IBRAHIM OMAYAYI OMODAMILOLA

INTEGRATION ARTIFICIAL INTELLIGENCE POWERED DETECTION OF TUBERCULOSIS

DOCTORATE THESIS

2025

NEAR EAST UNIVERSITY

INSTITUTE OF GRADUATE STUDIES

DEPARTMENT OF BIOMEDICAL ENGINEERING

ARTIFICIAL INTELLIGENCE POWERED DETECTION OF TUBERCULOSIS

Ph.D. THESIS

IBRAHIM OMOYAYI OMODAMILOLA

NICOSIA

JANUARY, 2025

NEAR EAST UNIVERSITY
INSTITUTE OF GRADUATE STUDIES
DEPARTMENT OF BIOMEDICAL ENGINEERING

ARTIFICIAL INTELLIGENCE POWERED DETECTION OF TUBERCULOSIS

Ph.D. THESIS

IBRAHIM OMOYAYI OMODAMILOLA

Supervisor
Assoc. Prof. Dr. ABDULLAHI UMAR IBRAHIM
Assoc. Prof. Dr. SÜLYMAN AŞIR

Nicosia
JANUARY, 2025

Approval

We certify that we have read the thesis submitted by Ibrahim Omoyayi Omodamilola titled **"Artificial Intelligence-Powered Detection of Tuberculosis"** and that in our combined opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Doctor of Biomedical Engineering.

Examining Committee	Name-Surname	Signature

Head of the Committee: PROF.DR. MEHMET OZSOZ

Committee Member*: ASST.PROF. DR PWADUBASHIYI COSTON

Committee Member*: ASST.PROF. DR EMMANUEL HALILU

Committee Member*: ASST.PROF. DR AUWAL MUBRAK

Supervisor: ASSOC.PROF. DR ABDULLAHI UMAR IBRAHIM

Approved by the Head of the Department

20/01/2025

....

Assoc. Prof. Dr. Sülyman Aşir

Head of Department

Approved by the Institute of Graduate Studies

....../....../20...

Prof. Dr. Kemal Hüsnü Can Başer

Prof. Dr. Kemal Hüsnü Can Başer

Declaration

I hereby declare that all information, documents, analysis and results in this thesis have been collected and presented according to the academic rules and ethical guidelines of Institute of Graduate Studies, Near East University. I also declare that as required by these rules and conduct, I have fully cited and referenced information and data that are not original to this study.

Ibrahim Omoyayi Saleh

ACKNOWLEGMENTS

I would like to express my profound gratitude to my supervisor Assoc. Professor Abdullahi Umar and Assoc. Professor Sülyman Aşir for their thorough dedication to the success of this thesis.

I would like to express my special thanks to my wife Amirah Omoyayi and our beautiful children Zayden and Zeyada Omoyayi for their motivation. To my mother, father and sisters, I want to say a big thank you for all your sacrifices.

.

ABSTRACT

ARTIFICIAL INTELLIGENCE POWERED DETECTION OF TUBERCULOSIS

IBRAHIM OMOYAYI OMODAMILOLA

Ph.D., Department of Biomedical Engineering

November 2024, 120 pages

The exponential increase in the incidence and mortality rate of patients suffering from tuberculosis coupled with the incidence of false positive result, inaccuracy, time consuming, high cost of diagnostic or screening technique for detection of tuberculosis and the prevalence of multidrug resistant type especially in low and middle income countries continue to exert burden on global healthcare system. To address these resolute challenges, scientists adopted Artificial Intelligence (AI) based techniques. AI-based models have been deployed for the screening of several diseases which include cancer, pneumonia and tuberculosis. Moreover, majority of current studies focused on the implementation of DL-based models for detection of tuberculosis using single type of medical images. Another challenges of existing studies include the use of DL-models designed using SoftMax as the output classifier. Thus, in this study, we proposed the application of several pre-trained DL CNN models which include MobileNet, 2 VGGNet variants and 2 ResNet variants, coupled with 2 classical ML classifiers. In order to test the generalizability of the pre-trained models, 2 distinct dataset curated from publicly accessible domains are employed which include 7000 Chest X-ray (CXR) images and 1893 Microscopic Slide (MS) images. Evaluation of the performances of the models and comparison has shown that ResNet101-KNN attained the best result with 0.9970 sensitivity, 0.9970 specificity, 0.9970 accuracy, 0.9971 precision, 0.9971 F1 score, 0.9970 Younden index and 0.9970 AUC using CXR images and 1.00 sensitivity, 0.9960 specificity, 0.9980 accuracy, 0.9959 precision, 0.9979 F1 score, 0.9960 Youden index and 0.9980% AUC using MS images. The optimum performance achieved can be attributed to ensemble learning or hybrid models where DL models are coupled with ML classifiers.

Keywords: Pneumonia; Tuberculosis; Artificial Intelligence; Deep Learning; X-ray Images; Microscopic slide Images;

TABLE OF CONTENTS

CHAPTER I

CHAPTER II

CHAPTER III

CHAPTER IV

CHAPTER V

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Abbreviations	Full meaning
AFB:	Acid Fast Bacillus
AI:	Artificial Intelligence
AIDS:	Acquired Immune Deficiency syndrome
ANN:	Artificial Neural Networks
AUC:	Area Under the Curve
CADe:	Computer Aided Detection
CADx:	Computed Aided Diagnosis
CDC:	Center for Disease Control and Prevention
CNNs:	Convolutional Neural Networks
COVID-19:	Coronavirus Diseases 2019
CSF:	Cerebral Spinal Fluid
CT:	Computed Tomography
CXR:	Chest X-ray
DL:	Deep Learning
DQN:	Deep Q-network
DT:	Decision Tree
EHR:	Electronic Health Records
EPTB:	Extra-Pulmonary TB
FAS:	Fatty Acid Synthase
FN:	False Negative
FNLM:	Fast Nonlocal Means
FP:	False positive
FPR:	False Positive Rate
GPU:	Graphical Processing Unit
HIV:	Human Immunodeficiency Virus
IGRAs:	Interferon-Gamma Release Assays
ILSVRC:	ImageNet Large Scale Visual Recognition Challenge
IoHT:	Internet of Healthcare Things

IoT:	Internet of Things
IoMT:	Internet of Medical things
KNN:	K-Nearest Neighbor
LDA:	Linear Discriminant Analysis
LR:	Logistic Regression
MBTB:	Mycobacterium tuberculosis
ML:	Machine Learning
MLNNs:	Multi-layer Neural Networks
MRI:	Magnetic Resonance Imaging
NB:	Naïve Bayes
NLM:	National Library of Medicine
NLP:	Natural Language Programming
NPV:	Negative Predictive Value
PCA:	Principal Component Analysis
PCR:	Polymerase Chain Reaction
PET:	Position Emission Tomography
PPD:	Purified Protein Derivative
PPV:	Positive Predictive Value
PTB:	Pulmonary Tuberculosis
QDA:	Quadratic Discriminant Analysis
QFT:	QuantiFERON
RAM:	Random Access Memory
RELU:	Rectified Linear Unit
RF:	Random Forest
RNNs:	Recurrent Neural Network
ROC:	Receiver Operating Characteristic
ROIs:	Regions of Interests
ResNets:	Residual Networks
SARSA:	state–action–reward–state–action
SE:	squeeze-and-excitation
SGD:	Stochastic Gradient Descent
SPECT:	Single Emission Positron Tomography

SVD:	Singular Value Decomposition
SVM:	Support Vector Machine
TB:	Tuberculosis
TL:	Transfer Learning
TN:	True negative
TP:	True positive
TPR:	True Positive Rate
TT:	Tuberculin Test
VGGNET:	Visual Geometry Group network
WHO:	World Health Organization
YI:	Youden Index
ZN:	Zihel-Nelson
ZNSM-iDB:	Ziehl-Neelsen Sputum Smear Microscopy image Database

CHAPTER I
Introduction

1.0 Overview

This chapter start with a background on tuberculosis, transmission, epidemiology, pathology, diagnosis, and treatment. The chapter also covers the statement of the problem which is tailored toward medical diagnostic errors and inaccuracies. The statement of the problem is aimed towards connecting the role of artificial intelligence as a technique to address the outline limitations as the basis of the significant of the study. Next section pinpointed the study's aims and objectives followed by limitations and lastly structure of the thesis.

1.1 Tuberculosis Overview

Tuberculosis is one of the most fatal lung infections caused by a bacteria know as Mycobacterium tuberculosis (MBTB). The bacteria is slender with size length ranging between 1-10mm. MBTB is an airborne disease which can be transferred from infected person to healthy person. The bacteria is strictly aerobic which means it require oxygen to survive (Dye, 2006). In terms of structure, the bacteria possess a waxy cell wall which makes it acid fast due to the formation of mycolic acid. This feature signifies that MBTB can retain on to stain or dye despite been exposed to alcohol, the resulting staining using Zeihl-Nelson stain makes the bacteria to appear red in color. Notwithstanding, the nature of MBTB waxy wall contribute to their ability to deterred weak disinfectants and tend to survive longer on dry surfaces (Tsai et al., 2013).

Tuberculosis is characterized as one of the most fatal infections with the second number of cases after COVID-19. The disease is ranked 13th leading cause of death worldwide. According to the WHO, every year more than 10 people become infected and more than 1.5 million people died as a result of the disease in 2021. The disease is characterized as the leading cause of death for people suffering with HIV. Moreover, the WHO also estimated that treatment of TB has saved approximately 74 million people between 2000 to 2021. Tuberculosis is predominant in many low and middle income countries which include India, Bangladesh, Pakistan, Nigeria, China, South Africa, Indonesia, Philippines. According to the CDC, there are over 13 million

Americans with latent TB (CDC, 20024). Some people can still harbor the disease throughout their lifetime without becoming active. However, latent TB can be reactivated due to the weakening of the immune system

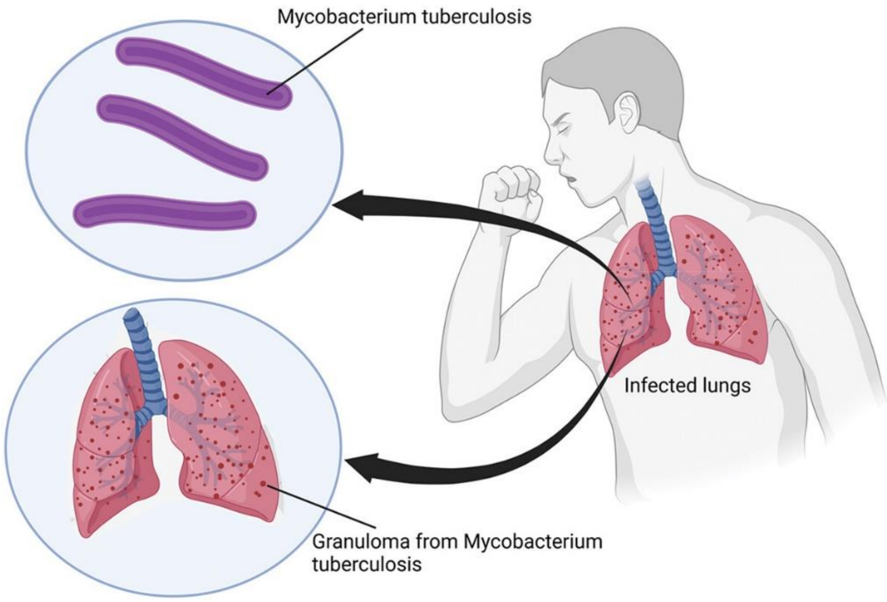
1.1.1 Transmission and Pathology

In terms of transmission, the bacteria can be transmitted from infected patient through coughing, sneezing, talking, singing where droplet that contain the bacteria can be transmitted through the air. People at risk of TB are those with other immune disorders or weakened immune system which include HIV/AIDS. The disease can easily spread where people are residing in crowded places, jails and prisons, correction centers, hospitals, traveling to regions where TB is common (Druszczynska et al., 2012). Tuberculosis can be categorized into 3 which include primary, latent and active. Primary tuberculosis is characterized by the continuous multiplication of the bacteria despite the presence of immune cells. The immune cells identify and destroy the bacteria. However, some can overpower the immune cells and continue to multiply resulting in primary tuberculosis. Many people suffering from primary TB does not show any symptoms (asymptomatic) (Long and Schwartzman, 2014; Basaraba and Hunter, 2017; Mathema et al., 2017).

Latent TB is the next stage after primary TB. This stage is characterized by the development of wall around the lung tissue as a result of the action of the immune cells. Like primary TB, latent TB is also asymptomatic. Active TB on the other hand, is characterized by the overpowering of the immune cells by the bacteria. The bacteria continue to multiply and spread to the lungs and other part of the body (kidneys, lymph node, liver, heart muscles, skin etc.) which is termed as extrapulmonary as extrapulmonary as shown in **Figure 1** (Long and Schwartzman, 2014; Basaraba and Hunter, 2017; Mathema et al., 2017).

Figure 1

Tuberculosis infection



1.1.2 Symptoms

The symptom of tuberculosis varies depending on the stage of infection (primary, latent, active and military). The common symptoms of TB comprises of cough (which can extend longer than 14 days) coughing blood, chest pain, fever, night sweats, tiredness or feeling exhausted, chills, elevated temperature, loss of appetite, weight loss, vomiting etc. (Panic et al., 2020). **Figure 2** present the symptoms of TB.

Figure 2

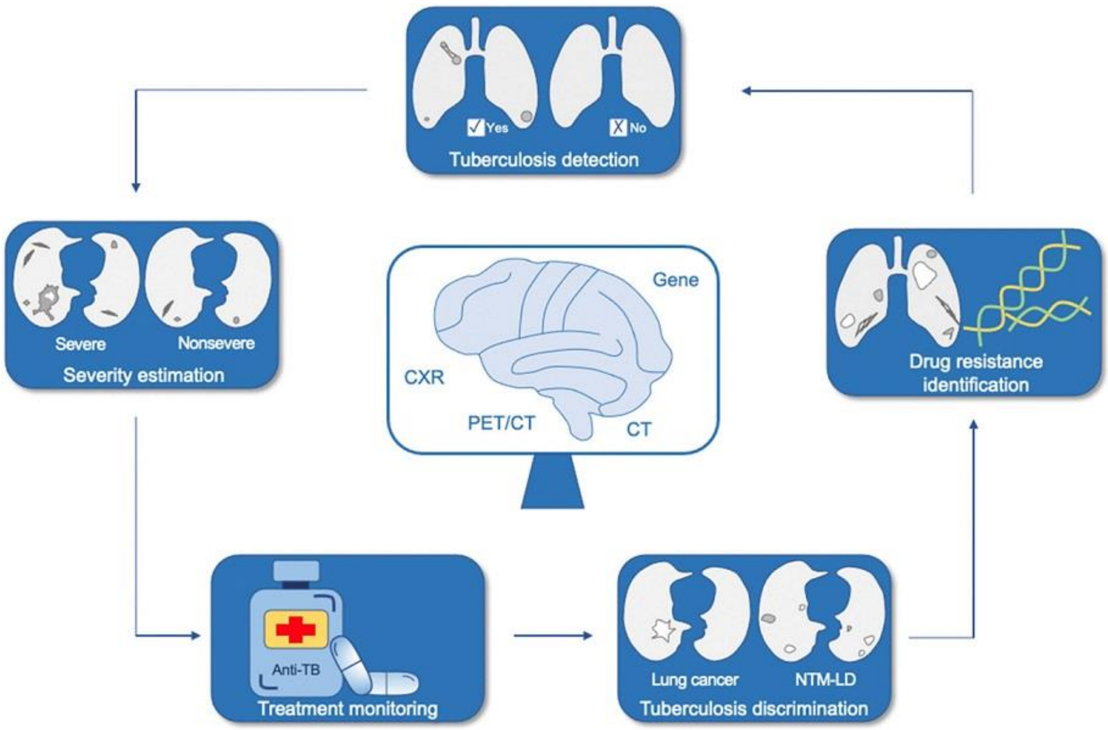
Symptoms of Tuberculosis (DC Health, 2024)

1.1.3 Testing

Accurate and reliable diagnosis or screening of tuberculosis is crucial for appropriate treatment. Medical expert employs several screening testing ranging from laboratory test to medical imaging. Tests include tuberculin test, acid-fast bacillus staining assay, interferon-gamma release assays (IGRAs), GeneXpert, X-ray, and CT scan (Brodie and Schluger, 2005; Campbell and Bah-Sow, 2006; Suárez et al., 2017). Accurate management of TB revolves around proper diagnosis through nucleic acid-based detection (GeneXpert) and medical imaging such as X-ray and PET. However, if patients continue to exhibit symptoms, there is need for screening in order to detect multidrug resistant type in order to prescribe the right medications as demonstrated in **Figure 3.**

Figure 3

Management of TB through accurate diagnosis and treatment



1.1.3.1 Mantoux tuberculin skin test (TST)

Mantoux tuberculin skin test (TST) is a primary screening test conducted for the detection of latent TB. TST is used by medical expert in order to determine if a suspected patient has been exposed to TB or not. The procedure involves the intradermal injection of small quantity of purified protein derivative (PPD) tuberculin under the skin especially the arm. The reaction is check after 2-3 days. Induration is the terminology use to explain the reading of the TST which means detecting a raised, thickened local area around the place of injection. Physicians measures the diameter of the induration in millimeters. The result can be classify as positive when the induration is greater than 15mm in heathy person with normal immune system. However, the result can also be deem positive if the induration is around 10mm or above in people that migrated from high-prevalence regions and employees or residents working or living in high-risk areas, people who work in clinical laboratory handling the bacteria and children less than 4 years. Moreover, the induration of 5mm can also be considered positive in people with low immune cells or weak immune system such as HIV patients, people in contact with active TB patients and patients that received transplants. Furthermore, negative result doesn't often indicate suspect is negative due to false positive results as a result of comprise of the immune system by chemotherapy, steroid therapy or AIDs (Tsiouri et al., 2009; Abdel-Samea et al., 2013).

Another limitations of the TST is that approximately 25% of patients diagnosed with TB exhibit negative result due to weak immune system, poor diet. Moreover, more than 50% of patients suffering with miliary TB (which spread through the body) also exhibit negative results. However, testing positive using the Mantoux test does not necessarily indicate that the patient is suffering from active tuberculosis but rather than the patient has been infected with TB at one point. Despite several drawbacks of the test, it has several advantages which include inexpensive and readily available (Geldenhuys et al., 2010; Abdel-Samea et al., 2013).

1.1.3.2 Interferon gamma release assay (IGRA)

IGRA also known as QuantiFERON test abbreviated as QFT is one of the most popular blood test conducted by medical expert to screen patient suspected with mycobacterium species (*M. tuberculosis, M. bovis, M. africanum*). The IGRA screening test function by measuring the patient's immune response against the bacteria. The IGRA test measure the levels of interferon gamma (IFN-g) released by white blood cells due to presence of MBTB. The test result is usually available after 5-7 days. The test result can be positive or negative. However positive test only implies that the patient at one point in time was infected with TB. The IGRA test is more specific compared to TST test as it gives a quantitative value compare with TST that produce visual response. The test is recommended to be conducted along with either PCR or acid-fast bacillus test. Another limitation of the test includes the probability of false positive results. Advantages of the test include less expensive or affordable, simple and provide more accurate result compare with TST. (Clifford et al., 2015; Pai and Behr, 2016).

1.1.3.3 Acid fast bacillus (AFB)

AFB testing is another most widely used test employ by medical personnel for the detection of TB and monitoring how patients response to treatment. The test is carried out by collecting 3 sputum samples from the patient or collection of fluid using bronchoscope. Other samples that can be use include cerebral spinal fluid (CSF), urine, and other bodily fluids. The samples collected are further stain using the different types of dyes such as the Zihel-Nelson (ZN) stain. Sputum test revolves around examining the presence of the bacteria under the microscope. Initially sputum sample or mucus coughed up from the lungs is used. The AFB test looks for the present of MBTB using the microscope which appear to be red, rod-shaped bacteria. One of the advantages of AFB over IGRA and TST is that it can detect active TB. The main advantage of sputum test is that it can distinguish between active and latent TB and thus can dictate treatment selection (Selvakumar et al., 2006; Nwalozie et al., 2024).

1.1.3.4 Chest X-ray

CXR is another most widely technique adopted by healthcare expert for the detection of TB. The fundamental mechanism lies on the fact that X-ray machine can provide structural information and abnormalities in the lungs. Thus, X-ray must diagnose TB compared to other techniques but can indicate changes in the lungs as a result of the presence of the bacteria which causes cavities, nodule or fluid. Therefore, X-ray imaging only confirm the presence of the bacteria but cannot distinguish between latent and active TB. Moreover, they may fail to detect initial stages (Van Cleeff et al., 2005; Jaeger et al., 2013; Triasih et al. 2015).

1.1.3.5 GeneXpert

GeneXpert also known as Xpert MTB/RIF is regarded as one of the most rapid, automated, sensitive, and specific screening technique for the detection of extrapulmonary TB and resistant to Rifampin. GeneXpert is nucleic acid-based amplification test in which sputum samples is collected from suspected patients which is further mixed with chemical reagent. The mixture is placed in a cartridge and future place in the GeneXpert machine which is automated.

One of the advantages of this test is that it is amazingly fast usually 2 hours compare to IGRA which takes days. Another advantage of the test includes a minimal training is needed to run the machine. Moreover, the technique can be used to detect MBTB that are resistant to both isoniazid and rifampin. However, positive results need to be interpreted along with confirmatory and supporting test such as CXR.

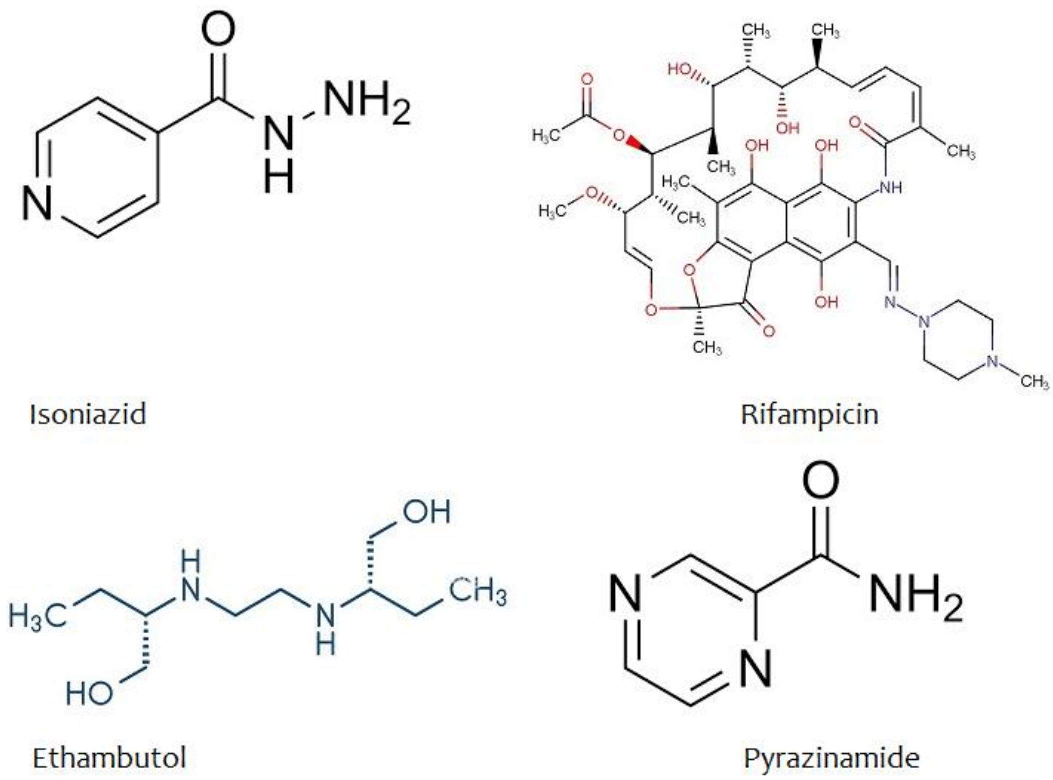
1.1.4 Treatment

Appropriate treatment of tuberculosis is highly critical for extend life span of patients and preventing widespread of the disease. Thus, treatment relies on the stage of the infection (primary, latent and active). Patients are prescribed combination of different antibiotics which lasted for several months depending on the severity of the symptoms. It is very crucial for patients to complete the dosage in order to complete destroy the bacteria (Campbell and Bah-Sow, 2006; Suárez et al., 2017).

The most common drugs administered to patients suffering from latent tuberculosis include isoniazid, rifapentine, and rifampin. While combination of different drugs such as isoniazid, ethambutol, pyrazinamide, rifampin as shown in **Figure 4** are used for the treatment of active tuberculosis. However, one of the challenges facing treatment of tuberculosis revolves around drug resistant types. Some strains of MBTB can become resistant to the drugs due to their genetic makeup or as a result of mutation. Other factors that influence resistant include incorrect prescription, not finishing the drugs and over prescription bacteria (Campbell and Bah-Sow, 2006; Suárez et al., 2017).

Figure 4

Anti-tubercular agents/medications



1.1.4.1 Isoniazid

Isoniazid is one of the most common drug use for the treatment of tuberculosis. Isoniazid is antibacterial drug that is prescribed for patients suffering from bacteria diseases and commonly used for the treatment of mycobacterium diseases such as active TB. The drug is used against MBTB, *M. bovis* and *M. kansasii*. The chemical formula of isoniazid is C6H7N3O. The mechanism of action revolves around the activation by bacterial catalase. The activation is linked with reduction of the mycobacterial ferric KatG catalase-peroxidase by hydrazine and reaction with oxygen to form an oxyferrous enzyme complex. After activation the drug act by inhibiting the formation of mycolic acid which is crucial component of the bacterial cell wall. There are several brands that produces isoniazid drugs (Unissa et al., 2016). These brands include Isotamine, Rifamate and Isonarif etc.

The complications of isoniazid is that it may interact with certain food such as cheeses, red wine, and some fish species due to the presence of tyramine/histamine. This interaction can lead to increase in blood pressure, headache, flushing of the skin, dizziness, or faster heartbeat. Side effects of the drug include weakness, loss of appetite, vomiting, nausea (McIlleron et al., 2007; Rangaka et al., 2014).

1.1.4.2 Ethambutol

Ethambutol is another anti tubercular drug that use in combination with other medications such as isoniazid for the treatment of TB. Another brand name of Ethambutol is called Myambutol. However, the mode of action of the drug is still unknown. But there is evidence that the anti-tubercular agent exerts its bacteriostatic activity by inhibiting an enzyme known as arabinosyl transferase which polymerizes arabinose into arabinan and then arabinogalactan, a mycobacterial cell wall component. Side effects include swollen joints, chills, headache, loss of appetite, nausea and vomiting, vision problems, confusion etc. (Yee et al., 2003; Goude et al., 2009; Diallo et al., 2018).

1.1.4.3 Rifampin

Rifampin with the chemical formula: C43H58N4O12 is another common drug that is prescribed for the treatment of bacterial infections which include tuberculosis (latent TB) and meningitis caused by *Neisseria meningitides*. Some of the common rifampin brands include Rimactane and Rifadin (Alfarisi et al., 2017). Rifampin some of the side effects include nausea and vomiting, skin rash, loss of appetite, diarrhea etc. (Törün et al., 2005; Diallo et al., 2018).

1.1.4.4 Pyrazinamide

Pyrazinamide is another anti-tubercular drug that is used for the treatment of TB. Like other anti-tubercular medications, it is used in combination with other drugs. The mode of action of the drug relies on slightly acidic pH where it becomes activated to Pyrazinoic acid in the bacilli where it interferes with fatty acid synthase FAS I. This process intercedes with the ability of the bacteria to produce new fatty acid necessary for growth and replication (Nijre et al., 2016). Some pyrazinamide side effects include joint or muscle pain, nausea and vomiting, lack of energy (Mohan and Sharma, 2004; Zhang et al., 2021).

1.2 Statement of the problem

TB continue to wreak havoc in many underdeveloped countries with substandard healthcare facilities, lack of proper screening tests and treatment as well as cost of medications. Even though there are several screening techniques employed by healthcare practitioners for the screening of diseases, however, each technique has its own drawbacks. In spite of the high reliance of ZN staining technique and X-ray imaging, both techniques are hampered by low accuracy, low sensitivity for early stage detection such as primary and latent TB, the need for trained staffs, biohazard and contaminations, the high workload of carrying out several tests and interpretation which can be prone to miss-interpretation and errors. Screening of TB cases using microscopy is hampered by numerous challenges which include the small size of the bacteria, the overlaps between bacilli in the slides or dish, the heterogeneous shape and irregular appearance of the bacteria, faint boundaries, and low background contrast etc. (Ibrahim et al., 2021).

Notwithstanding, conventional diagnosis of tuberculosis continue to suffer from several setbacks due to the continuous development of multidrug resistance types, Extra-Pulmonary TB (EPTB) and patients diagnosed with both TB and Human Immunodeficiency Virus (HIV). The invention of molecular testing based on serological testing such as antigen and antibody tests and nucleic acid-based techniques such as PCR and GeneXpert have shown improved specificity and sensitivity, minimal processing time, modern sample processing techniques etc. However, despite the edge of these approaches over conventional techniques such as the TST, AFB, they are limited in terms of generating false positive results and high cost of the procedures especially in underdeveloped countries (Singh et al., 2015; MacGregor-Fairlie et al., 2020).

To address these tenacious issues regarding accurate, safe, environmentally friendly diagnosis, scientists adopted AI-based techniques which include both ML and DL-based approaches. AI-based techniques are designed using algorithms that can detect underlying features from images. These underlying features can be extracted using the sophisticated architecture of deep neural networks which are designed using several hidden layers (Jogin et al., 2018). Several DL models have been designed or repurposed for improving diagnostic accuracy of tuberculosis from both AFB and CXR images.

1.3 Significance of the Study

The exponential growth of human population couple with aging population, sedentary lifestyles, stress, and increase in drug addiction is contributing to increase in the growing number of patients which in return contribute to increase workload and burnout faced by medical experts. Medical errors are regarded as one of the leading cause of death globally. Majority of public healthcare facilities are operating with insufficient staffs and substandard medical tools which increase the risk of miss-diagnosis and treatment (Brady et al., 2019; Regina et al., 2021). Moreover, the pharmaceutical companies are facing recall of drugs due to impotency, adverse side effects and expiration (Miglani et al., 2022).

TB is a fatal communicable and contagious disease that affect millions of people globally. The growing amount of multi-resistant drug type is l contributing to growing number and mortality rate (Wulandari et al., 2024). Early and accurate screening of patients suffering from the disease is highly crucial for prevention and treatment. Medical specialists rely on several diagnostic techniques which include TST, AFB and CXR. Conversely, these procedures have several limitations which include low accuracy, false positive results and can be prone to miss diagnosis due to overload (Brodie and Schluger, 2005; Campbell and Bah-Sow, 2006; Suárez et al., 2017). Secondly, the treatment of TB relies on the combination of several antibiotics. However, the growing number of MDR types are limiting the use of medications which contribute to the virulence and contagiousness of the bacteria (Millard et al., 2015; Wulandari et al., 2024). Therefore, there is need to develop a viable solution that can address these limitations.

Scientists have proposed solutions that can be used to mitigate the problem of medical errors. Among these solutions is the application of AI which employs algorithms to find similarities, features and patterns from a particular dataset and provide outcome in terms of classification, prediction. In line with this, this research proposed the deployment of AI-based technologies. The technique involves the use of ensemble DL and ML models for the detection of TB from 2 clinical images: AFB and CXR. Thus, this study will not only contribute to the growing literature or knowledge but also provide a background for real-time diagnosis of TB in healthcare settings.

1.4 Aim and Objectives

The main aim of this thesis is to provide accurate detection of TB from both X-ray and AFB images.

The objectives of this study include the following:

- Binary classification of AFB images into TB and Normal cases using DL-based models
- Binary classification of CXR images into TB and Normal cases using DL-based models
- Fusion of DL and ML models and comparison between 2 classifiers
- Performance evaluation of proposed ensemble models and comparison with state-of-the-art.

1.5 Assumption and Limitations of the Study

In this thesis, several pre-trained models are used which include MobileNet, ResNet-50, ResNet-101, VGG-16 and VGG-19 fused with either SVM or KNN for the 2-way classification of TB from microscopic slides and radiographic images. However, apart from these pre-trained models, there are other high performing CNNs such as Inceptionv3, DenseNet-121, EfficientNet, UNet, ResUNet, NANSNet, InceptionResNet etc. which are not part of the scope of this thesis. Secondly, only 2 classifiers are employed while other classical machine learning classifiers such as decision tree, random forest, linear and logistic regression are not used.

1.6 Structure of the Thesis

This thesis is structure or organized into 6 chapters. Chapter 1 covers the background on tuberculosis, epidemiology, transmission, pathology, diagnosis, and treatment. The chapter also present the thesis statement of problem, significant of the study, outlined aims and objectives as well as assumptions and limitations. Chapter provide detail information on the concept of AI, ML, DL, ANNs CNNs. Moreover, the chapter also covers the application of artificial intelligence in medicine and computer-aided detection. Furthermore, the chapter provide literature review on the implementation of AI and CAD for detection of TB from both microscopic slide and radiographic images. Chapter 3 covers the overall methodology, which include theoretical and practical framework, data collection, pre-processing, split, pre-trained models, classifiers, training, and evaluation using performance metrics. Chapter 4 present the result and discussion using tables and figures while chapter provide a holistic conclusion based on the performance, challenges, and future perspective.

.

CHAPTER II
Literature Review

2.0 Introduction

This chapter outlines the concept of AI and its sub-branches which include ML such as Supervised ML, Semi-supervised ML Unsupervised ML, Reinforcement ML as well as DL and subtypes which include CNNs and ANNs and TL. The chapter also reviewed existing studies that implemented AI-assisted techniques for the screening of TB from both MS images and CXR slices.

2.1 Artificial Intelligence (AI)

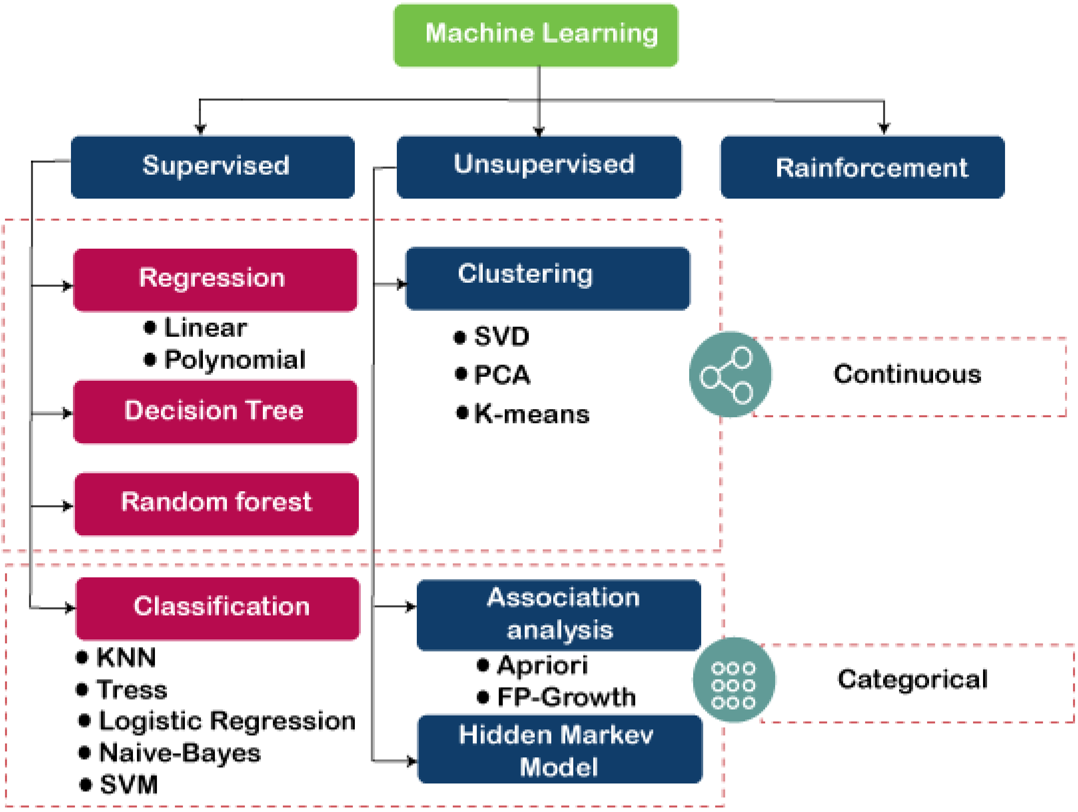
AI is a sub-field or a branch of computer science that concentrates on the development of intelligent systems that can mimic human cognitive function such as learning, decision-making, memory, understanding patterns etc. or that can solve tasks peculiar to humans. AI uses mathematical algorithms to analyze, classify and draw prediction from a given data. AI leverages several technologies which include machine learning, deep learning, autonomous machines, visions, robotics, Neuro-Linguistic Programming (NLP) etc. (Zhang and Lu, 2021). AI is designed to use algorithms to find similarities, features and patterns from a given dataset and provide result in terms of classification, prediction or decision making. AI have emerged as innovative technology that is transforming every sector and all facet of life. Some of the tasks AI can perform include image recognition, handwriting and voice recognition, language translation, prediction of stock market, autonomous vehicle navigation, search engines, social media etc. (Cioffi et al., 2020; Javaid et al., 2022).

2.1.1 Machine Learning

ML is a branch of AI and computer science that concentrates on the application of algorithms using data to enable machines replicate human cognitive function such as learning, memory, problem solving and gradually improving its accuracy (Helm et al., 2020). The basic idea revolves around the supply of data input (labeled or unlabeled) and subsequent classification or prediction using machine learning algorithms. ML can be sub-divided into 3 or 4 types which include supervised, semi-supervised, unsupervised and reinforcement learning (Jakhar and Kaur, 2020) as displayed in Figure 5.

Figure 5

Classification of Machine Learning (https://www.javatpoint.com/machine-learning-algorithms)



2.1.1.1 Supervised Machine Learning

Supervised ML is defined as the use of ML algorithms and labeled dataset with the aim of reaching an output such as classification (according to the quantity of categories) or prediction. In supervised ML, the labeled dataset are fed into ML algorithms where the model optimize its weights until it become fitted applicably. SML algorithms are used in different domain which include computer vision, weather forecast, detecting spam email, prediction and classification of diseases etc. Example of supervised ML algorithms include SVM, KNN, Random Forest (RF), Naive Bayes (NB), logistic regression, linear regression, neural networks etc. (Jiang et al., 2020; Burkart and Huber, 2021; Shetty et al., 2022).

2.1.1.2 Unsupervised Machine Learning

Unsupervised ML is characterized by the use an algorithm fed with unlabeled dataset (I.e., unstructured, unsorted, or scattered). The algorithms is task with clustering dataset with shared similarities or features. In another word, unsupervised ML algorithms identify patterns within a given dataset which result in grouping (Ghahramani, 2003). The ability of unsupervised ML algorithms to find underlying patterns within a data make it an ideal for several applications which include image and pattern recognition, exploratory data analysis, customer segmentation. Unsupervised ML algorithms are based on clustering algorithms which cluster data that share similar features together (Wong et al., 2021). An example of this type of algorithms includes Principal component analysis (PCA), K-means clustering, probabilistic clustering, and singular value decomposition (SVD) etc. (Alzubi et al., 2018; Naeem et al., 2023).

2.1.1.3 Semi-supervised Machine Learning

Semi-supervised ML is the type of algorithms that bridge both supervised and unsupervised learning and the combination of the two. Semi-supervised ML both used labeled and unlabeled dataset (Reddy et al., 2018; Van Engelen and Hoos, 2020). In this technique, models are initially trained with small portion of the dataset before iteratively running the model on a larger dataset. During training, the algorithms is fed with small ratio of labeled data samples in order to serve as a guide for the clustering or classification of unlabeled dataset. Semi-supervised ML play a vital role in solving limitations of both task that require the use of sorted and unsorted dataset. It offer an edge over unsupervised ML by using labeled dataset to guide accurate clustering. Another advantage is time and cost saving (Hady and Schwenker, 2013; Zhou and Zhou, 2021).

Semi supervised ML are used for several applications revolving around clustering, classification, prediction and regression. Semi supervised ML are used in speech recognition and analysis, classification of web contents, text document classification, spam filtering, customer segmentation as well as drug analysis and protein sequence classification. Some of the advantages of this type of model include time and resource saving especially when it comes to annotation and manual labeling of datasets. Some of the common techniques use if semi supervised ML include pseudo labeling, self-training and label Propagation (Huang et al., 2006; Reddy et al., 2018).

2.1.1.4 Reinforcement Machine Learning

Reinforcement ML is another type of ML that share close similarities with supervised ML. Unlike the use of labeled data, reinforcement ML use trial and error to enable models to learn overtime. A series of accomplished task will be reinforced or rewarded in order to develop the best performance (Li et al., 2017; François-Lavet et al., 2018). In other words, Reinforcement ML is a sub type of ML that enable computer agents to perform certain actions in each environment based on trial-and-error methods while providing a reward based on feedback responses. Reinforcement ML enables a computer agent to learn to achieve tasks based on reward-feedback actions. Example of Reinforcement ML include Q-learning, Monte Carlo, Deep Q-network (DQN) and state–action–reward–state–action (SARSA). Reinforcement ML algorithms are use in self-driving vehicles, gaming, signal traffic control, robotics, etc. (Qiang and Zhongli, 2011; Mahesh, 2020; Mehta, 2020).

2.1.2 Deep Learning

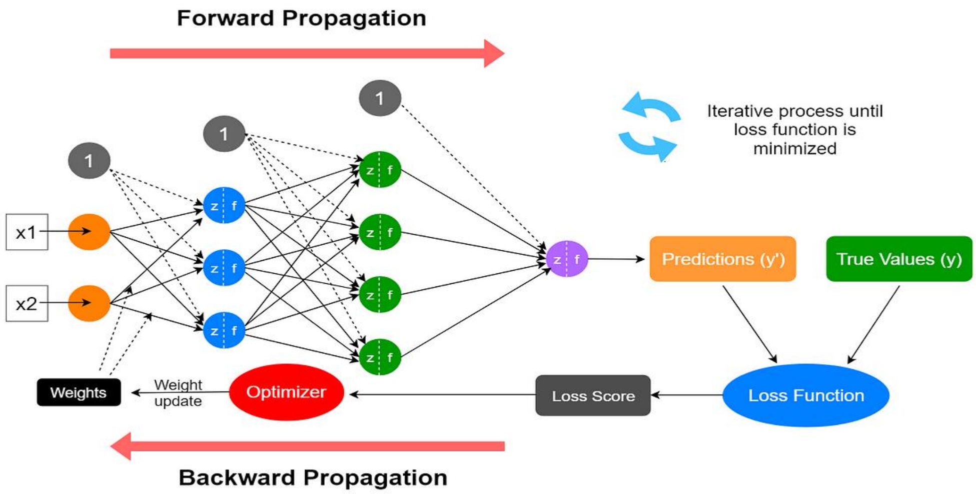
DL is a subsidiary of ML that utilizes architectures such as ANNs, CNNs, recurrent neural network (RNNs) for solving tasks related to computer vision, NLP, audio or speech recognition, machine translation etc. The term deep learning is characterized by neural networks that comprises of more than 3 layers inclusive of the input and output layers. DL-based models have been credited for improving performances in several applications which include computer vision, speech recognition and NLP (Michelucci, 2018; Janiesch et al., 2021; Razavi, 2021).

2.1.2.1 Artificial Neural Networks

ANNs are type of deep NNs which comprises of 3 or more layers, an input layer, one or more hidden layers and output layer. Each neuron in the network is connected to another neuron and has an linked weight and threshold. The node can be activated when the output raise above the specified threshold value and thereby transmitting data to the next layer of the network. The training process in which neural network models learn is based on iterative process where optimization is conducted on both forward and backward through each layer in the network the loss function is minimize (Guresen and Kayakutlu, 2011; Krenker et al., 2011; Samek et al., 2021). The overall process can be classified into 3 which include forward pass (forward propagation), optimization of the loss function and back propagation (backward pass) as displayed in **Figure 6**.

Figure 6

Architecture of Artificial Neural Network (Rukshan, 2022)



2.1.2.2 Convolutional Neural Network (CNN)

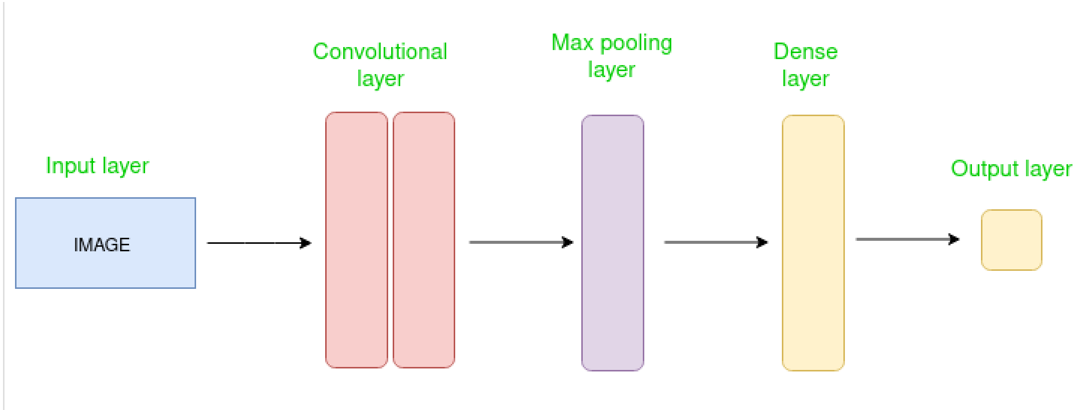
CNN also known as ConvNet is a sub-type of NNs or DL architecture that is developed to solve tasks which include image recognition, detection and classification, segmentation. ConvNet deeper with traditional ML models such as Decision Trees and SVMs by their prowess to independently extract features at large scale thereby enhancing efficiency and bypassing manual feature engineering. The stacks of convolutional layers enable CNN their translation-invariant characteristics which enables them to automatically identify and extract features and patterns from a given data irrespective of variations in scale, orientation, translation or position (O'shea and Nash, 2015; Jacovi et al., 2018; Wang et al., 2020).

CNN models are powerful tools that require thousands or millions of labeled data, high power process such as Graphical Processing Unit (GPU) for training in order to achieve significance performance. CNN comprises of several layers and operations which include input layer, convolutional layer, pooling layer, fully connected layers and output layers (O'shea and Nash, 2015; Teuwen and Moriakov, 2020). The convolutional operations revolve around the use of filters (filter map or kernels) to extract features from the input image while pooling layer down sample or reduce the image which minimize computation, prevent overfitting and reduce memory. The activation layer function by adding nonlinearity to the network.

Examples of activation function include RELU: max (0, x), Tanh, Leaky RELU, etc. (Aghdam and Heravi, 2017; Wang et al., 2020; Teuwen and Moriakov, 2020). The fully connected layer use probability approach to make prediction. A simple representation of CNN is presented in **Figure 7**.

Figure 7

Schematic Representation of Convolutional Neural Network (CNN)

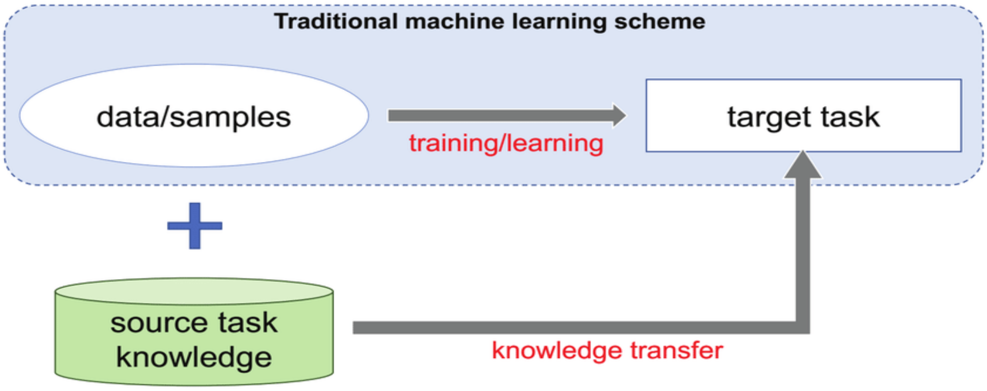


2.1.3 Transfer Learning (Pre-trained Models)

The concept of TL evolves because of the fact that training DL architectures require tremendous amount of data to produce significant result. Moreover, the process is time consuming, require computing power and require intensive process. To address this issue, scientist proposed the use pre-trained models (Marcelino, 2018; Neyshabur et al., 2020). TL is a DL-based approach in which a model pre-trained on a task mostly with large volume or large amount of dataset is repurpose or fine-tune on a new or related task. In another word, TL is smart technique in ML and DL where a model employ knowledge such as weight from previous task to solve problem regarding new task. For example, models trained to classify ImageNet dataset such as dog and cat and be fine-tuned to classify medical images (Ribani and Marengoni, 2019; Kansert et al., 2019). **Figure 8** highlight the concept of TL.

Figure 8

Transfer Learning (TL) (Shashank Srivastava, 2019)



Some of the advantages of pre-trained models include improve efficiency and performance, reduce cost and time and overfitting. TL models play a crucial function in achieving high performance in the case of data scarcity. Thus, TL models are employed in computer vision and NLP. Scientists employ several pre-trained CNN models, which include Le-Net, AlexNet, VGGNets, ResNet, InceptionNets, DenseNets, MobileNets, and EfficientNet etc. (Marcelino, 2018; Ribani and Marengoni, 2019: Salehi et al., 2023).

2.2 Application of AI in Diagnostic and Treatment

The increase in global population coupled with urban migration, over crowdedness, deforestation, global warming, lack of social amenities and poverty in underdeveloped countries is contributing to emergence and re-emergence of both communicable and non-communicable diseases. The healthcare sector is burden by growing populations and the need of innovative technology that can minimize the workload face by medical expert and improve diagnostic and treatment efficiency (Sarkar et al., 2012; Lee et al., 2013). The growing healthcare data is fueling artificial intelligence engines which have shown to aid in analyzing structural and unstructured data (Kaur and Singh, 2017; Van Hartskamp et al., 2019).

AI is perceived by many scholars as game-changer or transformer in healthcare sector. AI technologies such as intelligent automation, virtual assistant cognitive computing etc. has shown to aid scientists developed models that can detect diseases at early stage which is critical for preventive treatment. AI diagnostic applications is aiding in improving quality of healthcare sector by bridging the gap between demand and supply (Zhang et al., 2019; Summerton and Cansdale, 2019; Ao et al., 2020). AI-based methods have now been incorporated for the screening of several types of cancer. AI is employed to identify cancer at an early stage more precisely and faster compare to conventional approaches (Ström et al., 2020; Haung et al., 2020). AI models also play a vital role in electrophysiology where electrocardiographs and electroencephalograms are inputted into DL models for classification or enhancement. AI-based techniques have been applied for screening of pneumonia (which include bacterial pneumonia, COVID-19 pneumonia and non-COVID-19 viral pneumonia such as influenza virus) (Ibrahim et al., 2021a; Umar Ibrahim et al., 2022a).

AI is transforming treatment by helping scientists find cure of diseases such as cancer and genetic diseases or gene therapy. AI is set to make difference in various medical field ranging from precision medicine, patient centric medicine, surgery etc. (Jiang et al., 2017; Secinaro et al., 2021). The field of AI is driving industrial growth within the healthcare sector as it is widely used in clinical research, drug development, electronic health records (EHR) analysis and management. By next year (2025), the global market of AI in healthcare is estimated to grow over 13 billion USD (Horgan et al., 2020; Zhukovska et al., 2023).

Thus, the incorporation of AI in diagnostic medicine and medical imaging can't only help improve diagnostic accuracy but also predict the possibility of diseases. AI models have started outperforming medical expert in terms of diagnostic, however, it is still a long way before AI-based models replaced medical expert completely (Pesapane et al., 2018; Tang, 2019; Shuaib et al., 2020; Rony et al., 2024). Another significant limitation of AI in diagnostic medicine and treatment involves the issue and concern related to data privacy, confidentiality, data breach, cyber-attack etc. (Murdoch, 2021; Khalid et al., 2023; Gabriel, 2023).

2.3 Computer-aided Detection (CAD)

CAD often referred to either Computer Aided Detection (CADe) or Computed Aided Diagnosis (CADx) is a computer-assisted technology that aid medical specialists in reaching decision. In the medical settings, healthcare workers employ medical images in order to analyze information such as irregularity for accurate diagnosis. Thus, proper interpretation of clinical images as positive or negative, healthy or unhealthy, normal or abnormal and multi classification is highly crucial considering the fact that any miss-diagnosis can result in serious consequences (Castellino, 2005; Fujita et al., 2008; Petrick et al., 2013). Different branches or department in the clinics utilized various medical imaging techniques which include microscopes in microbiology, histopathological stain slides and biopsies in pathology and oncology, X-ray, CT scans, MRIs by radiologists, Position emission tomography (PET), Single emission positron tomography (SPECT) and hybrid (PET/CT, SPECT/CT) in nuclear medicine. These images contribute to the large volume of data generated by clinics which can fuel CAD and AI-based technology (Irkham et al., 2022; Barua and Mondal, 2023).

Consequently, the large medical imaging data produce can be process, classify, segmented, extracted using CAD. The main function of CAD in clinical applications revolves around the identification of anomaly in clinical images such as the detection of potential Regions of Interests (ROIs) as well as providing computed image metrics to calculate probabilities of different screening (Cicerone & Camp, 2019). CAD are currently use for detection and classification of different types of cancers (breast, brain, colon, colorectal, prostate, skin etc.) into malignant and benign or into grades using histopathological images, ultrasound and radiographic images. Moreover, the technology is also adopted for the detection of microbial infections such as pneumonia and TB using microscopic slide images and radiographic images (Jorritsma et al., 2015; Oakden-Rayner et al., 2019). Other applications of CAD include detection of eye diseases such as glaucoma and diabetic retinopathy (Zhang et al., 2014; Tamim et al., 2021), coronary artery diseases (Faust et al., 2017), Alzheimer's disease (Illán et al., 2011) etc.

2.4 Application of AI in Detection of Tuberculosis

The current literature is growing due continuous contributions made by scientist in the field of microbiology and radiology. Computational microbiology and radiology are helping scientists make accurate diagnosis and proper decision making. AI-based models are employed for the screening of TB from radiographic images and MS images (Umar Ibrahim et al., 2022b). Majority of AI or CAD techniques employ revolve around the use of pre-trained models, which are already trained using large volume of data. Other studies reported in the literature implemented either pre-trained models such as ResNets (18, 50, 101 and 152), AlexNet, VGGNets (16 and 18), Inception, DenseNet, EfficientNet etc. or segmentation models such as UNet and UNet++ or models developed from scratch. However, the last 5 years have witnessed the raise of ensemble models, which combined both DL models and ML models such as traditional algorithms or classifiers (SVM, KNN, DT, RF, LR etc.) (Ganaie et al., 2022; Mahajan et al., 2023).

2.4.1 AFB Images

The study proposed by Kant et al., 2018 reported the application of Deep NN and ML models by for the detection of TB from MS images. To achieve this objective, the study designed as experiments based on following techniques and steps which include data acquisition, data pre-processing, deployment of DL-based models and performance assessment. The first step includes collection of Ziehl-Neelsen Sputum smear Microscopy image Database (ZNSM-iDB) available dataset which is publicly accessible. The dataset comprises of 202 images which split into 60% for training, 20% for validation and 20% for testing. The acquired images are trained and evaluated using 5-layered CNN model designed using several convolution operations, ReLu as activation function and SoftMax as the classifier. The process includes the use of SVM with HoG features as the baseline. Both CNN and SVM are employed to classify the images (based on patchwise classification). The performance evaluation of proposed approach resulted in 74.79% F1-Score, 68.55% precision and 83.78% recall.

Panicker et al., 2018 proposed the automatic screening of TB based on image binarization and the use of CNN for subsequent. The proposed approach is implemented in 2 steps. The study acquired 22 sputum MS images with different background which include low-density and high-density images from Instituto Nacional de Pesquisas da Amazonia) lab, Manaus, Brazil. The images are further processed by denoising via fast nonlocal means (FNLM), binarized using Otsu's method and segmented to differentiate the foreground and background of the images. The acquired images are further augmented to increase the volume via rotation, vertical and horizontal reflections which resulted into 1800 patches (900 positive and 900 negative cases). The segmented images of the foreground that contain the bacilli are fed into the CNN model designed using 2 convolutional layer, batch normalization, a fully connected layer, ReLu activation and a sigmoid neuron for classification. Performance assessment of the proposed approach resulted F1-Score of 86.76%, precision of 78.4% and recall of 97.13%.

The use of DL based on TL technique for the screening of TB from MS images is proposed by Ibrahim et al., 2021b. The study conducted a 2-way classification of healthy and infected cases using pre-trained AlexNet. The overall experimental set is designed according to numerous steps which include data collection, data pre-processing, data augmentation, model implementation, training and testing and performance evaluation. The study acquired 530 images from Near East University which comprises of 200 positive cases and 330 negative cases. The images are pre-processed via labelling, data partitioning into 70% for training and 30% for testing and resizing to fit into the model input size requirement (227 x 227 x 3). Moreover, data augmentation via rotation at 3 angles (90, 180 and 270) degrees and cropping were conducted which enlarge the training set to 2444 images 1320 positive and 1144 negative). The pre-trained CNN is trained using the training set and evaluated according to 10k-cross validation. The average result achieved based on several evaluation metrics resulted average accuracy of 98.73%, average sensitivity of 98.59 % and average specificity of 98.84% respectively on testing set. Moreover, the study also conducted machine vs human analysis based on pre-trained model against human pathologists which include beginners, certified 1 (less than 10 years of experience) while certified 2 (more than 10 years of experience). The result of the comparison has shown that achieved the same accuracy with certified 1 and 2 while outperforming beginners.

The study proposed by Fu et al., 2022 reported the implimentation of AI-based framework for the automatic screening and detection of TB from AFB under acid-fast staining. The proposed approach is designed based on steps which include development of CNN architecture, collection of data, processing, evaluation performance and comparison between machine and manual approach. The first step revolves around the development of a CNN architecture which is developed by µ-Scan 1.1, Wallen Medica, Kaohsiung, Taiwan. The second step revolves around the collection of AFB images which comprises of 5930 smears. The acquired images were processed by varying the number of smears from 120 to 200 and eliminating disqualifying smears produced by poor staining quality and smear preparation. The next stage involve training of the developed CNN using the training set and subsequently performance assessment to measure the generalizability of the architecture. The performance evaluation indicated that the developed CNN achieved 95.2% accuracy, 85.7% sensitivity, and 96.9% specificity. Subsequently, the comparison between the developed CNN and manual screening has shown that CNN model recovered 85 positive results which are classified inaccurately using manual screening.

The use of Ziehl-Neelsen stain AFB images for the accurate classification of tuberculosis into infected and non-infected using AI-based method named RegNetX4 architecture was reported by Zurac et al., 2022. The whole methodology was structured based on data acquisition, data split, data augmentation, model training and assessment of proposed approach. In the case of data acquisition, the study acquired MS images from Department of Pathology of Colentina University Hospital which comprises of 510 images in which 400 are negative cases and 110 are positive cases. The images were split into training (550 images) and testing (60 images). The training set images were further annotated prior to carrying out numerous data augmentation techniques on the training set which include rotation, random shifting, saturation, contrasting and brightening. The overall data augmentation processes increase the dataset into approximately 1 million patches (700,000 negative and 260, 000 positive). The testing set are also annotated to yield 286, 000 patches. Evaluation of the model according 5 performance metrics resulted in 98.33% accuracy, 95.65% sensitivity, 100% specificity and 0.977 ROC.

Another study proposed by the same author (Umar Ibrahim et al., 2022b) reported the binary classification of MS images into tuberculosis and healthy cases. Unlike previous study that proposed the use of pre-trained AlexNet embedded with SoftMax, this study compare the performance of pre-trained AlexNet ensemble with SVM and pre-trained AlexNet embedded with SoftMax. The experimental set-up is structured based on data collection from Near East Hospital in which 530 images are acquired and merged with 1263 positive cases acquired from Kaggle domain (which led to sum of 1893 images). The acquired images undergo pre-processing based on size adjustment, labelling and data split into training (1325) and testing (568) according to 70:30 split respectively. The next stage involve training of the two techniques, performance evaluation and comparison to establish the model with better performance. The result has shown that ensemble Pre-trained AlexNet merged with SVM attained the top performance with 98.73% accuracy, 99.42% sensitivity and 98.03% specificity compared to ensemble Pre-trained AlexNet merged with SoftMax, which achieved 98.14% accuracy, 96.89% sensitivity and 99.38% specificity.

The study proposed by Lee and Lee, 2023 proposed the application of AI-based approach in order to solve the issue regarding time consuming and labor intensive surrounding microscopic examinations of slides at 1000× magnification. To achieve this, the study structured the methodology based on sample collection, image pre-processing, implementation of pre-trained models and performance evaluation using holdout set. The research acquired 40 slides with a 400x magnification which are collected from the National Forensic Service of Korea. The acquired slide are partitioned into 75% for training (30 slides) and 25% for testing (10 slides). Rather than direct classification using the 40 slides, the study extracted 98,034 spot patches in which 47,017 are positive and 47,017 are negative. The spot patches are fed into 9 pre-trained models, which include Vision Transformer, Swin Transformer, RegNet, NASNet, XceptionNet, and Inception v3, EfficientNet, DenseNet and ResNet-50. The evaluation of the model performances and comparative analysis between the 9 models has ranked NASNet as the top-best model which achieved optimum accuracy of 99.777 ± 0.0231, recall of 99.771 ± 0.0175, precision of 99.728 ± 0.0356 and FI-Score of 99.749 ± 0.0260 per patch.

Another study that employed Ziehl-Neelsen AFB images for the 2-way screening of MS images into tuberculosis and non-tuberculosis cases using AI-based method was reported by Gupta et al., 2023. The methodology of the study was organized based on collection of datasets, pre-processing, data split, training and testing and performance evaluation. The first step revolves around the collection of 400 clinically suspected cases of tuberculosis from 3 different centers in northern part of India. The acquired undergo several processing steps such as normalization and resizing. The pre-processed images are further fed into AI-based microscopy and microscopist platform which is designed from scratch. The subsequent assessment of the proposed technique using testing set resulted in 91.53% accuracy, 89.25% sensitivity, 92.15% specificity, 75.45% Positive Predictive Value (PPV) and 96.94% Negative Predictive Value (NPV).

The study proposed by Waluyo et al., 2023 reported an automated method for the screening of TB from AFB sputum samples. In order to achieved accurate binary classification, the research experiment was set up according to 6 phases which include curation of image dataset, image pre-processing, color segmentation, oversampling based on data augmentation techniques, feature extraction and classification of images into binary classes. The first phase revolves around data curation of 220 sputum images from the Semarang City Health Center. The curated dataset was processed based on data cleaning, resizing and color segmentation. The pre-processed images were further partitioned into 82% for training (which contain 180 images in which 115 are TB cases and 65 non-TB cases) and 18% for testing (which contain 40 images in which 20 are TB cases and 20 non-TB cases) The images are trained and tested using 6-layered CNN developed from scratch embedded with SoftMax as a classifier and 6-layered CNN fused with KNN. Performance assessment of the 2 proposed approach has shown that 6-layered CNN merged with KNN achieved the best result with 92.5% accuracy, 92.5% recall, 93.5% precision and 92.5% F1-Score. The summary of the related work on the application of AI-based techniques for detection of TB and non-TB from MS images is presented table 1.

Table 1

Summary of Related work on the application of AI-based techniques for detection of Tuberculosis from MS images

References	Number of Images	Number of classes	AI-based model used	Result achieved
Kant et al., 2018	202	2	5-layered CNN	74.79% F1-Score, 68.55% PR and 83.78% RC
Panicker et al., 2018	1800	2	CNN	86.76% F1-Score, 78.4% PR and 97.13% RC
Ibrahim et al., 2021b	530	2	Pre-trained AlexNet	98.73% ACC, 98.59 % SV and 98.84% SP
Zurac et al., 2022	510 (1.1 million patches)	2	RegNetX4	98.33% ACC, 95.65% SV, 100% SP and 97.70 ROC
Umar Ibrahim et al., 2022b	1893	2	Pre-trained AlexNet + SVM	98.73% ACC, 99.42% SV and 98.03% SP
Fu et al., 2022	5930	2	CNN	95.2% ACC, 85.7% SV and 96.9% SP
Lee and Lee, 2023	40 slides (98, 034 patches)	2	NASNet	99.777 ACC, 99.771 RC, 99.728 PR and 99.749 FI-Score
Gupta et al., 2023	400	2	AI-based microscopy	91.53% ACC, 89.25% SV, 92.15% SP, 75.45% PPV and 96.94% NPV
Waluyo et al., 2023	220	2	CNN + KNN	92.5% ACC, 92.5% RC, 93.5% PR and 92.5% F1-Score

*ACC: Accuracy; RC: Recall; PR: Precision; SP: Specificity; SV: Sensitivity; PPV: Positive Predictive Value; NPV: Negative Predictive Value.

2.4.2 Chest X-ray (CXR) Images

As one of the first studies on the implementation of DL-based TL approach for the screening of TB from CXR images, Lakhani and Sundaram (2017) proposed the application of models developed from scratch and pre-trained models for the 2-way screening of CXR images into TB and healthy cases. The overall system is organized according to several steps which include curation of data, data pre-processing, data split, data augmentation, training and testing, evaluation, and comparison. The first step includes the curation of image dataset which include Maryland and Shenzhen and Montgomery count dataset which are publicly accessible and Belarus TB portal and Thomas Jefferson University hospital. The combination of the two datasets produces 1007 total images. The second step involve image data pre-processing based on labelling and resizing. The third step include data partition into training (68%), validation (17.1%) and testing (14.9%). To maximize the amount of training set, 2 data augmentation methods which include rotation and cropping were conducted. In order to distinguish between TB and healthy cases, untrained ensemble AlexNet and GoogleNet and pre-trained ensemble AlexNet and GoogleNet were employed. The performance evaluation and comparison between pre-trained and untrained models has shown that pre-trained ensemble AlexNet and GoogleNet accomplished the optimum result with 97.3% sensitivity, 94.7% specificity and 99.0% AUC score.

The study proposed by Hooda et al., 2017 reported the use of DL-based approach for the discrimination of CXR images into positive (TB) and negative (non-TB) cases. To achieve this objective, the study developed a feed forward 19-layer CNN model which is designed with 7 convolutional layers, 7 ReLu layers, 2 dropout layers and 3 fully connected layers. The process revolves around feature extraction technique and the use of extracted features to train the CNN model using different optimization parameters. The study acquired 2 dataset which include the Montgomery which contain 138 CXR images (58 TB and 130 healthy cases) and Shenzhen dataset which comprises of 662 CXR images (336 TB and 326 healthy cases). The acquired datasets combined sum of to 800 images. The total images are further split into 75% for training (600) and 25% for testing. The proposed approach is evaluated based on 3 optimization techniques which include Adam optimizer, stochastic gradient descent (SGD) optimizer and momentum optimizer. The performance assessment of the developed CNN based on different optimization techniques has shown that Adam optimizer achieved the highest score across all the metrics with 94.73% accuracy.

Xiong et al., 2018 proposed an automated method based on AI-based technique for the screening of TB and healthy cases using CXR image dataset. The overall procedure which outlines the steps and technique employed to achieve accurate detection are organized according to 5 phases. Phase 1 is characterized by the curation of CXR image dataset from Peking University hospital which comprises of 246 images. Phase 2 involved data pre-processing based on sorting and resizing. The next phase is characterized by the deployment of pre-trained CNN for training and testing. Contemporary to the majority of existing studies that trained models using large ratio and testing using smaller ratio, this study trained the pre-trained CNN model using 45 images (30 positive and 15 negative cases) and evaluated the performance of the model using reserve dataset which comprises of 201 images (108 positive and 93 negative cases). Second to the last phase involve performance evaluation of pre-trained CNN using holdout dataset. The result of the model assessment produced sensitivity of 97.94% and specificity of 83.65%. The last phase is characterized by comparative study between developed model and medical specialist based on 2 runs. In the first run, medical specialist outperformed the model while on the second run, the model achieved optimum result similar with medical specialist after undergoing modifications.

Norval et al., 2019 proposed the implementation of hybrid method which combine CAD technique and DL-based technique by using CNN to detect TB from CXR images. The methodology is organized based on steps which involve data collection, data pre-processing, execution of model through training and testing and performance evaluation. The study curated 2 datasets which include Shenzhen Hospital dataset (which comprises 662 in which 336 are healthy cases and 326 TB cases) and Montgomery dataset (which comprises of 138 images in which 80 are healthy cases and 58 TB cases) leading to 800 images. Several pre-processing techniques were employed which include data enhancement, histogram equalization, sharpening, reducing color Chanel and extraction of ROIs. The next step include the use of the extracted lung ROI for training D-CNN designed using multi-layer neural networks (MLNN). The performance evaluation of the hybrid method (CAD-MLNN) resulted in an accuracy of 92.54%.

The study proposed by Rahman et al., 2020 conducted an intensive procedure which involve data collection, data per-processing, data augmentation, image segmentation, classification, and model assessment. In terms of data collection, the study curated several publicly accessible datasets which include National Library of Medicine (NLM) dataset, RSNA CXR dataset, NIAID TB dataset and Belarus dataset which comprises of 3500 positive and 3500 negative cases. The acquired images undergo pre-processing based on resizing and Z-score normalization. The experiment was conducted based on (1) using segmented images extracted using 2 different U-Net models (original and modified U-Nets), (2) classification using X-ray images and (3) classification of segmented images. The images are trained and assessed using 9 pre-trained DL-based models, which include SqueezeNet, ResNet-101, ResNet-50, ResNet-18, MobileNet, DenseNet201, VGG19, InceptionV3 and ChexNet. The comparison between direct classification of X-ray images and the use segmented images has displayed that DenseNet201 attained the top result on segmented images with an accuracy of 98.6%, precision of 98.57%, sensitivity of 98.56% , 98.54% specificity and F1 score of 98.56%. While binary classification of CXR images has shown that ChexNet accomplished the top result with an accuracy of 96.47%, precision of 96.62%, sensitivity of 96.47%, specificity of 96.51% and F1-score of 96.47%.

The combination of both segmentation and classification techniques for the accurate detection of Pulmonary TB (PTB) and Extra Pulmonary TB (EPTB) using CXR images is reported by Sharma et al., 2021. The overall process was structure according to several procedure which include image curation, data pre-processing, segmentation, feature extraction, classification and performance assessment. The first procedure is characterized by the curation of image dataset from TB Specialists of Jalandhar region, Punjab, India in which 1000 CXR images are collected. The second procedure involved data pre-processing through enhancement and restoration. The third procedure is characterized by segmentation in order to extract ROIs. In order to salvage the segmented images, feature extraction as the fourth stage is triggered based on Histogram Filter and Median Filter combined with the CLAHE technique. The extracted images are subsequently partitioned into 80% for training and 20% for testing. The next procedure involved the implementation of ML classifiers which include DT, NB and SVM for the discrimination of CXR images into (TB and non-TB. The assessment of the models generalizability and subsequent comparison has placed DT classifier as the best classification model with 98% accuracy, 98% recall, 98% precision, 97.9% F1-Score and 99% AUC score.

The study presented by Umar Ibrahim et al., 2022B reported the deployment of AI-based TL model for the discrimination between TB and normal cases from CXR images. The research set up the experiment based on 4 stages which include the collection of CXR images, image pre-processing, model deployment and performance evaluation. The first stage is characterized by data curation of 7000 images which comprises of 3500 positive and 3500 negative cases from Kaggle repository which is publicly accessible. The second stage involve image pre-processing which involve data cleaning, sorting and resizing. The third stage involve the deployment of pre-trained AlexNet model embedded with SoftMax and pre-trained AlexNet fused with SVM. The ensemble models are trained using 70% of the image dataset (4900 images) and tested using 30% hold-out set (2100 images). Evaluation of the pre-trained models ensemble with 2 classifiers and comparative analysis has shown that the pre-trained AlexNet ensemble with SVM achieved significant result with an accuracy of 98.38%, sensitivity of 98.71% and specificity of 98.04% in comparison with pre-trained AlexNet embedded with SoftMax with 98.19% accuracy, 99.62% sensitivity and 96.76% specificity.

The study proposed by Akbari and Azizi (2023) developed highly efficient and time-saving AI-based technique for screening of tuberculosis from digitalized CXR images. The overall procedure is organized according to several stages ranging from image collection, image pre-processing, generation of synthetic data, model development and deployment and performance assessment. The first stage which involve data collection revolves around the curation of publicly accessible dataset from Kaggle repository which comprises of 4200 images (3500 negative cases and 700 positive cases). The next stage involves data processing which include resizing and labelling. The third stage revolves around oversampling of classes labelled as training set via data augmentation. This stage utilized several methods which include zooming, rotation, flipping and scaling in order to maximize the training set to 6155 images (i.e., 1630 additional negative images and 324 additional positive images). The fourth stage involved the development of CNN model from scratch and it deployment for training and testing of reserve dataset. The evaluation of the deployed CNN model, which was developed from scratch, produced an accuracy of 97%, precision of 98%, and recall of 90% and F1-Score of 90%.

The implementation of segmentation models and several DL-based models for the accurate screening of TB from CXR images is reported by Nafisah and

Muhammad (2024). The series of steps designed to achieve accurate detection revolves around image dataset collection, augmentation, segmentation, classification, and performance evaluation of proposed techniques. The study employed 3 publicly accessible datasets which include Shenzhen with 662 images, Belarus with 304 images and Montgomery with 138 images which led to 1098 total images. The overall acquired image dataset is pre-processed and maximized via data augmentation based on rotation. The next step involves the implementation of segmentation models in order to extract ROIs followed by the implementation of several pre-trained models which include MobileNet, ResNeXt-50, Inception-ResNet-v2, XceptionNet and EfficientNet-B3. The overall frameworks which involved both segmentation and classification are evaluated based on 10k cross validation and the result has shown that EfficientNet-B3 achieved better result compare to other proposed approaches with an average accuracy of 99.1%, recall of 98.3%, precision of 98.3%, ROC of 99.9% and Kappa score of 99.1% using reserve or testing set.

The summary of the related work on the application of AI-based techniques for detection of TB and non-TB from CXR images is presented table 2.

Table 2

Summary of Related work on the application of AI-based techniques for detection of Tuberculosis from CXR images

References	Number of Images	Number of classes	AI-based model used	Result achieved
Lakhani and Sundaram (2017)	1007	2	Pre-trained ensemble AlexNet and GoogleNet	97.3% SV, 94.7% SP and 99.0% AUC
Hooda et al., 2017	800	2	19-layer CNN	94.73% ACC
Xiong et al., 2018	246	2	Pre-trained CNN	97.94% SV and 83.65% SP
Norval et al., 2019	800	2	CAD-MLNN	92.54% ACC
Rahaman et al., 2020	7000	2	DenseNet201	98.6% ACC, 98.57% PR, 98.56% SV,

				98.54% SP and 98.56% F1 score
Sharma et al., 2021	1000	2	DT	98% ACC, 98% PR, 98% RC, 99% AUC and 97.9% F1-Score
Umar Ibrahim et al., 2022b	7000	2	Pre-trained AlexNet + SVM	98.38% ACC, 98.71% SV and 98.04% SP
Akbari and Azizi (2023)	4200	2	CNN	97% AC, 98% PR, 83% RC and 90% F-1 score
Nafisah and Muhammad (2024)	1098	2	EfficientNet-B3	99.1% ACC, o99.9% ROC, 98.3% RC, 98.3% PR, 98.3% F1-score, 99.0% SP and 97.2% kappa

*ACC: Accuracy; RC: Recall; PR: Precision; SP: Specificity; SV: Sensitivity;

2.5 Limitation of Existing Studies

Despite the growing implementation of DL models in the field of diagnosis of TB, yet studies on CAD in detection of classification of TB from MS images and CXR images are sparse. Limitations of the few studies reported in the literature include the application of DL models for classification of single type of image dataset which are mostly acquired from publicly accessible repositories. Secondly, majority of existing studies implemented DL models embedded with SoftMax as the model classifier. Thirdly, only handful of studies reported the use of more than 3 DL models. To tighten these gaps, this study proposed the use of 2 different image datasets which include MS and CXR images. Secondly, this study implemented 5 pre-trained models, which include ResNet-101, ResNet-50, VGG-19, VGG-16 and MobileNet. Thirdly, rather than using SoftMax as the main classifier, 2 classical ML classifiers which include SVM and KNN are used.

CHAPTER III

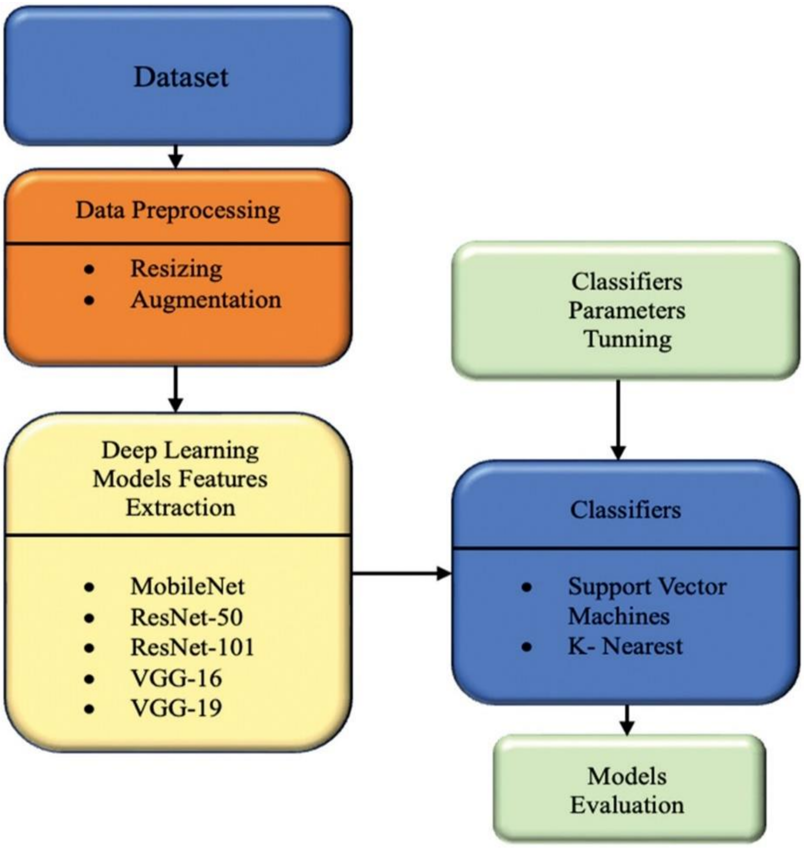
Experimental Set Up

3.0 Introduction

This chapter cover the overall experimental set up. The section is divided into 2 experiments. Experiment A revolves around the use of 5 pre-trained models which include MobileNet, VGG-19, VGG-16, ResNet-101 and ResNet-50 ensemble with either SVM or KNN as classifiers for the binary classification of CXR images into tuberculosis and healthy cases. While experiment B involve the use of same pre-trained model fused with either SVM or KNN as classifiers for the binary classification of MS images into tuberculosis and healthy cases The overall experimental set-up is illustrated in Figure 9. The chapter also covers performance evaluation of ensemble models based on evaluation metrics.

Figure 9

Flow chart of Experimental Set-up



3.1 Datasets

Training of DL networks require large volume of training set in order to enable the model to learn underlying features and patterns and perform efficiently in prediction or classification of unseen, reserve or test set. Unlike developing model from scratch that is full appetite for millions of data points, we employ pre-trained models, which are already trained using millions of images and thus can performed efficiently using small dataset. This study employs 2 distinct datasets which include CXR images and MS images.

3.1.1 CXR Images

In terms of the curation of the CXR images, the study acquired 2 dataset which include the Tawsifurrahman TB dataset and Tuberculosis (TB) Chest X-ray Cleaned Database which are publicly accessible on Kaggle domain. The total dataset used in this study comprises of 7000 images. The Tawsifurrahman TB contain 1500 positive and 3500 negatives. The overall weight of the folder is 802mb. The dataset can be access through this link: https://www.kaggle.com/datasets/legrande/tbdata. TB CXR Cleaned Database comprises of thousands of images. However, we only downloaded the positive cases which amount to 2650 images. In order to balance the dataset and avoid overfitting we selected 2000 images which sum it up to 3500 images. The overall weight of the entire dataset is 9.17GB. The dataset can be access through this link:
https://www.kaggle.com/datasets/scipygaurav/tuberculosis-tb-chest-x-ray-cleaned-database. Both positive and negative samples are presented in Figure 10 while the summary of the dataset is summarized in Table 3.

Figure 10

Left: Tuberculosis. Right: Healthy/Normal

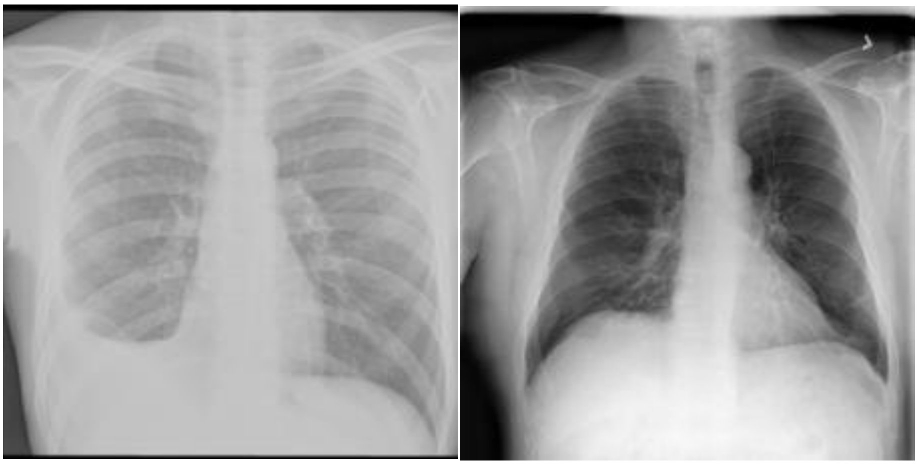


Table 3

CXR Images

Classes	Number of images
Positive	3500
Negative	3500

3.1.2 MS Images

In terms MS images, the study curated the Tuberculosis Image Dataset. The dataset which weighs approximately 480mb comprises of 1893 images in which 1363 images are positive and 530 images are negative. The dataset can be access through this link: https://www.kaggle.com/saife245/tuberculosis-image-datasets. Both positive and negative samples are presented in Figure 11 while the summary of the dataset is summarized in Table 4.

Figure 11

Microscopic slide image. Left: Positive. Right: Negative tuberculosis

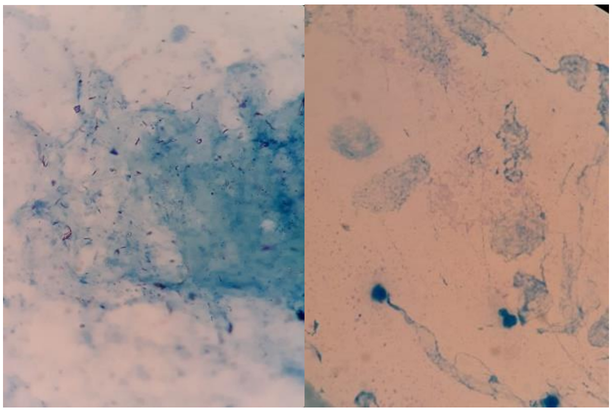


Table 4

MS Images

Classes	Number of images
Positive	1363
Negative	530

3.2 Data split and Data processing

Data split also known as data partition is a widespread practice in ML where dataset are categorized into 2 partitioning groups (which include training and testing) and 3 partitioning groups (which include training, testing and validation). Splitting dataset play a key role in the learning rate and efficiency of the model. Several data scientists have proposed partitioning. The lowest data partitioning includes 50:50 ratio followed by 60: 40 for training and testing respectively. However, majority of the studies in the literature reported the use of either 70:30, 75:25 or 80:20 for training and testing, respectively. In terms of 3 partitioning set, the most common data split includes 70:15:15 or 80:10:10 for training, testing and validation respectively (Joseph and Vakayil, 2022; Medyakov et al., 2023). However, in this study, we opted for 80:20 split for training and testing respectively in order to allow the models learn from large quantity of the dataset. The data split for both CXR and MS images are presented in table 5. As a result of noise, image pre-processing techniques were implemented. The important portion of the image is cropped and processed by removing the unwanted region.

Table 5

Data split for CXR and MS images

Dataset	Label	Training	Testing
MS Images	Positive	954	409
	Negative	371	159
CXR Images	Positive	2450	1050
	Negative	2450	1050

3.3 Pre-trained Models

Pre-trained models also known as TL models are the most widely models used in real-life applications and in the literature. Pre-trained model can be described as ML or DL model that has been trained using large amount of dataset, which can be fine-tuned for solving similar or related task. Pre-trained models offer benefits which include the ability to transfer the experience and knowledge of existing models for solving issues related with scarcity of dataset. Secondly, pre-trained models have been attributed to time and resource savings. Thirdly, pre-trained models produce high performances compare to models developed from scratch. There are different forms of pre-trained models which include image classification and segmentation, object detection, NLPs (Marcelino, 2018; Han et al., 2021).

Pre-trained CNNs are used in computer vision task via object detection, image recognition and classification. Thus, the advent of TL models combined with the growing of Big Biomedical Data is changing the landscape of medical diagnostic into a more reliable, accurate, smart system. Several pre-trained models have been developed and implemented for detection of diseases from medical data such as EHR, wearable signals, biosensors, and medical images (Yu et al., 2022; Wang et al., 2023; Bi et al., 2024).

Thus, in this thesis, 5 pre-trained models were selected which include ResNet-101, MobileNet, ResNet-50, VGG-19 and VGG-16. MobileNet is selected due to it lightweight and application in mobile devices. The ResNets variants (ResNet-50 and ResNet-101) were selected due to their excellence performance in computer vision tasks and depth of the models (several layers). Lastly, the VGG variants were selected as baseline for comparison between shallow model (less depth) and deeper models.

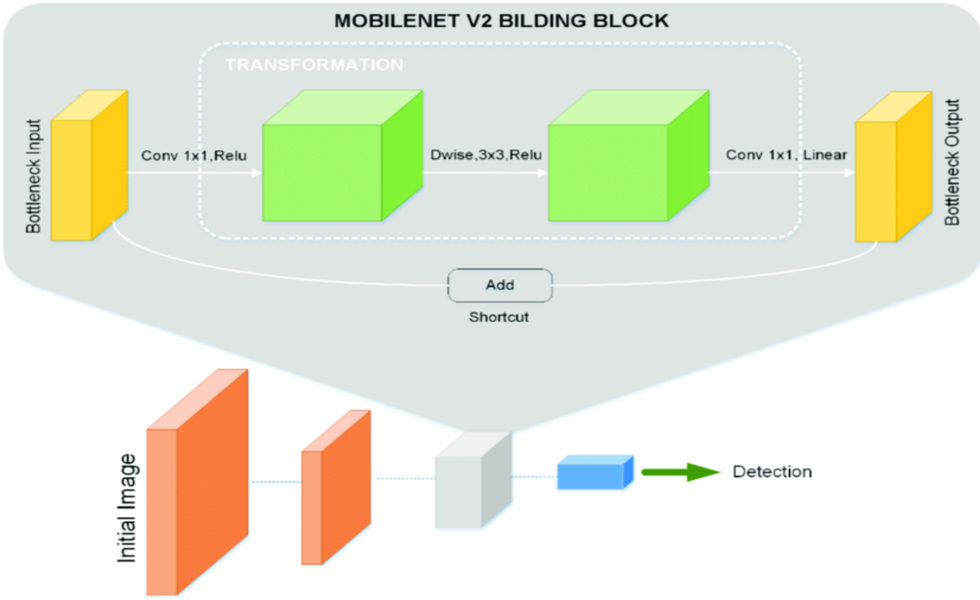
3.3.1 MobileNet

MobileNet has emerged as one of the most reliable and efficient DL model for computer vision applications. The model which is developed by Howard et al., 2017 achieved 70.6% accuracy on the ImageNet dataset. MobileNet as lightweight CNN is developed by Google researchers for mobile applications due to its striking ability and balance between size and performance which makes ideal for resource-constrained devices. Among different versions, MobileNetv2 has become the most outstanding MobileNet architecture employ by scientists for image classification tasks.

The architecture of MobileNetv2 comprises of several key features which play a significant role in maintaining the network's efficiency and minimizing computational complexity. Some of these features include inverted residuals, depth wise separable convolution, squeeze-and-excitation (SE) blocks, bottleneck design and linear bottlenecks as shown in **Figure 12**. The network operates using several processes which include convolution, striding, batch normalization, ReLu as activation function and SoftMax as the output classifier. The network employs a Depth wise separable convolution (3x3 convolution) to minimize computational cost. The technique can be divided into layers which include depth wise convolution and pointwise convolution. While inverted residuals in the other hand are employ in order to improve the model's performance (Talahua et al., 2021).

Figure 12

MoblileNetv2 Architecture (Talahua et al., 2021)



The introduction of bottleneck (1x1 convolution) contributes to the expansion of the network which allows the network to capture more complex features and enhance its representation power. While SE Blocks are included in the model design in order intensify its feature representation abilities. Lastly, linear bottleneck are introduced into the network to solve the problem of information loss. Thus, by introducing linear activation in place of non-linear activations, the model enhance its ability to capture fine-grained details and preserves more information (Talahua et al., 2021).

MobileNetv2 offer advantages over contemporary networks which include it lightweight architecture which enables the deployment of the model in embedded and mobile devices with limited computational resources. Another advantage of the model includes it competitive performance in comparison with other deeper and more computational models. Moreover, the network is extremely fast due to its small size which makes it appropriate for real-time applications (Howard et al., 2017; Talahua et al., 2021).

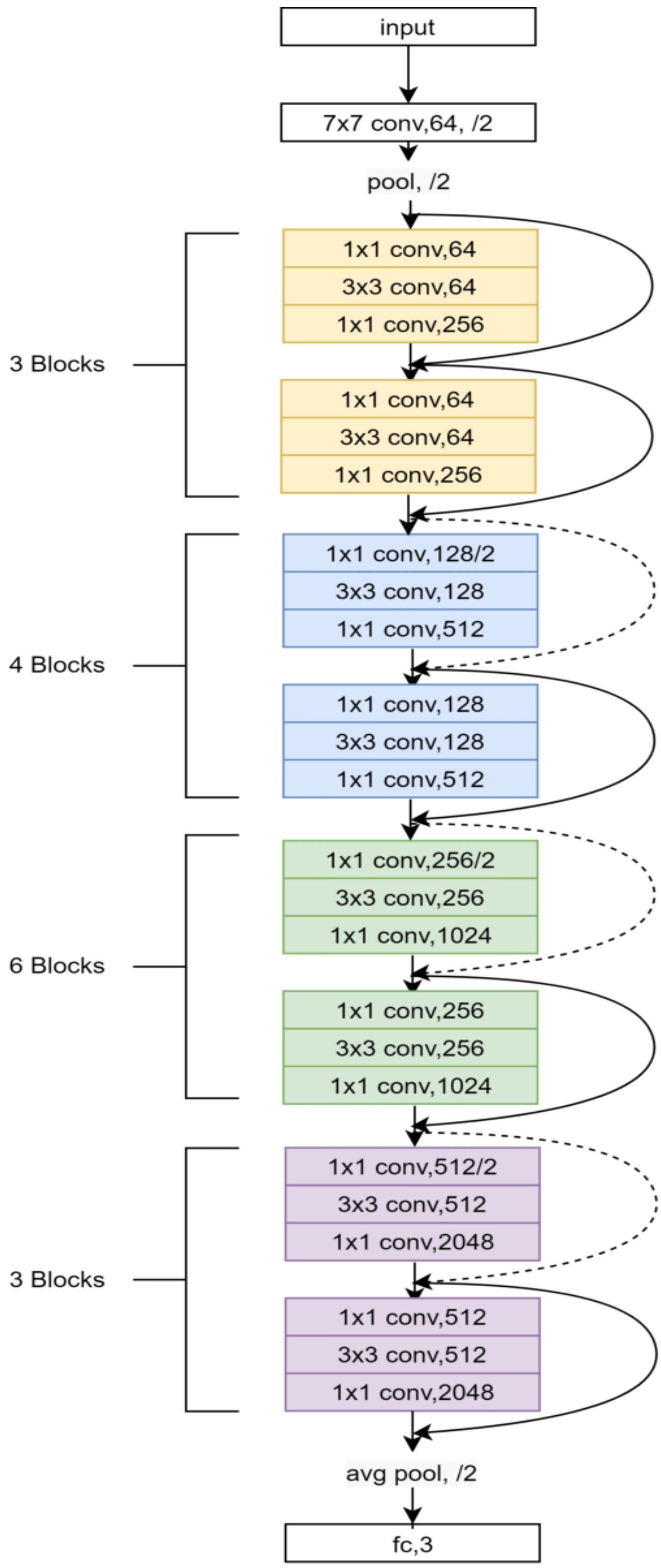
3.3.2 Residual Networks (ResNets)

The notion that "the deeper the better" was the basis behind the development of residual networks (ResNet). Scientists believed that the more layers a network has the better the performance in terms of accuracy. However, this was not the case as increasing layers result in degrading result as models suffer from vanishing gradient or started to lose their generalization capability. To address these issues, Kaiming He and colleagues who are working at Microsoft Research Asia developed ResNet in 2015 by introducing residual/ identity mapping (block). The function of this block is to skip convolutional operations that exists between 2 ReLu activation functions. This skipping process enable the model to learn the residual function rather than directly learning the underlying mapping which can result in more effective learning and increase in performance (He et al., 2016; Targ et al., 2016; Shin et al., 2016).

The network became the best performing model in the ILSVRC 2015 classification competition with a top 5 mistake rate of 3.57%. ResNet are deep CNN architecture that are used for computer vision applications. The model accepts images with an input size of 224 x 224 x 3. The network has several versions based on number layers which include ResNet-18, ResNet-34, ResNet-50, ResNet-101 and ultra-deep ResNet-152. For example, ResNet-50 is one of the most common model uses for image classification tasks, the model is designed with 50 layers, 16 residual blocks where each blocks comprises of several operations which include residual connections, convolution layers, batch normalization, ReLu activation, pooling layers. The rest of the architecture include flattening, global average pooling, fully connected layer, and the use of SoftMax as the model classifier. The first step involves 7x7 convolution followed by pooling. The subsequent steps include 8 convolutional blocks (1x1, 3x3 and 1x1) separated by pooling layers followed by average pooling, fully connected layer and classification using SoftMax as shown in **Figure 13**. ResNet-101 comprises of 33 residual blocks and 96 convolutional layers, 1 max pooling, 1 average pooling layer followed by fully connected layer and the use of SoftMax for classification (He et al., 2016; Targ et al., 2016).

Figure 13

Architecture ResNet-50



3.3.3 VGGNets

Since the development of AlexNet which achieved significant result in the 2012 ILSVRC competition, scientists have been attempting to increase the depth of

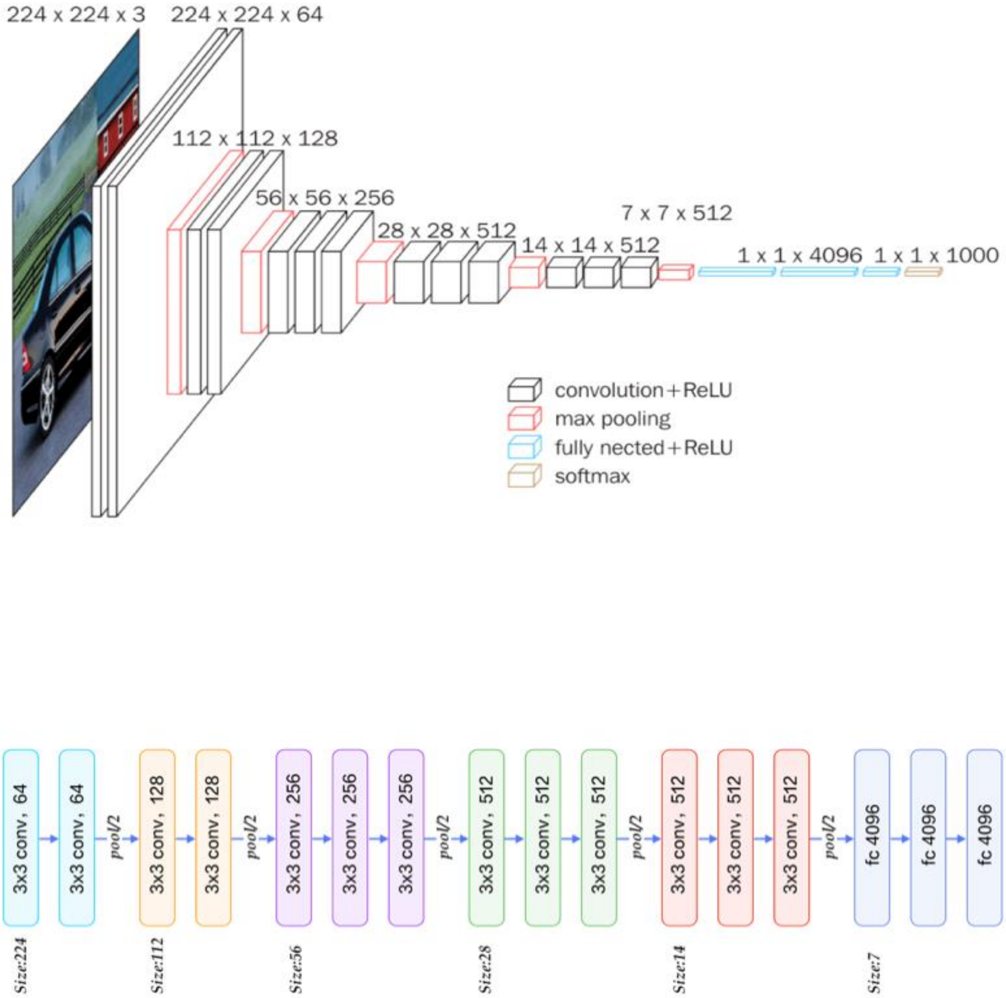
DL models with the objective of increasing performance. Visual Geometry Group network (VGGNET) is developed by scientists known as A. Zisserman and K.

Simonyan from who were working at Oxford University. The DL model is employed by scientists for image classification task. The main contribution of VGGNET compared to AlexNet is the replacement of large kernel filters with 3x3 filters. The model attained top-5 test accuracy of 92.7% using the ImageNet dataset and runner up in the ILSVRC 2014 competition. The model accepts image input with size of 224x224x3 (Simonyan and Zisserman, 2014; Jaderberg et al., 2016).

There are 2 variants: VGG16 and VGG19. As the name implies, VGG16 is made of 16 layers. The first layer is the input layer followed by 6 blocks. The first block consists of 2 convolutional layers with 3x3 filter and 2x2 max pooling. The second block comprises same convolution and max pooling layers. The third, fourth and fifth block are made of 3 convolutional layers (3x3) followed by 2x2 pooling layer each. The final block comprises of 3 fully connected layers in which the last layer use SoftMax for classification. In summary, VGG16 comprises 13 convolutional layer and 3 fully connected layers as shown in **Figure 14**. VGG19 on the other hand is an extension of VGG16 with extra block which contain 3 layers (3 convolutional layers and 1 pooling layer).

Figure 14

Architecture of VGG16 (Paras, 2020).



3.4 Machine Learning Classifiers

Classifiers are ML algorithms that are used for classification tasks. These classifiers are fed with labeled textual data, numerical data, an image, or even an audio with the sim of categorizing the dataset into predefined groups or classes. Initially, the classifiers are trained using large chunk of dataset where they identify categories and classify subsequent data accordingly. Therefore, ML classification algorithms or classifiers basically identify and recognize features in the training data and use the learned patterns or features to classify data in the future or unseen or reserve dataset. There are several classifiers use directly or merged with DL models for classification of dataset into categories. Some of the most common ML classifiers include KNN, SVM, Logistic regression (LR), DT, Quadratic Discriminant Analysis (QDA), Linear Discriminant Analysis (LDA), Naive Bayes (NB), and ANNs etc. (Abro et al., 2021; Burkart and Huber, 2021; Shetty et al., 2022).

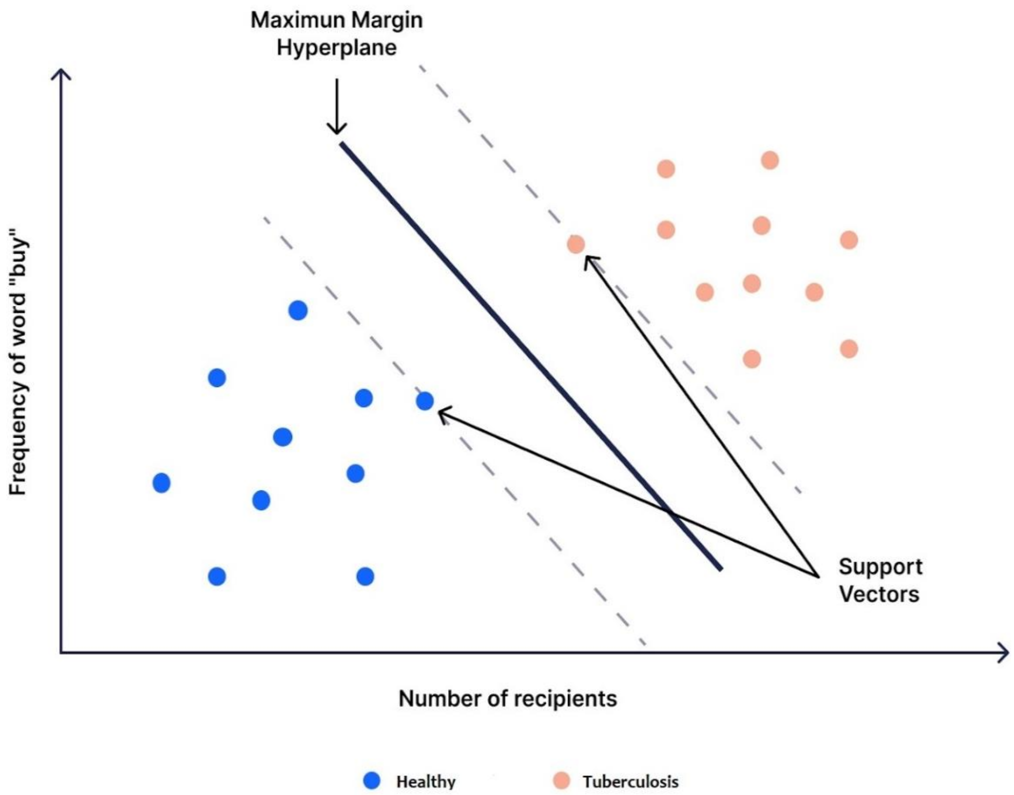
3.4.1 Support Vector Machines (SVM)

SVM is a classical supervised ML algorithms that is employ for classification of data mostly into 2 classes as well regression. The ML algorithms were developed in 1963 by 2 scientists namely Vladimir N. Vapnik and Alexey Ya Chervonenkis. The fundamental mechanism behind SVM revolves around the use of hyperplane which discrete the data point into 2 boundaries. In another word, SVM function by discovering the best hyperplane that splits data points into various denominations. The hyperplane attempts to locate the boundary between the nearby points of different classes. The dimension of the hyperplane varies according to the number of features. For instance, there will only be single hyperplane for 2 classes (as shown in **Figure 15**) while for 3 classes it becomes 2D plane. Some of the features of SVM include support vectors, hyperplane, and margin (Suthaharan and Suthaharan, 2016; Pisner and Schnyer, 2020).

The main advantages of SVM over other traditional supervised ML and neural networks revolves around their excellent performance and higher speed when it comes to classification of limited amount of data. Some of the disadvantages of the classifier or algorithms such as non-linear SVM include sensitivity to outliers and noise, requirement of large memory and difficulty in terms of addressing problem related to big data. SVM are predominantly use in various fields such as computer vision (image analysis and pattern recognition) and NLP (Suthaharan and Suthaharan, 2016; Pisner and Schnyer, 2020).

Figure 15

Binary Classification using Support Vector Machine (SVM)



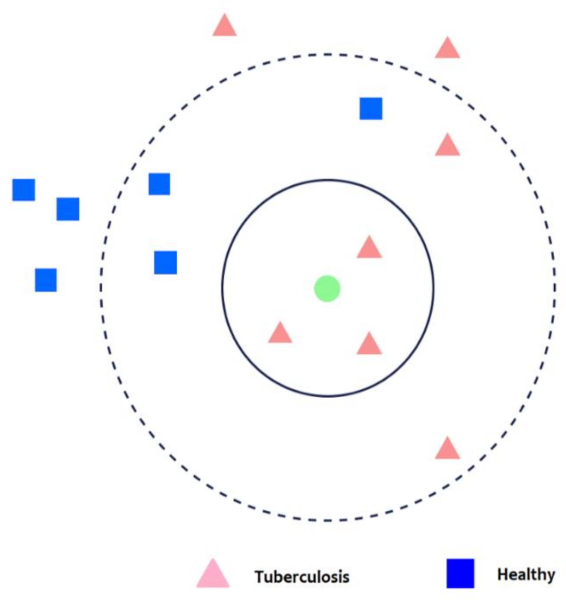
3.4.2 K-Nearest Neighbor (KNN)

There are several types of ML algorithms that are used for solving problems that often require manual labor and time. Among these algorithms, KNN is characterized as one of the most widely used, versatile and intuitive ML algorithms. The ML algorithms are often tagged as "lazy learner". KNN was first presented by Evelyn fix and Joseph Hodges in 1951. KNN is characterized as non-parametric because unlike other ML algorithms that learn through vigorous training to understand underlying features or patterns within a dataset in order to predict or classify reserve or unseen dataset, KNN operate based on memorizing the training datasets (Peterson, 2009; Zhang et al., 2016). Even though the ML algorithms can be used for solving regression and classification tasks, however, it is widely adopted for classification of data points. The only major difference between the use of KNN for classification and regression is that in classification, discrete data are use while in regression, continuous data are used. It is widely used in banking and finance, customer care, medicine and agriculture (Kramer and Kramer, 2013).

The basic working principle of KNN is based on using proximity to make predictions or classification of categories of individual data points. In another word, the algorithms consider that similar data point stay close to each other. As a supervised ML algorithm, the data of interest are labeled or assigned a “majority voting” (Cunningham and Delany, 2021). The algorithms use Euclidean distance to measure the distance between a data point and other point. These measurements form decision boundaries which separate query points into different classes. The k value in KNN algorithm signifies how many neighbors will be examine in order to ascertain the class of a specific query point. Similarly, the algorithms determine the value of k” which present the number of neighbors it will check before assigning a value to any new observations. The value of K starts from 1 and above. For example if k = 1, the algorithms only cross check the nearest neighbor for each prediction) against the sum of number of data points in the dataset (Mucherino et al., 2009; Zhang, 2016). The use of KNN for the binary classification of TB is illustrated in **Figure 16**.

Figure 16

Binary classification Using K-Nearest Neighbor (KNN)

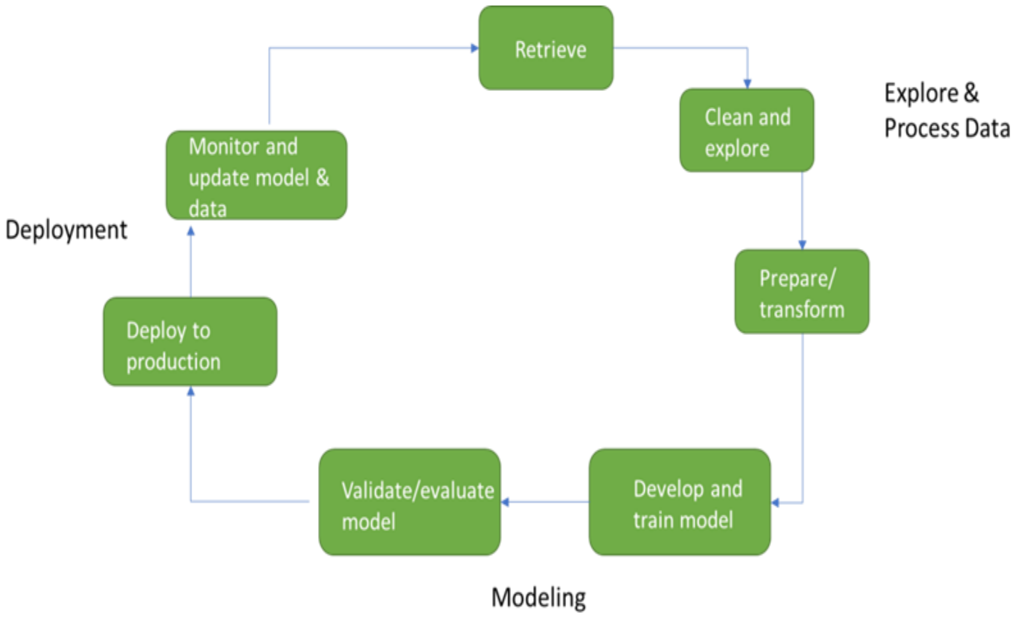


3.5 Training

The training of ML models requires enormous volume of data in order to achieved significant result. The first step in ML workflow revolves around the collection of data from private or public domains such as online repositories. The acquired dataset is further pre-processed based on normalization, sorting, cleaning, labelling, resizing etc. The pre-processed data is subsequently fed into a ML model which learn underlying features and pattern and use classifier such as SoftMax to classify the images into categories or classes. The next stage involves performance evaluation of models using evaluation metric such as Accuracy, AUC, recall, precision etc. Once the model is evaluated and the result are promising, the model can be deployed in clinical settings or website for real-time detection as illustrated in Figure 17.

Figure 17

Machine Learning Workflow



In this study, the pre-trained model's ensemble with ML classifiers are trained and evaluated using a computer with the following specifications: Dell inspiron15 with 8GB, core i7, GPU NVidia 1060ti. The dataset is trained using 80% and tested using 20% which also used to assess the performance of the model. Minibatch optimization is a gradient descent that is employed to optimize the model. The training is conducted based on 20 epochs with 0.0001 as learning rate. The models were tuned based on the parameters presented in Table 6.

Table 6

Parameters used for fine-tuning the models

SVM	
Kernel	RBF
C (Regularization)	0.1
Gamma	Auto
KNN	
Number of Neighbours	20
Weights	Uniform
Metric	Euclidean

3.6 Evaluation metrics and Confusion Matrix

Assessment of models trained from scratch and pre-trained models are crucial for understanding the model efficiency in testing real-world data or reserve dataset. There are several evaluation metrics use for classification algorithms such as accuracy, sensitivity, specificity, precision, recall, AUC or ROC curve, F1-score, Younden Index, Jaccard Index etc. in data analysis, these evaluation metrics are used to test model performance, however, the concepts are very similar but yet different (Liu et al., 2014; Hossin and Sulaiman, 2015; Gong, 2021; Vujović, 2021).

In binary classification which are characterized by 2 classes such as true and false, positive and negative, right and wrong, normal and healthy etc. we can evaluate outcomes of ML models based on true, correct predictions which include True positive (TP) and True negative (TN) and model errors which include False positive (FP) and False Negative (FN) as shown in **Figure18.**

TP: can be characterized as positive cases that are accurately classify as positive.

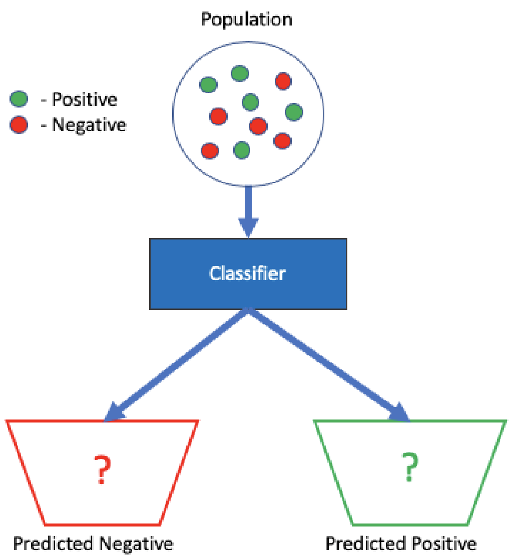
TN: can be characterized as negative cases that are accurately classify as negative cases.

FP: are characterized as negative cases that incorrectly classify as positive. For example FP occur when the model mistakenly classify healthy patients as one suffering from TB.

FN: is characterized as positive cases that are incorrectly classify as negative cases. For example, FN occur when the model mistakenly classifies a patient that is suffering from TB as healthy.

Figure 18

Confusion Matrix



		Real	
		Positive	Negative
Predicted	Positive	True Positive (tp)	False Positive (fp)
	Negative	False Negative (fn)	True Negative (tn)

3.6.1 Accuracy

Accuracy is the most common evaluation metrics that is used to assess the performance of ML models. Accuracy measures how ML models correctly classify or predict outcomes. The metrics is calculated based on the proportion of the number of correct prediction/classifications over the total amount of predictions/classification. Essentially, accuracy answer the question of how often ML is accurate. Accuracy is measure based on percentage or on a scale between 0 and 1. With 1 showing 100% accuracy (Hossin and Sulaiman, 2015). Accuracy can be represented mathematically as:

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

$$Accuracy = \frac{Number\ of\ correctly\ classified\ \ images}{Total\ number\ of\ images}$$

One of the issues related with accuracy is called the accuracy paradox which explain the downside of relying on accuracy alone as the evaluation of the model generalizability. The main issue is that accuracy treats all classes as equally significant and looks at all correct predictions while ignoring the issue of imbalance dataset especially in real-world applications. Therefore, is an ideal evaluation metrics when it comes to balance dataset (Hossin and Sulaiman, 2015; Gong, 2021).

3.6.2 Precision

Precision is another evaluation metrics that is use in data science to evaluate among all the cases that are classify as positive Gong, 2021; Vujović, 2021). How many are truly positive? Mathematically precision can be represented as:

$$\boldsymbol{precision} = \frac{tp}{tp + fp}$$

3.6.3 Recall

Recall is another common metrics that address the question: Out of all the positive cases, how many are predicted as positive? Gong, 2021; Vujović, 2021). Precision can be represented mathematically as:

$$\boldsymbol{recall} = \frac{tp}{tp + fn}$$

3.6.4 Sensitivity

Sensitivity also regarded as “true positive rate (TPR) or the number of positive (tuberculosis) cases that are predicted or classified as tuberculosis. Consequently, based on probability, the sum of true positive (sensitivity) and false positive is equal to 1 (Vujović, 2021). Sensitivity can be presented mathematically as:

$$\boldsymbol{sensitivity} = \frac{tp}{tp + fn}$$

3.6.5 Specificity

Specificity is another common evaluation metrics that seeks to answer the question: how many patients are actually or correctly predicted or classified as negative cases (Gong, 2021; Vujović, 2021). Specificity can be represented mathematically as:

$$\boldsymbol{specificity} = \frac{tn}{tn + fp}$$

3.6.6 F1-Score

F1 score is another conventional evaluation metrics that sum of both precision and recalls. It shares close similarities with accuracy in answering the question on how well model classify both positive and negative cases (Gong, 2021; Vujović, 2021).

$$F1 = \frac{Precison\ X\ Recall}{Precison + Recall}\ Or\ \frac{2TP}{(2TP+FP+FN)}$$

3.6.7 Area under the Curve (AUC)

AUC abbreviated as Area under the Receiver Operating Characteristic (ROC) curve present the cumulative measure of performance across all classification thresholds. The metric is widely used to evaluate the capability of model to discriminate between binary classes; positive and negative. The value of AUC ranges between 0 and 1. A model with an AUC of 1 indicate 100% correct predictions while 0 indicates 0% prediction. Unlike accuracy that measure correct prediction over total number of predictions, AUC provide information on the model's prediction irrespective of what classification threshold is selected. Mathematically, ROC can be represented as the ratio of true positive rate (TPR) vs the false positive rate (FPR) at different classification thresholds (Liu et al., 2014; Zhang et al., 2015).

3.6.8 Youden Index

Youden Index is another performance metrics that is used in conjunction with AUC. YI also known as Youden's J statistic is a value that maximizes the specificity and sensitivity of any continuous variables and thereby enabling selection of suitable cut point for dichotomization (Luo and Xiong, C 2013).

CHAPTER IV
Performance Evaluation

4.0 Introduction

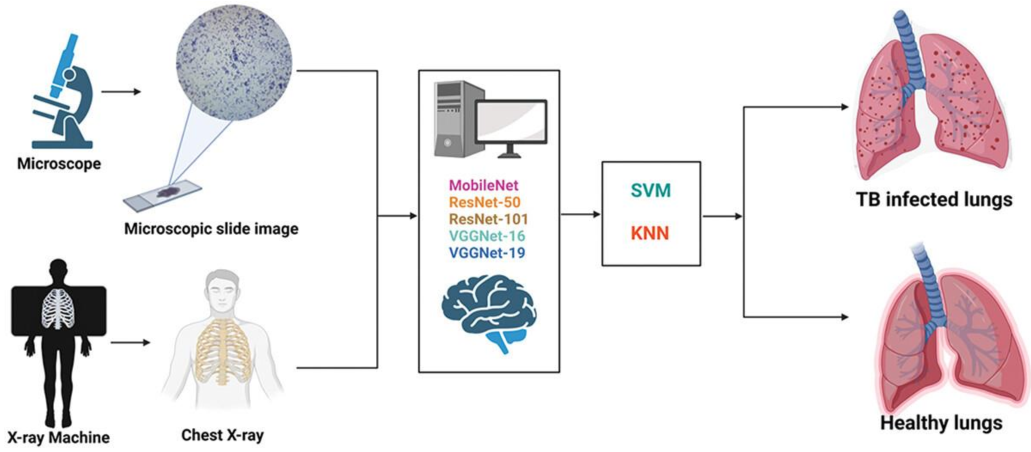
This chapter present the result achieved based on evaluating the ensemble DL and ML models using reserve dataset (i.e., 20%) of both MS and CXR images. The chapter also discussed the overall result achieved by comparing performance evaluation of ensemble models based on (1) fusion of DL models with SVM and KNN (2) performance evaluation of models on MS and CXR images and (3) comparison between best performing models with existing studies.

4.1 Performance evaluation of pre-trained Models

The expansion and exponential growth of CAD which integrate DL models and medical imaging continue to transform medical diagnosis into a more reliable, efficient, accurate, time-saving, affordable systems. The increase in the volume of medical data which include textual data from digital devices, audio data, and images obtained from microscope, endoscopy, mammography, MRI, CT scans, X-ray, PET, SPECT etc. continue to fuel ML and DL models. These images are currently employed or use as input for AI-based models for accurate classification of diseases and contribute significantly to minimize errors and high workload faced by medical expert. By conforming to these techniques, we employed 5 pre-trained models fused with either KNN or SVM for the binary classification of TB and Healthy cases from both MS and CXR images as shown in Figure 19. To assess the generalizability of the proposed approach, several performance metrics were measured which include Youden Index, AUC score, F1-Score, Precision, specificity, sensitivity are accuracy. The performance evaluation result is presented in tables and figures.

Figure 19

Proposed frameworks



4.2 Performance Evaluation of ensemble models using MS Images

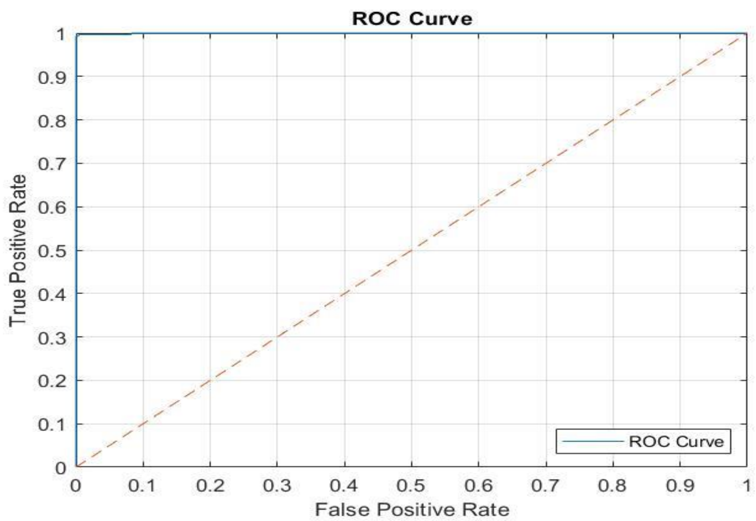
4.2.1 Performance Evaluation of Pre-trained Models fused with SVM using MS Images

MobileNet + SVM

The fusion of MobileNet with SVM resulted in an accuracy of 97.50%, sensitivity of 97.33%, specificity of 97.69%, F1-Score of 97.33%, precision of 97.33%, Youden Index of 95.02% and AUC score 97.51%. The ROC curve for the 2-way screening of MS images using MobileNet + SVM is presented in Figure 20

Figure 20

ROC Curve for the 2-way screening of MS images using MobileNet + SVM

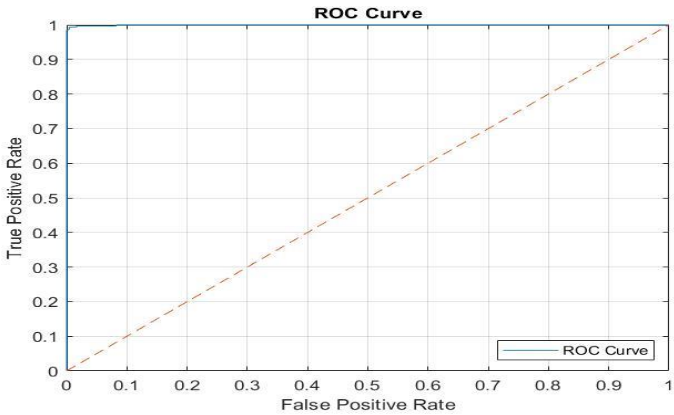


ResNet-50 + SVM

The fusion of ResNet-50 with SVM resulted in an accuracy of 99.40%, sensitivity of 99.50%, specificity of 99.20%, F1-Score of 99.34%, precision of 99.18%, Youden Index of 98.70% and AUC score 99.35%. The ROC curve for the 2-way screening of MS images using ResNet-50 + SVM is presented in Figure 21.

Figure 21

ROC Curve for 2-way screening of MS images using ResNet-50 + SVM

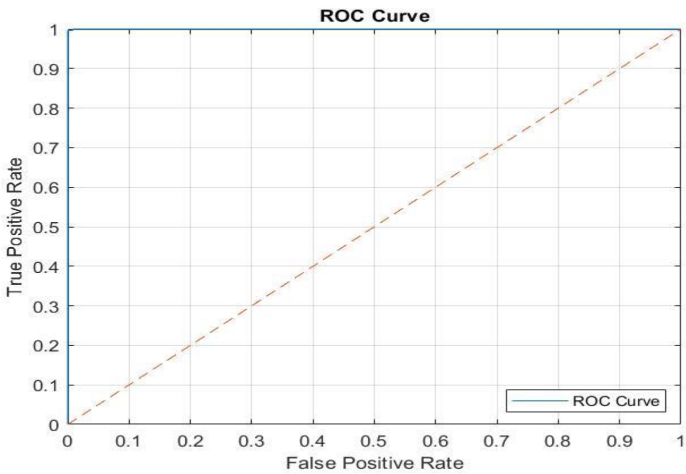


ResNet-101 + SVM

The fusion of ResNet-101 with SVM resulted in an accuracy of 99.60%, sensitivity of 99.50%, specificity of 99.60%, F1-Score of 99.54%, precision of 99.59%, Youden Index of 99.10% and AUC score 99.55%. The ROC curve for the binary classification of MS images using ResNet-101 + SVM is presented in Figure 22.

Figure 22

ROC Curve for the binary classification of MS images using ResNet-101 + SVM

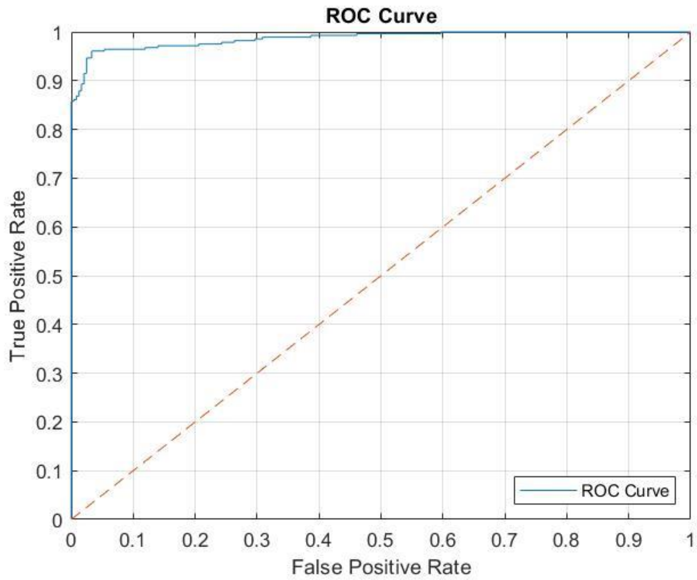


VGG-16 + SVM

The fusion of VGG-16 with SVM yielded in an accuracy of 95.00%, sensitivity of 93.40%, specificity of 96.40%, F1-Score of 94.58%, precision of 95.78%, Youden Index of 89.80% and AUC score 94.90%. The ROC curve for the 2-way screening of MS images using VGG-16+ SVM is presented in Figure 23.

Figure 23

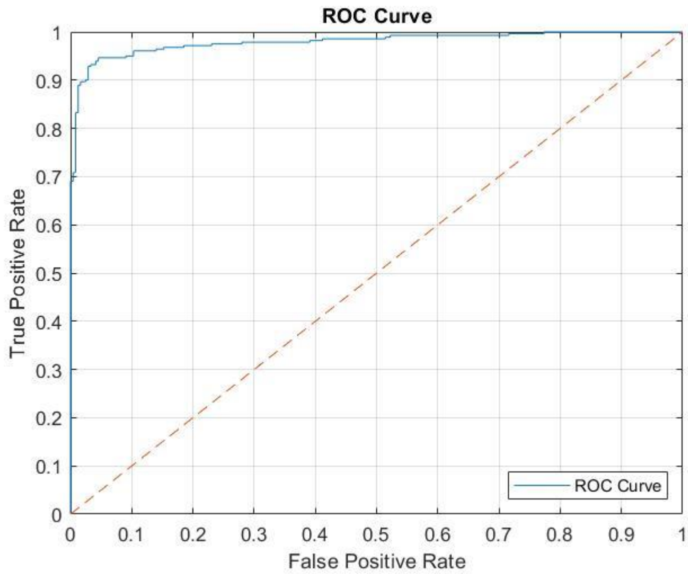
ROC Curve 2-way screening of MS images using VGG-16 + SVM



The fusion of VGG-19 with SVM yielded in an accuracy of 94.80%, sensitivity of 94.60%, specificity of 94.00%, F1-Score of 94.43%, precision of 94.26%, Youden Index of 88.60% and AUC score 94.30%. The ROC curve for 2-way screening of MS images using VGG-19+ SVM is presented in Figure 24.

Figure 24

ROC Curve for the 2-way screening of MS images using VGG-19 + SVM



The overall result achieved by training and testing pre-trained models ensemble with SVM are shown in Table 7 and Figure 30.

Table 7

Performance Evaluation of Pre-trained models fused with SVM on MS Images

Models	Performance Metrics (%)						
	ACC	AUC	F1-Score	PR	SP	SV	YI
MobileNet	97.50	97.51	97.33	97.33	97.69	97.33	95.02
ResNet-50	99.40	99.35	99.34	99.18	99.20	99.50	98.70
ResNet-101	99.60	99.55	99.54	99.59	99.60	99.50	99.10
VGG-16	95.00	94.90	94.58	95.78	96.40	93.40	89.80
VGG-19	94.80	94.30	94.43	94.26	94.00	94.60	88.60

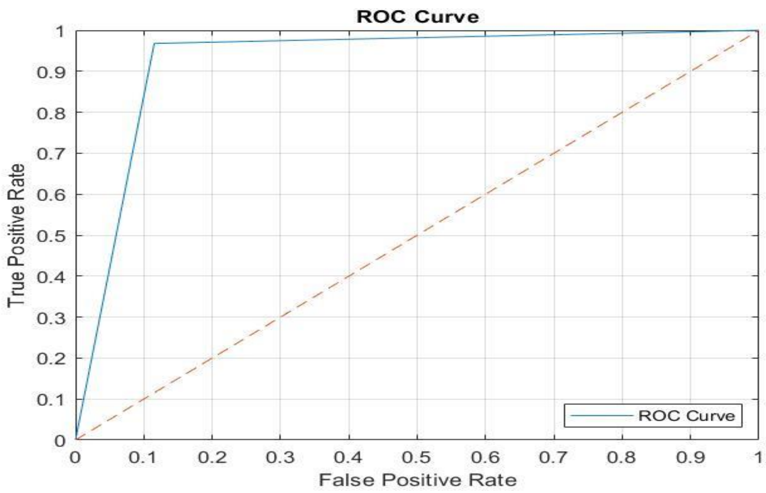
4.2.2 Performance Evaluation of Pre-trained Models fused with KNN using MS Images

MobileNet + KNN

The fusion of MobileNet with KNN yielded in an accuracy of 92.90%, sensitivity of 88.07%, specificity of 97.15%, F1-Score of 92.05%, precision of 96.40%, Youden Index of 85.22% and AUC score 92.61%. The ROC curve for the 2-way screening of MS images using MobileNet + KNN is presented in Figure 25.

Figure 25

ROC Curve for the 2-way screening of MS images using MobileNet + KNN

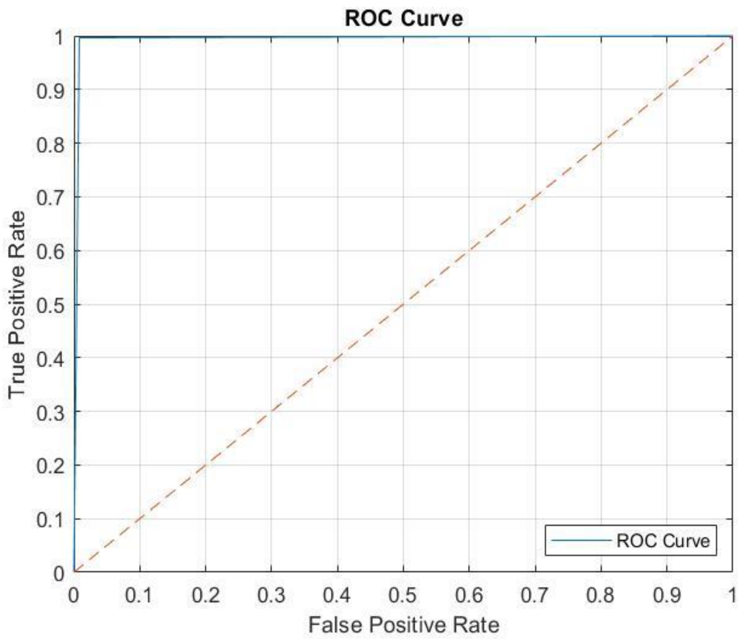


ResNet-50 + KNN

The fusion of ResNet-50 with KNN resulted in an accuracy of 99.43%, sensitivity of 99.64%, specificity of 99.18%, F1-Score of 99.47%, precision of 99.29%, Youden Index of 98.82% and AUC score 99.41%. The ROC curve for the 2-way screening of MS images using ResNet-50 + KNN is presented in Figure 26.

Figure 26

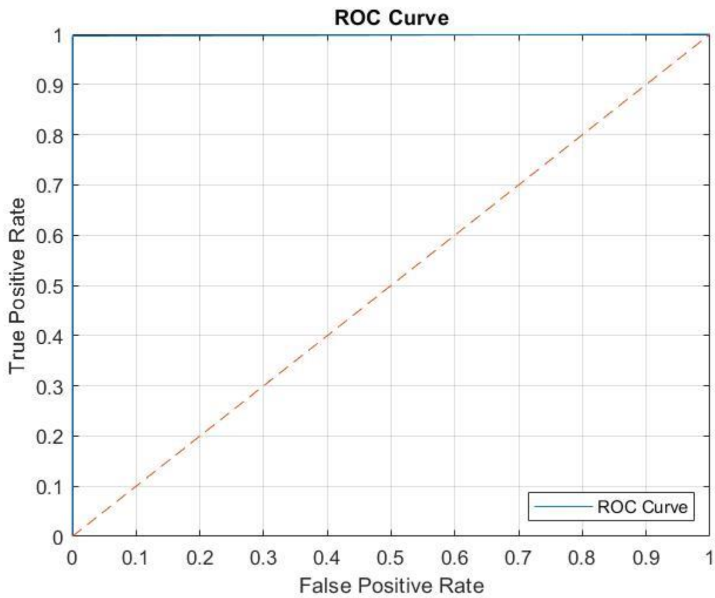
ROC Curve for the 2-way screening of MS images using ResNet-50 + KNN



The fusion of ResNet-101 with KNN resulted in an accuracy of 99.8%, sensitivity of 100%, specificity of 99.60%, F1-Score of 99.79%, precision of 99.59%, Youden Index of 99.60% and AUC score 99.80%. The ROC curve for the 2-way screening of MS images using ResNet-101 + KNN is presented in Figure 27

Figure 27

ROC Curve for the 2-way screening of MS images using ResNet-101 + KNN

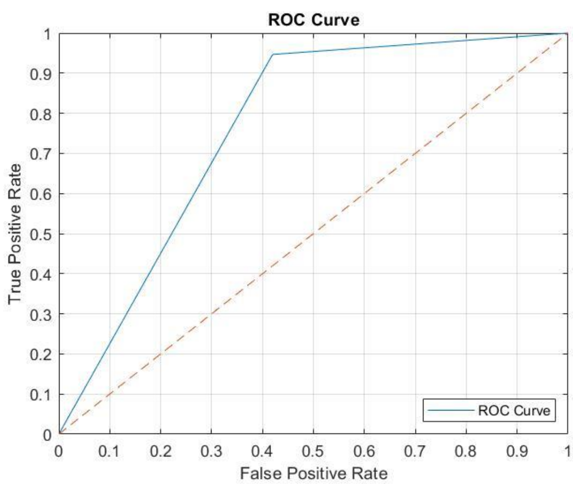


VGG-16 + KNN

The fusion of VGG-16 with KNN yielded in an accuracy of 79.60%, sensitivity of 60.90%, specificity of 95.70%, F1-Score of 73.45%, precision of 92.50%, Youden Index of 56.60 % and AUC score 78.30%. The ROC curve for the 2-way screening of MS images using VGG-16 + KNN is presented in Figure 28.

Figure 28

ROC Curve for the 2-way screening of MS images using VGG-16 + KNN

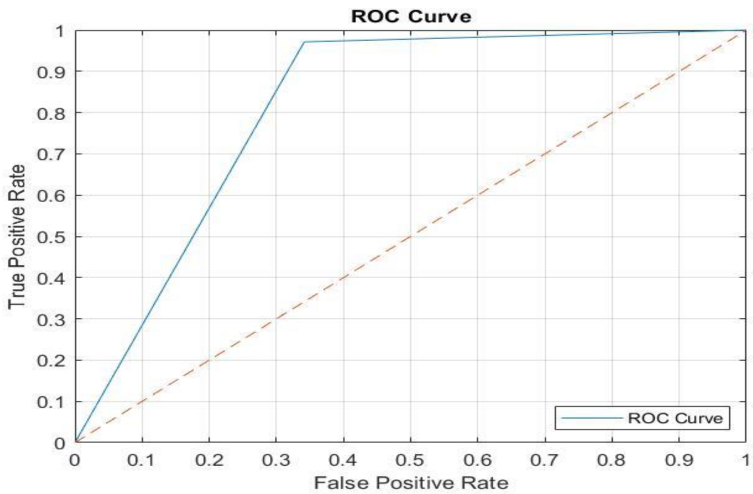


VGG-19 + KNN

The fusion of VGG-19 with KNN resulted displayed an accuracy of 80.90%, sensitivity of 62.15%, specificity of 97.15%, F1-Score of 75.13%, precision of 94.97%, Youden Index of 59.30% and AUC score 79.65%. The ROC curve for the 2-way screening of MS images using VGG-19 + KNN is presented in Figure 29.

Figure 29

ROC Curve for the 2-way screening of MS images using VGG-19 + KNN



The overall result achieved by training and testing pre-trained models ensemble with SVM are shown in Table 8 and Figure 30.

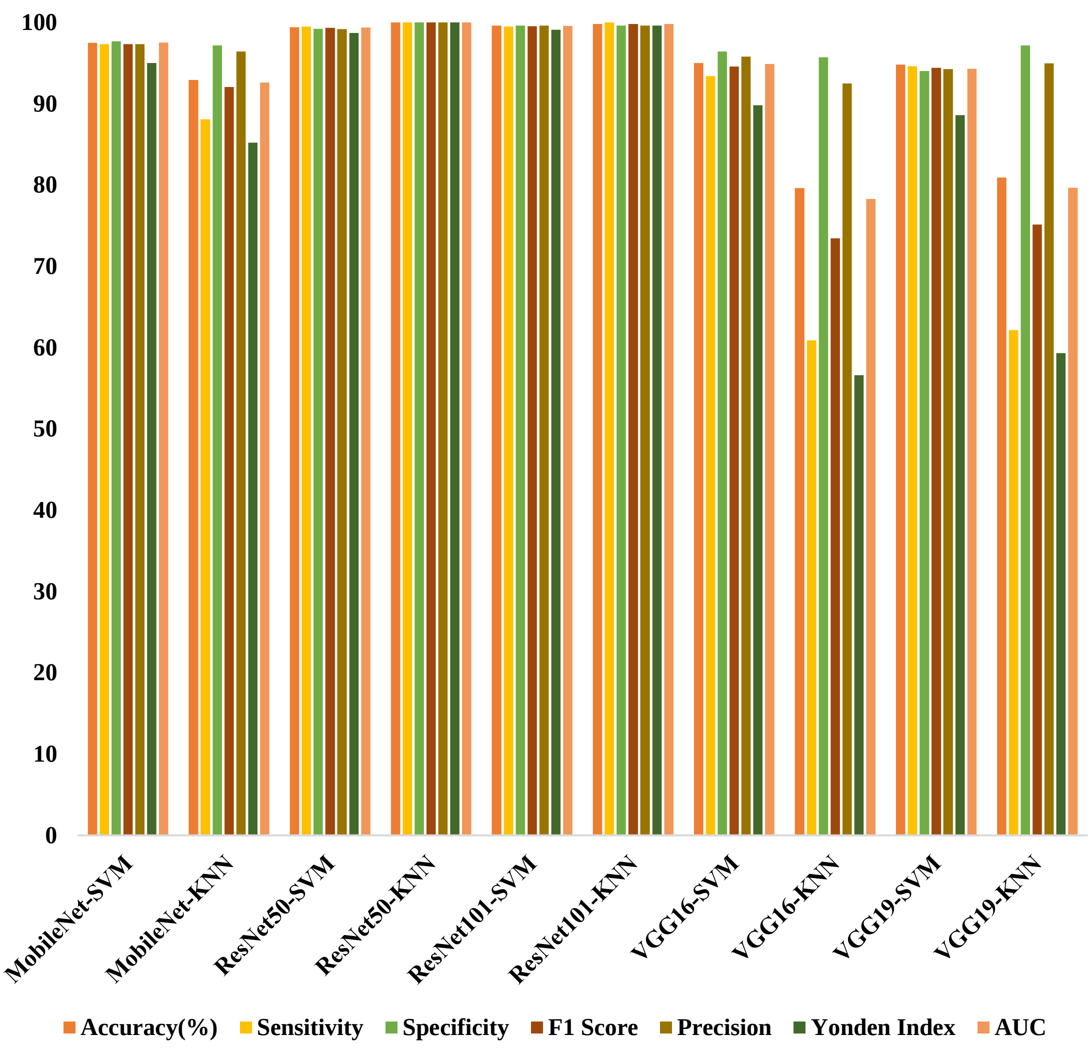
Table 8

Performance Evaluation of Pre-trained models fused with KNN on MS Images

Models	Performance Metrics (%)						
	ACC	AUC	F1-Score	PR	SP	SV	YI
MobileNet	92.90	92.61	92.05	96.40	97.15	88.07	85.22
ResNet-50	99.43	99.41	99.47	99.29	99.18	99.64	98.82
ResNet-101	99.80	99.80	99.79	99.59	99.60	100.00	99.60
VGG-16	79.60	78.30	73.45	92.50	95.70	60.90	56.60
VGG-19	80.90	79.65	75.13	94.97	97.15	62.15	59.30

Figure 30

Performance Evaluation of ensemble models using MS images



4.3 Performance Evaluation of ensemble models using CXR Images

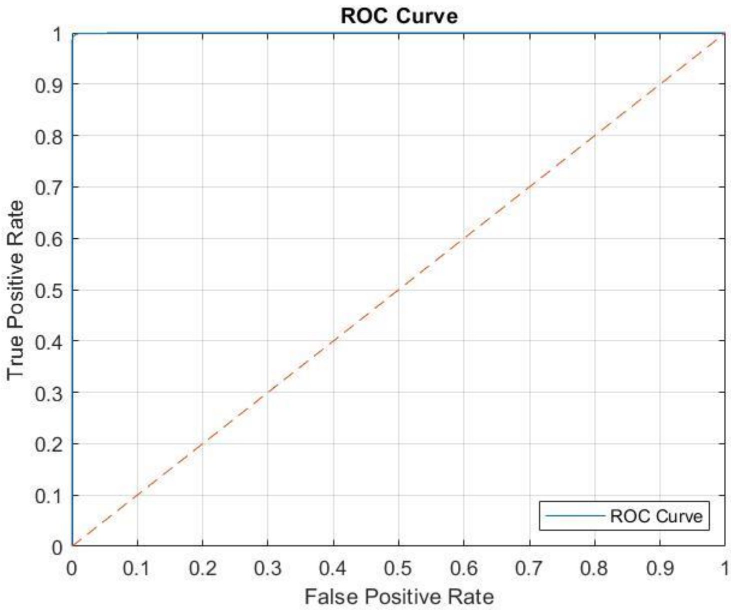
4.3.1 Performance Evaluation of Pre-trained Models fused with SVM using CXR Images

MobileNet + SVM

The fusion of MobileNet with SVM resulted in an accuracy of 99.6%, sensitivity of 99.00%, specificity of 99.50%, F1-Score of 99.36%, precision of 99.71%, Youden Index of 98.50% and AUC score 99.25%. The ROC curve for the 2-way screening of MS images using MobileNet + SVM is presented in Figure 31.

Figure 31

ROC Curve for the 2-way screening of MS images using MobileNet + SVM

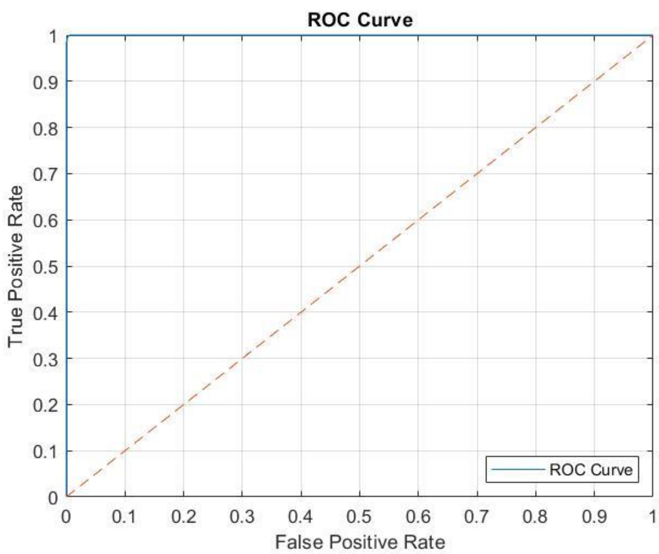


ResNet-50 + SVM

The fusion of ResNet-50 with SVM resulted in an accuracy of 99.40%, sensitivity of 99.43%, specificity of 99.20%, F1-Score of 99.36%, precision of 99.29%, Youden Index of 98.63% and AUC score 98.31%. The ROC curve for the 2-way screening of MS images using ResNet-50 + SVM is presented in Figure 32.

Figure 32

ROC Curve for the 2-way screening of MS images using ResNet-50 + SVM



ResNet-101 + SVM

The fusion of ResNet-101 with SVM resulted in an accuracy of 99.70%, sensitivity of 99.70%, specificity of 99.70%, F1-Score of 99.71%, precision of 99.71%, Youden Index of 99.40% and AUC score 99.70%. The ROC curve for the 2-way screening of MS images using ResNet-101 + SVM is presented in Figure 33.

Figure 33

ROC Curve for the 2-way screening of MS images using ResNet-101 + SVM

VGG-16 + SVM

The fusion of VGG-16 with SVM resulted in an accuracy of 99.60%, sensitivity of 99.40%, specificity of 99.70%, F1-Score of 99.56%, precision of 99.71%, Youden Index of 99.10% and AUC score 99.55%. The ROC curve for the 2-way screening of MS images using VGG16 + SVM is presented in Figure 34.

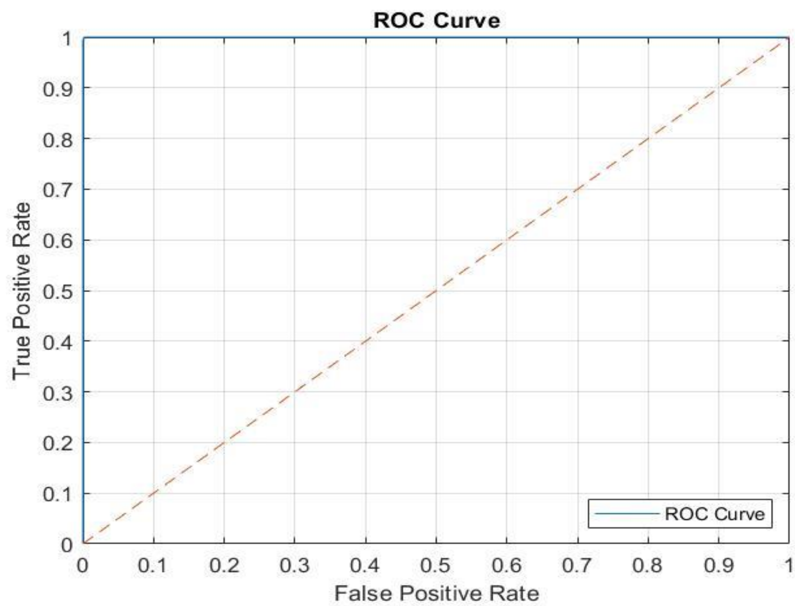


Figure 34

ROC Curve for the 2-way screening of MS images using VGG16 + SVM

VGG-19 + SVM

The fusion of VGG-19 with SVM resulted in an accuracy of 99.60%, sensitivity of 99.50%, specificity of 99.70%, F1-Score of 99.61%, precision of 99.71%, Youden Index of 99.20% and AUC score 99.60%. The ROC curve for the 2-way screening of MS images using ResNet-50 + SVM is presented in Figure 35.

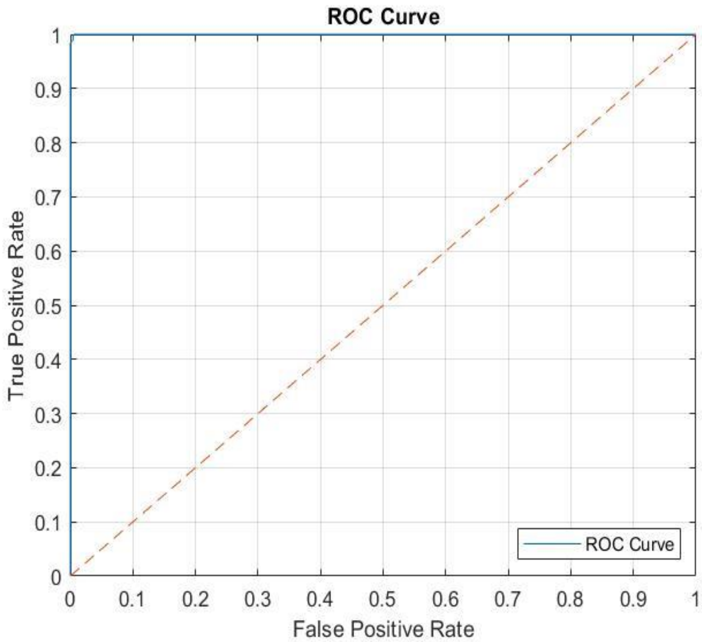
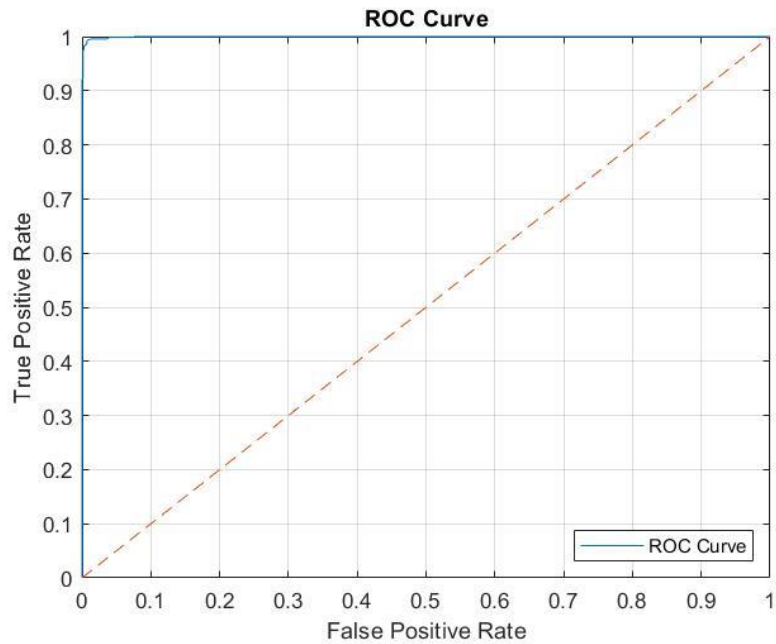


Figure 35

ROC Curve for the 2-way screening of MS images using ResNet-50 + SVM



The overall result achieved by training and testing pre-trained models' ensemble with SVM are shown in Table 9 and Figure 41.

Table 9

Performance Evaluation of Pre-trained models fused with SVM on CXR Images

Models	Performance Metrics (%)						
	ACC	AUC	F1-Score	PR	SP	SV	YI
MobileNet	99.60	99.25	99.36	99.71	99.50	99.00	98.50
ResNet-50	99.40	99.32	99.36	99.29	99.20	99.43	98.63
ResNet-101	99.70	99.70	99.71	99.71	99.70	99.70	99.70
VGG-16	99.60	99.55	99.56	99.71	99.70	99.40	99.10
VGG-19	99.60	99.60	99.61	99.71	99.70	99.50	99.20

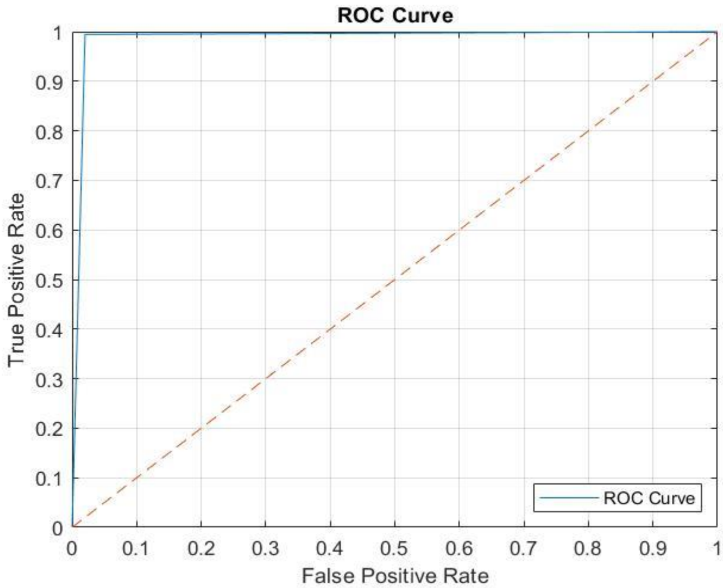
4.3.2 Performance Evaluation of Pre-trained Models fused with KNN using CXR Images

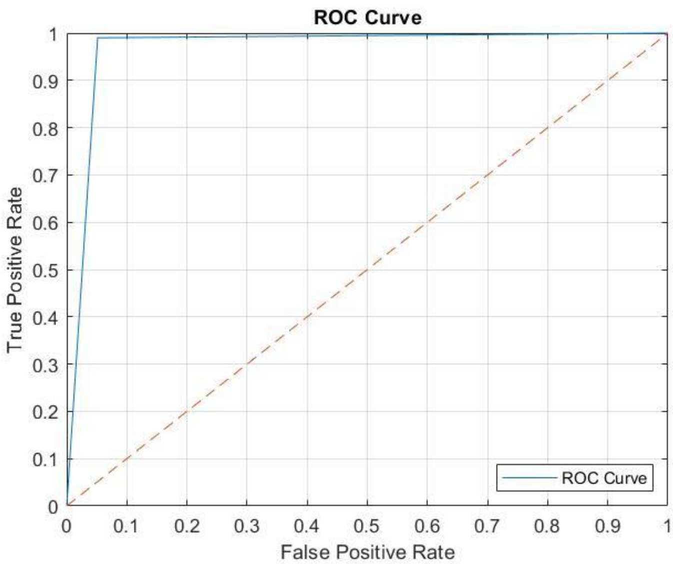
MobileNet + KNN

The fusion of MobileNet with KNN displayed an accuracy of 96.40%, sensitivity of 94.60%, specificity of 98.00%, F1-Score of 96.24%, precision of 97.94%, Youden Index of 92.60% and AUC score 96.30%. The ROC curve for the 2-way screening of CXR images using MobileNet + KNN is presented in Figure 36.

Figure 36

ROC Curve for the 2-way screening of MS images using MobileNet + KNN





ResNet-50 + KNN

The fusion of ResNet-50 with KNN resulted in an accuracy of 98.50%, sensitivity of 98.40%, specificity of 98.50%, F1-Score of 98.48%, precision of 98.57%, Youden Index of 96.90% and AUC score 98.45%. The ROC curve for the 2-way screening of CXR images using ResNet-50 + KNN is presented in Figure 37.

Figure 37

ROC Curve for the 2-way screening of MS images using ResNet-50 + KNN

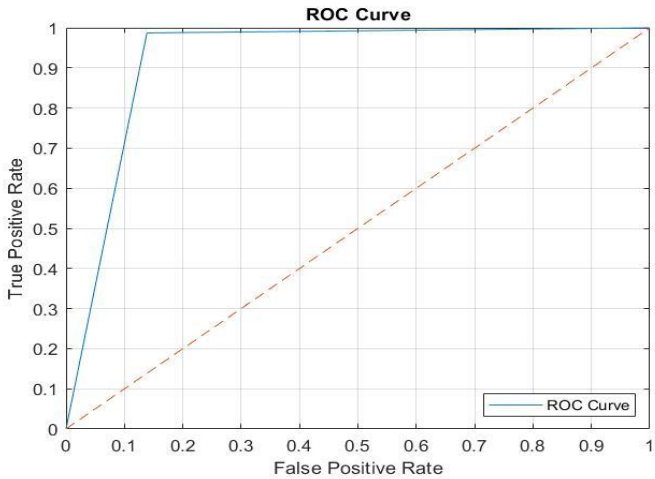
ResNet-101 + KNN

The fusion of ResNet-101 with KNN resulted in an accuracy of 98.10%, sensitivity of 98.00%, specificity of 98.10%, F1-Score of 98.07%, precision of 98.14%, Youden Index of 96.10% and AUC score 98.05%. The ROC curve for the 2-way screening of CXR images using ResNet-101 + KNN is presented in Figure 38.

Figure 38

ROC Curve for the 2-way screening of MS images using ResNet-101 + KNN

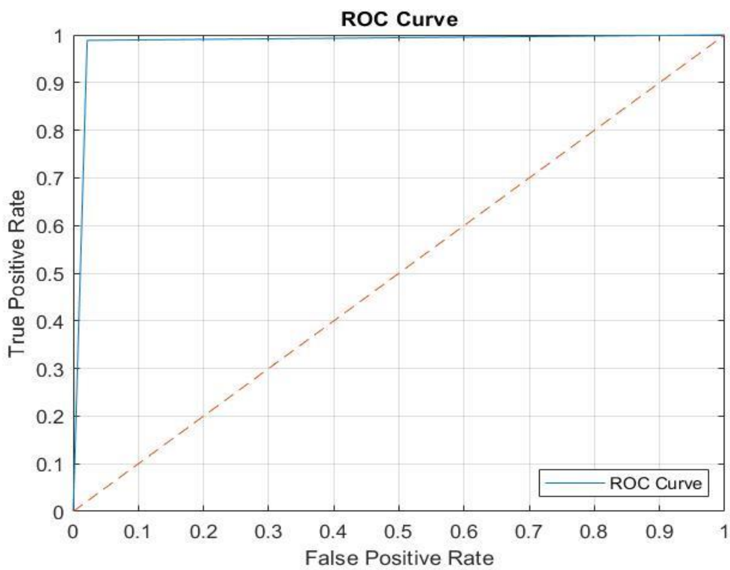
VGG-16 + KNN



The fusion of VGG-16 with KNN produced an accuracy of 93.10%, sensitivity of 87.40%, specificity of 98.70%, F1-Score of 92.64%, precision of 98.55%, Youden Index of 86.10% and AUC score 93.05%. The ROC curve for the 2-way screening of CXR images using VGG-16 + KNN is presented in Figure 39.

Figure 39

ROC Curve for the 2-way screening of MS images using VGG-16 + KNN



VGG-19 + KNN

The fusion of VGG-19 with KNN yielded in an accuracy of 93.9%, sensitivity of 88.70%, specificity of 99.00%, F1-Score of 93.52%, precision of 98.89%, Youden Index of 87.70% and AUC score 93.85%. The ROC curve for the 2-way screening of CXR images using VGG-19 + KNN is presented in Figure 40.

Figure 40

ROC Curve for the binary classification of MS images using VGG-16 + KNN

The overall result achieved by training and testing pre-trained models' ensemble with SVM are presented in Table 10 and Figure 41.

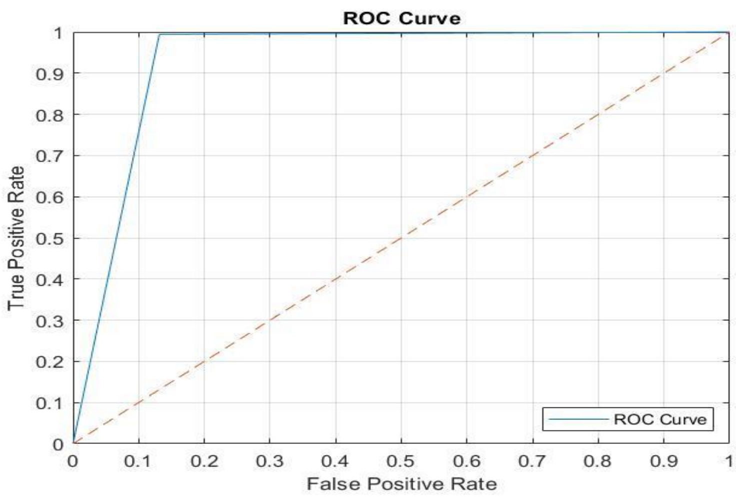
Table 10

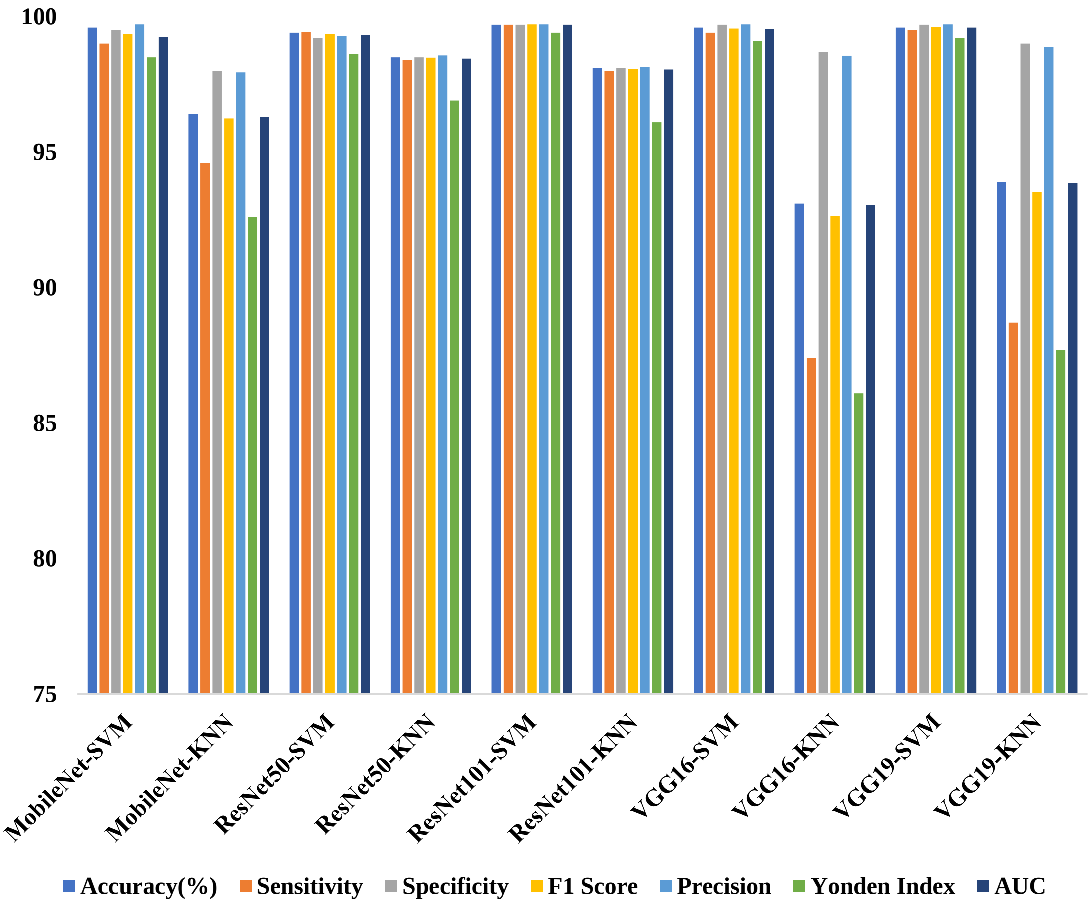
Performance Evaluation of Pre-trained models fused with KNN on CXR Images

Models	Performance Metrics (%)						
	ACC	AUC	F1-Score	PR	SP	SV	YI
MobileNet	99.64	96.30	96.24	97.94	98.00	94.60	92.60
ResNet-50	98.50	98.45	98.48	98.57	98.50	98.40	96.90
ResNet-101	98.81	98.05	98.07	98.14	98.10	98.00	96.10
VGG-16	93.10	93.05	92.64	98.55	98.70	87.40	86.10
VGG-19	93.90	93.85	93.52	98.89	99.00	88.70	87.70

Figure 41

Performance Evaluation of ensemble models using CXR images





4.4 Discussion

The advancement in the field of computing sciences and technology, software engineering, internet and networking, cloud computing as well as growing volume of big data (BD) is changing the landscape and transforming several fields and our day to day lives. The field of AI as a sub field of computer discipline is improving several applications in banking and finance, education, industries, transportation, medicine and agriculture. Narrowing down to the healthcare sector, AI coupled with IoT or IoMT is transforming the sector to a more productive, accurate and reliable system. AI-based system are used in conjunction with various types of medical data such as digital, audio, imagery, video etc. for detection, classification, prediction, enhancement, modeling, processing of diseases, prognosis, diagnosis of severity of diseases , treatment, generation of new drug candidates and repurposing of drugs etc. Across healthcare settings, AI has been applied for diagnosis of disease which include different types of cancer such as metastatic breast cancer, lung cancer, skin cancer, brain cancer etc. parasites, microbes (pneumonia, tuberculosis).

4.4.1 Evaluation of Best Performing models on MS images

As shown in Table 4.1, ResNet-101 coupled with SVM achieved the highest score across all metrics for the classification of MS images. The next model after ResNet-101 is ResNet-50 fused with SVM which achieved similar score with ResNet-101 coupled with SVM in terms of sensitivity (99.50%). MobileNet ranked third followed by VGG-16 while VGG-19 ranked last. The result achieved by both ResNet-101 and ResNet-50 can be attributed to the depth of the models and strong classification capability of SVM.

The performance analysis of 5 pre-trained models fused with KNN for the discrimination of TB and normal cases from MS images has shown that ResNet-101 ensemble with KNN achieved the best result across all metrics followed by ResNet-50 and MobileNet while VGG-19 and VGG-16 ranked fourth and fifth respectively as shown in Table 4.2.

4.4.2 Evaluation of Best Performing models on CXR images

As shown in Table 4.3, ResNet-101 coupled with SVM achieved the highest score across all metrics for the classification of MS images. The next models after ResNet-101 are VGG-16 and VGG-19 which achieved similar score with ResNet-101 coupled with SVM in terms of precision (99.71%) and specificity (99.70). MobileNet and ResNet-50 ranked bottom despite achieving excellent score across all metrics. The result achieved by ResNet-101 can be attributed to the depth of the models and strong classification capability of SVM.

The performance analysis of 5 pre-trained models fused with KNN for the discrimination of TB and normal cases from CXR images has shown that ResNet-50 ensemble with KNN accomplished the best result across 4 metrics which include AUC (98.45%), F1-Score (98.48%), Sensitivity (98.40%) and Youden Index (96.90%) followed by ResNet-101 and VGG-19 which achieved the highest result in terms of precision (98.89%) and specificity (99.00%). Moreover, MobileNet and VGG-16 ranked fourth and fifth respectively as shown in Table 10.

4.4.3 Comparative Analysis

4.4.3.1 MS Images

The comparison between models fused with SVM and KNN for the classification of TB and Healthy cases from MS images has shown that ResNet-101 fused with KNN achieved significantly higher result compared with ResNet-101 fused with SVM as shown in Table 11.

Table 11

Comparison between best performing models trained and tested using MS images

Performance Metrics	ResNet-101 + SVM	ResNet-101 + KNN
Accuracy (%)	99.60	99.80
AUC (%)	99.55	99.80
F1-Score (%)	99.54	99.79
Precision (%)	95.59	99.59
Specificity (%)	99.60	99.60
Sensitivity (%)	95.50	100.00
Youden Index (%)	99.10	99.60

4.4.3.2 CXR Images

The comparison between models fused with SVM and KNN for the discrimination of TB and normal cases from CXR images has shown that ResNet-101 fused with SVM achieved significantly higher result compared with ResNet-50 fused with KNN as shown in Table 12.

Table 12

Comparison between best performing models trained and tested using CXR images

Performance Metrics	ResNet-101 + SVM	ResNet-50 + KNN
Accuracy (%)	99.70	98.50
AUC (%)	99.70	98.45
F1-Score (%)	99.71	98.48
Precision (%)	99.71	98.57
Specificity (%)	99.70	98.50
Sensitivity (%)	99.70	98.40
Youden Index (%)	99.70	96.90

4.4.4 Comparison between Best performing ensemble models on MS and CXR images

Comparison between best performing modes has shown that ResNet-101 + KNN as the best approach for the discrimination of TB and normal cases from MS images outperformed ResNet-101 + SVM as the best approach for the discrimination of TB and Healthy cases from CXR images. ResNet-101 + KNN ranked higher in terms of 4 metrics (4/7) which include Accuracy, AUC, F1-Score and Sensitivity while ResNet-101 + SVM ranked higher in terms of 3 metrics (3/7) which include Precision, specificity and Youden Index as summarized in Table 13.

Table 13

Comparison between Best performing ensemble models on MS and CXR images

Performance Metrics	CXR Images	MS Images
	ResNet-101 + SVM	ResNet-101 + KNN
Accuracy (%)	99.70	99.80
AUC (%)	99.70	99.80
F1-Score (%)	99.71	99.79
Precision (%)	99.71	99.59
Specificity (%)	99.70	99.60
Sensitivity (%)	99.70	100.00
Youden Index (%)	99.70	99.60

4.5 Comparison with Related work

In order to assess the efficiency of models, it is critical to conduct realistic comparison with other models by taking into account the amount of training dataset, the number of layers of each model (I.e., depth), the classification layer use (SoftMax, SVM, KNN, RF, DT etc.

4.5.1 MS Images

The study conducted by Kant et al., 2018 implemented 5-layered CNN model fused with SVM for the binary classification of MS images. The model is trained and tested using 202 images and the model achieved 74.79% F1-Score, 68.55% precision and 83.78% recall. Compared to this study, our best performing model (ResNet-101 fused with SVM) achieved higher result with 99.79% F1-Score and 99.59% precision. This can be attributed to the depth (number of layers) of ResNet-101 compared to the 5-layered CNN and training and testing the models with larger dataset (1893) compared to 202 utilized by Kant et al., 2018. The study presented by Panicker et al., 2018 implemented a simple CNN model which is trained and evaluated using 1800 images. The model attained F1-Score of 86.76%, precision of 78.4% and recall of 97.13%. Despite using slightly more images (i.e., 93 images more), our model achieved better result (99.79% F1-Score and 99.59% precision) compare with the model developed by Panicker et al., 2018. This may be attributed to the depth of the ResNet-101 and fusion of KNN as classifier.

The study reported by Ibrahim et al., 2021b implemented pre-trained AlexNet for binary classification of MS images. The study initially curated 530 images in which the training set was maximized to 2444 images. The pre-trained model was able to attain significant result with an average accuracy of 98.73%, average sensitivity of 98.59 % and average specificity of 98.84%. Our model achieved slightly higher result owing to the number layers in our model and the use of KNN as the classifier. The study presented by Fu et al., 2022 on the other hand employed large dataset (which comprises of 5930 smears). The study developed CNN model achieved significant result owing to the extensive pre-processing steps and large number of images use. However, our best performing model achieved better result across all similar metrics despite using a smaller number of images.

Zurac et al., 2022 implemented DL-CNN model known as RegNetX4 for the screening of TB. Initially, the research collected 510 images in which several data augmentation was carried and annotation which yield approximately 1 million patches. Compared to ResNet-101 + KNN, RegNetX4 only achieved slightly higher sensitivity (100%) compared with ResNet-101 + KNN (99.60). However, ResNet-101 + KNN achieved slightly higher result in terms of accuracy and sensitivity. The study conducted by Umar Ibrahim et al., 2022b utilized same number of the dataset which present an opportunity for realistic evaluation. The model achieved excellent result using AlexNet + SoftMax. However, by comparison, ResNet-101 + KNN achieved slightly higher result across all metrics. While the study reported by Lee and Lee, 2023 evaluated the performance of 9 models using nearly 100, 000 patches. Comparison of the performance has revealed that NASNet reached the optimum performance with an accuracy of 99.777, recall of 99.771, precision of 99.728 and FI-Score of 99.749. In comparison with ResNet + KNN, NASNet achieved almost similar result across all metrics.

Notwithstanding, Gupta et al., 2023 employed 400 images for detection of TB using AI-based microscopy. Despite achieving moderate result using small number of images, our model achieved significantly higher result across all similar metrics. The study proposed by Waluyo et al., 2023 opted for 6-layered model developed from scratch. The model is designed using SoftMax as the classifier is trained and tested using 220 images. For a model developed from scratch, the result was quite impressive, however, compare to our pre-trained ResNet fused with KNN, the CNN model achieved significantly lower result. The summary of comparison between models reported in previous studies and ResNet + KNN for the detection of TB from MS images is summarized in Table 14.

Table 14

Comparison between best performing model evaluated using MS images with state of the art

References	Number of Images	Number of classes	AI-based model used	Result achieved
Kant et al., 2018	202	2	5-layered CNN	74.79% F1-Score, 68.55% PR and 83.78% RC
Panicker et al., 2018	1800	2	CNN	86.76% F1-Score, 78.4% PR and 97.13% RC
Ibrahim et al., 2021b	530	2	Pre-trained AlexNet	98.73% ACC, 98.59 % SV and 98.84% SP
Zurac et al., 2022	510 (1.1 million patches)	2	RegNetX4	98.33% ACC, 95.65% SV, 100% SP and 97.70 ROC
Umar Ibrahim et al., 2022b	1893	2	Pre-trained AlexNet + SVM	98.73% ACC, 99.42% SV and 98.03% SP
Fu et al., 2022	5930	2	CNN	95.2% ACC, 85.7% SV and 96.9% SP

Lee and Lee, 2023	40 slides (98, 034 patches)	2	NASNet	99.777 ACC, 99.771 RC, 99.728 PR and 99.749 FI-Score
Gupta et al., 2023	400	2	AI-based microscopy	91.53% ACC, 89.25% SV, 92.15% SP, 75.45% PPV and 96.94% NPV
Waluyo et al., 2023	220	2	CNN + KNN	92.5% ACC, 92.5% RC, 93.5% PR and 92.5% F1-Score
This study		2	ResNet-101 + KNN	99.80% ACC, 99.80% AUC, 99.79% F1-Score, 99.59% PR, 99.60% SP, 100.00% SV and 99.60% YI

4.5.2 CXR Images

The study conducted by Lakhani and Sundaram (2017) proposed the use of ensemble AlexNet and GoogleNet for the discrimination of TB and Healthy cases using CXR images. The ensemble network is trained and verified using 1007 images which resulted in 97.3% sensitivity, 94.7% specificity and 99.0% AUC score. However, our best performing model ResNet-101 + SVM achieved higher score across all metrics. Thus, this can be ascribed to the use of deeper model fused with SVM and training using larger volume of images. The study reported by Hooda et al., 2017 utilized 800 images to trained and tested FF-CNN-19 model. The result achieved by the model yielded 94.73% top accuracy which fall short of ResNet + SVM with 99.70% accuracy.

The study presented by Xiong et al., 2018 trained and assessed pre-trained DL-based CNN model with 246 images. In comparison with ResNet + SVM, pre-trained DL-based CNN achieved slightly lower sensitivity (97.94%) and significantly lower specificity (83.65%) compared with ResNet-101 + SVM (99.70% sensitivity) and (99.70% specificity) respectively.

Norval et al., 2019 reported DL technique based on MLNN for the detection of TB using 800 images. The proposed model was able to attain 92.54% accuracy which is lower than the accuracy achieved by ResNet-101 + SVM. Rahman et al., 2020 on the other hand utilized 7000 images which is similar to the amount of dataset employed in this study, thus, this presents an opportunity for realistic comparison. The study employed several models, and the resulting comparison has revealed that ChexNet attained the optimum result with an accuracy of 96.47%, precision of 96.62%, and sensitivity of 96.47%, specificity of 96.51% and F1-score of 96.47%. However, ResNet-101 + SVM outperformed ChexNet across all metrics.

The study reported by reported by Sharma et al., 2021 utilized 1000 CXR images in order to trained and tested proposed framework which combines segmentation, feature extraction and classification using DT. Evaluation of the proposed framework result in 98% accuracy, 98% precision, 97.9% F1-Score and 99% AUC score. However, despite Akbari and Azizi (2023) proposed the implementation of developed CNN from scratch for the detection of TB and non-TB cases. The model which is trained and evaluated using 4200 CXR images achieved an accuracy of 97%, precision of 98% and F1-Score of 90%. However, the developed model achieved slightly lower score compared to ResNet + SVM. Finally, the study proposed by Nafisah and Muhammad (2024) utilized 1098 CXR images to trained and tested several DL models. Among these models, EfficientNet-B3 achieved best result with an average accuracy of 99.1% and precision of 98.3%. However, the model achieved slightly lower result compared to ResNet + SVM. The summary of comparison between models reported in previous and ResNet + SVM for the screening of TB from CXR images is presented in Table 15.

Table 15

Comparison between best performing model evaluated using CXR images with state of the art

References	Number of Images	Number of classes	AI-based model used	Result achieved

Lakhani and Sundaram (2017)	1007	2	Pre-trained ensemble AlexNet and GoogleNet	97.3% SV, 94.7% SP and 99.0% AUC
Hooda et al., 2017	800	2	19-layer CNN	94.73% ACC
Xiong et al., 2018	246	2	Pre-trained CNN	97.94% SV and 83.65% SP
Norval et al., 2019	800	2	CAD-MLNN	92.54% ACC
Rahaman et al., 2020	7000	2	DenseNet201	98.6% ACC, 98.57% PR, 98.56% SV, 98.54% SP and 98.56% F1 score
Sharma et al., 2021	1000	2	DT	98% ACC, 98% PR, 98% RC, 99% AUC and 97.9% F1-Score
Umar Ibrahim et al., 2022b	7000	2	Pre-trained AlexNet + SVM	98.38% ACC, 98.71% SV and 98.04% SP
Akbari and Azizi (2023)	4200	2	CNN	97% AC, 98% PR, 83% RC and 90% F-1 score
Nafisah and Muhammad (2024)	1098	2	EfficientNet-B3	99.1% ACC, o99.9% ROC, 98.3% RC, 98.3% PR, 98.3% F1-score, 99.0% SP and 97.2% kappa
This study	7000	2	ResNet-101 + SVM	99.70% ACC, 99.70% AUC, 99.71% F1-Score, 99.71% PR, 99.70% SP, 99.70 % SV and 99.70% YI

CHAPTER V
CONCLUSION

TB is classified as one of the major courses of mortality globally. Despite the growing incidence of the disease, one of the major issues is the growing prevalence of drug-resistant tuberculosis (DR-TB) which lead to upsurge in number of cases, cost of diagnosis and management. Medical practitioners rely on several techniques for screening of suspected patients which include TST, IGRA or QuantiFERON, AFB, GeneXpert and CXR imaging. The TST approach is hindered by inaccuracy, lack specificity for active TB. The IGRA method is limited due to longer processing time and required extra test to confirm positive cases. The AFB is time-consuming, The GeneXpert procedure is restricted to laboratory, require the use of chemicals and machine. The CXR imaging technique is comparatively available and more affordable in majority of underdeveloped countries, however, the technique suffer from the lack of sufficient radiologist that can run and interpret the results which can lead to high workload and risk of miss-diagnosis. Therefore, there is need to develop automated, smart, accurate, reliable, fast and cheap approach for the screening of TB.

To address these issues, scientist adopted AI/CAD techniques which include ML and DL models such as classical ML algorithms which include DT, KNN, LR, SVM etc. Neural networks such as ANNs, CNNs, RNNs etc. image processing, feature extraction, segmentation etc. The last decade has witnessed the raise of DL models, a sub-class of ML and AI that run mathematical algorithms which can detect underlying patterns and features from data input. Several DL-based models have been developed from scratch as well as the reuse of models known as pre-trained models for image recognition, detection, segmentation, and classification. AI-based models have been reported to aid healthcare workers in conducting accurate diagnosis, minimizing error rate and miss-diagnosis, and reducing workload.

In line with this application, we proposed an automated CAD of TB and Healthy cases from both MS and CXR images. The methodology of this work is divided into experiments (A and B). Experiment A is characterized by the curation of 1893 MS images (1363 TB and 530 Healthy cases). While experiment B is characterized by the curation of 7000 CXR images (3500 TB and 3500 healthy cases). Both the images in experiment A and B are trained and verified using 5 pre-trained models which include ResNet-101, ResNet-50, MobileNet, VGG-19 and VGG-16 ensemble with 2 classical ML models which include KNN and SVM. Evaluation and comparison between performance of ensemble models has revealed that ResNet-101 + KNN accomplished the optimum performance with 99.80% accuracy, 99.80% AUC, 99.79% F1-Score, 99.59% precision, 99.60% specificity, 100.00% sensitivity (%) and 99.60%Youden index on MS images. While ResNet-101 + SVM reached the optimum scores across all metrics with 99.70% accuracy, 99.70% AUC, 99.71% F1-Score, 99.71% precision, 99.70% specificity, 99.70 % sensitivity (%) and 99.70%Youden index on CXR images. Despite achieving nearly optimum result, the performance can be improved by training the models with larger datasets, oversampling using several data augmentation techniques and development of Internet of Things (IoT) system for real-time detection.

5.1 Limitations and Future Direction

ML architectures are fuel by enormous amount of data in order to achieve significant result. Even though there is no specific amount of data required, however, training ML models especially the ones developed from scratch require thousands if not hundred thousand data samples or features. One of the limitations of this study is training ensemble models with only X-ray and MS images, therefore, future studies will explore other clinical images such as 3D, CT scans, and MRI scans etc. Another limitation of this study is training ensemble models with less than 2000 images for MS images and 7000 for CXR images. Even though, we attained state of the art result owing to the implementation of pre-trained models. TL models have been used to address the issue of shortage of data samples. TL or pre-trained CNNs are models that are already trained using substantial volume of data point and can be repurposed for training using small number of datasets. Despite implementing pre-trained models, training using large amount of image data samples can improve performance and efficiency.

Consequently, scientists also proposed another approach to resolve the issue of scarcity of dataset through synthetic data generation. The use of synthetic data generated via several data augmentation methods such as rotation, shifting, zooming, cropping, shearing, color enhancement etc. have shown it maximize training set which contribute significantly to the performance of the models. Another, limitations of this study involve the absent of data augmentation. Image dataset obtained from online repositories are use directly. Thus, future studies will employ several data augmentation in order maximize training set.

Notwithstanding, the last decades has witnessed the development of several ML and DL models which include Le-Net, AlexNet-8, VGG-16 and VGG-19, GoogleNet or Inception and their variations, ResNet and their variants and other recent models such as NASNet, EfficientNet, DenseNet etc. The third limitation of this research is the implementation of 5 pre-trained models. Therefore, future studies will attempt to employ other state of art models or deeper models such as ResNet-152, Inception, Xception, InceptionResNetV2, EfficientNet, DenseNets and its variants and its variants etc. Subsequently only 2 classical ML classifiers which include KNN and SVM were employed. Thus, future studies will attempt to ensemble DL models with other classifiers such as DT, NB, LR, RF etc.

IoMT also known as IoHT is another cutting-edge technology that is revolutionizing healthcare, communication, industries, banking and finance, agriculture, military, security sector. Several IoT-based frameworks integrated with AI-based technology have been established and implemented for the real-time screening of diseases. Thus, future work will attempt to develop AI/IoT-based framework that will enable patients and medical experts to upload MS and CXR images and obtain result in real-time.

REFERENCES

Abdel-Samea, S. A., Ismail, Y. M., Fayed, S. M. A., & Mohammad, A. A. (2013). Comparative study between using QuantiFERON and tuberculin skin test in diagnosis of Mycobacterium tuberculosis infection. Egyptian Journal of Chest Diseases and Tuberculosis, 62(1), 137-143.

Abro, A. A., Khan, A. A., Talpur, M. S. H., Kayijuka, I., & Yaşar, E. (2021). Machine learning classifiers: a brief primer. University of Sindh Journal of Information and Communication Technology, 5(2), 63-68.

Aghdam, H. H., & Heravi, E. J. (2017). Guide to convolutional neural networks. New York, NY: Springer, 10(978-973), 51.

Alfarisi, O., Alghamdi, W. A., Al-Shaer, M. H., Dooley, K. E., & Peloquin, C. A. (2017). Rifampin vs. rifapentine: what is the preferred rifamycin for tuberculosis?. Expert Review of Clinical Pharmacology, 10(10), 1027-1036.

Alzubi, J., Nayyar, A., & Kumar, A. (2018, November). Machine learning from theory to algorithms: an overview. In Journal of physics: conference series (Vol. 1142, p. 012012). IOP Publishing.

Ao, C., Jin, S., Ding, H., Zou, Q., & Yu, L. (2020). Application and development of artificial intelligence and intelligent disease diagnosis. Current pharmaceutical design, 26(26), 3069-3075.

Barua, R., & Mondal, J. (2023). Study of the Current Trends of CAD (Computer-Aided Detection) in Modern Medical Imaging. Machine Learning and AI Techniques in Interactive Medical Image Analysis, 35-50.

Basaraba, R. J., & Hunter, R. L. (2017). Pathology of tuberculosis: how the pathology of human tuberculosis informs and directs animal models. Tuberculosis and the Tubercle Bacillus, 117-129.

Bi, Z., Dip, S. A., Hajialigol, D., Kommu, S., Liu, H., Lu, M., & Wang, X. (2024). AI for Biomedicine in the Era of Large Language Models. arXiv preprint arXiv:2403.15673.

Brady, J., Cohen, E., El-Kareh, R., Gleason, K., Kathryn McDonald, M. M., Olson, A., & Simpson, L. (2019). The Diagnostic Error in Medicine 12th Annual International Conference.

Brodie, D., & Schluger, N. W. (2005). The diagnosis of tuberculosis. Clinics in chest medicine, 26(2), 247-271.

Burkart, N., & Huber, M. F. (2021). A survey on the explainability of supervised machine learning. Journal of Artificial Intelligence Research, 70, 245-317.

Campbell, I. A., & Bah-Sow, O. (2006). Pulmonary tuberculosis: diagnosis and treatment. Bmj, 332(7551), 1194-1197.

Castellino, R. A. (2005). Computer aided detection (CAD): an overview. Cancer Imaging, 5(1), 17.

Center for Disease amd Prevention (CDC) (2024). Tuberculosis. Available at https://www.cdc.gov/tb/index.html. Accessed on 12th July, 2024.

Cicerone, M. T., & Camp Jr, C. H. (2019). Potential roles for spectroscopic coherent Raman imaging for histopathology and biomedicine. In Neurophotonics and biomedical spectroscopy (pp. 547-570). Elsevier.

Cicerone, M. T., & Camp Jr, C. H. (2019). Potential roles for spectroscopic coherent Raman imaging for histopathology and biomedicine. In Neurophotonics and Biomedical Spectroscopy (pp. 547-570). Elsevier.

Cioffi, R., Travaglioni, M., Piscitelli, G., Petrillo, A., & De Felice, F. (2020). Artificial intelligence and machine learning applications in smart production: Progress, trends, and directions. Sustainability, 12(2), 492.

Clifford, V., He, Y., Zufferey, C., Connell, T., & Curtis, N. (2015). Interferon gamma release assays for monitoring the response to treatment for tuberculosis: a systematic review. Tuberculosis, 95(6), 639-650.

Cunningham, P., & Delany, S. J. (2021). K-nearest neighbour classifiers-a tutorial. ACM computing surveys (CSUR), 54(6), 1-25.

DC Health (2024). Tuberculosis Basics. Available at https://dchealth.dc.gov/page/tuberculosis-basics. Accessed on 13th July, 2024.

Diallo, T., Adjobimey, M., Ruslami, R., Trajman, A., Sow, O., Obeng Baah, J., ... & Menzies, D. (2018). Safety and side effects of rifampin versus isoniazid in children. New England Journal of Medicine, 379(5), 454-463.

Druszczynska, M., Kowalewicz-Kulbat, M., Fol, M., Wlodarczyk, M., & RuDnICKA, W. (2012). Latent M. tuberculosis infection--pathogenesis, diagnosis, treatment and prevention strategies. Pol J Microbiol, 61(1), 3-10.

Dye, C. (2006). Global epidemiology of tuberculosis. The Lancet, 367(9514), 938-940.

Faust, O., Acharya, U. R., Sudarshan, V. K., San Tan, R., Yeong, C. H., Molinari, F., & Ng, K. H. (2017). Computer aided diagnosis of coronary artery disease, myocardial infarction and carotid atherosclerosis using ultrasound images: a review. Physica Medica, 33, 1-15.

François-Lavet, V., Henderson, P., Islam, R., Bellemare, M. G., & Pineau, J. (2018). An introduction to deep reinforcement learning. Foundations and Trends® in Machine Learning, 11(3-4), 219-354.

Fu, H. T., Tu, H. Z., Lee, H. S., Lin, Y. E., & Lin, C. W. (2022). Evaluation of an AI-based TB AFB smear screening system for laboratory diagnosis on routine practice. Sensors, 22(21), 8497.

Fujita, H., Uchiyama, Y., Nakagawa, T., Fukuoka, D., Hatanaka, Y., Hara, T., ... & Zhou, X. (2008). Computer-aided diagnosis: The emerging of three CAD systems induced by Japanese health care needs. Computer methods and programs in biomedicine, 92(3), 238-248.

Gabriel, O. T. (2023). Data privacy and ethical issues in collecting health care data using artificial intelligence among health workers (Master's thesis, Center for Bioethics and Research).

Ganaie, M. A., Hu, M., Malik, A. K., Tanveer, M., & Suganthan, P. N. (2022). Ensemble deep learning: A review. Engineering Applications of Artificial Intelligence, 115, 105151.

Geldenhuys, H., Verver, S., Surtie, S., Hatherill, M., Van Leth, F., Kafaar, F., ... & Mahomed, H. (2010). The tuberculin skin test: a comparison of ruler and calliper readings. The International journal of tuberculosis and lung disease, 14(10), 1266-1271.

Ghahramani, Z. (2003). Unsupervised learning. In Summer school on machine learning (pp. 72-112). Berlin, Heidelberg: Springer Berlin Heidelberg.

Gong, M. (2021). A novel performance measure for machine learning classification. International Journal of Managing Information Technology (IJMIT) Vol, 13.

Goude, R., Amin, A. G., Chatterjee, D., & Parish, T. (2009). The arabinosyltransferase EmbC is inhibited by ethambutol in Mycobacterium tuberculosis. Antimicrobial agents and chemotherapy, 53(10), 4138-4146.

Gupta, P., Khare, V., Srivastava, A., Agarwal, J., Mittal, V., Sonkar, V., ... & Jain, A. (2023). A prospective observational multicentric clinical trial to evaluate microscopic examination of acid-fast bacilli in sputum by artificial intelligence-based microscopy system. Journal of Investigative Medicine, 71(7), 716-721.

Guresen, E., & Kayakutlu, G. (2011). Definition of artificial neural networks with comparison to other networks. Procedia Computer Science, 3, 426-433.

Hady, M. F. A., & Schwenker, F. (2013). Semi-supervised learning. Handbook on Neural Information Processing, 215-239.

Han, X., Zhang, Z., Ding, N., Gu, Y., Liu, X., Huo, Y., ... & Zhu, J. (2021). Pre-trained models: Past, present and future. AI Open, 2, 225-250.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

Helm, J. M., Swiergosz, A. M., Haeberle, H. S., Karnuta, J. M., Schaffer, J. L., Krebs, V. E., ... & Ramkumar, P. N. (2020). Machine learning and artificial intelligence: definitions, applications, and future directions. Current reviews in musculoskeletal medicine, 13, 69-76.

Hooda, R., Sofat, S., Kaur, S., Mittal, A., & Meriaudeau, F. (2017, September). 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). IEEE.

Horgan, D., Romao, M., Morré, S. A., & Kalra, D. (2020). Artificial intelligence: power for civilisation–and for better healthcare. Public health genomics, 22(5-6), 145-161.

Hossin, M., & Sulaiman, M. N. (2015). A review on evaluation metrics for data classification evaluations. International journal of data mining & knowledge management process, 5(2), 1.

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M. & Adam, H. (2017). MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications (cite arxiv:1704.04861)

Huang, T. M., Kecman, V., & Kopriva, I. (2006). Semi-supervised learning and applications. Kernel Based Algorithms for Mining Huge Data Sets: Supervised, Semi-supervised, and Unsupervised Learning, 125-173.

Ibrahim, A. U., Guler, E., Guvenir, M., Suer, K., Serte, S., & Ozsoz, M. (2021). Automated detection of Mycobacterium tuberculosis using transfer learning. The Journal of Infection in Developing Countries, 15(05), 678-686.

Ibrahim, A. U., Guler, E., Guvenir, M., Suer, K., Serte, S., & Ozsoz, M. (2021). Automated detection of Mycobacterium tuberculosis using transfer learning. The Journal of Infection in Developing Countries, 15(05), 678-686.

Ibrahim, A. U., Ozsoz, M., Serte, S., Al-Turjman, F., & Yakoi, P. S. (2021a). Pneumonia classification using deep learning from chest X-ray images during COVID-19. Cognitive computation, 1-13.

Illán, I. A., Górriz, J. M., López, M. M., Ramírez, J., Salas-Gonzalez, D., Segovia, F., ... & Puntonet, C. G. (2011). Computer aided diagnosis of Alzheimer's disease using component based SVM. Applied Soft Computing, 11(2), 2376-2382.

Irkham, I., Ibrahim, A. U., Nwekwo, C. W., Al-Turjman, F., & Hartati, Y. W. (2022). Current technologies for detection of COVID-19: Biosensors, artificial intelligence and internet of medical things (IOMT). Sensors, 23(1), 426.

Jacovi, A., Shalom, O. S., & Goldberg, Y. (2018). Understanding convolutional neural networks for text classification. arXiv preprint arXiv:1809.08037.

Jaderberg, M., Simonyan, K., Vedaldi, A., & Zisserman, A. (2016). Reading text in the wild with convolutional neural networks. International journal of computer vision, 116, 1-20.

Jaeger, S., Karargyris, A., Candemir, S., Folio, L., Siegelman, J., Callaghan, F., ... & McDonald, C. J. (2013). Automatic tuberculosis screening using chest radiographs. IEEE transactions on medical imaging, 33(2), 233-245.

Jakhar, D., & Kaur, I. (2020). Artificial intelligence, machine learning and deep learning: definitions and differences. Clinical and experimental dermatology, 45(1), 131-132.

Janiesch, C., Zschech, P., & Heinrich, K. (2021). Machine learning and deep learning. Electronic Markets, 31(3), 685-695.

Javaid, M., Haleem, A., Singh, R. P., & Suman, R. (2022). Artificial intelligence applications for industry 4.0: A literature-based study. Journal of Industrial Integration and Management, 7(01), 83-111.

Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., ... & Wang, Y. (2017). Artificial intelligence in healthcare: past, present and future. Stroke and vascular neurology, 2(4).

Jiang, T., Gradus, J. L., & Rosellini, A. J. (2020). Supervised machine learning: a brief primer. Behavior therapy, 51(5), 675-687.

Jogin, M., Madhulika, M. S., Divya, G. D., Meghana, R. K., & Apoorva, S. (2018, May). Feature extraction using convolution neural networks (CNN) and deep learning. In 2018 3rd IEEE international conference on recent trends in electronics, information & communication technology (RTEICT) (pp. 2319-2323). IEEE.

Jorritsma, W., Cnossen, F., & van Ooijen, P. M. (2015). Improving the radiologist–CAD interaction: designing for appropriate trust. Clinical radiology, 70(2), 115-122.

Joseph, V. R., & Vakayil, A. (2022). SPlit: An optimal method for data splitting. Technometrics, 64(2), 166-176.

Kant, S., & Srivastava, M. M. (2018, November). Towards automated tuberculosis detection using deep learning. In 2018 IEEE Symposium Series on Computational Intelligence (SSCI) (pp. 1250-1253). IEEE.

Kaur, S., & Singh, I. (2017). Artificial intelligence based clinical data management systems: A review. Informatics in Medicine Unlocked, 9, 219-229.

Kensert, A., Harrison, P. J., & Spjuth, O. (2019). Transfer learning with deep convolutional neural networks for classifying cellular morphological changes. SLAS Discovery: Advancing Life Sciences R&D, 24(4), 466-475.

Khalid, N., Qayyum, A., Bilal, M., Al-Fuqaha, A., & Qadir, J. (2023). Privacy-preserving artificial intelligence in healthcare: Techniques and applications. Computers in Biology and Medicine, 158, 106848.

Kramer, O., & Kramer, O. (2013). K-nearest neighbors. Dimensionality reduction with unsupervised nearest neighbors, 13-23.

Krenker, A., Bešter, J., & Kos, A. (2011). Introduction to the artificial neural networks. Artificial Neural Networks: Methodological Advances and Biomedical Applications. InTech, 1-18.

Lee, C. S., Nagy, P. G., Weaver, S. J., & Newman-Toker, D. E. (2013). Cognitive and system factors contributing to diagnostic errors in radiology. American Journal of Roentgenology, 201(3), 611-617.

Lee, J., & Lee, J. (2023). A Study of mycobacterium tuberculosis detection using different neural networks in autopsy specimens. Diagnostics, 13(13), 2230.

Li, Y. (2017). Deep reinforcement learning: An overview. arXiv preprint arXiv:1701.07274.

Liu, Y., Zhou, Y., Wen, S., & Tang, C. (2014). A strategy on selecting performance metrics for classifier evaluation. International Journal of Mobile Computing and Multimedia Communications (IJMCMC), 6(4), 20-35.

Long, R., & Schwartzman, K. (2014). Pathogenesis and transmission of tuberculosis. Canadian tuberculosis standards, 25.

Luo, J., & Xiong, C. (2013). Youden index and associated cut-points for three ordinal diagnostic groups. Communications in Statistics-Simulation and Computation, 42(6), 1213-1234.

MacGregor-Fairlie, M., Wilkinson, S., Besra, G. S., & Goldberg Oppenheimer, P. (2020). Tuberculosis diagnostics: overcoming ancient challenges with modern solutions. Emerging topics in life sciences, 4(4), 435-448.

Mahajan, P., Uddin, S., Hajati, F., & Moni, M. A. (2023, June). Ensemble learning for disease prediction: A review. In Healthcare (Vol. 11, No. 12, p. 1808). MDPI.

Mahesh, B. (2020). Machine learning algorithms-a review. International Journal of Science and Research (IJSR).[Internet], 9(1), 381-386.

Marcelino, P. (2018). Transfer learning from pre-trained models. Towards data science, 10(330), 23.

Mathema, B., Andrews, J. R., Cohen, T., Borgdorff, M. W., Behr, M., Glynn, J. R., ... & Wood, R. (2017). Drivers of tuberculosis transmission. The Journal of infectious diseases, 216(suppl_6), S644-S653.

McIlleron, H., Meintjes, G., Burman, W. J., & Maartens, G. (2007). Complications of antiretroviral therapy in patients with tuberculosis: drug interactions, toxicity, and immune reconstitution inflammatory syndrome. The Journal of infectious diseases, 196(Supplement_1), S63-S75.

Medyakov, D., Molodtsov, G., Beznosikov, A., & Gasnikov, A. (2023, December). Optimal data splitting in distributed optimization for machine learning. In Doklady Mathematics (Vol. 108, No. Suppl 2, pp. S465-S475). Moscow: Pleiades Publishing.

Mehta, D. (2020). State-of-the-art reinforcement learning algorithms. International Journal of Engineering Research and Technology, 8, 717-722.

Michelucci, U. (2018). Applied deep learning. A case-based approach to understanding Deep neural networks.

Miglani, A., Saini, C., Musyuni, P., & Aggarwal, G. (2022). A review and analysis of product recall for pharmaceutical drug product. Journal of Generic Medicines, 18(2), 72-81.

Millard, J., Ugarte-Gil, C., & Moore, D. A. (2015). Multidrug resistant tuberculosis. Bmj, 350.

Mohan, A., & Sharma, S. K. (2004). Side effects of antituberculosis drugs. American journal of respiratory and critical care medicine, 169(7), 882-883.

Mucherino, A., Papajorgji, P. J., Pardalos, P. M., Mucherino, A., Papajorgji, P. J., & Pardalos, P. M. (2009). K-nearest neighbor classification. Data mining in agriculture, 83-106.

Murdoch, B. (2021). Privacy and artificial intelligence: challenges for protecting health information in a new era. BMC Medical Ethics, 22, 1-5.

Naeem, S., Ali, A., Anam, S., & Ahmed, M. M. (2023). An unsupervised machine learning algorithms: Comprehensive review. International Journal of Computing and Digital Systems.

Nafisah, S. I., & Muhammad, G. (2024). Tuberculosis detection in chest radiograph using convolutional neural network architecture and explainable artificial intelligence. Neural Computing and Applications, 36(1), 111-131.

Neyshabur, B., Sedghi, H., & Zhang, C. (2020). What is being transferred in transfer learning?. Advances in neural information processing systems, 33, 512-523.

Njire, M., Tan, Y., Mugweru, J., Wang, C., Guo, J., Yew, W., ... & Zhang, T. (2016). Pyrazinamide resistance in Mycobacterium tuberculosis: Review and update. Advances in medical sciences, 61(1), 63-71.

Norval, M., Wang, Z., & Sun, Y. (2019, December). Pulmonary tuberculosis detection using deep learning convolutional neural networks. In Proceedings of the 3rd International Conference on Video and Image Processing (pp. 47-51).

Nwalozie, R., Nnokam, B. A., Kareem, J. A., & Obi-Thomas, J. N. (2024). Detection of Mycobacterium tuberculosis using Acid-Fast Bacilli (AFB) Test among Subjects in Rivers State of Nigeria. Journal of Advances in Microbiology, 24(4), 1-11.

Oakden-Rayner, L. (2019). The rebirth of CAD: how is modern AI different from the CAD we know?. Radiology: artificial intelligence, 1(3), e180089.

O'shea, K., & Nash, R. (2015). An introduction to convolutional neural networks. arXiv preprint arXiv:1511.08458.

Pai, M., & Behr, M. (2016). Latent Mycobacterium tuberculosis infection and interferon-gamma release assays. Microbiology spectrum, 4(5), 10-1128.

Panic, N., Maetzel, H., Bulajic, M., Radovanovic, M., & Löhr, J. M. (2020). Pancreatic tuberculosis: a systematic review of symptoms, diagnosis and treatment. United European gastroenterology journal, 8(4), 396-402.

Panicker, R. O., Kalmady, K. S., Rajan, J., & Sabu, M. K. (2018). Automatic detection of tuberculosis bacilli from microscopic sputum smear images using deep learning methods. Biocybernetics and Biomedical Engineering, 38(3), 691-699.

Paras, V. (2020). VGGNet-16 Architecture: A Complete Guide. Available at https://www.kaggle.com/code/blurredmachine/vggnet-16-architecture-a-complete-guide. Accessed on 15th July, 2024.

Pesapane, F., Codari, M., & Sardanelli, F. (2018). Artificial intelligence in medical imaging: threat or opportunity? Radiologists again at the forefront of innovation in medicine. European radiology experimental, 2, 1-10.

Peterson, L. E. (2009). K-nearest neighbor. Scholarpedia, 4(2), 1883.

Petrick, N., Sahiner, B., Armato III, S. G., Bert, A., Correale, L., Delsanto, S., ... & Chan, H. P. (2013). Evaluation of computer-aided detection and diagnosis systems a. Medical physics, 40(8), 087001.

Pisner, D. A., & Schnyer, D. M. (2020). Support vector machine. In Machine learning (pp. 101-121). Academic Press.

Qiang, W., & Zhongli, Z. (2011, August). Reinforcement learning model, algorithms and its application. In 2011 International Conference on Mechatronic Science, Electric Engineering and Computer (MEC) (pp. 1143-1146). IEEE.

Rahman, T., Khandakar, A., Kadir, M. A., Islam, K. R., Islam, K. F., Mazhar, R., ... & Chowdhury, M. E. (2020). Reliable tuberculosis detection using chest X-ray with deep learning, segmentation and visualization. Ieee Access, 8, 191586-191601.

Rangaka, M. X., Wilkinson, R. J., Boulle, A., Glynn, J. R., Fielding, K., Van Cutsem, G., ... & Maartens, G. (2014). Isoniazid plus antiretroviral therapy to prevent tuberculosis: a randomised double-blind, placebo-controlled trial. The Lancet, 384(9944), 682-690.

Razavi, S. (2021). Deep learning, explained: Fundamentals, explainability, and bridgeability to process-based modelling. Environmental Modelling & Software, 144, 105159.

Reddy, Y. C. A. P., Viswanath, P., & Reddy, B. E. (2018). Semi-supervised learning: A brief review. Int. J. Eng. Technol, 7(1.8), 81.

Regina, M. L., Vecchié, A., Bonaventura, A., & Prisco, D. (2021). Patient safety in internal medicine. Textbook of Patient Safety and Clinical Risk Management, 213-252.

Ribani, R., & Marengoni, M. (2019, October). A survey of transfer learning for convolutional neural networks. In 2019 32nd SIBGRAPI conference on graphics, patterns and images tutorials (SIBGRAPI-T) (pp. 47-57). IEEE.

Rony, M. K. K., Parvin, M. R., Wahiduzzaman, M., Debnath, M., Bala, S. D., & Kayesh, I. (2024). "I Wonder if my Years of Training and Expertise Will be Devalued by Machines": Concerns About the Replacement of Medical Professionals by Artificial Intelligence. SAGE Open Nursing, 10, 23779608241245220.

Rukshan P., (2022). Overview of a Neural Network's Learning Process. Available at https://medium.com/data-science-365/overview-of-a-neural-networks-learning-process-61690a502fa. Accessed on July 12th, 2024.

Salehi, A. W., Khan, S., Gupta, G., Alabduallah, B. I., Almjally, A., Alsolai, H., ... & Mellit, A. (2023). A study of CNN and transfer learning in medical imaging: Advantages, challenges, future scope. Sustainability, 15(7), 5930.

Samek, W., Montavon, G., Lapuschkin, S., Anders, C. J., & Müller, K. R. (2021). Explaining deep neural networks and beyond: A review of methods and applications. Proceedings of the IEEE, 109(3), 247-278.

Sarkar, U., Bonacum, D., Strull, W., Spitzmueller, C., Jin, N., López, A., ... & Singh, H. (2012). Challenges of making a diagnosis in the outpatient setting: a multi-site survey of primary care physicians. BMJ quality & safety, 21(8), 641-648.

Secinaro, S., Calandra, D., Secinaro, A., Muthurangu, V., & Biancone, P. (2021). The role of artificial intelligence in healthcare: a structured literature review. BMC medical informatics and decision making, 21, 1-23.

Selvakumar, N., Ravikumar, M., Sivagamasundari, S., Gopi, P. G., & Narayanan, P. R. (2006). A novel method of staining acid-fast bacilli in sputum containers. Indian Journal of Medical Research, 123(Jun), 776-780.

Sharma, A., Sharma, A., Malhotra, R., Singh, P., Chakrabortty, R. K., Mahajan, S., & Pandit, A. K. (2021). An accurate artificial intelligence system for the detection of pulmonary and extra pulmonary Tuberculosis. Tuberculosis, 131, 102143.

Shashank Srivastava (2019). Transfer Learning. Available at https://medium.com/@skshashankkumar41/transfer-learning-8386e1a9e34a. Accessed on July 7th, 2024.

Shetty, S. H., Shetty, S., Singh, C., & Rao, A. (2022). Supervised machine learning: algorithms and applications. Fundamentals and methods of machine and deep learning: algorithms, tools and applications, 1-16.

Shin, H. C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., ... & Summers, R. M. (2016). Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. IEEE transactions on medical imaging, 35(5), 1285-1298.

Shuaib, A., Arian, H., & Shuaib, A. (2020). The increasing role of artificial intelligence in health care: will robots replace doctors in the future?. International journal of general medicine, 891-896.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

Singh, S., Dey, B., Sachdeva, K. S., Kabra, S. K., Chopra, K. K., Chaudhary, V. K., ... & Katoch, V. M. (2015). Challenges in tuberculosis diagnosis and management: recommendations of the expert panel. Journal of laboratory physicians, 7(01), 001-003.

Ström, P., Kartasalo, K., Olsson, H., Solorzano, L., Delahunt, B., Berney, D. M., ... & Eklund, M. (2020). Artificial intelligence for diagnosis and grading of prostate cancer in biopsies: a population-based, diagnostic study. The Lancet Oncology, 21(2), 222-232.

Suárez, I., Fünger, S. M., Kröger, S., Rademacher, J., Fätkenheuer, G., & Rybniker, J. (2019). The diagnosis and treatment of tuberculosis. Deutsches Aerzteblatt International, 116(43).

Summerton, N., & Cansdale, M. (2019). Artificial intelligence and diagnosis in general practice. British Journal of General Practice, 69(684), 324-325.

Suthaharan, S., & Suthaharan, S. (2016). Support vector machine. Machine learning models and algorithms for big data classification: thinking with examples for effective learning, 207-235.

Talahua, J. S., Buele, J., Calvopiña, P., & Varela-Aldás, J. (2021). Facial recognition system for people with and without face mask in times of the covid-19 pandemic. Sustainability, 13(12), 6900.

Tamim, N., Elshrkawey, M., & Nassar, H. (2021). Accurate diagnosis of diabetic retinopathy and glaucoma using retinal fundus images based on hybrid features and genetic algorithm. Applied Sciences, 11(13), 6178.

Tang, X. (2019). The role of artificial intelligence in medical imaging research. BJR| Open, 2(1), 20190031.

Targ, S., Almeida, D., & Lyman, K. (2016). Resnet in resnet: Generalizing residual architectures. arXiv preprint arXiv:1603.08029.

Teuwen, J., & Moriakov, N. (2020). Convolutional neural networks. In Handbook of medical image computing and computer assisted intervention (pp. 481-501). Academic Press.

Törün, T., Güngör, G., Özmen, I., Bölükbaşı, Y., Maden, E., Bıçakçı, B., ... & Tahaoğlu, K. (2005). Side effects associated with the treatment of multidrug-resistant tuberculosis. The International Journal of Tuberculosis and Lung Disease, 9(12), 1373-1377.

Triasih, R., Robertson, C., De Campo, J., Duke, T., Choridah, L., & Graham, S. M. (2015). An evaluation of chest X-ray in the context of community-based screening of child tuberculosis contacts. The International Journal of Tuberculosis and Lung Disease, 19(12), 1428-1434.

Tsai, K. S., Chang, H. L., Chien, S. T., Chen, K. L., Chen, K. H., Mai, M. H., & Chen, K. T. (2013). Childhood tuberculosis: epidemiology, diagnosis, treatment, and vaccination. Pediatrics & Neonatology, 54(5), 295-302.

Tsiouri, G., Gaitanis, G., Kiorpelidou, D., Dionysiou, A., Efthymiou, A., Daskalopoulos, G., ... & Bassukas, I. D. (2009). Tuberculin skin test overestimates tuberculosis hypersensitivity in adult patients with psoriasis. Dermatology, 219(2), 119-125.

Umar Ibrahim, A., Al-Turjman, F., Ozsoz, M., & Serte, S. (2022). Computer aided detection of tuberculosis using two classifiers. Biomedical Engineering/Biomedizinische Technik, 67(6), 513-524.

Umar Ibrahim, A., Ozsoz, M., Serte, S., Al-Turjman, F., & Habeeb Kolapo, S. (2022). Convolutional neural network for diagnosis of viral pneumonia and COVID-19 alike diseases. Expert Systems, 39(10), e12705.

Unissa, A. N., Subbian, S., Hanna, L. E., & Selvakumar, N. (2016). Overview on mechanisms of isoniazid action and resistance in Mycobacterium tuberculosis. Infection, Genetics and Evolution, 45, 474-492.

Van Cleeff, M. R. A., Kivihya-Ndugga, L. E., Meme, H., Odhiambo, J. A., & Klatser, P. R. (2005). The role and performance of chest X-ray for the diagnosis of tuberculosis: a cost-effectiveness analysis in Nairobi, Kenya. BMC infectious diseases, 5, 1-9.

Van Engelen, J. E., & Hoos, H. H. (2020). A survey on semi-supervised learning. Machine learning, 109(2), 373-440.

Van Hartskamp, M., Consoli, S., Verhaegh, W., Petkovic, M., & Van de Stolpe, A. (2019). Artificial intelligence in clinical health care applications. Interactive journal of medical research, 8(2), e12100.

Vujović, Ž. (2021). Classification model evaluation metrics. International Journal of Advanced Computer Science and Applications, 12(6), 599-606.

Waluyo, W. N., Isnanto, R. R., & Rochim, A. F. (2023). Comparison of mycobacterium tuberculosis image detection accuracy using CNN and combination CNN-KNN. Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi), 7(1), 80-87.

Wang, B., Xie, Q., Pei, J., Chen, Z., Tiwari, P., Li, Z., & Fu, J. (2023). Pre-trained language models in biomedical domain: A systematic survey. ACM Computing Surveys, 56(3), 1-52.

Wang, Z. J., Turko, R., Shaikh, O., Park, H., Das, N., Hohman, F., ... & Chau, D. H. P. (2020). CNN explainer: learning convolutional neural networks with interactive visualization. IEEE Transactions on Visualization and Computer Graphics, 27(2), 1396-1406.

Wong, P. C. (2021). Unsupervised machine learning. Computational Psychometrics: New Methodologies for a New Generation of Digital Learning and Assessment: With Examples in R and Python, 173-193.

World Health Organization (WHO). Tuberculosis. Available at https://www.who.int/health-topics/tuberculosis#tab=tab_1. Accessed on July 8, 2024.

Wulandari, D. A., Hartati, Y. W., Ibrahim, A. U., & Pitaloka, D. A. E. (2024). Multidrug-resistant tuberculosis. Clinica Chimica Acta, 559, 119701.

Yee, D., Valiquette, C., Pelletier, M., Parisien, I., Rocher, I., & Menzies, D. (2003). Incidence of serious side effects from first-line antituberculosis drugs among patients treated for active tuberculosis. American journal of respiratory and critical care medicine, 167(11), 1472-1477.

Yu, X., Wang, J., Hong, Q. Q., Teku, R., Wang, S. H., & Zhang, Y. D. (2022). Transfer learning for medical images analyses: A survey. Neurocomputing, 489, 230-254.

Zhang, C., & Lu, Y. (2021). Study on artificial intelligence: The state of the art and future prospects. Journal of Industrial Information Integration, 23, 100224.

Zhang, N., Savic, R. M., Boeree, M. J., Peloquin, C. A., Weiner, M., Heinrich, N., ... & Dooley, K. E. (2021). Optimising pyrazinamide for the treatment of tuberculosis. European Respiratory Journal, 58(1).

Zhang, X., Li, X., Feng, Y., & Liu, Z. (2015). The use of ROC and AUC in the validation of objective image fusion evaluation metrics. Signal processing, 115, 38-48.

Zhang, Y., Weng, Y., & Lund, J. (2022). Applications of explainable artificial intelligence in diagnosis and surgery. Diagnostics, 12(2), 237.

Zhang, Z. (2016). Introduction to machine learning: k-nearest neighbors. Annals of translational medicine, 4(11).

Zhang, Z. (2016). Introduction to machine learning: k-nearest neighbors. Annals of translational medicine, 4(11).

Zhang, Z., Srivastava, R., Liu, H., Chen, X., Duan, L., Kee Wong, D. W., ... & Liu, J. (2014). A survey on computer aided diagnosis for ocular diseases. BMC medical informatics and decision making, 14, 1-29.

Zhou, Z. H., & Zhou, Z. H. (2021). Semi-supervised learning. Machine Learning, 315-341.

Zhukovska, A., Zheliuk, T., Shushpanov, D., Brych, V., Brechko, O., & Kryvokulska, N. (2023, September). Management of the Development of Artificial Intelligence in Healthcare. In 2023 13th International Conference on Advanced Computer Information Technologies (ACIT) (pp. 241-247). IEEE.

Zurac, S., Mogodici, C., Poncu, T., Trăscău, M., Popp, C., Nichita, L., ... & Bastian, A. (2022). A new artificial intelligence-based method for identifying mycobacterium tuberculosis in Ziehl–Neelsen stain on tissue. Diagnostics, 12(6), 1484.

APPENDICES

APPENDIX A

Confusion Matrix FOR Microscopic Slide Images

Figure 42
MobileNet-SVM on MSI

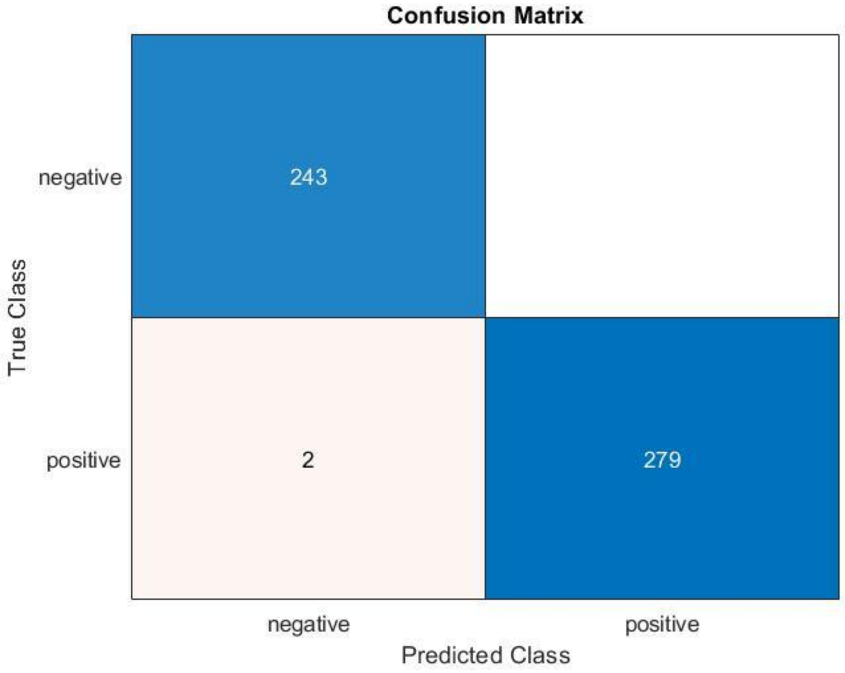


Figure 43
MobileNet-KNN on MSI

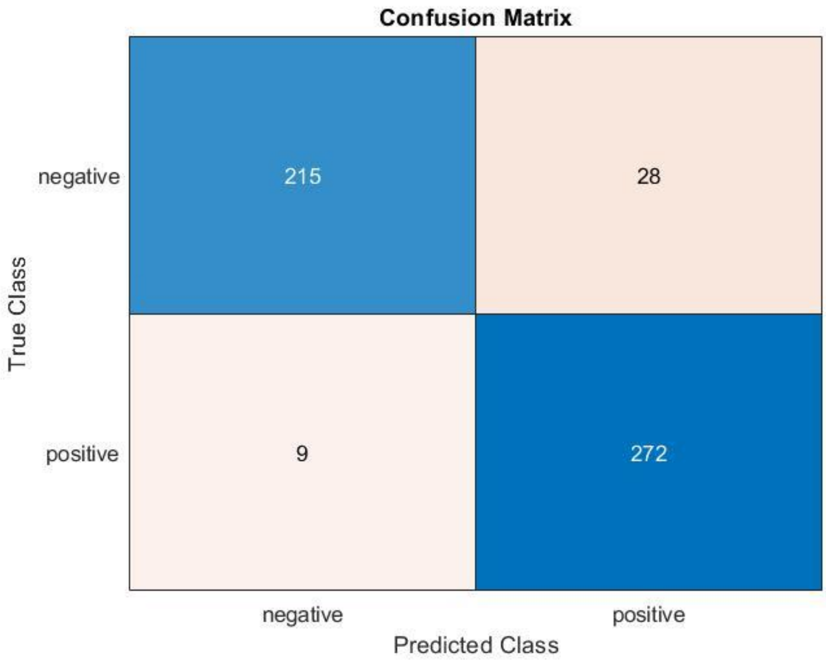


Figure 44

ResNet50-SVM on MSI

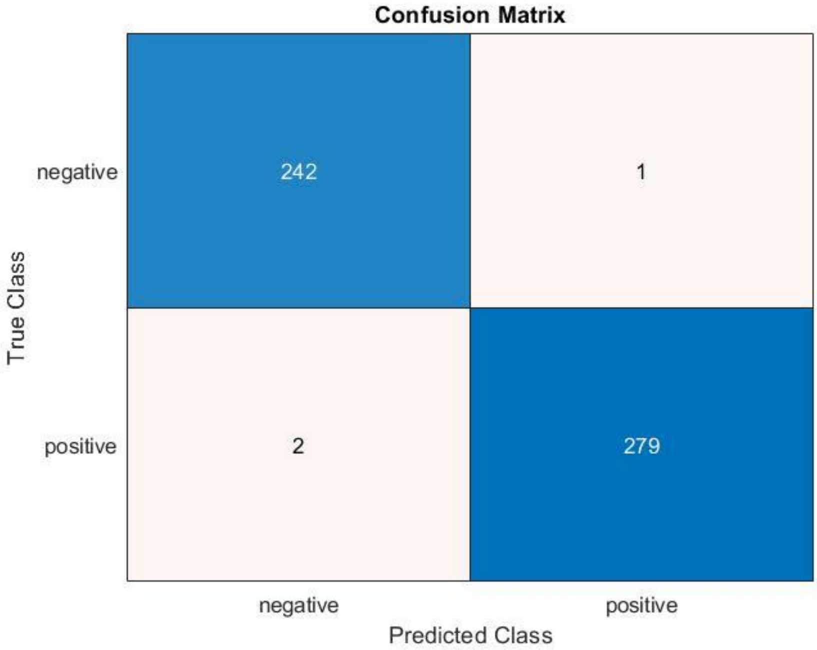


Figure 45

ResNet50-KNN on MSI

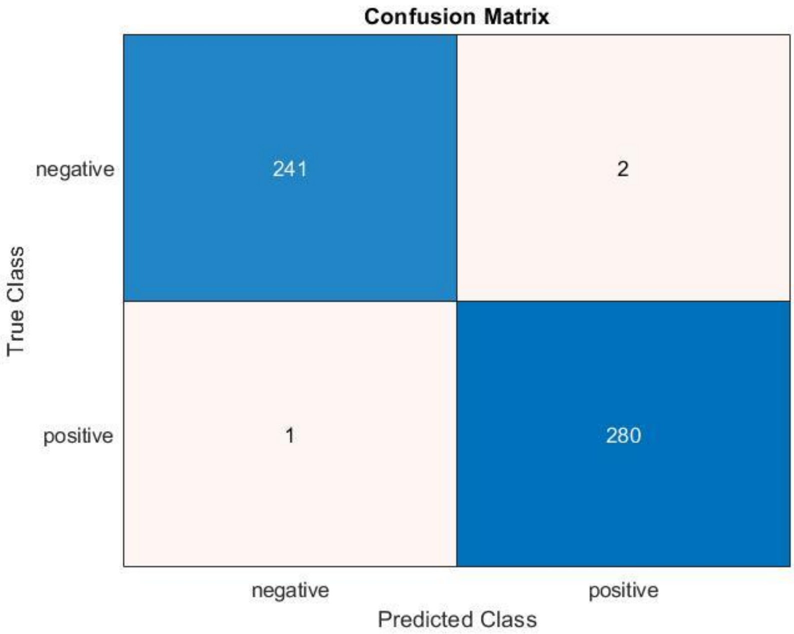


Figure 46

ResNet101-SVM on MSI

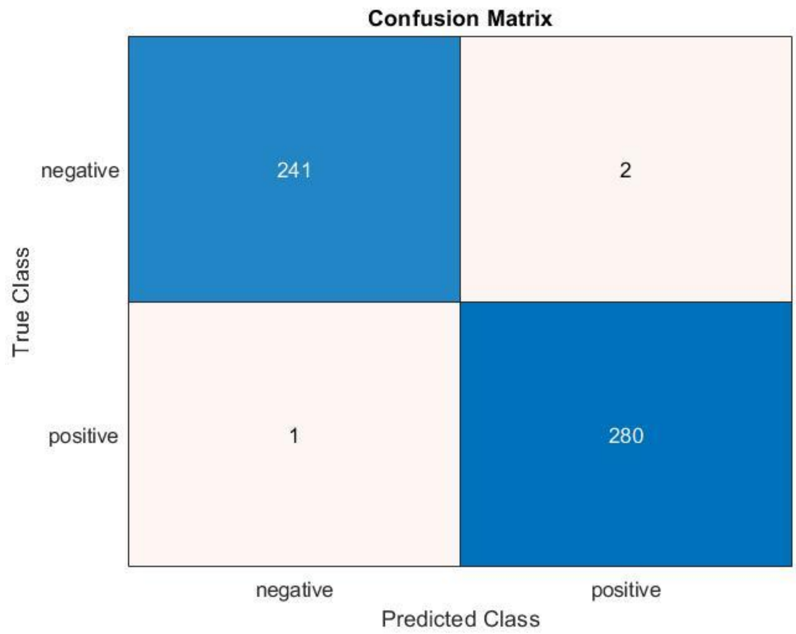


Figure 47

ResNet101-KNN on MSI

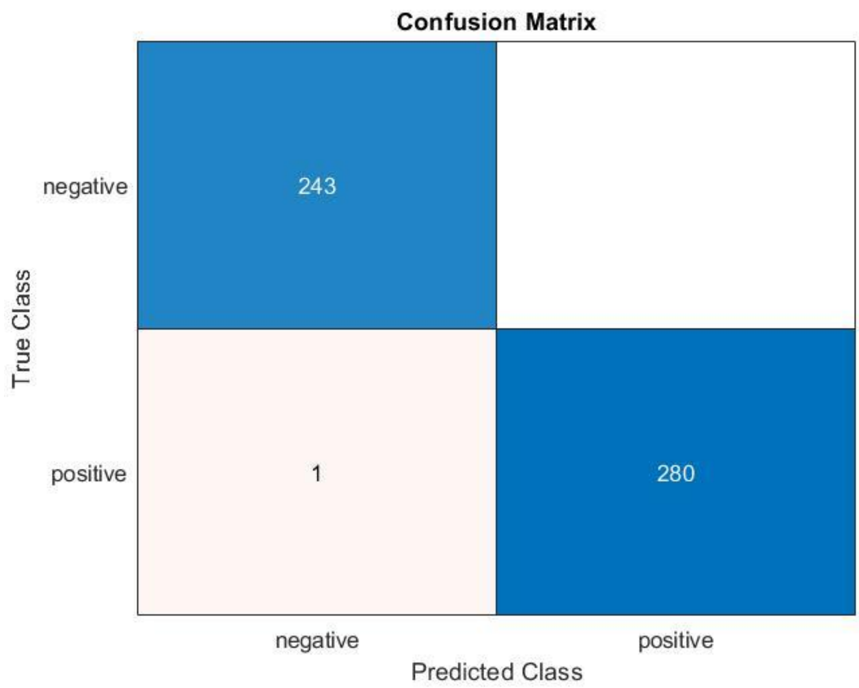


Figure 48

VGG16-SVM on MSI

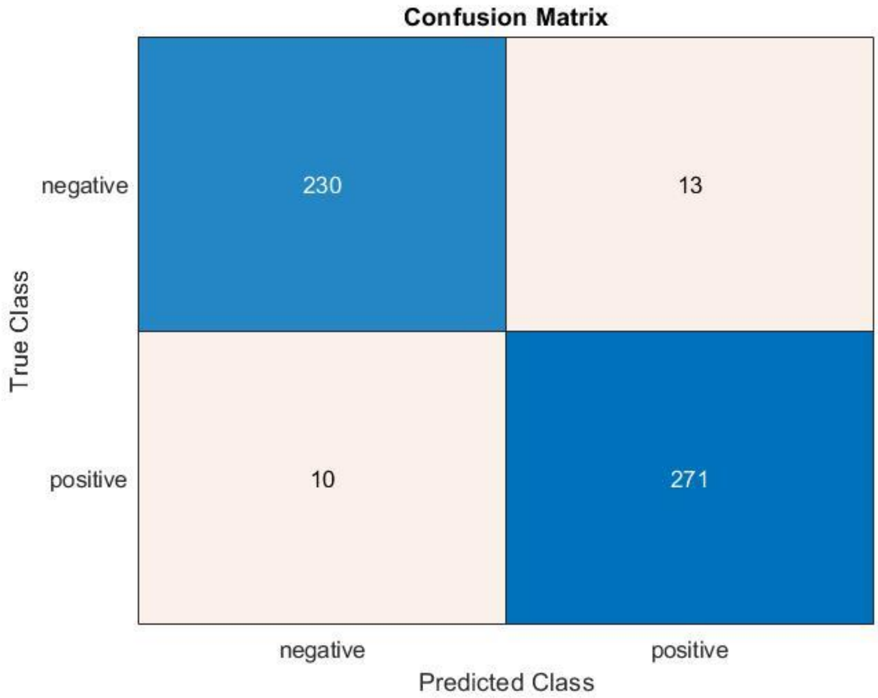


Figure 49

VGG16-KNN on MSI

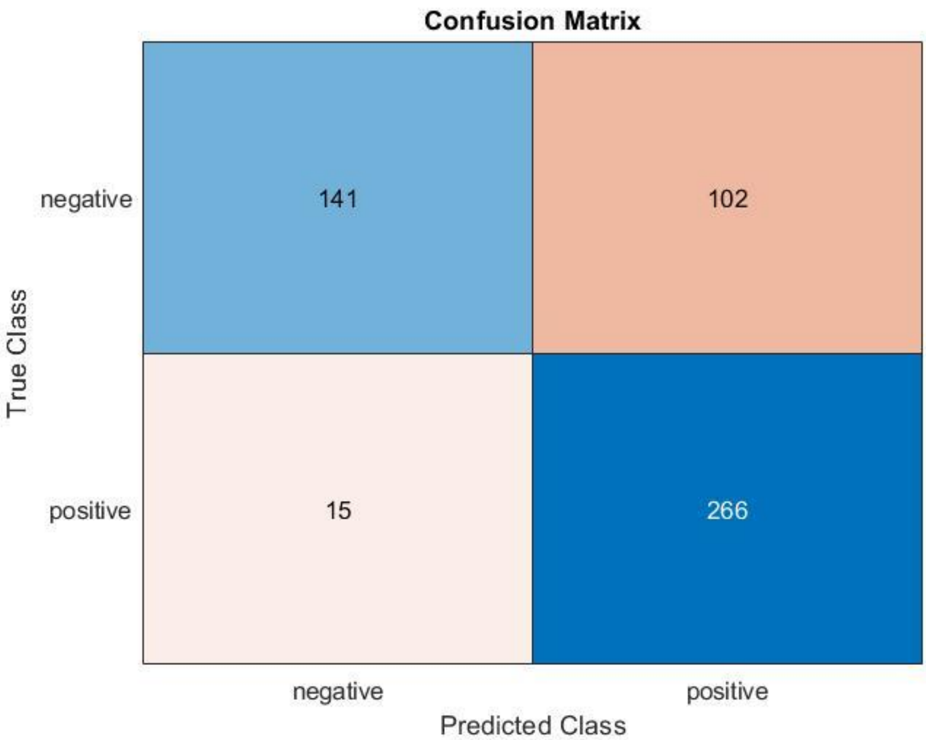


Figure 50

VGG19-SVM on MSI

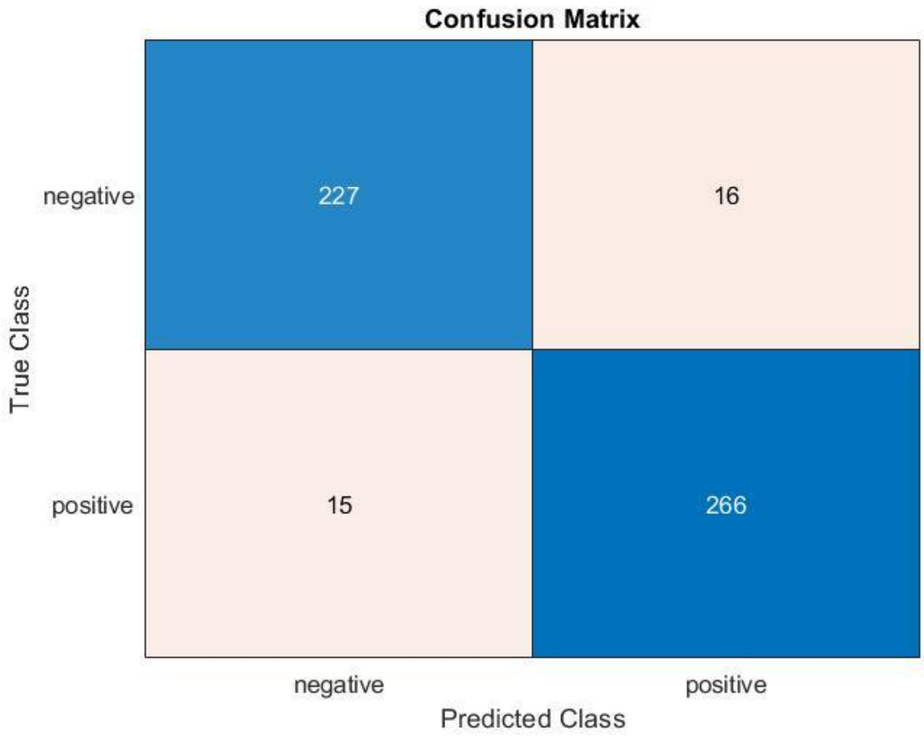
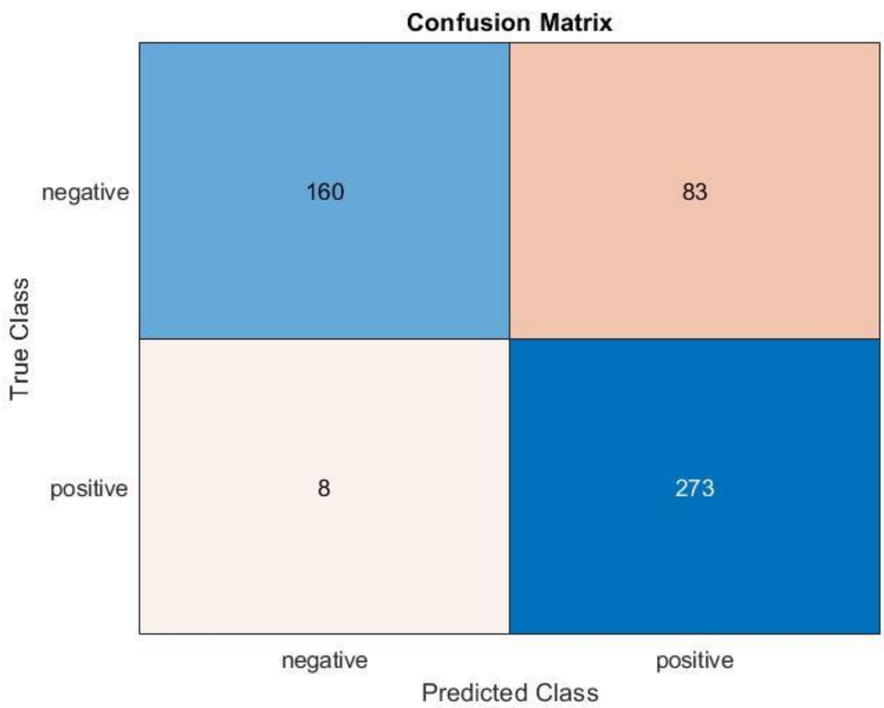


Figure 51

VGG19-KNN on MSI



APPENDIX B

Confusion Matrix For Chest X-Ray Images (CXI)

Figure 52

MobileNet-SVM on CXI

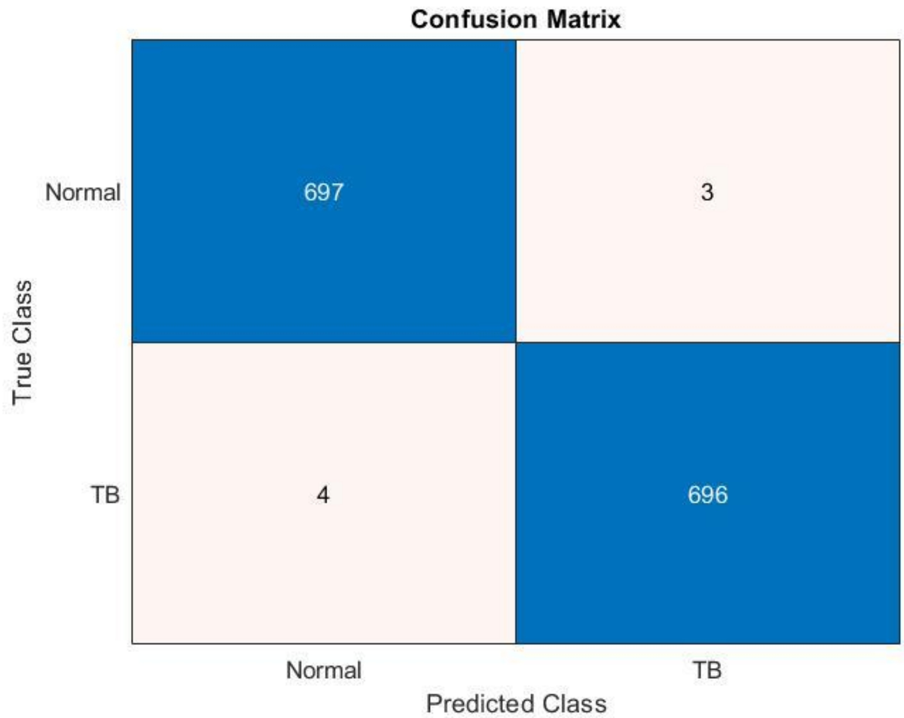


Figure 53
MobileNet-KNN on CXI

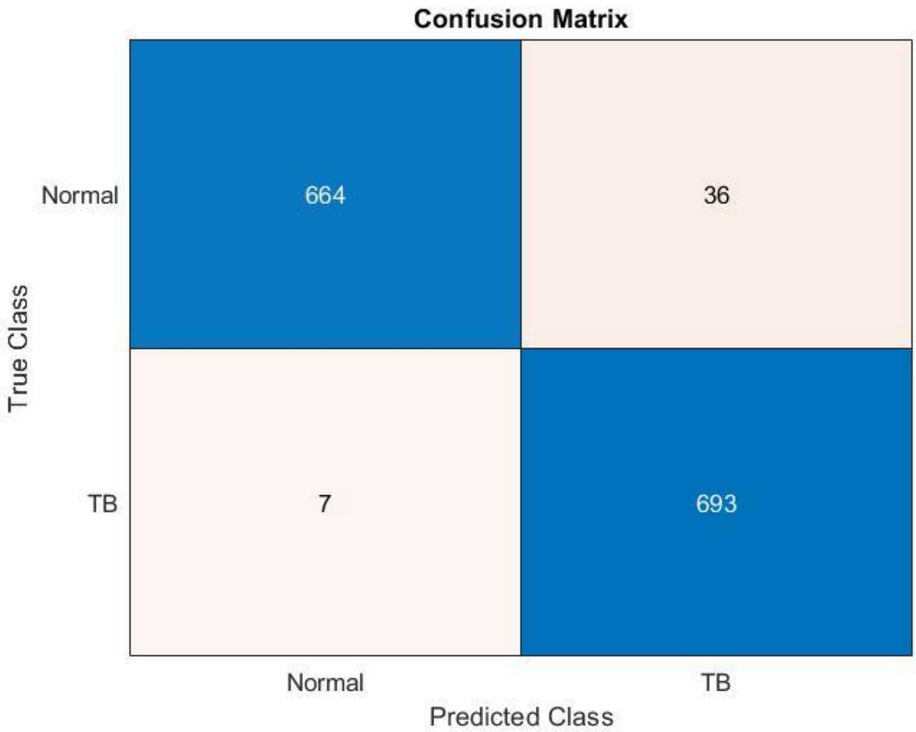


Figure 54
ResNet50-SVM on CXI

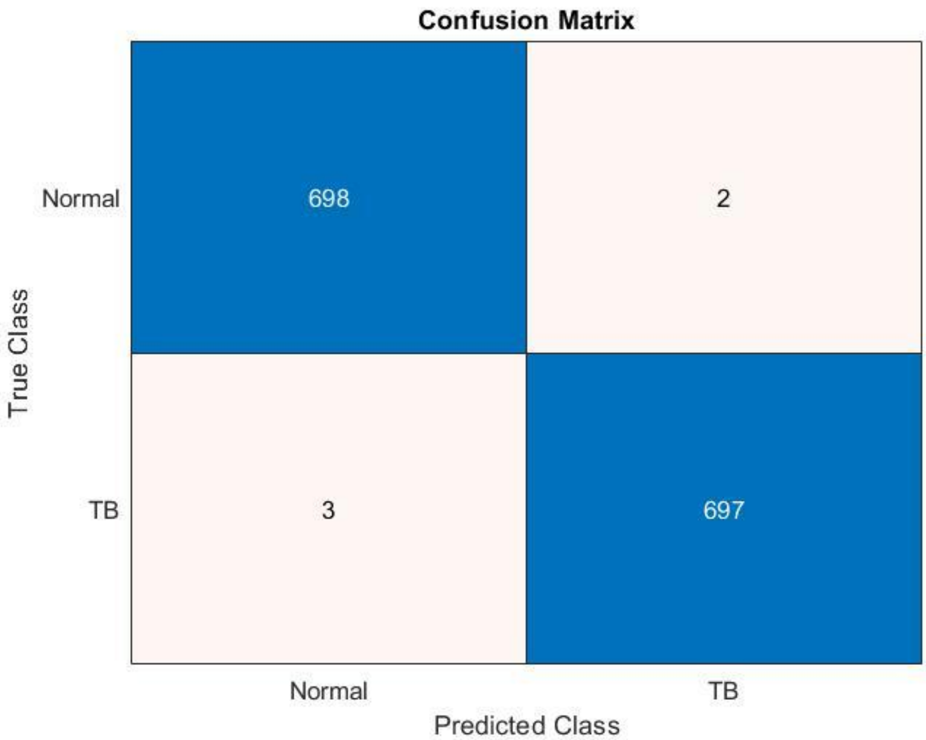


Figure 55

ResNet50-KNN on CXI

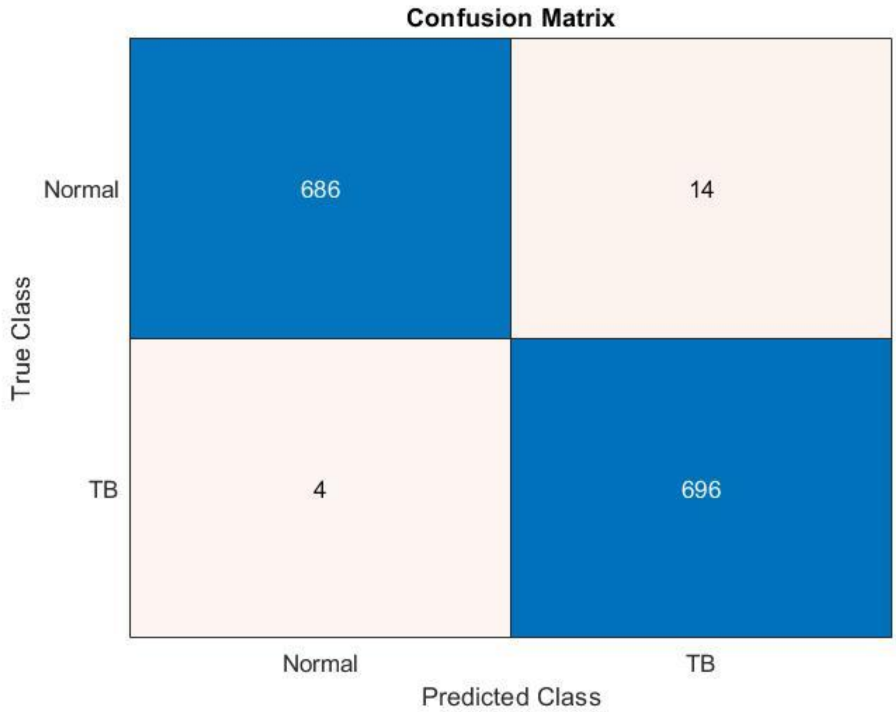


Figure 56

ResNet101-SVM on CXI

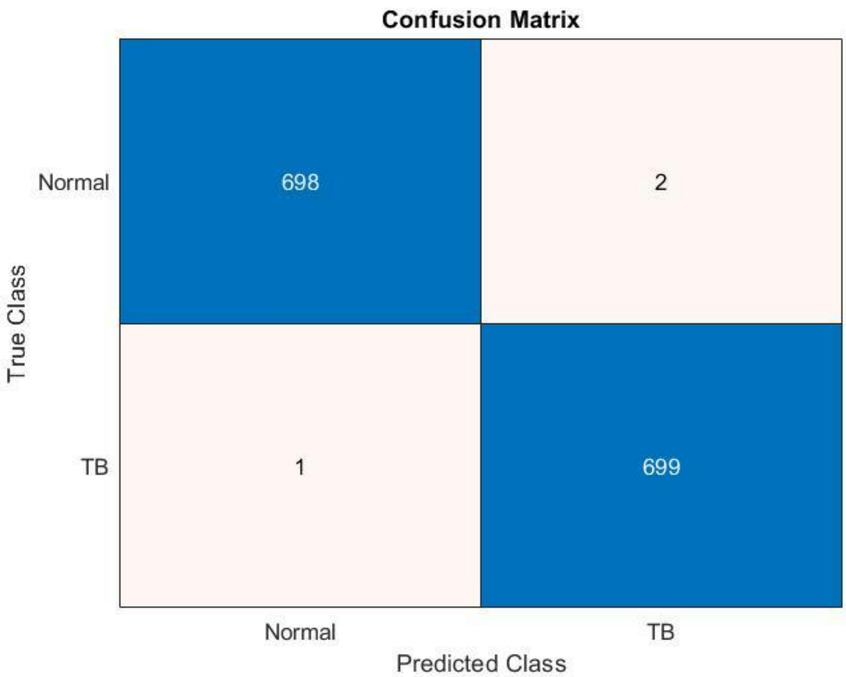


Figure 57

ResNet101-KNN on CXI

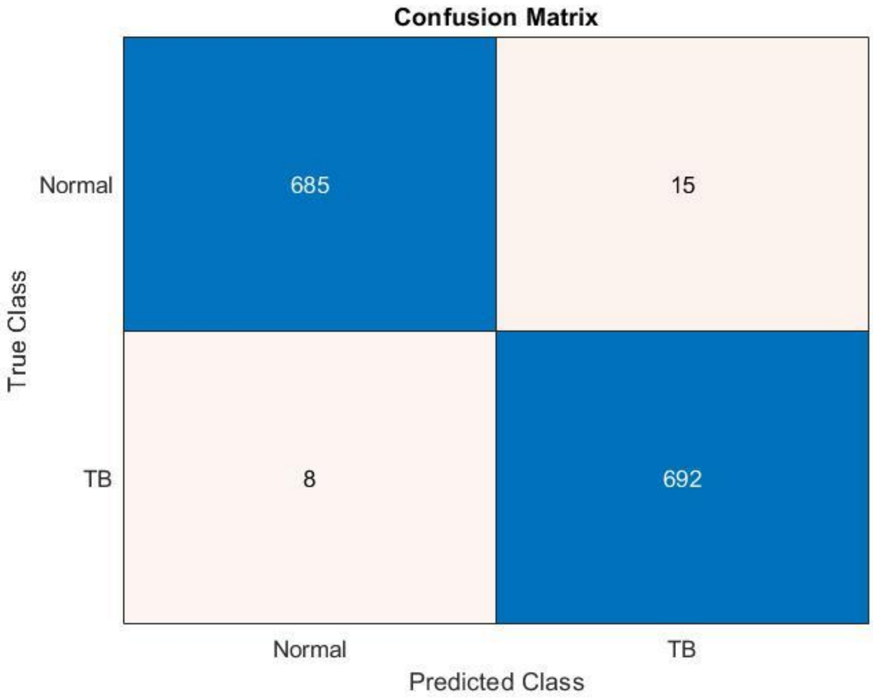


Figure 58

VGG16-SVM on CXI

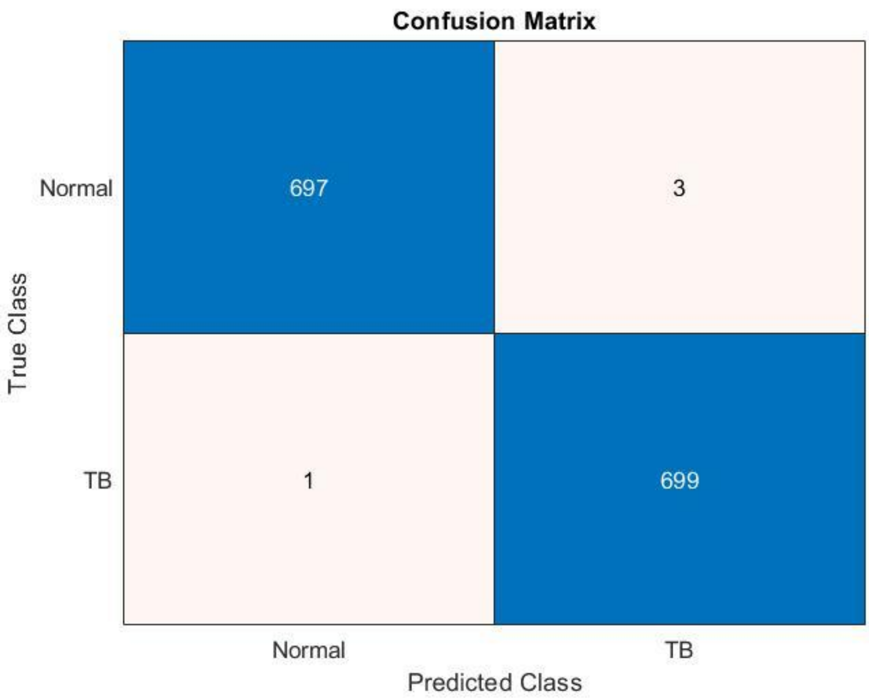


Figure 59

VGG16-KNN

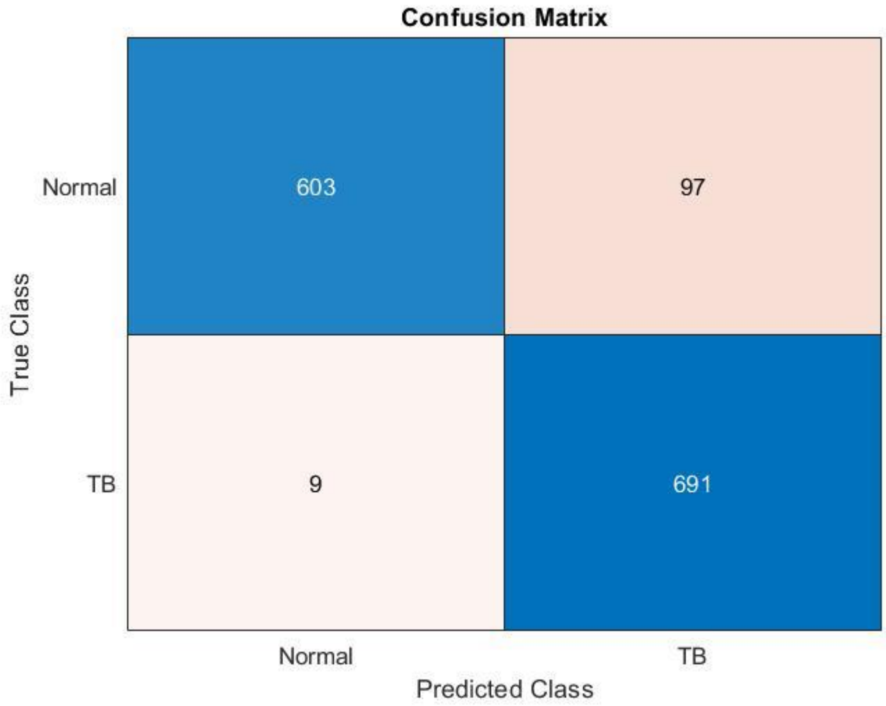


Figure 60

VGG19-SVM on CXI

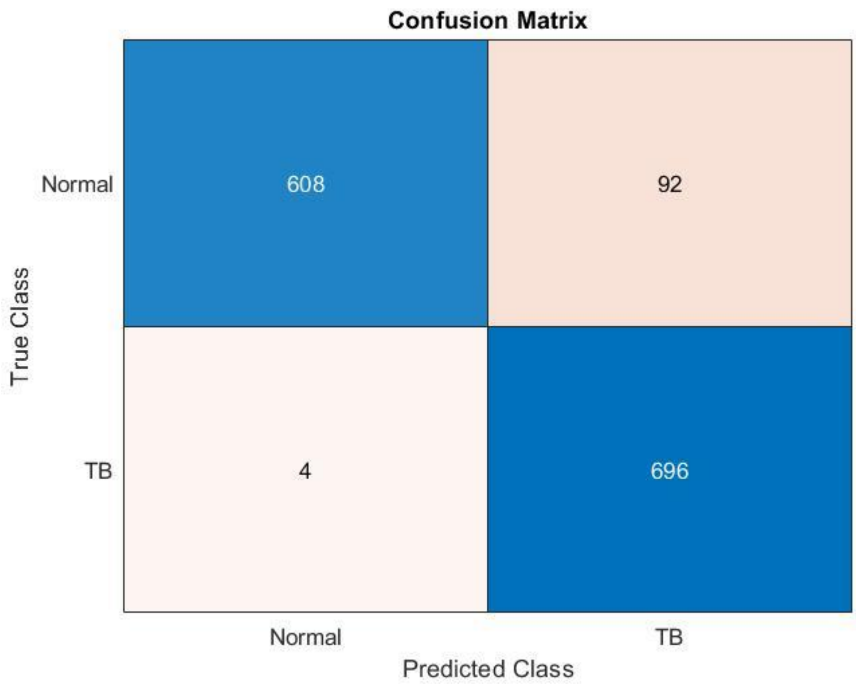
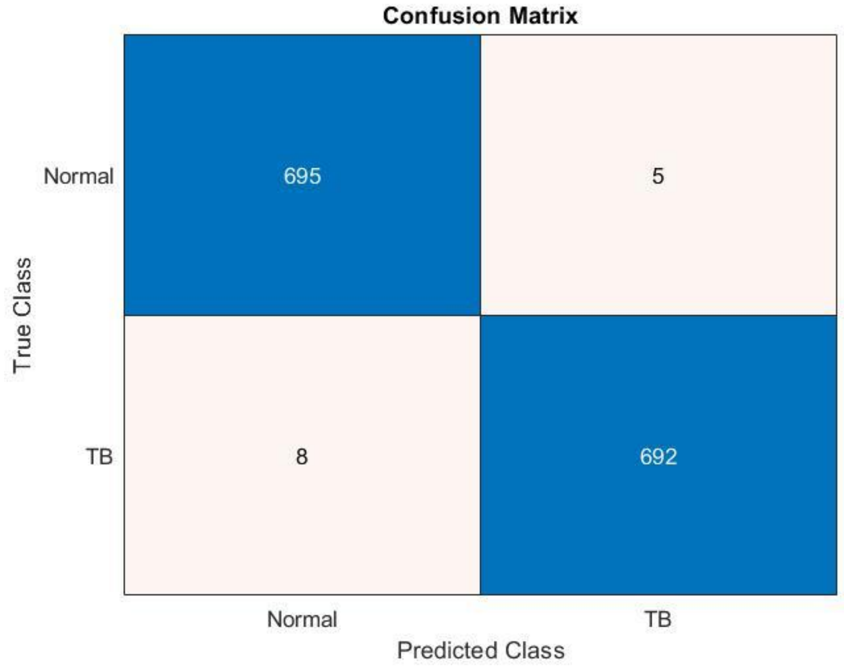


Figure 61
VGG19-KNN on CXI



APPENDIX C

Plagiarism Check

Yakın Doğu Üniversitesi

QUICK SUBMIT | NOW VIEWING: ALL PAPERS ▾

Submit

AUTHOR	TITLE	SIMILARITY	FILE	PAPER ID
Ibrahim Omoyayi Omod...	Abstract	0%		2559719497
Ibrahim Omoyayi Omod...	Conclusion	0%		2559722490
Ibrahim Omoyayi Omod...	Chapter I	6%		2559719500
Ibrahim Omoyayi Omod...	Chapter II	8%		2559724955
Ibrahim Omoyayi Omod...	All Chapters	9%		2559732666
Ibrahim Omoyayi Omod...	Chapter III	9%		2559731921
Ibrahim Omoyayi Omod...	Chapter IV	9%		2559722487