of features extracted from the images across the classes, using a custom feature descriptor, and constructs a visual vocabulary. In this study, speeded-up robust features (SURF) are used as feature descriptors that detect interesting key points in a given image by using an integer approximation of the determinant of a blob detector based on the Hessian matrix [57]. The feature point locations across the CXR images are selected through a grid method and SURF are extracted from the selected locations. A grid step is chosen and the features are extracted from the CXR images across the normal and TB-positive categories. The number of features across the image categories is balanced to improve clustering. A visual vocabulary is created by reducing the dimension of the features through feature space quantization, using K-means clustering. Images are encoded into feature vectors and the encoded training samples across the image categories are fed into the SVM classifier to be classified into TB-positive and normal categories. An [8 8] grid step and 500 clusters are used in this study

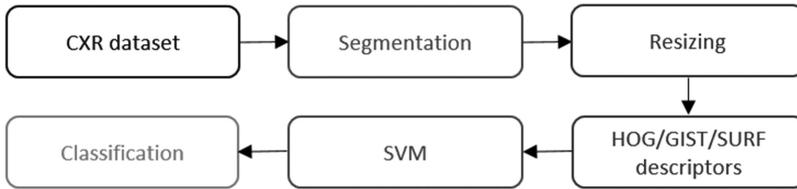


FIGURE 1.4
Steps involved in Proposal 1.

across the datasets. For this proposal, we performed a nested cross-validation. In the outer loop, we performed five-fold cross-validation for all the datasets. In the inner loop, we performed Bayesian optimization [58] to minimize classification error by varying the parameters for the SVM classifier which includes the box constraint, kernel scale, kernel function, and order of the polynomial. The chosen ranges include $[1e-3 \ 1e3]$, $[1e-3 \ 1e3]$, and $[2 \ 4]$ for box constraint, kernel scale, and order of the polynomial respectively. For the kernel function, the optimization process searched among linear, Gaussian, RBF, and polynomial kernels. Figure 1.4 shows the steps involved in this proposal.

1.2.3 Proposal 2—Feature Extraction and Classification Using a Customized CNN

In the second proposal (P2), we evaluated customized CNN models toward the task of TB detection. As stated earlier, we were interested in optimizing a customized model to learn task-specific features. Train/validation splits are randomized (70/30). Images are down-sampled to 224×224 pixel resolutions and training samples are augmented with horizontal and vertical translations in the range of $[-5 \ 5]$ pixels and rotations in the range of degrees $[-10 \ 10]$, toward preventing model overfitting. We made sure to augment only the training data to suit the deployment scenario where abrupt mirroring, flipping, and a huge degree of rotations are not viable. Figure 1.5 shows the steps involved in this proposal. We applied Bayesian optimization [58] to find the optimal network parameters and training options for the custom CNNs trained on different datasets. Bayesian optimization is applied to optimize

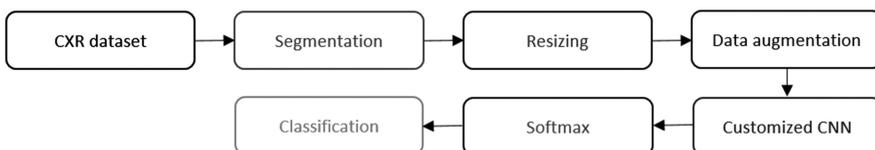


FIGURE 1.5
Steps involved in Proposal 2.

non-differentiable, discontinuous functions by maintaining a Gaussian process model of the objective function to be minimized and to perform objective function evaluations for training to find the optimal model parameters for the underlying data.

The framework for the customized CNN is specified. Each CNN block has a convolutional layer, followed by batch normalization [59], and Rectified Linear Units (ReLU) layer [34]. Padding is added to the convolutional layers to ensure that the spatial output dimensions match the original input. The number of filters is increased by a factor of two, and every time a max-pooling layer is used to ensure the amount of computation roughly remains the same across the convolutional layers. A filter size of 3×3 is used uniformly across the layers. The number of filters in a given layer is chosen to be $1/\sqrt{\text{network depth}}$ so that the customized CNNs with different network depths have roughly the same number of parameters and demand the same computational cost per iteration. The initial number of filters in each convolutional layer is chosen to be $\left(\text{round}\left(\frac{\text{image size}}{\sqrt{\text{network depth}}}\right)\right)$. The variables to be optimized are chosen and search ranges are specified. These ranges include [1 3], [1e-3 5e-2], [0.8 0.99], and [1e-10 1e-2] for the network depth, learning rate, stochastic gradient descent (SGD) momentum, and L2-regularization parameters respectively. An objective function for the Bayesian optimization process that takes, as its inputs, the values of the optimization variables, is used for training the customized CNN across different datasets, and a classification error is returned. Bayesian-optimized parameters for the least classification error are recorded.

1.2.4 Proposal 3—Feature Extraction Using Pre-Trained CNNs and Classification Using SVM

In the third proposal (P3), we evaluated the performance of pre-trained CNNs as feature extractors toward classifying TB-positive and healthy CXR images. Figure 1.6 shows the steps involved in this proposal. We evaluated state-of-the-art CNN models including AlexNet, VGG-16, GoogLeNet, and ResNet-50. The segmented ROI constituting the lungs are down-sampled to match the input dimensions of the pre-trained models. Each layer of the

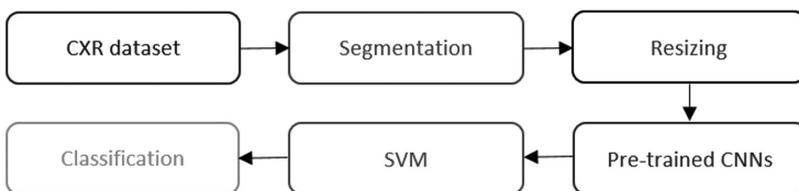


FIGURE 1.6

Steps involved in Proposal 3.

pre-trained CNNs produces an activation for the given image. Earlier layers capture primitive features that include blobs, edges, and colors which are abstracted by the deeper layers to form higher-level features to present a more affluent image representation. These features are extracted from the layer before the classification layer [33] and used to train an SVM classifier. Figure 1.6 shows the steps involved in this proposal. As in P1, we performed a nested cross-validation. In the inner loop, we performed Bayesian optimization to minimize the cross-validation error by varying the SVM parameters. The chosen ranges include $[1e^{-3} \ 1e^3]$, $[1e^{-3} \ 1e^3]$, and $[2 \ 4]$ for box constraint, kernel scale, and order of the polynomial respectively. For the kernel function, the optimization process searched among linear, Gaussian, RBF, and polynomial kernels.

1.2.5 Proposal 4—Constructing Stacked Model Ensembles

Literature reveals the usage of local/global feature descriptors, and customized and pre-trained CNNs in classifying medical images. However, there are no state-of-the-art methods available that evaluate the performance of a stacked generalization of these approaches toward TB detection. It has been shown that, in general, a classifier's performance could be improved using a stacked model ensemble that combines multiple, diverse base-learners which make independent errors via a meta-learner [51]. These base-learners use different learning algorithms to overfit different regions in the feature space; thus, the stacked ensemble is often heterogeneous. In the fourth and final proposal (P4), we created a stacked generalization of models from different proposals to find an optimal stacked ensemble model that improved the accuracy of TB detection. Stacked ensembles are created for the models in P1 (E [P1]), P1 and P2 (E [P1, P2]), P1 and P3 (E [P1, P3]), P2 and P3 (E [P2, P3]), and P1, P2, and P3 (E [P1, P2, P3]). Figure 1.7 shows the steps involved in this proposal.

Figure 1.8 shows the schematic of the stacked ensemble of models from different proposals. Stacked model ensembles are implemented with the best parameter values for the models from different proposals and evaluated through five-fold cross-validation. The model ensembles consist of two levels. At the first level (Level-0), models from different proposals

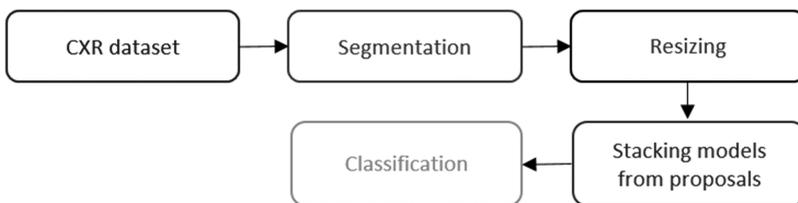


FIGURE 1.7
Steps involved in Proposal 4.

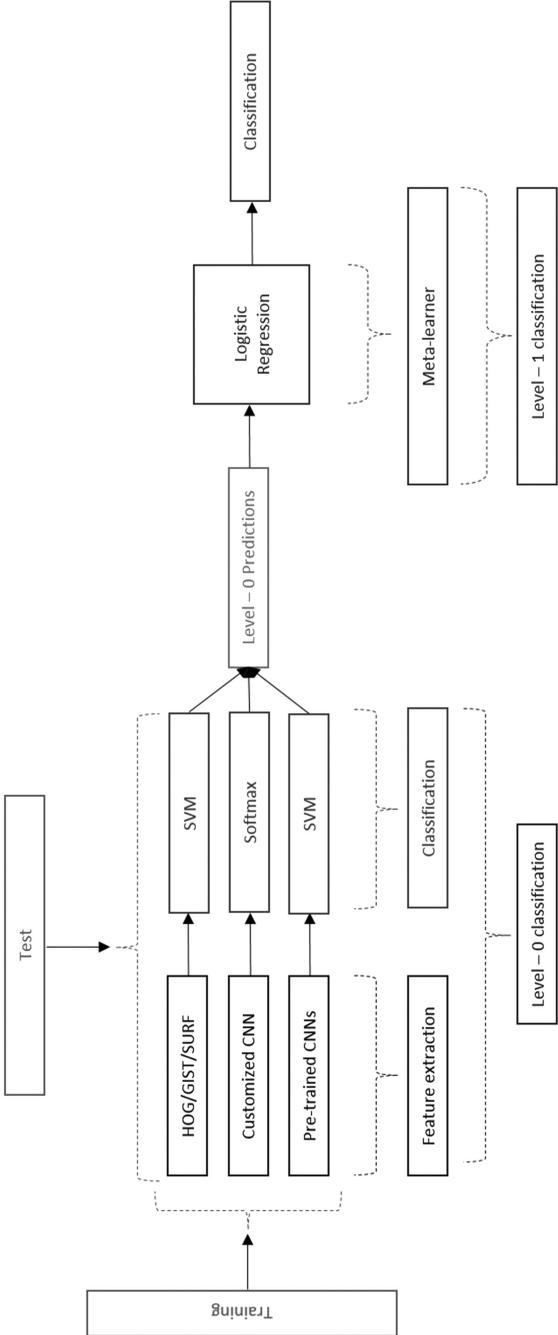


FIGURE 1.8 Stacked generalization of models from different proposals.

constitute diverse, individual base-learners. Let M denote the number of rows in the training data and N denote the number of base-learners. An $M \times N$ matrix, along with the original responses for the training samples, constitutes Level-0 predictions. The meta-learner is trained on the predictions of the base-learners. The meta-learner is a logistic regression classifier that estimates the probability for a binary response. The stacked ensemble of base-learners and meta-learner is used to predict on the test data. In this study, we used a Windows® system with Intel® Xeon® CPU E5-2640v3 2.60-GHz processor, 1 TB of Hard Disk space, 16 GB RAM, a CUDA-enabled Nvidia® GTX 1080 Ti 11GB graphical processing unit (GPU), Matlab® R2017b, Weka® 3 ML software, and CUDA 8.0/cuDNN 5.1 dependencies for GPU acceleration.

1.3 Results and Discussion

In P1, we used global descriptors including GIST and HOG, and local descriptors including SURF, toward identifying TB manifestations. The images are down-sampled to 3072×3072 , 4096×4096 , 2048×2048 , and 1024×1024 pixel resolutions for Shenzhen, Montgomery, Kenya, and India collections respectively. Table 1.1 shows the results obtained with different feature descriptors in terms of accuracy and AUC. For the Shenzhen dataset, the best results are obtained with the GIST features and SVM/RBF, with an accuracy of 0.845 and AUC of 0.921. For the Montgomery dataset, the BOW model using SURF and SVM/RBF demonstrated superior performance with an accuracy of 0.775 and AUC of 0.845. For the Kenya dataset, HOG features and SVM/Gaussian showed better performance in terms of accuracy, but the GIST features and RBF kernel-based SVM classifier gave the best AUC of 0.748. For the India dataset, GIST features and SVM/RBF demonstrated superior performance with an accuracy of 0.882 and AUC of 0.961. We could observe that no feature descriptor performed equally well on the underlying data. The chest radiographic images across the

TABLE 1.1

P1— GIST, HOG, and SURF-Based Feature Extraction and SVM-Based Classification

Datasets	Accuracy			AUC		
	HOG	GIST	SURF	HOG	GIST	SURF
Shenzhen	0.841	0.845	0.816	0.917	0.921	0.890
Montgomery	0.708	0.750	0.775	0.772	0.817	0.845
Kenya	0.683	0.667	0.672	0.741	0.748	0.747
India	0.880	0.882	0.864	0.947	0.961	0.938

datasets are collected with different machinery and at different pixel resolutions. The local/global feature descriptors are rule-based feature extraction mechanisms that are built and optimized to improve performance on individual datasets. For this reason, they don't perform equally well across the datasets. It can be noted here that the results obtained with the India dataset are superior to those obtained with the other datasets. A similar pattern is observed in the result tables for different proposals. A noteworthy factor with the India dataset is that though the dataset is limited, TB manifestations are obvious and are distributed throughout the lungs, which gives the feature descriptors the opportunity to capture highly discriminative features across normal and abnormal categories. The lowest performance is observed with the Kenya dataset, the principal reason being that the dataset has a highly imbalanced distribution of instances across the classes, with 238 abnormal CXRs in comparison to 729 healthy controls. The patients were all HIV+ with a low immune response. For this reason, the expression of the disease, even in severe cases, is significantly weaker than with ordinary TB. Also, the CXRs are obtained as a result of mobile truck-based screening, the images are cassette-based, and hence the image resolution is not commendable even after CLAHE enhancement, which further impaired the performance of feature extraction and classification. With the Montgomery dataset, performance limitation may be attributed to the limited size of the dataset and also to the degree of imbalance across the classes, where 40% of the samples are TB-positive as compared to 60% of the healthy controls.

In P2, we evaluated a customized CNN model for each dataset, optimized to learn task-specific features toward the task of TB detection. The images are down-sampled to 224×224 pixel resolution. The architecture of the optimized customized CNN models is shown in Figure 1.9. We applied Bayesian optimization to find the optimal network parameters and training options for the customized models. Table 1.2 shows the optimal values for the parameters learned by the customized CNN for the different datasets. Table 1.3 presents the performance measures of the customized CNNs across the datasets. The performance of the customized CNN, with respect to the India dataset, is superior to that with the other datasets, obtaining an accuracy of 0.860 and AUC of 0.937. The custom model showed the least performance with the Kenya dataset for the reasons already discussed. Also, the customized CNN is not able to completely reap the benefits of task-specific learning and classification due to data scarcity. The training samples are augmented only to resolve overfitting concerns, but this did not improve the validation accuracy.

We visualized the task-specific features and salient network activations in the optimized CNN model towards understanding the learning dynamics. For instance, we took the optimized CNN model for the India dataset and visualized the features in the second, fourth, sixth, and ninth convolutional layers, as shown in Figure 1.10 (a)–(g). The convolutional layers outputs

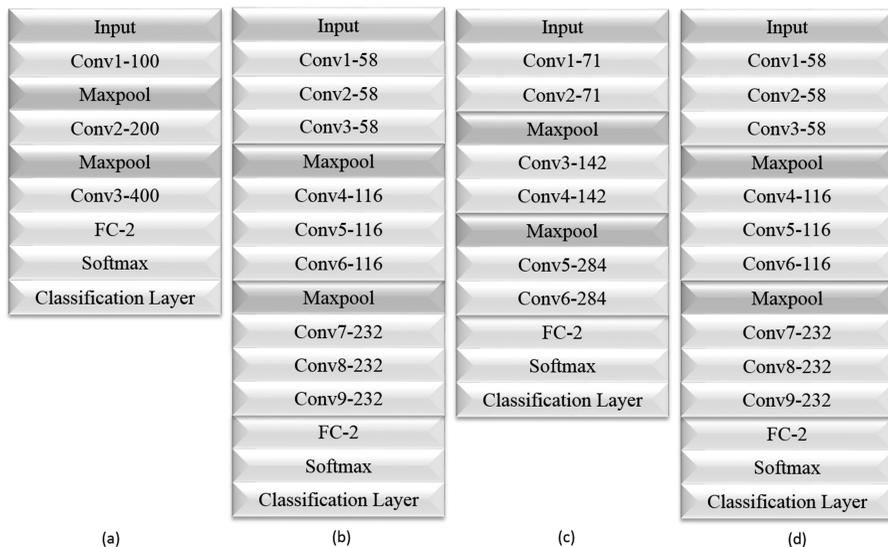


FIGURE 1.9
Optimized CNNs for different datasets: (a) Shenzhen, (b) Montgomery, (c) Kenya, (d) India.

TABLE 1.2

Optimal Parameters for the Customized CNN Model

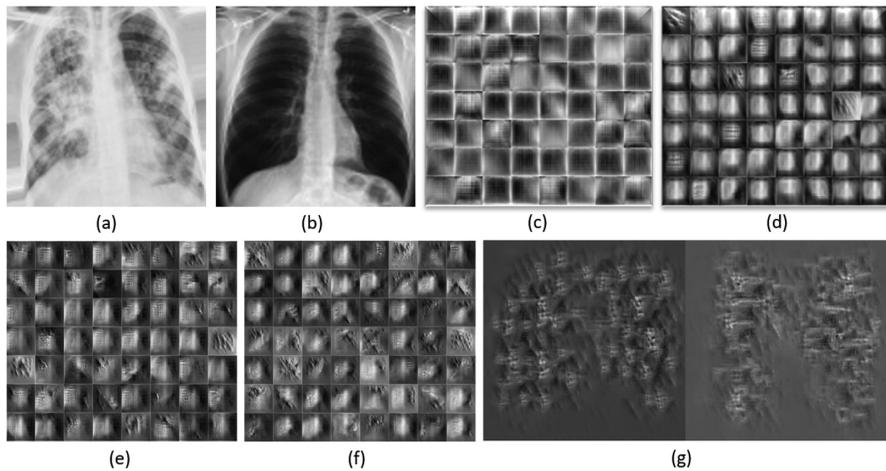
Datasets	Parameters			Performance		
	Momentum	Learning Rate	L2-Decay	Network Depth	Accuracy	AUC
Shenzhen	0.841	3.615e-04	1.608e-10	3	0.820	0.894
Montgomery	0.841	9.353e-04	1.448e-08	9	0.750	0.817
Kenya	0.845	3.949e-04	2.058e-09	6	0.698	0.761
India	0.942	1.994e-04	3.600e-03	9	0.860	0.937

TABLE 1.3

P2— Customized DL Model-Based Feature Extraction and Classification

Datasets	Accuracy	AUC
Shenzhen	0.820	0.894
Montgomery	0.750	0.817
Kenya	0.698	0.761
India	0.860	0.937

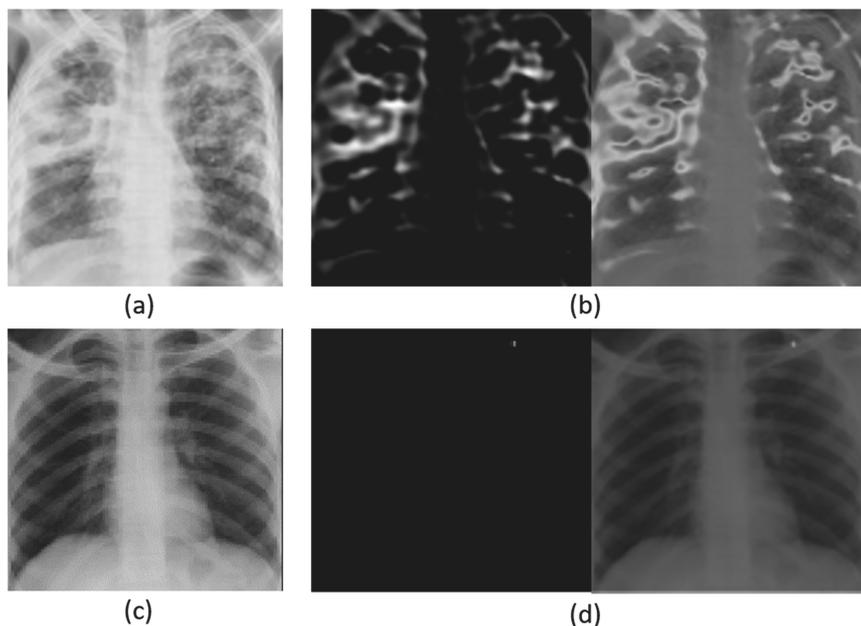
multiple channels, each corresponding to a filter applied to the input layer. The fully connected layer abstracts the features from the earlier layers and outputs channels corresponding to the image categories. The second convolutional layer appears to learn mostly colors and edges, indicating that the channels

**FIGURE 1.10**

Visualizing features. From left to right: (a) abnormal CXR, (b) normal CXR, (c) second convolution layer, (d) fourth convolution layer, (e) sixth convolution layer, (f) ninth convolution layer, (g) fully connected layer.

are color filters and edge detectors. As we progressed to the fourth convolutional layer, we observed that the customized CNN began to learn the edges and orientations. As we progressed to the sixth and ninth convolutional layers, we visualized task-specific features leading to the formation of shapes by abstracting the primitive features from the earlier layers. The fully connected layer towards the end of the model loosely resembled the abnormal and normal classes respectively. The model understands a decomposition of its visual input space as a hierarchical-modular network of filters and a probabilistic mapping between combinations of filters and a set of labels. This is not analogous to “seeing,” as programmed in the visual cortices of humans. The exact nature of the convolutional filters, the hierarchy, and the process through which they learn have very little in common with the human visual cortices, which are not convolutional and are structured into cortical columns whose purpose is yet to be fully explored. The visual perception of humans is sequential, active, and involves motor control; thus it is very different from that of the convolutional filters.

We also visualized the salient network activations to discover the learned features by comparing the areas of activation against the original image. An abnormal CXR is fed into the optimized CNN model and the activations of different layers are analyzed. The performance of the model is evaluated by investigating the activation of the channels on a set of input images. The resulting activations are compared with that of the original image, as shown in Figure 1.11 (a)–(d). CNN learns to detect complicated features in deeper convolutional layers that build up their features by combining features from the earlier layers. Thus, the channels of the deepest convolutional layer in the model are analyzed to

**FIGURE 1.11**

Visualizing the highest channel activations in the deepest convolutional layer: (a) CXR showing bilateral pulmonary TB, (b) activations/heat maps, (c) normal lung, (d) activations/heat maps.

observe the areas activating on the image and compared to the corresponding areas in the original image. The channel showing the highest activation for the abnormal locations in the input images is investigated. All activations are scaled to the range $[0, 1]$. When the activations are strongly positive, they are observed as white pixels, and when strongly negative, by black pixels. When the activations are not strong enough, they are observed as gray pixels. The position of a pixel in the channel activation corresponds to that in the original image. The channels show both positive and negative activations. However, only positive activations are investigated because of the ReLU non-linearity following the convolutional layers. The activations of the ReLU layer show the location of abnormalities. CXRs showing bilateral pulmonary TB and normal lung are input to the model, and the saliency maps and the grayscale image showing the network channel activations are extracted. A pseudo-color image is generated to get a clearer and more appealing representation from the perceptual aspect by using the “jet” colormap, so that the activations higher than a given threshold appear bright red, with discrete color transitions in between. The threshold is selected to match the range of activations and achieve the best visualization effect. The resulting heat maps are overlaid onto the original image, and the black pixels in the heat maps are made fully transparent. It is observed from the heat maps that the customized model precisely activated the location of abnormalities and showed no activation for the normal lung

image. This implies that the customized model learns task-specific features and localizes the abnormalities precisely to help distinguish between normal and abnormal classes.

Table 1.4 presents the results of P3 using pre-trained CNNs for feature extraction and SVM/RBF-based classification. Images are down-sampled to 224×224 and 227×227 pixel resolutions to suit the input requirements for the different pre-trained CNNs, across all the datasets. For the Shenzhen dataset, AlexNet obtained the best accuracy of 0.859 and AUC of 0.924. The same pattern is observed across Montgomery, Kenya, and India datasets. For the Montgomery dataset, AlexNet obtained the best accuracy of 0.725 and AUC of 0.817. For the India dataset, AlexNet outperformed the other pre-trained CNNs with an accuracy of 0.872 and AUC of 0.950. Only for the Kenya dataset did we observe that the AUC of VGG-16 is slightly better than that of AlexNet; however, the accuracy of AlexNet was higher than that of the other pre-trained CNNs. It can be noted that the results obtained with the India dataset are superior to the results obtained with the other datasets for the reasons discussed earlier. Among the pre-trained CNNs evaluated in this study, AlexNet outperformed the other models across the datasets. The deeper layers of ResNet-50 and GoogLeNet are progressively more complex, specific to the ImageNet dataset, and not suitable for the underlying task of binary medical image classification. For large-scale datasets such as ImageNet, deeper networks outperform shallow counterparts for the reason that the data is diverse and the networks learn abstractions for a huge selection of classes. In our case, for the binary task of TB detection, the variability in data is several orders of magnitude smaller and deeper networks do not seem to be a fitting tool. Also, literature studies reveal that convolutional features from shallow networks lead to higher accuracy than do the features of the deeper networks. Shallow models such as AlexNet provided high accuracy in the detection task [60]. Also, the top layers of pre-trained CNNs such as GoogLeNet and ResNet-50 are probably too specialized, progressively more complex, and not the best candidate to re-use for the task of our interest. This explains the difference in performance in our case.

Table 1.5 shows the results of the final proposal using ensembles of models from different proposals. The results obtained are promising in comparison

TABLE 1.4

P3—Pre-Trained CNNs-Based Feature Extraction and SVM-Based Classification

Datasets	Accuracy				AUC			
	AlexNet	VGG-16	GoogLeNet	ResNet-50	AlexNet	VGG-16	GoogLeNet	ResNet-50
Shenzhen	0.859	0.829	0.768	0.819	0.924	0.901	0.870	0.893
Montgomery	0.725	0.717	0.678	0.676	0.817	0.757	0.648	0.616
Kenya	0.693	0.691	0.674	0.678	0.776	0.777	0.750	0.753
India	0.872	0.812	0.796	0.812	0.950	0.892	0.888	0.902

Note: Bold numerical values indicate the performance measures of the best-performing model/s.

to that obtained from the other proposals. The stacked ensemble of local/global features (E[P1]) had similar accuracy values of 0.875 and 0.960 across the selection of ensembles for the Montgomery and India datasets. However, the ensemble of all the proposals (E[P1P2P3]) had the highest AUC of 0.986 and 0.995 in comparison to other stacked ensembles for the Montgomery and India datasets. The results are superior to those obtained from the other individual proposals. The same pattern is observed across all the datasets. One of the most significant requirements for the creation of stacked ensembles is that the base-learners be accurate and make diverse errors; i.e. the errors must have a low correlation [61]. Since we have a collection of models from different proposals, we benefit from the fact that their outputs are diverse and accurate, with less correlation in their errors, thus enhancing the performance of the ensembles. This is exactly what happened, as observed from the results.

Tables 1.6 and 1.7 compare the results obtained across the ensembles of different proposals presented in this study and literature on TB detection. The stacked ensemble results are, in almost all the cases, superior to those from different proposals. For this reason, the stacked ensemble of models from all the proposals (E[P1P2P3]) outperformed the other ensembles under study. In terms of accuracy, as shown in Table 1.6, the stacked ensemble of models from all the proposals (E[P1P2P3]) outperformed the state-of-the-art. The proposed ensemble demonstrated the highest accuracy of 0.960 for India, followed by 0.959 for the Shenzhen, 0.875 for the Montgomery, and 0.810 for

TABLE 1.5

P4—Ensemble of Models from Different Proposals

Datasets	E[P1]		E[P1P2]		E[P1P3]		E[P2P3]		E[P1P2P3]	
	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC
Shenzhen	0.934	0.955	0.944	0.980	0.934	0.991	0.944	0.78	0.959	0.994
Montgomery	0.875	0.875	0.875	0.927	0.875	0.962	0.708	0.927	0.875	0.986
Kenya	0.733	0.825	0.784	0.826	0.776	0.826	0.767	0.765	0.810	0.829
India	0.960	0.960	0.940	0.958	0.960	0.965	0.940	0.974	0.960	0.995

Note: Bold numerical values indicate the performance measures of the best-performing model/s.

TABLE 1.6

Comparison with the Literature—Accuracy

Datasets	Literature				Proposed Approaches				
	[5]	[43]	[45]	[21]	E[P1]	E[P1P2]	E[P1P3]	E[P2P3]	E[P1P2P3]
Shenzhen	0.840	0.837	0.847	–	0.934	0.944	0.934	0.944	0.959
Montgomery	0.783	0.674	0.826	–	0.875	0.875	0.875	0.708	0.875
Kenya	–	–	–	–	0.733	0.784	0.776	0.767	0.810
India	–	–	–	0.943	0.960	0.94	0.960	0.940	0.960

Note: Bold numerical values indicate the performance measures of the best-performing ensemble/s.

TABLE 1.7

Comparison with the Literature—AUC

Datasets	Literature				Proposed Approaches				
	[5]	[43]	[45]	[21]	E[P1]	E[P1P2]	E[P1P3]	E[P2P3]	E[P1P2P3]
Shenzhen	0.900	0.926	0.926	–	0.955	0.98	0.991	0.780	0.994
Montgomery	0.869	0.884	0.926	–	0.875	0.927	0.962	0.927	0.986
Kenya	–	–	–	–	0.825	0.826	0.826	0.765	0.829
India	–	–	–	0.960	0.960	0.958	0.965	0.974	0.995

Note: Bold numerical values indicate the performance measures of the best-performing ensemble/s.

the Kenya datasets. We could observe similar patterns with the AUC values as shown in Table 1.7, where (E[P1P2P3]) clearly demonstrated a high AUC value in comparison to the results discussed in the literature. The results for the India dataset are superior, with an AUC of 0.995, followed by 0.994 for the Shenzhen, 0.986 for the Montgomery, and 0.829 for the Kenya datasets. As observed from these result tables, (E [P1P2P3]) achieved superior results across all the datasets.

1.4 Conclusion and Future Work

We have discussed four different proposals for improving the performance of TB detection. In P1, we used local and global feature descriptors to extract discriminative features from the radiographic images. The extracted features are used to train an SVM classifier. In P2, we optimized the architecture and hyper-parameters of customized CNNs using Bayesian optimization toward learning task-specific features for the underlying data. The customized model is highly compact, and has fewer trainable parameters and less architectural flexibility. The learned features and salient network activations are visualized to understand the learning dynamics. In P3, we used four different pre-trained CNNs to extract features from the datasets, and trained an SVM classifier on the extracted features. Under circumstances when the data availability is sparse, it is not recommended to fine-tune the pre-trained CNNs due to overfitting concerns. Literature studies [30] reveal that pre-trained CNNs could be used as a promising feature-extraction tool, especially for biomedical imagery. In P4, we performed a stacked ensemble of different proposals to find the optimal ensemble model that improved the accuracy of TB detection. Model stacking optimizes the combination of several diverse and accurate base-learners and reduces generalization errors. From our current studies, we believe that the stacked ensemble of diverse and accurate models using local/global feature descriptors and customized and pre-trained CNN models could

be a promising option for improving the detection accuracy, particularly when the data are sparse. An appealing use case is to apply this method in applications with sparse data, particularly in biomedical imagery where the usage of only CNNs would lead to overfitting. The proposed ensemble could serve as triage, to minimize patient loss and reduce delays in resource-constrained settings; it could also be adapted to improving the accuracy of screening for other health-related applications. With regard to advancements in TB detection, recent works [62] demonstrate that the future demands large-scale biomedical datasets. The performance of the stacked ensemble could be highly promising with such a large collection of data.

Acknowledgments

This work was supported by the Intramural Research Program of the Lister Hill National Center for Biomedical Communications (LHNCBC), the National Library of Medicine (NLM), and the U.S. National Institutes of Health (NIH).

Conflict of Interest

The authors have no conflict of interest to report.

References

1. "WHO Global Tuberculosis Report 2018," 2018, Available at: https://www.who.int/tb/publications/global_report/en/, Accessed on 05/15/2018.
2. F. A. Mettler, W. Huda, T. T. Yoshizumi, and M. Mahesh, "Effective doses in radiology and diagnostic nuclear medicine: a catalog," *Radiology*, vol. 248, no. 1, pp. 254–263, 2008.
3. S. Rajaraman, S. Candemir, I. Kim, G. Thoma, S. Antani, "Visualization and interpretation of convolutional neural network predictions in detecting pneumonia in pediatric chest radiographs," *Applied Sciences*, vol. 8, no. 10, p. 1715, 2018.
4. M. F. Iademarco, J. O'Grady, and K. Lönnroth, "Chest radiography for tuberculosis screening is back on the agenda," *International Journal of Tuberculosis and Lung Disease*, vol. 16, no. 11, pp. 1421–1422, 2012.

5. S. Jaeger, A. Karargyris, S. Candemir, L. Folio, J. Siegelman, F. Callaghan, Z. Xue, K. Palaniappan, R. K. Singh, S. Antani, G. Thoma, Y. X. Wang, P. Lu, and C. J. McDonald, "Automatic tuberculosis screening using chest radiographs," *IEEE Transactions on Medical Imaging*, vol. 33, no. 2, pp. 233–245, 2014.
6. Y. Balabanova, R. Coker, I. Fedorin, S. Zakharova, S. Plavinskij, N. Krukov, R. Atun, and F. Drobniowski, "Variability in interpretation of chest radiographs among Russian clinicians and implications for screening programmes: observational study," *BMJ*, vol. 331, no. 7513, pp. 379–382, 2005.
7. S. Jaeger, A. Karargyris, S. Candemir, J. Siegelman, L. Folio, S. Antani, and G. Thoma, "Automatic screening for tuberculosis in chest radiographs: a survey," *Quantitative Imaging in Medicine and Surgery*, vol. 3, no. 2, pp. 89–99, 2013.
8. P. Maduskar, M. Muyoyeta, H. Ayles, L. Hogeweg, L. Peters-Bax, and B. van Ginneken, "Detection of tuberculosis using digital chest radiography: automated reading vs. interpretation by clinical officers," *International Journal of Tuberculosis and Lung Disease*, vol. 17, no. 12, pp. 1613–1620, 2013.
9. S. Rajaraman, K. Silamut, M. Hossain, I. Ersoy, R. Maude, S. Jaeger, G. Thoma, and S. Antani, "Understanding the learned behavior of customized convolutional neural networks toward malaria parasite detection in thin blood smear images," *Journal of Medical Imaging*, vol. 5, no. 3, pp. 34501–34511, 2018.
10. S. Rajaraman, S. Antani, M. Poostchi, K. Silamut, M. Hossain, R. Maude, S. Jaeger, and G. Thoma, "Pre-trained convolutional neural networks as feature extractors toward improved malaria parasite detection in thin blood smear images," *PeerJ*, vol. 6, p. e4568, 2018.
11. R. Sivaramakrishnan, S. Antani, and S. Jaeger, "Visualizing deep learning activations for improved malaria cell classification," in *First Workshop Medical Informatics and Healthcare (MIH 2017)*, held in Halifax, Nova Scotia, Canada, pp. 40–47, 2017.
12. A. Karargyris, J. Siegelman, D. Tzortzis, S. Jaeger, S. Candemir, Z. Xue, K. C. Santosh, S. Vajda, S. Antani, L. Folio, and G. Thoma, "Combination of texture and shape features to detect pulmonary abnormalities in digital chest X-rays," *International Journal of Computer Assisted Radiology and Surgery*, 2016.
13. K. C. Santosh, S. Vajda, S. Antani, and G. Thoma, "Edge map analysis in chest X-rays for automatic pulmonary abnormality screening," *International Journal of Computer Assisted Radiology and Surgery*, vol. 11, no. 9, pp. 1637–1646, 2016.
14. K. C. Santosh and S. Antani, "Automated chest x-ray screening: can lung region symmetry help detect pulmonary abnormalities?" *IEEE Transactions on Medical Imaging*, 2018.
15. S. Vajda, A. Karargyris, S. Jaeger, K. C. Santosh, S. Candemir, Z. Xue, S. Antani, and G. Thoma, "Feature selection for automatic tuberculosis screening in frontal chest radiographs," *Journal of Medical Systems*, 2018.
16. F. Ahmad Khan, T. Pande, B. Tessema, R. Song, A. Benedetti, M. Pai, K. Lönnroth, and C. M. Denkinger, "Computer-aided reading of tuberculosis chest radiography: moving the research agenda forward to inform policy," *European Respiratory Journal*, 2017.
17. R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-3, no. 6, pp. 610–621, 1973.

18. B. Van Ginneken, S. Katsuragawa, B. M. Ter Haar Romeny, K. Doi, and M. A. Viergever, "Automatic detection of abnormalities in chest radiographs using local texture analysis," *IEEE Transactions on Medical Imaging*, vol. 21, no. 2, pp. 139–149, 2002.
19. L. Hogeweg, C. Mol, P. A. De Jong, R. Dawson, H. Ayles, and B. Van Ginneken, "Fusion of local and global detection systems to detect tuberculosis in chest radiographs," *Lecture Notes in Computer Science* (including subseries *Lecture Notes in Artificial Intelligence* and *Lecture Notes in Bioinformatics*), vol. 6363 LNCS, no. PART 3, pp. 650–657, 2010.
20. S. Jaeger, S. Candemir, S. Antani, Y. J. Wang, P.X. Lu, and G. Thoma, "Two public chest X-ray datasets for computer-aided screening of pulmonary diseases," *Quantitative Imaging in Medicine and Surgery*, vol. 4, no. 6, pp. 475–477, 2014.
21. A. Chauhan, D. Chauhan, and C. Rout, "Role of gist and PHOG features in computer-aided diagnosis of tuberculosis without segmentation," *PLoS One*, vol. 9, no. 11, pp. 1–12, 2014.
22. J. Melendez, C. I. Sánchez, R. H. H. M. Philipsen, P. Maduskar, R. Dawson, G. Theron, K. Dheda, and B. V. Ginneken, "An automated tuberculosis screening strategy combining X-ray-based computer-aided detection and clinical information," *Scientific Reports*, vol. 6, p. 25265, 2016.
23. T. Pande, C. Cohen, M. Pai, and F. Ahmad Khan, "Computer-aided detection of pulmonary tuberculosis on digital chest radiographs: a systematic review," *International Journal of Tuberculosis and Lung Disease*, vol. 20, no. 9, 2016.
24. S. Candemir, S. Jaeger, K. Palaniappan, J. P. Musco, R. K. Singh, Z. Xue, A. Karargyris, S. Antani, G. Thoma, and C. J. McDonald, "Lung segmentation in chest radiographs using anatomical atlases with nonrigid registration," *IEEE Transactions on Medical Imaging*, vol. 33, no. 2, pp. 577–590, 2014.
25. J. H. Tan, U. R. Acharya, C. Tan, K. T. Abraham, and C. M. Lim, "Computer-assisted diagnosis of tuberculosis: a first order statistical approach to chest radiograph," *Journal of Medical Systems*, vol. 36, no. 5, pp. 2751–2759, 2012.
26. U. Avni, H. Greenspan, E. Konen, M. Sharon, and J. Goldberger, "X-ray categorization and retrieval on the organ and pathology level, using patch-based visual words," *IEEE Transactions on Medical Imaging*, vol. 30, no. 3, pp. 733–746, 2011.
27. J. Schmidhuber, "Deep learning in neural networks: an overview," *Neural Networks*, vol. 61, pp. 85–117, 2015.
28. Y. LeCun, B. Yoshua, and H. Geoffrey, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
29. N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, pp. 1929–1958, 2014.
30. A. Ben Abacha, S. Gayen, J. J. Lau, S. Rajaraman, and D. Demner-Fushman, "NLM at ImageCLEF 2018 Visual Question Answering in the medical domain," held in Avignon, France, in *CEUR Workshop Proceedings*, 2018.
31. S. Rajaraman, S. Antani, Z. Xue, S. Candemir, and S. Jaeger, "Visualizing abnormalities in chest radiographs through salient network activations in deep learning," in *Life Sciences Conference (LSC), 2017 IEEE*, held in Sydney, NSW, Australia, pp. 71–74.
32. J. Deng, Wei Dong, R. Socher, Li-Jia Li, Kai Li, and Li Fei-Fei, "ImageNet: a large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, held in Miami, Florida, pp. 248–255.

33. A. S. Razavian, H. Azizpour, J. Sullivan, and S. Carlsson, "CNN features off-the-shelf: an astounding baseline for recognition," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, held in Columbus, Ohio, pp. 512–519.
34. A. Krizhevsky, I. Sutskever and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.
35. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *International Conference on Learning Representations*, held in San Diego, California, 2015.
36. C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, held in Las Vegas, Nevada, pp. 2818–2826, 2016.
37. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *arXiv Prepr. arXiv1512.03385v1*, vol. 7, no. 3, pp. 171–180, 2015.
38. F. Bousetouane and B. Morris, "Off-the-shelf CNN features for fine-grained classification of vessels in a maritime environment," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9475, pp. 379–388, 2015.
39. Y. Bar, I. Diamant, L. Wolf, S. Lieberman, E. Konen, and H. Greenspan, "Chest pathology detection using deep learning with non-medical training," *Proceedings – International Symposium on Biomedical Imaging*, held in Brooklyn Bridge, New York, vol. 2015–July, pp. 294–297, 2015.
40. T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, 2002.
41. A. Oliva and A. Torralba, "Modeling the shape of the scene: a holistic representation of the spatial envelope," *International Journal of Computer Vision*, vol. 42, no. 3, pp. 145–175, 2001.
42. A. Bergamo and L. Torresani, "PiCoDes: learning a compact code for novel-category recognition," *Advances in Neural Information Processing Systems*, vol. 24, pp. 2088–2096, 2011.
43. S. Hwang, H.E. Kim, J. Jeong, and H.J. Kim, "A novel approach for tuberculosis screening based on deep convolutional neural networks," in *SPIE Medical Imaging*, held in San Diego, California, 2016, p. 97852W.
44. R. Singh, M. K. Kalra, C. Nitiwarangkul, J. Patti, F. Homayounieh, A. Padole, P. Rao, P. Putha, V. V. Muse, A. Sharma, and S. R. Digumarthy, "Deep learning in chest radiography: detection of findings and presence of change," *PLoS One*, vol. 13, no. 10, p. e0204155, 2018.
45. U. K. Lopes and J. F. Valiati, "Pre-trained convolutional neural networks as feature extractors for tuberculosis detection," *Computers in Biology and Medicine*, vol. 89, pp. 135–143, 2017.
46. L. K. Hansen and P. Salamon, "Neural network ensembles," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 10, pp. 993–1001, 1990.
47. L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, no. 2, pp. 123–140, 1996.
48. Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of Computer and System Sciences*, vol. 55, no. 1, pp. 119–139, 1997.

49. L. Lam and S. Y. Suen, "Application of majority voting to pattern recognition: an analysis of its behavior and performance," *IEEE Transactions on Systems, Man, and Cybernetics – Part A*, vol. 27, no. 5, pp. 553–568, 1997.
50. G. Fumera and F. Roli, "A theoretical and experimental analysis of linear combiners for multiple classifier systems," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 6, pp. 942–956, 2005.
51. D. H. Wolpert, "Stacked generalization," *Neural Networks*, vol. 5, no. 2, pp. 241–259, 1992.
52. P. Lakhani and B. Sundaram, "Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks," *Radiology*, vol. 000, no. 0, p. 162326, 2017.
53. C. Liu, J. Yuen, A. Torralba, J. Sivic, and W. T. Freeman, "SIFT flow: dense correspondence across different scenes," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 5304 LNCS, no. PART 3, pp. 28–42, 2008.
54. Y. Boykov and G. Funka-Lea, "Graph cuts and efficient N-D image segmentation," *International Journal of Computer Vision*, vol. 70, no. 2, pp. 109–131, 2006.
55. S. E. Grigorescu, N. Petkov, and P. Kruizinga, "Comparison of texture features based on Gabor filters.," *IEEE Trans. Image Process.*, vol. 11, no. 10, pp. 1160–1167, 2002.
56. N. Dalal and W. Triggs, "Histograms of oriented gradients for human detection," 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition CVPR05, held in San Diego, California, vol. 1, no. 3, pp. 886–893, 2004.
57. H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, 2008.
58. J. Mockus, "On Bayesian methods for seeking the extremum," *IFIP Tech. Conf. Optimization Techniques*, pp. 400–404, 1974.
59. S. Ioffe and C. Szegedy, "Batch normalization: accelerating deep network training by reducing internal covariate shift," in *Proceedings of the 32nd International Conference on Machine Learning*, held in Lille, France, pp. 448–456, 2015.
60. K. Ashraf, B. Wu, F. N. Iandola, M. W. Moskewicz, and K. Keutzer, "Shallow networks for high-accuracy road object-detection," in *Proceedings of the 3rd International Conference on Vehicle Technology and Intelligent Transport Systems, VEHITS 2017, Porto, Portugal, April 22-24*, pp. 33–40, 2017.
61. T. G. Dietterich, "Ensemble methods in machine learning," *Multiple Classifier Systems*, vol. 1857, pp. 1–15, 2000.
62. A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.