Tuberculosis Diagnosis from Pulmonary Chest X-Ray using Deep Learning

Author
Mustapha Olayemi OLOKO-OBA

Supervisor
Prof. Serestina VIRIRI

A thesis submitted to the University of KwaZulu-Natal, College of Agriculture, Engineering and Science, in fulfillment of the requirements for the degree of Doctor of Philosophy in Computer Science

School of Mathematics, Statistics and Computer Science,
University of KwaZulu-Natal, Durban, South Africa.

Declaration of Authorship

I, Mustapha Olayemi OLOKO-OBA, declare that this thesis titled, "Tuberculosis Diagnosis from Pulmonary Chest X-Ray using Deep Learning" and the work presented in it are my own. I declare that:

1. The research reported in this thesis, except where otherwise indicated or acknowledged, is my original work;
2. This thesis has not been submitted in full or in part for any degree or examination to any other university;
3. This thesis does not contain other person's data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons;
4. This thesis does not contain other person's writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:
   (a) their words have been re-written but the general information attributed to them has been referenced;
   (b) where their exact words have been used, their writing has been placed inside quotation marks, and referenced;
5. This thesis does not contain text, graphics or tables copied and pasted from the Internet, unless specifically acknowledged, and the source being detailed in the thesis and in the references sections.

Candidate:
Mustapha Olayemi OLOKO-OBA

Signed: MO.Oloko-Oba

Date: 30/05/2022

As the candidate's supervisor I approve the submission of this thesis for examination.

Supervisor:
Prof. Serestina VIRIRI

Signed: [redacted]

Date: 05/06/2022

List of Publications

I, Mustapha Olayemi OLOKO-OBA, declare that the following are publications from this thesis:

1. **Mustapha Oloko-Oba and Serestina Viriri**. "Diagnosing Tuberculosis Using Deep Convolutional Neural Network", *Image and Signal Processing*. LNCS Springer, vol 12119, pp 151-161 **(2020)**. DOI: `https://doi.org/10.1007/978-3-030-51935-3_16`.

2. **Mustapha Oloko-Oba and Serestina Viriri**. "Tuberculosis Abnormality Detection in Chest X-Rays: A Deep Learning Approach". *Computer Vision and Graphics*. LNCS Springer, vol 12334, pp 121-132 **(2020)**. DOI: `https://doi.org/10.1007/978-3-030-59006-2_11`.

3. **Mustapha Oloko-Oba and Serestina Viriri**. "Pre-trained Convolutional Neural Network for the Diagnosis of Tuberculosis". *Advances in Visual Computing*. LNCS Springer, vol 12510, pp 558-569 **(2020)**. DOI: `https://doi.org/10.1007/978-3-030-64559-5_44`.

4. **Mustapha Oloko-Oba and Serestina Viriri**. "Ensemble of Convolution Neural Networks for Automatic Tuberculosis Classification". *Computational Collective Intelligence*. LNCS Springer, vol 12876, pp 549-559 **(2021)**. DOI: `https://doi.org/10.1007/978-3-030-88081-1_41`.

5. **Mustapha Oloko-Oba and Serestina Viriri**. "Ensemble of EfficientNets for the Diagnosis of Tuberculosis". *Computational Intelligence and Neuroscience*.vol. 2021, **(2021)**. DOI: `https://doi.org/10.1155/2021/9790894`.

6. **Mustapha Oloko-Oba and Serestina Viriri**. "A systematic review of Deep Learning Techniques for Tuberculosis Detection from Chest Radiograph". *Frontiers in Medicine Journal*, **(2022)**. DOI: `https://doi.org/10.3389/fmed.2022.830515`.

7. **Mustapha Oloko-Oba and Serestina Viriri**. "Tuberculosis Detection: A Comprehensive Review of State-of-the-Art". *International Journal of Computer Science*, **(2022)**
   (Under peer review)

Abstract

Tuberculosis (TB) remains a life-threatening disease, and it is one of the leading causes of mortality in developing countries. This is due to poverty and inadequate medical resources. While treatment for TB is possible, it requires an accurate diagnosis first. Several screening tools are available, and the most reliable is Chest X-Ray (CXR), but the radiological expertise for accurately interpreting the CXR images is often lacking. Over the years, CXR has been manually examined; this process results in delayed diagnosis, is time-consuming, expensive, and is prone to misdiagnosis, which could further spread the disease among individuals. Consequently, an algorithm could increase diagnosis efficiency, improve performance, reduce the cost of manual screening and ultimately result in early/timely diagnosis. Several algorithms have been implemented to diagnose TB automatically. However, these algorithms are characterized by low accuracy and sensitivity leading to misdiagnosis. In recent years, Convolutional Neural Networks (CNN), a class of Deep Learning, has demonstrated tremendous success in object detection and image classification task. Hence, this thesis proposed an efficient Computer-Aided Diagnosis (CAD) system with high accuracy and sensitivity for TB detection and classification. The proposed model is based firstly on novel end-to-end CNN architecture, then a pre-trained Deep CNN model that is fine-tuned and employed as a features extractor from CXR. Finally, Ensemble Learning was explored to develop an Ensemble model for TB classification. The Ensemble model achieved a new state-of-the-art diagnosis accuracy of 97.44% with a 99.18% sensitivity, 96.21% specificity and 0.96% AUC. These results are comparable with state-of-the-art techniques and outperform existing TB classification models.

Keywords: Tuberculosis, Chest X-Ray, Deep Learning, Convolutional Neural Network, Pre-processing, Ensemble, Computer-Aided Diagnostic.

Acknowledgements

The work presented in this thesis has been a meaningful experience, and it would not have been possible without ALLAH, the encouragement, support and counsel that I have received from many people; hence I would like to use this opportunity to extend my profound appreciation to everyone who has made this PhD thesis a reality.

My sincere gratitude goes to my supervisor Prof Serestina Viriri for suggesting the topic and introducing me to the thrilling field of Deep Learning for Computer Vision. His professional advice, mentorship, and care throughout this research cannot be overemphasized. I cannot forget his strict scrutiny that has helped improved my academic writing skills, my critical thinking for research and financial support towards the articles published in this thesis.

My kind appreciation to COPINE-NASRDA for granting the request to pursue this research, the University of KwaZulu-Natal for offering me a PhD bursary from the Big Data in Science and Society and the School of Mathematics, Statistics, and Computer Science staff for their warm reception and support.

My gratitude also goes out to my colleagues and friends Kehinde Aruleba, Olalekan Adisa, Obaido George, Ogbuokiri Blessing, Micheal Ayawei, Jude Ewemade, Abdulsalam Shakirudeen, Anil Pise, Mazi Ojukwu Nkwere, Dr Mutiu, Moshood Onifade, Badru Rahmon, Abdulgafari Lukumon, Kehinde Oladele, Taiwo Aruleba, Kehinde Oladele, Sarafadeen Olatunji, Idowu Aruleba, Saheed A Ogunjobi, Jerimiah Olugbenga, Olatunbosun Agbo-Ajala, Sunday Ekundayo, Olutosin Taiwo, Akeem Agunbiade, Ibrahim Gafar and the lovely inner caucus (Samuel Samson and Popoola Oladimeji) for being good friends and colleagues. You guys are awesome.

Finally, my most profound appreciation goes to my loving family for their encouragement and prayers throughout the PhD journey. I am grateful to the Oloko-Oba's, Olowo's, Onipe's, and Ayilara's. Special thanks to my siblings Taibat, Hameed, Ridwan and Aishat for your constant support. Most of all, to my darling wife (Kafayat Olabisi) and loving kids (Mukhtar and Mustakeem) for your love, patience and encouragement throughout the years of studies. I say JAZAKALLAHU KHAIRA.

Dedication

This research work is dedicated to the loving memory of my late parents Alhaji Abubakar Sidiq and Alhaja Rodiat. May Allah continue to bless you in death.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Introduction

Tuberculosis (TB) is one of the global sources of ill health and among the leading causes of death, as reported by the World Health Organization (WHO). TB mostly attacks the lungs (pulmonary) but sometimes attacks other parts of the body (extra-pulmonary) [36]. TB disease is contagious and can be transmitted through the air by sneezing, spitting and coughing from an infected individual [23]. Most of the infected individuals usually have TB disease in a latent state, which gradually advances into active TB if left undetected and untreated [21]. Typical indications of TB include persistent coughing, weight loss, fatigue, night sweats, and low fever [3, 42, 6]. Both men and women are at risk of contracting TB, although it is more prominent among men. As of 2018, about 1.2 million deaths were recorded from the 10 million estimated TB infected individuals, which was a decrease from the previous 1.3 million and 1.7 million deaths recorded in 2016 and 2011 respectively [36, 35].

Meanwhile, the WHO 2021 global TB report already estimates that 4.1 million people are currently infected with TB, increasing from 2.9 million in previous years. The burden of TB is far more prevalent in developing regions [27] where high death tolls resulting from TB infections are recorded compared to developed regions where sophisticated diagnosis equipment is readily available. TB patients lose their lives yearly due to diagnostic delay, misdiagnosis, and lack of appropriate treatments [57]. Although TB is a global challenge, the mortality rate is more prevalent in low and middle-income nations [46].

TB is certainly treatable if diagnosed early enough for appropriate treatment. Early diagnosis is essential for successful treatment, preventing further spread, and significantly reducing the mortality rate in line with the World Health Organization (WHO) End TB Strategy [11]. The gold standard for TB screening is Sputum culture. However, posterior-anterior chest radiograph (CXR) is an effective technique with low-cost and moderately low radiation doses for screening lung abnormalities

to achieve prompt results [34]. CXR has been adequately employed in developed countries for analyzing individuals exhibiting symptoms of active TB. At the same time, its application is limited in developing countries where TB is most prevalent [26, 39]. High TB burden regions lack the skilled radiological expertise required to interpret CXR images adequately [28, 41].

In the last decades, several efforts have been made using Artificial Intelligence (AI) to develop a Computer-Aided Detection (CAD) system to advance automatic object/image recognition tasks and overcome the challenges of a skilled workforce. Machine Learning (ML) and Deep Learning (DL) are the predominant AI techniques employed to develop CAD systems for analyzing CXR images. Both methods have had a significant impact, but the DL approach, such as Convolutional Neural Network (CNN), has become more prominent for analyzing different pulmonary abnormalities in the medical domain, most notably in diagnosing TB. The application of an efficient classification tool is vital for improving the quality of diagnosis while reducing the time taken to analyze a large volume of CXRs [47]. This is an endeavour to achieve a global decline in TB incidence to about 5% annually compared to the current 2% yearly as part of the World Health Organization's strategy to End TB [37].

More recently, CNN algorithms have been broadly used in Computer Vision for object detection, classification, and video analyzing tasks because of their ability to distinguish one object from another by assigning important learnable biases and weights to various images [2, 20]. CNN's are feed-forward neural networks and are renowned for performing better than other classification networks due to taking advantage of the inherent properties of the images. Also, the image pre-processing required in a CNN is less compared to other classification networks [12, 51]. The availability of large datasets used for training and the increased numbers of high-end computing power also help us to choose the deep CNN models for the classification task. CNN models can learn compact and discriminating pulmonary features to obtain the relevant information needed for the classifications of TB. In the light of these encouraging signs of progress in DL and CNN, we propose a novel end-to-end DL-based classification model that detects pulmonary abnormalities relating to TB from CXR and efficiently classify it as healthy or infected. See the WHO website[1] for updated facts about TB.

1.2 Motivation and Problem Statement

Tuberculosis (TB) is a global health challenge, and the mortality rate is more prevalent in low and middle-income nations. TB patients lose their lives yearly due to diagnostic delay, misdiagnosis, lack of appropriate treatments, and lack of or inadequate expert radiologists. Years of progress in providing essential TB

[1] `https://www.who.int/teams/global-tuberculosis-programme/tb-reports/global-tuberculosis-report-2021.`

services and reducing the TB disease burden has been hampered by the COVID-19 pandemic. Identifying the extent of TB infection either as latent or active is fundamental as these conditions require different treatment procedures. Diagnosing TB is sometimes difficult especially in cases where the patients exhibit symptoms that are similar to other illnesses and scenarios where the patients exhibit co-infections such as TB + HIV makes it even more problematic as the presence of active TB could be masked, resulting in an undetected and untreated case which increases the likelihood of death in patients. Due to the deadly nature of TB disease and the rate at which it can easily be spread, WHO has emphasized more proactive measures for the continuous reduction of TB cases and deaths. Also, the decision to embark on a mission to put an end to the global TB epidemic by the year 2030 as contained in the 2019 Global Tuberculosis report [38] is underway and requires urgent attention.

TB disease is certainly curable but needs to be detected early for appropriate treatment. Traditional approaches employed in screening TB include Tuberculin Skin Test [7], Smear Microscopy [10], and Interferon Gamma Release Assays [58], among others. Regrettably, most of these approaches have limitations, especially in developing countries, ranging from high cost for general adoption, inadequate equipment, requires stable power supply, inaccurate results due to low sensitivity, and the inability to distinguish between latent and active TB. Several medical imaging modalities are available to screen TB, but WHO has recommended CXR as a diagnostic tool that adequately identifies pulmonary abnormalities, reduces diagnosis errors, and ultimately increases the quality of diagnosis. Regrettably, the lack of expert radiologists to adequately interpret CXR images in high TB-burden regions has limited its applications in the past [55]. These limitations have motivated various efforts in developing CAD systems for the automatic detection and classification of TB.

Existing CAD systems are characterized by low sensitivity and accuracy, which remain significant concerns that lead to misdiagnosis. These systems are also computationally complex and expensive, making them unaffordable in a low-budget TB burden economy. To address these challenges, we propose an efficient system with high accuracy and sensitivity for detecting and classifying TB from CXR.

1.3 Research Aims and Objectives

This research aims to design a model for an efficient diagnosis of TB from pulmonary CXR. The specific objectives are to:

(i) Extensively study and investigate the development of the existing CAD system for TB diagnosis with the primary focus on data acquisition, methods implemented, performance accuracy and limitations.

(ii) Conduct a critical survey of the state-of-the-art DL methods related to TB detection and classification.

(iii) Model a CNN-based framework for timely and accurate classification of pulmonary CXR as healthy or infected with TB.

(iv) Develop a novel Ensemble model to improve classification accuracy of TB.

1.4 Technical Contributions

The technical contributions of this research to the field of medical imaging diagnosis and computer vision are summarized below:

(i) The study presents a comprehensive review of various traditional TB diagnostic tests. It discusses popular ML and DL architectures that have been employed on medical imaging to detect pulmonary abnormalities, summarize prominent CXR datasets and briefly discuss the steps involved in detection models from image pre-processing, features extractions, and features classifications. The study highlights state-of-the-art DL (CNN-based) architectures stating their strengths and weaknesses, evaluation metric, and performance of the current TB classifiers. https://doi.org/10.3389/fmed.2022.830515.

(ii) The study designed a robust image processing algorithm that handles the CXR pre-processing such as data augmentation, contrast enhancement, edge features, and noise removal before feeding it to the design CNN model. https://doi.org/10.1007/978-3-030-64559-5_44, https://doi.org/10.1007/978-3-030-59006-2_11, https://doi.org/10.1007/978-3-030-51935-3_16.

(iii) Developed novel CNN architectures and fine-tuned pre-trained CNN to detect abnormalities relating to TB from CXR and effectively classify it as healthy or infected. https://doi.org/10.1007/978-3-030-64559-5_44, https://doi.org/10.1007/978-3-030-51935-3_16, https://doi.org/10.1007/978-3-030-88081-1_41, https://doi.org/10.1007/978-3-030-59006-2_11.

(iv) The study demonstrate that fine-tuning a pre-trained model on the pulmonary CXR produced good generalization on the test set and leads to high accuracy. This approach is important in the medical field, where annotated datasets are limited. https://doi.org/10.1007/978-3-030-64559-5_44, https://doi.org/10.1155/2021/9790894.

(v) This thesis demonstrate that performance accuracy and sensitivity of a model can be improved through ensemble learning (bagging). https://doi.org/10.1007/978-3-030-88081-1_41, https://doi.org/10.1155/2021/9790894.

(vi) Demonstrate that training a model on a dataset and testing it on a different dataset that does not originate from the training subset gives better generalization on the diagnostic accuracy. (i.e. train using Shenzhen and test using Montgomery) https://doi.org/10.1007/978-3-030-88081-1_41.

1.5 Thesis Organisation

This thesis is divided into five chapters, and it follows a format submitted by publications. The thesis is organized as follows:

Chapter 2 Covers the background and related work. It presents a comprehensive review of the existing methods for developing computer-aided detection systems. This section further reviews the state-of-the-art DL techniques for TB diagnosis, stating the CXR datasets, evaluation metrics, and results.

Chapter 3 Presents the publications of the proposed models for detecting and classifying TB. It shows the implementation of the models.

Chapter 4 Contains the comprehensive analysis and discussion of the experimental results, including a comparison with results in the literature.

Chapter 5 Concludes this thesis with a summary of the presented method and recommendations for further research to improve the diagnostic accuracy of TB.

Chapter 2

Literature Review and Related Works

2.1 Tuberculosis Detection: A Comprehensive Review of State-of-the-Art

2.1.1 Introduction

This section presents a comparative analysis of previous research in TB detection. The review paper presented a thorough discussion on Machine Learning (ML) and Deep Learning (DL) architectures that have been employed on medical imaging to detect pulmonary abnormalities. It summarised prominent CXR datasets and briefly discussed the steps involved in detection models from image pre-processing, features extractions, and features classifications. The study highlights state-of-the-art DL architectures stating their strengths and weaknesses, evaluation metric, and performance of the current TB classifiers.

The paper suggests some aspects that may improve the medical imaging analysis for TB diagnosis in the future. The literature review manuscript is in the final revision stage of peer-reviewed in the IAENG International Journal of Computer Science (IJCS).

Tuberculosis Detection: A Comprehensive Review of State-of-the-Art

Mustapha Oloko-Oba and Serestina Viriri

***Abstract*—Tuberculosis is one of the global sources of deaths according to the World Health Organization. Tuberculosis is certainly curable but needs to be detected early. While several screening techniques exist, chest x-rays have been recommended by the World Health Organization as an important diagnostic aid. Therefore, the automatic detection of Tuberculosis from chest x-rays has become a crucial research focus in the Machine Learning and Deep Learning domain. In this paper, we present various traditional Tuberculosis diagnosis tests with more focus on the state-of-the-art models. We discuss popular Machine learning and Deep learning architectures that have performed excellently on medical imaging to detect chest x-ray abnormalities, summarise some prominent chest x-ray datasets and briefly discuss the steps used in detection models ranging from image pre-processing, features extractions, and features classifications. We also highlight the common evaluation metrics used to measure different models in terms of performance accuracy, strengths and weaknesses. Lastly, we suggest some aspects that may bring about additional improvement in the area of medical imaging analysis for Tuberculosis diagnosis in the future.**

***Index Terms*—Tuberculosis. Chest X-Ray. Machine Learning. Deep Learning. Computer-Aided Diagnostic.**

I. Introduction

ONE of the global sources of death as reported by the World Health Organization (WHO) is Tuberculosis (TB) which mostly attacks the lungs but sometimes attacks other body parts [1]. TB infection is contagious and can be transmitted through the air by means of sneezing, spitting and coughing from an infected individual [2]. The Majority of the infected individuals usually have TB in a latent state which gradually advances into active TB if left undetected and untreated [3]. Typical indications of TB include persistent coughing, weight loss, fatigue, night sweats, and low fever [4], [5], [6]. Both men and women are at risk of contracting TB although; it is more prominent in men. As of 2018, about 1.2 million deaths were recorded from the 10 million estimate of TB infected individuals, which is a decrease from the previous 1.3 million and 1.7 million deaths which were recorded in 2016 and 2011 respectively [1], [7]. The burden of TB is far more prevalent in developing regions [8] where high death tolls resulting from TB infections are recorded compared to developed regions where sophisticated diagnosis equipment is readily available.

Accurate and timely diagnosis of TB is ultimately essential to treating and eradicating this global epidemic disease. Identifying the extent of TB infection either as latent or active is fundamental as these conditions require different treatment procedures. The issue of resistance to drugs makes finding a cure for TB more problematic as most of the available drugs are becoming ineffective and hindering the goal of extinguishing the infection [9]. Diagnosing TB is sometimes difficult in cases where the patients exhibit symptoms that are similar to other illness and scenario where the patients exhibit co-infections such as TB + HIV makes it even more problematic as the presence of active TB could be masked, resulting in an undetected and untreated case which increases the likelihood of death in patients [2].

Recently, classification methods have become popular for screening various abnormalities in medical imaging as the utilization of an efficient classification tool is important for improving the quality of diagnosis while reducing the time taken to diagnose [10]. There are several means of screening for TB in a patient such as the chest x-ray (CXR), Tuberculin Skin Test (TST), Sputum Microscopy; and Mycobacteria Growth Indicator Tube (MGIT), among others [11] but the CXR is a less expensive way of diagnosing for TB manifestation [9], [12]. Regrettably, interpreting the CXR is dependent on the skill and expertise of a radiologist and is subject to misdiagnosis [13], [14]. Diagnosing a large number of patients with TB symptoms requires being x-rayed and screened for the type of TB (either latent or active) to make certain that appropriate therapy is administered.

Also, manually diagnosing a large number of patients is a laborious and time-consuming task which could result in misdiagnosis. In most developing regions, for instance; in Africa where the resources required to undertake adequate diagnosis are lacking, it is a major challenge to curb the TB epidemic. Consequently, developing Computer Aided Detection (CAD) systems to automatically detect TB infection in CXR is a promising way of reducing diagnosis errors and ultimately increasing the quality of diagnosis.

The rationale for this review study is to:

- Present a comprehensive background that serves as a guide for the application of Deep Learning in the development of an existing computer-aided diagnostic (CAD) system for Tuberculosis detection;
- Highlight the strengths and weaknesses of current state-of-the-art Deep Learning architectures.
- Provide detailed information on the various CXR dataset; and finally
- Identify future directions that may bring about further improvements in analyzing medical imaging and developing efficient CAD for diagnosing TB and other pulmonary diseases.

This study contains recent and relevant details that are specifically useful for researchers interested in applying Deep learning architectures to real-life problems, academia, and governments interested in the extent of effort invested in developing CAD for solving the global Tuberculosis epidemic.

Mustapha Oloko-Oba is a PhD candidate at the School of Mathematics, Statistics and Computer Sciences, University of KwaZulu-Natal, Durban, South Africa e-mail: 219098624@stu.ukzn.ac.za.

Serestina viriri is a Professor of Computer Science at the School of Mathematics, Statistics and Computer Sciences, University of KwaZulu-Natal, Durban, South Africa. e-mail: viriris@ukzn.ac.za.

II. Survey Methodology

This section discuss the comprehensive procedures for the articles reviewed in this study, including article database, search keywords, and standards for inclusion.

A. Databases

The databases utilized for searching articles were IEEE Xplore, Scopus, and Science Direct, basic and advanced searches for related journals, conference proceedings and books in the computer science and medical domain. The focus was more on the journals and conference proceedings articles because they often contain recent research discovery related to the study.

B. Search Keywords

Searches were performed across the three databases using the following search terms: tuberculosis, Deep Learning application, chest x-ray, diagnosis of tuberculosis, segmentation of lung fields from a chest x-ray, tuberculosis screening, computer-aided detection, and classification.

C. Standards for Inclusion/Exclusion

- Article language is an important criterion considered in this study. The articles included were written entirely in English while papers presented in other languages were excluded.
- Articles selection was based on a comprehensive search for publication in the medical domain related to pulmonary tuberculosis abnormalities and computer science domain with respect to (Deep Learning and Machine Learning) methods.
- The articles reported cover the period between 2000 to 2020. This period was suitable for exploring the research evolution and findings on the conventional and state-of-the-art methods of tuberculosis diagnosis.
- Articles that did not sufficiently report methods and detail results in terms of performance accuracy were excluded.
- Articles were searched using the appropriate keywords as listed above, the search results were then scanned to filter topics highly related to the study. An overview reading of the abstract, methods, and conclusion was followed to further screen relevant articles for full-text reading and better understanding.

371 articles were subjected to title and abstract scans, out of which 200 papers were found suitable for full-text reading. The detailed reading further streamlined the paper to 160, which were finally included in the study.

III. Conventional Tuberculosis Diagnosis Test

Before the advent of Deep Learning in diagnosing TB from CXR, several methods were used in screening TB manifestation in patients. These methods are briefly discussed below:

A. Tuberculin Skin Test

Tuberculin Skin Test (TST) is one of the common methods of diagnosing TB. In this method, a small amount of Tuberculin which contains some inactive TB protein is injected into the foream skin. The patient will then return after 2-3 days to observe if there is a reaction or not. Mostly, if the patient has some swelling or a hard lump on the arm, this indicates that there is the presence of TB bacteria in the body otherwise the patient is considered negative for TB bacteria when there is no reaction in the arm. TST is vital in identifying the presence of TB bacteria [15] but cannot determine the state of the TB infection as either latent or active TB [2].

B. Interferon Gamma Release Assays

Interferon Gamma Release Assays (IGRA) is a TB examination procedure that is dependent on blood to measure patient immune response to TB bacteria. In the IGRA screening, interferon-gamma is released in the white blood cells of an infected individual when combined with TB protein derivative [16]. Just like the TST, IGRA screening is also unable to distinguish latent TB from active TB but is more explicit than the TST because it is not affected by prior Bacillus CalmetteGurin vaccines [17].

C. Smear Microscopy

Smear Microscopy involves placing of sputum (a mixture of mucus, spit, saliva) under the microscope for a period of six weeks for the detection of mycobacterial growth. This method is one of the primary means of TB diagnosis in a low-income region where TB is most prevalent [18]. This method is mostly inaccurate as a result of low sensitivity in co-infected cases, it has a low track report in paediatric TB, and extra pulmonary TB and takes too long to give a result [19], [2].

D. GeneXpert/Xpert MTB/RIF

GeneXpert/Xpert MTB/RIF is an amplification examination for detecting mycobacteria TB (MTB) and resistance to a TB drug known as Rifampicin. The Xpert MTB/RIF is an initial diagnostic screening for TB infections and diagnoses faster and better than the Smear Microscopy but it is very expensive and requires a constant electricity supply [20] which makes it a challenge in developing regions where stable electricity is lacking.

E. Mycobacteria Growth Indicator Tube

The Mycobacteria Growth Indicator Tube (MGIT) developed by Becton Dickinson [21] is used to screen the possibility of TB bacteria growing when subjected to different TB drugs to determine if the TB bacteria are resistant to a particular drug or not. This is an expensive screening method; it requires sufficient testing equipment and skilled personnel as it needs to be isolated in a laboratory [22], [23].

Another TB drug resistant screening method that requires being isolated in the laboratory is the Genotype MTBDR and INNO-LiPA Rif TB [24]. This testing technique also requires sufficient equipment and skilled personnel as in the case of MGIT.

All the generally known TB diagnostic approaches have limitations especially in co-infection cases as evident above, hence the struggle continues to ensure the rapid reduction in the mortality rate recorded and total eradication of TB epidemic [1] in which early detection is a decisive factor [9].

The constraints of the existing TB diagnosis techniques contribute to diagnostic delays resulting in severe consequences for the global efforts of controlling and eradicating the epidemic [25]. CXR can be employed as both a measurement and screening medium to identifying all the various manifestations and abnormalities with respect to pulmonary infections [26].

The advent of Computer-Assisted Detection (CAD) system for the detection of TB manifestation is a promising tool that will assist health care centres to eliminate the delay arising from the conventional methods of diagnosing patients and also provide accurate diagnoses in the developing regions where TB is most prevalent.

IV. Challenges Associated with TB Diagnosis in Chest X-Ray

Delay in screening is the very first challenge that impacts on the quality of diagnosis. Screening delay is a key factor that can diminish the quality of diagnosis and contribute hugely to the high rate of death associated with TB. Delays could be patient-based (arising from the onset of clinical symptoms until the initial visit to the medical facility), or health personnel based (arising from the patient's initial visit to a medical facility until when the diagnosis is established). Inadequate screening equipment, high cost of screening, lack of skilled medical personnel are among other diagnostic delays that are mostly observed in high TB burden regions such as Africa,and Asia [2]. The global impact of TB liability is on the increase in the developing regions compared to developed regions. Individuals that show TB manifestation can be clinically categorized into latent and active TB. The latent (dormant) TB is easily advanced into active TB when left untreated or detected.

The common radiological entities of TB manifestations on CXRs are:

- Miliary Pattern: appears in the lungs in a sand-like formation affecting around 1% - 7% TB infected persons. The miliary pattern usually displays manifestation between 6 months from the onset of exposure [27], [28].
- Adenopathy: results in enlargement of the nodes. It usually affects the right-side region of the lung and is mostly observed in children (up to 83%-96%) as compared to adults (between 10%–43%) [29].
- Pleural Effusion: This is excessive fluid building up within the medial and lateral regions. The pleural effusion usually produces a small amount of fluid that acts as a lubricant that facilitates breathing but once infected, it produces excess fluid [30].
- Airways Enlargement: is usually visible as tabular rings extending from the lung [31].
- Airspace Consolidation: is a condition that is visible in the lobe as opacity; it is often diagnosed as pneumonitis [32].

Most of the attributes associated with TB infection may also be present in some other diseases. In some cases, the chest screening of an infected patient looks perfect to the human eye, while some may exhibit faint abnormalities that can only be sported by a skilled radiologist [33].

As reported in [26] CXR is a vital screening procedure in detecting abnormalities as well as the manifestation of TB in patients. CXR can be employed as a guide to determine when additional screening of infected individuals is required, and provide reassuring proof to confirm TB manifestation when the bacteriologic outcome is undetermined or uncertain. Interpreting CXR as both measurement and screening procedure is of utmost importance.

Screening is carried out for identifying any abnormalities relating to TB. However, it is pertinent to include both TB specific and non TB CXR in the screening procedures to improve the quality of the result. The utilization of computer-aided detection (CAD) systems in analyzing CXR can improve the quality, timeliness and accuracy of diagnosis thereby surpassing manual human interpretation [26].

CXR has some limitations such as requiring skilled professionals for proper screening exercise and interpretation, standard equipment with stable power (electricity), and exposure of ionizing radiation. Recent efforts such as decreasing the dose of radiation, improving the archiving facility, advancing quality of image, utilization of CAD, and harmonizing scan interpretation and reports have assisted in maintaining CXR as a vital tool in detecting TB cases.

V. State-of-the-art Convolutional Neural Network Architectures

In this section we present prominent Convolutional Neural Network (CNN) models that are capable of extracting and learning distinctive features/patterns from images. These models have been successfully and widely utilized for various tasks such as diagnosis, recognition, classification, prediction, and estimation, etc. The AlexNet, GoogLeNet, VGGNet, ResNet, SqueezeNet, Xception, CapsuleNet, and ZFNet architectures are described below.

A. AlexNet Architecture

In 2012, Krizhevsky et al [34] presented the AlexNet structure which is made up of five successively convolutional layers and three fully connected layers. The architecture achieved the topmost performance in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC-2012) competition against all other competitors. Their achievement brought about a fundamental advancement in the field of Machine Learning for recognition and classification tasks which lead to increased interest in the Deep Learning domain. The AlexNet architecture employed augmentation and dropout techniques to drastically reduce overfitting and was the first to implement the Rectified Linear Unit (ReLU) activation function. AlexNet is, however, computationally expensive. Figure 1 shows a representation of the AlexNet structure.

B. Visual Geometry Group (VGGNet) Architecture

The Visual Geometry Group (VGG), was presented by [35] and emerged as the runner up after GoogLeNet in the ImageNet ILSVRC 2014 classification challenge. The VGG structure consists of convolutional layers which utilize the ReLU activation function followed by a max-pooling layer

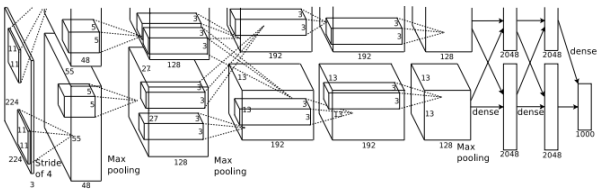


Fig. 1. The AlexNet Architecture. [34]

and then the fully connected layers that also utilize the ReLU activation function except for the final layer which uses Softmax activation function for classification. This model exhibits that the depth of a network is important for obtaining better performance accuracy. The VGGNet uses 3x3 kernel size filters as a modification to the AlexNet 11x11 and 5x5 filters. There are different variants of VGG like VGG-11, VGG-16, and VGG-19, all of which have the same number of fully connected layers but different convolutional layers specific to each variation. This model has established effectiveness as evident in [36], [37], [38] for TB detection with better generalization even though it is time-consuming when training from scratch, sluggish to train and the weights of the model are quite large. See Figure 2 for the VGGNet architecture.

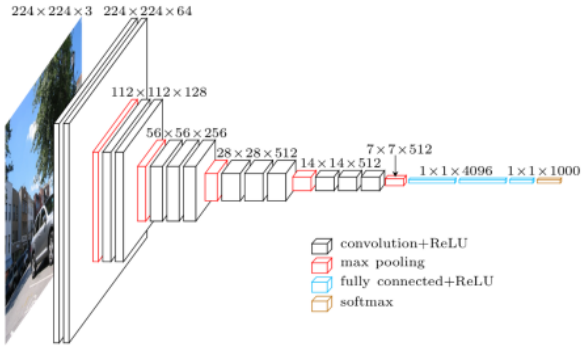


Fig. 2. The VGGNet Architecture [35]

C. GoogLeNet Architecture

GoogLeNet, otherwise known as Inception V1, was presented at the 2014 edition of the ImageNet Large-Scale Visual Recognition Challenge by [39] and won the best architecture with respect to detection and classification of images. The architecture which consists of 22 deep network layers is aimed at reducing computational complicacy of previous architectures to achieve improved detection and classification of objects.

The 22 layers depth of GoogLeNet learns faster and surpasses the previous architectures. However, it has lower computation and lesser network parameters of just about 7 million compared to VGG-19's 138 million and AlexNet's 60 million. GoogLeNet has successfully been used in [40], [41], [36] for detecting TB from CXR. The GoogLeNet architecture is presented in Figure 3.

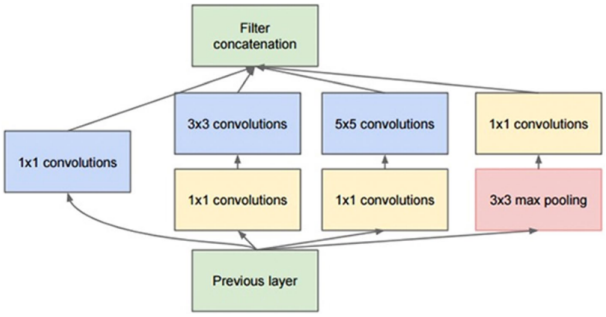


Fig. 3. The GoogLeNet Architecture [39]

D. ResNet Architecture

Residual Network, popularly known as ResNet, was presented by [42] to resolve the vanishing gradient issue and simplify the complexity of deeper networks. Rather than learning the unreferenced functions, the ResNet framework distinctly recomposes layers as learning residual functions with reference to the layers input. The network offers the idea of identity mappings and residual section and consists of different residual blocks.

The residual section consists of two parts: firstly, an alternative which links the input to the second part, and then convolutions with an activation part. The result of the residual network won the 2015 edition of the ILSVRC classification challenge. ResNet is easier to optimize and attained better accuracy from deeper network using the Deeper Bottleneck structure, an extension of the residual section. Nevertheless, ResNet takes lots of time to train and is quite expensive computationally. ResNet has a deep variant for difficult tasks and a shallow variant for simple tasks. Other variations of residual models are proposed in [43], [44], [45]. Figure 4 shows a residual network architecture.

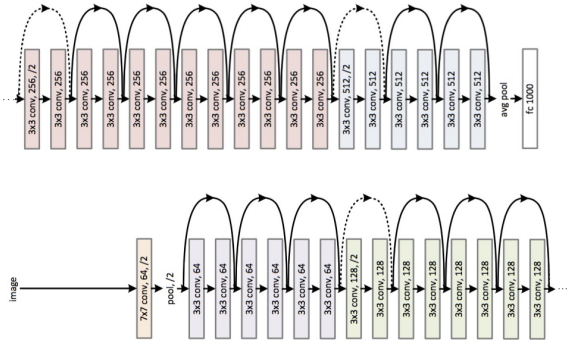


Fig. 4. A ResNet Architecture with 34 parameter layers [42]

E. SqueezeNet Architecture

A mini CNN model named SqueezeNet was proposed by [46] with the aim of achieving the same accuracy level as the robust architectures rather than only focusing on increased accuracy with deeper models. The authors believed that building mini architectures were more feasible to deploy

across networks with little memory size, required smaller bandwidth to ease exporting, as well as limiting communication during training across servers. SqueezeNet architecture is less than 0.5 mb model size and has fewer parameters yet attained the same accuracy level with a deeper network AlexNet on the same ImageNet datasets.

The SqueezeNet architecture is developed out of fire modules. The architecture begins with inputting an image into the convolutional layer which is then passed through 8 sequential fire modules while the number of filters per module are gradually increased. Then the process is moved to another convolutional layer.

Max-pooling operation is performed in the model with strides of 2 at different intervals as shown in Figure 5. Global average pooling is employed to replace the fully connected layers before finally passing on to the SoftMax classifier. Although, SqueezeNet can achieve better accuracy but the small nature of the architecture will impact on its generalization performance.

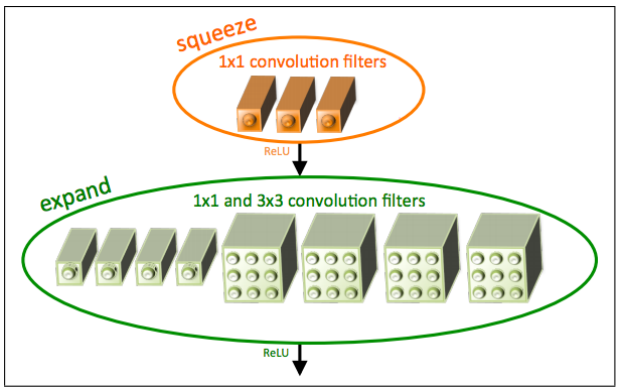


Fig. 5. SqueezeNet Architecture [46]

F. Xception Architecture

Xception, which stands for Extreme Inception, is completely dependent on the depth wise separable convolution [47] and is presented in [48] with the assumption that spatial correlation and cross channels correlations mapping in the feature maps of CNNs can be completely decoupled. The feature extraction part of the Xception model is composed of 36 convolutional layers followed by logistic regression layers. The convolutional layers within the Xception model are arranged into 14 layers consisting of linear residual connections around them excluding only the first and last modules. The linear stack with residual connections mode makes defining and modifying the Xception model easier than the Inception V3. The experimental results of the Xception model showed exceptional performance over the Inception V3 model on a larger JFT dataset but was a slight improvement on the ImageNet datasets which the Inception V3 model was originally built upon. The Xception model is presented in Figure 6.

G. CapsuleNet Architecture

CapsNet architecture presented in [49] is a shallow network that consists of two convolutional layers and one fully

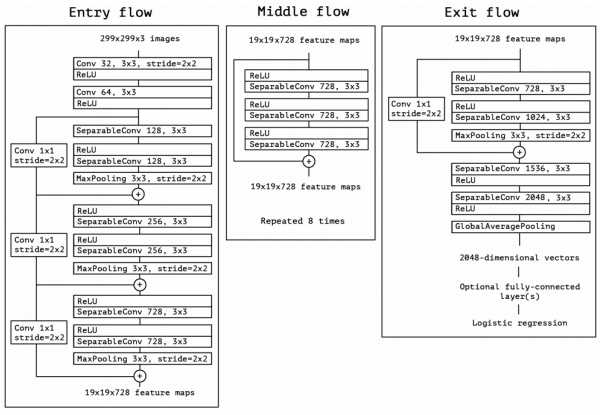


Fig. 6. The Xception Architecture [48]

convolutional layer. The first convolutional layer of CapsNet changes pixel intensities to the task of local feature detectors that are further pushed as inputs to the primary capsules modules and then passed to the second convolutional layer before passing on to the DigitCaps. The idea behind CapsNet is to add a capsule module to CNN; the output from these capsules is then used to form steady representation for greater capsules to output a vector. CNN does not consider the relationship between objects as well as affine transformation which is the strength of CapsNet. Training CapsNet from scratch on MultiMNIST dataset attained a higher test classification accuracy than baseline convolutional structures and also performed well on the CIFAR-10 dataset. Also, in [50], the authors employed customized CapsNet along with VGG-16 and AlexNet for classifying TB in CXR. The result showed that CapsNet structure outperformed VGG-16 and AlexNet with respect to predicting affine images. The general architecture of CapsNet is presented in Figure 7 while Figure 8 is a customized CapsNet architecture for TB detection.

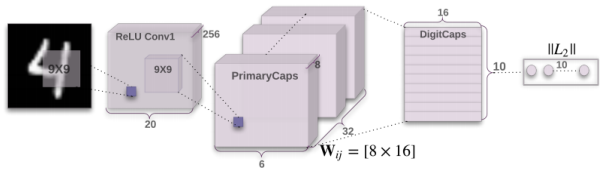


Fig. 7. A simple 3 layers CapsNet Architecture. The occurence of every class is represented in DigitCaps layer that is utilized for computing classification loss. W_ij represents the weights between primary capsule layer and DigitCaps layer [49]

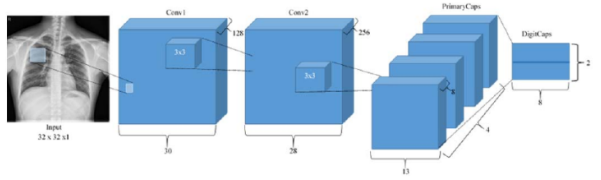


Fig. 8. CapsNet architecture for tuberculosis detection [50]

H. ZFNet/Clarifai Architecture

In the 2013 edition of the ImageNet classification challenge, the authors in [51] present ZFNet, an 8 layers model which improved on the AlexNet architecture. ZFNet showed significant improvement by tweaking the AlexNet model. They introduced a visualization technique which provided insight into the intermediate layers to identify problems of the model. The modifications that resulted in improving the classification errors were (a) substituting the 11x11 filters in the AlexNet first convolutional layer with a 7x7 filters; and (b) utilizing strides of 2 as substitute for the original 4 strides. These modifications were important to the ZFNet architecture for outperforming the AlexNet on the same ImageNet benchmark dataset and won the 2013 ILSVRC challenge. The ZFNet also showed better generalization and outperformed modern results on both CalTech-101 and 256 datasets.

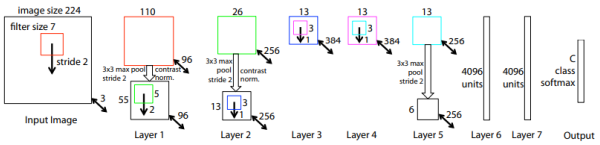


Fig. 9. Architecture of ZFNet [51]

Other architectures include FractalNet [52], Densely Connected Network (DenseNet) [53], Network in Network (NiN) [54], and Faster R-CNN [55].

VI. Overview of Prominent CXR Datasets

Data is crucial for the development of algorithms required to solve life-threatening diseases including TB which is one of the leading causes of death worldwide [1]. The various prominent CXR datasets that have been employed in training and testing TB detection algorithms are described below. These datasets are made available to the research communities to foster state of the art researches into finding lasting solutions to the early diagnosis of TB manifestations.

A. Montgomery County Dataset

The Montgomery County (MC) CXR dataset is a TB specific dataset made available by the National Library of Medicine in conjunction with the Department of Health Services Maryland, U.S.A., for research intent. This dataset is composed of 58 abnormal samples and 80 normal samples. The naming of each image ends with either "0" denoting normal sample or "1" denoting abnormal sample. All samples are of the size 4020 x 4892 pixels and available as portable network graphic (png) file format. This dataset is accompanied with clinical readings that give details about each of the samples with respect to sex, age, and manifestations. The MC dataset is accessible from the NIH website[1].

[1]https://lhncbc.nlm.nih.gov/publication/pub9931

B. The Shenzhen Dataset

The Shenzhen dataset is specific to TB and is publicly available for the purpose of research. It is made up of 336 abnormal samples and 326 normal samples. The naming of each image ends with either a 0 denoting a normal sample or a 1 denoting an abnormal sample. All images are approximately 3000 x 3000 pixels saved as portable network graphic (png) file format. This dataset is accompanied by clinical readings that give details about each of the samples with respect to sex, age, and diagnosis. The Shenzhen dataset is accessible from the NIH website.

C. JSRT Dataset

The JSRT dataset is put together by the Japanese Society of Radiological Technology in collaboration with the Japanese Radiological Society. The dataset is composed of a total of 247 CXR of which 93 are without lung nodules and 154 are with long nodules. The images are 2048 x 2048 pixels size in a universal image format (uif). This dataset also has additional patient information such as gender, age, diagnosis, nodule location diagram, degree of subtle, and X and Y coordinates of nodules.The JSRT dataset can be requested from the Japanese Society of Radiological Technology website[2].

D. Indiana (OpenI) Dataset

The Indiana CXR dataset was collected from two hospitals by the Indiana University Network for Patient Care. The dataset is composed of a total of 3996 posterior-anterior CXR images of varying sizes ranging from 1024 pixels minimum dimension to a maximum dimension of 4248 pixels. It contains 1526 normal images and 2470 abnormal images of 10 different manifestations including pleural effusion, cardiomegaly, opacity, etc. and is accompanied by radiologist reports. The dataset can be downloaded at: https://openi.nlm.nih.gov/faq#collection.

E. The Peruvian Dataset

The Peruvian dataset is provided by the Socios en Salud, Partners In Health in Lima, Peru [41]. It consists of a total of 4701 CXR images out of which 453 are normal i.e. healthy while 4248 are abnormal class, that is unhealthy with various manifestations. The manifestations are classified into six such as cavitation lymphadenopathy, miliary pattern, alveolar infiltration, ghon focus, and others. The images are stored as JPEG file format.

F. KIT Dataset

The KIT is a TB specific dataset and is made up of 10848 CXR images provided by the Korea Institute of Tuberculosis under the Korean National Tuberculosis Association (KNTA), South Korea [56]. There are 7020 normal samples and 3828 abnormal samples. The images are provided in the Digital Imaging and Communications in Medicine (DICOM) file format.

[2]https://db.jsrt.or.jp/eng.php

G. Belarus Dataset

The Belarus dataset is put together and maintained by Belarus TB Public Health. The dataset consists of 304 CXR images all of which are entirely positive for TB with different abnormalities such as miliary pattern, pleural effusion, and cavitation manifestations. Out of the 304 images, 194 are associated with male patients while 110 are associated with female patients. The CXR images vary in sizes as 2248 x 2248 to 2724 pixels and are stored as Digital Imaging and Communications in Medicine (DICOM) file format. The dataset is usually requested from the Belarus TB portal but was unavailable at the time of this report. The images were used in [40], [57].

H. India Dataset

The India dataset is obtained from private hospitals in India. The set is composed of a total of 397 CXR images with different dimensions like 1772 x 1430 pixels, 2010 x 1572 pixels, and 2446 x 2010 pixels respectively in a 12-bit gray scale. The images are 0.175mm pixels spacing in a horizontal and vertical directions. The India datasets is reported in [58].

I. CheXpert Dataset

CheXpert [59] is a huge CXR dataset collected at Stanford Hospital in California between the years 2002 to 2017. The dataset is composed of 224,316 CXR images along with their associated radiological interpretations from 65,240 patients and is labeled as positive, negative, and uncertain for the presence of 14 clinical observations. The dataset also features labeled radiologist tests and validation sets which serve as standard references, as well as expert scores evaluating distinctive models. The dataset is available for downloading after accepting the CheXpert research use agreement by the Stanford University School of Medicine to complete the registration form. The download link, which may not be shared with others, will be sent to the registered email for downloads. The agreement form to request the CheXpert datasets can be found at the Standford ML Group website[3].

J. MIMIC-CXR Dataset

The MIMIC-CXR dataset [60] is sourced from the Beth Israel Deaconess Medical Center between 2011 and 2016. The dataset consists of 377,110 CXR originally in Digital Imaging and Communications in Medicine (DICOM) file format and then exported in the JPEG standard format which is now MIMIC-CXR-JPG that is entirely derived from the MIMIC-CXR dataset, and aims to provide a convenient processed version of MIMIC-CXR, as well as provide a standard reference for image labels and data splits. The dataset is available to all researchers for use after completing a data use agreement which states that, the researcher will not attempt to re-identify the patients because all the images are de-identified to protect patients' privacy, the data is not shared with others, and relevant codes emanating from any publications for the use of the dataset be made available. A detailed description of the dataset and download can be found on the PhysioNet website[4].

[3]https://stanfordmlgroup.github.io/competitions/chexpert/

[4]https://physionet.org/content/mimic-cxr-jpg/2.0.0/

K. NIH Dataset

The NIH CXR [61] dataset was provided by the National Institute of Health, Bethesda, MD, USA and consists of 108,948 de-identified CXR including radiological reports saved as PNG format. The images are labeled with the presence of 8 pathologies as follows: Mass, Cardiomegaly, Pneumothorax, Effusion, Nodule, Atelectasis, Pneumonia, and Infiltration or otherwise labeled as normal if no manifestation is observed. Some images can have one or more pathology. This dataset is often referred to as ChestX-ray8 or Chest X-ray14 and is available for downloading at the NIH website [5] with no restrictions for use.

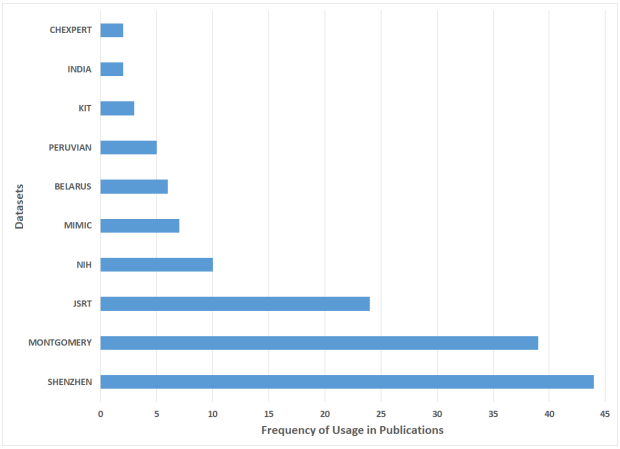


Fig. 10. A chart of CXR datasets by frequency of use in publications

Apart from the prominent dataset mentioned above, there are other independents datasets that have been utilized for researches. Some of these datasets include the TB specific datasets of the Sehatmand Zindagi Healthy Life Centres [62], CXR scans from the University of Alberta Hospital [63], Imagine dataset of the Seoul National University Hospital [64], the Boramae Medical Center datasets, Daejeon Eulji Medical Center datasets, and the Kyunghee University Hospital at Gangdong datasets[64].

Table I presents a brief summary of the datasets studied in this paper.

L. Summary

All the above datasets are de-identified in order to protect the privacy of the patients. In otherwords, the identity of the patients is not disclosed. Most of the CXR in the dataset are accompanied with radiological interpretations of observed manifestations. The Montgomery and the Shenzhen, for instance, also have a ground truth of the images to assist researchers in a segmentation task. The datasets are freely available to researchers for various research tasks such as segmentation of the lungs, detection of pulmonary abnormalities and so on. The MIMIC and CheXpert, for instance, require researchers to complete a data users agreement form before the datasets can be downloaded for use.

[5]https://nihcc.app.box.com/v/ChestXray-NIHCC

TABLE I
SUMMARY OF CXR DATASETS

Dataset	Dataset size	Dataset provider
Montgomery County	138	National Library of Medicine, Maryland, USA
Shenzhen	662	National Library of Medicine, Maryland, USA
JSRT	247	Japanese Society of Radiological Technology
Indiana (OpenI)	3996	Indiana University Network for Patient Care
Peruvian	4701	Socios en Salud, Partners In Health in Lima, Peru
KIT	10,848	Korea Institute of Tuberculosis, South Korea
Belarus	304	Belarus Tuberculosis Public Health
India	397	National Institute of Research in Tuberculosis, India
CheXpert	224,316	Stanford Hospital, California
MIMIC	377,110	Beth Israel Deaconess Medical Center (BIDMC), U.S.A
NIH	108,948	National Institutes of Health, Bethesda, MD, USA

VII. TECHNICAL ADVANCES AND THE NEED FOR TIMELY AND CORRECTLY DIAGNOSIS OF TB

In spite of the various efforts to control the TB epidemic, the rate of detecting active TB in developing regions remains a concern. As reported in [65], almost half of the cases of TB remain undiagnosed in the African region of World Health Organization (WHO), where only about 60% of cases of TB are detected, leaving about 40% of cases undetected.

In most TB + HIV co-infection regions, fewer than 18 out of 50 cases are detected. Differentiating and detecting latent or active TB in HIV co-infection cases in developing regions remains a serious problem. Also, out of about 500,000 new multi-drug resistant TB infected individuals estimated yearly, only 7% of cases are detected after a lengthy delay before diagnosis [66].

The optimal diagnosis of TB is faced with the crucial and logistic difficulty to purchase and provide the needed equipment especially in the developing nations when compared to the case of HIV where the maximum effort was involved to curb the HIV menace with huge investments that resulted in development of Point of Care (POC) appropriate for field use [67]. An important aspect that has hugely complicated the quality of diagnosing TB is the desire to differentiate active and latent TB in patients and also, the negligence of developing on-the-spot screening tools as compared to the case of HIV which contributes to diagnosis delays and errors. Various commercial screening products are available for HIV because of the awareness and fear of its deadliness, meanwhile TB, which is ranked more deadly than HIV is receiving less attention with respect to on-the-spot commercial screening products.

For effective and timely diagnosis of TB, there is a need to rapidly develop an efficient Point of Care (POC) toolkit suitable for on-the-spot diagnosis to eliminate diagnostic delays and facilitate rapid treatment for infected patients. It is predicted that the provision of an accurate and broadly acceptable diagnostic POC for TB detection leading to immediate treatment might perhaps prevent about 625,000 deaths yearly arising from the TB epidemic [68].

VIII. EVALUATION METRICES

It is important to highlight the various metrics that are commonly used to evaluate the quality of models and/ or how a model is compared with another in terms of performance, strength and weaknesses. The popular evaluation metrices that are used in TB detection task are discussed briefly below.

A. Accuracy

This refers to the rate of the correct outcome from the total number of examined samples [69], [70]. This is given as:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{1}$$

where

TP : is the case where the model predicts a sample as positive and it is actually positive.
TN : is the case where the model predicts a sample as negative and it is actually negative.
FP : is the case where the model predicts a sample as positive whereas the actual output is negative.
FN : is the case where the model predicts a sample as negative whereas the actual output is positive.

B. Precision

Precision, otherwise known as positive predictive value is the ratio of positive samples that are accurately predicted [69]. This is given as:

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

C. Sensitivity

This is the proportion of the actual positive samples that are correctly identified as positive [71], [72]. This metric is given as:

$$Sensitivity = \frac{TP}{TP + FN} \tag{3}$$

D. Specificity

Specificity is the ratio of the actual negative samples that are correctly identified as negative [71], [72]. This metric is given as:

$$Specificity = \frac{TN}{TN + FP} \tag{4}$$

E. AUC-ROC Curve

The AUC-ROC sometimes written as AUROC is an evaluating metric that indicates the ratio at which a model distinguishes between classes [70]. That is, it tells the probability of separating negative samples from positive samples. Thresholds are set to determine the ROC curve which is the plot of True Positive Rate (TPR) against False Positive Rate (FPR) given as:

$$TPR = sensitivity = \frac{TP}{TP + FN} \tag{5}$$

$$FPR = 1 - specificity = \frac{FP}{FP + TN} \tag{6}$$

IX. Pre-processing

Pre-processing is a vital image processing task that is aimed at transforming raw data into meaningful and efficient data [73], [74], [75] by removing/reducing/correcting unwanted distortions such as noise elimination, missing pixels etc. from the image [12]. In medical image processing, pre-processing technique is employed to enhance the quality of the image in such a way that the region of interest is clear to reveal the anomaly associated with the raw image for better accomplishment of the subsequent process such as features extraction, segmentation etc. to achieve optimal performance [9]. Image Enhancement, Bone Suppression, and Segmentation are typical approaches for pre-processing CXR images. These approaches are summarised below.

A. Image Enhancement

It is important to carry out image enhancement so as to reveal the uncertainty and low variation associated with CXR images [76]. The process is generally useful in minimizing diagnosis error and preserving the image details. Types of image enhancement usually applied to CXR include contrast enhancement, edge features (sharpening, filtering), and noise removal [9], [77], [78].

Contrast enhancement [79], [80] is used to boost the quality of an image by exposing the unclear details associated with it. It ensures the image is suitable for subsequent action such as feature extraction. An example of contrast ehnancement is shown in Figure 11.

Noise removal [77], [81] is applied to reduce or eliminate noise without distorting the image details. Edge features such as image filtering [82] an operation that utilizes the neighborhood pixel of an image to decide the output value of that pixel and image sharpening [76] are used to improve the edge of image quality.

B. Bone Suppression

Bone suppression is an essential pre-processing procedure to ease features extraction and lung segmentation on CXR [83]. Abnormalities in the lung can be blocked either by the clavicle or ribs which hinders feature extraction stage in CAD; as a result of this complexity, it is vital to extricate obstructing bony structures to improve visibility for accurate features extraction [9]. A bone suppression technique is proposed in [84] to suppress the possible overlap between the ribs and lung nodules which makes it difficult to accurately detect abnormalities in the lungs. Another bony structure suppression method was presented in [85] by utilizing k-nearest regression to estimate soft-tissue image such as scapulae and the ribs that overlay the pulmonary region which ultimately reduce the rate of misclassification. A comparison of two models was carried out in [86] to identify abnormalities on CXR. One of the models was trained using bone suppression x-ray images while the other was trained on the original CXR images. The authors found the bone suppression models outperformed the original image model. The utilization of bony suppression method has been proven to improve the performance of pulmonary nodule detection as well as other irregularities like detecting pneumonia [87]. Some other research work that employed bone suppression techniques includes recognition of lung cancer [88], separation of ribs from other parts of pulmonary images using ICA [89], skeletal bone suppression [90], enhancement of lung nodule [83], and bony thorax decomposition [91] among others.

C. Segmentation

Sectioning the region of interest is important in processing CXR for quality extraction of desired features for specific objectives. The essential taskof performing segmentation on images is to group together regions with same pixel value, and isolate regions of similarity or a specific part of interest from the original image for further analysis task [12]. Segmentation techniques have been applied on medical images for various purposes such as pulmonary nodule detection, abnormalities recognition, lung cancer estimation, pneumonia, detection of TB manifestation, and cardiomegaly among others. Much research work has been carried out to segment different regions but segmenting the lungs is more prominent as a result of its accuracy in delineating fields where certain radiological characteristics show increased density as a result of one or more instances of consolidation, mass, atelectasis, and interstitial [76]. Segmentation techniques can be grouped into Rule-Based, Pixel Based, Deformable, and Hybrid models [9], [92]. these are explained below.

Rule-based model: Lung fields are segmented in a rule-based models by applying a series of steps and rules for edge detection, region expansion, thresholding, geometrical fitting and many more. The rule- based approach rarely presents a dependable outcome but is a perfect starting point for advanced segmentation techniques [93], [94], [95].

The study in [94], for lung field identification from CXR, is based on the first derivative of the vertical and/or horizontal image profiles to achieve an accuracy of 95.2% and 96.0% respectively for both the left and right lung fields. The author claims that their proposed segmentation algorithm is useful for lung nodule and rib structure detection although the algorithm was applied on only forty (40) samples and was capable of segmenting the image one at a time in an average of 0.775 seconds. Hence it will be difficult to determine how the algorithm will perform on large datasets.

An algorithm presented in [95] is based upon thresholding for CXR segmentation. This algorithm set a threshold for segmenting the image pixels into black where it is lower than the threshold and white where it is higher than the set threshold. The procedure in this work was firstly to determine the left and right lungs frame, then partition the lung field from the entire image and further eliminate noise in the image before finally refining the lung edges. Although the author reported an effective result this was not validated. The study in [93] also utilized grey-level thresholding approach for lung segmentation and obtained 79% accuracy.

A portable CXR segmentation algorithm that is capable of identifying salient points which are inherently interpolated through Bezier curves was presented in [96]. This algorithm

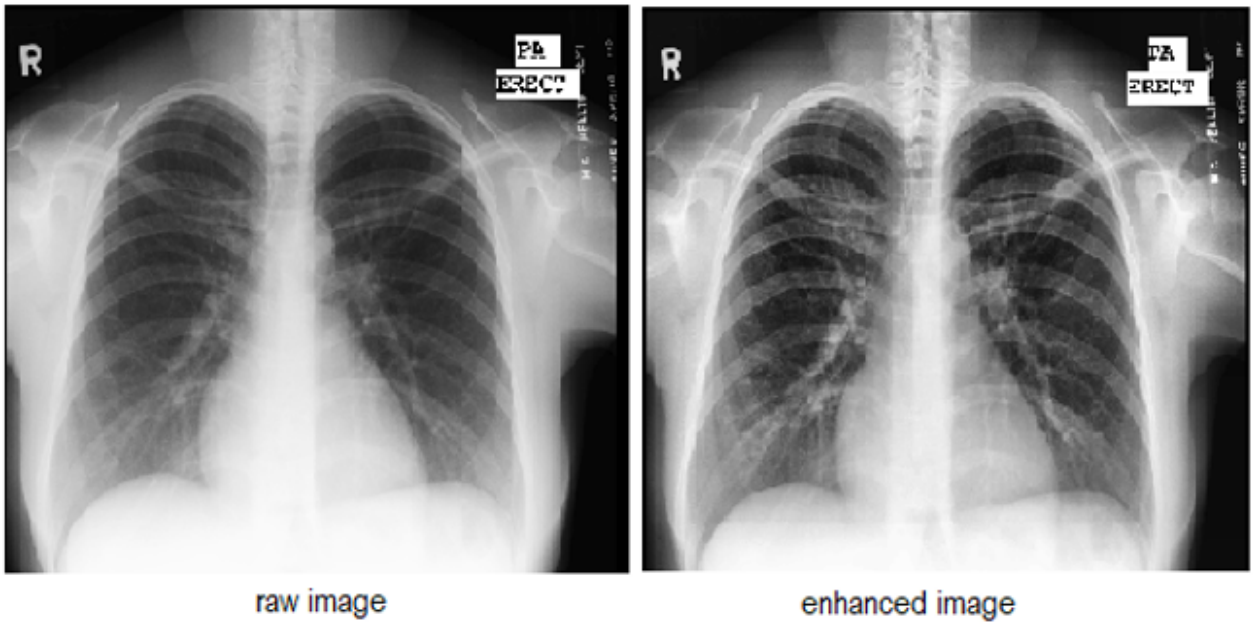


Fig. 11. Samples of the original and enhanced CXR images

is able to delineate lung field smoothly compared to graph cut methods, segment portable CXR irrespective of a patient's laying position which is an issue in some other methods, and is ensured not to exempt the heart overlapping region in the detection of lung abnormalities. The efficacy of this model with respect to specificity and sensitivity was measured on only twenty-four sample CXR.

Pixel based model: The pixel model is employed to classify every image pixel into its matching anatomical group such as diaphram, heart, and mediastinum [97], [98], [99]. The pixel based technique is more resilient [100] and has proven to have performed better than the rule-based technique [101].

Deformable models: Deformable models are more robust for anatomical boundary segmentation as a result of their flexibility, although impaired by local minima [102], [103]. Examples of deformable models include the Active Appearance Models (AAM) [104], [105], and the Active Shape Models (ASM) [106], [107] briefly discussed below.

Active Shape Model (ASM): introduced in [106] is a statistical model that consistently learns patterns of an object from a given image. Unlike the active contour model, ASM has the ability to capture natural irregularities in images of the same category although it is time-consuming and expensive because it requires multiple landmark samples to represent shape variations. Also, the result of segmentations arising from ASM application are available to local search fields throughout the landmarks.

In [108], a fusing shape and thresholding framework is presented for lung segmentation by effective integration of shape priors with intensity information which utilized an active shape model for the final fitting. Although the algorithm performed effectively in comparison with some other algorithms it cannot extract tissues overlapping with the heart and also, it is mostly dependent on manual extraction of the lung field by experts.

Active Appearance Model (AAM): [105] is another deformable model that is used widely for image analysis. AAM makes use of the statistical information to learn shape structure from set of training images [109]. AAM usually consists of a statistical model of the structure as well as grey-level appearance of the image that establishes credible examples. A deformable model is presented in [110] for segmentation of lung region in CXR using patients' shape statistics and is population based. The authors introduced a modified version of the Scale Invariant Feature (SIFT) descriptors to delineate image features around every pixel. The empirical result was reported to proffer a more accurate and robust segmentation of lung region. The segmentation model presented in [111] is geared with optimal features that is different to the approach in [106] to find optimal displacement for the landmarks with the use of KNN rather than linear Mahalanobis distance. This model is reported to have achieved a more significant performance than the pioneer active shape model [106] with respect to overlap errors.

Hybrid model: What is the best segmentation technique for CXR? It is difficult to adequately answer this query since every technique has their own essential strength and weakness. In other to answer this query, one may resort to trying different segmentation techniques and compare their accuracy. The strength and weakness of each technique can be found by comparing them and hence combining techniques that complement each other to form a hybrid technique. Hybrid segmentation technique is a combination of two or more techniques to take advantage of the strength of each technique and is combined not only to attain a better and powerful performance accuracy but also to become robust [101].

An automatic algorithm for lung region segmentation was

presented in [103] by comparing varieties of segmentation techniques. The authors found that maximum performance was achieved by combining the pixel method with rule-based segmentation method which resulted in a robust performance accuracy.

A comparison of active appearance model, pixel classification and active shape model was presented in [101] to segment lung region, clavicle, and heart in CXR. The three models were applied for lung segmentation and the pixel-based model showed the highest performance accuracy with no significant difference when evaluated against human observers. Active shape model performed better when applied to clavicle segmentation although the human performance was better. As for the heart segmentation, the different models performed worse than the human observer but the authors concluded that improved performance could be obtained by combining the models with majority voting.

Another hybrid method that used low-level features with prior shape was presented in [112] to solve the local minima challenge arising as a result of strong edge, and shading effects due to varying clavicle and rib cage. The hybrid model was optimized and validated upon several images from multiple scanners and various medical sites.

Three mask models were combined in [113] to segment the lung for the detection of TB and other lung abnormalities. After the segmentation task, the study proceeded to extract curvatures, texture, and shape features for training a Support Vector Machine (SVM) aimed at classifying the CXR images into normal and abnormal classes. The combination of different masks presented a super segmentation of the lung, according to the authors.

An unsupervised model [114] utilized oriented Gaussian derivatives filter for segmenting a lung on mobile and standard CXR. This model integrated thresholding and fuzzy C-Means for lung field refinement. Accuracy, precision, f-score, specificity, and sensitivity of the model were measured on the Japanese Society of Radiological Technology (JSRT) dataset. This is a solely unsupervised model that does not require any form of training to perform lung segmentation.

In [58] a robust nonrigid registration lung segmentation is presented. The model first searches for lung atlases that are similar to the patient's CXR, it then calculates the patient's definite lung model, and finally utilizes the graph cut approach to carry out detection of the lung region. The authors reported an accuracy of 95.4%, 94.1%, and 91.7% for the respective JSRT, MC, and India datasets.

In [115] a method for highlighting region of interest on CXR for lung segmentation with the use of K-Means method to analyze the lung features is presented. The patients physique, gender, and age can be described from the result of the model. Sets of fragmented images were used for their experiments; the image was preprocessed and grouped into 36 x 36, 24 x 24, and 12 x 12 classes to determine the image cluster error. The result reported standing at 28.06%, 30.61%, and 38.49% for the respective clustering errors.

Other conventional segmentation methods include: Gaussian Kernel Fuzzy Clustering [116], Markov Random Field Modeling [117], graph-cut based [118], pattern classification [119], Watershed [120], and Otsu thresholding [121].

Deep Learning based Segmentations

Some Deep Learning based Segmentations are discussed below:

Convolutional Neural Network (CNN) was combined with region proposal for segmentation and object detection purposes in [122]. This hybrid model referred to as Region-based Convolutional Network began with accepting input images, then the extraction of region of interest from the image and then employed CNN to compute features for every proposal after which the classification of each region was carried out. This model boasted of a 62.9% increased mean average precision which is about 50% better compared to the unified framework reported in [123].

An end-to-end segmentation technique for multi-class anatomical structures in CXR using fully connected neural network was described in [124]. The technique performed better when compared to the human observer for the heart and lung region segmentation. The combination of multiple classes ensured the model surpassed modern techniques in heart, clavicle and lung field segmentation.

In [125], the study exploited fully a convolutional network to train pixel-to-pixel for semantic segmentation. The model fine-tuned the novel classification approach GoogLeNet [39], [126], VGGNet [127], and AlexNet [34] to the segmentation task and transferred their learned representation into FCNN for the production of a thorough and precise image segmentation to a 20% improvement of PASCAL VOC.

The study in [128] discussed a system that automatically locates, and segments nuclei as well as cells in micro images. Structure Correcting Adversarial Network, otherwise known as SCAN, was proposed in [129] for segmenting lung and heart regions in CXR. SCAN integrates a critical network which learns the structures from highest to lowest so as to differentiate between the ground truth and the synthesized network. The critical network provided an adequate guide to the segmentation network via the adversarial procedures to achieving realistic and accurate results similar to human observers without the intervention of any pre-trained approach.

the Deep Learning approach was presented in [130] to achieve a fast and accurate lung segmentation that surpasses the complex techniques employed in modern studies. Varieties of dataset were compared for training their model and illustrated its significance. The authors drew some conclusions among which are that generic, semantic, and segmentation techniques are sufficient enough for lung segmentation task thereby emphasizing the fundamental need for a universally dependable lung segmentation tool to further research on serious pulmonary diseases.

A framework for automatic learning of x-ray scans from 3D computer tomography (CT) is proposed in [131] using task driven based generative adversarial networks. This framework trained a deep image-to-image model for multi-organ segmentation fulfilled from CT volumes followed by the development of a task driven GANs for obtaining simultaneous synthesis and parsing for invisible X-ray image. The model achieved 89% accuracy as well as a dice average of 86%.

In [132] a model to augment three CNN architectures with organs contour information for segmentation from CXR was proposed. The model was evaluated on the JSRT dataset and reported to perform better than previous models which utilized the same dataset with respect to segmentation of lung region, clavicle, and heart.

TABLE II
SUMMARY OF CXR SEGMENTATION

Reference	Evaluation Metric	Results
[119]	Accuracy	LDA = 70%, KNN = 70%, NN = 76%
[99]	Accuracy, Sensitivity, Specificity	94.8%, 90.7%, 97.2%
[93]	Accuracy	79%
[97]	Accuracy	94%
[94]	Accuracy	Left lung = 95.2%, Right lung = 96.0%
[111]	Overlap score	Left lung = 0.887%, Right lung = 0.929
[110]	Average overlay	ASM Sift = 87%, ASM Intensity = 92%
[96]	Sensitivity , Specificity	94.3%, 95.3%
[116]	Accuracy	97.8%
[108]	Overlap percentage, Contour distance in pixel	92,%, 94% , 3.71%, 2.46%
[58]	Accuracy	JSRT = 95.4%, MC = 94.1%, India = 91.7%
[114]	Accuracy, Sensitivity, Specificity, F-score	90%, 90%, 90%, 90%
[125]	Mean IU	62.2%
[122]	mean Average Precision (mAP)	62.4%
[129]	Intersection over union	Lungs = 94.7%, Heart = 86.6%
[124]	Dice coefficient, \\jaccard coefficient	Lungs = 97.4%, Colarbone = 92.9%, Heart = 93.7%
[115]	Clustering error	12x12 = 38.49%, 24x24 = 30.61%, 36x36 = 28.06%
[131]	Dice average,Accuracy	86%, 89%
[132]	Jaccard overlap coefficient	Lungs = 0.971%, Clavicle = 0.903%, Heart = 0.933%

Generally, the challenges of segmenting the lung region have received substantial attention. Diverse approaches have been employed resulting in good results. Although we can say that lung field segmentation issues have been solved, nevertheless, there is need to validate the reported results on large datasets to ascertain that various approaches are reliable enough to generate satisfactory results.Table II presents the summary of CXR segmentation.

X. AUTOMATIC TB DETECTION TECHNIQUES: FEATURES EXTRACTION AND CLASSIFICATION

There have been various of researches that employed different approaches towards the detection of pulmonary TB and more techniques are currently being studied to ensure a real-time diagnosis accuracy for total eradication of the deadly disease that is most prevalent in the developing nations. These approaches and processes are discussed as follows.

A CNN model was used in [133] to automatically diagnose TB from the Montgomery County CXR dataset. The CXR was fed into the model which was made up of four convolutional layers and a fully connected layer. The model was configured and trained using Stochastic Gradient Descent at 0.001 learning rate with an L2 regularizer to control overfitting. The model was trained with 75% of dataset and tested on the remaining 25% to record an accuracy rate of 87.1%

The study presented in [134] evaluated the effect of three image enhancement algorithms on the performance of Deep Learning method to address the low contrast to improve visualization of CXR. The authors performed enhancement on the Shenzen dataset and used the image to train EfficientNets and ResNet architectures for transfer learning. The experiments performed to detect TB from the enhanced images achieved 89.92% and 94.8% accuracy. The study concluded that the use of image enhancement algorithms to pre-process CXR images will thus allow the tested pre-trained network to learn a better model.

A method using wavelets transformation was proposed in [135] for the detection of TB by decomposing CXR scans into low and high frequency components. The study investigated the CXR of a patient that was confirmed to have TB, then assigned coordinates and subset the image based on the coordinates assigned to obtain about thirty (30) line profiles each of which was represented by 26 Daubechies coefficient. Statistical analysis is then performed using SPSS and hierarchical clustering to determine the features for identifying the presence of TB in CXR images.

Transfer learning of nine pre-trained CNN models was employed in [136] to detect TB from CXR and classify it into TB and non-TB. The CXR was first subjected to pre-processing, segmentation and augmentation before classification. Three experiments were carried out: segmentation of CXR using U-net models, then the classification of CXR with and without segmented lung fields. Out of the nine pre-trained models, ChexNet performed better to obtain an overall accuracy of 96.47% for the classification without segmentation, while DensNet201 obtained the best result in the category with segmentation. The study further employed visualization techniques to confirm that CNN learns dominantly from the segmented lung regions, resulting in higher detection accuracy. The study showed that the effect of performing image pre-processing and segmentation of lung fields can improve classification accuracy. The evaluation metrics examined in the study were accuracy, F1-score, precision, sensitivity, and specificity.

An experiment was done in [137] to compare the performance of three different methods in detecting TB in the lungs. The obtained results from the experiment showed the K-Nearest Neighbor (KNN) classifier had the maximum accuracy of 80%, Simple linear regression 79% accuracy, and the Sequential minimal optimizer obtained 75% accuracy.

The approach presented in [9] for TB detection began by employing graph cut approach to segment the lung fields, then computed shape features and textures using binary classifier (BC) and SVM to classify the CXR images into healthy and unhealthy classes. The method boasted of 78.3% and 84% accuracy for both datasets respectively.

Detecting the presence of TB early is presented in [138]. The homomorphic filter, median filter, and histogram equalizers were applied firstly to pre-process the input image, the pre-process image was then segmented using the active contour and finally obtained the classification output using the mean values. Although there was no accuracy result reported in the work the authors affirmed the impact and

contribution of CAD as an assistive tool to guide radiologists and doctors in reaching an accurate and timely diagnosis decision with respect to TB detection.

In [139], the authors present a TB identification system using a bright field microscope. The study further compared pixel, SVM, KNN, Linear Regression, Quadratic Discriminant, and Probabilistic Neural Networks classifiers which produced a result above 95% with respect to accuracy, sensitivity and specificity. The use of this system for TB screening may reduce the involvement of a technician according to the authors.

A comprehensive study that combined geometrical and advanced textural features for the detection of TB cavity is presented in [63]. The study first extracted cavity field from CXR scans with the use of histogram of oriented gradients, local binary pattern, and Gaussian model template matching at the coarse scale, then applied active contour snake-based and eigenvalues of the Hessian matrix to enhance the extracted field and lastly, utilized SVM classification to reduce false positives on a fine scale with gradient inverse coefficient of variation, Kullback-Leibler divergence, and circularity measure. The rate of true cavity achieved in the experiment is reported to perform better than existing methods.

In an attempt to limit the time taken by radiologists to examine patient CXR scans, [140] employed texture features to identify and classify thoracic x-ray scans as infected and non-infected. Several statistical features of the x-ray histogram were calculated and reduced with principal component analysis and then finally classified the image using minimum distance classifier. The result of 95.7% accuracy was presented.

Tuberculosis Index [TI] which focused on segmented lung texture features is presented in [141]. The method applied a decision tree to classify the CXR into abnormal and normal classes and achieved a 94.9% accuracy. Other metrics measured included: specificity, sensitivity, precision, and area under the curve (AUC).

In [142], a Bayesian classification method is presented for automatic diagnosis of TB cavities. Circularity measures and gradient inverse coefficient of variations were used to classify the diagnosed features and ascertain positive TB cavities.

[143] employed CNN architecture to detect TB from the Shenzen CXR dataset. The study began with preprocessing the CXR before propagating it through the CNN models which were composed of 6 convolutional layers with 2 fully connected layers. 70% of the dataset was utilized for model training while 30% was used to evaluate the model which achieved 87.8% accuracy. Data augmentation and some hyper-parameters employed in the study helped to control model overfitting.

In [144] granulometry and correlation technique was utilized to detect the miliary TB region. Template matching and correlation was performed by transforming the template and the image into a frequency domain. The authors strived to minimize false positives and negatives in the training set by correlating every image with 16 templates of varying sizes and respective thresholds. The technique achieved a 94% accuracy.

In [145] a hybrid of several methods including edge detection, rib image processing, boundary tracing and information extracting methods was used to establish a localized frame of reference in CXR. Binarization of the CXR was done for ease and quick extraction of the rib line followed by implementation of other algorithms such as seed growth and morphological openings to achieve about 85% accuracy for the automatic detection of focal opacities.

The method presented in [113] is aimed at TB diagnosis system from PA CXR. The system combined log gabor mask, statistical lung model mask, and intensity mask to segment the lung region followed with extraction of curvatures, textures, and shape features. These extracted features were then utilized in training an SVM classifier to differentiate the healthy images from the infected images. The system performed at 75% accuracy and about 83.12% AUC.

CXR interpretation using statistical approach for identifying lung TB is proposed in [14]. Wavelet transformation was first applied to each x-ray image region of interest to convert the image into four labelled subsets representing the vertical, diagonal, trend, and horizontal coefficients. The next procedure was to apply principal component analysis on 12 measures' texture in order to minimize dimensionality and finally use discriminant functions with probability ellipsoids to determine the misclassified probabilities. The system achieved a 94% classification accuracy.

A computer-aided detection system is presented in [146] based on multi-varied dissimilarities. The method estimated global dissimilarities that were classified by adopting linear discriminant classifier to codify the CXR image into normal and suspected TB class. The result of the global dissimilarity classification reached 0.81% AUC compared to 0.83% when combined with the local classification.

The algorithm proposed in [147] is aimed at enhancing CXR images to remove noise by segmenting the lung field and applying contrast enhancement that makes it easier to extract features associated with the identification of TB manifestations. There was no report of the system performance and validation accuracy. Hence, the reliability of the system could not be ascertained.

The CNN structure employed in [37] consists of seven convolutional layers and three fully connected layers to perform classification experiments on CXR for diagnosing TB. The authors compared three variety optimizers in their experiments and found the Adam optimizer performed better with a validation accuracy of 82% and loss of 0.40%. The result of their experiment was validated on the Shenzhen and Montgomery datasets.

The research work presented in [56] wa one of the first to employ deep learning techniques on medical images. The work was based on popular AlexNet architecture and transfer learning for screening the system performance on different datasets. The cross-dataset performance analysis carried out shows the system's accuracy of 78.3% and 86.9% on the Montgomery dataset with 84.1% accuracy and AUC of 0.90% on the Shenzhen dataset.

The study carried out in [36] presented proposals to improve feature extractors in detecting diseases using pre-trained CNN. This research is famous for combining multiple instance learning algorithms with pre-trained CNN, assessment of classifiers trained on features extracted, and the comparison of performance analysis of existing models for extracting features from the CXR dataset. The experimental result carried out for the Montgomery dataset, GoogLenet

achieved the highest accuracy while the best result for the Shenzhen dataset was obtained with the ResNet model.

An experiment for TB screening was presented in [38] using AlexNet and VGGNet architectures for the classification of CXR into positive and negative classes. The analysis carried out on the Montgomery and Shenzhen CXR datasets showed VGGNet outperformed Alexnet as a result of a deeper network of VGGNet. The performance accuracy of 80.4% was obtained for Alexnet while VGGNet reached 81.6% accuracy. The authors concluded that improved performance accuracy was possible by increasing the dataset size used for the experiment.

In [148], a framework is made up of a set of segments that must be executed in sequence to detect TB manifestation in CXR. The procedure begins with pre-processing based on Pyramid Histogram of Oriented Gradients (PHOG) and GIST descriptor, followed by a selection of features based on chi-square and then utilizes the SVM classifier in building the model to obtain the result. The authors further developed a toolkit in Matlab for training and predicting new CXR images.

In [149], a method is described to segment lung region on CXR based on Watershed. They incorporated some assumptions that the background area and the lung were greater than 30% and 20% of the image frame and a fixed sized window scan found the maximum grey levels and the average of the image pixels and then employed intensity threshold to classify the window midpoint as a suspected nodule.

A deep CNN is presented in [40] for pulmonary TB classification as either healthy or unhealthy. The images were classified using GoogLeNet and AlexNet architectures after applying pre-processing technique and performing ensemble on the image. The authors also employed the services of a professional radiologist to interpret the images in cases where both classifiers differed. AUC, specificity and sensitivity were used as metrics to measure the mode.

A CNN model that involved classifications of different manifestations of TB was presented in [41]. This work looked at unbalanced and less categorized CXR scans and incorporated cross-validation with sample shuffling in training their model. It was reported to have obtained 85.6% accuracy in classifying the various manifestations of TB using the Peruvian dataset.

CheXNeXt is an algorithm developed by the authors in [150] for the identification of 14 various pathologies in CXR. The algorithm, which employed CNN approach, was validated on the NIH dataset and compared the result with the interpretations of 9 professional radiologists. The result showed that CheXNeXt achieved an equivalent performance with the radiologists in 10 different pathologies, a best performance in 1 pathology and it underperformed in 3 pathologies. The algorithm took less than 2 minutes to identify the various pathologies while it took the radiologists about 240 minutes.

Another study that employed CNN in detecting various pathologies in CXR is presented in [151]. The study, which fused GIST and CNN descriptors, was evaluated on non-medical image dataset to obtain the system AUC,and the sensitivity and specificity of the system was between 0.87% - 0.94%.

The authors in [152] present an algorithm to assess the efficacy of deep learning architecture for abnormalities detection in CXR. NIH CXR 8 dataset [61] was used to train their algorithm that was based on CNN to identify certain CXR abnormalities such as infiltration, nodules, pneumonia, effusion, etc. in order to ascertain the standard of reference for the study. The service of a professional radiologist was employed to assess the presence of abnormalities in the CXR. The data analysis was done using SPSS and Excel. The result of the experiment showes AUC that ranged between 0.693-0.923% for the radiologist's interpretation and 0.837-0.929% for Deep Learning approach for all abnormalities. The study further concluded that although the use of Deep Learning was important in interpreting accuracy and could improve the level of detection of abnormalities but it was unlikely to entirely replace the input of a radiologist.

In [64], an automatic Deep Learning algorithm for diagnosis of active lung TB aimed at streamlining the process of screening and detection was presented. The algorithm was developed using CNN comprising about 27 layers. The CXR images were resized, then geometric and photometric operation was applied on them before passing on to the CNN architecture for training. The algorithm was evaluated on two public datasets and four custom datasets and compared with the interpretation of a certified radiologist and physicians. The performance of the algorithm showed significantly higher classification accuracy compared to the radiologist's and physicians performance on all the datasets.

Deep Learning was also employed in [153] for the detection of TB manifestation on CXR. Their CNN approach incorporates demographic information to train images containing 1000 normal and infected samples each using five different CNNs architecture as feature extractors. The experimental result in comparing the output of DCNN to I-CNN shows DCNN performed better but generally, the authors demonstrated that incorporating demographic information such as gender, age, weight, and height of the patients with CNN model resulted in an improved and effective diagnosis of TB.

A model that is composed of data acquisition and recognition stage for the detection of TB is presented in [154]. The images used in this model consisted of field of view digital images acquired from a microscope. The images were fed into the recognition stage that used a transfer learning model which learns from fine-tuned pre-trained weights. SVM was then employed to classify the images into infected and non-infected classes. The model attained a 95.05% accuracy.

The work presented in [70] aimed to detect about 20 abnormalities from CXR. The study employed three different public datasets and found that a hybrid of ensemble model with Deep Convolutional Neural Network (DCNN) outperformed utilizing the DCNN alone which did not show better performance across the various abnormalities reported in the study. The result of their ensemble model showed greater improvement when compared to the AlexNet, VGG-16, 19, ResNet-50,101 and 152 interms of accuracy, specificity, sensitivity and AUC.

In [155], a study that employed pre-trained CNN (GoogLeNet) was presented to automatically identify and classify CXR images as either normal, which means healthy, or abnormal with the presence of one or more instances of pleural effusion, consolidation, pneumothorax, cardiomegaly,

and pulmonary edema. Area Under the Curve (AUC) was the metric used to measure the performance of the model with respect to the interpretation of a board certified radiologist. The results obtained were 0.964% AUC, 91% sensitivity and specificity for the healthy classes while different percentages were obtained for the various abnormalities classes.

In [156], the authors participated in the 2019 ImageCLEF challenge and presented a Deep Learner model (LungNet) that focused on automatic analysis of TB from computer tomography CT scans. The CT scans employed were firstly decompressed and the slices were extracted having 512 images for the X and Y dimensions and between 40-250 images for the Z dimension. Filters were then introduced on the slices to eliminate the slices that dis not contain valuable information required for classifying the samples. The proposed LungNet along with ResNet-50 architecture were employed as the Deep Learner whose outputs are regarded as the preliminary results. The Deep Learner model was trained on 70% and 50% training sets and achieved AUC performance of 63% and 65% for the ImageCLEF CT report and severity scoring task respectively. Although these performances were not the best presented in the challenge the authors believed if they were subjected to advanced preprocessing techniques such as data augmentation and masking, these could provide a better performance.

The study in [157] was conducted to assess the detection accuracy of qXR which is a computer aided diagnosis software based on CNN. The authors utilized microbiologically established lung TB images as the standard for reference and made use of kappa coefficient along with confidence interval as the statistical tools to analyse the data and examine the inter-rater reliability of radiologists in detecting certain lung abnormalities. The study also used radiologist interpretation as a standard to validate the detection accuracy of the qXR in terms of generation ROC curves and calculating AUC. The qXR system achieved 0.81% AUC for detection of lung TB, with 71% sensitivity and 80% specificity.

Development of a Deep Learning model is presented in [158] for identifying abnormalities in CXR scans. The model was trained using large datasets up to about 2 million images using CNN. The model development began with down sampling, resizing, normalizing and augmenting the CXR prior to feeding them into the training network for features extractions and classification after which validation was performed to compute accuracy metrics. The model attained AUC of 0.92 for the general abnormalities detection and 0.96 for specific manifestations. The study confirmed that having large data to train a deep learning model will result in obtaining a higher detection rate.

The study presented in [62] evaluates and compares a commercial software CAD4TB for detecting TB manifestation on CXR. The software has different versions ranging from version 3, 4, 5, and the most recent version 6. Each of the versions uses different classifiers like SVM, KNN, and DCNN and is composed of varieties of features that are not common to all versions such as symmetry check, shape analysis, heat map and texture analysis.

The CAD4TB version 6 is said to outperform other versions obtaining 76% specificity and 90% sensitivity, while comparing version 6 to standard radiological inference attained a higher performance at specificity of 98% and 90% sensitivity.

The study presented in [159] employed a Bayesian Convolutional Neural Network (BCNN) for identifying uncertainty with low perceptibility associated with normal and abnormal TB manifestation on CXR. The fusion of Bayesian with CNN was aimed at overcoming the limitation of Softmax classifier in CNN models. The BCNN model was evaluated on both Shenzhen and Montgomery datasets to obtain 86.46% and 96.42% identification accuracy.

A model that combines transfer learning with CNN is presented in [160] to develop a system to detect and classify TB culture tests. The model was trained using small and imbalanced data samples that were obtained from the Tao-Yuan General Hospital in Taiwan. The study employed SMOTE to resolve the imbalance and introduced a two-stage classification method to improve recall of non-negative which eventually reached 98% recall. The preliminary result of this approach showed a promising output. A summary of automatic TB approaches is presented in Table III.

TABLE III
AUTOMATIC TUBERCULOSIS DETECTION

Reference	Computation Technique	Evaluation metric and Results
[133]	CNN	acc = 87.1%
[137]	Sequential Minimal Optimizer, Simple Linear Regression, KNN	eff = 75%, 79%, 80%
[9]	SVM, Binary Classifier	acc = 78.3%, 84%
[139]	Bright-field Microscope	acc = 95%, sen = 95% , spe = 95%\
[63]	Histogram of Oriented Gradients, SVM	acc = 82.8%
[140]	Textural Features, PCA, Minimum Distance Classifier	acc = 95.7%, far = 3.33% , frr = 6.67%
[141]	Decision Tree, Texture Features, User-guided Snake Algorithm	acc = 94.9%, pre = 92.9%, spe = 95.4%, sen = 92.8%, auc = 91%
[142]	Bayesian Classification, Gradient Inverse Coefficient of Variation, Circularity Measures	fp = 0.237%, acc = 82.35%
[143]	CNN	acc = 87.8%
[144]	Granulometry, Template Matching, Fourier Domain Correlation, Threshold, Rolling Ball	acc = 94%
[145]	Edge Detection, Boundary Tracing, Seed Growth, Morphological Openings	acc = 85%
[113]	Log Gabor Mask, Intensity Mask, Lung Model Mask, SVM	acc = 75%, 83.12%
[14]	Wavelet Coefficients, Maximum Column Energy, PCA	acc = 94%
[146]	Gaussian Derivatives, Multivalued Dissimilarities, Voting Rule, Central Moments	auc = 0.81%, 0.83%
[37]	CNN	acc = 97.4%
[56]	CNN, Transfer Learning	auc = 86.9%, 90%, acc = 78.3%, 84.1%

Figure 12 present the analysis of the techniques using a

Reference	Computation Technique	Evaluation metric and Results
[36]	CNN, Bag of features, Ensembles, SVM	acc =0.834%, auc = 0.912%
[38]	Pre-trained CNN	acc = 80.4%, 81.6%
[148]	Wavelet Denoising, Chi square, GIST, SVM	acc = 94%, 92%
[40]	Deep CNN, Ensemble	auc = 0.99%, sen = 97.3%, spe = 100%
[41]	CNN	acc = 85.6%
[150]	CNN	auc = 0.851%,
[151]	Pre-trained CNN,SVM,PCA, GIST	auc = 0.87%, sen = 0.94%, spe= 0.94%
[152]	CNN, SPSS	auc = 0.923%, 0.929%
[64]	CNN	auc = 0.977%, sen = 94%, spe = 91.1%
[70]	Ensemble, Deep CNN	acc = 90%, sen = 96%, spe = 96%, auc = 0.94%
[155]	CNN	auc = 0.964%, sen = 91%, spe = 91%
[153]	CNN,Deep CNN	auc = 0.9075%, sen = 0.815%, spe = 0.962%
[156]	CNN, Supervised Filtering, SVM Unsupervised filtering,	acc = 60%, auc = 65%
[157]	CNN,Kappa Coefficient, Confidence Interval, Younda Index	acc = 0.81%, sen = 71%, spe = 80%
[158]	CNN	auc = 0.96%
[134]	CNN, CLAHE, HEF	acc = 89.92%, 94.80%
[62]	CNN, SVM, KNN	spe = 98 %, sen = 90%

bar chart to show the most prominent classifiers used in CAD development for TB detection.

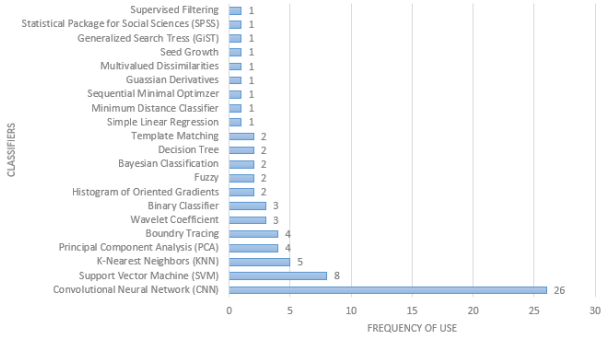


Fig. 12. bar chart showing classifiers for used TB detection.

XI. Discussion

In this study, we have presented both the conventional methods and state-of-the-art approaches of diagnosing TB. It is evident that automatic detection of TB has received tremendous attention over the years with diverse research papers being published with respect to abnormalities detections in CXR. It is difficult to judiciously compare and contrast the performances of the various algorithms employed for the task of identifying TB manifestations due to different datasets used that may have been collected under different conditions with different devices as well as severity of abnormalities. Some datasets are accompanied with radiological interpretation which serves as groundtruth to assist researchers to justify their findings. These algorithms have assisted medical experts and professional radiologist to substantiate their diagnosis opinions as well as providing quicker/faster screening of the patients CXR to shed excessive workloads.

It is shown from literature that the use of Deep Learning algorithms [56], [70], [155] has performed more accurately in classification tasks such as TB detection compared to other automatic methods like SVM, KNN, ANN, and Bayesian classifier. Also, earlier methods can only detect one or few diseases at the same time whereas Deep Learning algorithms are capable of detecting many diseases at the same time directly from CXR which agree with the interpretation of a skilled radiologist. Despite the varieties of potentials offered by Deep Learning algorithms for classifying images, there are still some challenges like the inability of SoftMax layer to detect uncertainties in classification and unbalanced distribution of samples with respect to limited datasets. Since it involves extracting distinct features from images, having limited datasets could often result in poor generalization of the network. Therefore, there is a need for testing a large amount of data samples that are collected from different regions and captured with different devices under different atmosphere so as to ensure versatility and better model generalization.

The implementation of Deep Learning algorithms in training big data has resulted in considerable progress, although it is time consuming as a result of features extraction which progresses from the shallow stage to deeper level via convolutions. Machine Learning algorithms like KNN, SVM, etc. also have potential for improvement and should be considered to attain high performance accuracy. Most of the algorithms have reported tremendous results in literature but are not transformed into CAD tools to be used in real life scenarios. Also, the systems should be evaluated against independent observers to assess if they can attain the same level as the human performance or surpass expert radiologists at some point. Deep Learning methods also require quite extensive hyper-parameter tuning to train a generalized model and resizing the input image to a much lower resolution may result in the loss of some important diagnostic features and could limit performance.

XII. Conclusion and Future Directions

A comprehensive survey of automatic methods for the diagnosis of pulmonary TB from CXR is presented in this paper. The emphasis is more on the CAD systems involving image pre-processing, features extractions, and features classifications. Image enhancement and segmentation are very important procedures in image pre-processing to boost the quality of images making them clearer for extracting distinct features in other to attain quality classifications. All these techniques are discussed and summarized in Table II and Table III. It is evident that the Deep Learning models discussed in this work have performed excellently in image classification tasks over the traditional Machine Learning

approaches like SVM, ANN and the like. Deep Learning models have become the main stream these days and promise more potential for the future for image classification tasks and specifically in medical imaging to detect various abnormalities to support radiological examinations.

However, there are more interesting future studies that may bring about additional improvement in the area of medical imaging analysis for TB diagnosis. Some of these are:

- Development of algorithms that could distinctly identify foreign objects like rings, buttons, pieces of bone, and coins that may be found on CXR images. These foreign objects could lead to misclassification and impede performance of detection systems. For instance round objects like buttons, bone pieces or rings can be mistaken for nodules.
- Abilities of CAD systems to simultaneously detect different diseases.
- A deeper network architecture, with sufficient and evenly distributed training CXR images, should be investigated in future works.
- Some diseases usually exhibit similar manifestations on the CXR and this could be a problem leading to misdiagnosis. Several studies are encouraged to look into adequately differentiating these diseases regardless of their common manifestations.
- A hybrid of traditional Machine Learning and Deep Learning models could complement each other in terms of strengths and weaknesses to develop a powerful CAD system.
- Studies that incorporate other information such as patients' level of cough, and body temperatures to CAD system could impact on the general performance accuracy.

The comparative investigation of different models helps in having a better understanding of a project and its implementation. Hence, this study serves as a guide in selecting the right method and techniques more suitable for early TB detection, which could help avert millions of deaths. It is our belief that this extensive review will proffer a better perception of TB detection domain while providing insights for future research.

XIII. Abbreviations

CAD: Computer-Aided Diagnosis; CNN: Convolutional Neural Network; TB: Tuberculosis; PCA: Principal Component Analysis; SVM: Support Vector Machine; AUC: Area Under Curve; ROC: Receiver Operating Characteristic; KNN: K-Nearest Neighbour; ANN: Artificial Neural Network; TP: True Positive; FP: False Positive, TN: True Negative; FN: False Negative; CT: Computed Tomography; DCNN: Deep Convolutional Neural Network; BCNN Bayesian Convolutional Neural Network; PHOG: Pyramid Histogram of Oriented Gradients; VGGNET: Visual Geometry Group Network; Xception: Extreme Inception; CXR: Chest X-Ray; PC: Phase Congruency; TI: Tuberculosis Index; POC: Point of Care; ILSVRC: ImageNet Large Scale Visual Recognition Competition; WHO: World Health Organizations, JSRT: Japanese Society of Radiological Technology; SPSS: Statistical Package for the Social Sciences; KIT: Korea Institute of Tuberculosis.

References

[1] W. H. Organization, *Global status report on alcohol and health 2018*. World Health Organization, 2019.

[2] M. Levin, M. Kafarou, and L. Coin, "Method of detecting active tuberculosis using minimal gene signature," Oct. 24 2019, uS Patent App. 16/326,828.

[3] N. A. Knechel, "Tuberculosis: pathophysiology, clinical features, and diagnosis," *Critical Care Nurse*, vol. 29, no. 2, pp. 34–43, 2009.

[4] M. G. Ametembun, "Iddf2019-abs-0238 follow-up of small bowel and dry type peritoneal tuberculosis treatment," 2019.

[5] K. Potts, M. Ragland, A. Hans, and M. Kearns, "Empyema necessitans caused by mycobacterium tuberculosis," in *D58. PLEURAL INFECTION CLINICAL CASE AND STUDIES*. American Thoracic Society, 2019, pp. A6842–A6842.

[6] A. T. Brennan, M. Maskew, B. A. Larson, I. Tsikhutsu, M. Bii, L. Vezi, M. P. Fox, W. D. Venter, P. Ehrenkranz, and S. Rosen, "Who is seeking antiretroviral treatment for hiv now? characteristics of patients presenting in kenya and south africa in 2017-2018," *Journal of the International AIDS Society*, vol. 22, no. 9, p. e25358, 2019.

[7] W. H. Organization, *Global tuberculosis report 2017*. World Health Organization, 2017. [Online]. Available: https://www.who.int/tb/publications/global_report/gtbr2017_main_text.pdf

[8] N. Nair, F. Wares, and S. Sahu, "Tuberculosis in the who south-east asia region," pp. 164–164, 2010.

[9] 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, p. 89, 2013.

[10] I. Livieris, T. Kotsilieris, I. Anagnostopoulos, and V. Tampakas, "Dtco: An ensemble ssl algorithm for x-rays classification," *Advances in Experimental Medicine and Biology; Springer: Berlin/Heidelberg, Germany*, 2018.

[11] C. Lange and T. Mori, "Advances in the diagnosis of tuberculosis," *Respirology*, vol. 15, no. 2, pp. 220–240, 2010.

[12] W. Y. N. Naing and Z. Z. Htike, "Advances in automatic tuberculosis detection in chest x-ray images," *Signal & Image Processing*, vol. 5, no. 6, p. 41, 2014.

[13] K. Nakamura, A. Ohmi, T. Kurihara, S. Suzuki, and M. Tadera, "Studies on the diagnostic value of 70 mm radiophotograms by mirror camera and the reading ability of physicians," *Kekkaku (Tuberculosis)*, vol. 45, no. 4, pp. 121–128, 1970.

[14] N. M. Noor, O. M. Rijal, A. Yunus, A. A. Mahayiddin, G. C. Peng, and S. Abu-Bakar, "A statistical interpretation of the chest radiograph for the detection of pulmonary tuberculosis," in *2010 IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES)*. IEEE, 2010, pp. 47–51.

[15] D. L. Cohn, R. J. OBrien, L. J. Geiter, F. Gordin, E. Hershfield, C. Horsburgh *et al.*, "Targeted tuberculin testing and treatment of latent tuberculosis infection," *MMWR Morb Mortal Wkly Rep*, vol. 49, no. 6, pp. 1–54, 2000.

[16] A. Zwerling, S. van den Hof, J. Scholten, F. Cobelens, D. Menzies, and M. Pai, "Interferon-gamma release assays for tuberculosis screening of healthcare workers: a systematic review," *Thorax*, vol. 67, no. 1, pp. 62–70, 2012.

[17] A. Lalvani and M. Pareek, "Interferon gamma release assays: principles and practice," *Enfermedades infecciosas y microbiologia clinica*, vol. 28, no. 4, pp. 245–252, 2010.

[18] P. Desikan, "Sputum smear microscopy in tuberculosis: is it still relevant?" *The Indian Journal of Medical Research*, vol. 137, no. 3, p. 442, 2013.

[19] P. C. Hopewell, M. Pai, D. Maher, M. Uplekar, and M. C. Raviglione, "International standards for tuberculosis care," *The Lancet infectious diseases*, vol. 6, no. 11, pp. 710–725, 2006.

[20] G. Theron, R. Venter, G. Calligaro, L. Smith, J. Limberis, R. Meldau, D. Chanda, A. Esmail, J. Peter, and K. Dheda, "Xpert mtb/rif results in patients with previous tuberculosis: can we distinguish true from false positive results?" *Clinical Infectious Diseases*, vol. 62, no. 8, pp. 995–1001, 2016.

[21] J. Palomino, H. Traore, K. Fissette, and F. Portaels, "Evaluation of mycobacteria growth indicator tube (mgit) for drug susceptibility testing of mycobacterium tuberculosis," *The International Journal of Tuberculosis and Lung Disease*, vol. 3, no. 4, pp. 344–348, 1999.

[22] E. Macondo, F. Ba, A. Gaye-Diallo, N. Touré-Kane, O. Kaï ré, A. Gueye-Ndiaye, C. Boye, and S. Mboup, "Rapid susceptibility testing of mycobacterium tuberculosis by the mycobacteria growth indicator tube (mgit ast sire)," *Clinical microbiology and infection*, vol. 6, no. 7, pp. 361–365, 2000.

[23] M. Palaci, S. Ueki, D. N. Sato, M. D. S. Telles, M. Curcio, and E. Silva, "Evaluation of mycobacteria growth indicator tube for

recovery and drug susceptibility testing of mycobacterium tuberculosis isolates from respiratory specimens." *Journal of Clinical Microbiology*, vol. 34, no. 3, pp. 762–764, 1996.

[24] D. I. Ling, A. A. Zwerling, and M. Pai, "Genotype mtbdr assays for the diagnosis of multidrug-resistant tuberculosis: a meta-analysis," *European Respiratory Journal*, vol. 32, no. 5, pp. 1165–1174, 2008.

[25] D. G. Storla, S. Yimer, and G. A. Bjune, "A systematic review of delay in the diagnosis and treatment of tuberculosis," *BMC public health*, vol. 8, no. 1, p. 15, 2008.

[26] W. H. Organization *et al.*, *Tuberculosis prevalence surveys: a handbook*. World Health Organization, 2011.

[27] J. Burrill, C. J. Williams, G. Bain, G. Conder, A. L. Hine, and R. R. Misra, "Tuberculosis: a radiologic review," *Radiographics*, vol. 27, no. 5, pp. 1255–1273, 2007.

[28] A. F. Gelb, C. Leffler, A. Brewin, V. Mascatello, and H. A. Lyons, "Miliary tuberculosis," *American Review of Respiratory Disease*, vol. 108, no. 6, pp. 1327–1333, 1973.

[29] A. C. Nachiappan, K. Rahbar, X. Shi, E. S. Guy, E. J. Mortani Barbosa Jr, G. S. Shroff, D. Ocazionez, A. E. Schlesinger, S. I. Katz, and M. M. Hammer, "Pulmonary tuberculosis: role of radiology in diagnosis and management," *Radiographics*, vol. 37, no. 1, pp. 52–72, 2017.

[30] J. H. Woodring, H. Vandiviere, A. Fried, M. Dillon, T. Williams, and I. Melvin, "Update: the radiographic features of pulmonary tuberculosis," *American Journal of Roentgenology*, vol. 146, no. 3, pp. 497–506, 1986.

[31] H. P. McAdams, J. Erasmus, and J. A. Winter, "Radiologic manifestations of pulmonary tuberculosis." *Radiologic Clinics of North America*, vol. 33, no. 4, pp. 655–678, 1995.

[32] E. Castañer, X. Gallardo, J. M. Mata, and L. Esteba, "Radiologic approach to the diagnosis of infectious pulmonary diseases in patients infected with the human immunodeficiency virus," *European Journal of Radiology*, vol. 51, no. 2, pp. 114–129, 2004.

[33] S. D. Greenberg, D. Frager, B. Suster, S. Walker, C. Stavropoulos, and A. Rothpearl, "Active pulmonary tuberculosis in patients with aids: spectrum of radiographic findings (including a normal appearance)." *Radiology*, vol. 193, no. 1, pp. 115–119, 1994.

[34] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing systems*, 2012, pp. 1097–1105.

[35] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *ArXiv Preprint ArXiv:1409.1556*, 2014.

[36] U. 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.

[37] R. Hooda, S. Sofat, S. Kaur, A. Mittal, and F. Meriaudeau, "Deep-learning: A potential method for tuberculosis detection using chest radiography," in *2017 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*. IEEE, 2017, pp. 497–502.

[38] A. Rohilla, R. Hooda, and A. Mittal, "Tb detection in chest radiograph using deep learning architecture," *ICETETSM-17*, pp. 136–147, 2017.

[39] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1–9.

[40] P. Lakhani and B. Sundaram, "Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks," *Radiology*, vol. 284, no. 2, pp. 574–582, 2017.

[41] C. Liu, Y. Cao, M. Alcantara, B. Liu, M. Brunette, J. Peinado, and W. Curioso, "Tx-cnn: Detecting tuberculosis in chest x-ray images using convolutional neural network," in *2017 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2017, pp. 2314–2318.

[42] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.

[43] M. Abdi and S. Nahavandi, "Multi-residual networks: Improving the speed and accuracy of residual networks," *ArXiv Preprint ArXiv:1609.05672*, 2016.

[44] X. Zhang, Z. Li, C. Change Loy, and D. Lin, "Polynet: A pursuit of structural diversity in very deep networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 718–726.

[45] A. Veit, M. J. Wilber, and S. Belongie, "Residual networks behave like ensembles of relatively shallow networks," in *Advances in Neural Information Processing Systems*, 2016, pp. 550–558.

[46] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "Squeezenet: Alexnet-level accuracy with 50x fewer parameters and¡ 0.5 mb model size," *ArXiv Preprint ArXiv:1602.07360*, 2016.

[47] F. Mamalet and C. Garcia, "Simplifying convnets for fast learning," in *International Conference on Artificial Neural Networks*. Springer, 2012, pp. 58–65.

[48] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 1251–1258.

[49] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic routing between capsules," in *Advances in Neural Information Processing Systems*, 2017, pp. 3856–3866.

[50] T. Karnkawinpong and Y. Limpiyakorn, "Chest x-ray analysis of tuberculosis by convolutional neural networks with affine transforms," in *Proceedings of the 2018 2nd International Conference on Computer Science and Artificial Intelligence*, 2018, pp. 90–93.

[51] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *European Conference on Computer Vision*. Springer, 2014, pp. 818–833.

[52] G. Larsson, M. Maire, and G. Shakhnarovich, "Fractalnet: Ultra-deep neural networks without residuals," *ArXiv Preprint ArXiv:1605.07648*, 2016.

[53] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 4700–4708.

[54] M. Lin, Q. Chen, and S. Yan, "Network in network," *ArXiv Preprint ArXiv:1312.4400*, 2013.

[55] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in Neural Information Processing Systems*, 2015, pp. 91–99.

[56] S. Hwang, H.-E. Kim, J. Jeong, and H.-J. Kim, "A novel approach for tuberculosis screening based on deep convolutional neural networks," in *Medical imaging 2016: computer-aided diagnosis*, vol. 9785. International Society for Optics and Photonics, 2016, p. 97852W.

[57] F. Pasa, V. Golkov, F. Pfeiffer, D. Cremers, and D. Pfeiffer, "Efficient deep network architectures for fast chest x-ray tuberculosis screening and visualization," *Scientific Reports*, vol. 9, no. 1, pp. 1–9, 2019.

[58] 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, 2013.

[59] J. Irvin, P. Rajpurkar, M. Ko, Y. Yu, S. Ciurea-Ilcus, C. Chute, H. Marklund, B. Haghgoo, R. Ball, K. Shpanskaya *et al.*, "Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, 2019, pp. 590–597.

[60] A. E. Johnson, T. J. Pollard, S. Berkowitz, N. R. Greenbaum, M. P. Lungren, C.-y. Deng, R. G. Mark, and S. Horng, "Mimic-cxr: A large publicly available database of labeled chest radiographs," *ArXiv Preprint ArXiv:1901.07042*, vol. 1, no. 2, 2019.

[61] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "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, pp. 2097–2106.

[62] K. Murphy, S. S. Habib, S. M. A. Zaidi, S. Khowaja, A. Khan, J. Melendez, E. T. Scholten, F. Amad, S. Schalekamp, M. Verhagen *et al.*, "Computer aided detection of tuberculosis on chest radiographs: An evaluation of the cad4tb v6 system," *Scientific Reports*, vol. 10, no. 1, pp. 1–11, 2020.

[63] T. Xu, I. Cheng, R. Long, and M. Mandal, "Novel coarse-to-fine dual scale technique for tuberculosis cavity detection in chest radiographs," *EURASIP Journal on Image and Video Processing*, vol. 2013, no. 1, p. 3, 2013.

[64] E. J. Hwang, S. Park, K.-N. Jin, J. I. Kim, S. Y. Choi, J. H. Lee, J. M. Goo, J. Aum, J.-J. Yim, C. M. Park *et al.*, "Development and validation of a deep learning–based automatic detection algorithm for active pulmonary tuberculosis on chest radiographs," *Clinical Infectious Diseases*, vol. 69, no. 5, pp. 739–747, 2019.

[65] W. H. Organization *et al.*, "Who report 2011: global tuberculosis control. geneva: Who," HTM/TB/2011.16, Tech. Rep., 2011.

[66] ——, "World health organization multidrug and extensively drug-resistant tb (m/xdr-tb): 2010 global report on surveillance andresponse. worldhealthorganization, geneva, switzerland," 2010.

[67] K. Weyer, S. Carai, and P. Nunn, "Viewpoint tb diagnostics: what does the world really need?" *Journal of Infectious Diseases*, vol. 204, no. suppl_4, pp. S1196–S1202, 2011.

[68] E. Keeler, M. D. Perkins, P. Small, C. Hanson, S. Reed, J. Cunningham, J. E. Aledort, L. Hillborne, M. E. Rafael, F. Girosi *et al.*, "Reducing the global burden of tuberculosis: the contribution of improved diagnostics," *Nature*, vol. 444, no. 1, pp. 49–57, 2006.

[69] D. L. Olson and D. Delen, *Advanced data mining techniques*. Springer Science & Business Media, 2008.

[70] M. T. Islam, M. A. Aowal, A. T. Minhaz, and K. Ashraf, "Abnormality detection and localization in chest x-rays using deep convolutional neural networks," *ArXiv Preprint ArXiv:1705.09850*, 2017.

[71] A. G. Lalkhen and A. McCluskey, "Clinical tests: sensitivity and specificity," *Continuing Education in Anaesthesia Critical Care & Pain*, vol. 8, no. 6, pp. 221–223, 2008.

[72] D. G. Altman and J. M. Bland, "Diagnostic tests. 1: Sensitivity and specificity." *BMJ: British Medical Journal*, vol. 308, no. 6943, p. 1552, 1994.

[73] S. A. Alasadi and W. S. Bhaya, "Review of data preprocessing techniques in data mining," *Journal of Engineering and Applied Sciences*, vol. 12, no. 16, pp. 4102–4107, 2017.

[74] M. Sonka, V. Hlavac, and R. Boyle, *Image processing, analysis, and machine vision*. Cengage Learning, 2014.

[75] S. García, J. Luengo, and F. Herrera, *Data preprocessing in data mining*. Springer, 2015, vol. 72.

[76] C. Qin, D. Yao, Y. Shi, and Z. Song, "Computer-aided detection in chest radiography based on artificial intelligence: a survey," *Biomedical Engineering Online*, vol. 17, no. 1, p. 113, 2018.

[77] L. Xiao, C. Li, Z. Wu, and T. Wang, "An enhancement method for x-ray image via fuzzy noise removal and homomorphic filtering," *Neurocomputing*, vol. 195, pp. 56–64, 2016.

[78] A. G. Webb, *Introduction to biomedical imaging*. John Wiley & Sons, 2017.

[79] B. Kwan and H. K. Kwan, "Improved lung nodule visualization on chest radiographs using digital filtering and contrast enhancement," *World Acad Sci Eng Technol*, vol. 110, pp. 590–3, 2011.

[80] A. Singh, S. Yadav, and N. Singh, "Contrast enhancement and brightness preservation using global-local image enhancement techniques," in *2016 Fourth International Conference On Parallel, Distributed and Grid Computing (Pdgc)*. IEEE, 2016, pp. 291–294.

[81] L. Wang, J. Lu, Y. Li, T. Yahagi, and T. Okamoto, "Noise removal for medical x-ray images in wavelet domain," *Electrical Engineering in Japan*, vol. 163, no. 3, pp. 37–46, 2008.

[82] C. Behrenbruch, S. Petroudi, S. Bond, J. Declerck, F. Leong, and J. Brady, "Image filtering techniques for medical image post-processing: an overview," *The British Journal of Radiology*, vol. 77, no. suppl_2, pp. S126–S132, 2004.

[83] X. Li, S. Luo, Q. Hu, J. Li, and D. Wang, "Rib suppression in chest radiographs for lung nodule enhancement," in *2015 IEEE International Conference on Information and Automation*. IEEE, 2015, pp. 50–55.

[84] K. Suzuki, H. Abe, H. MacMahon, and K. Doi, "Image-processing technique for suppressing ribs in chest radiographs by means of massive training artificial neural network (mtann)," *IEEE Transactions on Medical Imaging*, vol. 25, no. 4, pp. 406–416, 2006.

[85] M. Loog and B. van Ginneken, "Bony structure suppression in chest radiographs," in *International Workshop on Computer Vision Approaches to Medical Image Analysis*. Springer, 2006, pp. 166–177.

[86] P. Maduskar, L. Hogeweg, R. Philipsen, S. Schalekamp, and B. Van Ginneken, "Improved texture analysis for automatic detection of tuberculosis (tb) on chest radiographs with bone suppression images," in *Medical Imaging 2013: Computer-Aided Diagnosis*, vol. 8670. International Society for Optics and Photonics, 2013, p. 86700H.

[87] F. Li, R. Engelmann, L. Pesce, S. G. Armato, and H. MacMahon, "Improved detection of focal pneumonia by chest radiography with bone suppression imaging," *European Radiology*, vol. 22, no. 12, pp. 2729–2735, 2012.

[88] Y. Gordienko, P. Gang, J. Hui, W. Zeng, Y. Kochura, O. Alienin, O. Rokovyi, and S. Stirenko, "Deep learning with lung segmentation and bone shadow exclusion techniques for chest x-ray analysis of lung cancer," in *International Conference on Computer Science, Engineering and Education Applications*. Springer, 2018, pp. 638–647.

[89] H. X. Nguyen and T. T. Dang, "Ribs suppression in chest x-ray images by using ica method," in *5th International Conference on Biomedical Engineering in Vietnam*. Springer, 2015, pp. 194–197.

[90] W. Yang, Y. Chen, Y. Liu, L. Zhong, G. Qin, Z. Lu, Q. Feng, and W. Chen, "Cascade of multi-scale convolutional neural networks for bone suppression of chest radiographs in gradient domain," *Medical Image Analysis*, vol. 35, pp. 421–433, 2017.

[91] J. von Berg, C. Levrier, H. Carolus, S. Young, A. Saalbach, P. Laurent, and R. Florent, "Decomposing the bony thorax in x-ray images," in *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2016, pp. 1068–1071.

[92] B. Van Ginneken, B. T. H. Romeny, and M. A. Viergever, "Computer-aided diagnosis in chest radiography: a survey," *IEEE Transactions on Medical Imaging*, vol. 20, no. 12, pp. 1228–1241, 2001.

[93] S. G. Armato III, M. L. Giger, and H. MacMahon, "Automated lung segmentation in digitized posteroanterior chest radiographs," *Academic Radiology*, vol. 5, no. 4, pp. 245–255, 1998.

[94] L. Li, Y. Zheng, M. Kallergi, and R. A. Clark, "Improved method for automatic identification of lung regions on chest radiographs," *Academic Radiology*, vol. 8, no. 7, pp. 629–638, 2001.

[95] D. Cheng and M. Goldberg, "An algorithm for segmenting chest radiographs," in *Visual Communications and Image Processing'88: Third in a Series*, vol. 1001. International Society for Optics and Photonics, 1988, pp. 261–268.

[96] D. K. Iakovidis and G. Papamichalis, "Automatic segmentation of the lung fields in portable chest radiographs based on bézier interpolation of salient control points," in *2008 IEEE International Workshop on Imaging Systems and Techniques*. IEEE, 2008, pp. 82–87.

[97] B. Van Ginneken and B. M. ter Haar Romeny, "Automatic segmentation of lung fields in chest radiographs," *Medical Physics*, vol. 27, no. 10, pp. 2445–2455, 2000.

[98] A. Mittal, R. Hooda, and S. Sofat, "Lung field segmentation in chest radiographs: a historical review, current status, and expectations from deep learning," *IET Image Processing*, vol. 11, no. 11, pp. 937–952, 2017.

[99] N. F. Vittitoe, R. Vargas-Voracek, and C. E. Floyd Jr, "Markov random field modeling in posteroanterior chest radiograph segmentation," *Medical Physics*, vol. 26, no. 8, pp. 1670–1677, 1999.

[100] Z. Yue, A. Goshtasby, and L. V. Ackerman, "Automatic detection of rib borders in chest radiographs," *IEEE Transactions on Medical Imaging*, vol. 14, no. 3, pp. 525–536, 1995.

[101] B. Van Ginneken, M. B. Stegmann, and M. Loog, "Segmentation of anatomical structures in chest radiographs using supervised methods: a comparative study on a public database," *Medical Image Analysis*, vol. 10, no. 1, pp. 19–40, 2006.

[102] T. McInerney and D. Terzopoulos, "Deformable models in medical image analysis: a survey," *Medical Image Analysis*, vol. 1, no. 2, pp. 91–108, 1996.

[103] 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.

[104] I. Matthews and S. Baker, "Active appearance models revisited," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 135–164, 2004.

[105] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 681–685, 2001.

[106] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models-their training and application," *Computer Vision and Image Understanding*, vol. 61, no. 1, pp. 38–59, 1995.

[107] N. N. K. Chi, "Active shape models their training and application," Ph.D. dissertation, International University HCMC, Vietnam, 2011.

[108] A. Dawoud, "Lung segmentation in chest radiographs by fusing shape information in iterative thresholding," *IET Computer Vision*, vol. 5, no. 3, pp. 185–190, 2011.

[109] N. Babaii Rizvandi, A. Pizurica, and W. Philips, "Active appearance model (aam)-from theory to implementation," in *3rd International Conference on Computer Vision Theory and Applications*. INSTICC, 2008, pp. 539–542.

[110] Y. Shi, F. Qi, Z. Xue, L. Chen, K. Ito, H. Matsuo, and D. Shen, "Segmenting lung fields in serial chest radiographs using both population-based and patient-specific shape statistics," *IEEE Transactions on Medical Imaging*, vol. 27, no. 4, pp. 481–494, 2008.

[111] B. Van Ginneken, A. F. Frangi, J. J. Staal, B. M. ter Haar Romeny, and M. A. Viergever, "Active shape model segmentation with optimal features," *IEEE Transactions on Medical Imaging*, vol. 21, no. 8, pp. 924–933, 2002.

[112] P. Annangi, S. Thiruvenkadam, A. Raja, H. Xu, X. Sun, and L. Mao, "A region based active contour method for x-ray lung segmentation using prior shape and low level features," in *2010 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*. IEEE, 2010, pp. 892–895.

[113] S. Jaeger, A. Karargyris, S. Antani, and G. Thoma, "Detecting tuberculosis in radiographs using combined lung masks," in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2012, pp. 4978–4981.

[114] W. S. H. M. W. Ahmad, W. M. D. W. Zaki, and M. F. A. Fauzi, "Lung segmentation on standard and mobile chest radiographs using

oriented gaussian derivatives filter," *Biomedical Engineering Online*, vol. 14, no. 1, p. 20, 2015.
[115] N. Y. Ilyasova, A. Shirokanev, and N. Demin, "Segmentation of lung images using textural features," in *Journal of Physics: Conference Series*, vol. 1438. IOP Publishing, 2020, p. 012015.
[116] Z. Shi, P. Zhou, L. He, T. Nakamura, Q. Yao, and H. Itoh, "Lung segmentation in chest radiographs by means of gaussian kernel-based fcm with spatial constraints," in *2009 Sixth International Conference on Fuzzy Systems and Knowledge Discovery*, vol. 3. IEEE, 2009, pp. 428–432.
[117] N. F. Vittitoe, R. Vargas-Voracek, and C. E. Floyd Jr, "Identification of lung regions in chest radiographs using markov random field modeling," *Medical Physics*, vol. 25, no. 6, pp. 976–985, 1998.
[118] S. Candemir, S. Jaeger, K. Palaniappan, S. Antani, and G. Thoma, "Graph-cut based automatic lung boundary detection in chest radiographs," in *IEEE Healthcare Technology Conference: Translational Engineering in Health & Medicine*, 2012, pp. 31–34.
[119] M. F. McNitt-Gray, J. W. Sayre, H. Huang, and M. Razavi, "Pattern classification approach to segmentation of chest radiographs," in *Medical Imaging 1993: Image Processing*, vol. 1898. International Society for Optics and Photonics, 1993, pp. 160–170.
[120] H. Ng, S. Ong, K. Foong, P. Goh, and W. Nowinski, "Medical image segmentation using k-means clustering and improved watershed algorithm," in *2006 IEEE Southwest Symposium on Image Analysis and Interpretation*. IEEE, 2006, pp. 61–65.
[121] C. H. Bindu and K. S. Prasad, "An efficient medical image segmentation using conventional otsu method," *International Journal of Advanced Science and Technology*, vol. 38, no. 1, pp. 67–74, 2012.
[122] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Region-based convolutional networks for accurate object detection and segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 1, pp. 142–158, 2015.
[123] C. Gu, J. J. Lim, P. Arbeláez, and J. Malik, "Recognition using regions," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2009, pp. 1030–1037.
[124] A. A. Novikov, D. Lenis, D. Major, J. Hladuvka, M. Wimmer, and K. Bühler, "Fully convolutional architectures for multiclass segmentation in chest radiographs," *IEEE Transactions on Medical Imaging*, vol. 37, no. 8, pp. 1865–1876, 2018.
[125] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3431–3440.
[126] M. Günel, "Googlenet," 2016.
[127] H. Jun, L. Shuai, S. Jinming, L. Yue, W. Jingwei, and J. Peng, "Facial expression recognition based on vggnet convolutional neural network," in *2018 Chinese Automation Congress (CAC)*. IEEE, 2018, pp. 4146–4151.
[128] F. Ning, D. Delhomme, Y. LeCun, F. Piano, L. Bottou, and P. E. Barbano, "Toward automatic phenotyping of developing embryos from videos," *IEEE Transactions on Image Processing*, vol. 14, no. 9, pp. 1360–1371, 2005.
[129] W. Dai, N. Dong, Z. Wang, X. Liang, H. Zhang, and E. P. Xing, "Scan: Structure correcting adversarial network for organ segmentation in chest x-rays," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Springer, 2018, pp. 263–273.
[130] J. Hofmanninger, F. Prayer, J. Pan, S. Rohrich, H. Prosch, and G. Langs, "Automatic lung segmentation in routine imaging is a data diversity problem, not a methodology problem," *ArXiv Preprint ArXiv:2001.11767*, 2020.
[131] Y. Zhang, S. Miao, T. Mansi, and R. Liao, "Unsupervised x-ray image segmentation with task driven generative adversarial networks," *Medical Image Analysis*, vol. 62, p. 101664, 2020.
[132] M. Kholiavchenko, I. Sirazitdinov, K. Kubrak, R. Badrutdinova, R. Kuleev, Y. Yuan, T. Vrtovec, and B. Ibragimov, "Contour-aware multi-label chest x-ray organ segmentation," *International Journal of Computer Assisted Radiology and Surgery*, vol. 15, no. 3, pp. 425–436, 2020.
[133] M. Oloko-Oba and S. Viriri, "Diagnosing tuberculosis using deep convolutional neural network," in *International Conference on Image and Signal Processing*. Springer, 2020, pp. 151–161.
[134] K. Munadi, K. Muchtar, N. Maulina, and B. Pradhan, "Image enhancement for tuberculosis detection using deep learning," *IEEE Access*, vol. 8, pp. 217 897–217 907, 2020.
[135] N. M. Noor, O. Rijal, and C. Y. Fah, "Wavelet as features for tuberculosis (mtb) using standard x-ray film images," in *6th International Conference on Signal Processing, 2002.*, vol. 2. IEEE, 2002, pp. 1138–1141.
[136] T. Rahman, A. Khandakar, M. A. Kadir, K. R. Islam, K. F. Islam, R. Mazhar, T. Hamid, M. T. Islam, S. Kashem, Z. B. Mahbub *et al.*, "Reliable tuberculosis detection using chest x-ray with deep learning, segmentation and visualization," *IEEE Access*, vol. 8, pp. 191 586–191 601, 2020.
[137] B. Antony and N. B. PK, "Lung tuberculosis detection using x-ray images," *International Journal of Applied Engineering Research*, vol. 12, no. 24, pp. 15 196–15 201, 2017.
[138] I. Gabriella *et al.*, "Early detection of tuberculosis using chest x-ray (cxr) with computer-aided diagnosis," in *2018 2nd International Conference on Biomedical Engineering (IBIOMED)*. IEEE, 2018, pp. 76–79.
[139] R. Khutlang, S. Krishnan, R. Dendere, A. Whitelaw, K. Veropoulos, G. Learmonth, and T. S. Douglas, "Classification ofmycobacterium tuberculosisin images of zn-stained sputum smears," *IEEE Transactions on Information Technology in biomedicine*, vol. 14, no. 4, pp. 949–957, 2009.
[140] R. N. Rohmah, A. Susanto, and I. Soesanti, "Lung tuberculosis identification based on statistical feature of thoracic x-ray," in *2013 International Conference on QiR*. IEEE, 2013, pp. 19–26.
[141] 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.
[142] R. Shen, I. Cheng, and A. Basu, "A hybrid knowledge-guided detection technique for screening of infectious pulmonary tuberculosis from chest radiographs," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 11, pp. 2646–2656, 2010.
[143] M. Oloko-Oba and S. Viriri, "Tuberculosis abnormality detection in chest x-rays: A deep learning approach," in *International Conference on Computer Vision and Graphics*. Springer, 2020, pp. 121–132.
[144] A. Koeslag and G. de Jager, "Computer aided diagnosis of miliary tuberculosis," *Proceedings of the Pattern Recognition Association of South Africa*, 2001.
[145] Y.-L. Song and Y. Yang, "Localization algorithm and implementation for focal of pulmonary tuberculosis chest image," in *2010 International Conference on Machine Vision and Human-machine Interface*. IEEE, 2010, pp. 361–364.
[146] Y. Arzhaeva, L. Hogeweg, P. A. de Jong, M. A. Viergever, and B. van Ginneken, "Global and local multi-valued dissimilarity-based classification: Application to computer-aided detection of tuberculosis," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2009, pp. 724–731.
[147] S. Hariharan, A. Ray, and M. Ghosh, "An algorithm for the enhancement of chest x-ray images of tuberculosis patients," in *Proceedings of IEEE International Conference on Industrial Technology 2000 (IEEE Cat. No. 00TH8482)*, vol. 2. IEEE, 2000, pp. 107–112.
[148] 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, 2014.
[149] K. Le, "Automated detection of early lung cancer and tuberculosis based on x-ray image analysis," in *Proc. WSEAS International Conference on Signal, Speech and Image Processing*, 2006, pp. 1–6.
[150] P. Rajpurkar, J. Irvin, R. L. Ball, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. P. Langlotz *et al.*, "Deep learning for chest radiograph diagnosis: A retrospective comparison of the chexnext algorithm to practicing radiologists," *PLoS Medicine*, vol. 15, no. 11, p. e1002686, 2018.
[151] Y. Bar, I. Diamant, L. Wolf, S. Lieberman, E. Konen, and H. Greenspan, "Chest pathology detection using deep learning with non-medical training," in *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2015, pp. 294–297.
[152] R. Singh, M. K. Kalra, C. Nitiwarangkul, J. A. Patti, F. Homayounieh, A. Padole, P. Rao, P. Putha, V. V. Muse, A. Sharma *et al.*, "Deep learning in chest radiography: detection of findings and presence of change," *PloS One*, vol. 13, no. 10, 2018.
[153] S.-J. Heo, Y. Kim, S. Yun, S.-S. Lim, J. Kim, C.-M. Nam, E.-C. Park, I. Jung, and J.-H. Yoon, "Deep learning algorithms with demographic information help to detect tuberculosis in chest radiographs in annual workers health examination data," *International Journal of Environmental Research and Public Health*, vol. 16, no. 2, p. 250, 2019.
[154] R. D. J. Samuel and B. R. Kanna, "Tuberculosis (tb) detection system using deep neural networks," *Neural Computing and Applications*, vol. 31, no. 5, pp. 1533–1545, 2019.
[155] M. Cicero, A. Bilbily, E. Colak, T. Dowdell, B. Gray, K. Perampaladas, and J. Barfett, "Training and validating a deep convolutional neural network for computer-aided detection and classification of abnormalities on frontal chest radiographs," *Investigative Radiology*, vol. 52, no. 5, pp. 281–287, 2017.
[156] A. Hamadi, N. B. Cheikh, Y. Zouatine, S. M. B. Menad, and M. R. Djebbara, "Imageclef 2019: Deep learning for tuberculosis ct image analysis," *CLEF2019 Working Notes*, vol. 2380, pp. 9–12, 2019.

[157] M. Nash, R. Kadavigere, J. Andrade, C. A. Sukumar, K. Chawla, V. P. Shenoy, T. Pande, S. Huddart, M. Pai, and K. Saravu, "Deep learning, computer-aided radiography reading for tuberculosis: a diagnostic accuracy study from a tertiary hospital in india," *Scientific Reports*, vol. 10, no. 1, pp. 1–10, 2020.

[158] P. Putha, M. Tadepalli, B. Reddy, T. Raj, J. A. Chiramal, S. Govil, N. Sinha, M. KS, S. Reddivari, A. Jagirdar *et al.*, "Can artificial intelligence reliably report chest x-rays?: Radiologist validation of an algorithm trained on 2.3 million x-rays," *ArXiv Preprint ArXiv:1807.07455*, 2018.

[159] Z. U. Abideen, M. Ghafoor, K. Munir, M. Saqib, A. Ullah, T. Zia, S. A. Tariq, G. Ahmed, and A. Zahra, "Uncertainty assisted robust tuberculosis identification with bayesian convolutional neural networks," *IEEE Access*, vol. 8, pp. 22 812–22 825, 2020.

[160] R.-I. Chang, Y.-H. Chiu, and J.-W. Lin, "Two-stage classification of tuberculosis culture diagnosis using convolutional neural network with transfer learning," *The Journal of Supercomputing*, pp. 1–16, 2020.

2.1.2 Conclusion

This comprehensive review presents the analysis of automatic methods in TB diagnosis from CXR. It is evident from the literature that DL models have performed excellently in image classification tasks over traditional ML models like Support Vector Machines (SVM), Artificial Neural Networks (ANN), etc. DL models have become mainstream and promise more potential for image classification tasks, specifically in medical imaging, to detect various pulmonary abnormalities and support radiological examinations. The review concludes with highlights of some interesting areas that may improve medical imaging analysis for TB diagnosis.

2.2 A Systematic Review of Deep Learning Techniques for Tuberculosis Detection from Chest Radiograph

2.2.1 Introduction

This section extends the review paper in section 2.1. The paper presents the results of a systematic review based on the preferred reporting items for systematic reviews and meta-analyses (PRISMA) procedures that investigate the most recent DL techniques for developing a CAD system for rigorously screening pulmonary TB. The systematic review was conducted using an extensive selection of scientific databases as reference sources that granted access to distinctive articles in the field. Four scientific databases were searched to retrieve related articles. Inclusion and exclusion criteria were defined and applied to each article to determine those included in the study. The systematic review paper was published in Frontiers in Medicine [1]

[1] **Mustapha Oloko-Oba and Serestina Viriri**. "A Systematic Review of Deep Learning Techniques for Tuberculosis Detection From Chest Radiograph". *Frontiers in Medicine*, Vol. 9, **(2022)**. DOI: `https://doi.org/10.3389/fmed.2022.830515`

frontiers in Medicine

SYSTEMATIC REVIEW
published: 10 March 2022
doi: 10.3389/fmed.2022.830515

A Systematic Review of Deep Learning Techniques for Tuberculosis Detection From Chest Radiograph

*Mustapha Oloko-Oba and Serestina Viriri**

Computer Science Discipline, School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Durban, South Africa

The high mortality rate in Tuberculosis (TB) burden regions has increased significantly in the last decades. Despite the possibility of treatment for TB, high burden regions still suffer inadequate screening tools, which result in diagnostic delay and misdiagnosis. These challenges have led to the development of Computer-Aided Diagnostic (CAD) system to detect TB automatically. There are several ways of screening for TB, but Chest X-Ray (CXR) is more prominent and recommended due to its high sensitivity in detecting lung abnormalities. This paper presents the results of a systematic review based on PRISMA procedures that investigate state-of-the-art Deep Learning techniques for screening pulmonary abnormalities related to TB. The systematic review was conducted using an extensive selection of scientific databases as reference sources that grant access to distinctive articles in the field. Four scientific databases were searched to retrieve related articles. Inclusion and exclusion criteria were defined and applied to each article to determine those included in the study. Out of the 489 articles retrieved, 62 were included. Based on the findings in this review, we conclude that CAD systems are promising in tackling the challenges of the TB epidemic and made recommendations for improvement in future studies.

Keywords: tuberculosis, chest radiograph, computer-aided diagnosis, deep learning, systematic review

OPEN ACCESS

Edited by:
Emmanuel Twumasi Osei, University of British Columbia Okanagan, Canada

Reviewed by:
Anjali Agrawal, Teleradiology Solutions, India
Muhammad Usman Akram, National University of Sciences and Technology (NUST), Pakistan

***Correspondence:**
Serestina Viriri
viriris@ukzn.ac.za

Specialty section:
This article was submitted to Pulmonary Medicine, a section of the journal Frontiers in Medicine

Received: *07 December 2021*
Accepted: *14 February 2022*
Published: *10 March 2022*

Citation:
Oloko-Oba M and Viriri S (2022) A Systematic Review of Deep Learning Techniques for Tuberculosis Detection From Chest Radiograph. Front. Med. 9:830515. doi: 10.3389/fmed.2022.830515

INTRODUCTION

Tuberculosis (TB) is ranked among the leading causes of death. About 10 million persons fell ill globally from TB infections in 2019 (1). TB is triggered by the Mycobacterium bacteria that usually affect the lungs (pulmonary) but sometimes affect other parts of the body (extrapulmonary) (2). Many TB patients lose their lives yearly due to diagnostic delay, misdiagnosis, and lack of appropriate treatments (3, 4). Although TB is a global challenge, the mortality rate is more prevalent in low and middle-income nations (5).

TB is certainly treatable if diagnosed early for appropriate treatment. Early diagnosis is essential for successful treatment, preventing further spread, and significantly reducing the mortality rate in line with the World Health Organization (WHO) End TB Strategy (1). The gold standard for TB screening is Sputum culture. However, posterior-anterior chest radiographs (CXR) are an effective technique with low-cost and moderately low radiation doses for screening lung abnormalities to achieve prompt results (6). CXR has been adequately employed in developed countries to analyze individuals exhibiting active TB symptoms. At the same time, its application is limited in developing countries where TB is most prevalent (7, 8). High TB burden regions lack the skilled and radiological expertise required to interpret CXR images adequately (9, 10).

In the last decades, several efforts have been made using Artificial Intelligence (AI) to develop a Computer-Aided Detection (CAD) system to advance automatic object/image recognition tasks and overcome the challenges of a skilled workforce. Machine Learning (ML) and Deep Learning (DL) are the predominant AI techniques employed to develop CAD systems for analyzing CXR images. Both techniques have had a significant impact, but the DL approach, such as Convolutional Neural Network (CNN), has become more prominent for analyzing different pulmonary abnormalities in the medical domain, most importantly in diagnosing TB. The application of an efficient classification tool is vital for improving the quality of diagnosis while reducing the time taken to analyze a large volume of CXRs (11). This endeavor is to achieve the global decline in TB incidence to about 5% annually compared to the current 2% yearly as part of the World Health Organization strategy to end TB (1).

The contribution of this systematic review is to present an extensive summary of the various state-of-the-art CAD system proposed in the literature for the classification of TB. Ultimately, only the CAD system developed using Deep Learning models is considered in this study detailing the diagnostic accuracy between 2017 and 2021. The rest of the paper is structured as follows: section Methodology presents the study methodology. The results are presented in section Results. Section Discussion presents the discussion, while the conclusion is expressed in section Conclusion and Recommendations.

METHODOLOGY

This systematic review aims to establish various CAD systems related to Tuberculosis diagnosis from CXR using DL techniques. The study followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) procedures (12) to identify the standards for inclusion and exclusion, as shown in **Figure 1**. These standards were formulated based on the present study objectives and the research questions. All articles that satisfied the following conditions were included and excluded if otherwise:

- Articles that considered only pulmonary tuberculosis disease.
- Employed at least one deep learning technique as a classifier
- CXR is the only medical imagine for screening tuberculosis.
- Articles published between January 2017 and September 2021.
- Articles are entirely written in English.
- Articles must be full text. All others, such as abstract, preprints are excluded.

Search Strategy

The search strategy was developed to identify relevant published articles in Scopus (https://www.scopus.com/), IEEEXplore (https://ieeexplore.ieee.org/), Web of Science (www.webofscience.com), and PubMed (https://www.ncbi.nlm.nih.gov/pubmed/). Searches were performed across the four databases using the following search keywords: "tuberculosis," "chest x-ray," "classification," "artificial intelligence," "computer-aided diagnosis," "deep learning." These keywords were used to

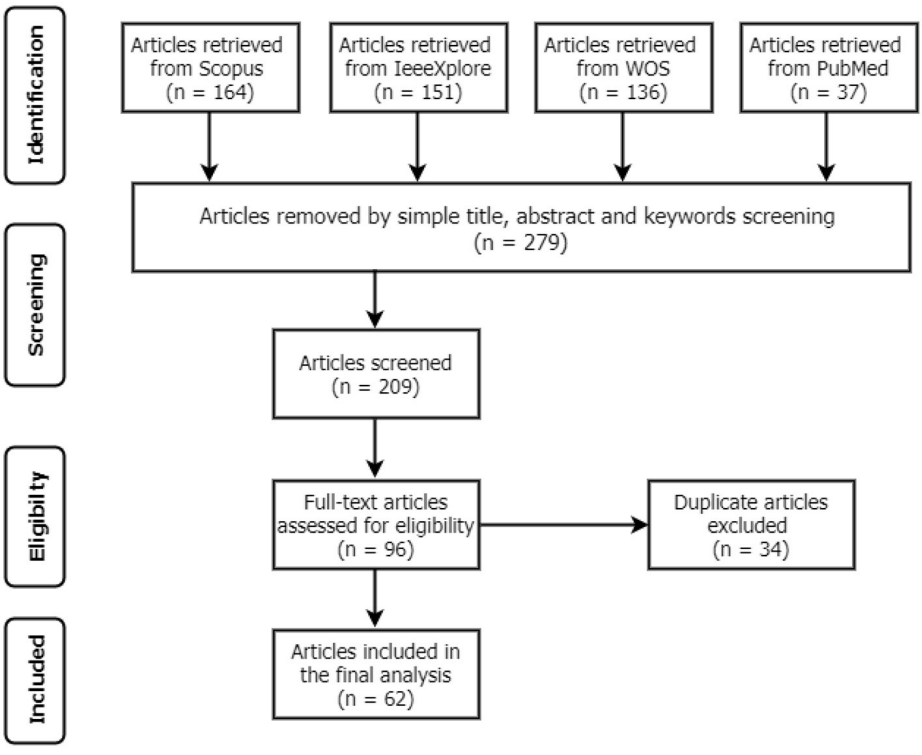


FIGURE 1 | The PRISMA structure for the study selection process.

TABLE 1 | Construction of search keywords.

Databases	Search keywords
Scopus	(TITLE-ABS-KEY ("Tuberculosis" AND "Chest X-Ray") AND TITLE-ABS-KEY ("Deep learning" OR "Machine learning" OR "Artificial Intelligence" OR "Classification")
IEEEXplore	(Tuberculosis) AND ("Chest X-Ray") AND ("Deep learning" OR "Machine learning" OR "classification" OR "artificial intelligence") AND ("CAD" OR "computer-aided detection")
Web of Science	((Tuberculosis AND Chest x-ray) AND ("Machine learning" OR "Deep learning" OR "Artificial intelligence") AND ("classification" OR "classify") AND ("computer-aided diagnosis" OR "CAD"))
PubMed	("Tuberculosis") AND ("chest x-ray") AND ("deep learning" OR "convolutional neural network") AND ("classify" or "classification") OR ("computer-aided diagnosis" OR "computer-aided detection" OR "CAD")

form Boolean search strings according to searching standards on different databases. The configuration of keywords used to retrieve all relevant articles from each database search engine is presented in **Table 1**.

Study Selection

A total of 488 articles were retrieved from the initial search on all the databases. The next step then scans the article topics and keywords to identify highly related papers from the irrelevant ones. This step is followed by the overview reading of the abstract, methods, and Conclusion to further screen for relevant papers for full-text reading and better understanding. Thus, 209 articles were considered for the title and abstract screening, out of which 96 papers were found suitable for full-text reading, and a total of 62 were finally included in the analysis after removing 34 duplicates. The detailed structure for the study selection is presented in **Figures 1**, **2** shows the numbers of articles included per year and databases. It is necessary to note that some relevant articles might have been unintentionally omitted.

Data Extraction

The details extracted from each article are presented in **Tables 2**–**5**. Each table represents the search results from the Scopus, IEEE Xplore, PubMed, and Web of Science databases. Data extracted include study aim and scope, computational techniques (DL models), CXR datasets, evaluation metrics/results achieved, and publication year. The database search shows that most IEEEXplore articles are conference proceedings and included in this systematic review, along with a few conference articles from springer and ACM. This is due to the articles' quality and contribution to the subject matter.

RESULTS

This study reviewed computer-aided diagnosis systems articles to detect pulmonary TB from January 2017 until September 2021. It was observed from the articles that the development of CAD follows a standard framework involving four steps as follows: "pre-processing" is the first step which deals with cleaning up the CXR images by eliminating noise and enhancing for clarity. The second step is "segmentation" of a region of interest from the entire image, which is the lung field region in the case of CXR. The third step is the "Feature extraction," where discriminative features are identified and selected for further analysis in the "classification" step, where the various images are categorized as normal or abnormal (infected) with TB. Several techniques such as handcraft, machine learning, and deep learning have been employed to diagnose TB, but DL has recorded more success in this regard; hence our interest was to analyze the CAD system based on one or more DL techniques as the classifier for TB detection.

The descriptive analysis of the results is presented in **Tables 2**–**5**. These tables show the computational technique, study scope, datasets, evaluation criteria, and results. Articles that utilized CXR as the only imaging modality and employed DL as the only computational technique for developing CAD are considered.

Imaging Modalities

Different screening procedures are used to confirm the presence of TB. Still, Chest Radiograph, otherwise referred to as Chest X-ray (CXR), is a radiograph tool used to detect abnormalities in the lungs and nearby structures. In recent years, CXR has remained a vital method for screening TB and other lung diseases and hence recommended by WHO due to its high sensitivity, wide availability, and relatively less expensive (70, 71).

Deep Learning Techniques

Many DL techniques have been used for screening, predicting, and diagnosing TB. In most classification algorithms, a dataset is required for training and testing with many samples of inputs and outputs to learn from. Model is developed using the training set to calculate how best to map examples of input data to a specific class label; then, the model validation is accomplished using the test set (72).

In **Tables 2**–**5**, only the results obtained from the test set are extracted. Some studies employed more the one DL technique to find the optimal results for diagnosing TB diseases. Only the best results are documented in this study if more than one accuracy is reported in an article.

Datasets

Data is crucial for developing CAD required to solve life-threatening diseases, including TB, as one of the leading causes of worldwide death. The various popular datasets that have been used in developing TB detection algorithms contain de-identified CXR images to protect the privacy of the patients. In other words, the identity of the patients is not disclosed. Most of the datasets are accompanied by radiological interpretation of the observed manifestation that can serve as groundtruth. These datasets are made available to the research communities to foster state-of-the-art research into finding lasting solutions to the early diagnosis of TB manifestations. Some of these public datasets include Montgomery County (73), Shenzhen (73), Peruvian (49), KIT (74), MIMIC-CXR (75), Belarus, NIH (chest x-ray 8, 14) (76), JSRT. **Figure 3** shows the datasets frequency of use.

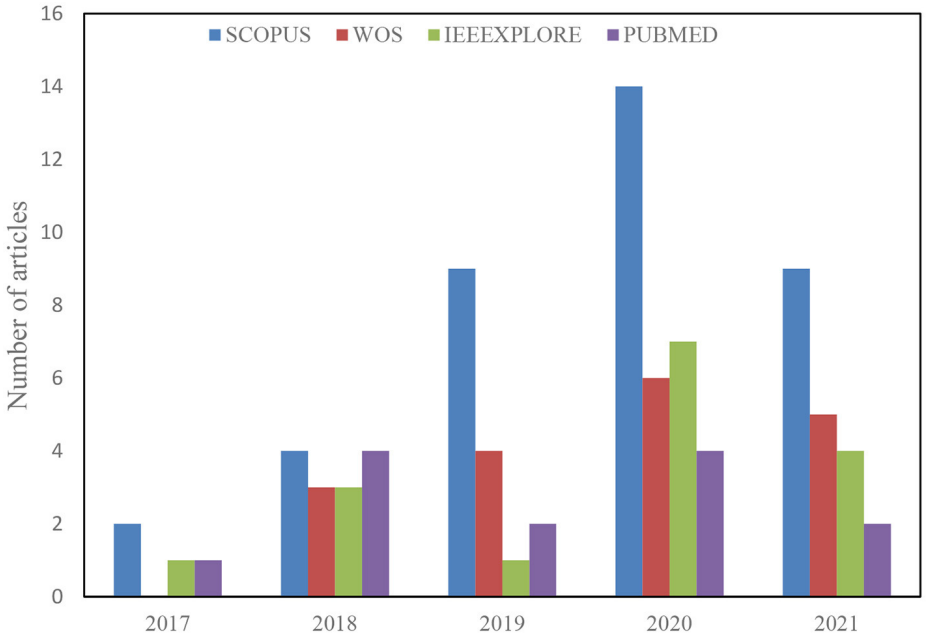


FIGURE 2 | Articles selection by database and year.

Evaluation Metrics

Once a model is trained using some training images, the test dataset is then employed to assess the quality of the model. It is evident from the data extraction process that different evaluation metrics exist for assessing model performance. Most of the popular evaluation metrics applied to the development of CAD system includes accuracy, sensitivity, specificity, and AUC. These evaluation metrics are briefly explained as follows:

Accuracy is the rate of the correct samples from the total number of samples examined (77). The equation gives the accuracy:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

Sensitivity is the proportion of the actual positive samples that are correctly identified as positive (78). This metric is given as:

$$Sensitivity = \frac{TP}{TP + FN}$$

Specificity is the ratio of the actual negative samples that are correctly identified as negative (78). This metric is given as:

$$Specificity = \frac{TN}{TN + FP}$$

Precision, otherwise known as a positive predictive value, is the ratio of positive samples that are accurately predicted (77). This is given as:

$$Precision = \frac{TP}{TP + FP}$$

AUC-ROC Curve: sometimes written as AUROC, specify the rate at which a model distinguishes between classes [84]. It tells the probability of separating negative samples from positive samples. Thresholds are set to determine the ROC curve, which is the plot of True Positive Rate (TPR) against False Positive Rate (FPR) given as:

$$TPR = Sensitivity = \frac{TP}{TP + FN}$$

$$FPR = 1 - Specificity = \frac{FP}{FP + TN}$$

Where TP, TN, FP, FN, TPR, and FPR are true positive, true negative, false positive, false negative, true positive rate, and false positive rate.

DISCUSSION

This systematic review extensively searched the various Deep Learning classifiers for diagnosing TB from CXR. Automatic detection of TB has received mammoth attention in the last decades resulting in many publications with state-of-the-art techniques. **Figure 4** presents a hierarchical chart of computational techniques categorized according to the frequency of usage in CAD systems for TB based on the included articles. Despite some good accuracies reported in some studies, we found some limitations in the existing studies concerning methods and reported accuracy, which should be a point of consideration for developing CAD systems in the future. Many studies measured diagnostic accuracy without evaluating the risk of bias emerging from the datasets that were used. It is essential that the accuracy

TABLE 2 | Scopus search results.

References	Computational techniques	Study aim/scope	Datasets	Results	Year
Duong et al. (13)	EfficientNet, Vision Transformer	Classification of CXR as normal, pneumonia and TB	Montgomery, Shenzhen, Belarus, RSNA, A COVID-19 CXR	Acc = 97.72%, Auc = 100%	2021
Norval et al. (14)	AlexNet, VGG16, and VGG19	Improve accuracy and Classification of CXR as Has TB and No TB	Montgomery, Shenzhen, NIH	Acc = 89.99%	2021
Rahman et al. (15)	ResNet101, VGG19, and DenseNet201 XGBoost	The study utilizes three Deep CNN models as features extractor then classify using eXtreme Gradient Boosting	Montgomery, Shenzhen, Belarus, RSNA, Private	Acc = 99.92% Auc = 99.93% Pre = 99.85% Sen = 100% Spe = 99.85% F1 = 99.92%	2021
Govindarajan and Swaminathan (16)	ELM, OSELM	Identify and classify TB conditions from healthy subjects in chest radiographs using integrated local feature descriptors and variants of extreme learning machine	Montgomery	Acc = 99.2% Sen = 99.3% Spe = 99.3% Pre = 99.0% F1 = 99.20%	2021
Alawi et al. (17)	CNN	Study proposed automated technique to diagnose TB from CXR	NLM, Belarus, NIAID TB and RSNA	Acc = 98.71% Sen = 98.86% Spe = 98.57%	2021
Khatibi et al. (18)	Complex networks and stacked ensemble (CCNSE), CNN	A multi –instance classification model to detect TB from CXR is proposed	Shenzhen, Montgomery	Acc = 99.26%, Auc = 99.00%	2021
Ayaz et al. (19)	Deep CNN	Present a hybrid method for TB detection	Montgomery, Shenzhen	Auc = 0.99%	2021
Priya and Vimina (20)	VGG-19, RestNet50, DenseNet121, InceptionV3	The study employs transfer learning for TB diagnosis.	Montgomery, Shenzhen	Auc = 0.95%	2021
Dasanayaka and Dissanayake (21)	DC-GAN, VGG16 and InceptionV3	generate, segment, and classify CXR for TB using three deep architectures	Montgomery, Shenzhen	Acc = 97.10% Sen = 97.90% Spe = 96.20%	2021
Msonda et al. (22)	AlexNet, GoogLeNet, ResNet50 and Spatial Pyramid Pooling (SPP)	Integrate SPP with Deep CNN to improve performance for TB detection	Montgomery, Shenzhen, Private (KERH)	Acc = 0.98% Pre = 1.00% Spe = 1.00% F1 = 1.00%	2020
Owais et al. (23)	Fusion-based deep classification network	The study proposes a CAD system for the effective diagnosis of TB and provides visual with descriptive information that is useful to the radiologist	Montgomery, Shenzhen	Acc = 0.928% Auc = 0.965% Pre = 0.937% Recall = 0.921% F1 = 0.929%	2020
Yoo et al. (24)	ResNet18	Classify CXR into Normal, TB and Non-TB	Montgomery, Shenzhen, NIH	Acc = 0.98% Sen = 0.98% Spe = 0.97% Auc = 0.965% Pre = 0.97%	2020
Sathitratanacheewin et al. (25)	DCNN	To examine the generalization of deep CNN models for classification of CXR as normal or abnormal with different manifestations	Shenzhen, NIH (ChestX-ray8)	Acc = 0.985% Sen = 72% Spe = 82%	2020
Sahlol et al. (26)	MobileNet Artificial Ecosystem-based Optimization (AEO)	Classification of CXR to detect TB	Shenzhen, Private	Acc = 94.1%	2020
Das et al. (27)	InceptionV3	Screening CXR for TB abnormalities	Montgomery, Shenzhen	Acc = 91.7% Auc = 0.96% Sen = 0.89% Spe = 0.93% Pre = 0.93%	2020
Rahman et al. (28)	ResNet, ChexNet, InceptionV3, Vgg19, DenseNet201, SqueezeNet, MobileNet, and Ensemble	Automatic detection of TB from the CXR.	NLM, Belarus, NIAID TB, and RSNA	Acc = 98.6% F1 = 98.56% Sen = 98.56% Spe = 98.54% Pre = 98.57%	2020

(Continued)

TABLE 2 | Continued

References	Computational techniques	Study aim/scope	Datasets	Results	Year
Munadi et al. (29)	ResNet and EfficientNet	Enhances CXR images for improving TB detection accuracy	Shenzhen	Acc = 91.7% Auc = 0.96%	2020
Oloko-Oba and Viriri (30)	CNN	Detection of TB from CXR and classification as normal and abnormal	Shenzhen	Acc = 87.8%	2020
Xie et al. (31)	Faster RCNN	Detection of multiple categories of TB lesions in CXR	Montgomery, Shenzhen	Acc = 0.926% Auc = 0.977%	2020
Verma et al. (32)	InceptionV3, faster RCNN	Classify CXR as pulmonary TB and Pneumonia	Shenzhen	Acc = 99.01%	2020
Tasci (33)	AlexNet, VGGNet	classifying CXR ROI for TB detection	Montgomery, Shenzhen	Acc = 88.32% Auc = 0.92%	2020
Rajaraman and Antani (34)	Inception-V3, ResNet-V2, VGG-16, Xception, DenseNet-121, Ensemble	Improve state-of-the-art architecture for TB detection from CXR	Shenzhen, RSNA, Indiana	Acc = 0.941% F1 = 0.941% Sen = 0.926% Spe = 1.00% Auc = 0.990%	2020
Abideen et al. (35)	B-CNN	Identification and classification of CXR as TB and Non-TB	Montgomery, Shenzhen	Acc = 96.42%	2020
Hijazi et al. (36)	Ensemble of VGG16 InceptionV3	Detection of TB from CXR	Montgomery, Shenzhen	Acc = 89.77% Sen = 90.91% Spe = 88.64%	2019
Pasa et al. (37)	CNN	Developed a faster TB detection algorithm	Montgomery, Shenzhen, Belarus	Acc = 84.4% Auc = 0.900%	2019
Meraj et al. (38)	VGGNet, RestNet50, GoogLeNet	Detection of TB abnormalities from CXR	Montgomery, Shenzhen	Acc = 86.74% Auc = 92.0%	2019
Ahsan et al. (39)	VGG16	Screening of CXR to identify the presence of TB	Montgomery, Shenzhen	Acc = 81.25%	2019
Nguyen et al. (40)	ResNet-50, VGGNet, DenseNet-121, Inception, ResNet.	Improving detection rate of TB	Montgomery, Shenzhen, NIH-14	Auc = 0.99%	2019
Ho et al. (41)	InceptionResNetV2, ResNet150, DenseNet-121	Classification of CXR as pulmonary TB or healthy.	ChestX-ray14, Shenzhen, Montgomery	Auc = 0.95%	2019
Heo et al. (42)	VGG19, InceptionV3, ResNet-50, DenseNet-121, InceptionResNetV2.	Detection of TB from CXR	Privete (Yonsei University)	Auc = 0.9213% Sen = 0.815% Spe = 0.962%	2019
Hernández et al. (43)	Ensemble of VGG19, InceptionV3, ResNet-50	Automatic classification of CXR for TB detection.	Private	Acc = 0.8642.%	2019
Hijazi et al. (44)	Ensemble of InceptionV3, VGG16	Detection of TB from CXR without segmentation	Shenzhen, Montgomery	Acc = 91.0% Sen = 89.6% Spe = 90.7%	2019
Abbas and Abdelsamea (45)	AlexNet	Classification of CXR as healthy or having TB manifestation	Montgomery	Auc = 0.998% Sen = 0.997% Spe = 0.999%	2018
Karnkawinpong and Limpiyakorn (46)	AlexNet, VGG16, CapsNet	CAD for early diagnosis of TB	Private (Thai), Shenzhen, Montgomery	Acc = 90.79% Sen = 89.07% Spe = 92.50%	2018
Stirenko et al. (47)	DCNN	Prediction of the presence of TB from CXR	Shenzhen	———	2018
Becker et al. (48)	CNN	Detection and classification of different TB pathologies from CXR	Private	Auc = 0.98%	2018
Liu et al. (49)	AlexNet, GoogLeNet	Detection and classification of TB manifestations in CXR images	Peruvian	Acc = 85.68%	2017
Hooda et al. (3)	DCNN	Detect and classify TB from CXR as normal and abnormal	Shenzhen, Montgomery	Acc = 82.09%	2017

of CAD systems is evaluated using a different set of datasets (CXR images) from the set used for training. In other words, avoid

- Using the same set of CXR images for training and testing.
- Testing with CXR images that were not used for training but originated from the same image subset.
- Using images with class imbalance, and
- Using unannotated CXR images.

Otherwise, the diagnostic accuracy evaluation is likely to be exaggerated and could impact the overall generalization of the system. In general, about 80% of the studies used the

TABLE 3 | PubMed search results.

References	Computational techniques	Study aim/scope	Datasets	Results	Year
Oloko-Oba and Viriri (50)	Ensemble of VGG-16, ResNet50, Inception V3	Automatic detection of TB from CXR	Shenzhen, Montgomery	Acc = 96.14% Sen = 90.03% Spe = 92.41%	2021
Lee et al. (51)	DLAD	Detection of active TB and classification of relevant abnormalities on CXR	Private	Auc = 0.967% Sen = 0.821% Spe = 0.997%	2021
Zhang et al. (52)	Convolutional Block Attention Module (CBAM)	Classification of TB from CXR	Private	Acc = 87.7% Auc = 94.3% Recall = 89.7% Spe = 85.9.7%	2020
Hwang et al. (53)	DLAD	Developed a Deep Learning-based automatic detection algorithm (DLAD) for active Pulmonary TB on CXR and validate its performance using various datasets compared to physicians' results.	Shenzhen, Montgomery, Private (SNUH)	Auc = 0.977% Sen = 94.3% Spe = 91.1%	2019
Rajpurkar et al. (54)	CNN (CheXNeXt)	To evaluate the effectiveness of CheXNeXt in detecting TB and other abnormalities from CXR	ChestX-ray14	Auc = 0.862% Sen = 0.594% Spe = 0.927%	2018
Lakhani and Sundaram (55)	Ensemble of AlexNet, GoogLeNet	Evaluates the efficacy of deep models for detecting TB on CXR	Belarus, Shenzhen, Montgomery	Auc = 0.99% Sen = 97.3% Spe = 94.7%	2017

TABLE 4 | IEEE Xplore search results.

References	Computational techniques	Study aim/scope	Datasets	Results	Year
Cao et al. (56)	DenseNet121 VGGNet16, VGGNet19, ResNet152	Evaluates the performance of deep learning for classification of CXR for TB	Shenzhen, Montgomery	Acc = 90.38% F1 = 90.36% Pre = 90.33% Recall = 90.53%	2021
Karaca et al. (57)	VGG16, VGG19, DenseNet121, MobileNet, InceptionV3	Development of a TB detection system	Montgomery	Acc = 98.9% Auc = 1.00%	2021
Saif et al. (58)	DenseNet169, ResNet-50, InceptionV3	Detection of TB from CXR	Shenzhen, Montgomery	Acc = 99.7% Sen = 97.5% Spe = 98.4%	2021
Das et al. (27)	InceptionV3	Screening TB from CXR to eliminate patents diagnosis delay	Shenzhen, Montgomery	Acc = 91.7% Auc = 0.96%	2021
Imam et al. (59)	Modified Inception	They analyzed patients' CXR to determine those infected with TB or not.	Shenzhen, Montgomery	Acc = 91%	2020
Griffin et al. (60)	R-CNN	Location of TB manifestations on CXR	Peruvian	Auc = 0.753% Sen = 0.922% Spe = 0.666%	2020
Rashid et al. (61)	Ensemble of ResNet, Inception-ResNet, DenseNet	Development of a CAD system to classify CXR as normal and infected with TB	Shenzhen	Acc = 90.5%	2018
Abbas and Abdelsamea (45)	AlexNet	Classification of CXR as healthy and unhealthy with TB manifestation	Shenzhen, Montgomery	Auc = 0.998% Sen = 0.999% Spe = 0.997%	2018

same CXR images for training and testing or did not report on it.

CONCLUSION AND RECOMMENDATIONS

This systematic review intends to inform researchers of the existing Deep Learning classifiers and assist them in developing a CAD system for the efficient diagnosis of TB. Different state-of-the-art Deep Learning techniques that have been used to detect TB have been explored and presented in **Tables 2–5**. Generally, it is challenging to compare the methods with each other extensively. Factors like the types of datasets used for evaluation, number of image samples used, evaluation metrics approach, and model parameters tunning make the comparison complex.

TABLE 5 | Web of Science search results.

References	Computational techniques	Study aim/scope	Datasets	Results	Year
Fehr et al. (62)	CAD4TBv5	Implement CAD4TB to screen for TB on CXR	Private	Sen = 82.8% Spe = 68.0%	2021
Vats et al. (63)	CNN (iDoc-X)	Diagnosis of TB manifestations from CXR and classify images as TB and Non-TB	Private	Acc = 91.10%	2021
Khatibi et al. (18)	CNNs, complex networks and stacked ensemble	TB recognition from CXR images	Shenzhen, Montgomery	Acc = 99.26% Auc = 99.00%	2021
Rajpurkar et al. (64)	DenseNet121	Development of CheXaid for diagnosing TB	Private	Acc = 0.78% Auc = 0.83%	2020
Grivkov and Smirnov (65)	InceptionV3	Screening of CXR to detect TB pathologist	Shenzhen, Montgomery	Acc = 0.868%	2020
Msonda et al. (22)	AlexNet, GooLeNet, ResNet50, SPP	Integrate SPP with DCNN for the diagnosis of TB on CXR.	Shenzhen, Montgomery, Private (KERH)	Auc = 0.98% Spe = 1.0% Pre = 1.0% F1 = 1.0% Recall = 0.99%	2020
Gozes and Greenspan (66)	DenseNet121	Learned specific features from CXR to detect TB	Chest X-ray14	Auc = 0.965%	2019
Karnkawinpong and Limpiyakorn (67)	AlexNet, VGG-16, CapsNet	Classification of TB from CXR	NLM, Private	Acc = 94.56% Sen = 92.83% Spe = 96.06%	2019
Sivaramakrishnan et al. (68)	AlexNet, VGG16, VGG19, Xception, ResNet-50	evaluate the performance of Deep models toward improving the accuracy of TB screening from CXR	Shenzhen, Montgomery, Private	Acc = 0.855% Auc = 0.956%	2018
Vajda et al. (69)	CNN	Screening CXR to determine which CXR images are normal or abnormal with TB.	Shenzhen, Montgomery	Acc = 97.03% Auc = 0.99%	2018

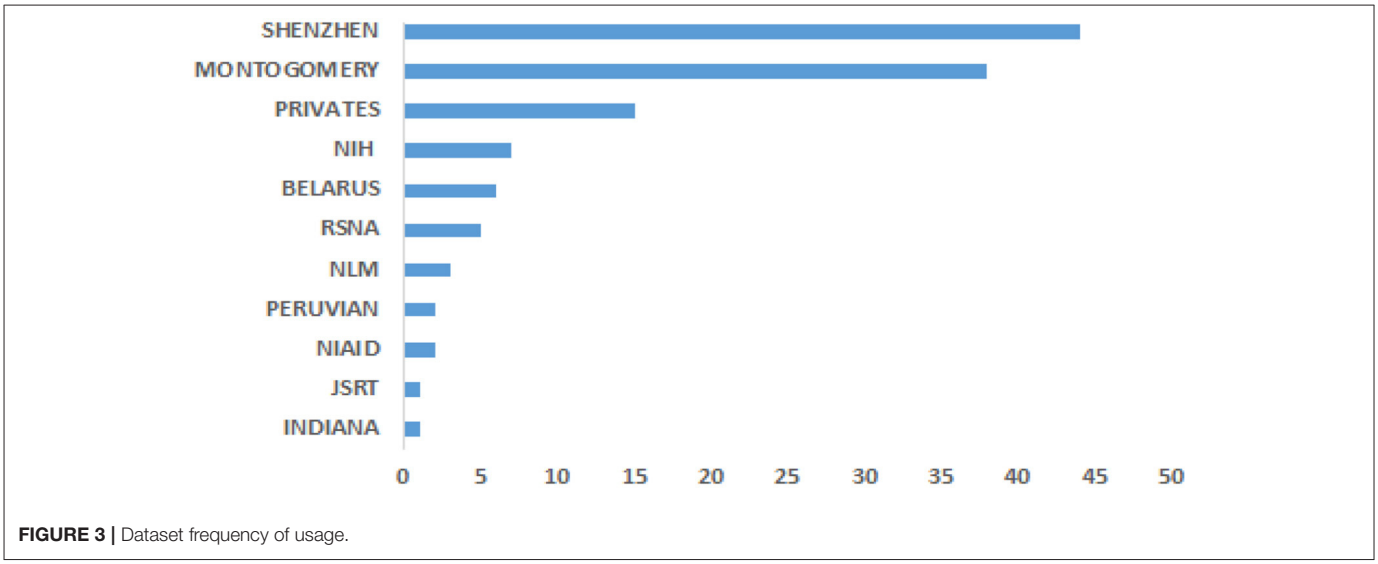


FIGURE 3 | Dataset frequency of usage.

As evident from the literature, these pre-trained models (VggNet, ResNet, AlexNet, DenseNet, and Inception) are the most popular and have been extensively explored for the classification of TB, as shown in **Figure 4**. Despite the effectiveness of Deep Learning models in detection and classification tasks, CAD systems for clinical diagnosis are still challenging in a real-world scenario. The physicians and radiologists see CAD intervention as a threat to their jobs rather

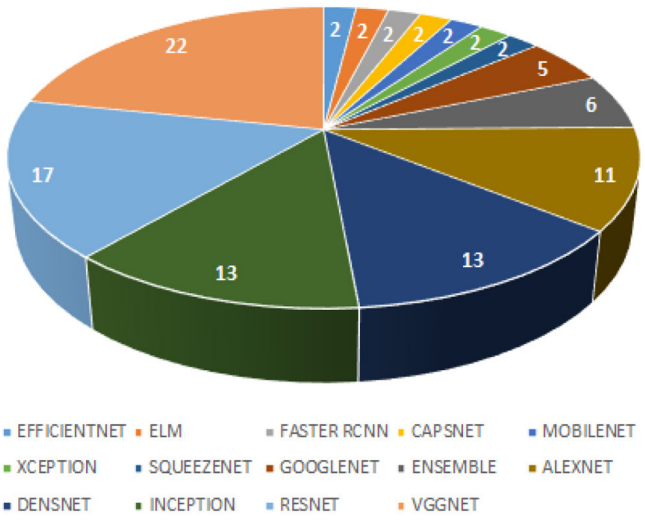


FIGURE 4 | Hierarchical chart of computational techniques according to the frequency of usage.

than a supporting system to improve physicians' performance in terms of time, effort, efficiency, and affordability, especially in developing countries. This review found that most existing works focused on development studies rather than clinical studies.

One of the likely weaknesses of this review is that it is limited to only the studies written in English because there might be other high-quality studies written in other languages. Also, we restricted the computational techniques to only Deep Learning classifiers, which could increase the risk of classifiers bias where studies that employed a hybrid of both Machine Learning and Deep Learning could have achieved better performance. Furthermore, this study did not undertake meta-analyses due to variations of algorithms used. Also, the raw data required to meta-analyze the diagnostic accuracy for most studies were unavailable.

However, it is recommended that studies be carried out using standardized public datasets that contain additional masks of the images that can be used as groundtruth to detect the infected aspect of the pulmonary images. It is highly recommended that models are trained with a set of images and evaluated on a different set of images. For instance, training a model with the Shenzhen datasets and evaluating it on the Montgomery dataset will validate better generalization. It is also recommended that future CAD systems focus more on clinical evaluation and should be able to identify foreign objects such as buttons and rings that look like a nodule on the CXR images, which may lead to misclassification.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

REFERENCES

1. *World Health Organization. Global Tuberculosis Report 2020: Executive Summary*. Available online at: https://apps.who.int/iris/handle/10665/337538 (accessed November 29, 2021). License: CC BY-NC-SA 3.0 IGO.
2. Baral SC, Karki DK, Newell JN. Causes of stigma and discrimination associated with tuberculosis in Nepal: a qualitative study. *BMC Public Health.* (2007) 7:1–0. doi: 10.1186/1471-2458-7-211
3. Hooda R, Sofat S, Kaur S, Mittal A, Meriaudeau F. Deep-learning: a potential method for tuberculosis detection using chest radiography. In: *2017 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)* (IEEE) (2017), 497–502.
4. Zumla A, George A, Sharma V, Herbert RH, Oxley A, Oliver M. The WHO 2014 global tuberculosis report—further to go. *Lancet Global Health.* (2015) 3:e10–2. doi: 10.1016/S2214-109X(14)70361-4
5. Sathitratanacheewin S, Pongpirul K. Deep learning for automated classification of tuberculosis-related chest x-ray: dataset specificity limits diagnostic performance generalizability. *arXiv[Prepint].arXiv:1811.07985*. (2018).
6. World Health Organization. *Chest Radiography in Tuberculosis Detection: Summary of Current WHO Recommendations and Guidance on Programmatic Approaches*. (2016). World Health Organization. Available online at: https://apps.who.int/iris/handle/10665/252424 (accessed November 29, 2021).
7. Williams FH. The use of X-ray examinations in pulmonary tuberculosis. *Boston Med Surg J.* (1907) 157:850–3. doi: 10.1056/NEJM190712261572602
8. Pande T, Pai M, Khan FA, Denkinger CM. Use of chest radiography in the 22 highest tuberculosis burden countries. *Eur Respir J.* (2015) 46:1816–9. doi: 10.1183/13993003.01064-2015
9. Noor NM, Rijal OM, Yunus A, Mahayiddin AA, Peng GC, Abu-Bakar SA. A statistical interpretation of the chest radiograph for the detection of pulmonary tuberculosis. In: *2010 IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES)* (IEEE). (2010), 47–51.
10. Pedrazzoli D, Lalli M, Boccia D, Houben R, Kranzer K. Can tuberculosis patients in resource-constrained settings afford chest radiography? *Eur Respir J.* (2017) 49:1601877. doi: 10.1183/13993003.01877-2016
11. Schmidhuber J. Deep learning in neural networks: an overview. *Neural Netw.* (2015) 61:85–117. doi: 10.1016/j.neunet.2014.09.003
12. Moher D, Liberati A, Tetzlaff J, Altman DG, Prisma Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med.* (2009) 6:e1000097. doi: 10.1371/journal.pmed.1000097
13. Duong LT, Le NH, Tran TB, Ngo VM, Nguyen PT. Detection of tuberculosis from chest X-ray images: boosting the performance with vision transformer and transfer learning. *Expert Syst Appl.* (2021) 184:115519. doi: 10.1016/j.eswa.2021.115519
14. Norval M, Wang Z, Sun Y. Evaluation of image processing technologies for pulmonary tuberculosis detection based on deep learning convolutional neural networks. *J Adv Inform Technol.* (2019) 12:253–9. doi: 10.12720/jait.12.3.253-259
15. Rahman M, Cao Y, Sun X, Li B, Hao Y. Deep pre-trained networks as a feature extractor with XGBoost to detect tuberculosis from chest X-ray. *Comput Elect Eng.* (2021) 93:107252. doi: 10.1016/j.compeleceng.2021.107252
16. Govindarajan S, Swaminathan R. Extreme learning machine based differentiation of pulmonary tuberculosis in chest radiographs using integrated local feature descriptors. *Comput Methods Progr Biomed.* (2021) 204:106058. doi: 10.1016/j.cmpb.2021.106058

17. Alawi AE, Al-basser A, Sallam A, Al-sabaeei A, Al-khateeb H. Convolutional neural networks model for screening tuberculosis disease. In: *2021 International Conference of Technology, Science and Administration (ICTSA)* (IEEE). (2021), 1–5.
18. Khatibi T, Shahsavari A, Farahani A. Proposing a novel multi-instance learning model for tuberculosis recognition from chest X-ray images based on CNNs, complex networks and stacked ensemble. *Phys Eng Sci Med.* (2021) 44:291–311. doi: 10.1007/s13246-021-00980-w
19. Ayaz M, Shaukat F, Raja G. Ensemble learning based automatic detection of tuberculosis in chest X-ray images using hybrid feature descriptors. *Phys Eng Sci Med.* (2021) 44:183–94. doi: 10.1007/s13246-020-00966-0
20. Priya PA, Vimina ER. Tuberculosis detection from CXR: an approach using transfer learning with various CNN architectures. In: *International Conference on Communication, Computing and Electronics Systems: Proceedings of ICCCES 2020*. Vol. 733 (Springer Nature). (2021), 407p.
21. Dasanayaka C, Dissanayake MB. Deep learning methods for screening pulmonary tuberculosis using chest X-rays. *Comput Methods Biomech Biomed Eng Imaging Vis.* (2021) 9:39–49. doi: 10.1080/21681163.2020.1808532
22. Msonda P, Uymaz SA, Karaagaç SS. Spatial pyramid pooling in deep convolutional networks for automatic tuberculosis diagnosis. *Traitement du Signal.* (2020) 37:1075–1084. doi: 10.18280/ts.370620
23. Owais M, Arsalan M, Mahmood T, Kim YH, Park KR. Comprehensive computer-aided decision support framework to diagnose tuberculosis from chest X-ray images: data mining study. *JMIR Med Inform.* (2020) 8:e21790. doi: 10.2196/21790
24. Yoo SH, Geng H, Chiu TL, Yu SK, Cho DC, Heo J, et al. Study on the TB and non-TB diagnosis using two-step deep learning-based binary classifier. *J Instrument.* (2020) 15:P10011. doi: 10.1088/1748-0221/15/10/P10011
25. Sathitratanacheewin S, Sunanta P, Pongpirul K. Deep learning for automated classification of tuberculosis-related chest X-Ray: dataset distribution shift limits diagnostic performance generalizability. *Heliyon.* (2020) 6:e04614. doi: 10.1016/j.heliyon.2020.e04614
26. Sahlol AT, Abd Elaziz M, Tariq Jamal A, Damaševičius R, Farouk Hassan O. A novel method for detection of tuberculosis in chest radiographs using artificial ecosystem-based optimization of deep neural network features. *Symmetry.* (2020) 12:1146. doi: 10.3390/sym12071146
27. Das D, Santosh KC, Pal U. Inception-based deep learning architecture for tuberculosis screening using chest X-rays. In: *2020 25th International Conference on Pattern Recognition (ICPR)*. Vol. 10 (IEEE) (2021), 3612–9.
28. Rahman T, Khandakar A, Kadir MA, Islam KR, Islam KF, Mazhar R, et al. Reliable tuberculosis detection using chest X-ray with deep learning, segmentation and visualization. *IEEE Access.* (2020) 8:191586–601. doi: 10.1109/ACCESS.2020.3031384
29. Munadi K, Muchtar K, Maulina N, Pradhan B. Image enhancement for tuberculosis detection using deep learning. *IEEE Access.* (2020) 8:217897–907. doi: 10.1109/ACCESS.2020.3041867
30. Oloko-Oba M, Viriri S. Tuberculosis abnormality detection in chest X-rays: a deep learning approach. In: *International Conference on Computer Vision and Graphics*. Vol. 14 (Cham: Springer) (2020), 121–32.
31. Xie Y, Wu Z, Han X, Wang H, Wu Y, Cui L, et al. Computer-aided system for the detection of multicategory pulmonary tuberculosis in radiographs. *Journal of Healthcare Engineering.* (2020) 2020:1–12. doi: 10.1155/2020/9205082
32. Verma D, Bose C, Tufchi N, Pant K, Tripathi V, Thapliyal A. An efficient framework for identification of tuberculosis and pneumonia in chest X-ray images using Neural Network. *Procedia Computer Science.* (2020) 171:217–24. doi: 10.1016/j.procs.2020.04.023
33. Tasci E. Pre-processing effects of the tuberculosis chest x-ray images on pre-trained cnns: an investigation. *In The International Conference on Artificial Intelligence and Applied Mathematics in Engineering. Vol. 20 (Cham: Springer).* (2019), 589-596. doi: 10.1007/978-3-030-36178-5_48
34. Rajaraman S, Antani SK. Modality-specific deep learning model ensembles toward improving TB detection in chest radiographs. *IEEE Access.* (2020) 8:27318–26. doi: 10.1109/ACCESS.2020.2971257
35. Abideen ZU, Ghafoor M, Munir K, Saqib M, Ullah A, Zia T, et al. Uncertainty assisted robust tuberculosis identification with bayesian convolutional neural networks. *IEEE Access.* (2020) 8:22812–25. doi: 10.1109/ACCESS.2020.2970023
36. Hijazi MH, Hwa SK, Bade A, Yaakob R, Jeffree MS. Ensemble deep learning for tuberculosis detection using chest X-Ray and canny edge detected images. *IAES Int J Artif Intel.* (2019) 8:429. doi: 10.11591/ijai.v8.i4.pp429-435
37. Pasa F, Golkov V, Pfeiffer F, Cremers D, Pfeiffer D. Efficient deep network architectures for fast chest X-ray tuberculosis screening and visualization. *Sci Rep.* (2019) 9:1–9. doi: 10.1038/s41598-019-42557-4
38. Meraj SS, Yaakob R, Azman A, Rum SN, Shahrel A, Nazri A, et al. Detection of pulmonary tuberculosis manifestation in chest x-rays using different convolutional neural network (CNN) models. *Int J Eng Adv Technol.* (2019) 9:2270–5. doi: 10.35940/ijeat.A2632.109119
39. Ahsan M, Gomes R, Denton A. Application of a convolutional neural network using transfer learning for tuberculosis detection. In: *2019 IEEE International Conference on Electro Information Technology (EIT)* (IEEE) (2019), 427–33.
40. Nguyen QH, Nguyen BP, Dao SD, Unnikrishnan B, Dhingra R, Ravichandran SR, et al. Deep learning models for tuberculosis detection from chest x-ray images. In: *2019 26th International Conference on Telecommunications (ICT)* (IEEE) (2019), 381–5.
41. Ho TK, Gwak J, Prakash O, Song JI, Park CM. Utilizing pretrained deep learning models for automated pulmonary tuberculosis detection using chest radiography. In: *11th Asian Conference on Intelligent Information and Database Systems, ACIIDS* (Springer Verlag) (2019), 395–403.
42. Heo SJ, Kim Y, Yun S, Lim SS, Kim J, Nam CM, et al. Deep learning algorithms with demographic information help to detect tuberculosis in chest radiographs in annual workers' health examination data. *Int J Environ Res Public Health.* (2019) 16:250. doi: 10.3390/ijerph16020250
43. Hernández A, Panizo Á, Camacho D. An ensemble algorithm based on deep learning for tuberculosis classification. In: *International Conference on Intelligent Data Engineering and Automated Learning* (Cham: Springer). (2019), 145–54.
44. Hijazi M, Yang L, Alfred R, Mahdin H, Yaakob R. Ensemble deep learning for tuberculosis detection. *Indonesian J Elect Eng Comput Sci.* (2020) 8:429. doi: 10.11591/ijeecs.v17.i2.pp1014-1020
45. Abbas A, Abdelsamea MM. Learning transformations for automated classification of manifestation of tuberculosis using convolutional neural network. In: *2018 13th International Conference on Computer Engineering and Systems (ICCES)* (IEEE). (2018), 122–6.
46. Karnkawinpong T, Limpiyakorn Y. Chest X-ray analysis of tuberculosis by convolutional neural networks with affine transforms. In: *Proceedings of the 2018 2nd International Conference on Computer Science and Artificial Intelligence.* (2018), 90–3.
47. Stirenko S, Kochura Y, Alienin O, Rokovyi O, Gordienko Y, Gang P, et al. Chest X-ray analysis of tuberculosis by deep learning with segmentation and augmentation. In: *2018 IEEE 38th International Conference on Electronics and Nanotechnology (ELNANO)* (IEEE) (2018), 422–8.
48. Becker AS, Blüthgen C, Sekaggya-Wiltshire C, Castelnuovo B, Kambugu A, Fehr J, et al. Detection of tuberculosis patterns in digital photographs of chest X-ray images using deep learning: feasibility study. *Int J Tubercul Lung Dis.* (2018) 22:328–35. doi: 10.5588/ijtld.17.0520
49. Liu C, Cao Y, Alcantara M, Liu B, Brunette M, Peinado J, et al. TX-CNN: Detecting tuberculosis in chest X-ray images using convolutional neural network. In: *2017 IEEE International Conference on Image Processing (ICIP)* (IEEE). (2017), 2314–8.
50. Oloko-Oba M, Viriri S. Ensemble of convolution neural networks for automatic tuberculosis classification. In: *International Conference on Computational Collective Intelligence* (Cham: Springer) (2021), 549–59. doi: 10.1007/978-3-030-88081-1_41
51. Lee JH, Park S, Hwang EJ, Goo JM, Lee WY, Lee S, et al. Deep learning–based automated detection algorithm for active pulmonary tuberculosis on chest radiographs: diagnostic performance in systematic screening of asymptomatic individuals. *Eur Radiol.* (2021) 31:1069–80. doi: 10.1007/s00330-020-07219-4
52. Zhang R, Duan H, Cheng J, Zheng Y. A study on tuberculosis classification in chest X-ray using deep residual attention networks. In: *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)* (IEEE) (2020), 1552–5.
53. Hwang EJ, Park S, Jin KN, Kim JI, Choi SY, Lee JH, et al. Development and validation of a deep learning–based automatic detection algorithm for

active pulmonary tuberculosis on chest radiographs. *Clin Infect Dis.* (2019) 69:739–47. doi: 10.1093/cid/ciy967
54. Rajpurkar P, Irvin J, Ball RL, Zhu K, Yang B, Mehta H, et al. Deep learning for chest radiograph diagnosis: a retrospective comparison of the CheXNeXt algorithm to practicing radiologists. *PLoS Med.* (2018) 15:e1002686. doi: 10.1371/journal.pmed.1002686
55. Lakhani P, Sundaram B. Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. *Radiology.* (2017) 4:574–82. doi: 10.1148/radiol.2017162326
56. Cao K, Zhang J, Huang M, Deng T. X-ray classification of tuberculosis based on convolutional networks. In: *2021 IEEE International Conference on Artificial Intelligence and Industrial Design (AIID)* (IEEE). (2021), 125–9. doi: 10.1109/AIID51893.2021.9456476
57. Karaca BK, Güney S, Dengiz B, Agildere M. Comparative study for tuberculosis detection by using deep learning. In: *2021 44th International Conference on Telecommunications and Signal Processing (TSP)* (IEEE) (2021), 88–91.
58. Saif AF, Imtiaz T, Shahnaz C, Zhu WP, Ahmad MO. Exploiting cascaded ensemble of features for the detection of tuberculosis using chest radiographs. *IEEE Access.* (2021) 9:112388–99. doi: 10.1109/ACCESS.2021.3102077
59. Imam OT, Haque M, Shahnaz C, Imran SA, Islam MT, Islam MT. Detection of tuberculosis from chest X-ray images based on modified inception deep neural network model. In: *2020 IEEE International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE)* (IEEE). (2020), 360–3.
60. Griffin T, Cao Y, Liu B, Brunette MJ. Object detection and segmentation in chest X-rays for tuberculosis screening. In: *2020 Second International Conference on Transdisciplinary AI (TransAI)* (IEEE) (2020), 34–42.
61. Rashid R, Khawaja SG, Akram MU, Khan AM. Hybrid RID network for efficient diagnosis of tuberculosis from chest X-rays. In: *2018 9th Cairo International Biomedical Engineering Conference (CIBEC)* (IEEE) (2018), 167–70.
62. Fehr J, Konigorski S, Olivier S, Gunda R, Surujdeen A, Gareta D, et al. Computer-aided interpretation of chest radiography reveals the spectrum of tuberculosis in rural South Africa. *NPJ Dig Med.* (2021) 4:1–10. doi: 10.1101/2020.09.04.20188045
63. Vats S, Singh S, Kala G, Tarar R, Dhawan S. iDoc-X: an artificial intelligence model for tuberculosis diagnosis and localization. *J Discrete Math Sci Cryptogr.* (2021) 24:1257–72. doi: 10.1080/09720529.2021.1932910
64. Rajpurkar P, O'Connell C, Schechter A, Asnani N, Li J, Kiani A, et al. CheXaid: deep learning assistance for physician diagnosis of tuberculosis using chest x-rays in patients with HIV. *NPJ Dig Med.* (2020) 3:1–8. doi: 10.1038/s41746-020-00322-2
65. Grivkov AV, Smirnov AA. Application of convolutional neural networks for diagnostics of tuberculosis. *AIP Conf Proc.* (2020) 2313:080011.
66. Gozes O, Greenspan H. Deep feature learning from a hospital-scale chest x-ray dataset with application to TB detection on a small-scale dataset. In: *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (IEEE) (2019), 4076–9.
67. Karnkawinpong T, Limpiyakorn Y. Classification of pulmonary tuberculosis lesion with convolutional neural networks. *J Phys Conf Series.* (2019) 1195:012007. doi: 10.1088/1742-6596/1195/1/012007
68. Sivaramakrishnan R, Antani S, Candemir S, Xue Z, Abuya J, Kohli M, et al. Comparing deep learning models for population screening using chest radiography. *Med Imaging 2018 Comput Aided Diag.* (2018) 10575:105751E. doi: 10.1117/12.2293140
69. Vajda S, Karargyris A, Jaeger S, Santosh KC, Candemir S, Xue Z, et al. Feature selection for automatic tuberculosis screening in frontal chest radiographs. *J Med Syst.* (2018) 42:1–11. doi: 10.1007/s10916-018-0991-9
70. Puddy E, Hill C. Interpretation of the chest radiograph. Continuing education in anaesthesia. *Crit Care Pain.* (2007) 7:71–5. doi: 10.1093/bjaceaccp/mkm014
71. Gilpin C, Korobitsyn A, Migliori GB, Raviglione MC, Weyer K. The world health organization standards for tuberculosis care and management. *Eur Respir J.* (2018) 51:1800098. doi: 10.1183/13993003.00098-2018
72. Brownlee J. *4 Types of Classification Tasks in Machine Learning.* Machine Learning Mastery, 4p. Available online at: https://machinelearningmastery.com/types-of-classification-in~machine-learning (accessed November 25, 2021).
73. Jaeger S, Candemir S, Antani S, Wáng YX, Lu PX, Thoma G. Two public chest X-ray datasets for computer-aided screening of pulmonary diseases. *Quant Imaging Med Surg.* (2014) 4:475. doi: 10.3978/j.issn.2223-4292.2014.11.20
74. Irvin J, Rajpurkar P, Ko M, Yu Y, Ciurea-Ilcus S, Chute C, et al. Chexpert: a large chest radiograph dataset with uncertainty labels and expert comparison. *Proc AAAI conf Artif Intell.* (2019) 33:590–7. doi: 10.1609/aaai.v33i01.3301590
75. Johnson AE, Pollard TJ, Greenbaum NR, Lungren MP, Deng CY, Peng Y, et al. MIMIC-CXR-JPG, a large publicly available database of labeled chest radiographs. *arXiv[Prepint].arXiv:1901.07042.* (2019).
76. Hwang S, Kim HE, Jeong J, Kim HJ. A novel approach for tuberculosis screening based on deep convolutional neural networks. *Med Imaging Comput Aided Diag.* (2016) 9785:97852W. doi: 10.1117/12.2216198
77. Lalkhen AG, McCluskey A. Clinical tests: sensitivity and specificity. *Continuing Educ Anesth Crit Care Pain.* (2008) 8:221–3. doi: 10.1093/bjaceaccp/mkn041
78. Davis J, Goadrich M. The relationship between precision-recall and ROC curves. In: *Proceedings of the 23rd International Conference on Machine Learning 2006 Jun 25*, 233–40.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

2.2.2 Conclusion

This systematic review aims to identify modern DL algorithms for CAD development in TB diagnosis and serve as a bedrock for future diagnostic systems. Different DL techniques that have been used to detect and classify TB were explored and presented. Comparing these techniques was challenging due to variations in datasets, model parameters tuning, and evaluation metrics. We found that pretrained models like ResNet, VggNet, DenseNet, Inception, and AlexNet had been extensively explored for TB classification. The review analysis critically states the CXR datasets, evaluation metrics, and results. It also shows that most of the existing CAD focused on developmental studies and therefore encourages more clinical studies in the future.

Chapter 3

Detection of Tuberculosis from Chest X-Ray

The previous chapter discussed an extensive review of the approaches employed for the diagnosis of TB. It showed the progress and highlighted some limitations thus far. This chapter presents published work of our contribution to the development of CAD for TB diagnosis.

3.1 Diagnosing Tuberculosis Using Deep Convolutional Neural Network

3.1.1 Introduction

This section introduces a research paper that presented a method to classify the Montgomery CXR images. The paper proposed a CNN-based model comprising features extraction and classification. The model consisted of four lightweight network layers that handled the extraction of distinct TB features from CXR and a fully connected layer consisting of biases and weights that connected neurons required by the softmax layer; the softmax then determined the probability of the extracted features and provide the classification output. The work employed techniques that catered for the limited dataset and a regularisation algorithm for tuning the learning rate during the model training, which helped boost performance.

The research paper was presented at the [1] and published in [2] while the Montgomery dataset can be downloaded at [3].

[1] 9th International Conference on Image and Signal Processing ICISP 2020 June 4-6, 2020, Marrakech, Morocco.

[2] **Mustapha Oloko-Oba and Serestina Viriri**. "Diagnosing Tuberculosis Using Deep Convolutional Neural Network", *Image and Signal Processing*. LNCS Springer, vol 12119, pp 151-161 **(2020)**. DOI: `https://doi.org/10.1007/978-3-030-51935-3_16`.

[3] `https://lhncbc.nlm.nih.gov/publication/pub9931`

Diagnosing Tuberculosis Using Deep Convolutional Neural Network

Mustapha Oloko-Oba and Serestina Viriri( )

School of Mathematics, Statistics and Computer Science,
University of KwaZulu-Natal, Durban, South Africa
mustaphasmo@yahoo.com, viriris@ukzn.ac.za

Abstract. One of the global topmost causes of death is Tuberculosis (TB) which is caused by mycobacterium bacillus. The increase rate of infected people and the recorded deaths from TB disease is as a result of its transmissibility, lack of early diagnosis, and inadequate professional radiologist in developing regions where TB is more prevalent. Tuberculosis is unquestionably curable but needs to be detected early for necessary treatment to be effective. Many screening techniques are available, but chest radiograph has proven to be valuable for screening pulmonary diseases but hugely dependent on the interpretational skill of an expert radiologist. We propose a Computer-Aided Detection model using Deep Convolutional Neural Networks to automatically detect TB from Montgomery County (MC) Tuberculosis radiographs. Our proposed model performed at 87.1% validation accuracy and evaluated using confusion matrix and accuracy as metrics.

Keywords: Chest radiograph · Convolutional neural network · Preprocessing · Feature extraction · Tuberculosis · Classification

1 Introduction

Tuberculosis (TB) is a contagious disease that is considered worldwide as a significant source of death from a single transmittable agent as well as among the top 10 sources of death [1,2]. The TB disease is caused by the mycobacterium tuberculosis bacillus which is easily contractible by having close contact with infected individuals. This disease mostly affects the lungs but can as well affects other parts of the body [3,4].

World Health Organization (WHO) estimated about 10 million individuals fell sick as a result of Tuberculosis disease in 2018 which resulted in about 1.2 million deaths from the previous 1.3 million deaths recorded in 2017 [2].

Tuberculosis disease is more prevalent in developing regions and can affect both males and females but more prominent in males. Among the total number of individuals infected with tuberculosis in 2017, 1 million cases were reported in children aged less than 14, 5.8 million cases in males, and 3.2 million cases in females [1].

A. El Moataz et al. (Eds.): ICISP 2020, LNCS 12119, pp. 151–161, 2020.
https://doi.org/10.1007/978-3-030-51935-3_16

Tuberculosis disease is certainly curable but needs to be detected early for appropriate treatment. Several lung examination techniques are available but the Chest radiographs conversationally known as chest X-ray or CXR for short is a prominent screening tool for detecting abnormalities in the lungs [5,6,14]. Basically TB manifestation can be detected on CXR, however, quality CXR imaging equipment along with skilled radiologists to accurately interpret the CXR is either limited or not available in TB prevailing regions [7,8].

A geographical report by the World Health Organization of most Tuberculosis cases for 2018 is shown in Table 1.

Table 1. Geographical TB cases (WHO, 2018)

Region	Cases (%)
South-East Asia	44
Africa	24
Western Pacific	18
Eastern Mediterranean	8
The Americas	3
Europe	3

Due to the deadly nature of TB disease and the rate at which it can easily be spread, WHO has laid emphasis on more proactive measures for continuous reduction of TB cases and deaths [9]. Also, the decision to embark on a mission to put an end the universal TB epidemic by the year 2030 is underway as contained in the 2019 Global Tuberculosis report [2].

The lack of skilled radiologist, high number of patients waiting in line to be screen and mostly outdated equipment which results in a high rate of errors in properly screening the CXR remain a major problem that requires prompt attention.

As a result, to profer solution to the issue of limited or lack of expert radiologist and misdiagnosis of CXR, we propose a Deep Convolutional Neural Networks (CNN) model that will automatically diagnose large numbers of CXR at a time for TB manifestation in developing regions where TB is most prevalent. The proposed model will eliminate the hassle of patients waiting in line for days to get screened, guarantee better diagnosis, performance accuracy and ultimately minimize cost of screening as opposed to the process of manual examination of the CXR which is costly, time-consuming, and prone to errors due to lack of professional radiologist and huge number of the CXR pilled up to be diagnosed.

2 Related Work

The evolution of computer-aided detection and investigative systems has offered a new boost attributed to the emerging digital CXR and the ability of computer

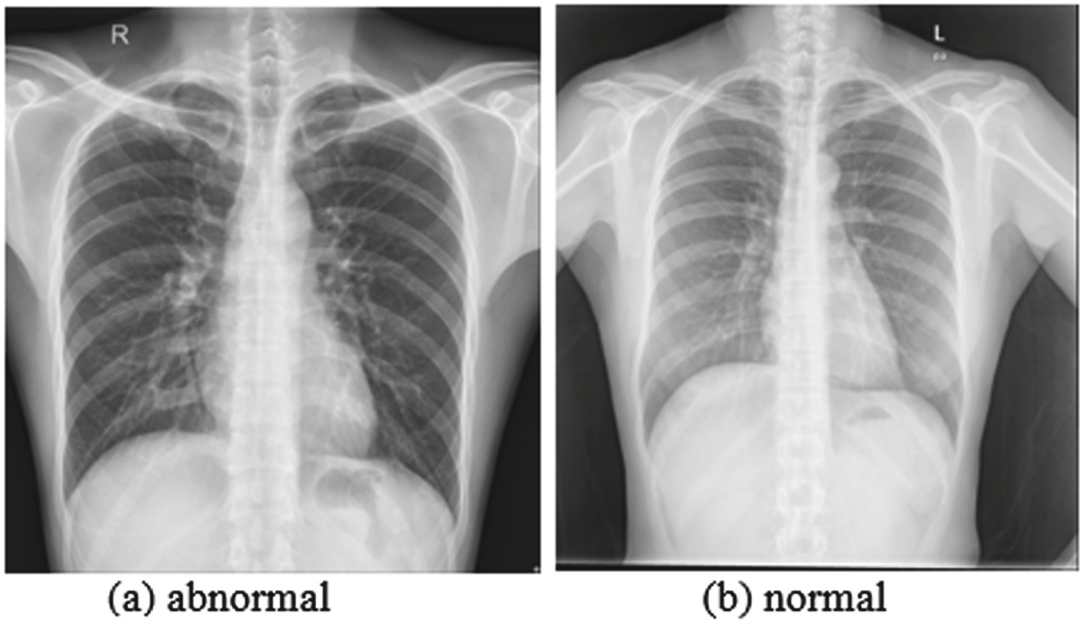


Fig. 1. Sample of the normal and abnormal chest X-ray.

vision for screening varieties of health diseases and conditions. Although much impressive research has been carried out in the last few years regarding computer-aided detection, nevertheless lots more finding is required in the field medical imaging to improve the existing methods and find convenient lasting solutions to deadly medical conditions as the case of Tuberculosis and many more.

A processing method that combines the Local Binary Pattern with laplacian of gaussian was employed in [10] for the manual detection of Tuberculosis nodules in CXR. This research centers on accentuating nodules by the Laplacian of Gaussian filter, lung segmentation, ribs suppression and the use of local binary pattern operators for texture classification. Computer-aided diagnosis system presented by [11] for screening Tuberculosis using two different Convolutional Neural Networks architectures (Alexnet and VGGNet) to classify CXR into positive and negative classes. Their experiment which is based on Montgomery and Shenzhen CXR datasets found VGGNet outperformed Alexnet as a result of a deeper network of VGGNet. The performance accuracy of 80.4% was obtained for Alexnet while VGGNet reached 81.6% accuracy. The authors conclude that improved performance accuracy can be achieved by increasing the dataset size used for the experiment.

One of the first research papers that utilized deep learning techniques on medical images is shown in [12]. The work was based on popular Alexnet architecture and transfer learning for screening the system performance on different datasets. The cross dataset performance analysis carried out shows the system accuracy of 67.4% on the Montgomery dataset and 83.7% accuracy on the Shenzhen dataset. A ConVnet model involving classifications of different manifestations of tuberculosis was presented in [13]. This work looked at unbalanced, less

categorized X-ray scans and incorporate cross-validation with sample shuffling in training their model. The Peruvian Tuberculosis datasets comprising of a total of 4701 image samples with about 4248 samples marked as abnormal containing six manifestation of tuberculosis and 453 samples marked as normal were used for the experiment to obtain 85.6% performance accuracy.

CNN has also been applied by the authors of [15] for extracting discriminative and representative features from X-ray radiographs for the purpose of classifying different parts of the body. This research has exhibited the capabilities of CNN models surpassing traditional hand-crafted method of feature extraction.

An approach based on deep learning for the classification of chest radiographs into positive and negative classes is depicted in [16]. The CNN structure employed in this work consists of seven convolutional layers and three fully connected layers to perform classification experiments. The authors compared three variety optimizers in their experiments and found the Adam optimizer to perform better with validation accuracy of 0.82% and loss of 0.4013.

Other methods that have been utilized for TB detection and classification includes: Support Vector Machine [17,18], K-Nearest Neighbor [19], Adaptive Thresholding, Active Contour Model, and Bayesian Classifier [20], Linear Discriminant Analysis [21].

It is evident from the related work that more effort is required in dealing with the Tuberculosis epidemic that has continued as one of the topmost causes of death. In view of this, we have presented an improved performance validation accuracy concerning to detecting and classifying CXR for TB manifestation.

3 Materials

3.1 Datasets

The Montgomery County (MC) CXR dataset was employed in this research. The MC dataset is a TB specific dataset made available by the National Library of Medicine in conjunction with the Department of Health Services Maryland, U.S.A for research intent. This dataset composed of 58 abnormal samples labeled as "1" and 80 normal samples labeled as "0". All samples are of the size 4020 by 4892 pixels saved as portable network graphic (png) file format as shown in Fig. 1. This dataset is accompanied by clinical readings that give details about each of the samples with respect to sex, age, and manifestations. The dataset can be accessed at https://lhncbc.nlm.nih.gov/publication/pub9931 [23].

3.2 Preprocessing

Since deep neural networks are hugely dependent on large data size to avoid overfitting and achieve high accuracy [22], we performed data augmentation on the MC dataset as a way of increasing the size from 138 samples to 5000 samples. The following types of augmentation were applied: horizontal left and right flip with a probability = 0.3, random zoom = 0.3 with an area = 0.8, top and bottom

flip = 0.3, left and right rotation = 0.5. Other preprocessing task employed here includes image resizing, noise removal and histogram equalization. The data augmentation procedure used in this work is not such that gives room for data redundancy.

4 Proposed Model

A model based on Deep Convolutional Neural Network (CNN) structure has been proposed in this work for the detection and classification of Tuberculosis. CNN models are based upon feed-forward neural network structures for automatic features selection and extraction as a result of taking advantage of the inherent properties of images. The depth of a CNN model has an impact on the performance of the features extracted from an image. CNN models have many layers but the Convolutional layer, MaxPooling, and the Fully connected layer are regarded as the main layers [15]. At the time of model training, diverse parameters are optimized in the convolution layers for extracting meaning features before is been pass on to the fully connected layer where the extracted features are then classified into the target classes which in this case is "normal and abnormal" classes.

Our proposed CNN structure is composed of feature extraction and features classification stages. The feature extraction stage consists of convolution layers, batch normalization, relu activation function, dropout, and max pooling while the classification stage contains the fully connected layer, flatten, dense and a softmax activation function. There are 4 convolution layers in the network for extracting distinct features from the input image with shape $224 \times 224 \times 3$ that is passed to the first convolutional layer learning 64, 3×3 filters, the same as the second convolutional layer. Both the third and fourth convolutional layers learn 128, 3×3 filters. Relu activation function and batch normalization were employed in all the convolutional layers but only the second and fourth layer uses max pooling with a 2×2 pooling size and 25% dropout. The fifth layer which is the fully connected layers output 512 feature that is mapped densely to 2 neurons required by the softmax classifier for classifying our images into normal and abnormal classes. The detail representation of our proposed TB detection model is presented in Table 2.

4.1 Convolution Layer

At each convolution layer, the feature maps from the preceding layer convolute with kernels which are then fed through the ReLu activation function to configure the feature output maps. Also, each output map can be formulated with respect to several input maps. This can be mathematically written as:

$$y_j^i = f\left(\sum_{i \in N_j} y_i^{l-1} * M_{ij}^l + a_j^l\right)$$

where y_j^i depicts the j^{th} output feature of the l^{th} layer, $f(.)$ is a nonlinear function, N_j is the input map selection, y_i^{l-1} refer to the i^{th} input map of $l-1^{th}$ layer, M_{ij}^l is the kernel for the input i and output map M in the j^{th} layer, and a_j^l is the addictive bias associated with the j^{th} map output.

Table 2. Details representation of our proposed TB detection model

Layers	Output size	Filter
INPUT IMAGE	$224 \times 224 \times 3$	
CONVO1	$112 \times 112 \times 64$	3×3
ACTN	$112 \times 112 \times 64$	
BATCHNORM	$112 \times 112 \times 64$	
CONVO2	$112 \times 112 \times 64$	3×3
ACTN	$112 \times 112 \times 64$	
BATCHNORM	$112 \times 112 \times 64$	
MAXPOOL	$56 \times 56 \times 64$	2×2
DROPOUT	$56 \times 56 \times 64$	
CONVO3	$56 \times 56 \times 128$	3×3
ACTN	$56 \times 56 \times 128$	
BATCHNORM	$56 \times 56 \times 128$	
CONV4	$56 \times 56 \times 128$	3×3
ACTN	$56 \times 56 \times 128$	
BATCHNORM	$56 \times 56 \times 128$	
MAXPOOL	$28 \times 28 \times 128$	2×2
DROPOUT	$28 \times 28 \times 128$	
FULLY CONN	512	
ACTN	512	
BATCHNORM	512	
DROPOUT	512	
SOFTMAX	2	

4.2 MaxPooling Layer

MaxPooling layer carryout a downsampling operation of the input map by calculating the maximum activation value in each feature map. The downsampling of MaxPooling is done to partly control overfitting and is formally written as:

$$y_j^l = f\left(\alpha_j^l down(y_i^{l-1} * M_{ij}^l) + a_j^l\right)$$

where α_j^l represent the multiplicative bias of every feature output map j that scale the output back to its initial range, $down(.)$ can be substituted for either $avg(.)$ or $max(.)$ over an $n \times n$ window effectively scaling the input map by n times in every dimension.

Our model is trained using the Stochastic Gradients Descents (SGD) optimizer with an initial learning rate set to 0.001, batch size of 22 samples, momentum equals 0.9, regularization L2 equals 0.0005 to control overfitting and training loss. The SGD optimizer is given below as:

$$\alpha = \alpha - n. \bigtriangledown_\alpha \ J(\alpha; x^{(i)}; y^{(i)})$$

where $\bigtriangledown_\alpha J$ is the gradient of the loss w.r.t α, n is the defined learning rate, α is the weight vector while x and y are the respective training sample and label.

4.3 Softmax Classifier

The Softmax classifier is used to process and classify the features that have been extracted from the convolutional stage. Softmax determine the probability of the extracted features and classify them into normal and abnormal classes defined as:

$$\sigma(q)_i = \frac{e^{q_i}}{\sum_{k=1}^{k} e_k^q}$$

where q mean the input vector to the output layer i that is depicted from the exponential element.

The structure of our model is presented in Fig. 2.

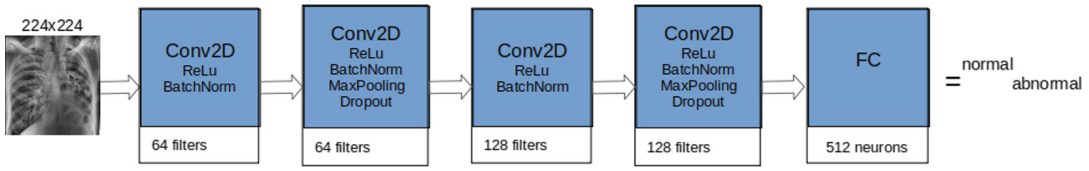


Fig. 2. Architecture of our ConvNet model.

5 Result

Training and validation of our model were performed on the Montgomery County (MC) TB specific dataset. The dataset which originally consists of 138 normal and abnormal samples were augmented to 5000 samples and split into 3,750 (75%) for training and 1,250 (25%) validation. The samples in the training set are indiscriminately selected i.e (shuffled) during the training to ensure features are extracted across all samples. Of course, the samples in the training set are

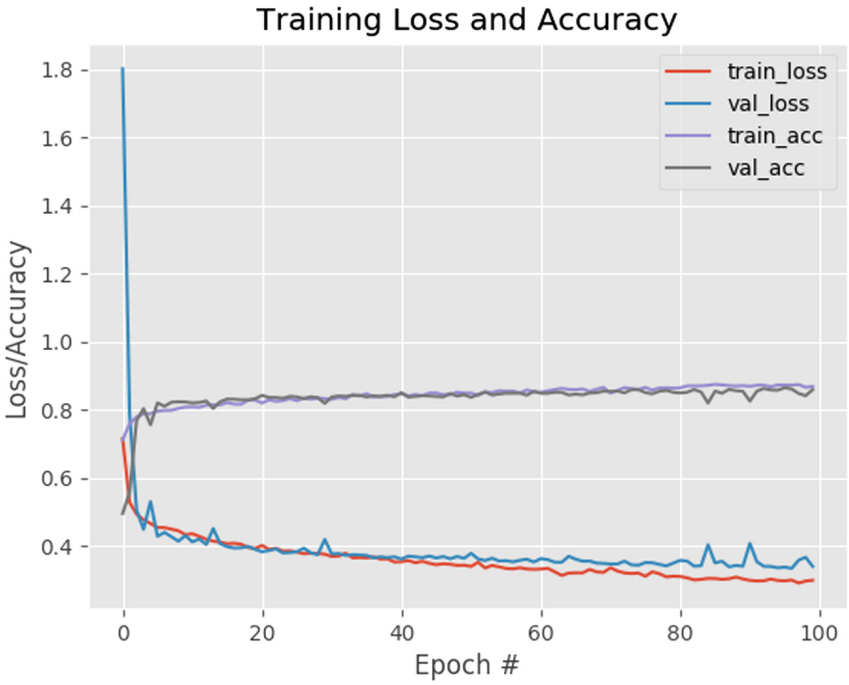


Fig. 3. Model accuracy and loss

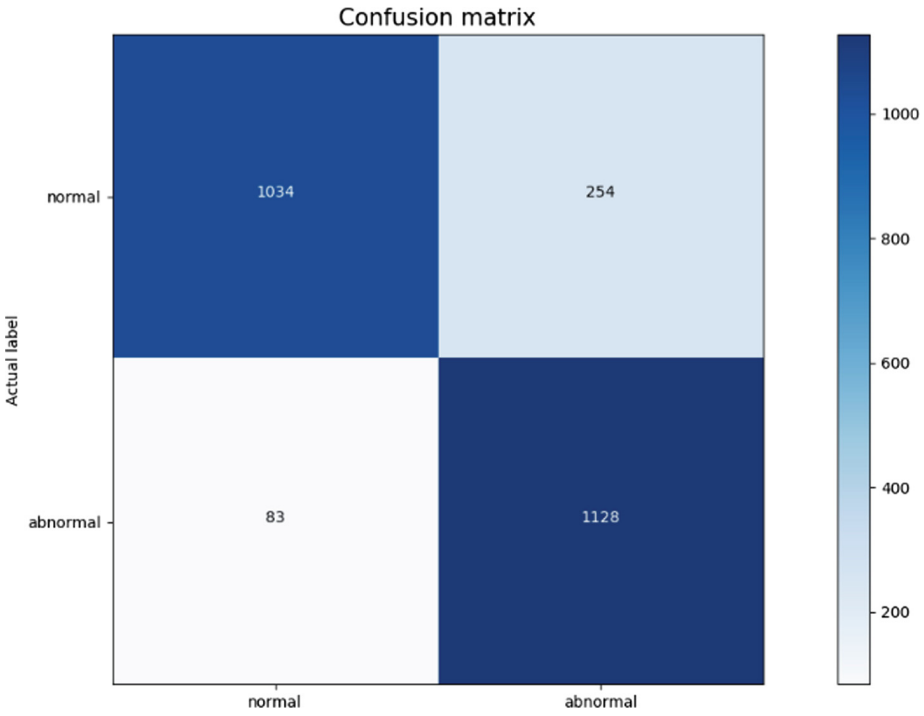


Fig. 4. Model confusion matrix

Table 3. Proposed model as compared

Ref	Accuracy (%)
23	82.6, 84.7
12	83.7
11	80.4, 81.6
16	82
13	85
24	83.7
Proposed	**87.1**

entirely different from the samples contained in the validation set which are samples that the model has not seen before during model training. We use confusion matrix and accuracy as the metrics for evaluating the model. The 224×224 sample is fed into the model which was trained for 100 iterations with a batch size of 22 for SGD optimizer at an initial 0.001 learning rate. The model employs cross entropy loss function for weight update at every iteration and use Softmax function for processing and classifying the samples into normal and abnormal classes labeled 0 and 1. The proposed model achieved 87.1% validation accuracy. The Performance accuracy of the model and confusion matrix are presented in Figs. 3 and 4.

6 Conclusion

Presented in this paper is a model that aids early detection of Tuberculosis using CNN structure to automatically extract distinctive features from chest radiographs and classify them into normal and abnormal categories. We did not test other architectures in this research; instead, their performance accuracy is reported in Table 3. The histogram equalizer we applied to enhance the visibility of the data samples, which makes the extracted features more evident, is one of the contributing factors responsible for the improved performance. We will consider the Shenzhen, JSRT, KIT, and Indiana datasets in our future work while we continue to aim for optimal performance accuracy.

References

1. World Health Organization: Global tuberculosis report. WHO (2018)
2. World Health Organization: Global tuberculosis report. WHO (2019)
3. Zaman, K.: Tuberculosis: a global health problem. J. Health Popul. Nutr. **28**(2), 111 (2010)
4. Baral, S.C., Karki, D.K., Newell, J.N.: Causes of stigma and discrimination associated with tuberculosis in Nepal: a qualitative study. BMC Public Health **7**(1), 211 (2007)

5. World Health Organization: Tuberculosis prevalence surveys: a handbook. Report No.: WHO/HTM/TB/2010.17. World Health Organization, Geneva (2011)
6. World Health Organization: Chest radiography in tuberculosis detection: summary of current WHO recommendations and guidance on programmatic approaches (No. WHO/HTM/TB/2016.20). World Health Organization (2016)
7. Lakhani, P., Sundaram, B.: Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. Radiology **284**(2), 574–582 (2017)
8. Pedrazzoli, D., Lalli, M., Boccia, D., Houben, R., Kranzer, K.: Can tuberculosis patients in resource-constrained settings afford chest radiography? Eur. Respir. J. **49**(3), 1601877 (2017)
9. Lönnroth, K., Migliori, G.B., Abubakar, I., et al.: Towards tuberculosis elimination: an action framework for low-incidence countries. Eur. Respir. J. **45**, 928–952 (2015)
10. Leibstein, J.M., Nel, A.L.: Detecting tuberculosis in chest radiographs using image processing techniques. University of Johannesburg (2006)
11. Rohilla, A., Hooda, R., Mittal, A.: TB detection in chest radiograph using deep learning architecture. In: Proceeding of 5th International Conference on Emerging Trends in Engineering, Technology, Science and Management (ICETETSM-2017), pp. 136–147 (2017)
12. Hwang, S., Kim, H.E., Jeong, J., Kim, H.J.: A novel approach for tuberculosis screening based on deep convolutional neural networks. In: Computer-Aided Diagnosis, vol. 9785, p. 97852W. International Society for Optics and Photonics (2016)
13. Liu, C., et al.: TX-CNN: detecting tuberculosis in chest X-ray images using convolutional neural network. In: IEEE International Conference on Image Processing (ICIP), pp. 2314–2318 (2017)
14. Konstantinos, A.: Testing for tuberculosis. Aust. Prescr. **33**(1), 12–18 (2010)
15. Srinivas, M., Roy, D., Mohan, C.K.: Discriminative feature extraction from X-ray images using deep convolutional neural networks. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 917–921 (2016)
16. Hooda, R., Sofat, S., Kaur, S., Mittal, A., Meriaudeau, F.: Deep-learning: a potential method for tuberculosis detection using chest radiography. In: IEEE International Conference on Signal and Image Processing Applications (ICSIPA), pp. 497–502 (2017)
17. Jaeger, S., Karargyris, A., Antani, S., Thoma, G.: Detecting tuberculosis in radiographs using combined lung masks. In: Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 4978–4981 (2012)
18. Karargyris, A., et al.: Combination of texture and shape features to detect pulmonary abnormalities in digital chest X-rays. Int. J. Comput. Assist. Radiol. Surg. **11**(1), 99–106 (2015). https://doi.org/10.1007/s11548-015-1242-x
19. Van Ginneken, B., Katsuragawa, S., ter Haar Romeny, B.M., Doi, K., Viergever, M.A.: Automatic detection of abnormalities in chest radiographs using local texture analysis. IEEE Trans. Med. Imaging **21**(2), 139–149 (2002)
20. Shen, R., Cheng, I., Basu, A.: A hybrid knowledgeguided detection technique for screening of infectious pulmonary tuberculosis from chest radiographs. IEEE Trans. Biomed. Eng. **57**(11), 2646–2656 (2010)
21. Hogeweg, L., Mol, C., de Jong, P.A., Dawson, R., Ayles, H., van Ginneken, B.: Fusion of local and global detection systems to detect tuberculosis in chest radiographs. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010. LNCS, vol. 6363, pp. 650–657. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-15711-0_81

22. Shorten, C., Khoshgoftaar, T.M.: A survey on image data augmentation for deep learning. J. Big Data **6**(1) (2019). Article number. 60. https://doi.org/10.1186/s40537-019-0197-0
23. National Center for Biomedical Communications. http://lhncbc.nlm.nih.gov

3.1.2 Conclusion

As presented in the paper, the proposed CNN-based method achieved 2.1% improvement over other compared methods. The improved performance was due to the data augmentation, parameter tuning, and pre-processing techniques employed.

3.2 Tuberculosis Abnormality Detection in Chest X-Rays: A Deep Learning Approach

3.2.1 Introduction

The research paper presented in this section is an improvement on the publication in Section 3.1. Consequent to the understanding that the depth of a CNN model and large number of datasets contributes to high performance, we developed a six-layered end-to-end CNN model and used the Shenzhen dataset, which is four times the size of the Montgomery set in Section 3.1 to screen for TB manifestations. The experimental results showed a significant improvement which justified the theory that depth and large number of data samples are essential to improving performance.

The research paper was presented at [4] and published in [5], while the Shenzhen dataset can be downloaded at [6].

[4] International Conference on Computer Vision and Graphics ICCVG 2020. September 14-16, 2020, Warsaw, Poland.

[5] **Mustapha Oloko-Oba and Serestina Viriri**. "Tuberculosis Abnormality Detection in Chest X-Rays: A Deep Learning Approach". *Computer Vision and Graphics*. LNCS Springer, vol 12334, pp 121-132 **(2020)**. DOI: `https://doi.org/10.1007/978-3-030-59006-2_11`

[6] `https://lhncbc.nlm.nih.gov/publication/pub9931`

Tuberculosis Abnormality Detection in Chest X-Rays: A Deep Learning Approach

Mustapha Oloko-Oba and Serestina Viriri( )

School of Mathematics, Statistics and Computer Sciences,
University of KwaZulu-Natal, Durban, South Africa
219098624@stu.ukzn.ac.za, viriris@ukzn.ac.za

Abstract. Tuberculosis has claimed many lives, especially in developing countries. While treatment is possible, it requires an accurate diagnosis to detect the presence of tuberculosis. Several screening techniques exist and the most reliable is the chest X-ray but the necessary radiological expertise for accurately interpreting the chest X-ray images is lacking. The task of manual examination of large chest X-ray images by radiologists is time-consuming and could result in misdiagnosis as a result of a lack of expertise. Hence, a computer-aided diagnosis could perform this task quickly, accurately and drastically improve the ability to diagnose correctly and ultimately treat the disease earlier. As a result of the complexity that surrounds the manual diagnosis of chest X-ray, we propose a model that employs the use of learning algorithm (Convolutional Neural Network) to effectively learn the features associated with tuberculosis and make corresponding accurate predictions. Our model achieved 87.8% accuracy in classifying chest X-ray into abnormal and normal classes and validated against the ground-truth. Our model expresses a promising pathway in solving the diagnosis issue in early detection of tuberculosis manifestation and, hope for the radiologists and medical healthcare facilities in the developing countries.

Keywords: Tuberculosis · Chest X-ray · Classification · Deep learning · Pulmonary · Convolutional neural network

1 Introduction

Tuberculosis, known as TB, is regarded as a significant health challenge across the world but more prevalent in developing countries [1–3]. TB is caused by the Mycobacterium tuberculosis bacteria that generally affect the lungs (pulmonary), and various parts of the body (extrapulmonary) [4]. Several TB patients lose their lives yearly as a result of diagnosis error and lack of treatments [1,5].

Pulmonary Tuberculosis is a transmissible infection mostly visible around the collarbone. It is contractible via the air when people who have an active

L. J. Chmielewski et al. (Eds.): ICCVG 2020, LNCS 12334, pp. 121–132, 2020.
https://doi.org/10.1007/978-3-030-59006-2_11

tuberculosis infection sneeze, cough, or transmit their saliva through the air or other means [6]. In other words, persistence and closeness of contact are major factors of the risk of contracting TB which make individuals inhabiting in the same household at higher risk of contracting TB than casual contacts. Lack of diagnosis and treatment of affected individuals will as well increase the rate of transmission.

Tuberculosis disease is certainly treatable if identified/diagnosed early for appropriate treatment. Diverse test procedures may be employed to confirm a diagnosis of suspected pulmonary tuberculosis. These test techniques include a Computed Tomography (CT) scan, Chest X-ray (CXR), Magnetic Resonance Imaging (MRI) scan or ultrasound scan of the affected part of the body. Among the various screening techniques in existence, chest X-Ray is renowned for evaluating the lungs, heart and chest wall to diagnose symptoms as reported by [7] and recommend their application for screening patients in order to exclude the usual instance of a costly test.

World Health Organization affirms that Tuberculosis is, in fact, one of the topmost causes of death across the world most especially in developing countries. As reported in 2016, 1.8 million deaths were recorded resulting from about 10.4 million cases of people affected with tuberculosis. As much as tuberculosis can be detected on chest X-ray, tuberculosis prevalent regions usually suffer the expertise of radiologists required to accurately diagnose and or interpret the X-ray results [2].

In this paper, we propose a model that will accurately improve the quality and timely diagnosis of chest X-rays for the manifestation of tuberculosis. The proposed model will increase diagnosis efficiency, enhance performance and eventually minimize cost as opposed to the process of manually examining the chest X-ray scan which is costly, time-consuming, and prone to errors due to lack of professional radiologist and volume of the chest X-rays.

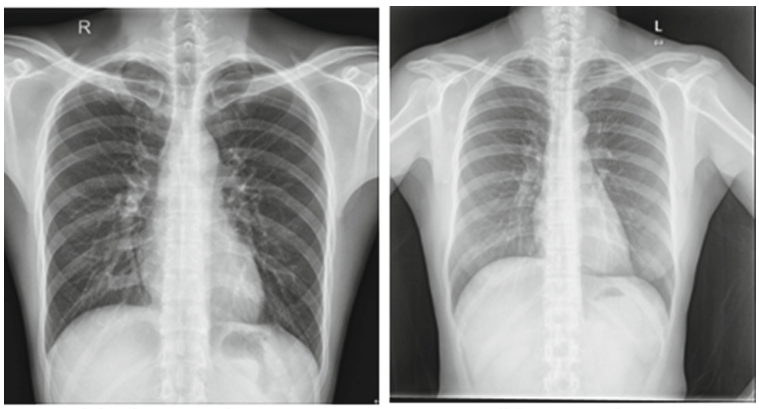


Fig. 1. Sample of the normal and abnormal chest X-rays.

2 Related Work

The strengths and possibilities offered by machine learning has provided a boost to computer vision especially for the diagnosis and screening of various health diseases and conditions. In spite of the global application of machine learning techniques in the biomedical domain, chest X-ray is still a very important and renowned tool [7] among others for evaluating pulmonary diseases that require rapt attention. There is always a need to improve the existing methods and proposing new techniques for stability, global growth, and better performance.

A handful of investigation has been done by [8–10] in assessing the ability of the existing computer-aided detection CAD systems in biomedical domain to diagnose pulmonary nodules. Notable improvement of 0.986 area under curve (AUC) was reported in [8] with the use of a computer-aided design system as opposed to 0.924 AUC without CADs. Similarly, 0.923 AUC improvement with CADs was observed by [9] over 0.896 AUC without CADs. The results of their assessment ultimately show the significant impact of CADs in assisting radiologists to improve the diagnosis from chest X-rays.

An experiment for tuberculosis screening was presented in [24] using Alexnet and VGGNet architectures for the classification of CXR into positive and negative classes. The analysis carried out on the Montgomery and Shenzhen CXR datasets reported that VGGNet outperformed Alexnet as a result of a deeper network of VGGNet. The performance accuracy of 80.4% was obtained for Alexnet while VGGNet reached 81.6% accuracy. The authors concluded that improved performance accuracy is possible by increasing the dataset size used for the experiment.

Another article that deals with detecting the presence of tuberculosis early is presented in [14] which applied a median filter, histogram equalization and homomorphic filter to preprocess the input image before segmentation is applied utilizing the active contour and finally arriving at the classification using the mean values. Although, the research does not report any accuracy attained but affirm the impact and contribution of computer-aided diagnosis as an assisting tool to guide radiologists and doctors in reaching an accurate and timely diagnosis decision with respect to tuberculosis detection.

The study carried out in [11] presented proposals that will improve feature extractors in detecting diseases using pre-trained CNN. This research is famous for combining multiple instance learning algorithms with pre-trained CNN, assessment of classifiers trained on features extracted, and the comparison of performance analysis of existing models for extracting features from the radiograph dataset and achieved an accuracy of 82.6%.

An experiment was done [15] to compare the performance of three different methods in detecting tuberculosis in the lungs. The obtained results show the K-nearest neighbor (KNN) classifier with the maximum accuracy of 80%, Simple linear regression 79%, and the Sequential minimal optimizer 75% accuracy.

An approach to discover tuberculosis from a radiograph image where the X-ray machine is placed behind the patients (posteroanterior) was presented in [12]. This study employs a graph cut segmentation approach for extracting the

lung region on the chest X-ray and then computes sets of features such as (edge, texture, and shape) that allow classification into abnormal and normal classes respectively.

A CNN model that involved classification of different manifestations of tuberculosis was presented in [19]. This work looked at unbalanced and less categorized X-ray scans and incorporated cross-validation with sample shuffling in training the model and reported to have obtained 85.6% accuracy in classifying the various manifestation of tuberculosis using the Peruvian datasets composed of a total of 4701 samples with about 4248 samples labeled as abnormal containing six manifestation of tuberculosis and 453 samples labeled as normal. The authors affirm that their model surpasses previous classification models and is promising in tuberculosis diagnosis.

The research work presented in [13] is one of the first to employ deep learning techniques on medical images. The work was based on popular Alexnet architecture and transfer learning for screening the system performance on different datasets. The cross dataset performance analysis carried out shows the system accuracy of 67.4% on the Montgomery dataset and 83.7% accuracy on the Shenzhen dataset.

In [25], the authors participated in the 2019 ImageCLEF challenge and presented a deep learner model (LungNet) that focus on automatic analysis of tuberculosis from computer tomography CT scans. The CT scans employed is firstly decompressed and the slices are extracted having 512 images for the X and Y dimensions and between 40–250 images for the Z dimension. Filters were then introduced on the slices to eliminate the slices that do not contain valuable information required for classifying the samples. The proposed LungNet along with ResNet-50 architecture was employed as a deep learner whose outputs are regarded as the preliminary results. The deep learner model was trained on 70(%) and 50(%) training sets and achieved AUC performance of 63(%) and 65(%) for the ImageCLEF CT report and severity scoring task respectively. Although, these performances were not the best presented in the challenge but the authors believed if subjected to advanced pre-processing techniques such as data augmentation, and masking could provide a better performance.

CheXNeXt is an algorithm developed by the authors in [26], for identification of 14 various pathologies in chest X-rays. The algorithm which employed convolutional neural network approach was validated on the NIH dataset and compared the result with interpretation of 9 professional radiologists. The results show that CheXNeXt achieved equivalent performance with radiologists in 10 different pathologists, best performance on 1 pathology (atelectasis) attaining 0.862(%) AUC and underperformed in 3 pathologists. The algorithm took less than 2 min to identify the various pathologists while it took the radiologist about 240 min.

The study in [27] is conducted to assess the detection accuracy of qXR, a computer aided diagnosis software based on convolutional neural network. The authors utilized microbiologically established lung tuberculosis images as the standard for reference and made use of kappa coefficient along with confidence

interval as the statistical tools to analyse the data and examine the inter-rater reliability of radiologist in detecting certain lung abnormalities. The study also used radiologist interpretation as standard to validate the detection accuracy of the qXR in terms of generating ROC curves and calculating AUC. The qXR system achieved 0.81(%) AUC for detection of lung tuberculosis, 71(%) sensitivity and 80(%) specificity.

As with existing models, most of which basically show the performance accuracy of their models without a view of the predicted sample. We, however in this work present a model evaluated on the Shenzhen datasets to provide an improved performance accuracy thereby showing predictions of the model validated on the groundtruth data. The groundtruth data will ultimately assist us to further develop a tuberculosis diagnosis system to be deployed to health facilities in the developing countries to improve quality and timely diagnosis.

3 Methods and Techniques

3.1 Datasets

The dataset used in training our model is the Shenzhen tuberculosis dataset. This dataset is specific to tuberculosis and is publicly available for the purpose of research. The Shenzhen dataset is made up of 336 abnormal samples labeled as "1" and 326 normal samples labeled as "0". All samples are of the size 3000 by 3000 pixels saved as portable network graphic (png) file format as shown in Fig. 1. This dataset is accompanied by clinical readings that gives details about each of the samples with respect to sex, age, and diagnosis. The dataset can be accessed at https://lhncbc.nlm.nih.gov/publication/pub9931.

3.2 Convolutional Neural Networks

A broadly used deep learning algorithm for object classification, detection, and video analyzing in computer vision is the Convolutional Neural Network otherwise known as ConvNet or CNN. ConvNet is able to distinguish one object

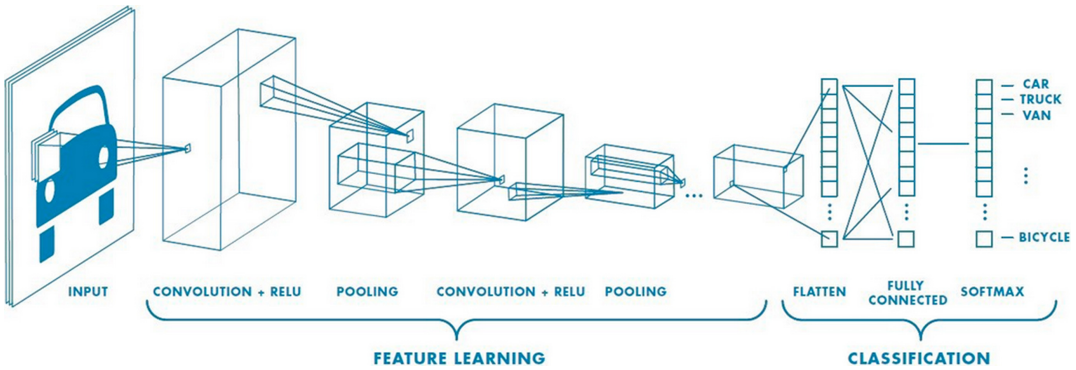


Fig. 2. CNN architecture extracted from [18].

from another by assigning important learnable biases and weights to various images [19]. ConvNets which are feed-forward neural networks are famous for performing better than other classification networks as a result of taking advantage of the inherent properties of the images. The required preprocessing in a convolutional neural network is less compared to other classification neural networks. The ConvNet is made of an input object, features learning layers such as (conv2d, pooling, normalization) with ReLu activation function and finally the classification layers such as (flatten, fully connected, dense) with softmax or sigmoid activation functions as shown in Fig. 2.

3.3 Preprocessing

ConvNet models performs better with large dataset [17], hence we applied augmentation [16] as a preprocessing methods to increase the size of data for better performance. The augmentation types we applied to enlarge the size of our data are: flip_left_right with a probability = 0.5, zoom_random = 0.5, with an area = 0.8, and flip_top_bottom = 0.5, rotate left_right = 5. With the use of data augmentation, we were able to enlarge our dataset to 10,000 samples.

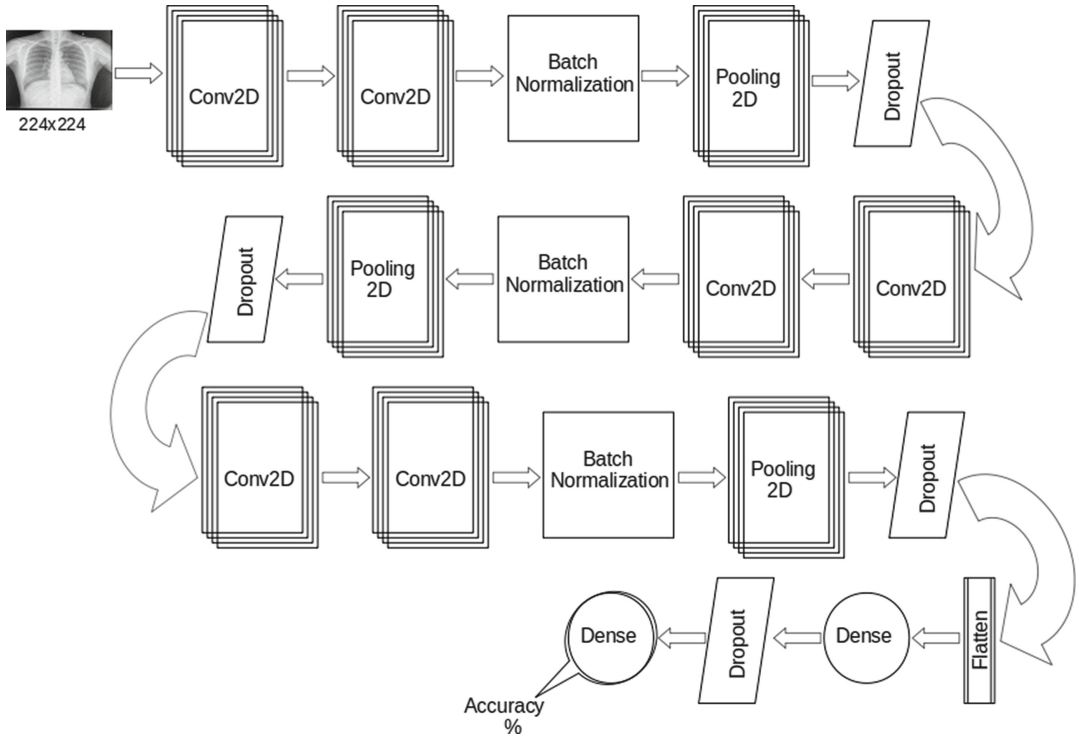


Fig. 3. Architecture of our 6-layer network for tuberculosis classification.

3.4 Model Architecture

Our ConvNet model is a sequential model that is made up of 6 convolutional layers, batch normalization, dense layer sas shown in Fig. 3. Each layer in the model use the Rectified Linear Unit (ReLu) activation function except the second dense layer which serves as the output layer and used the Softmax activation function written mathematically as:

$$\beta(v)_i = \frac{e^{v_i}}{\sum_{n=1}^{n} e_n^v}$$

where v mean the input vector to the output layer i that is depicted from the exponential element.

The input shape is $224 \times 224 \times 3$ which is fed to the 1st convolutional layer using a $32 \times 3 \times 3$ filters, the output from the first layer is pass to the 2nd convolutional layer which also makes use of a $32 \times 3 \times 3$ filters. This is followed by the next two convolutional layers which employ a $64 \times 3 \times 3$ filters and the output from here is fed to the final 2 convolutional layers with a $128 \times 3 \times 3$ filters respectively. The output from the last convolutional layer is normalized, flatten and pass to the dense layer with 4096 neurons. A regularization technique dropout of (0.5) was added to control overfitting before passing to the last dense layer which is the output layer that uses softmax activation with 2 neurons.

The activities in the convolutional layer are given as:

$$p_j^i = f\left(\sum\nolimits_{i \in N_j} p_i^{l-1} * Q_{ij}^l + \alpha_j^l\right)$$

where p_j^i depicts the j^{th} output feature of the l^{th} layer, $f(.)$ is a nonlinear function, N_j is the input map selection, p_i^{l-1} refers to the i^{th} input map of $l-1^{th}$ layer, Q_{ij}^l is the kernel for the input i and output map Q in the j^{th} layer, and α_j^l is the addictive bias associated with the j^{th} map output.

3.5 Training

The dataset was split into a 70% training set, 10% validation set and 20% testing set and compile our model using cross-entropy loss function and the Stochastic Gradient Descent (SGD) optimizer along with varieties of hyper-parameters. The initial learning rate was set to = 0.01, momentum = 0.9 with regularization L2 = 0.0004 as a starting point to control overfitting and training loss and we measure our metrics by the accuracy obtained from the best experiment. The SGD optimizer is given below as:

$$\beta = \beta - x. \bigtriangledown_\beta J(\beta; n^{(i)}; m^{(i)})$$

where $\bigtriangledown_\beta J$ is the gradient of the loss w.r.t β, x is the defined learning rate, β is the weight vector while n and m are the respective training sample and label.

The model was trained using Keras deep learning framework running on tensorflow backend. The training is done entirely in the cloud on a Telsa GPU provided by Google which is available via google col laboratory notebook platform. Collaboratory is a free cloud Jupyter notebook platform that provides researchers with free GPU up to 12 GB of random access memory (RAM) and 360 GB of hard disc.

4 Results

We measure the performance of our model using the accuracy metric which determined the ratio of the correctly predicted samples from the entire predictions made. This metric is given as:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

where

TP: is the case where the model predicts a sample as positive and it is actually positive
TN: is the case where the model predicts a sample as negative and it is actually negative
FP: is the case where the model predicts a sample as positive whereas the actual output is negative
FN: is the case where the model predicts a sample as negative whereas the actual output is positive

The training and validation of our proposed model were performed using the Shenzhen TB specific dataset. The dataset which consists of a total of 662 samples were augmented to 10,000 samples out of which 70% were used for training while 30% was utilized for testing validation. The training samples were shuffled during model training to make certain that both positive and negative samples were included in the training to avoid biases. We performed several experiments and obtained 87.8% performance accuracy.

The initial experiment was performed with 662 samples and obtained 79(%) accuracy whereas after applying some image processing techniques, our model achieved 87.8(%) accuracy. We applied contrast limited adaptive histogram equalization on the chest X-ray samples to enhance the image quality which made features identification and extraction distinctly clearer and obvious to the classification module for accurate classification.

The data augmentation and CLAHE techniques were some factors responsible for obtaining an improved performance. Samples of the enhance images is presented in Fig. 6 which shows obvious clarity compared to the samples in Fig. 1.

The loss and performance accuracy of the model is presented in Fig. 4. In Table 1 is the summary of the existing and proposed CNN models with their accuracy while Fig. 5 shows few predictions of the proposed model.

Table 1. Summary of tuberculosis detection models showing the various datasets, assessment measure and the performance accuracy

Authors ref	Datasets	Evaluation metric	Accuracy (%)
[13]	Montgomery County and Shenzhen	Accuracy	67.4, 83.7
[11]	Shenzhen and Montgomery County	Accuracy	84.7, 82.6
[19]	Peruvian partners at "Socios en Salud". (Peru)	Accuracy	85.6
[24]	Montgomery County and Shenzhen	Accuracy	80.4, 81.6
[25]	Custom CT scan	Accuracy, AUC	63, 65
[26]	NIH Chest X-ray14	AUC	86.2
[27]	Custom: Chest x-ray at Kasturba hospital Manipal, India	AUC, Sensitivity, Specificity	81, 71, 80
Proposed	**Shenzhen**	**Accuracy**	**87.8**

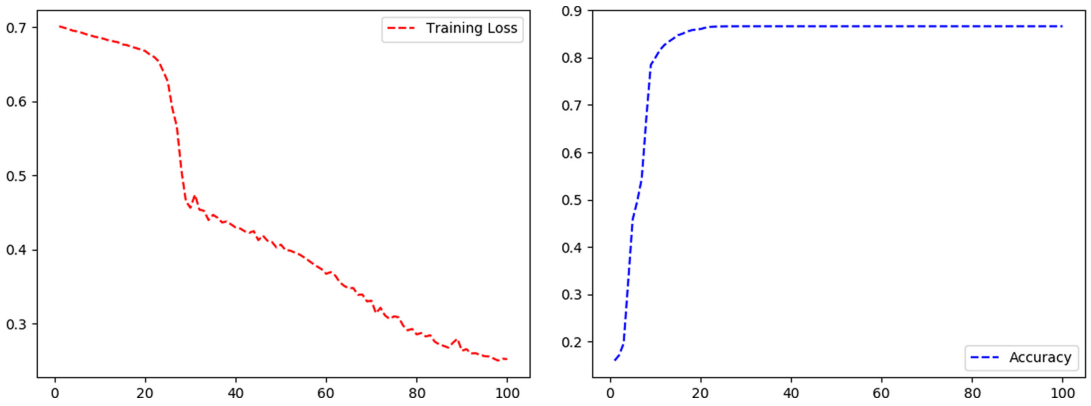


Fig. 4. Training loss and accuracy

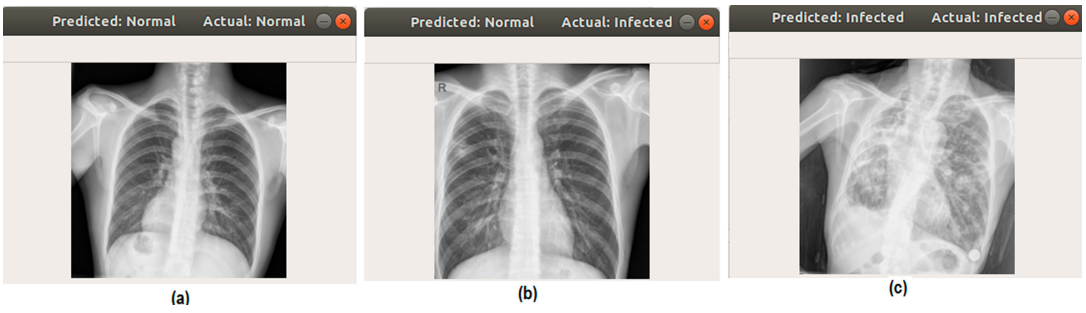


Fig. 5. Model predictions. These are few predictions of our model. (a) and (c) are samples that were correctly predicted while (b) is one of the very few misclassified sample that will be improve in future work

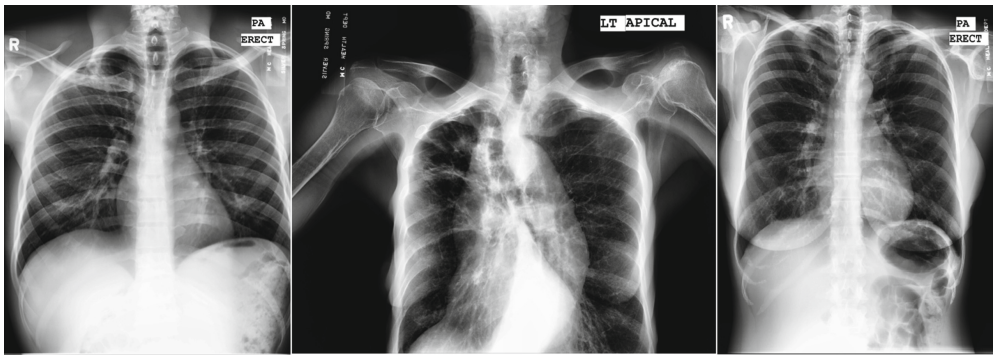


Fig. 6. Samples of the enhanced chest X-ray

5 Conclusion

We presented a ConvNet model in this paper for improving the diagnosis of pulmonary tuberculosis from chest X-ray images for prompt and effective treatment. The ConvNet model presented here achieved 87.8% classification accuracy and further shows the prediction. Comparing the predicted samples with the ground-truth has expressed an impressive performance which is promising to the thought of providing lasting solutions to the early detection of tuberculosis most especially in developing countries where tuberculosis is most prevalent.

As for the future advancement in the timely and accurate diagnosis of a large chest X-rays for the detection of tuberculosis and other pulmonary diseases. We intend to:

- Improve on our model to reach 100(%) accuracy and ultimately deploy a tuberculosis diagnosis system to health facilities where radiological expertise is lacking and hindering the quality of diagnosis and delay in diagnosing large number of chest X-ray images.
- Develop an algorithm that could distinctly identify foreign objects like rings, buttons, pieces of bone, and coins that may be found on chest x-ray images. This foreign objects can results in misclassification and will impede performance of detection systems. For instance round objects like buttons, bone piece or rings can be mistaken for nodules.
- Deploy a computer-aided detection systems to simultaneously detect different pulmonary diseases.

References

1. Zumla, A., George, A., Sharma, V., Herbert, R.H.N., Oxley, A., Oliver, M.: The WHO 2014 global tuberculosis report-further to go. Lancet Glob. Health **3**(1), e10–e12 (2015)
2. Lakhani, P., Sundaram, B.: Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. Radiology **284**(2), 574–582 (2017)

3. Sathitratanacheewin, S., Pongpirul, K.: Deep learning for automated classification of tuberculosis-related chest X-ray: dataset specificity limits diagnostic performance generalizability. arXiv preprint arXiv:1811.07985 (2018)
4. Baral, S.C., Karki, D.K., Newell, J.N.: Causes of stigma and discrimination associated with tuberculosis in Nepal: a qualitative study. BMC Public Health **7**(1), 211 (2007)
5. Hooda, R., Sofat, S., Kaur, S., Mittal, A., Meriaudeau, F.: Deep-learning: a potential method for tuberculosis detection using chest radiography. In: 2017 IEEE International Conference on Signal and Image Processing Applications (ICSIPA), pp. 497–502 (2017)
6. Bhatt, M.L.B., Kant, S., Bhaskar, R.: Pulmonary tuberculosis as differential diagnosis of lung cancer. South Asian J. Cancer **1**(1), 36 (2012)
7. World Health Organization: Chest radiography in tuberculosis detection: summary of current WHO recommendations and guidance on programmatic approaches (No. WHO/HTM/TB/2016.20). World Health Organization (2016)
8. Kakeda, S., et al.: Improved detection of lung nodules on chest radiographs using a commercial computer-aided diagnosis system. Am. J. Roentgenol. **182**(2), 505–510 (2004)
9. Sakai, S., et al.: Computer-aided nodule detection on digital chest radiography: validation test on consecutive T1 cases of resectable lung cancer. J. Digit. Imaging **19**(4), 376–382 (2006). https://doi.org/10.1007/s10278-006-0626-4
10. Shiraishi, J., Abe, H., Li, F., Engelmann, R., MacMahon, H., Doi, K.: Computer-aided diagnosis for the detection and classification of lung cancers on chest radiographs: ROC analysis of radiologists' performance. Acad. Radiol. **13**(8), 995–1003 (2006)
11. Lopes, U.K., Valiati, J.F.: Pre-trained convolutional neural networks as feature extractors for tuberculosis detection. Comput. Biol. Med. **89**, 135–143 (2017)
12. Jaeger, S., et al.: Automatic tuberculosis screening using chest radiographs. IEEE Trans. Med. Imaging **33**(2), 233–245 (2013)
13. Hwang, S., Kim, H.E., Jeong, J., Kim, H.J.: A novel approach for tuberculosis screening based on deep convolutional neural networks. In: International Society for Optics and Photonics: Computer-Aided Diagnosis, vol. 9785, p. 97852W (2016)
14. Gabriella, I.: Early detection of tuberculosis using chest X-Ray (CXR) with computer-aided diagnosis. In: 2018 2nd International Conference on Biomedical Engineering (IBIOMED), pp. 76–79 (2018)
15. Antony, B., Nizar Banu, P.K.: Lung tuberculosis detection using x-ray images. Int. J. Appl. Eng. Res. **12**(24), 15196–15201 (2017)
16. Hussain, Z., Gimenez, F., Yi, D., Rubin, D.: Differential data augmentation techniques for medical imaging classification tasks. In: AMIA Annual Symposium Proceedings. AMIA Symposium, vol. 2017, pp. 979–984 (2018)
17. Tajbakhsh, N., et al.: Convolutional neural networks for medical image analysis: full training or fine tuning? IEEE Trans. Med. Imaging **35**(5), 1299–1312 (2016)
18. Saha, S.: A comprehensive guide to convolutional neural networks-the ELI5 way (2018)
19. Liu, C., et al.: TX-CNN: detecting tuberculosis in chest X-ray images using convolutional neural network. In: 2017 IEEE International Conference on Image Processing (ICIP), pp. 2314–2318 (2017)
20. Basheer, S., Jayakrishna, V., Kamal, A.G.: Computer assisted X-ray analysis system for detection of onset of tuberculosis. Int. J. Sci. Eng. Res. **4**(9), 2229–5518 (2013)

21. Le, K.: Automated detection of early lung cancer and tuberculosis based on X-ray image analysis. In Proceedings of the WSEAS International Conference on Signal, Speech and Image Processing, pp. 1–6 (2006)
22. Khutlang, R., et al.: Classification of Mycobacterium tuberculosis in images of ZN-stained sputum smears. IEEE Trans. Inf. Technol. Biomed. **14**(4), 949–957 (2009)
23. Alcantara, M.F., et al.: Improving tuberculosis diagnostics using deep learning and mobile health technologies among resource-poor communities in Peru. Smart Health **1**, 66–76 (2017)
24. Rohilla, A., Hooda, R., Mittal, A.: TB detection in chest radiograph using deep learning architecture. In: Proceeding of 5th International Conference on Emerging Trends in Engineering, Technology, Science and Management (ICETETSM-17), pp. 136–147 (2017)
25. Hamadi, A., Cheikh, N.B., Zouatine, Y., Menad, S.M.B., Djebbara, M.R.: ImageCLEF 2019: deep learning for tuberculosis CT image analysis. In: CLEF2019 Working Notes, vol. 2380, pp. 9–12 (2019)
26. Rajpurkar, P., et al.: Deep learning for chest radiograph diagnosis: a retrospective comparison of the CheXNeXt algorithm to practicing radiologists. PLoS Med. **15**(11), e1002686 (2019)
27. Nash, M., et al.: Deep learning, computer-aided radiography reading for tuberculosis: a diagnostic accuracy study from a tertiary hospital in India. Sci. Rep. **10**(1), 1–10 (2020)

3.2.2 Conclusion

The techniques employed in this paper improved the diagnosis accuracy by about 0.7%. The predicted samples, when compared with the ground-truth data, expressed promising attributes in solving TB related problems. The model generalized well on the test images, although it recorded some misclassification, forming the basis for the next framework. The misclassified samples could result from inadequate hyper-parameter tuning and limited features learned from the training samples to identify certain features from the CXR images.

3.3 Pre-trained Convolutional Neural Network for the Diagnosis of Tuberculosis

3.3.1 Introduction

The research work presented here improved the research published in Section 3.2, despite knowing that the depth of a model contributes to its performance, as evident in the model presented in Section 3.2 to achieve 87.8% classification accuracy. This accuracy was above the results obtained with previous models in literature but was not satisfactory due to the life-threatening nature of TB disease. Hence, we employed a deep CNN model (Vgg16) [49] as a feature extractor rather than an end-to-end network. Vgg16 is one of many Deep models pre-trained on the ImageNet [9] dataset for identifying millions of objects. This technique transfers the rich and discriminative features learned from varieties of objects to the target CXR dataset.

The research paper was presented at the [7] and published in [8].The ImageNet dataset can be downloaded at [9]

[7] 15th International Symposium on Visual Computing ISVC 2020. November 05-07, 2020, San Diego, California, United States.

[8] **Mustapha Oloko-Oba and Serestina Viriri**. "Pre-trained Convolutional Neural Network for the Diagnosis of Tuberculosis". *Advances in Visual Computing*. LNCS Springer, vol 12510, pp 558-569 **(2020)**. DOI: `https://doi.org/10.1007/978-3-030-64559-5_44`

[9] `https://www.image-net.org/download`

Pre-trained Convolutional Neural Network for the Diagnosis of Tuberculosis

Mustapha Oloko-Oba and Serestina Viriri(✉)

School of Mathematics, Statistics and Computer Sciences,
University of KwaZulu-Natal, Durban, South Africa
219098624@stu.ukzn.ac.za, viriris@ukzn.ac.za

Abstract. Tuberculosis (TB) is an infectious disease that claimed about 1.5 million lives in 2018. TB is most prevalent in developing regions. Even though TB disease is curable, it necessitates early detection to prevent its spread and casualties. Chest radiographs are one of the most reliable screening techniques; although, its accuracy is dependent on professional radiologists interpretation of the individual images. Consequently, we present a computer-aided detection system using a pre-trained convolutional neural network as features extractor and logistic regression classifier to automatically analyze the chest radiographs to provide a timely and accurate interpretation of multiple images. The chest radiographs were pre-processed before extracting distinctive features and then fed to the classifier to detect which image is infected. This work established the potential of implementing pre-trained Convolutional Neural Network models in the medical domain to obtained good results despite limited datasets.

Keywords: Tuberculosis · Chest radiograph · Pre-processing · Features extractor · Logistic regression · Convolutional neural network.

1 Introduction

Tuberculosis (TB) is ranked among the topmost ten deadly diseases that resulted in 1.5million deaths in 2018 from about 10 million people falling it globally. Tuberculosis is predominant in developing nations, which contributes to about 95(%) of infected cases and deaths [1]. Everyone is at risk of contracting Tuberculosis even though it is common in men, with about 5.7million cases reported in 2018. Tuberculosis is curable, and most of the causalities could be averted if only the disease is diagnosed early [1]. Traditional approaches employed in screening Tuberculosis includes Tuberculin Skin Test [2], Smear Microscopy [3], Interferon Gamma Release Assays [4], among others. Still, regrettably, most of these approaches have limitations, especially in developing countries ranging from high cost for general adoption, inadequate equipment, requires a stable power supply, inaccurate results because of low sensitivity, and inability to distinguish between latent and active Tuberculosis [5].

G. Bebis et al. (Eds.): ISVC 2020, LNCS 12510, pp. 558–569, 2020.
https://doi.org/10.1007/978-3-030-64559-5_44

More recently, Chest Radiograph (CXR) has been identified as a prominent and less expensive technique for detecting abnormalities in the lungs relating to Tuberculosis [6–8]. World Health Organization recommends the application of CXR as a screening tool for pulmonary diseases [9]. However, interpreting the CXR requires skilled and expert radiologists who are lacking in developing regions and usually results in delay and misdiagnosis [10,11].

Over the years, Computer-aided classification tasks using Convolutional Neural Network (CNN), a Deep Learning technique, have become prominent for analyzing different pulmonary abnormalities in the medical domain. The application of an efficient classification tool is vital for improving the quality of diagnosis while reducing the time taken to diagnose a large volume of CXRs [12,13]. This endeavor is to achieve the global decline in Tuberculosis incidence to about 4–5% annually compared to the current 2% yearly as part of the World Health Organization "End TB Strategy" [1].

Convolutional neural networks have demonstrated tremendous success for several computer vision tasks, including face detection, age and gender classification, object detection, autonomous cars, and even medical imaging analysis. Although, it still requires continuous effort to develop a sophisticated and efficient computer-aided detection system to assist the medical practitioner in the early diagnosis and treatment of deadly diseases such as Tuberculosis.

Application of a convolutional neural network can be made either by building the system from scratch or by using the weight of a pre-trained model, usually referred to as "Transfer Learning" [15]. In transfer learning, the weights of a pre-trained model can be used as features extractor upon which a classifier is trained, or the weights can be fine-tuned [14]. The former is the focus of this study, which is to employ a pre-trained model as a features extractor to train a more effective and accurate classifier.

This study aims to present VGG16 variants of VGGNet [25] pre-trained CNN architecture as features extractor then train a Logistic Regression (LR) [41] classifier to determine which images constitute Tuberculosis manifestations. To the best of our knowledge, only few research work has employed a pre-trained CNN models as a features extractor and trained a machine learning classifier to detect the presence of Tuberculosis. An example is [16].

2 Related Work

The possibilities and strength offered by deep learning and machine learning models have been tremendous in the last years. Their popularity, achievement, and interest in their application for object detection and classification task is partly as a result of the availability of large datasets for training models (either from scratch or transferring the weights of state-of-the art models available from ImageNet repository) [18] and sophisticated computing power infrastructures like GPUs, cloud services and other related frameworks [17,19]. In this section, we will review most of the all-important study that employed transfer learning CNN architectures for detecting Tuberculosis on Chest Radiograph.

It is evident from [18,21] that a convolutional neural network built on the ImageNet extensive database can learn complicated and distinctive features. This capability makes it deployable as features extractor in different domains where large datasets are not readily available such as Remote Sensing [20], Maritime Vessels [21], Medical Field [16,22,23] to achieve competing outcomes or sometimes better performance.

One of the recent studies that employed a pre-trained convolutional neural network for classifying chest radiographs to identify TB abnormalities was [24]. The authors applied the VGG16 model for training 75% of the Shenzhen and Montgomery datasets, while 25% was used as a testing set. Two experiments were performed where augmentation was applied in one instance and not applied in the other case. The result for the example with augmentation achieved 81.25% accuracy, while 80% accuracy was obtained for the case without augmentation. These results prove that increased performance can be achieved by employing preprocessing techniques on the images. Although, only augmentation technique was applied in this case and yet shows an improvement of 1.25%.

The work presented in [16], applied three pre-trained CNNs models to extract features from chest radiographs to identify TB infection. The authors employed each of "Bag of CNN Features," "Simple Features Extractor," and "Ensembles" as proposals to evaluate the potential of GoogLeNet [28], VggNet [25], and, ResNet [29] as features extractor from Montgomery and Shenzhen datasets. The images were segmented and resized according to the original input size of the pre-trained models except for the implementation of "Bag of Features" where the images were not resized. The proposals were evaluated in terms of AUC and Accuracy, where either of GoogLeNet, ResNet, and VggNet shows improvement on different datasets and also at different performance metrics. For instance, under the implementation of "Simple Features" GoogLeNet obtained the best accuracy of 0.782% with 0.838% AUC on the Montgomery datasets meanwhile on the Shenzhen dataset, VggNet attained a higher AUC of 0.912% with Resent having the best accuracy at 0.834% still, the overall best accuracy and AUC stood at 0.847% and 0.926%, respectively. It was evident in this study that the size of datasets and class imbalance can impact the performance of a model.

GoogLeNet [28] architecture was used in [26] for training CXR to detect different TB manifestations. The authors classified the CXR as healthy or unhealthy with one or more of consolidation, cardiomegaly, pulmonary edema, pneumothorax, or pleural effusion. The input images were resized using the half-fill, half-crop technique to normalize the images with large dimensions. Then the images with shorter dimensions were padded with noise to deal with aspect ratio imbalance. The evaluation metric employed were sensitivity, specificity, and AUC with validation of 2 board-certified radiologists, each having 8 and 3 years of experience. The best performance of their experiments was achieved on the healthy class obtaining 91% sensitivity and specificity with 0.964% AUC higher than the unhealthy classes. The disparity could be as a result of the class and data imbalance since the datasets used is a private dataset that is not publicly available and might not have been appropriately labeled.

The authors in [27] are the first to apply deep convolutional neural networks on chest radiographs to diagnose TB. They adopt prominent AlexNet [17] architecture and tweak it by adding a new convolutional layer to handle the high-resolution image since the chest radiographs resolutions are higher than the input images applied in AlexNet as well as a minimum number of nodes in the classifier layer. The system was trained on large private datasets initially using random weight and later trained using AlexNet learned weights. The experiment with the random weight achieved 0.77% accuracy and 0.82% AUC. Meanwhile, the experiment conducted with AlexNet learned weights attained 0.90% accuracy with an AUC of 0.96%. Another experiment was then performed to classify two benchmark datasets (Montgomery and Shenzhen). The average class probabilities of the training were taking to attain ensemble outcomes on both datasets. The Montgomery produced an accuracy of 0.674% and AUC of 0.884%, while accuracy and AUC stand at 0.837% and 0.926% respectively for the Shenzhen dataset. A comparison to evaluate the performance of CNN with and without application of transfer learning is then performed. The result found that CNN with transfer learning outperformed having an accuracy of 0.903% and AUC of 0.964% against the 0.773% accuracy and 0.816% AUC obtained for the CNN without transfer learning.

The study conducted in [30] evaluates the performance of some pre-trained convolutional neural networks concerning localization and classification of different TB abnormalities from chest radiographs. The study used heatmap as a way of localization in the chest radiograph images. Experiments were performed on three public datasets and showed that the performance of a particular pre-trained CNN model across all the abnormalities. In other words, a model can perform better on some manifestations and act woefully on other manifestations. The study also found that for classification accuracy, shallow layers perform better than deeper layers and finally concludes that Ensemble [32] of different pre-trained CNN models attained better performance than a single one. The result of their system was evaluated in terms of accuracy, sensitivity, specificity, and AUC, where the best performance obtained for the detection of the cardiomegaly abnormalities saw VggNet achieved 92.00% accuracy, 0.94% AUC, 96.00% sensitivity. At the same time, ResNet got better at a specificity of 96.00%.

Another study that applied transfer learning to screen chest radiographs for the presence of TB is presented in [31]. The authors modified the AlexNet and VggNet architectures to classify chest radiograph images into healthy and unhealthy classes. The modified architecture was trained on Shenzhen and Montgomery datasets; the experimental results show AlexNet achieved 80.4% while VggNet slightly outperformed the AlexNet to obtain 81.6% classification accuracy. The study further concludes that having a large number of datasets will boost accuracy as the model will be able to learn more discriminative features.

In [33], pre-trained CNN models were fine-tuned to train chest radiograph images collected at the Yonsei University to detect TB automatically. The images obtained are accompanied by demographic information such as patients' height, weight, age, and gender. Before training the modified model, the lungs region

in the images were segmented using U-Net [34], then resized to 256 256 pixels, and augmented as a way of pre-processing. The experiments performed to compare the performance of the model directly on the images and the performance of incorporating demographic information with the images. The authors used an equal number of image samples for the positive and negative classes for training the model. The results show the experiment which incorporates demographic information with the images obtained 0.9213% while the one based on the images alone achieved 0.9075% on the test set. The result clearly shows that incorporating demographic information like age, gender, height, and weight can slightly improve the performance of the model.

3 Methods and Techniques

In this section, we discussed the Chest Radiograph databases employed for the development of the Computer-aided detection system for identifying the presence of Tuberculosis.

3.1 Datasets

The following databases were used for the experiments done in this work. These databases are publicly available at: https://lhncbc.nlm.nih.gov/publication/pub9931 and have been utilized in many studies.

The Shenzhen database [35,36] is explicit to TB disease and is publicly available for research. The database is made up of 336 unhealthy samples and 326 healthy samples. The identification of each image ends either with a "0" signifying healthy sample or a "1" meaning unhealthy sample. All images are approximately 3000 × 3000 pixels in resolution. This database also contains clinical readings that give demographic details about each of the samples with respect to age, sex, and diagnosis. The chest radiographs in this database were collected at the Guandong Hospital in Shenzhen, China, and saved in (png) format. See Fig. 1 for samples of chest radiographs in this database.

The Montgomery County chest radiograph database [35,36] is a TB-specific database collected by the Department of Health Services, Maryland, U.S.A for research intent. This database is composed of 58 unhealthy samples and 80 healthy samples. The naming convention of each image ends either with "0" denoting a healthy sample or "1" indicating an unhealthy sample. All samples are of the size 4892 × 4020 pixels or 4020 × 4892 pixels available as (png) format. This database is also accompanied by clinical readings that give demographic details about each of the samples with respect to sex, age, and manifestations. Additionally, the database contains manually generated left and right mask segmentation of the chest radiographs. See Fig. 2 for samples of chest radiographs in this database.

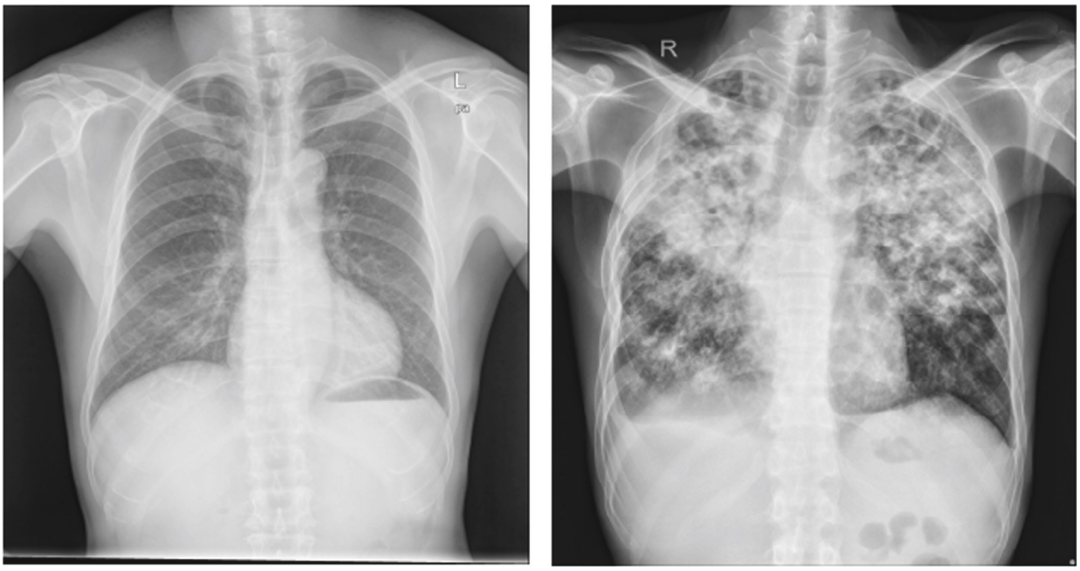


Fig. 1. Sample in the Shenzhen Database. **(a)** is a healthy chest radiograph of a 60 years old male patient while **(b)** is an unhealthy chest radiograph of a 36 years old male patient diagnosed with "bilateral pulmonary TB".

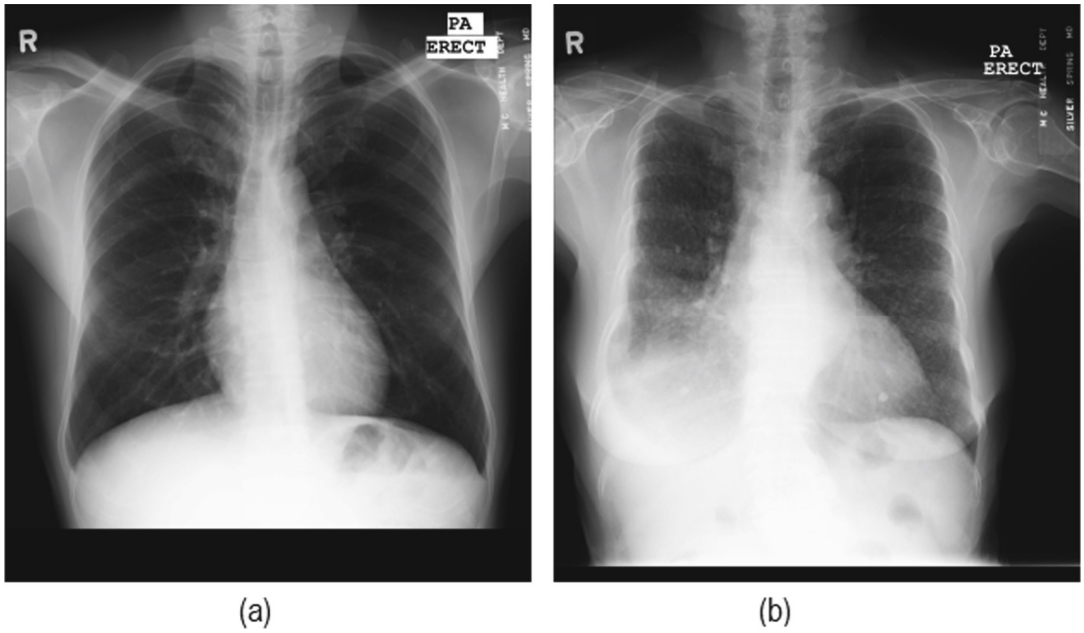


Fig. 2. Sample in the Montgomery Database. **(a)** is a healthy chest radiograph of a 33 years old male patient while **(b)** is an unhealthy chest radiograph of an 84 years old female patient diagnosed with "bilateral miliary nodules diffusely with right middle lobe and right pleural effusion".

3.2 Chest Radiographs Preprocessing

In this study, we applied "Data Augmentation [38], and CLAHE [39] as preprocessing techniques to boost and improve the quality of the chest radiographs.

Augmentation is an artificial data generation technique used in machine learning to control overfitting, improving generalization as well as the accuracy of the network. CNN models perform better when the size of the datasets is large [31,40], and since the number of the chest radiographs in our database is not large enough, we applied data augmentation to the original chest radiograph to generate new samples while preserving the labels of the samples. The augmentation operations applied to generate new samples includes: "rotation with probability = 1, max left and right probability = 2, flip right and left probability = 0.1, random zoom = 0.2, percentage area = 0.5 top and bottom flip probability of 0.2.

Contrast limited adaptive histogram equalization (CLAHE) has been applied to many medical images to improve the quality of the images for the purpose of analysis. We applied CLAHE to the chest radiographs to enhance the images making the details more visible and suitable for the feature extraction task. Figure 3 show the effect of the CLAHE on an image.

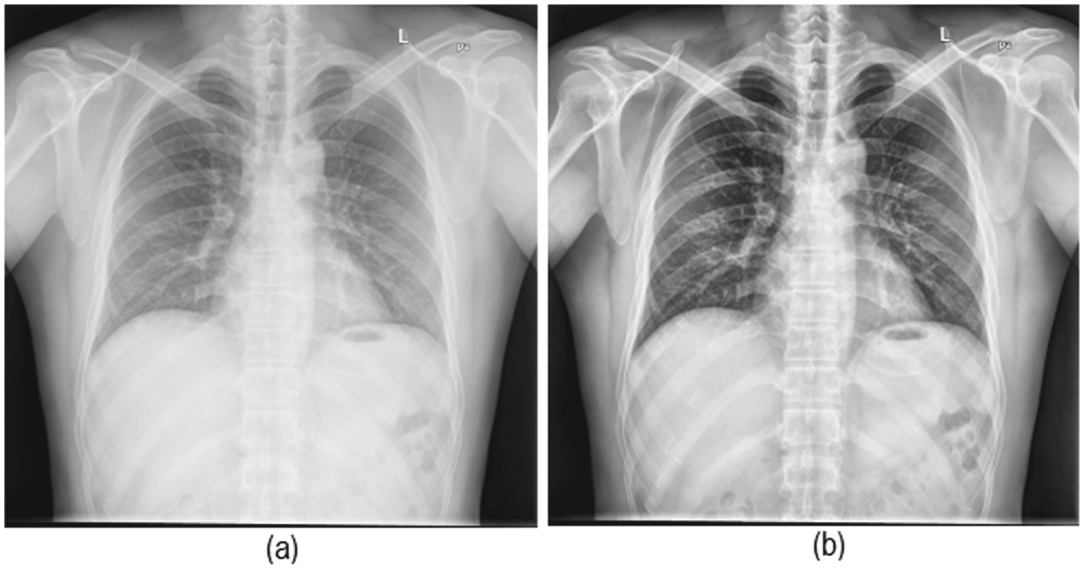


Fig. 3. **(a)** is the original image, while **(b)** is the enhanced version of the original.

3.3 Pre-trained Deep CNN Model for Feature Extraction

The concept of Transfer Learning is actually to employ the weight of a model that is trained on perticular objects and transfer it to train a different object to eliminate the complexity of training a model from scratch. There are several benchmark deep CNN learning models trained on large datasets like ImageNet

[18] to recognize several objects. Example of these models that have performed excellently well includes AlexNet [17], VggNet [25], GoogLeNet [28], ResNet [29] among other prominent ones. In this present work, we treated the Vgg16 variant of VggNet as a feature extractor rather than as end-to-end network as the case of the original model.

We truncated the three fully convolutional layers in the original vgg16, which resulted in making the "max-pooling" the final and output layer, as presented in Fig. 4. The chest radiographs images are resized to $224 \times 224 \times 3$ to conform to the input dimension of Vgg16 and then propagated through the network until the last max-pooling layer, which now serves as the output layer having a $7 \times 7 \times 512$ shape that is flattened into a vector map and fed as input to the classifier. Of course, the output can be extracted at any random layer, but we decided to output at the pooling layer, this is with respect to the analysis presented in [37].

Fig. 4. Truncated Vgg16 where "Max-pooling layer" is employed as the output layer.

After extracting features from the chest radiographs, these features are then fed as input to the Logistic Regression classifier for training to obtain classification details, as illustrated in Fig. 5. The LR classifier is powerful and efficient for binary classification. Its relation to exponential probability distribution makes them perform well in medical domain because it allows the dependence of class labels on features and gives room to carry out sensitivity analysis.

As a tradition, we split the dataset into training and validation set prior to training the classifier such that an indicator is configured as [:q] and [q:] specifying that all the images prior to ":q" are included in the training set while the images after "q:" are reserved for validation. The system 'C' parameters range between "0.0001 to 1000", the optimum parameter for the classifier is then selected by employing a grid search. Once the optimum parameter is established, the model is then evaluated against the testing data to determine the performance of the classifier in detecting whether an image is infected with TB or not.

4 Result and Discussion

This section presents the experimental results performed on both Montgomery and Shenzhen benchmark datasets. These datasets were split into 70(%) used for training the logistic regression classifier on the extracted features from the

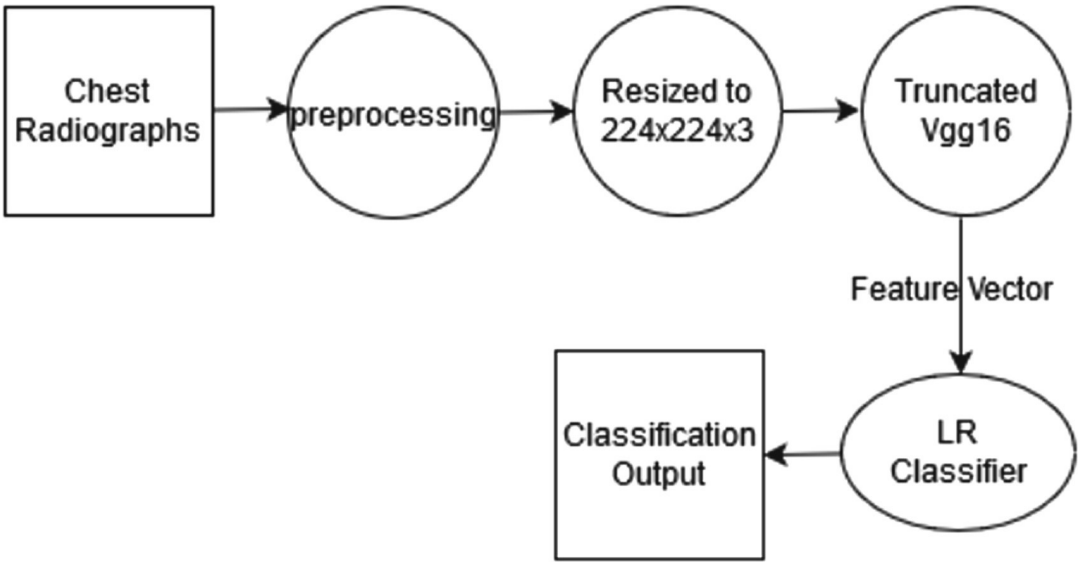


Fig. 5. Procedural flow diagram.

modified pre-trained vgg16 model. In contrast, 30(%) of the dataset was used to evaluate the performance of the classifier on both datasets.

In the first experiment carried out on the Montgomery set, the system obtained an accuracy of 89(%) and precision of 90.1(%) while on the Shenzhen set, the system performed at 95.8(%) accuracy with a precision of 96(%). Also, we took a step further to determine the Rank-1 accuracy of the system, which is the number of times where the predicted output conforms to the corresponding ground-truth. Hence, the Rank-1 accuracy obtained for both datasets is 92.40(%) and 99.25(%), respectively, as shown in Table 1.

Table 1. Results of the experiments

Database	Accuracy (%)	Precision (%)	Rank-1 (%)
Montgomery	89	90.1	92.40
Shenzhen	95.8	96	99.25

It can be observed from Table 1 that the result obtained in the experiment on the Montgomery set is lower compared to the Shenzhen set, as reported in related work. This outcome is mostly as a result of lesser dataset and class imbalance, Although, in this case since we performed augmentation on the dataset, we can say only class imbalance where the ratio of healthy to unhealthy samples is about "60 : 40" in the Montgomery set is the factor responsible for it (Table 2).

Table 2. Comparison of the proposed model with related work. The performance metrics are measured in (%)

Ref	Model	Accuracy	Precision
24	VggNet	81.25	0.80
16	GoogLeNet, VggNet, ResNet	0.847	-
27	AlexNet	90.03	-
30	AlexNet, VggNet, ResNet	92.00	-
31	VggNet, Inception, ResNet, DensNet	81.60	-
Proposed	**VggNet**	**95.80**	**96.00**

5 Conclusion

In this work, we explored pre-trained CNN and implemented Vgg16 as a features extractor. Logistic Regression classifier was then trained on the extracted features to detect Tuberculosis. This study and other related work have shown characteristics such as eliminating "training complexity, expenses, high computing power, and large datasets' that made implementing pre-trained CNN models powerful and relevant to medical and other fields where dataset are limited for classification. Although the approach applied in this work has obtained excellent results, but a better computer-aided detection system can still be developed if a large amount of annotated chest radiographs dataset can be created. These annotated datasets can then be utilized in building deeper models from the scratch to identify various pulmonary diseases.

References

1. World Health Organization.: Global status report on alcohol and health 2018. World Health Organization (2019)
2. Cohn, D.L., O'Brien, R.J., Geiter, L.J., Gordin, F., Hershfield, E., Horsburgh, C.: Targeted tuberculin testing and treatment of latent tuberculosis infection. MMWR Morb Mortal Wkly Rep **49**(6), 1–54 (2000)
3. Desikan, P.: Sputum smear microscopy in tuberculosis: is it still relevant? Indian J. Med. Res. **137**(3), 442 (2013)
4. Zwerling, A., van den Hof, S., Scholten, J., Cobelens, F., Menzies, D., Pai, M.: Interferon-gamma release assays for tuberculosis screening of healthcare workers: a systematic review. Thorax **67**(1), 62–70 (2012)
5. Leung, C.C.: Reexamining the role of radiography in tuberculosis case finding. Int. J. Tuberc. Lung Dis.: Official J. Int. Union Against Tuberc. Lung Dis. **15**(10), 1279 (2011)
6. Jaeger, S., et al.: Automatic screening for tuberculosis in chest radiographs: a survey. Quant. Imaging Med. Surg. **3**(2), 89 (2013)
7. Naing, W.Y.N., Htike, Z.Z.: Advances in automatic tuberculosis detection in chest x-ray images. Signal Image Process. **5**(6), 41 (2014)

8. World Health Organization.: Tuberculosis prevalence surveys: a handbook. World Health Organization (2011)
9. World Health Organization.: Chest radiography in tuberculosis detection: summary of current WHO recommendations and guidance on programmatic approaches (No.WHO/HTM/TB/2016.20). World Health Organization (2016)
10. Noor, N.M., Rijal, O.M., Yunus, A., Mahayiddin, A.A., Peng, G.C., Abu-Bakar, S.A.R.: A statistical interpretation of the chest radiograph for the detection of pulmonary tuberculosis. In: 2010 IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES), pp. 47–51 (2010)
11. Pedrazzoli, D., Lalli, M., Boccia, D., Houben, R., Kranzer, K.: Can tuberculosis patients in resource-constrained settings afford chest radiography? Eur. Respir. J. **49**(3), 1601877 (2017)
12. Schmidhuber, J.: Deep learning in neural networks: an overview. Neural Networks **61**, 85–117 (2015)
13. Livieris, I.E., Kanavos, A., Tampakas, V., Pintelas, P.: A weighted voting ensemble self-labeled algorithm for the detection of lung abnormalities from X-rays. Algorithms **12**(3), 64 (2019)
14. Al Hadhrami, E., Al Mufti, M., Taha, B., Werghi, N.: Transfer learning with convolutional neural networks for moving target classification with micro-Doppler radar spectrograms. In: IEEE International Conference on Artificial Intelligence and Big Data (ICAIBD), pp. 148–154 (2018)
15. Nogueira, K., Penatti, O.A., Dos Santos, J.A.: Towards better exploiting convolutional neural networks for remote sensing scene classification. Pattern Recogn. **61**, 539–556 (2017)
16. Lopes, U.K., Valiati, J.F.: Pre-trained convolutional neural networks as feature extractors for tuberculosis detection. Comput. Biol. Med. **89**, 135–143 (2017)
17. Krizhevsky, A., Sutskever, I., Hinton, G. E.: Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1097–1105 (2012)
18. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: a large-scale hierarchical image database. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255 (2009)
19. Dean, J., et al.: Large scale distributed deep networks. In: Advances in Neural Information Processing Systems, pp. 1223–1231 (2012)
20. Xie, M., Jean, N., Burke, M., Lobell, D., Ermon, S.: Transfer learning from deep features for remote sensing and poverty mapping. In: Thirtieth AAAI Conference on Artificial Intelligence (2016)
21. Bousetouane, F., Morris, B.: Off-the-shelf CNN features for fine-grained classification of vessels in a maritime environment. In: Bebis, G., et al. (eds.) ISVC 2015. LNCS, vol. 9475, pp. 379–388. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-27863-6_35
22. Liu, T., Xie, S., Yu, J., Niu, L., Sun, W.: Classification of thyroid nodules in ultrasound images using deep model based transfer learning and hybrid features. In: 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 919–923 (2017)
23. Ribeiro, E., Uhl, A., Wimmer, G., Häfner, M.: Exploring deep learning and transfer learning for colonic polyp classification. Comput. Math. Methods Med. (2016)
24. Ahsan, M., Gomes, R., Denton, A.: Application of a Convolutional Neural Network using transfer learning for tuberculosis detection. In: IEEE International Conference on Electro Information Technology (EIT), pp. 427–433 (2019)

25. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
26. Cicero, M., et al.: Training and validating a deep convolutional neural network for computer-aided detection and classification of abnormalities on frontal chest radiographs. Invest. Radiol. **52**(5), 281–287 (2017)
27. Hwang, S., Kim, H. E., Jeong, J., Kim, H.J.: A novel approach for tuberculosis screening based on deep convolutional neural networks. In: International Society for Optics and Photonics.: Computer-Aided Diagnosis, vol. 9785, p. 97852W (2016)
28. Szegedy, C., et al.: Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–9 (2015)
29. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
30. Islam, M.T., Aowal, M.A., Minhaz, A.T., Ashraf, K.: Abnormality detection and localization in chest x-rays using deep convolutional neural networks. arXiv preprint arXiv:1705.09850 (2017)
31. Rohilla, A., Hooda, R., Mittal, A.: Tb detection in chest radiograph using deep learning architecture. In: ICETETSM-17, pp. 136–147 (2017)
32. Rokach, L.: Ensemble-based classifiers. Artif. Intell. Rev. **33**(1–2), 1–39 (2010)
33. Heo, S.J., et al.: Deep learning algorithms with demographic information help to detect tuberculosis in chest radiographs in annual workers' health examination data. Int. J. Environ. Rese. Public Health **16**(2), 250 (2019)
34. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
35. Jaeger, S., et al.: Automatic tuberculosis screening using chest radiographs. IEEE Trans. Med. Imaging **33**(2), 233–245 (2013)
36. Candemir, S., et al.: Lung segmentation in chest radiographs using anatomical atlases with nonrigid registration. IEEE Trans. Med. Imaging **33**(2), 577–590 (2013)
37. Sharif Razavian, A., Azizpour, H., Sullivan, J., Carlsson, S.: CNN features off-the-shelf: an astounding baseline for recognition. In: Proceedings of the IEEE conference on Computer Vision and Pattern Recognition Workshops, pp. 806–813 (2014)
38. Bloice, M.D., Stocker, C., Holzinger, A.: Augmentor: an image augmentation library for machine learning. arXiv preprint arXiv:1708.04680 (2017)
39. Kurt, B., Nabiyev, V.V., Turhan, K.: Medical images enhancement by using anisotropic filter and clahe. In: IEEE International Symposium on Innovations in Intelligent Systems and Applications, pp. 1–4 (2012)
40. Tajbakhsh, N., et al.: Convolutional neural networks for medical image analysis: full training or fine tuning? IEEE Trans. Med. Imaging **35**(5), 1299–1312 (2016)
41. Hosmer Jr., D.W., Lemeshow, S., Sturdivant, R.X.: Applied logistic regression, vol. 398. John Wiley and Sons, New York (2013)

3.3.2 Conclusion

In this work, we truncated a pre-trained Vgg16 and implemented it to extract TB features from the CXR images; the extracted features were then fed as input to the classifier algorithm to obtain the final binary classification output. The classifier algorithm was constructed such that the "C" parameters ranged between "0.0001 to 1000"; the optimum parameter was then selected by employing a grid search. Once the optimum parameter was established, the model was then evaluated against the testing data samples to determine the performance of the classifier in detecting whether an image was infected with TB or not. The technique achieved 95.80% accuracy, with a precision of 96%. These results proved the effect of implementing transfer learning on training new models to improve accuracy, most importantly where large number of datasets are unavailable, like in the medical field.

3.4 Ensemble of Convolution Neural Networks for Automatic Tuberculosis Classification

3.4.1 Introduction

This section presents a research paper that proposed Ensemble learning for improving model accuracy. Three pre-trained models (ResNet50 [52], Vgg16 [49], and Inception V3 [12]) were combined to build an Ensemble classifier for the detection and classification of TB from CXR. This technique was motivated by the promising results obtained in Section 3.3. The Ensemble model was trained on the Shenzhen dataset and evaluated on the Montgomery dataset to achieve better generalization on the test set.

The research paper was presented at the [10] and published in [11].

[10] 13th International Conference on Computational Collective Intelligence ICCCI 2021. September 29-1 October 2021, Rhodes, Greece.

[11] **Mustapha Oloko-Oba and Serestina Viriri**. "Ensemble of Convolution Neural Networks for Automatic Tuberculosis Classification". *Computational Collective Intelligence*. LNCS Springer, vol 12876, pp 549-559 **(2021)**. DOI: `https://doi.org/10.1007/978-3-030-88081-1_41`

Ensemble of Convolution Neural Networks for Automatic Tuberculosis Classification

Mustapha Oloko-Oba and Serestina Viriri( )

School of Mathematics, Statistics and Computer Science,
University of KwaZulu-Natal, Durban, South Africa
viriris@ukzn.ac.za

Abstract. Tuberculosis (TB) is curable, and millions of deaths could be averted if diagnosed early. One of the sources of screening TB is through a chest X-ray. Still, its success depends on the interpretation of skilled and experienced radiologists, mostly lacking in high TB burden regions. However, with the intervention of a computer-aided detection system, TB can be automatically detected from chest X-rays. This paper presents an Ensemble model based on multiple pre-trained models to detect TB from chest X-rays automatically. The models were trained on the Shenzhen dataset and validated on the Montgomery dataset to achieve good generalization on a new (unseen) dataset. Improved classification accuracy was however achieved through the Ensemble model compared to the individual models. The proposed model indicates the strength of combining multiple models to improve model accuracy.

Keywords: Ensemble · Tuberculosis · Convolutional Neural Network · Chest X-Ray · Pre-processing

1 Introduction

Over the years, Tuberculosis (TB) has ranked among the highest causes of mortality. In 2019, about 10 million persons fell sick globally from TB [23]. TB disease is most widespread in developing regions. According to the world health organization (WHO), TB is a disease of economic distress, poverty and vulnerability, which is evident in the geographical incident of TB cases where South-East Asia and Africa accounted for about 69% of the total cases [24] as shown in Table 1.

TB is certainly preventable and curable but needs to be diagnosed early [24]. Early diagnosis is fundamental in effectively tackling TB, and significantly reducing the mortality rate in line with the WHO "End TB Strategy". This strategy has been hampered by the advent of COVID-19 pandemic, which exhibits similar symptoms with TB and could increase the mortality rate by about 0.2–0.4 million [24].

For early diagnosis of TB, proper diagnostic procedures are essential. Various test procedures have been employed to detect TB, but chest X-ray (CXR)

N. T. Nguyen et al. (Eds.): ICCCI 2021, LNAI 12876, pp. 549–559, 2021.
https://doi.org/10.1007/978-3-030-88081-1_41

Table 1. Geographical tuberculosis incidence

Region	South-East Asia	Africa	Western Pacific	Eastern Mediterranean	The Americas	Europe
Cases (%)	44	25	18	8.2	2.5	2.9

is more prominent and recommended due to high sensitivity in screening lung abnormalities [22,26]. Regardless of the CXR suitability in screening abnormalities, accurately interpreting CXR images requires a skilled and expert radiologist mostly due to different manifestations of TB on CXR images as well as in co-infection cases where other pulmonary diseases have similar symptoms with TB, such as lung cancer and COVID-19 [6].

Unfortunately, in TB epidemic regions, lack of skilled radiologist to adequately interpret CXR and other factors like outdated screening equipment, electricity, and lack of awareness have been among the reasons for the delay in diagnosing leading to a widespread of TB [19,33]. These challenges gave rise to the development of Computer-aided detection (CAD) system using Convolutional Neural Network (CNN), for screening CXR to detect pulmonary abnormalities. Various CAD has been developed to automatically detect TB, thereby resulting in many patients' early diagnosis and reducing the burden of an expert radiologist in the TB prevalent areas [11,18]. DL models (CNN) application to medical imaging has shown tremendous performance, especially in lung abnormalities screening, to detect/classify CXR images [12]. CNN architectures often require large data samples to learn distinctive features useful for predictions and attain high performance [30]. The first CNN model trained on CXR for TB classification was the ALexNet [12]. AlexNet [15] was initially introduced on the ImageNet dataset in 2012 and then fine-tuned to train CXR datasets. This process is called "Transfer Learning" and is very useful compared to training a model from scratch with small data samples [25] as evident from literature. Therefore, some of the CAD systems that employed CNN models are reviewed to highlight the strength and challenges that form the basis for this study.

The study presented in [28] modified the VggNet [31] and AlexNet architectures for screening and classifying CXR images as healthy and unhealthy. The architectures were independently trained on the Montgomery and Shenzhen dataset to evaluate the model. VggNet performed slightly above the AlexNet, achieving classification accuracy of 81.6% and 80.4% respectively. The study showed that having large data samples will increase the model accuracy as distinctive features are learned from different data samples. Three pre-trained CNN models were customized and employed in [14] to classify TB abnormalities in CXR. Image pre-processing such as "Data Augmentation" was applied to artificially increase the dataset size before training the models. Thereafter, the model's performance was evaluated to determine accuracy, specificity, and sensitivity where CapsNet [29] outperformed the other models concerning affine image prediction. The use of pre-trained CNN models to extract features from images has been an effective approach, especially in the medical field where annotated CXR are limited. This is evident in [20]. The study employed "Vgg16" to

extract distinctive features from CXR then train a classifier on the extracted features to identify the infected CXR from the healthy ones. Contrast limited adaptive histogram equalization and data augmentation were applied to boost the quality of the CXR to obtain a reasonable accuracy. The study further highlights that having a large dataset of annotated CXR could result in building a deeper and more robust CAD from scratch.

A CNN architecture consisting of five convolutional, pooling and Softmax layers is presented in [27] to screen for TB abnormalities in CXR. This model also incorporates a visualization procedure to assist radiologist with visual diagnosis while minimizing computational complexity to achieve a faster and efficient model. In [4], GoogLeNet [32] model was applied to diagnose different manifestation of TB. CXR was classified either as healthy or unhealthy with one or more of pleural effusion, consolidation, pneumothorax, cardiomegaly, and edema. The CXR were pre-processed where the images with shorter and larger dimensions were padded and normalized respectively before training. AUC, specificity and sensitivity were used to evaluate the model performance and then compared with the result of 2 board-certified radiologists. An approach proposed in [8] fine-tuned CNN to detect TB automatically. The CXR was collected alongside the patient's demographic details such as height, weight, gender, and age. Two experiments were conducted to compare the performance of CXR with demographic details and CXR without demographic details. The former shows higher performance in terms of AUC and sensitivity over the latter. The study in [1] employed Vgg16 to train CXR images for the detection of TB. Two experiments performed where data augmentation was applied to the CXR in one case and not applied in the other. The experiment with the augmented CXR performed slightly higher than the non-augmented experiment.

While different CNN models, either pre-trained or trained from scratch, have been used to predict objects, there have also been cases where multiple models are employed to improve/attain better and accurate predictions than any individual model; this phenomenal is called "Ensemble Learning". An Ensemble of ResNet [7], VggNet, and GoogLeNet was introduced in [17] along with "Bag of features and Simple feature extractor". Experiments were performed for each method where the input CXR were down-sampled to conform with different CNNs except for Bag of features where CXR images were not resized. Down-sampling may result in loss of vital detail, which will affect the performance of the model. The Ensemble model achieved better performance than the other two models. The study shows that class imbalance, down-sampling and dataset size can affect performance. In [2], deep CNN models were integrated with the hand-craft technique through Ensemble learning to extract features from CXR images. The extracted features were then employed as inputs to train a classifier for identifying infected CXRs. In all the experiments, the Ensemble model performed better compared to the individual model.

Similarly, the study proposed in [10] employed Ensemble technique where features were extracted from raw CXR images and edge detected images. The approach begins with image pre-processing followed by classifier development before

classification using different combinations to represent the Ensemble. The CNN architectures were applied to both image type using accuracy, sensitivity, and specificity as the evaluation metrics. The Ensemble model achieved the highest accuracy and sensitivity. An Ensemble classifier was proposed in [9] for classifying TB abnormalities in CXR. Individual CNN models are firstly employed for training using different hyper-parameters after which an Ensemble of these models is created and used for training. These models are then compared and the Ensemble attain an improved accuracy. Although the Ensemble model performed better in this study, but another single model [21] has achieved better accuracy.

In summary, efforts have been invested in developing CAD systems to detect TB abnormalities automatically. Evidently, from literature, deep learning models trained from scratch are computationally expensive and are hugely dependent on large dataset size which is limited in the medical field. However, studies have found Transfer Learning using pre-trained models as a useful approach to compliment limited datasets. Accurately diagnosing TB is a sensitive and challenging task due to its different manifestations and high risk to life, hence the need to develop an effective CAD that can efficiently detect different TB manifestation.

This paper proposed an Ensemble technique that takes advantage of the strength of different pre-trained CNN models with the expectation that a misclassified image from model "A" can be correctly classified by model "B" or "C". Therefore, fusing several classifiers can increase the rate of detection compared to any stand-alone classifier. The contributions of this study are as follows:

- Development of Ensemble model that leverage on different pre-trained models' strength to increase the TB detection rate.
- Models trained with one dataset (Shenzhen) and validated on a different dataset (Montgomery). The tradition has been using the same datasets for training and validation
- Data augmentation and contrast adaptive histogram equalization techniques to improve the quality of the datasets

The next section presents the proposed methodology, followed by the result section where the experimental outcome is presented before the Discussion section and finally, Conclusion.

2 Proposed Methodology

The proposed model is made up of Image pre-processing techniques, the generation of individual CNN classifiers, and combination of multiple models to build an Ensemble classifier for TB classification. The Ensemble implemented in this paper is the "Bagging" method because of its strength to achieve low variance, thereby controlling model overfitting. Bagging yield stability and achieve better performance by combining multiple models. In other words, it employs two

or more models to learn distinctive features independently of the other participating models; the output of each participating model is then combined and averaged to obtain a final prediction.

The flow diagram of the proposed model is shown in Fig. 1.

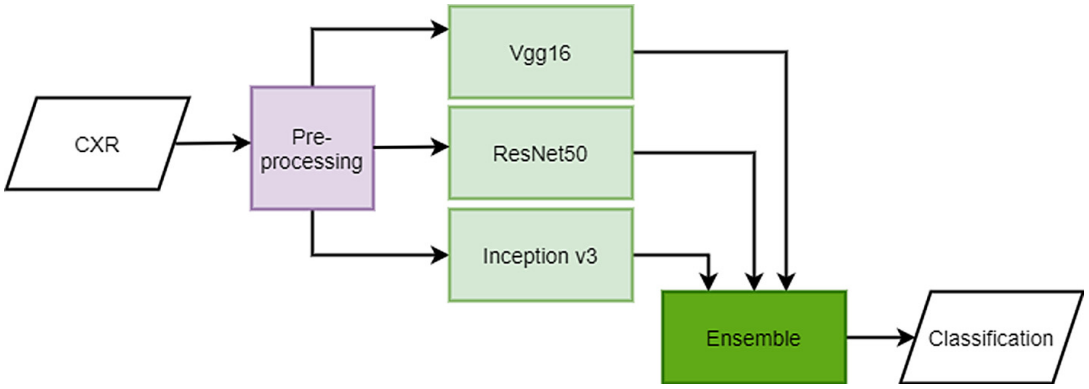


Fig. 1. Process diagram of the proposed model

2.1 Image Pre-processing

The pre-processing techniques applied to the CXR images in this study include Contrast Limited Adaptive Histogram Equalization [16], to improve the quality of the images in terms of clarity, then Data Augmentation [3], was applied to adequately increase the diversity of training samples by way of flipping, and padding. Finally, since we are training multiple models, the CXR images were resized to conform to the individual model's original input size. The input size for Inception v3 is $229 \times 229 \times 3$ while Vgg19 and ResNet50 have similar input size of $224 \times 224 \times 3$.

2.2 Building Ensemble of CNN

In constructing the proposed Ensemble model, we consider three powerful pre-trained CNN models, "Vgg16, ResNet50, and Inception v3," initially trained on the ImageNet [5], dataset. These models were trained to identify about 1000 classes of objects, hence adopting it to CXR dataset; their final layer (classifier) is substituted with a binary classifier. Each of the CNN models is trained independently using different hyper-parameters that perform best in terms of the optimizer, learning rate, and activation function. The models were trained each for 80 iterations (epochs) with a batch size of 32 to obtain the probability of each CXR samples belonging to either a positive or negative class representing TB and non-TB classes. The output from the individual model is then combined to obtain the final prediction through the Ensemble classifier. The performance

of the proposed model is measured in terms of the popular evaluation metrics defined as follows:

Accuracy: is the rate of correctly predicted samples from the overall samples examined and is defines as:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{1}$$

Sensitivity: refers to the measure of confirmed positive samples that are correctly identified as positives.

$$Sensitivity = \frac{TP}{TP + FN} \tag{2}$$

Specificity: is the rate of confirmed negative samples that are correctly identified as negatives.

$$Specificity = \frac{TN}{TN + FP} \tag{3}$$

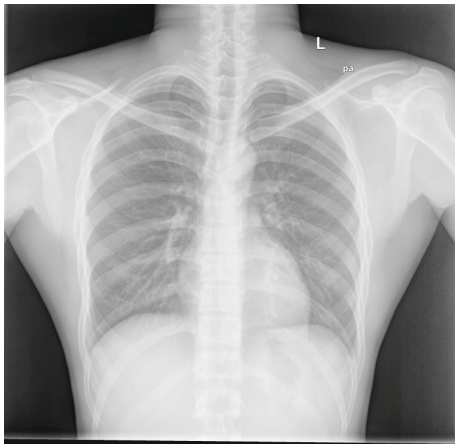
where:
TP = True Positive, FP = False Positive, TN = True Negative, and FN = False Negative respectively.

3 Results

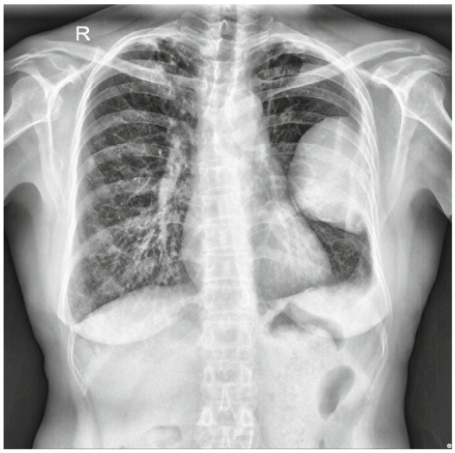
The proposed model was trained on the Shenzhen dataset and validated on the Montgomery dataset [13]. These datasets contain TB specific CXR images provided by the National Library of Medicine (NLM) and publicly made available for research purposes. The Shenzhen dataset is made up of 336 infected CXR samples and 326 healthy CXR samples. All the samples are 3000 × 3000 pixels in resolution. Meanwhile, the Montgomery dataset contains 58 infected CXR samples and 80 healthy samples having a size of either 4892 × 4020 pixels or 4020 × 4892 pixels. Both datasets are accompanied by clinical readings that provide demographic details about each of the CXR samples with respect to diagnosis, sex, and age of the patients. The diagnostic information given for each CXR sample serves as a standard ground truth to validate the proposed model's output. These datasets are accessible from the NIH website[1]. Figure 2 shows sample CXR from both datasets.

Several experiments were performed aimed at minimizing false negative and positives as much as possible. In this regard, Vgg16, ResNet50, and Inception v3 trained and evaluated for detection of TB abnormalities. Different outcomes are expected since different models exhibit diverse characteristics in terms of depth, parameters, and overall configurations. In the first experiment where the above models are individually employed to classify CXR, the best performance was obtained with the Inception v3 model at 93.50% followed by a 92.10% performance with the Vgg16 and the lowest accuracy of 89.51% with the ResNet50

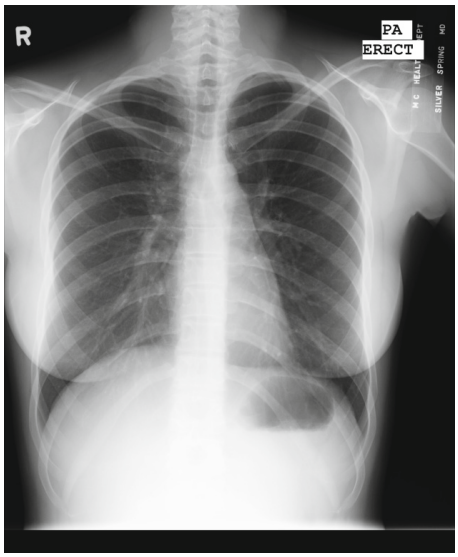
[1] https://lhncbc.nlm.nih.gov/publication/pub9931.



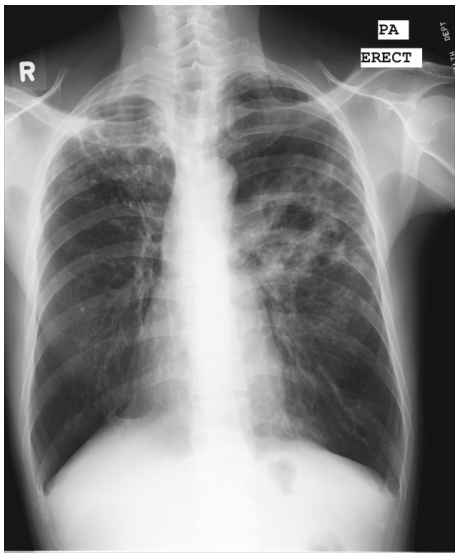
(a) Healthy CXR from train-set



(b) Unhealthy CXR from train-set



(c) Healthy CXR from test-set



(d) Unhealthy CXR from test-set

Fig. 2. Sample of CXR images from both datasets

model. Meanwhile, ResNet50 surpass the other models in terms of sensitivity. The next experiment is the Ensemble of all models from the previous experiment. The performance of the individual models is combined/averaged to build an Ensemble classifier for Tb detection. The Ensemble model output surpasses all the individual models to boost the overall classification accuracy. The detail experimental results are shown in Table 2 and Fig. 3 respectively.

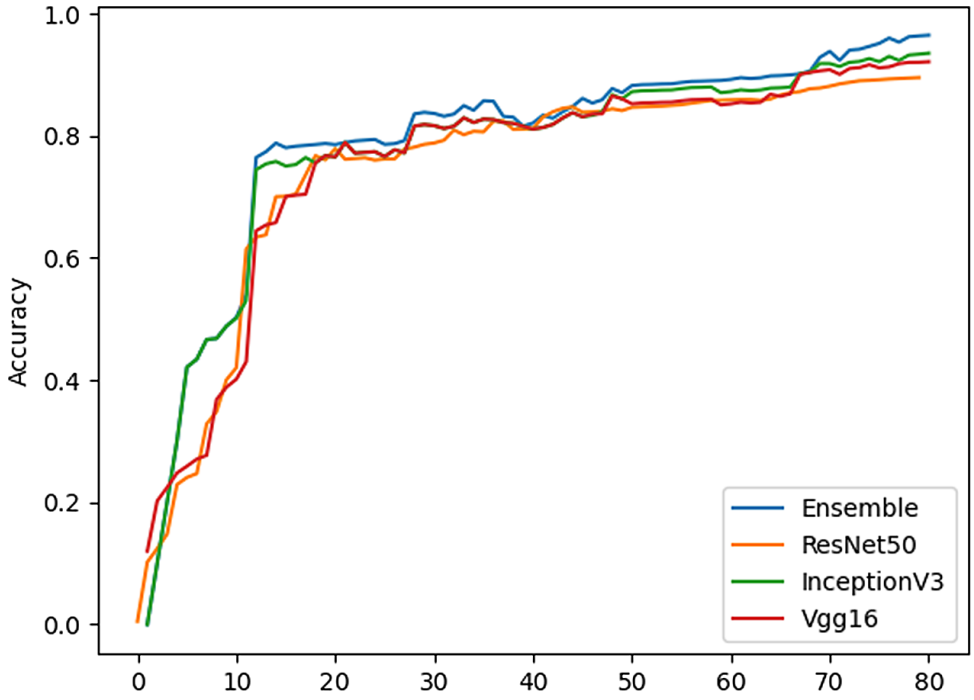


Fig. 3. Models accuracy performance

Table 2. Experimental results

	Ensemble	ResNet50	Vgg16	Inception v3
Accuracy (%)	96.14	89.51	92.10	93.50
Sensitivity (%)	90.03	89.09	88.25	79.31
Specificity (%)	92.41	90.08	91.00	83.36

4 Discussion

It is evident from Table 2, the performance of individual models shows the Inception V3 model with the best accuracy, Although the overall best performance is achieved through the Ensemble model. One of the obstacles that affect CNN model is overfitting which is common when a model is run on a new dataset. It is essential to control overfitting to achieve better generalization of the model on new datasets. In this study, we employed techniques such as "data augmentation, normalization, dropout, and least absolute deviations (LAD), also known as L1, to control overfitting. As a result of various manifestations of TB, depending on a single model might not be too effective. Hence Ensemble of multiple models employed in this study makes the CAD system more robust with an improved detection rate because with Ensemble, different models could compliment themselves such that where a model misclassifies an image, one or more of the other models can accurately classify the image. The proposed model generalized well

despite having the training, validation and test set from different datasets. The model performed well when compared with recent Ensemble CAD systems as shown in Table 3.

Table 3. Proposed model compared with related work. The performances are measured in terms of Accuracy (ACC) and Area Under Curve (AUC)

Authors ref	Models combined	Performance (%)
Hijazi et al. [10]	2	89.77
Hernandez et al. [9]	3	86.40
Ayaz et al. [2]	6	0.99
Proposed	3	96.14

Hijazi et al. [10], Hernandez et al. [9], and the **Proposed** model are evaluated using the accuracy, while Ayaz et al. [2] is measured using area under the curve.

5 Conclusion

An Ensemble model comprising of multiple pre-trained models is presented in this study to aid early and accurate TB diagnosis from CXR. The proposed model is trained on one dataset and tested on an alternative dataset, resulting in a good generalization. Due to limited datasets in medical fields, the dataset were augmented along with other techniques to control overfitting. However, the individual CNN classifier achieved good results and was improved through the Ensemble classifier. The proposed model can also be deployed to detect other pulmonary abnormalities. Future work will focus on developing a robust CAD system that could accurately identify foreign objects that may be seen on CXR. Objects such as "pieces of bones, coins, rings, or button found in the chest can be likened to one of many TB manifestations which will result in misclassification.

References

1. Ahsan, M., Gomes, R., Denton, A.: Application of a convolutional neural network using transfer learning for tuberculosis detection. In: 2019 IEEE International Conference on Electro Information Technology (EIT), pp. 427–433. IEEE (2019)
2. Ayaz, M., Shaukat, F., Raja, G.: Ensemble learning based automatic detection of tuberculosis in chest x-ray images using hybrid feature descriptors. Phys. Eng. Sci. Med. **44**, 1–12 (2021)
3. Bloice, M.D., Stocker, C., Holzinger, A.: Augmentor: an image augmentation library for machine learning. arXiv preprint arXiv:1708.04680 (2017)
4. Cicero, M., et al.: Training and validating a deep convolutional neural network for computer-aided detection and classification of abnormalities on frontal chest radiographs. Invest. Radiol. **52**(5), 281–287 (2017)

5. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: a large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255. IEEE (2009)
6. Hammen, I.: Tuberculosis mimicking lung cancer. Respir. Med. Case Rep. **16**, 45–47 (2015)
7. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
8. Heo, S.J., et al.: Deep learning algorithms with demographic information help to detect tuberculosis in chest radiographs in annual workers' health examination data. Int. J. Environ. Res. Public Health **16**(2), 250 (2019)
9. Hernández, A., Panizo, Á., Camacho, D.: An ensemble algorithm based on deep learning for tuberculosis classification. In: Yin, H., Camacho, D., Tino, P., Tallón-Ballesteros, A.J., Menezes, R., Allmendinger, R. (eds.) IDEAL 2019. LNCS, vol. 11871, pp. 145–154. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-33607-3_17
10. Hijazi, M.H.A., Hwa, S.K.T., Bade, A., Yaakob, R., Jeffree, M.S.: Ensemble deep learning for tuberculosis detection using chest x-ray and canny edge detected images. IAES Int. J. Artif. Intell. **8**(4), 429 (2019)
11. Hooda, R., Mittal, A., Sofat, S.: Tuberculosis detection from chest radiographs: a comprehensive survey on computer-aided diagnosis techniques. Curr. Med. Imaging **14**(4), 506–520 (2018)
12. Hwang, S., Kim, H.E., Jeong, J., Kim, H.J.: A novel approach for tuberculosis screening based on deep convolutional neural networks. In: Medical imaging 2016: Computer-Aided Diagnosis, vol. 9785, p. 97852W. International Society for Optics and Photonics (2016)
13. Jaeger, S., Candemir, S., Antani, S., Wáng, Y.X.J., Lu, P.X., Thoma, G.: Two public chest x-ray datasets for computer-aided screening of pulmonary diseases. Quant. Imaging Med. Surg, **4**(6), 475 (2014)
14. Karnkawinpong, T., Limpiyakorn, Y.: Chest x-ray analysis of tuberculosis by convolutional neural networks with affine transforms. In: Proceedings of the 2018 2nd International Conference on Computer Science and Artificial Intelligence, pp. 90–93 (2018)
15. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. Adv. Neural Inf. Process. Syst. **25**, 1097–1105 (2012)
16. Kurt, B., Nabiyev, V.V., Turhan, K.: Medical images enhancement by using anisotropic filter and clahe. In: 2012 International Symposium on Innovations in Intelligent Systems and Applications, pp. 1–4. IEEE (2012)
17. Lopes, U., Valiati, J.F.: Pre-trained convolutional neural networks as feature extractors for tuberculosis detection. Comput. Biol. Med. **89**, 135–143 (2017)
18. Meraj, S.S., Yaakob, R., Azman, A., Rum, S.N.M., Nazri, A.: Artificial intelligence in diagnosing tuberculosis: a review. Int. J. Adv. Sci. Eng. Inf. Technol. **9**(1), 81–91 (2019)
19. Oloko-Oba, M., Viriri, S.: Diagnosing tuberculosis using deep convolutional neural network. In: El Moataz, A., Mammass, D., Mansouri, A., Nouboud, F. (eds.) ICISP 2020. LNCS, vol. 12119, pp. 151–161. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-51935-3_16
20. Oloko-Oba, M., Viriri, S.: Pre-trained convolutional neural network for the diagnosis of tuberculosis. In: Bebis, G., et al. (eds.) ISVC 2020. LNCS, vol. 12510, pp. 558–569. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-64559-5_44

21. Oloko-Oba, M., Viriri, S.: Tuberculosis abnormality detection in chest x-rays: a deep learning approach. In: Chmielewski, L.J., Kozera, R., Orłowski, A. (eds.) ICCVG 2020. LNCS, vol. 12334, pp. 121–132. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59006-2_11
22. Organization, W.H., et al.: Chest radiography in tuberculosis detection: summary of current who recommendations and guidance on programmatic approaches. World Health Organization, Technical report (2016)
23. Organization, W.H., et al.: Global tuberculosis report 2019: executive summary (2019)
24. Organization, W.H., et al.: Global tuberculosis report 2020: executive summary (2020)
25. Pan, S.J., Yang, Q.: A survey on transfer learning. IEEE Trans. Knowl. Data Eng. **22**(10), 1345–1359 (2010)
26. Parsons, L.M., et al.: Laboratory diagnosis of tuberculosis in resource-poor countries: challenges and opportunities. Clin. Microbiol. Rev. **24**(2), 314–350 (2011)
27. Pasa, F., Golkov, V., Pfeiffer, F., Cremers, D., Pfeiffer, D.: Efficient deep network architectures for fast chest x-ray tuberculosis screening and visualization. Sci. Rep. **9**(1), 1–9 (2019)
28. Rohilla, A., Hooda, R., Mittal, A.: Tb detection in chest radiograph using deep learning architecture. In: ICETETSM-17, pp. 136–147 (2017)
29. Sabour, S., Frosst, N., Hinton, G.E.: Dynamic routing between capsules. arXiv preprint arXiv:1710.09829 (2017)
30. Shorten, C., Khoshgoftaar, T.M.: A survey on image data augmentation for deep learning. J. Big Data **6**(1), 1–48 (2019)
31. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
32. Szegedy, C., et al.: Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–9 (2015)
33. Van't Hoog, A., et al.: High sensitivity of chest radiograph reading by clinical officers in a tuberculosis prevalence survey. Int. J. Tuberculosis Lung Dis. **15**(10), 1308–1314 (2011)

3.4.2 Conclusion

The proposed Ensemble technique takes advantage of the strength of different pre-trained CNN models such that a misclassified image from model "A" can be correctly classified by model "B" or "C". Therefore, fusing several classifiers has proven to achieve state-of-the-art performance and increase the rate of detection compared to any stand-alone classifier.

3.5 Ensemble of EfficientNets for the Diagnosis of Tuberculosis

3.5.1 Introduction

This section introduces a research paper that implemented a lightweight EfficientNets [53], a new CNN model that achieved state-of-the-art performance on ImageNet [9] for the classification of TB. The EfficientNets was explored due to being smaller with fewer parameters, faster, and generalized well on popular transfer learning tasks to obtain higher accuracy. Precisely, five variants of EfficientNets were fine-tuned on two benchmark CXR datasets and incorporated a global average pooling (GAP) to scale down the number of parameters and handle overfitting. The model pipeline included robust pre-processing techniques such as data augmentation that produced altered samples of the CXR and image enhancement that improved the visibility quality of the images. This paper is an extension and improvement to the performance of the work in Section 3.4.

The model was published in Computational Intelligence and Neuroscience Journal [12].

[12] **Mustapha Oloko-Oba and Serestina Viriri**. "Ensemble of EfficientNets for the Diagnosis of Tuberculosis". *Computational Intelligence and Neuroscience*, Vol. 2021, **(2021)**. DOI: `https://doi.org/10.1155/2021/9790894`

Hindawi
Computational Intelligence and Neuroscience
Volume 2021, Article ID 9790894, 12 pages
https://doi.org/10.1155/2021/9790894

Research Article

Ensemble of EfficientNets for the Diagnosis of Tuberculosis

Mustapha Oloko-Oba and Serestina Viriri

School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Durban 4000, South Africa

Correspondence should be addressed to Serestina Viriri; viriris@ukzn.ac.za

Received 12 August 2021; Revised 8 November 2021; Accepted 25 November 2021; Published 14 December 2021

Academic Editor: Giosuè Lo Bosco

Tuberculosis (TB) remains a life-threatening disease and is one of the leading causes of mortality in developing regions due to poverty and inadequate medical resources. Tuberculosis is medicable, but it necessitates early diagnosis through reliable screening techniques. Chest X-ray is a recommended screening procedure for identifying pulmonary abnormalities. Still, this recommendation is not enough without experienced radiologists to interpret the screening results, which forms part of the problems in rural communities. Consequently, various computer-aided diagnostic systems have been developed for the automatic detection of tuberculosis. However, their sensitivity and accuracy are still significant challenges that require constant improvement due to the severity of the disease. Hence, this study explores the application of a leading state-of-the-art convolutional neural network (EfficientNets) model for the classification of tuberculosis. Precisely, five variants of EfficientNets were fine-tuned and implemented on two prominent and publicly available chest X-ray datasets (Montgomery and Shenzhen). The experiments performed show that EfficientNet-B4 achieved the best accuracy of 92.33% and 94.35% on both datasets. These results were then improved through Ensemble learning and reached 97.44%. The performance recorded in this study portrays the efficiency of fine-tuning EfficientNets on medical imaging classification through Ensemble.

1. Introduction

According to the World Health Organization (WHO), tuberculosis (TB) is one of the leading causes of death [1, 2]. It is most prevalent in developing countries due to the high rate of economic distress [3]. TB can be treated if detected early to avert its further spread and mortality [4]. Early diagnosis of TB requires a reliable screening procedure and accurate interpretation of the screening outcome. To this end, chest X-ray (CXR) has been recommended by the WHO for screening pulmonary (lungs) abnormalities due to its high sensitivity, wide availability, and relatively less expensive. Despite this recommendation, adequate medical resources and skilled radiologists required to efficiently screen patients and interpret results are limited in high TB burden countries [5, 6]. Several algorithms varying from hand-crafted to support vector machines (SVMs) and convolutional neural networks (CNNs) have been deployed to automate TB screening in tackling these limitations. However, our focus is on the CNN algorithms.

Amongst these algorithms, CNNs have demonstrated promising performance, as evident in the literature primarily in transferring models trained on large datasets (ImageNet [7] and CIFAR-10 [8]), to medical imaging with fewer datasets. One of the challenges in the medical field is the lack of reliable annotated datasets and class imbalance; hence, transferring the weights of deep models helps address these challenges since CNN algorithms often depend on large datasets to learn various discriminating features contributing to attaining good results [9].

To date, many CNN models have been developed to address the challenges of TB detection, some of which are discussed here, ranging from the ones trained from scratch to pretrained CNNs and Ensemble of them.

The CNN architecture proposed in [10] is trained from scratch to diagnose TB. The architecture comprises five convolutional layers, each preceded by the pooling layers, and then the final output layer to obtain 86.2% accuracy and AUC of 0.925%. The authors incorporate Grad-CAMS and Saliency map visualizers to validate the output of the model

classification and identify areas where TB was visible on the CXR. The importance of a saliency map is further emphasized as a valuable tool to review and interpret the model decision.

A technique based on deep CNN models is presented in [11] for screening CXR to detect TB. Artificial CXR images were generated using deep convolutional generative adversarial networks (DC-GANs). These images and the CXR images from popular datasets were segmented into only the lung fields using UNET. The segmented CXR images are then used to train Vgg16 and InceptionV3 and then Ensemble both for classification. The Ensemble model achieved 97.1%, while the individual model obtained 93.8% and 96.3%, respectively. The authors affirm that the performance of the proposed model is impacted by the preprocessing technique, segmentation, augmentation, image denoising, and hyperparameter optimization.

In [12], VGG16, a variant of VGGNet [13], was employed as a feature extractor to extract discriminating features from CXR, and a logistic regression classifier was trained to predict the normal and infected images. Preprocessing techniques such as augmentation and contrast limited adaptive histogram equalization were applied to boost the quality of the CXR. The authors conclude that a large database of labeled CXR would assist in building a deeper and robust computer-aided detection system from scratch. The study shows the effectiveness of transferring the weights of a pretrained model to the medical imaging domain where there are limited annotated datasets.

GoogLeNet [14] was fine-tuned in [15] to classify different manifestations of TB as consolidation, pulmonary edema, cardiomegaly, pneumothorax, or pleural effusion. Image preprocessing was applied to normalize the CXR with shorter dimensions, while the CXR with larger dimensions was padded before training the model. The performance of the model was evaluated to determine its AUC score, sensitivity, and specificity. The model obtained 0.964 AUC and 91% for both specificity and sensitivity, which are further compared with the result of 2 board-certified radiologists.

The study presented in [16] used the pretrained CNN to confirm the presence of TB on CXR. In the study, the authors conducted two experiments by applying preprocessing such as data augmentation to CXR in the first experiment, while the second experiment was performed directly on the CXR. The model achieved an accuracy of 81.25% in the first instance and 80% in the second instance. This performance is low compared to what could be acceptable due to the sensitive nature of TB. However, it has shown data augmentation to increase data samples where necessary and proves that the more data samples you have, the more features can be learned and better the performance.

Ensemble classifier is presented in [17] for detecting TB abnormalities in the CXR. Firstly, three pretrained CNN models are trained individually with different hyperparameters to classify the CXR into normal and abnormal classes. The three models are then combined through Ensemble to improve the classification. The results are then compared, and the Ensemble model performed better than each model, achieving 86.42%.

Pretrained CNN models were customized and proposed in [18] to classify CXR abnormalities. The study applied data augmentation to generate more CXR. After the image preprocessing, CapsNet [19], VGG16, and AlexNet [20] were trained to classify the CXR into normal and abnormal classes. The classifiers were evaluated using different metrics, and the CapsNet performed best above the others concerning affine transformations. Lopes and Valiati [21] proposed three approaches (Bag of features, Simple feature extractor, and Ensemble) as feature extractors and compared their performance to detect TB.

The Ensemble in the study is made up of VggNet, ResNet [22], and GoogLeNet. The CXR was first downsampled to suit the different approaches excluding Bag of features; even though downsampling could impact the performance due to vital detail loss, experiments were then performed for all the proposed methods, and the Ensemble outperformed the other two approaches, achieving 84.6%. This performance could have been impacted due to downsampling, class imbalance, and limited datasets.

The study in [23] proposed an Ensemble of pretrained CNN models toward learning modality-specific features from different CXR collections. The knowledge learned through modality specific is transferred and fine-tuned for TB detection tasks on the Shenzhen CXR dataset. The predictions of the best performing models are combined using different ensemble methods to demonstrate improved performance over any individual model in classifying the CXR as infected or healthy. The evaluation of the proposed model achieved 94.1% accuracy.

Another study that makes use of Ensemble models is proposed in [24]. This study extracted distinctive features from edge detected images and raw CXR images for TB detection. Image preprocessing was applied before developing different classifiers to represent Ensemble and applied to both image categories. The performance was evaluated, and the Ensemble model performed best at 89.77% accuracy with 90.91% sensitivity.

The study presented in [25] is a proposed method that fine-tuned five pretrained CNNs as feature extractors to diagnose TB automatically. The CXR was collected with the patient's demographic details such as weight, height, weight, age, and gender. The authors conducted two experiments that compare the performance of the CXR with demographic information and the CXR without demographic details. The first experiment shows higher performance in terms of AUC and sensitivity over the second experience. The result proves that incorporating demographic information can positively impact performance.

In [26], VggNet and AlexNet architectures were modified to screen TB on CXR and classify them into healthy and unhealthy classes. The architectures were independently trained on the Montgomery and Shenzhen datasets to evaluate the performance of each model. VggNet achieved 81.6% classification accuracy, a slight superiority over the AlexNet architecture at 80.4%.

Deep CNN models proposed in [27] were integrated with the handcraft technique through Ensemble learning as feature extractors from CXR. The features extracted were then used as

inputs to train a classifier for detecting infected CXRs. The model was evaluated to compare the performance of both methods, and the Ensemble model performed better at 0.99 AUC.

In [28], an automatic deep learning algorithm for the diagnosis of active lung TB was presented to streamline the process of screening and detection. The algorithm was developed using the CNN comprising about 27 layers. The CXR images were resized; then, a geometric and photometric operation was applied to them before passing them on to the CNN architecture for training. The algorithm was evaluated on two public datasets and four custom datasets compared with certified radiologists and physicians' interpretations. The algorithm performance shows significant performance compared to the radiologist and physician's performance on all the datasets.

A hybrid model (MoblieNet-AEO) was proposed in [29] to classify CXR images as TB and Not-TB. The proposed model first employed MobileNet as feature extraction and then used the Artificial Ecosystem-Based Optimization (AEO) as the feature selector. The model was trained with the Shenzhen dataset and a private dataset to achieve 90.2% and 94.1% accuracy on both datasets.

Although several computer-aided detection models have been developed to diagnose TB, low accuracy and sensitivity are still major concerns that lead to misdiagnosis. Existing models are also computationally complex and expensive, which make them unaffordable in a low-budget economy. All these serve as motivation to propose an efficient system suitable and accessible for low-income economies. The contributions of this work are summarized as follows:

(1) Implementation of state-of-the-art EfficientNets to develop an effective and inexpensive TB detection system. It is the first time the EfficientNet model is being ensembled to classify CXR images for tuberculosis diagnosis.

(2) The proposed model achieved high sensitivity and accuracy.

(3) The model improved accuracy through Ensemble learning (Bagging).

(4) Finally, two benchmark datasets (Shenzhen and Montgomery) were used to evaluate the performance of our model. Despite the challenges of limited annotated datasets and class imbalance in medical fields, our approach produces significant improvements in TB classification accuracy over the related methods.

The rest of the work is structured as follows: Section 2 presents detailed data and methodology explored in this study. The experimental results and discussion are provided in Section 3, while Section 4 concludes and gives insight into future direction.

2. Data and Methods

2.1. Dataset. In this study, all the experiments were performed with the two public CXR datasets provided by the Department of Health and Human Service, Montgomery Country, Maryland, USA, and Shenzhen No. 3 People's Hospital in China [30]. These datasets were deidentified and exempted from IRB review. Both datasets described below can be accessed from the website https://lhncbc.nlm.nih.gov/publication/pub9931.

(1) The Montgomery County dataset: this dataset is a TB-specific frontal CXR dataset provided by the National Library of Medicine in conjunction with the Department of Health Services, Maryland, USA, for research intent. This dataset is composed of 58 abnormal samples and 80 normal samples. The naming of each sample ends with either 0 denoting normal sample or 1 denoting abnormal sample. All samples are 4020×4892 pixels and are available in portable network graphic (png) file format. This dataset is accompanied by clinical readings that detail each sample regarding age, sex, and manifestations.

(2) The Shenzhen dataset: this dataset is a TB-specific frontal CXR provided by Shenzhen No. 3 Hospital in Shenzhen, Guangdong, China, and is publicly available for research. The dataset is made up of 336 abnormal samples and 326 normal samples. The naming of each sample ends with either 0 denoting normal sample or 1 denoting abnormal sample. All samples are approximately 3000×3000 pixels saved in portable network graphic (png) file format. This dataset is accompanied by clinical readings that detail each sample concerning sex, age, and diagnosis. Some of the CXR in both datasets are shown in Figure 1.

2.2. Preprocessing. Image preprocessing is crucial in a deep-learning task. Since images are provided in different sizes, the input images need to be resized to conform to the different CNN models while preserving the important features. Also, the CNN model generally requires a large dataset to properly learn discriminating features suitable for making predictions and obtaining a reasonable performance [9]. Medical imaging is usually limited; hence, data augmentation [31] is applied in this study to generate additional unique CXR images. The data augmentation approach contributes toward controlling overfitting. Overfitting is a phenomenon where the model performs well on the training set but poorly on the new (unseen) test set. Contrast limited adaptive histogram equalization (CLAHE) [32] is also applied on the CXR to improve the visibility quality of the images. The augmentation types applied to the CXR images are presented in Table 1, while Table 2 shows the parameter values used to perform the CLAHE operation.

2.3. Transfer Learning. In this study, we adopt the concept of transfer learning. This phenomenon leverages the pretrained CNN model on large datasets (millions) to learn features of the target (diagnosing TB from the CXR in this case) with a limited dataset. In other words, we transfer the rich discriminative features learned by the deep CNN models on the ImageNet [7] dataset to our CXR dataset. The number of

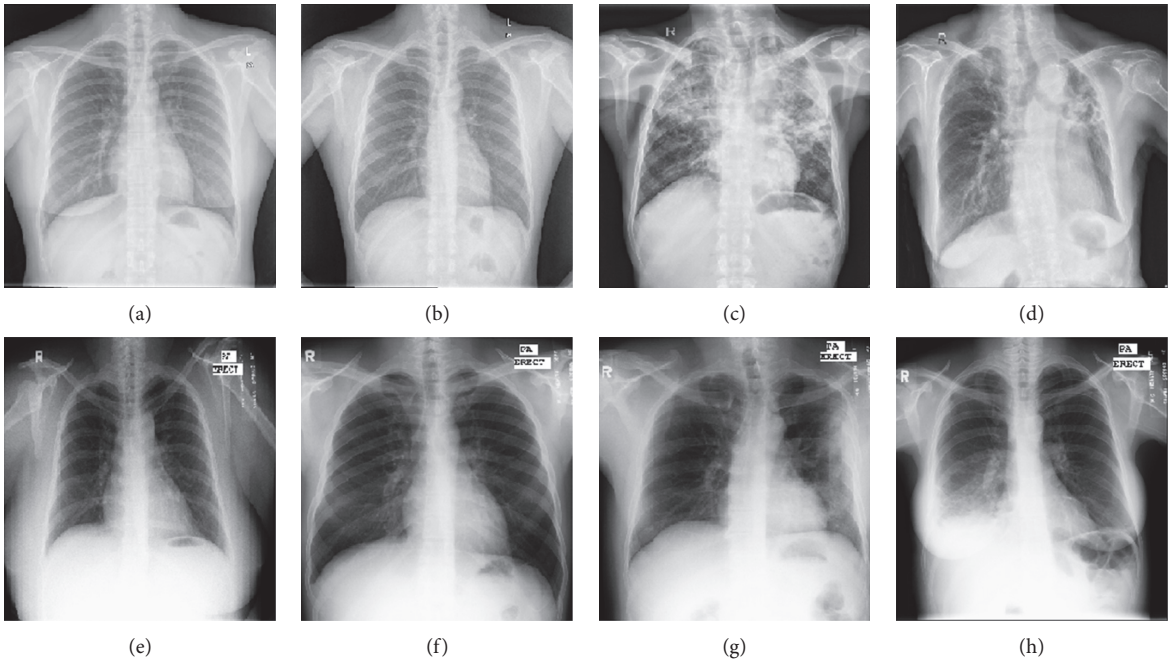


Figure 1: Training and validation CXR from both Shenzhen and Montgomery datasets. (a) Normal CXR of a 38 yr female. (b) Normal CXR of a 22 yr male. (c) Infected CXR of a 56 yr male. (d) Infected CXR of a 78 yr female. (e) Normal CXR of a 40 yr female. (f) Normal CXR of a 33 yr male. (g) Infected CXR of a 47 yr male. (h) Infected CXR of a 49 yr female. (a–d) Samples from the Shenzhen dataset. (e–h) Samples from the Montgomery dataset [30].

Table 1: Data augmentation types and probability values.

Augmentation type	Probability value
Rotation_right_left	0.5
Zoom range	0.3
Vertical flip	0.4
Horizontal flip	0.4
Height_shift_range	0.2
Width	0.2

Table 2: CLAHE parameter values.

Parameter	Value
clipLimit	2.0
tileGridSize	8×8

deep CNN model parameters increases as the network gets deeper to achieve improved efficiency. Hence, it requires many datasets for training, thereby making it computationally complex. So, applying these models directly on small and new datasets results in feature extraction bias, overfitting, and poor generalization. We, therefore, modified the pretrained CNN and fine-tuned its structure to suit the CXR dataset. The concept of transfer learning is computationally inexpensive, achieves less training time, overcomes limitations of the dataset as in the medical domain, improves performance, and is faster than training a model from scratch [33]. The pretrained CNN model fine-tuned in this work is the EfficientNets [34], and the structure of the proposed method is represented in Figure 2.

2.4. EfficientNet Architecture. EfficientNet [34] is a lightweight model based on the AutoML framework [35] to develop a baseline EfficientNet-B0 network and uniformly scaled up the depth, width, and resolutions using a simplified and effective compound coefficient to improve EfficientNet models B1–B7. The models performed efficiently and attained superiority over the existing CNN models on the ImageNet datasets, as shown in Figures 3 and 4, respectively. EfficientNets are smaller with few parameters, faster, and generalize well to obtain higher accuracy on other datasets popular for the transfer learning task. The proposed study fine-tuned EfficientNet models B0–B4 on CXR to detect TB. Due to limited annotated datasets in medical imaging, the data augmentation technique is applied to generate additional unique CXR images to control overfitting. In transferring the pretrained EfficientNets to the CXR dataset, we fine-tuned the models by adding a global average pooling (GAP) to reduce the number of parameters and handle overfitting. Two dense layers follow the GAP with a ReLU activation function and a dropout rate of 0.4 before a final dense layer that serves as output with SoftMax activation function to determine the probabilities of the input CXR representing the normal and infected classes. The SoftMax activation function is given as

$$\sigma(q)_i = \frac{e^{q_i}}{\sum_{y=1}^{N} e^{q_y}}. \quad (1)$$

where σ is the SoftMax activation function, q represents the input vector to the output layer i depicted from the

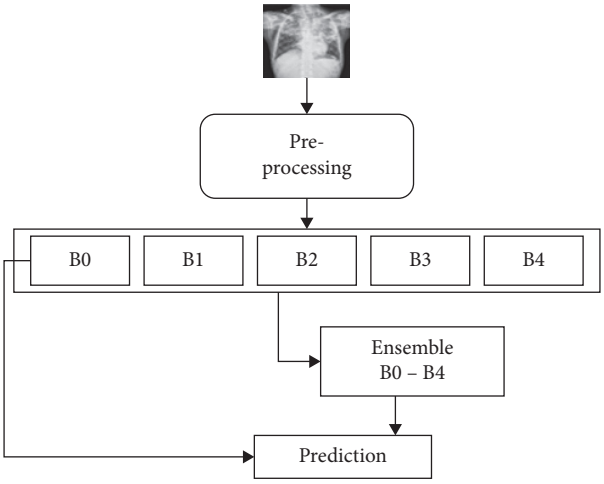


Figure 2: Block structure of the proposed model.

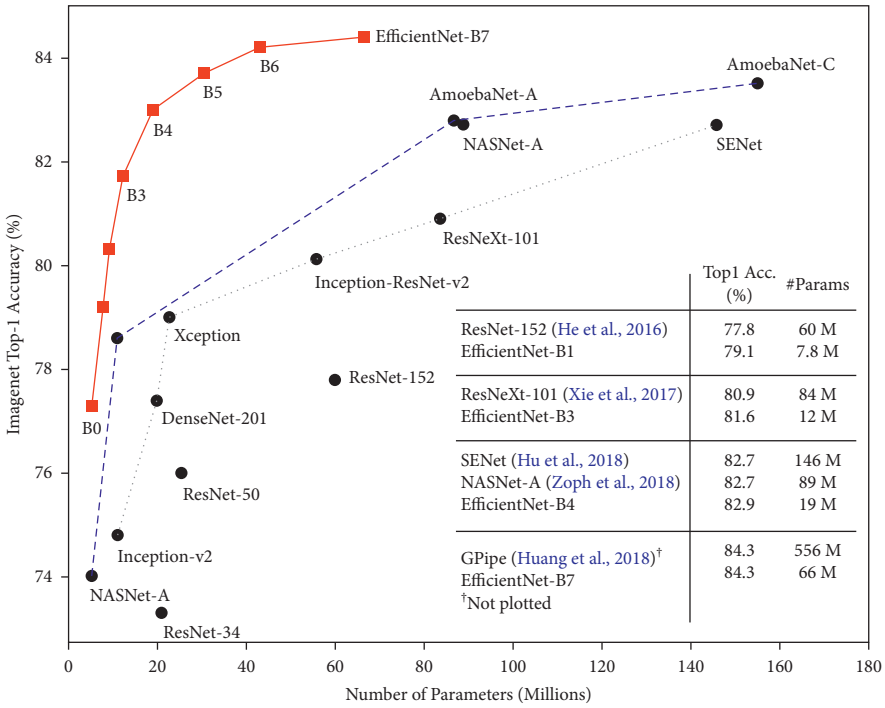


Figure 3: Comparison of EfficientNet model size with other models on ImageNet [34].

exponential element e^{q_i}, N is the number of classes, and e^{q_y} represents the output vector of the exponential function.

It is understood that too many iterations could lead to model overfitting, while too few iterations can cause model underfitting; hence, we invoked an early stopping strategy. About 120 training iterations were configured and then set up the early stopping to terminate the training once the performance stops improving on the hold-out validation set. The early stopping, L2 regularization, data augmentation, and dropout were implemented to control overfitting. The EfficientNet B0–B4 models were trained for 40 iterations (epochs) using a stochastic gradient descent (SGD) optimizer with a 0.0001 learning rate. The batch size for each iteration was 22, and momentum equals 0.9 and L2 regularizer. All these configurations help to control overfitting. At the same time, categorical cross-entropy is the loss function used to update weights at each iteration. Hyperparameters used were carefully evaluated and found to perform optimally. The optimizer employed is defined as

$$\alpha = \alpha - n \cdot \Delta_\alpha J\left(\alpha; x^{(i)}; y^{(i)}\right), \tag{2}$$

where $\Delta_\alpha J$ is the gradient of the loss w.r.t α, n is the defined learning rate, α is the weight vector, while x and y are the respective training sample and label, respectively.

2.5. Ensemble Modeling. Ensemble learning is the terminology for techniques that incorporate multiple models for decision making. The ultimate goal of Ensemble is such that by combining multiple models, the errors of a single model can be corrected (compensated) by other models, thereby making the overall result (prediction and classification) of the Ensemble better than any single participating model [36]. The concept of Ensemble is mainly considered the interpretation of machine learning for the intelligence of the majority.

In recent times, Ensembles are regarded as state-of-the-art methods for solving a more significant number of machines and deep learning challenges [37]. In other words, weighing and aggregating many individual viewpoints will be better than deciding on the judgment of a single individual.

(1) Strength of Ensemble learning: the efficiency of Ensemble models for obtaining improved performance could be due to the following [36, 38]:

Avoidance of overfitting: algorithms often find ways to perfectly predict the training set in a few datasets even though poor prediction is made on the test set. So when a different hypothesis is averaged, the risk of selecting the wrong hypothesis is minimized.
Computational expediency: individual models can get stuck in a local minimum, but through (Ensemble) combining models, the risk of local minimum is decreased as other models can get past a certain local minimum.
Representation: Ensemble models help improve the scenario where the optimal hypothesis is outside the scope of any individual model.
Class imbalance and dimensionality: they are major problems that affect the equality of results and are well handled by Ensemble model through resampling.

(2) Building an Ensemble model: building an Ensemble model culminates in methodology selection to train the individual participating models and select a suitable process for combining the individual predictions (output). The high diversity and predictive performance of a model are essential in selecting participating models [38]. Other considerations for model selections include input and output manipulations, partitioning, and hybridization. Furthermore, the methods of integrating all individual outputs into a single final result are crucial. There are different Ensemble types based on the generation process of different models and a combination strategy. The most prominent Ensemble types are Bagging [39], Boosting [40], and Stacking [41]. In this study, the Bagging method of Ensemble is implemented. This method is also known as Bootstrap Aggregation.

(3) Bagging [39] is an effective method for creating Ensembles of independent models. Many individual models are trained independently with a sample of data derived from the original datasets as replacements. By independently, we mean the training and output of a particular model do not affect the others. To ensure diversity, each replica sample has the same number of samples as the original dataset. Thus, some data samples may appear for the replica, while others may not or appear more than once for training individual models. Once the classification output of the models is obtained, bagging then combines them through majority voting to obtain the final prediction. Algorithm 1 presents bagging pseudocode. The bagging method is employed in this study because of the following:

It is suitable for a task with small training datasets.
It can achieve low variance without impacting the bias.
It handles classification tasks well
It has the ability to obtain improved accuracy.

3. Results and Discussion

Various EfficientNet (B0–B4) variants were fine-tuned on both Shenzhen and Montgomery TB-specific CXR datasets to detect TB. Each dataset is split into a 75% training set and 25% test set, respectively. The experiments were entirely performed using the Keras deep learning framework running on the TensorFlow backend. Each model was trained in the cloud on the Tesla graphics processing unit (GPU) and made available through the Google Collaboratory framework by Google. The Collaboratory framework provides up to 12 GB random access memory (RAM) and about 360 GB GPU in the cloud for research purposes. The models were evaluated using the popular evaluation metrics (accuracy, sensitivity, specificity, and area under the curve). The way these evaluation metrics are determined is given as follows:

$$\begin{aligned}
\text{accuracy} &= \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}},\\
\text{sensitivity} &= \frac{\text{TP}}{\text{TP} + \text{FN}},\\
\text{specificity} &= \frac{\text{TN}}{\text{TN} + \text{FP}},\\
\text{TPR} = \text{sensitivity} &= \frac{\text{TP}}{\text{TP} + \text{FN}},\\
\text{FPR} = 1 - \text{specificity} &= \frac{\text{FP}}{\text{FP} + \text{TN}},
\end{aligned} \tag{3}$$

where TP = true positive, FP = false positive, TN = true negative, FN = false negative, TPR = true-positive rate, and FPR = false-positive rate.

The experimental results showing each dataset's accuracy and loss are presented in Figures 5 and 6. The various EfficientNet models converge at the 40th iteration (epoch); then, the accuracy is determined using the test set. The models performed well at extracting and learning discriminative features from the CXR. Precisely, EfficientNetB4 attains the best accuracy (92.33% and 94.35%) for both datasets over the other variants. The performance details of

Model	Top-1 Acc. (%)	Top-5 Acc. (%)	#Params	Ratio-to-EfficientNet	#FLOPs	Ratio-to-EfficientNet
EfficientNet-B0	77.1	93.3	5.3 M	1 x	0.39 B	1 x
ResNet-50 (He et al., 2016)	76.0	93.0	26 M	4.9 x	4.1 B	11 x
DenseNet-169 (Huang et al., 2017)	76.2	93.2	14 M	2.6 x	3.5 B	8.9 x
EfficientNet-B1	79.1	94.4	7.8 M	1 x	0.70 B	1 x
ResNet-152 (He et al., 2016)	77.8	93.8	60 M	7.6 x	11 B	16 x
DenseNet-264 (Huang et al., 2017)	77.9	93.9	34 M	4.3 x	6.0 B	8.6 x
Inception-v3 (Szegedy et al., 2016)	78.8	94.4	24 M	3.0 x	5.7 B	8.1 x
Xception (Chollet, 2017)	79.0	94.5	23 M	3.0 x	8.4 B	12 x
EfficientNet-B2	80.1	94.9	9.2 M	1 x	1.0 B	1 x
Inception-v4 (Szegedy et al., 2017)	80.0	95.0	48 M	5.2 x	13 B	13 x
Inception-resnet-v2 (Szegedy et al., 2017)	80.1	95.1	56 M	6.1 x	13 B	13 x
EfficientNet-B3	81.6	95.7	12 M	1 x	1.8 B	1 x
ResNeXt-101 (Xie et al., 2017)	80.9	95.6	84 M	7.0 x	32 B	18 x
PolyNet (Zhang et al., 2017)	81.3	95.8	92 M	7.7 x	35 B	19 x
EfficientNet-B4	82.9	96.4	19 M	1 x	4.2 B	1 x
SENet (Hu et al., 2018)	82.7	96.2	146 M	7.7 x	42 B	10 x
NASNet-A (Zoph et al., 2018)	82.7	96.2	89 M	4.7 x	24 B	5.7 x
AmoebaNet-A (Real et al., 2019)	82.8	96.1	87 M	4.6 x	23 B	5.5 x
PNASNet (Liu et al., 2018)	82.9	96.2	86 M	4.5 x	23 B	6.0 x
EfficientNet-B5	83.6	96.7	30 M	1 x	9.9 B	1 x
AmoebaNet-C (Cubuk et al., 2019)	83.5	96.5	155 M	5.2 x	41 B	4.1 x
EfficientNet-B6	84.0	96.8	43 M	1 x	19 B	1 x
EfficientNet-B7	84.3	97.0	66 M	1 x	37 B	1 x
GPipe (Huang et al., 2018)	84.3	97.0	557 M	8.4 x	-	-

Figure 4: Performance of EfficientNets vs state-of-the-art models on ImageNet [34].

Input:
Data set $D = \{(x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m)\}$;
Base learning algorithm L;
Number of learning rounds T
Process:
for $t = 1, \ldots T$:
D_t = Bootstrap (D); % *Generate a bootstrap sample from D*
$h_t = L(D_t)$; % *Train a base learner* h_t from the bootstrap sample
end.
Output: $H(x) = \text{argmax}_{x_{y \in Y}} \sum_{t=1}^{T} 1(y = h_t(x))$ % the value of 1 (a) is 1 if a is true and 0 if otherwise

Algorithm 1: Bagging Ensemble.

the experiments are illustrated in Table 3, depicting accuracy, sensitivity, specificity, and AUC.

Despite achieving good performance, we further build an Ensemble model comprising the best performing individual models (B2, B3, and B4) from the initial experiments to improve accuracy. The performance of the three EfficientNets was averaged to build an Ensemble to classify the CXR. The Ensemble model outperformed the individual variants of EfficientNets in diagnosing TB, achieving 95.82% for the Montgomery dataset and 97.44% for the Shenzhen dataset. The accuracy and loss plots for the Ensemble model are presented in Figure 7. Also, the model obtains the highest sensitivity for TB detection.

One of the main advantages of EfficientNets is that they are smaller with fewer parameters and faster and generalize well on transfer learning datasets [34]. EfficientNet-B0 had the least performance, as shown in Table 3, despite having the fewest parameters. This low performance could result from image downsampling to conform to the model's 224×224 input size. A detailed number of parameters and input size for each EfficientNet are provided in Table 4. We observed that performance improves as the model gets deeper. For the Montgomery dataset, EfficientNet-B0 started poorly. It only began to converge from the 11th iteration, although with little noise, until the 25th iteration, where it began to stabilize until the 40th iteration. Overfitting starts to occur after the 40th iteration; hence, early stoppage was invoked. On the contrary, the training progressed well for the Shenzhen dataset but only began to overfit after the 40th iteration. Generally, the models performed better on the

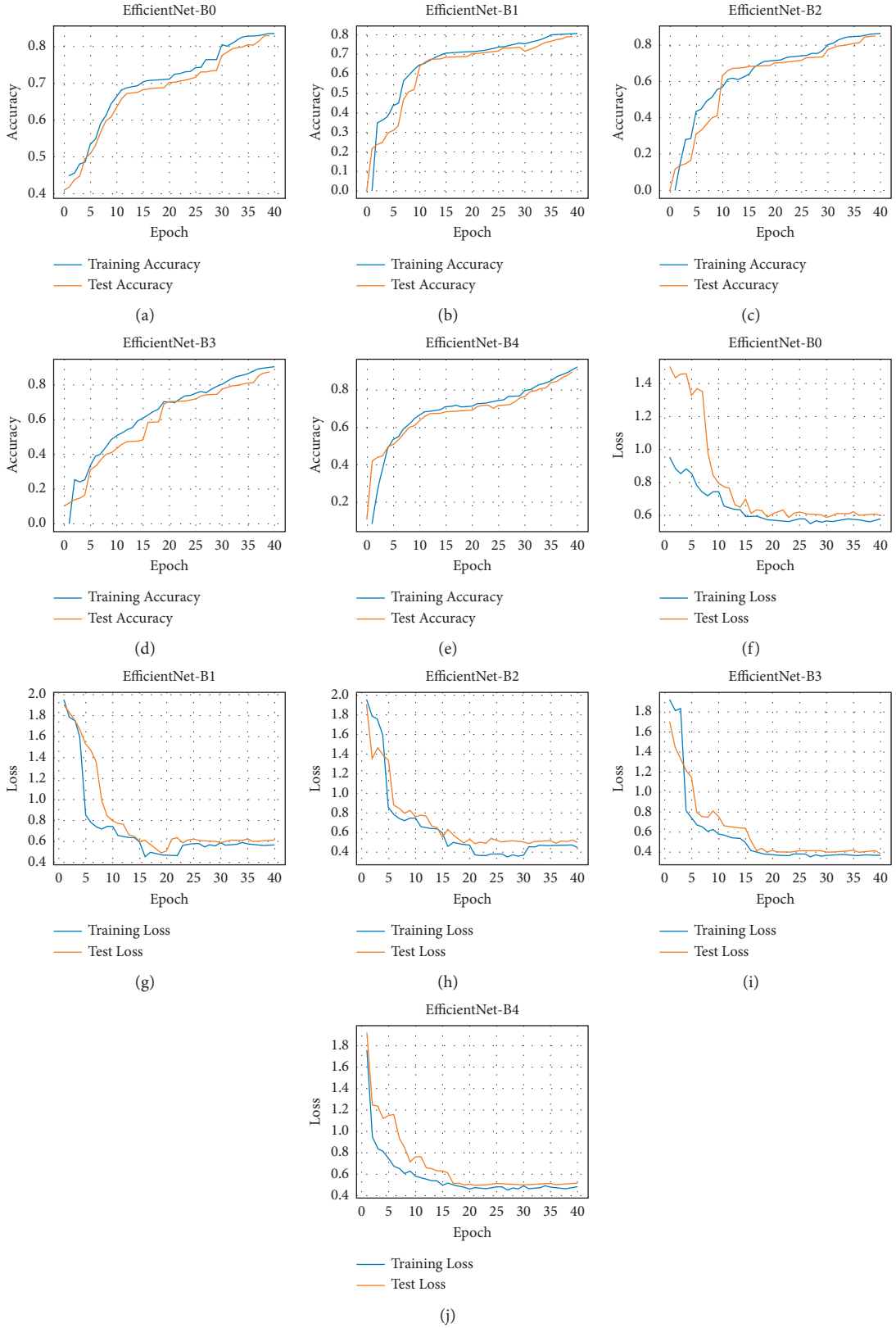


Figure 5: Accuracy and loss for the Shenzhen dataset: (a–e) performance accuracy of each EfficientNet and (f–j) the corresponding loss.

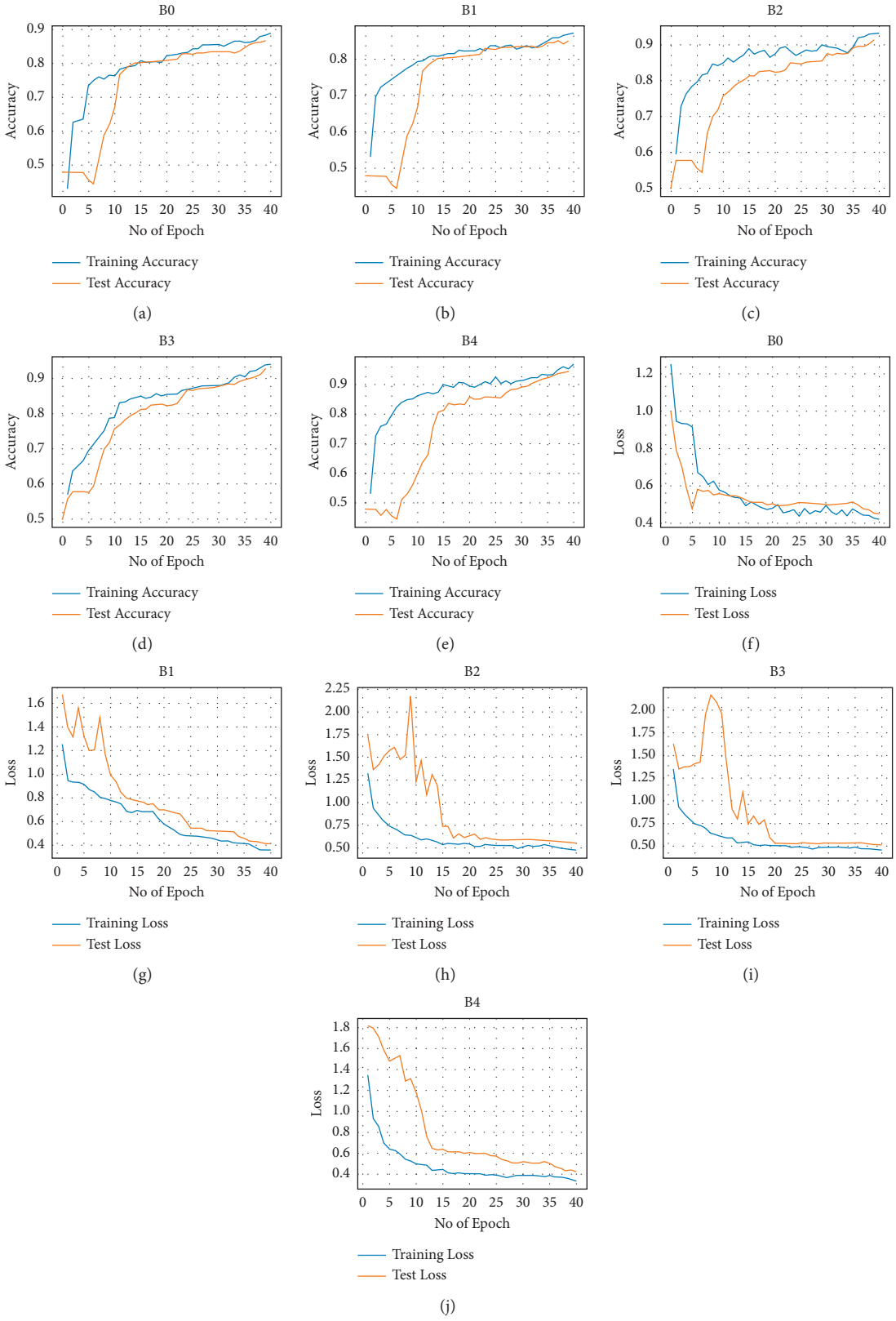


Figure 6: Accuracy and loss for the Montgomery dataset: (a–e) performance accuracy of each EfficientNet and (f–j) the corresponding loss.

Table 3: Experimental results.

EfficientNet model	Montgomery				Shenzhen			
	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC
B0	80.52	82.10	81.48	0.77	84.90	87.08	85.36	0.82
B1	83.46	83.78	83.77	0.80	86.62	89.12	90.81	0.86
B2	86.35	90.16	88.25	0.84	91.24	90.83	90.31	0.90
B3	90.40	93.01	92.00	0.90	92.83	93.64	89.53	0.92
B4	92.33	95.41	93.61	0.92	94.35	95.69	94.15	0.93
Ensemble	95.82	98.13	95.78	0.94	97.44	99.18	96.21	0.96

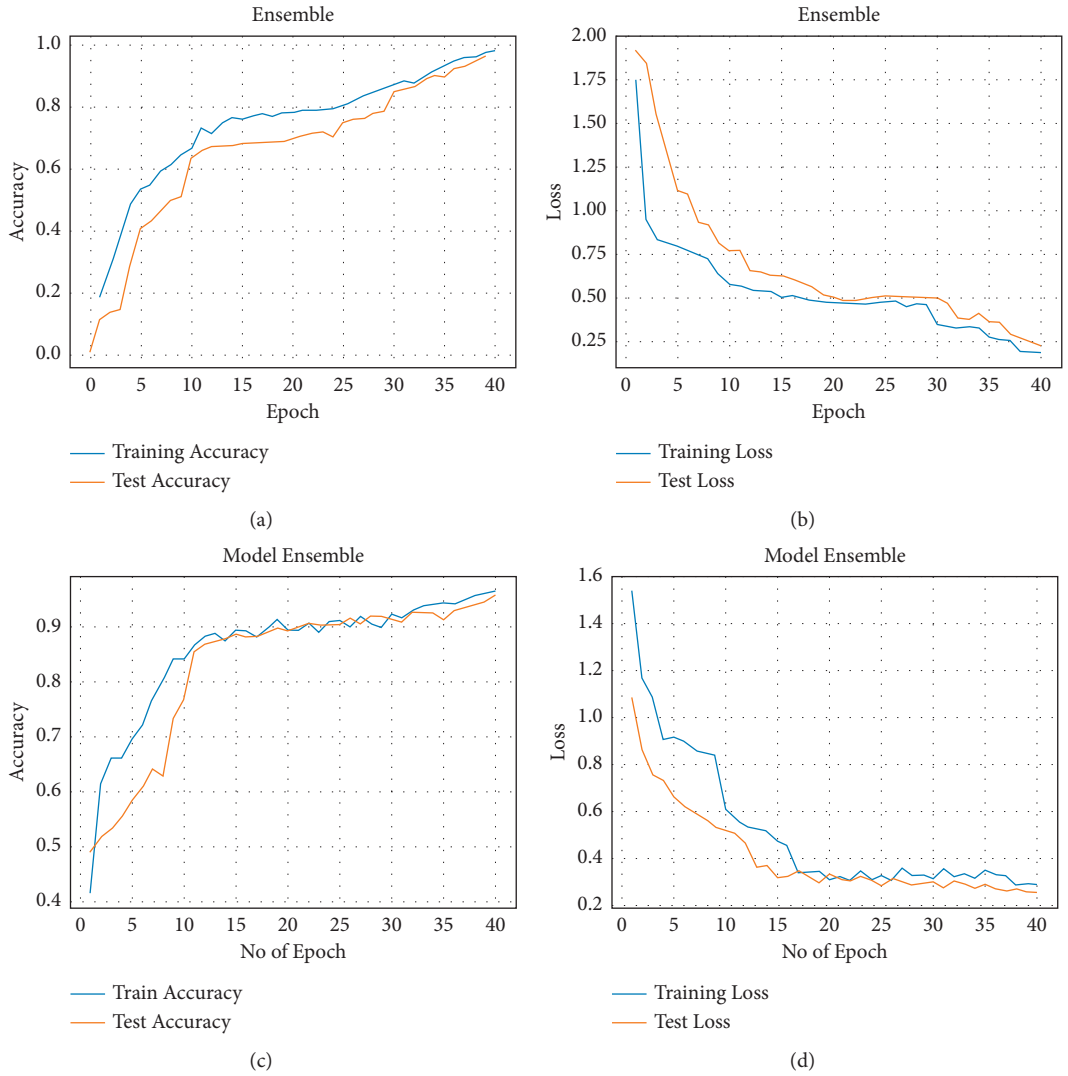


Figure 7: Accuracy and loss of the Ensemble model: (a, b) performance on the Shenzhen dataset and (c, d) performance on the Montgomery dataset.

Table 4: Parameter values and input sizes of each EfficientNet.

Model	Parameter (million) (M)	Input size
B0	5.3	224×224
B1	7.8	240×240
B2	9.2	260×260
B3	12	300×300
B4	19	380×380

Table 5: Proposed method compared with the related Ensemble study.

Author	Result (%)
Lopes and Valiati [21]	84.6
Rajaraman and Antani [23]	94.1
Hernández et al. [17]	86.4
Hijazi et al. [24]	89.7
Dasanayaka and Dissanayake [11]	97.1
Pasa et al. [10]	86.2
Sahlol et al. [29]	94.1
Proposed method	97.4

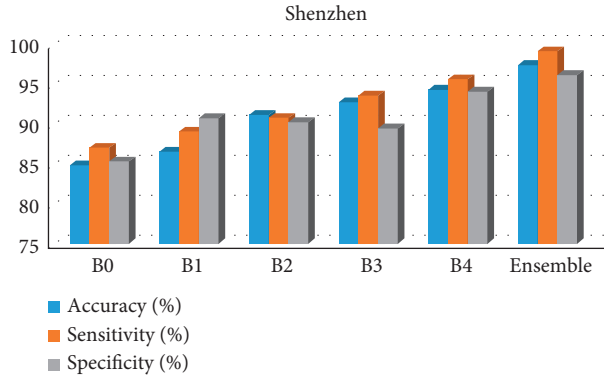


Figure 8: Best results achieved on the Shenzhen dataset.

Shenzhen dataset. This attribute could be due to the increased dataset and balanced class ratio of normal to abnormal CXR. One of the challenges faced during the experiments was determining the best combinations of hyperparameters suitable for these models and overfitting due to limited data samples. Hence, we explored data augmentation and other regularization techniques (dropout) to combat overfitting as much as possible. The result of the proposed method compared with the related Ensemble study is presented in Table 5.

4. Conclusion

This study investigated and implemented EfficientNet models for automatic diagnosis of TB on the two most prominent and publicly available CXR datasets. EfficientNets that achieved state-of-the-art performance over other architectures to maximize accuracy and efficiency were explored and fine-tuned on CXR images. The fine-tuning technique is a valuable way to take advantage of rich generic features learned from large dataset sources such as ImageNet to compliment the lack of annotated datasets affecting medical domains. The experimental results show the effectiveness of the EfficientNets in extracting and learning distinctive features from the CXR images and then classifying them into a healthy or infected class. Out of the five EfficientNet variants explored in this study, the EfficientNet-B4 outperformed among the others, as evident in Table 3 and depicted in Figure 8. The achieved result is improved through Ensemble of the best three (B2, B3, and B4). It is worthy of note that exploring EfficientNets for the classification of TB images helps save time while preserving accuracy. However, one of the major limitations of the proposed approach is training the model on small datasets and training on images with low resolutions. These limitations can easily result in significant overfitting. Hence, it is essential to explore effective image preprocessing techniques to overcome these challenges. For future study, we will consider Ensembling all variants of EfficientNets and compare with other state-of-the-art models for detecting foreign objects on CXR images that could affect performance or be misclassified as TB. Although the proposed methodology is specific to TB detection, it could be extended to detect other pulmonary abnormalities, given the right set of datasets.

Data Availability

Data used in this research are publicly available (Shenzhen and Montgomery TB-specific CXR datasets). The developed model can be shared upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] W. H. Organization, *Global Diffusion of eHealth: Making Universal Health Coverage Achievable: Report of the Third Global Survey on eHealth*, World Health Organization, Geneva, Switzerland, 2017.

[2] W. H. Organization, "Global status report on alcohol and health: executive summary," Technical Reports Series, World Health Organization, Geneva, Switzerland, 2018.

[3] J. R. Hargreaves, D. Boccia, C. A. Evans, M. Adato, M. Petticrew, and J. D. H. Porter, "The social determinants of tuberculosis: from evidence to action," *American Journal of Public Health*, vol. 101, no. 4, pp. 654–662, 2011.

[4] W. H. Organization, *Global Tuberculosis Report: Executive Summary*, World Health Organization, Geneva, Switzerland, 2020.

[5] A. H. v. t. Hoog, H. K. Meme, H. Van Deutekom et al., "High sensitivity of chest radiograph reading by clinical officers in a tuberculosis prevalence survey," *International Journal of Tuberculosis & Lung Disease*, vol. 15, no. 10, pp. 1308–1314, 2011.

[6] R. C. Malli, M. Aygun, and H. K. Ekenel, "Apparent age estimation using ensemble of deep learning models," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pp. 714–721, Las Vegas, NV, USA, June 2016.

[7] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255, IEEE, Floroida, FL, USA, June 2009.

[8] A. Krizhevsky and G. Hinton, *Learning Multiple Layers of Features from Tiny Images*, 2009.

[9] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, vol. 6, no. 1, pp. 1–48, 2019.

[10] F. Pasa, V. Golkov, F. Pfeiffer, D. Cremers, and D. Pfeiffer, "Efficient deep network architectures for fast chest x-ray tuberculosis screening and visualization," *Scientific Reports*, vol. 9, no. 1, pp. 6268-6269, 2019.

[11] C. Dasanayaka and M. B. Dissanayake, "Deep learning methods for screening pulmonary tuberculosis using chest x-rays," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, pp. 1–11, 2020.

[12] M. Oloko-Oba and S. Viriri, "Pre-trained convolutional neural network for the diagnosis of tuberculosis," in *Proceedings of the International Symposium on Visual Computing*, pp. 558–569, Springer, California, CA, USA, October 2020.

[13] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, https://arxiv.org/abs/1409.1556.

[14] C. Szegedy, W. Liu, Y. Jia et al., "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–9, Boston, MA, USA, June 2015.

[15] M. Cicero, A. Bilbily, E. Colak et al., "Training and validating a deep convolutional neural network for computer-aided detection and classification of abnormalities on frontal chest radiographs," *Investigative Radiology*, vol. 52, no. 5, pp. 281–287, 2017.

[16] M. Ahsan, R. Gomes, and A. Denton, "Application of a convolutional neural network using transfer learning for tuberculosis detection," in *Proceedings of the 2019 IEEE International Conference on Electro Information Technology (EIT)*, pp. 427–433, IEEE, Chicago, IL, USA, 2019.

[17] A. Hernández, Á. Panizo, and D. Camacho, "An ensemble algorithm based on deep learning for tuberculosis classification," in *Proceedings of the International Conference on Intelligent Data Engineering and Automated Learning*, pp. 145–154, Springer, Manchester, UK, November 2019.

[18] T. Karnkawinpong and Y. Limpiyakorn, "Chest x-ray analysis of tuberculosis by convolutional neural networks with affine transforms," in *Proceedings of the 2018 2nd International Conference on Computer Science and Artificial Intelligence*, pp. 90–93, Shenzhen, China, October 2018.

[19] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic routing between capsules," 2017, https://arxiv.org/abs/1710.09829.

[20] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, vol. 25, pp. 1097–1105, 2012.

[21] 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.

[22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, Las Vegas, NV, USA, June 2016.

[23] S. Rajaraman and S. K. Antani, "Modality-specific deep learning model ensembles toward improving tb detection in chest radiographs," *IEEE access: Practical Innovations, Open Solutions*, vol. 8, pp. 27318–27326, 2020.

[24] M. H. A. Hijazi, S. Kieu Tao Hwa, A. Bade, R. Yaakob, and M. Saffree Jeffree, "Ensemble deep learning for tuberculosis detection using chest x-ray and canny edge detected images," *IAES International Journal of Artificial Intelligence*, vol. 8, no. 4, p. 429, 2019.

[25] S.-J. Heo, Y. Kim, S. Yun et al., "Deep learning algorithms with demographic information help to detect tuberculosis in chest radiographs in Annual Workers' health examination data," *International Journal of Environmental Research and Public Health*, vol. 16, no. 2, p. 250, 2019.

[26] A. Rohilla, R. Hooda, and A. Mittal, "Tb detection in chest radiograph using deep learning architecture," in *Proceedings of the 5th International Conference on Emerging Trends in Engineering, Technology, Science and Management (ICETETSM-17), Ghaziabad, India*, pp. 136–147, September 2017.

[27] M. Ayaz, F. Shaukat, and G. Raja, "Ensemble learning based automatic detection of tuberculosis in chest x-ray images using hybrid feature descriptors," *Physical and Engineering Sciences in Medicine*, vol. 44, no. 1, pp. 183–194, 2021.

[28] E. J. Hwang, S. Park, K.-N. Jin et al., "Development and validation of a deep learning-based automatic detection algorithm for active pulmonary tuberculosis on chest radiographs," *Clinical Infectious Diseases*, vol. 69, no. 5, pp. 739–747, 2019.

[29] A. T. Sahlol, M. Abd Elaziz, A. Tariq Jamal, R. Damaševičius, and O. Farouk Hassan, "A novel method for detection of tuberculosis in chest radiographs using Artificial ecosystem-based optimisation of deep neural network features," *Symmetry*, vol. 12, no. 7, p. 1146, 2020.

[30] S. Jaeger, S. Candemir, S. Antani, Y. X. Wáng, 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. 475–7, 2014.

[31] M. D. Bloice, C. Stocker, and A. Holzinger, "Augmentor: an image augmentation library for machine learning," 2017, https://arxiv.org/abs/1708.04680.

[32] B. Kurt, V. V. Nabiyev, and K. Turhan, "Medical images enhancement by using anisotropic filter and clahe," in *Proceedings of the 2012 International Symposium on Innovations in Intelligent Systems and Applications*, pp. 1–4, IEEE, Trabzon, Turkey, July 2012.

[33] N. Donges, "What is transfer learning? exploring the popular deep learning approach," *Built In*, 2019.

[34] M. Tan and Q. Le, "Efficientnet: rethinking model scaling for convolutional neural networks," in *Proceedings of the International Conference on Machine Learning*, pp. 6105–6114, PMLR, Long Beach, CA, USA, June 2019.

[35] M. Tan, *Mnasnet: Towards Automating the Design of Mobile Machine Learning Models*, 2018.

[36] T. G. Dietterich, "Ensemble learning," *The handbook of brain theory and neural networks*, vol. 2, pp. 110–125, 2002.

[37] O. Sagi and L. Rokach, "Ensemble learning: a survey," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 8, no. 4, Article ID e1249, 2018.

[38] R. Polikar, "Ensemble learning," *Ensemble Learning, in Ensemble Machine Learning*, , pp. 1–34, Springer, 2012.

[39] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, no. 2, pp. 123–140, 1996.

[40] Y. Freund, R. Schapire, and N. Abe, "A short introduction to boosting," *Journal of Japanese Society for Artificial Intelligence*, vol. 14, pp. 771–780, 1999.

[41] D. H. Wolpert, "Stacked generalization," *Neural Networks*, vol. 5, no. 2, pp. 241–259, 1992.

3.5.2 Conclusion

The fine-tuned EfficientNets in this study was a valuable way to take advantage of rich generic features learned from large dataset sources like ImageNet to complement the lack of annotated datasets affecting medical domains. The experimental results showed the effectiveness of the EfficientNets in extracting and learning distinctive features from the CXR images and then classifying them into a healthy or infected class. The three best results obtained from the experiment of the five variants were combined through bagging Ensemble to improve high accuracy and sensitivity. Implementing EfficientNets saves time, eliminates the complexity of training models from the beginning, eliminates high computing power and challenges of limited datasets.

Chapter 4

Results and Discussion

This chapter presents a comprehensive analysis and discussion of the experimental results of the proposed CNN models presented in chapter 3 for TB detection. The discussion of results includes detailed configurations of the models and the effect of the design choices on the classification performance, pre-processing techniques, features extraction, feature classification, and finally, a comparison of the achieved results with the results obtained in the literature. The TB detection and classification models discussed here include the novel CNN, modified pre-trained CNN, Ensembled of Deep CNN, and the state-of-the-art EfficientNet models.

4.1 Classification with proposed CNN models

This section discusses the results of our TB classification models described in the published papers [29, 31] and presented in Sections 3.1 and 3.2. These models have demonstrated the potential to achieve the best results because there were trained directly on TB-specific datasets. Still, their performances are hindered due to the lack of a large number of annotated datasets in the medical field.

Generally, CNN models require a large number of datasets to properly learn discriminating features suitable for making predictions and obtaining a reasonable performance [48]. Also, CXR is provided in different sizes; hence, it needs to be resized to conform to the input channel of different models. Figure 4.1 is the sample of CXR from the Montgomery dataset.

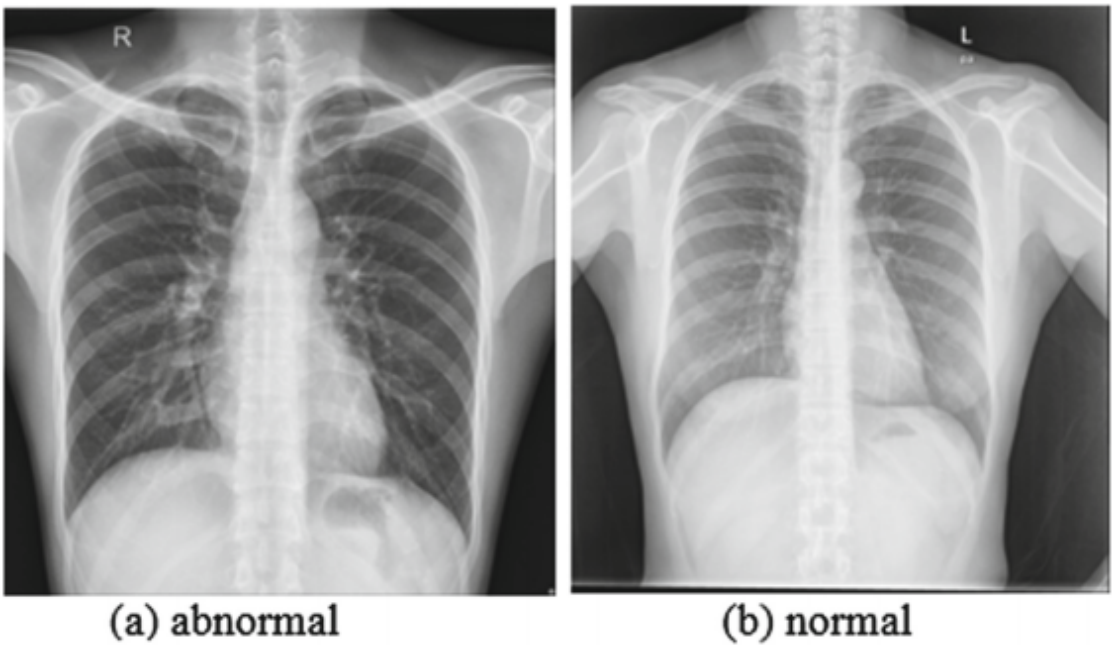


Figure 4.1: Normal and abnormal CXR

The results of our CNN model published in [29] employed pre-processing techniques through data augmentation [5]to considerably increase the diversity of the dataset and resize the images appropriately. The pre-processing augmentation categories applied were flipping, cropping, and rotation. The CXR images were then propagated through the convolutional layers of the proposed CNN architecture for features extraction. The Softmax classifier then processed the extracted features to determine probability and generate the prediction for each image as either normal or abnormal. Figure 4.2 shows the architecture of the model while the detailed configuration of the proposed TB detection model is presented in Table 4.1.

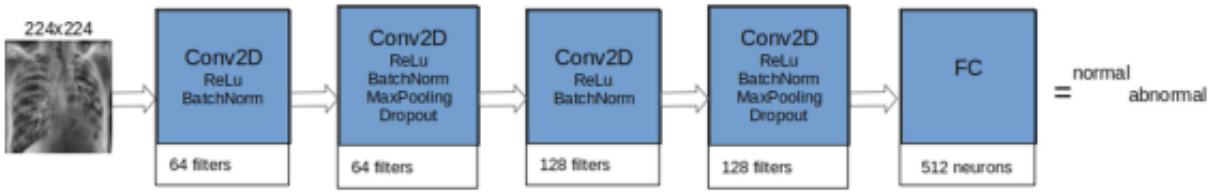


Figure 4.2: Proposed CNN model

The experimental results of the proposed model was evaluated to determine the classification performance using the accuracy metric. Figure 4.3 shows the validation accuracy and corresponding loss with the confusion matrix presented in Figure 4.4

Table 4.1: Detailed configuration of the proposed TB detection model

Layers	Output size	Filter
INPUT IMAGE	224 x 224 x 3	
CONVO1	112 x 112 x 64	3 x 3
ACTN	112 x 112 x 64	
BATCHNORM	112 x 112 x 64	
CONVO2	112 x 112 x 64	3 x 3
ACTN	112 x 112 x 64	
BATCHNORM	112 x 112 x 64	
MAXPOOL	56 x 56 x 64	2 x 2
DROPOUT	56 x 56 x 64	
CONVO3	56 x 56 x 128	3 x 3
ACTN	56 x 56 x 128	
BATCHNORM	56 x 56 x 128	
CONV4	56 x 56 x 128	3 x 3
ACTN	56 x 56 x 128	
BATCHNORM	56 x 56 x 128	
MAXPOOL	28 x 28 x 128	2 x 2
DROPOUT	28 x 28 x 128	
FULLY CONN	512	
ACTN	512	
BATCHNORM	512	
DROPOUT	512	
SOFTMAX	2	

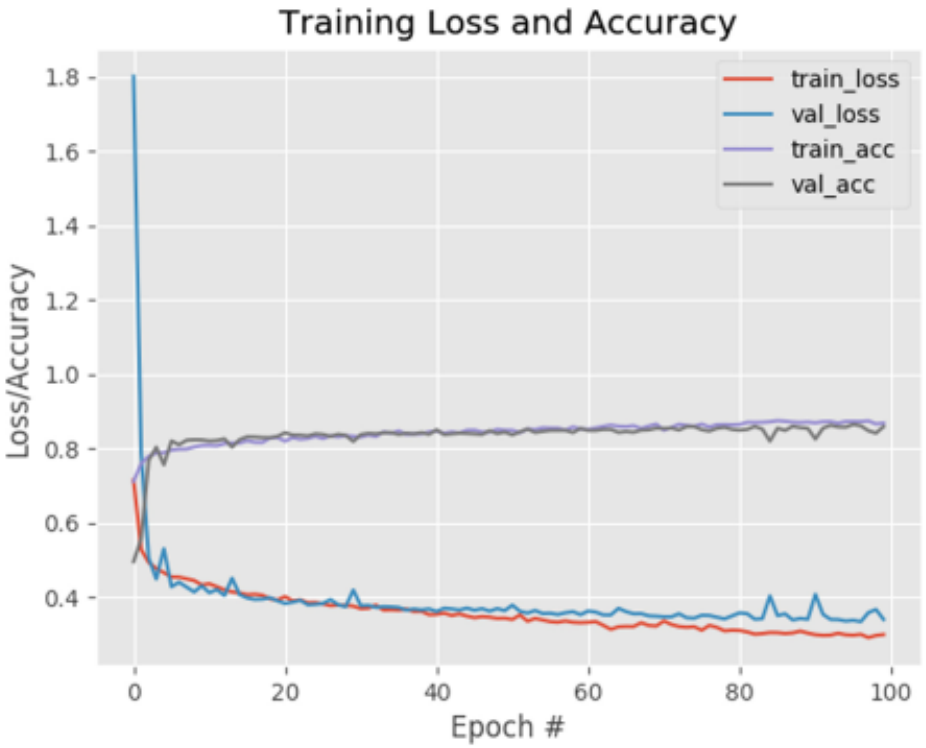


Figure 4.3: Model accuracy and loss [29]

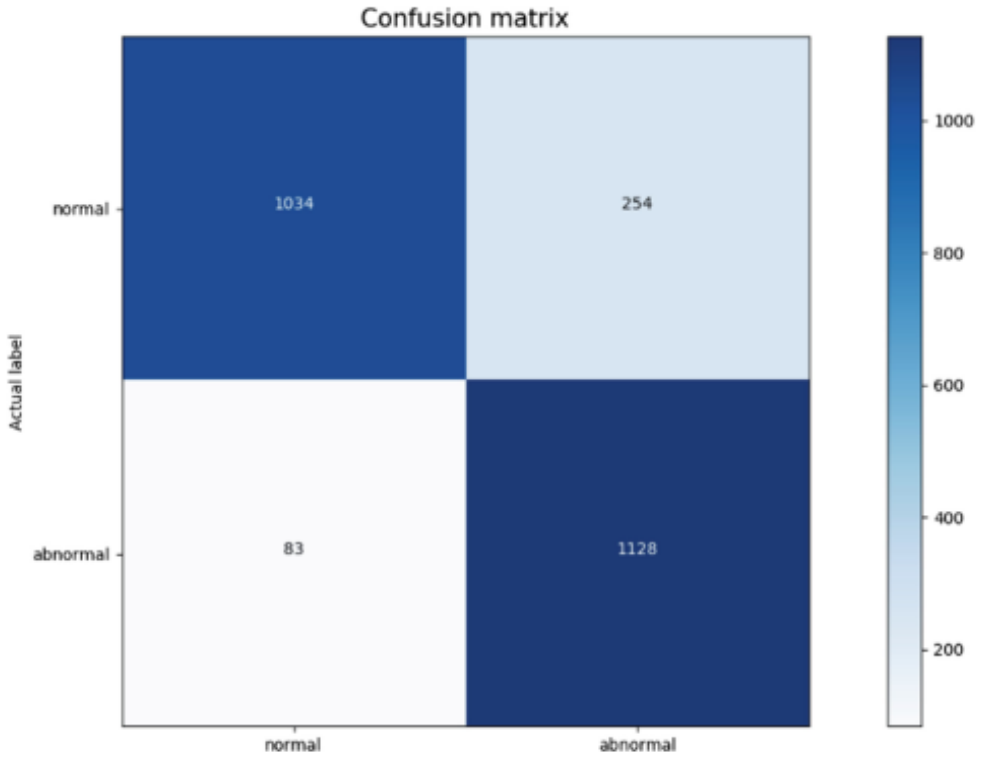


Figure 4.4: Model confusion matrix [29]

It can be observed from the accuracy graph in Figure 4.3 how overfitting started to occur initially but was controlled through pre-processing techniques and parameter tuning to reach a good convergence and achieved 87.1% classification accuracy. The results is further confirmed on the confusion matrix in Figure 4.4 that evaluated the performance of the TB classification model on the test sets and presented the summary in a Table using four combinations of actual versus predicted classes. This results was comparable to the performance of models reported in the literature some of which are discussed briefly as follows: A CNN approach consisting of seven convolutional layers and three fully connected layers to classify CXR as positive or negative for TB was presented in [15]. The model evaluation compared three different optimizers and reported the best accuracy on the Adam optimizer at 82%. A ConvNet model involving classifications of different manifestations of TB was presented in [24]. The work looked at unbalanced, less categorized CXR scans and incorporated cross-validation with sample shuffling in training the model. The TB datasets comprising 4701 image samples with about 4248 samples marked as abnormal containing six manifestations of TB and 453 samples marked as normal were used for the experiment to obtain 85.6% classification accuracy. Details of the proposed model as compared is shown in Table 4.2.

Another of our proposed CNN model presented in [31] was an improvement to the results obtained in [29]. The configuration followed a similar design with few modifications, leading to slight but meaningful improvement. The performance

Table 4.2: Model accuracy as compared

Ref	Accuracy (%)
Lopes and Valiati [25]	82.6, 84.7
Hwang *et al.* [17]	83.7
Rohilla *et al.* [44]	80.4, 81.6
Hooda *et al.* [15]	82
Liu *et al.* [24]	85
Oloko-Oba and Viriri [29]	**87.1**

accuracy of the model is presented in Figure 4.5 and Figure 4.7 presents some of the predicted images as confirmed with the ground-truth.

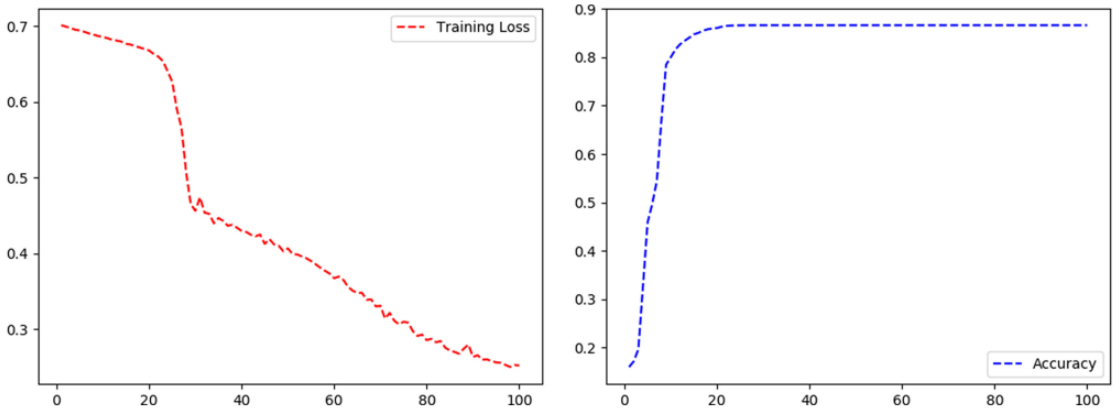


Figure 4.5: Model results [31]

The major differences between our classification model [29] in Section 3.1 and the improved model [31] in Section 3.2 is that, the latter

- Is deeper in terms of layers configuration
- Employed a different dataset (Shenzhen) that is four times the number of (Montgomery)
- Tweaks the hyper-parameters, and
- Improved the pre-processing techniques through Contrast Limited Adaptive Histogram Equalization (CLAHE) [22] to improve the visibility quality of the images for better feature extraction. Samples of the enhanced image are shown in Figure 4.6.

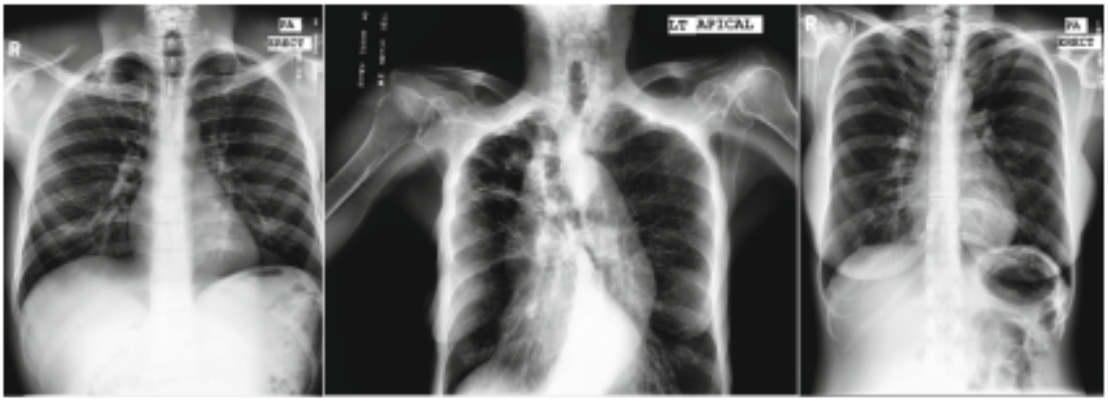


Figure 4.6: Samples of the enhanced CXR [31]

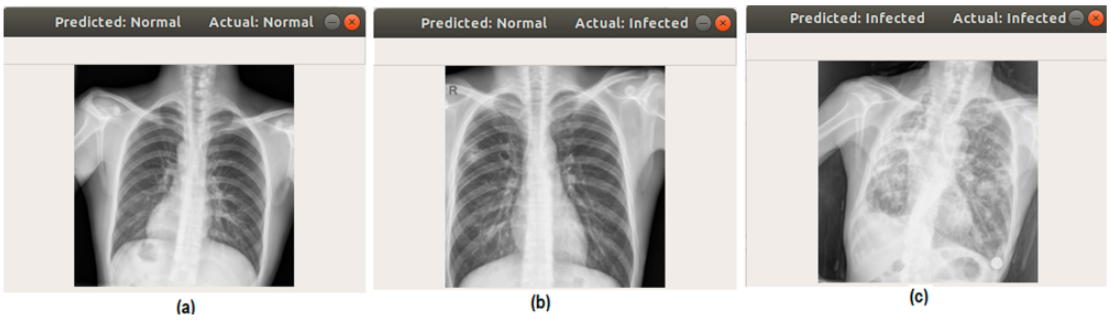


Figure 4.7: Model predictions [31].

Figure 4.7 shows a few predictions of our model. (a) and (c) are samples that were correctly predicted, while (b) is one of the very few misclassified samples that will be improved in future work.

4.2 Classification with Modified Pre-trained CNN

This section presents the experimental analysis and results of the novel modified CNN model for classifying TB. The findings of this model have been published in [30]. We observed from the previous models [29, 31] that training a CNN from the start is complex and time-consuming; hence we adopted the concept of Transfer Learning [54, 56] which involves the transfer of knowledge learned from a particular task to a new related task. This procedure helps achieve a better neural network performance and saves training time.

In this section, we employed deep pre-trained Vgg16 [49] as a feature extractor rather than an end-to-end network as in the case of the original model and used a Logistic Regression (LR) algorithm [50] to obtain the classification. The three fully connected layers in the original Vgg16 were truncated to make the max-pooling

the final output layer for the extracted features as shown in Figure 4.8. The CXR images were pre-processed through the data augmentation and CLAHE technique and resized to conform with the Vgg16 input dimension of 224x224x3. The effect of CLAHE in making the details on CXR more visible and suitable for the feature extraction task is expressed in Figure 4.10. The CXR images were then propagated through the network until the last max-pooling layer with a 7x7x512 shape flattened into a vector map to output the features extracted.

Figure 4.8: Truncated Vgg16.

The extracted features were fed as input to the LR algorithm for training to obtain the classification output as illustrated in Figure 4.9. The LR is a powerful and efficient algorithm for binary classification. Its relation to exponential probability distribution makes them perform well in the medical domain because it allows for the dependence of class labels on features and gives room to carry out sensitivity analysis.

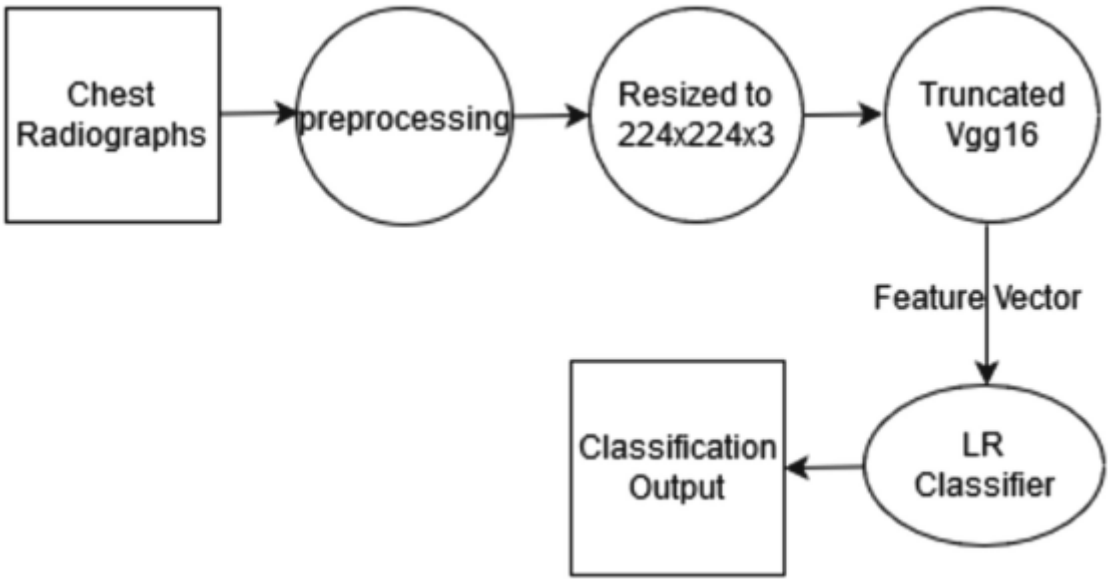


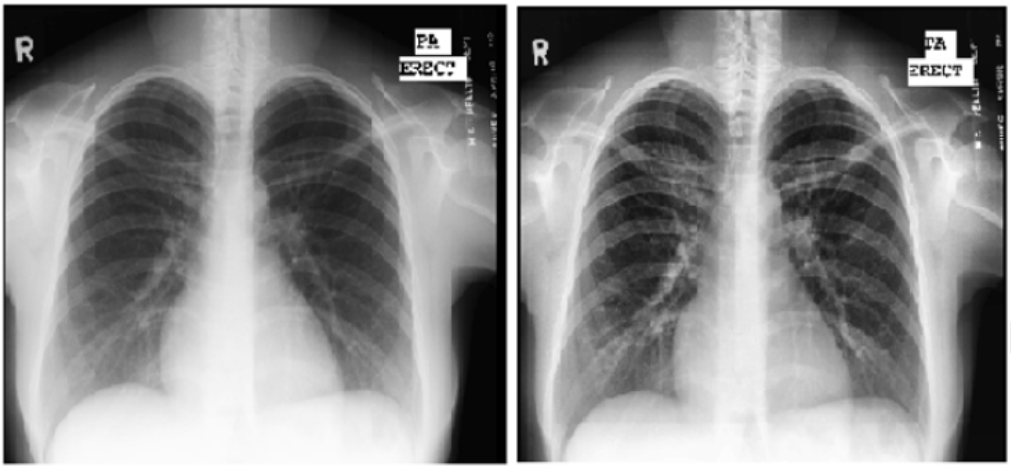
Figure 4.9: Procedural flow of the model.

The model was evaluated on two benchmark datasets (Shenzhen and Montgomery [19]) using standard evaluation metrics: accuracy, precision, and rank-1. The experimental results are presented in Table 4.3.

Table 4.3: Experimental results

Database	Accuracy (%)	Precision (%)	Rank-1 (%)
Montgomery	89	90.1	92.40
Shenzhen	95.8	96	99.25

It can be observed from Table 4.3 that the results obtained in the experiment on the Montgomery dataset is lower than the Shenzhen dataset. This attribute corresponds with the reported work in literature. This outcome is mainly due to fewer datasets and class imbalance. Although in this case, the augmentation technique applied had a positive impact on the dataset, so only the class imbalance where the ratio of healthy to unhealthy samples of the CXR is about "60:40" in the Montgomery set is the factor responsible for it. Table 4.4 presents the comparison of our model with related work.



(a) (b)

Figure 4.10: (a) is the original image, while (b) is the enhanced version of the original.

Table 4.4: Comparison of our model [30] with related work

Ref	Pre-trained Models	Accuracy (%)	Precision (%)
[1]	VggNet	81.25	0.80
[25]	GoogLeNet, VggNet, ResNet	0.847	-
[17]	AlexNet	90.03	-
[18]	AlexNet, VggNet, ResNet	92.00	-
[44]	VggNet, Inception, ResNet, DenseNet	81.60	-
Proposed [30]	**VggNet**	**95.80**	**96.00**

4.3 Classification with Bagging Ensemble Model

This section presents the experimental analysis and results of our novel Ensemble model [32] for the classification of TB from CXR. The Ensemble model is based on multiple pre-trained models comprising image pre-processing, generation of individual CNN classifiers, and combination of the generated models to build an Ensemble classifier for TB classification, as shown in Figure 4.11.

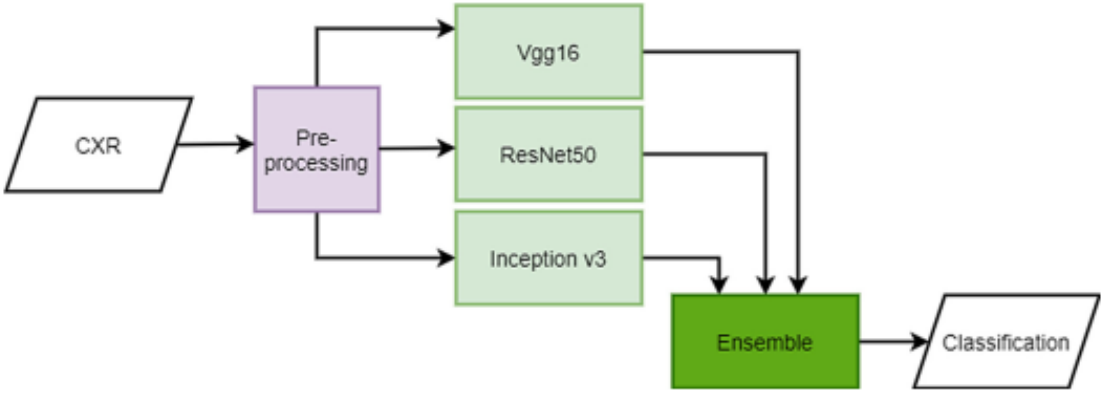


Figure 4.11: Ensemble model

Three pre-trained models (Vgg16, ResNet50, and Inception V3) were considered for building the Ensemble model. Each of these models has a different input dimension; hence the CXR images were pre-processed and resized to satisfy the model's input size. The participating pre-trained models were originally trained on the ImageNet datasets to classify about 1000 objects; it would be complicated to adopt them directly on the CXR image to identify two classes. We therefore constructed and substituted the final layers of the pre-trained models with a binary classifier that was suitable for the target CXR.

We trained the three models independently using the Bagging Ensemble approach such that the output from one model did not affect the other models. The result from the three models was then combined to obtain the final prediction through the Ensemble classifier. The Shenzhen dataset was employed for training the model and evaluated on the Montgomery dataset to determine accuracy, sensitivity, and specificity. The experimental results in Table 4.5 show that Inception V3 achieved the best accuracy among the three models. At the same time, ResNet50 obtained the highest sensitivity and specificity, but the overall highest performance was achieved with the Ensemble model. The graphical representation of the model's accuracy is presented in Figure 4.12

One of the challenges in training a CNN model is overfitting, which is common when a model is run on a new dataset. It is essential to control overfitting and achieve better generalization of the model on new datasets. In this section, we

Table 4.5: Ensemble experimental results of [32]

	Ensemble	ResNet50	Vgg16	Inception v3
Accuracy (%)	96.14	89.51	92.10	93.50
Sensitivity (%)	90.03	89.09	88.25	79.31
Specificity (%)	92.41	90.08	91.00	83.36

used techniques such as data augmentation, normalization, dropout, and the least absolute deviations (LAD) to control overfitting.

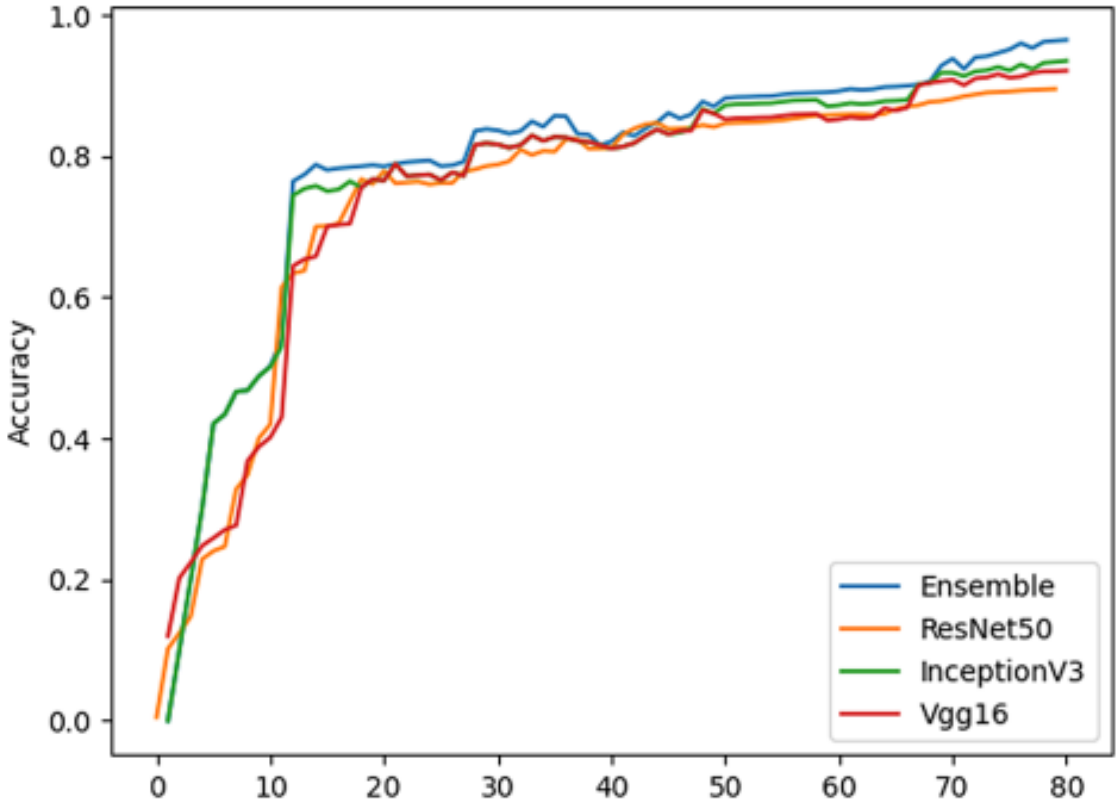


Figure 4.12: Accuracy performance of models

As a result of various manifestations of TB, depending on a single model might not be too effective. Hence Ensemble of multiple models makes the CAD system more robust with an improved detection rate. With Ensembling, different models complement themselves such that when a model misclassifies an image, one or more of the other models can accurately classify the image. The Ensemble model [32] generalized well despite having the training and test set from different datasets and surpassed existing Ensemble models as shown in Table 4.6.

Table 4.6: Ensemble model [32] compared with existing model

Authors ref	Models combined	Accuracy (%)
Hijazi et al. [14]	2	89.77
Hernandez et al. [13]	3	86.40
Ayaz et al. [4]	6	0.99
Oloko-Oba and Viriri [32]	**3**	**96.14**

4.4 Classification with state-of-the-art EfficientNets

This section presents the experimental analysis and results of our novel EfficientNets for detecting TB from CXR. The findings and results of this model have been published in [33]. Precisely, five variants of EfficientNets (B0-B4) were fine-tuned and implemented on the two prominent and publicly available TB datasets (Montgomery and Shenzhen). Each of these variants has a different input size, hence the need to apply pre-processing on the CXR and resize them to comply with the input parameters of each variant of the EfficientNets. Table 4.7 presents the data augmentation parameters.

Table 4.7: Data augmentation types and probability value

Augmentation type	Probability value
Rotation_righ_left	0.5
Zoom range	0.3
Vertical flip	0.4
Horizontal flip	0.4
Height_shift_range	0.2
Width	0.2

EfficientNets are a lightweight model based upon the AutoML framework to develop a network and uniformly scale up the depth, width, and resolutions through a simplified and effective compound coefficient. So, in transferring the pre-trained EfficientNets directly to the CXR images, we fine-tuned the models by adding a global average pooling (GAP) to reduce the number of parameters and handle overfitting. Two dense layers followed the GAP with a ReLU activation function and a dropout rate of 0.4 before a final dense layer that served as output with SoftMax activation function to determine the probabilities of the input CXR representing the healthy or infected classes. The structure of our fine-tuned EfficientNets is shown in Figure 4.13.

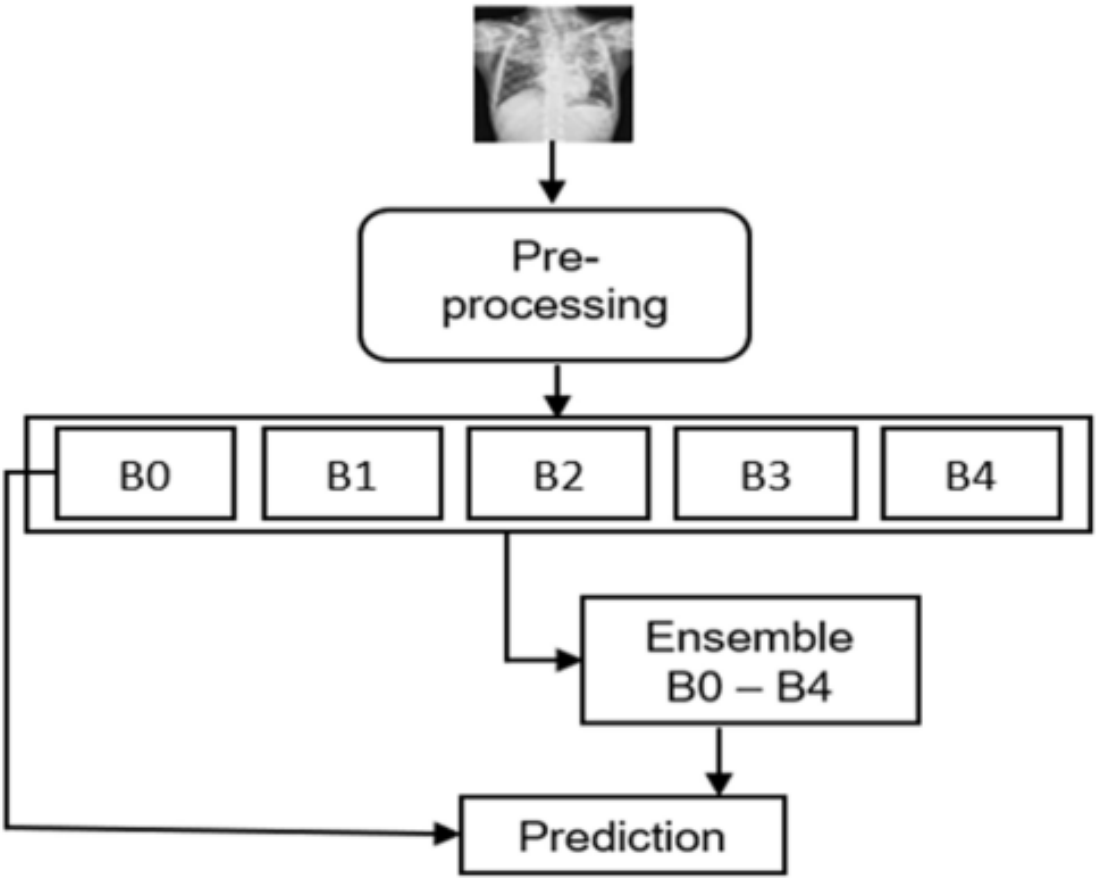


Figure 4.13: Block structure of EfficientNets model

Early stopping was configured and invoked in training the model as we understood that too many iterations could lead to model overfitting while too few iterations could cause model underfitting. About 120 training iterations were configured and we then set up the early stopping to terminate the training once the model stopped improving on the hold out validation set. The early stopping, L2 regularization, data augmentation and dropout were implemented to control overfitting.

The models were trained for 40 iterations (epochs) using a stochastic gradient descends (SGD) optimizer with a 0.0001 learning rate. The batch size for each iteration was set at 22, momentum equals 0.9 and L2 regularizer. At the same time, a categorical cross-entropy loss function was used to update weights at each iteration. Hyper-parameters used were carefully evaluated and found to perform optimally.

The experiments were performed using the Keras Deep Learning framework running on the TensorFlow backend. Each model was trained in the cloud on the Telsa graphics processing unit (GPU) made available through the Google Collaboratory framework. The Collaboratory provides up to 12GB random access memory

(RAM) and about 360GB GPU in the cloud for research purposes. The models were evaluated using the popular evaluation metrics and performed well at extracting and learning discriminative features from the CXR. EfficientNet-B4 achieved the best accuracy for both datasets over the other variants. The experimental results are detailed in Table 4.8 depicting accuracy, sensitivity, specificity, and AUC measured in percentage (%).

Table 4.8: Experimental results of EfficientNets

EfficientNet	Montgomery				Shenzhen			
Model	Acc	Sen	Spe	Auc	Acc	Sen	Spe	Auc
B0	80.52	82.10	81.48	0.77	84.90	87.08	85.36	0.82
B1	83.46	83.78	83.77	0.80	86.62	89.12	90.81	0.86
B2	86.35	90.16	88.25	0.84	91.24	90.83	90.31	0.90
B3	90.40	93.01	92.00	0.90	92.83	93.64	89.53	0.92
B4	92.33	95.41	93.61	0.92	94.35	95.69	94.15	0.93
Ensemble	95.82	98.13	95.78	0.94	97.44	99.18	96.21	0.96

Despite the excellent accuracy obtained from each model, an Ensemble model comprising EfficientNets-B2, B3, and B4 was developed to improve the classification accuracy. The Ensemble results on both datasets surpassed each of the EfficientNets variants employed in this study and literature, as evident in Table 4.9. The validation accuracy and loss is presented in Figure 4.14

Table 4.9: Proposed fine-tuned EfficientNets compared with related study

Authors	Results (%)
Lopes et al. [25]	84.6
Rajaraman and Antani [43]	94.1
Hernandez et al. [13]	86.4
Hijazi et al.[14]	89.7
Dasanyaka and Dissanayake [8]	97.1
Pasa et al. [40]	86.2
Sahlol et al. [45]	94.1
Proposed EfficientNets [33]	**97.4**

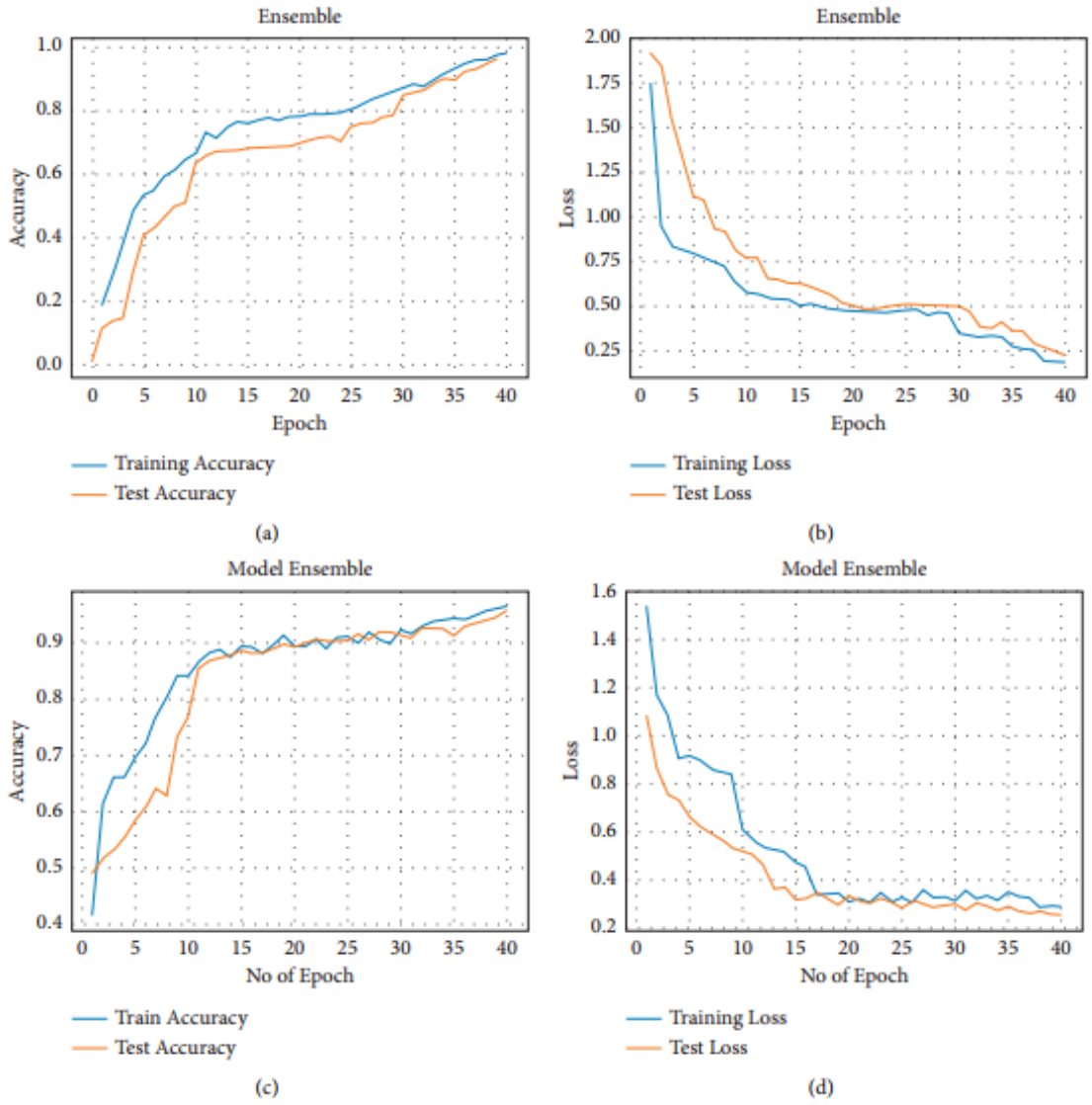


Figure 4.14: Validation accuracy and loss of the Ensemble EfficientNets: **(a)** and **(b)** are the performance with the Shenzhen dataset, while **(c)** and **(d)** represents the performance with the Montgomery dataset

4.5 Summary

In this chapter, the results of the novel end-to-end CNN models [29, 31] to screen TB from CXR images and classify them as healthy or infected was presented. The model achieved promising results but could not attain the optimal expectation due to the depth of the models and the lack of a large number of datasets and hence requires improvement due to the severity of TB disease. These limitations led to the exploration of transfer learning, where a pre-trained CNN model was fine-tuned and employed as a features extractor rather than an end-to-end network for classifi-

cation [30]. This approach improved the results of the previous models and proved the effect of implementing transfer learning to extract discriminating features from CXR. Having established the importance of transfer learning, an end-to-end Ensemble network of three pre-trained models was developed [32] that takes advantage of the strength of different pre-trained CNN models such that a misclassified image from model "A" can be correctly classified by model "B" and "C". The same approach was employed in [33] to implement state-of-the-art EfficientNets because these models are smaller with few parameters, faster, and more generalizable to achieve higher accuracy on a transfer learning task than other pre-trained models. Evidently, from the results achieved, the study found that combining several classifiers can increase the detection rate and improve classification accuracy compared to any stand-alone classifier.

Chapter 5

Conclusion and Future Work

This chapter summarises the work presented in previous chapters and discusses some future directions that will further enhance the quality of diagnostics to reach acceptable clinical implementations.

5.1 Conclusion

Early and accurate Tuberculosis (TB) diagnosis has become a significant challenge in TB burden regions. This is due to poverty, inadequate medical resources, and a lack of skilled medical practitioners. TB patients lose their lives yearly due to diagnostic delay, misdiagnosis, and lack of appropriate treatments. TB is curable, but it necessitates early diagnosis through reliable screening techniques. CXR is a recommended screening procedure for detecting pulmonary abnormalities. Although this recommendation requires the expertise of experienced radiologists to interpret the screening results, which forms part of the problem in rural communities. These challenges have resulted in efforts made in the last decades to develop Computer-Aided Detection (CAD) systems for the automatic diagnosis of TB. However, their performance has been characterized by low accuracy and sensitivity, resulting in misdiagnosis. Limited dataset, quality of CXR images, and model parameters have also hampered existing CAD performance. Due to the severity of TB and the World Health Organization's End TB strategy, the existing models need to be improved.

This thesis presented Deep Learning (DL) techniques to address the aforementioned challenges. The thesis began with an all-inclusive introduction of the life-threatening TB disease, stating typical indications of infections, rate of mortality and highlights the study objectives to be accomplished in Section 1.3. Chapter 2 provided a comprehensive review and analysis of the several Machine Learning (ML) and DL models that have been explored in developing CAD systems for automatic diagnosis of TB. The review's outcome led to the development of the following classifiers which are presented in chapter 3 of this thesis:

- An end-to-end CNN-based model comprising four convolutional layers for extracting TB-related features from the Montgomery CXR and a fully connected layer consisting of biases and weights connects neurons required by the softmax layer that determines features extraction probability for the classification results. The model achieved 2.1% accuracy over existing methods as shown in Table 4.2. The model presented in Section 3.2 refined the previous model's design to improve classification accuracy. The modifications included hyper-parameters tweaking, use of Shenzhen dataset, deeper model design, and enhanced pre-processing techniques. These models appeared in [29] and [31] and a limited number of training datasets hampered the performance of both models.

- A model that employed a learning concept to transfer the knowledge acquired from training a model as a starting point for the new model. A deep CNN model pre-trained on the ImageNet dataset was adopted as a features extractor on the CXR rather than an end-to-end network. The extracted features were then used to train a classifier to obtain the final classifications of the images. The model achieved 95.80% classification accuracy and precision of 96%. The transfer learning explored here helped overcome the challenges of the limited dataset encountered in Sections 3.1 and 3.2 and proved effective in maximizing accuracy. The approach was presented in [30].

- An Ensemble of three pre-trained CNN models was proposed to detect and classify TB from CXR. Each model was fine-tuned and trained separately on the Shenzhen dataset and evaluated with the Montgomery set. The results of these models was then averaged through bagging to build an Ensemble classifier. The Ensemble model takes advantage of the strengths of different pre-trained CNN models such that a misclassification from one model can be detected and corrected by other models. The results achieved with this model appeared in [32] and outperformed related existing Ensemble models as shown in Table 4.6.

- A lightweight EfficientNets model that achieved a state-of-the-art results on ImageNet was fine-tuned on the CXR images and implemented for the classification of TB. The model scaled down the number of parameters using a global average pooling to control overfitting and incorporated robust image pre-processing techniques that enhanced the visibility of the images and generated replica samples of the original datasets to increase and improve the performance of models. The results obtained with the fine-tuned EfficientNets B0 – B4 in this thesis and presented in [33] surpasses other models in Table 4.9. These models are smaller with fewer parameters, faster, and generalize well on popular transfer learning tasks to achieve state-of-the-art results. In general, the Ensembled of EfficientNets B2, B3 and B4 in this study achieved new state-of-the-art results of 97.44% accuracy, 99.18% sensitivity, 96.21% specificity and AUC of 0.96%. Finally, a detailed analysis of the results for

the CNN-based TB classification models compared with existing models was presented in chapter 4.

In general,this thesis presented robust image pre-processing techniques to handle the challenges of limited datasets, especially in the medical field and enhanced the quality of images used for training. It then developed end-to-end CNN models to deal with TB diagnostic delays and misdiagnosis. The study also fine-tuned and employed state-of-the-art EfficientNets models and achieved high accuracy and sensitivity compared to existing CAD systems. The performance of the models was then improved with the bagging Ensemble technique.

5.2 Future Work

The models developed in this thesis provide motivations for future research and suggest areas that can improve the development of future CAD systems for screening pulmonary abnormalities. The recommended areas include:

1. To develop models that could distinctly identify foreign objects like rings, buttons, pieces of bone, and coins that may be found on CXR images. Such foreign objects could lead to misclassification and affect the performance of detection systems. For instance, round objects like buttons, round bones or rings can be mistaken for nodules.

2. A robust CAD system to simultaneously detect different pulmonary diseases. Such a model could detect TB and lung cancer simultaneously on a CXR.

3. Some diseases usually exhibit similar manifestations on the CXR, leading to misdiagnosis. Several studies are encouraged to differentiate these diseases regardless of their common manifestations.

.

Bibliography

[1] Mostofa Ahsan, Rahul Gomes, and Anne Denton. Application of a convolutional neural network using transfer learning for tuberculosis detection. In *2019 IEEE International Conference on Electro Information Technology (EIT)*, pages 427–433. IEEE, 2019. 112

[2] Saad Albawi, Tareq Abed Mohammed, and Saad Al-Zawi. Understanding of a convolutional neural network. In *2017 International Conference on Engineering and Technology (ICET)*, pages 1–6. Ieee, 2017. 2

[3] Maria Goretti Ametembun. Iddf2019-abs-0238 follow-up of small bowel and dry type peritoneal tuberculosis treatment, 2019. 1

[4] Muhammad Ayaz, Furqan Shaukat, and Gulistan Raja. Ensemble learning based automatic detection of tuberculosis in chest x-ray images using hybrid feature descriptors. *Physical and Engineering Sciences in Medicine*, 44(1):183–194, 2021. 115

[5] Marcus D Bloice, Christof Stocker, and Andreas Holzinger. Augmentor: an image augmentation library for machine learning. *arXiv preprint arXiv:1708.04680*, 2017. 106

[6] Alana T Brennan, Mhairi Maskew, Bruce A Larson, Isaac Tsikhutsu, Margaret Bii, Lungisile Vezi, Matthew P Fox, Willem DF Venter, Peter Ehrenkranz, and Sydney Rosen. Who is seeking antiretroviral treatment for hiv now? characteristics of patients presenting in kenya and south africa in 2017-2018. *Journal of the International AIDS Society*, 22(9):e25358, 2019. 1

[7] David L Cohn, Richard J ОвГЄBrien, Lawrence J Geiter, F Gordin, E Hershfield, C Horsburgh, et al. Targeted tuberculin testing and treatment of latent tuberculosis infection. *MMWR Morb Mortal Wkly Rep*, 49(6):1–54, 2000. 3

[8] Chirath Dasanayaka and Maheshi Buddhinee Dissanayake. Deep learning methods for screening pulmonary tuberculosis using chest x-rays. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 9(1):39–49, 2021. 117

[9] Jia Deng, Wei Dong, Richard 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*, pages 248–255. Ieee, 2009. 66, 91

[10] Prabha Desikan. Sputum smear microscopy in tuberculosis: is it still relevant? *The Indian journal of medical research*, 137(3):442, 2013. 3

[11] Jana Fehr, Stefan Konigorski, Stephen Olivier, Resign Gunda, Ashmika Surujdeen, Dickman Gareta, Theresa Smit, Kathy Baisley, Sashen Moodley, Yumna Moosa, et al. Computer-aided interpretation of chest radiography reveals the spectrum of tuberculosis in rural south africa. *NPJ digital medicine*, 4(1):1–10, 2021. 1

[12] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 2, 79

[13] Alfonso Hernández, Ángel Panizo, and David Camacho. An ensemble algorithm based on deep learning for tuberculosis classification. In *International conference on intelligent data engineering and automated learning*, pages 145–154. Springer, 2019. 115, 117

[14] Mohd Hanafi Ahmad Hijazi, Stefanus Kieu Tao Hwa, Abdullah Bade, Razali Yaakob, and Mohammad Saffree Jeffree. Ensemble deep learning for tuberculosis detection using chest x-ray and canny edge detected images. *IAES International Journal of Artificial Intelligence*, 8(4):429, 2019. 115, 117

[15] Rahul Hooda, Sanjeev Sofat, Simranpreet Kaur, Ajay Mittal, and Fabrice Meriaudeau. Deep-learning: A potential method for tuberculosis detection using chest radiography. In *2017 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*, pages 497–502. IEEE, 2017. 108, 109

[16] Mohammad Hossin and MN Sulaiman. A review on evaluation metrics for data classification evaluations. *International Journal of Data Mining & Knowledge Management Process*, 5(2):1, 2015.

[17] Sangheum Hwang, Hyo-Eun Kim, Jihoon Jeong, and Hee-Jin Kim. A novel approach for tuberculosis screening based on deep convolutional neural networks. In *Medical imaging 2016: computer-aided diagnosis*, volume 9785, page 97852W. International Society for Optics and Photonics, 2016. 109, 112

[18] Mohammad Tariqul Islam, Md Abdul Aowal, Ahmed Tahseen Minhaz, and Khalid Ashraf. Abnormality detection and localization in chest x-rays using deep convolutional neural networks. *arXiv preprint arXiv:1705.09850*, 2017. 112

[19] Stefan Jaeger, Sema Candemir, Sameer Antani, Yì-Xiáng J Wáng, Pu-Xuan Lu, and George Thoma. Two public chest x-ray datasets for computer-aided screening of pulmonary diseases. *Quantitative imaging in medicine and surgery*, 4(6):475, 2014. 111

[20] Phil Kim. Convolutional neural network. In *MATLAB deep learning*, pages 121–147. Springer, 2017. 2

[21] Nancy A Knechel. Tuberculosis: pathophysiology, clinical features, and diagnosis. *Critical care nurse*, 29(2):34–43, 2009. 1

[22] Burçin Kurt, Vasif V Nabiyev, and Kemal Turhan. Medical images enhancement by using anisotropic filter and clahe. In *2012 International symposium on innovations in intelligent systems and applications*, pages 1–4. IEEE, 2012. 109

[23] Michael Levin, Myrsini Kafarou, and Lachlan Coin. Method of detecting active tuberculosis using minimal gene signature, October 24 2019. US Patent App. 16/326,828. 1

[24] Chang Liu, Yu Cao, Marlon Alcantara, Benyuan Liu, Maria Brunette, Jesus Peinado, and Walter Curioso. Tx-cnn: Detecting tuberculosis in chest x-ray images using convolutional neural network. In *2017 IEEE international conference on image processing (ICIP)*, pages 2314–2318. IEEE, 2017. 108, 109

[25] UK Lopes and João Francisco Valiati. Pre-trained convolutional neural networks as feature extractors for tuberculosis detection. *Computers in biology and medicine*, 89:135–143, 2017. 109, 112, 117

[26] Arun C Nachiappan, Kasra Rahbar, Xiao Shi, Elizabeth S Guy, Eduardo J Mortani Barbosa Jr, Girish S Shroff, Daniel Ocazionez, Alan E Schlesinger, Sharyn I Katz, and Mark M Hammer. Pulmonary tuberculosis: role of radiology in diagnosis and management. *Radiographics*, 37(1):52–72, 2017. 2

[27] Nani Nair, Fraser Wares, and Suvanand Sahu. Tuberculosis in the who south-east asia region, 2010. 1

[28] Norliza Mohd Noor, Omar Mohd Rijal, Ashari Yunus, Aziah A Mahayiddin, Gan Chew Peng, and SAR Abu-Bakar. A statistical interpretation of the chest radiograph for the detection of pulmonary tuberculosis. In *2010 IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES)*, pages 47–51. IEEE, 2010. 2

[29] Mustapha Oloko-Oba and Serestina Viriri. Diagnosing tuberculosis using deep convolutional neural network. In *International Conference on Image and Signal Processing*, pages 151–161. Springer, 2020. x, 105, 106, 107, 108, 109, 110, 118, 121

[30] Mustapha Oloko-Oba and Serestina Viriri. Pre-trained convolutional neural network for the diagnosis of tuberculosis. In *International Symposium on Visual Computing*, pages 558–569. Springer, 2020. ix, 110, 112, 119, 121

[31] Mustapha Oloko-Oba and Serestina Viriri. Tuberculosis abnormality detection in chest x-rays: A deep learning approach. In *International Conference on Computer Vision and Graphics*, pages 121–132. Springer, 2020. x, 105, 108, 109, 110, 118, 121

[32] Mustapha Oloko-Oba and Serestina Viriri. Ensemble of convolution neural networks for automatic tuberculosis classification. In *International Conference on Computational Collective Intelligence*, pages 549–559. Springer, 2021. ix, 113, 114, 115, 119, 121

[33] Mustapha Oloko-Oba and Serestina Viriri. Ensemble of efficientnets for the diagnosis of tuberculosis. *Computational Intelligence and Neuroscience,*, 2021, 2021. 115, 117, 119, 121

[34] World Health Organization. *Chest radiography in tuberculosis detection: summary of current WHO recommendations and guidance on programmatic approaches*. World Health Organization, 2016. 2

[35] World Health Organization. *Global tuberculosis report 2017*. World Health Organization, 2017. 1

[36] World Health Organization. *Global status report on alcohol and health 2018*. World Health Organization, 2019. 1

[37] World Health Organization. *Global tuberculosis report 2020: executive summary*. World Health Organization, 2020. 2

[38] World Health Organization et al. Geneva: Global tuberculosis report, 2019, 2019. 3

[39] Tripti Pande, Madhukar Pai, Faiz Ahmad Khan, and Claudia M Denkinger. Use of chest radiography in the 22 highest tuberculosis burden countries. *European Respiratory Journal*, 46(6):1816–1819, 2015. 2

[40] F Pasa, V Golkov, F Pfeiffer, D Cremers, and D Pfeiffer. Efficient deep network architectures for fast chest x-ray tuberculosis screening and visualization. *Scientific reports*, 9(1):1–9, 2019. 117

[41] Debora Pedrazzoli, Marek Lalli, Delia Boccia, Rein Houben, and Katharina Kranzer. Can tuberculosis patients in resource-constrained settings afford chest radiography? *European Respiratory Journal*, 49(3), 2017. 2

[42] KJ Potts, MF Ragland, A Hans, and MT Kearns. Empyema necessitans caused by mycobacterium tuberculosis. In *D58. PLEURAL INFECTION CLINICAL CASE AND STUDIES*, pages A6842–A6842. American Thoracic Society, 2019. 1

[43] Sivaramakrishnan Rajaraman and Sameer K Antani. Modality-specific deep learning model ensembles toward improving tb detection in chest radiographs. *IEEE Access*, 8:27318–27326, 2020. 117

[44] Anuj Rohilla, Rahul Hooda, and Ajay Mittal. Tb detection in chest radiograph using deep learning architecture. In *Proceeding of 5th International Conference on Emerging Trends in Engineering, Technology, Science and Management (ICETETSM-17)*, pages 136–147, 2017. 109, 112

[45] Ahmed T Sahlol, Mohamed Abd Elaziz, Amani Tariq Jamal, Robertas Damaševičius, and Osama Farouk Hassan. A novel method for detection of tuberculosis in chest radiographs using artificial ecosystem-based optimisation of deep neural network features. *Symmetry*, 12(7):1146, 2020. 117

[46] Seelwan Sathitratanacheewin and Krit Pongpirul. Deep learning for automated classification of tuberculosis-related chest x-ray: dataset specificity limits diagnostic performance generalizability. *arXiv preprint arXiv:1811.07985*, 2018. 1

[47] Jürgen Schmidhuber. Deep learning in neural networks: An overview. *Neural networks*, 61:85–117, 2015. 2

[48] Connor Shorten and Taghi M Khoshgoftaar. A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1):1–48, 2019. 105

[49] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. 66, 79, 110

[50] Sandro Sperandei. Understanding logistic regression analysis. *Biochemia medica*, 24(1):12–18, 2014. 110

[51] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015. 2

[52] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826, 2016. 79

[53] Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International Conference on Machine Learning*, pages 6105–6114. PMLR, 2019. 91

[54] Lisa Torrey and Jude Shavlik. Transfer learning. In *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques*, pages 242–264. IGI global, 2010. 110

[55] Anna H vant Hoog, Ikushi Onozaki, and Knut Lonnroth. Choosing algorithms for tb screening: a modelling study to compare yield, predictive value and diagnostic burden. *BMC infectious diseases*, 14(1):1–12, 2014. 3

[56] Wei Ying, Yu Zhang, Junzhou Huang, and Qiang Yang. Transfer learning via learning to transfer. In *International conference on machine learning*, pages 5085–5094. PMLR, 2018. 110

[57] Alimuddin Zumla, Andrew George, Virendra Sharma, Rt Hon Nick Herbert, Aaron Oxley, and Matt Oliver. The who 2014 global tuberculosis reportвГЎ-further to go. *The Lancet Global Health*, 3(1):e10–e12, 2015. 1

[58] Alice Zwerling, Susan van den Hof, Jerod Scholten, Frank Cobelens, Dick Menzies, and Madhukar Pai. Interferon-gamma release assays for tuberculosis screening of healthcare workers: a systematic review. *Thorax*, 67(1):62–70, 2012. 3