FIRAT UNİVERSİTY

GRADUATE SCHOOL OF NATURAL AND APPLİED SCİENCES

TÜRKİYE

CAD SYSTEMS IN MEDICAL APPLICATION: DETECTION OF TUBERCULOSIS FROM CXR-IMAGES USING CONVOLUTIONAL NEWURAL BASED NETWORKS (CASE STUDY: NİGERİAN PUBLİC HEALTH)

Muhammad Zaharaddeen ABUBAKAR

Master's Thesis

DEPARTMENT OF DİGİTAL FORENSİC ENGİNEERİNG

JULY 2022

FIRAT UNİVERSİTY

GRADUATE SCHOOL OF NATURAL AND APPLİED SCİENCES

TÜRKİYE

Department of Digital Forensic Engineering

Master's Thesis

CAD SYSTEMS IN MEDICAL APPLICATION: DETECTION OF TUBERCULOSIS FROM CXR-IMAGES USING CONVOLUTIONAL NEWURAL BASED NETWORKS (CASE STUDY: NİGERİAN PUBLİC HEALTH)

Author

Muhammad Zaharaddeen ABUBAKAR

Supervisor

Dr. Öğr. Üyesi Mustafa KAYA

JULY 2022

ELAZIG

FIRAT UNİVERSİTY
GRADUATE SCHOOL OF NATURAL AND APPLİED SCİENCES
TÜRKİYE

Department of Digital Forensic Engineering

Master's Thesis

Title: CAD Systems in Medıcal Applıcatıon: Detectıon of Tuberculosıs from CXR-Images usıng Convolutıonal Neural Based Networks (Case Study: Nigerian Public Health)

Author: Muhammad Zaharaddeen ABUBAKAR

Submission Date: 03 June 2022

Defense Date: 04 July 2022

THESİS APPROVAL

This thesis, which was prepared according to the thesis writing rules of the Graduate School of Natural and Applied Sciences, Fırat University, was evaluated by the committee members who have signed the following signatures and was unanimously approved after the defense exam made open to the academic audience.

Signature

Supervisor: Assist. Prof. Dr. Mustafa KAYA — Approved

Firat University, Faculty of Science

Chair: Prof. Dr. Galip AYDIN — Approved

Firat University, Faculty of Engineering.

Member: Assist. Prof. Dr. Hafzullah İŞ — Approved

Batman University, Faculty of Engineering

This thesis was approved by the Administrative Board of the Graduate School on

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

Signature

Prof. Dr. Kürşat Esat ALYAMAÇ
Director of the Graduate School

DECLARATION

I hereby declare that I wrote this master's thesis work titled "CAD systems ın medıcal applıcatıon: detectıon of tuberculosıs from cxr-ımages usıng convolutıonal neural based networks (Case study: Nigerian public health)" in consistent with the thesis writing guide of the Graduate School of Natural and Applied Sciences, Firat University. I also declare that all information in it is correct, that I acted according to scientific ethics in producing and presenting the findings, cited all the references I used, express all institutions or organizations or persons who supported the thesis financially. I have never used the data and information I provide here in order to get a degree in any way.

04 July 2022

Muhammad Zaharaddeen ABUBAKAR

PREFACE

Tuberculosis (TB) is a contagious bacterial infection that attacks the lungs region, a curable disease in which its early detection is pivotal to decrease the morbidity and mortality. The breakthrough of deep neural architectures and the availability of datasets in medical imaging has led to many researchers trying to come up with efficient ways through automated computer-aided diagnosis systems to serve as an alternative method to reduce the waiting time of patients and help clinicians and physicians to predict a more accurate result of the disease.

This study tries to optimize the binary classification performance of a densely affected region in Africa and to show the strength and weakness of different convolutional neural network techniques used in detecting active Tuberculosis on chest X-ray images together with ensemble methods on different datasets using several performance metrics after reviewing several studies. This research aims to help readers that are getting started or are interested in focusing their research on computer aided diagnosis system, precisely in TB with deep learning (CNN), and its techniques used for computer vision in medical field.

Special thanks go to my supervisor in person of **Associate Professor Mustapha KAYA,** for his immeasurable support and gratitude throughout my study and work. My profound gratitude goes to **Mustapha ERIS**, **Serkan KARKUS**, and all my and course mates and well-wishers; **Muhammad Mansur ABUBAKAR, Tanko DAHIRU, Bashir Zak, Bilyamin Muhammad, Khalid Jibril SANI, Dr. Abubakar Dahiru, Dr. Amina Sambo Magaji,** to mention just a few for their guidance and effort in making the research a success.

Special thanks go to my family, most especially my parent **Hajara USMAN**, **Hadiza Aliyu NGOBA**, and **Mr. Muhammad Ali,** for all their prayers, well wishes and support, may all your hidden prayers be answered.

Special thanks also go to **Dr Jiya** and **Mr. Abdul-Azeez Mahmud** of Nigerian **National Tuberculosis and leprosy Training Centre** and my sister **Doctor Fatima Zahra Sani Sambo,** as they helped in gathering the dataset I used for the research.

Finally, with utmost joy and happiness, I extend my special thanks to the former **CEO/DG Professor Isa Ali Ibrahim PANTAMI** for sanitizing and starting our batch and **Mr. Kashifu INUWA** for his continuous support after he took the office of his mentor at the Nigerian **National Information Technology Development Agency (NITDA)** and the whole scholarship committee for their fairness in the selection, and organization in exercising their duties, followed with continuous support to the end of my program.

Muhammad Zaharaddeen ABUBAKAR

ELAZIG, 2022

TABLE OF CONTENT

ABSTRACT

CAD Systems in Medical Application: Detection of Tuberculosis from CXR-Images using Convolutional Neural Based Networks (Case study: Nigerian Public Health)

Muhammad Zaharaddeen ABUBAKAR

Master's Thesis

FIRAT UNİVERSİTY

Graduate School of Natural and Applied Sciences

Department of Digital Forensic Engineering

July 2022, Page: x+ 37

Tuberculosis (TB) is a very dangerous contagious bacterial infection that usually attacks the lungs and has spread widely around the globe in the last 4 to 5 decades. Early detection of TB is pivotal to the decrease in morbidity and mortality. The diagnosis of TB is made through conventional methods such as blood tests, skin tests, CT scans, and sputum tests, which are tedious and take weeks or even more. Therefore, to lower the detection time and raise the accuracy of diagnosis, due to the time-consuming task that typically requires expert radiologists to read the test, leading to fatigue-based diagnostic error and lack of trained experts in areas of the world where radiologists are not available or limited. In recent years, the application of Artificial intelligence (AI) and the breakthrough of deep learning (DL) approaches like convolutional neural networks (CNN) architectures in object detection and classification have attracted many researchers to apply this sophisticated technology in different fields of science, primarily in the medical field to achieve expert-level performance in the interpretation of different diseases, which is powered by the availability of labeled digital dataset. Hence chest radiograph (CXR) has become critical in detecting thoracic diseases such as TB. This study used a public and private dataset of CXR images to build models that can effectively detect TB using pretrained deep neural networks. In our experiment, the best performing model was MobileNetV1 out of the models used, in which we were able to achieve an accuracy of 94%, specificity of 96%, sensitivity of 92%, with an f1-score, precision of 94% respectively on a private dataset we collected from a densely affected region of Africa, Nigeria to be precise. Meanwhile, the same model on a public dataset scored an accuracy of 100%, sensitivity of 99%, specificity, precision and F1-score of 100% respectively. Hence, the performance of the deep neural networks (DNN) model in our experiment shows the efficacy of DNN in detecting tuberculosis from CXR images.

Keywords: Tuberculosis (TB), Artificial intelligence (AI), Machine Learning (ML), deep learning (DL), convolutional neural networks (CNN), Binary Classification, Positive (+ve), Negative (-ve)

ÖZET

Tibbi Uygulamada CAD Sistemi: Evrensel Sinir Tabanli Ağlari Kullanarak CXR Görüntülerinden Tüberküloz Tespiti (Vaka Çalişmasi: Nijerya Halk Sağliği)

Muhammad Zaharaddeen ABUBAKAR

Yuksek Lisans Tezi

FIRAT ÜNİVERSİTESİ

Fen Bilimleri Enstitüsü

Adli Bilişim Mühendisliği Anabilim Dalı

Temmuz 2022, Sayfa: x + 37

Tüberküloz (TB), genellikle akciğerleri etkileyen ve son 40-50 yil içinde tüm dünyaya yayılmış olan çok tehlikeli, bulaşıcı bir bakteriyel enfeksiyondur. TB'nin erken tespiti, hastalik ve olumleri azaltmak için çok önemlidir. TB'nin teşhisi, kan testleri, cilt testleri, BT taramaları ve balgam testleri gibi zor ve ayni zamanda haftalarca sürebilen geleneksel yöntemlerle yapılmaktadir. Ayrica bu yontemlerle yapilan tanilarda cogu zaman bir uzmanin yapilan testleri yorumlamasi gerekmektedir. Bu durum ozellikle yeterli egitime sahip uzaman sayisinin yetersiz oldugu yerlerde, uzmanlarin daha fazla yorulmasina ve teshis suresinin uzamasina dolayisiyla hastaligin teshis surecinde hatalara sebep olabilmektedir. Tani suresinin kisaltilmasi bu tip hatalarin onune gecmek icin onemlidir. Son yıllarda, nesne tespiti ve goruntu sınıflandırmada evrişimli sinir ağları (CNN) mimarileri gibi derin öğrenme (DL) yaklaşımlarının atılımı, bu teknolojiyi farklı alanlarda uygulamak için birçok araştırmacıyı cezbetmiştir. Tıbbi alanda, etiketli dijital veri setlerinin olusturulmasi ve kullanılabilirliğinin artmasi ile farklı hastalıkların tespitinde uzman düzeyinde performans elde etmek için bu yontemler kullanilabilmektedir. Bu nedenle akciğer grafisi (CXR), TB gibi göğüs hastalıklarının saptanmasında kritik hale gelmistir. Bu çalışmada, oneğitimli derin sinir ağları kullanılarak tüberkülozu etkili bir şekilde tespit edebilen modeller oluşturmak için acik ve ozel CXR görüntülerinden olusan bir veri seti kullanilmistir. Deneyde, en iyi performansi gösteren MobileNetV1 modeli kullanilarak Afrika`da hastaligin cok yaygin oldugu Nijerya`dan topladigimiz ozel veri seti uzerinde sırasıyla %94 doğruluk (accuracy), %96 özgüllük (specificity), %92 duyarlılık (sensitivity) ve %94 f1 skoru (F1 Score) ile hassasiyet degeri elde edilirken acik veri seti uzerinde ise %100 dogruluk (accuracy), %99 ozgulluk (specificity) ve %100 duyarlilik (sensistivity),fi skoru (F1 Score) ve hassasiyet (precision) degeri elde edilmistir. Elde edilen sonuclar, DNN'nin CXR görüntüsünden tüberkülozu tespit etmedeki etkinliğini göstermektedir.

Anahtar Kelimeler: Tüberküloz (TB), Yapay zeka (AI), Makine Öğrenimi (ML), derin öğrenme (DL), evrişimli sinir ağları (CNN), İkili Sınıflandırma, Pozitif (+ve), Negatif (-ve)

LİST OF FİGURES

Page

LIST OF TABLES

Page

ABBREVİATİONS

AI : Artificial intelligence
ANN : Artificial Neural Network
CAD : Computer-aided diagnosis
CNN : Convolutional neural networks
CT : Computed tomography
C-DL : Custom Deep Learning
CXR : Chest X-ray or Chest Radiograph
DL : Deep learning
DNN : Deep Neural Network
FNR : False Negative Rate
FN : False negatives
FP : False positive
FPR : False positive rate
DOTS : Directly Observed Treatment, Short Course
H-CNN : Hybrid CNN
IGRA : Interferon Gamma Release Assay
TB : Tuberculosis
TL : Transfer Learning
TN : True negative
TP : True Positive
MC : Montgomery County
ML : Machine learning
SZ : Shenzhen
TPR ; True Positive Rate
TST : Tuberculin Skin Test
NLM : National Library of Medicine
JSRT : Japanese Society of Radiological Technology
KIT : Korean Institute of Tuberculosis
KNTA : Korean National Tuberculosis Association
SNUH : Seoul National University Hospital
TJ-UH : Thomas Jefferson University Hospital
PUH : Peking University Hospital
GFJH : GF Jooste Hospital
PIH : Partners in Health in Peru data

1 INTRODUCTİON

Tuberculosis (TB) is a contagious bacterial infection that usually attacks the lungs, which is known as Mycobacterium Tuberculosis. Morbidity and mortality of the disease can be reduced if it's been detected at its early stage. TB is a curable disease which is mostly cured using the global recommended strategy for TB control, known as DOTS (Directly Observed Treatment, Short Course), a strategy that Nigeria and other largely affected countries follow. The DOTS strategy is for patients on medication to have support from individuals to observe them while been on medication [1].

TB was declared a global public health emergency issue in 1993 after being initially neglected, and the DOTS strategy was launched as part of indicators in Millennium Development Goals; of the Stop TB Strategy [2] and [3]. TB is one of the top 10 leading causes of death globally, the first leading cause from a single infectious disease, and the leading killer of HIV positive people. In 2016, 10.4 million people approximately fell ill with TB and over 25% of those people are from Africa [4], [5], [6], and [7]. 24th March is a special day selected out in there hundred and sixty-five days (365) globally to commemorate and enlighten the public on one of the world's deadliest diseases [8], and its social and economic consequences as a measure in controlling and bringing an end to the global epidemic especially now as the maiden Covid-19 has put more people at increasing risk. TB is one of the major public health problems in Nigeria, a country of 169 million inhabitants, with the country currently ranking among the top 10, out of 22 high TB burden countries of the world and fourth highest in Africa (after South Africa, Ethiopia and DR Congo) [9] and [10]. Due to the rate of the disease in the region, TB is considered as a national emergency problem in Nigeria.

However, Various approaches of automated detection have been attempted in the last decade, and in recent years, Machine learning (ML) is not a strange term for the general public, ML based techniques with deep learning (DL) are attracting lots of attention in big data analysis. DL algorithms have become a popular choice for solving biomedical diagnostic problem [11], and [12], whereby data is filtered through a cascade of multiple layers using chest X-ray (CXR) images. The most recommended method of TB diagnosis is by using computer-aided diagnosis (CAD) framework, which plays a significant role in the mass screening of TB by classifying the CXR images, helping expert radiologists to read the images which is a time-consuming task that leads to fatigue-based diagnosis, error-prone, and lack of availability of diagnostic expertise.

DL techniques, such as convolutional neural networks (CNNs) have demonstrated great promise and performance in image classification and is now widely adopted by the research communities [13]. This models have been applied in different field of medicine such as; brain tumor

[14], breast cancer [15], Covid-19 [16], Heart failure [17], Diabetes detection [18], detection of thoracic pathologies [19].

1.1. Related Works

This sub-section, reviews different studies that have used either custom-based deep convolutional neural networks DCCN architectures, single pre-trained DCCN architectures, an ensemble of those DCCN architectures, or classical methods to detects of TB in different dataset.

F. Pasa et al. [20] in their study, build a Custom DCNN model that consists of five (5) convolutional blocks, where each block contains two (3 by 3) convolutions with ReLUs, followed by a (3 by 3) pooling layer with two strides similar to AlexNet [21] operation, and the whole block is followed by a global average pooling layer and a fully-connected SoftMax layer with two outputs. To preserve the input resolution, they zero-padded all the convolutions. Each convolutional layer also uses batch normalization to speed up the training procedure and reduce overfitting. They claim their model has an advantage over the second smallest network (GoogLeNet [22]) as it comes with 230,000 parameters compared to the 7million of GoogLeNet and hence reduced the computational and memory requirement significantly as in MobileNet [23], without sacrificing the classification performance. Computational complexity was also reduced compared to other deep learning models as their model achieved only a 4-fold reduction without affecting classification accuracy. They believe their model can be competitive, even when it is difficult to compare with other techniques. They also argued that saliency maps and grad-CAMs are useful tools for clinicians' in-depth visual explanation of tuberculosis, hence providing an approximate visual diagnosis that was never applied in most studies to the best of their knowledge. They finally claimed that increasing the amount of training data or performing pre-training on a bigger dataset would be possible to obtain similar results to those that have performed well in TB detection and achieve the goal behind our study.

Y. Xiong [24], in their study also designed a custom DCNN model called (TB-AI) intended to detect TB bacillus. The model was initially pretrained on the google CIFAR-10 dataset with the help of transfer learning (TL). The custom model was trained using 201 samples which were collected as a test set to examine TB-AI. Further pre-processing techniques such as image augmentation was also applied on the dataset to increase the number of training images to mitigate the occurrence of type one error and type two error having few or unbalanced datasets, even if they claimed to have a human pathologist on ground to analyze inconsistencies between there model and human, hence adjust the protocol accordingly which will slow the process of automation. In their study they compared the diagnosis of TB-AI to the ground truth result provided by a human pathologist and analyzed inconsistence between TB-AI and humans, hence adjusted the protocol accordingly where TB-AI achieved 97.94% sensitivity and specificity of 83.65%.

E. J. Hwang et al. [25] developed a custom DCNN-based automatic detection (DLAD) model of 27 layers with 12 residual connections, trained via a semi-supervised localization approach where only a portion of the training data was annotated. The last layer of the model was made up of image classification and lesion localization (lung segmentation), using 60,768 CXR images collected from Seoul National University Hospital (SNUH) to classify TB. Further performance of their model was tested on six independent datasets to confirm the generalization performance. Certified radiologists labeled and annotated the CXR images, hence the model's performance was further observed by other physicians. High cutoffs were set for Sensitivity and specificity, where they scored 98% through in-house validation, which demonstrated excellent performance in TB detection, outperforming physicians and several other studies such as P. Rajpurkar et al. [26].

Meanwhile, some other studies used pretrained DCNNs with the concept TL to binary classify TB into positive and negative. T. Rahman et al. [27] were able to detect TB and score state-of-the-art performance ahead of other studies reviewed, using some image preprocessing techniques, data augmentation, and image segmentation using two different U-Net models with the help of three available public datasets (NLM, Belarus, and RSNA). They used nine (9) pre-trained DL CNN techniques with the help of TL, from their pre-trained initial weights, they train, validate, and test to obtain a binary classification of TB from the combined datasets, where the result yielded the best results compared to other studies under review.

S. Hwang et al. [28] also believe DCNN for TB screening is a promising DL method for various visualization tasks since they enable end-to-end training from feature extraction to classification, as agreed with other various studies without the need for manual feature extraction. Hence, they developed a custom DCNN for automatic TB screening on AlexNet and used TL to binary detects TB infections on the three available public datasets. They scored performance of 96%, 93% and 88% in terms of AUC for three real field datasets. The study also shows that TL of pre-trained networks resolves the difficulties in handling high-resolution medical images and training huge parameters with a limited number of CXR images, and also states that transferring low-level filters from a pre-trained model based on large-scale images is very effective for training.

Also, an experiment was carried out using two different DCNNs, which were AlexNet [21] and GoogLeNet [22] in a study by P. Lakhani and B. Sundaram [29] where they used four deidentified HIPAA-compliant datasets that were exempted from review by the institutional review board (which is believed to be one of the problems of a medical dataset). They binary classify their dataset as either TB positive or negative using untrained (i.e., training from scratch) and pretrained networks on ImageNet with the concept of TL, which yielded better accuracy. As common to all the studies under review, augmentation techniques and other multiple preprocessing techniques were used to further increase their accuracy compared to studies like [13], [30] et al. Ensembles were further performed on the best-performing algorithms using AlexNet and GoogLeNet

architecture with AUC score of 99%. An independent board-certified cardiothoracic radiologist interpreted the result without seeing the model result for cases with type one error. Their study shows that their model performed better after the various preprocessing employed on the DCNNs architecture better than that of the other studies in this review study except [27] which had better performance than theirs after performing an ensemble of nine different DCNNs. Their study also shows that there is a need for cardiothoracic radiologist experts to further give verdicts on results even with CAD systems.

M. H. A. Hijazi et al. [30] also used the MC and SHZ public dataset, where they first preprocessed the CXRs images to retain only the Region Of Interest (ROI) and further used two different pretrained architectures; VGG 16 and InceptionV3, to develop a custom CNN architecture with 15 layers to detect TB from the preprocessed CXR images, where the process took less time to complete training. Furthermore, an ensemble of the architectures was performed, scoring the best accuracy of 91.0%, sensitivity of 89.6%, and specificity of 90.7%. than the individual architecture.

R. Hooda [31] presented an ensemble of three standard architectures used for binary image classification with AlexNet, GoogLeNet, and ResNet all trained from scratch. The dataset is a combination of four public datasets consisting of 1133 images that's were pre-processed using augmentation to increase the number of images as in [32] and other similar studies. Their ensemble model achieves an accuracy of 88% and AUC of 93%. They believe that in a DL based CAD system, segmenting the lung region is not mandatory as opposed to [27],

S. K. T. Hwa, M. H. A. Hijazi et al. [33] also used the publicly available MC and SHZ datasets similar to other studies under review, in which they first pre-processed the CXRs images to obtain edge features using the Canny Edge detector, with the belief that CXR images with more unusual edges can increase the classification detection rate. They perform an ensemble of two pretrained DCNN architectures as in [32] and [30]. They believe sensitivity is considered more in medical image analysis; hence their work was to get a high sensitivity score.

1.2. Motivation

The motivation behind this study is to use private dataset that was collated from a highly affected region and to compare it with a more prepared public datasets from a different region, hence examine both regions dataset by using different pretrained architectures, custom DCCNs algorithm, and classical machine learning methods to come up with a method that when deployed will be able to assist physicians in the diagnosis of TB in affected populations.

1.3. Innovations and Contributions

Most of the studies in the literature used different techniques for binary classification of CXR images using different public and private datasets and different techniques such as custom DCNNs, TL to binary classify TB.

In our study, we present the use of several pre-trained DCNNs architectures to make binary classifications on our private dataset from West Africa (Nigeria in particular), as most of the studies in the literature are from Europe and Asia. We further compare our results using the same models on a publicly available dataset from Europe and Asia. Hence, we achieved an accuracy of 94% for both VGG16 and mobile Net version one (MobileNetV1) as the best performing models, with precision, recall, an f1-score, specificity, and sensitivity at 94% respectively for VGG16, and 94% precision, 94% recall, 94% f1-score, 96% specificity, and 92% sensitivity for MobileNetV1. Whereas for our public dataset, we achieved 99% accuracy, sensitivity, specificity, precision, recall, and f1-score respectively with VGG16 and VGG19, and we achieved 99% sensitivity, with 100% accuracy, specificity, precision, recall, and f1-score respectively with MobileNetV1, as our best performing DCNN models while using the public dataset.

The concept of the study is presented in section 2. Hence the materials gathered for the study and the proposed approach are presented in Section 3 and 4 respectively. Meanwhile, our results and discussions were presented in Section 5. Finally, the conclusion and future work are also merge in Section 6.

2 CONCEPT DEFİNİTİON

In this section we briefly reviewed the knowledge behind our study in other to help readers have a better understanding of some various healthcare terminologies, CAD, and deep learning related to our work. We first briefly reviewed the knowledge and understanding of these concepts, and further cursorily discussed several classification techniques used in the different DCNNs architectures that have been used in different studies in the field of study and healthcare at large.

2.1. TB Screening Techniques

The screening tests aid in determining only if a person is infected with TB bacteria. The test only detects M. tuberculosis infections but does not distinguish between latent and active infections. The two types of tests are; Blood tests and skin tests. Further tests, such as Imaging screening, and a sample of sputum, are performed to see the possibility of a person been infected by TB disease. Chest radiography imaging screening technique is the most common type of imaging examination in the world [19], with over 2 billion procedures performed each year. This technique is critical for screening, diagnosis, and management of thoracic diseases, which is generally accepted by both healthcare experts and ML experts.

2.1.1. Blood Test

Interferon Gamma Release Assay (IGRA) also known as TB blood test, is one of the most common ways of detecting TB germs in the body. The TB blood test can be performed as the first precautionary move rather than the TB skin test (Mantoux). The blood tests come in two types; QuantiFERON®-TB and T-SPOT®-TB. It is advisable to have the blood test after having frequent close contact with someone with the disease or must have lived in a country where many people have TB, or places such as a nursing home, clinic, hospital, prison, or after diagnosed with HIV infection or having a week immune system. Where the result of the test is either “negative” or “positive”. Further investigations like the CXR to look for foreign bodies on the lungs are taking

2.1.2. Skin Test

Mantoux tuberculin skin test (TST) is referred to as the TB skin test. A TB skin test requires several visits to a health care provider. The test is done by injecting a fluid called tuberculin into the skin. which consists of two procedures; Firstly, the injection is giving, and then secondly within a speculated time, 48-72 hours a trained health care worker reads the reaction on the skin (forearm) as seen in figure 2.2. When the result is positive, the punctured forearm will be sensitive to the

injection, and a small hard red bump will swell at the site of the injection usually between 48-72 hours, whereas if you are not infected by the disease the skin will not react to the TST, therefore its recommended to be diagnosed again at a later time because the disease sometimes take time to develop. However, if the person under test has a strong skin reaction, further diagnosis such as CXR imaging is required.

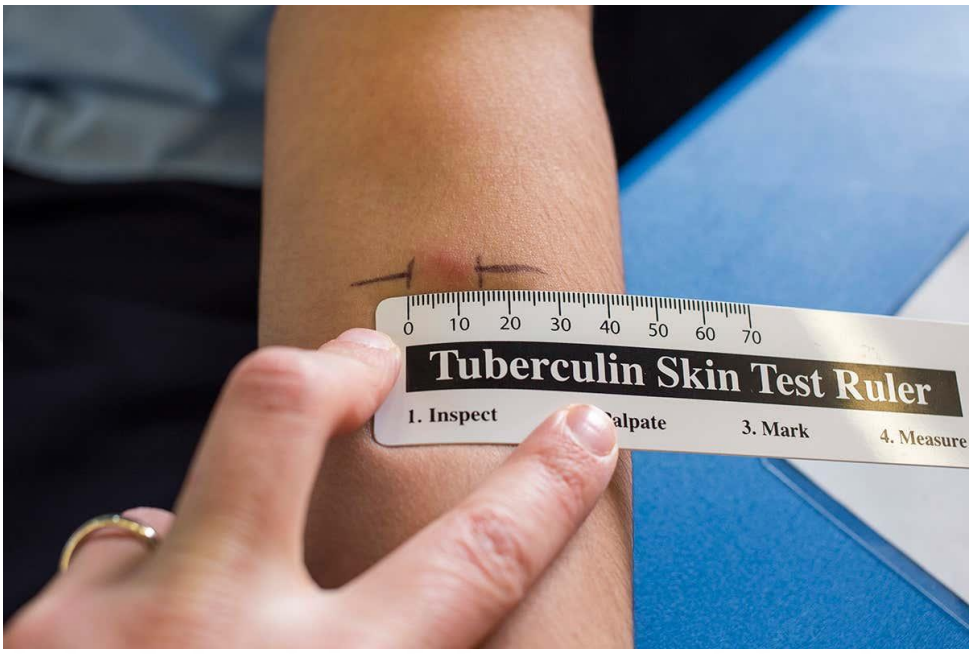

Figure 2.1 TB Skin test measuring technique by a health worker [34]

Figure 2.1 shows the injection of tuberculin which is one of the skin test measuring techniques of TB adopted worldwide.

2.2. Imaging Techniques

Following a positive skin test, an imaging technique is performed 7ort he detection of TB and its diagnosis for medical treatment using the CAD system; which is the process, procedure, and technique of visual representations of the inner part of the body. The test imaging techniques can see direct visualization of the foreign bodies in patient and hence observe the very small areas of anatomical and biological processes with different parameters. Different techniques have been developed for use in medical imaging and these techniques can be applied in scientific and industrial applications. There are several types of these imaging techniques that assist physicians in the diagnosis of TB which helps in choosing ideal treatment such as; CXR imaging, CT-Scan, Sputum test, et al. Each of the imaging techniques create images that assist in identifying some certain medical complications, not to forget that in medical image analysis, one of the bottlenecks is the lack of datasets being publicly available[31].

2.2.1. CXR

Chest radiography commonly called chest X-ray (CXR). It is commonly used to detect and evaluate abnormalities in the thoracic cavity. CXR detect some abnormalities in the heart, aorta, and the bones of the thoracic area. It is the most common imaging technique commonly used in diagnosing thoracic disease. The lungs, filled with air, are easily penetrated by electromagnetic radiation that passes through solid paths and output black images of the chest. It yields information about pulmonary, cardiac, and skeletal systems. It aids in the evaluation of cardiac, respiratory, tumors, and skeletal structure within the lung cavity and diagnoses multiple diseases such as pneumonia, congestive heart failure, Heart failure, Lung cancer, Emphysema, foreign body, TB, and so on. In recent times CXR is the most common type of test imaging used for computer vision in medical image analysis.

Figure 2.2. CXR Image [35]

Figure 2.2 shows the visual representation of CXR-Image after examination.

2.2.2. CT-SCAN

Computed tomography (CT) scan has revolutionized diagnostic medicine. Subsequently, the speed and accuracy of machines have improved tremendously. CT scan reveals the soft tissues and bones in the human body. Different annotations on the image can be adjusted to show tissues of similar weight, with the help of graphics software, and hence the result from multiple cross-sections can be assembled into 3-D images. CT also helps in the diagnosis and surgery.

2.2.3. Sputum Test

Sputum is a way of clinical specimen examination that produces mucus (thick, cloudy and sticky not saliva that comes out from the mouth, thin, clear, and watery) that is produced when a person cough. Healthcare workers collect sputum samples and test them for various strains of TB disease, including antibiotic-resistant types. The specimen is cultured in a laboratory specified for the testing. The results of sputum aid in choosing the best course of diagnosis. Sputum tests have a high risk as the aerosol produce is hazardous when it gets in contact with other patients or healthcare workers at close range. As such below are some generally accepted sputum collection method;

a) Coughing; This is the most common method of specimen collection.
b) Induced sputum; This is a method that helps patients that have difficulty producing sputum, the deep sputum-producing coughing is induced by inhalation of sterile, warm aerosol, hypertonic saline.
c) Bronchoscopy; This allows visualization of a person's inner airways. Which are the bronchial tubes. It's recommended when results have been nondiagnostic and doubt exists as to the diagnosis. Bronchoscopy is inserted through the mouth or nose directly into the desired portion of the lung, and the sputum is extracted.
d) Gastric aspiration Tube is inserted through the patient's mouth or nose and passed into the stomach to get a sample of gastric secretions that contain the sputum that has been coughed into the throat and then taken back, this procedure is mostly useful for diagnosis in children, who are often unable to cough up sputum.

2.3. Artificial Intelligence and Machine Learning

Artificial Intelligence (AI) is a branch of computer science that deals with developing smart machines that can perform tasks that normally require human intelligence [36]. AI encompasses Machine learning (ML), and deep learning as shown in Figure 3. ML was introduced in 1959 by Arthur Samuel to describe one of the technologies of AI, its provides the ability to automatically learn and improve from experience without been explicitly programmed [12].

AI and ML evolution in the past few decades have both changed the world around us, with their breakthrough innovations. The various techniques of deep learning (DL) have taken ML to a whole new level, where machines can learn to discern tasks, inspired by the human brain's neural network.

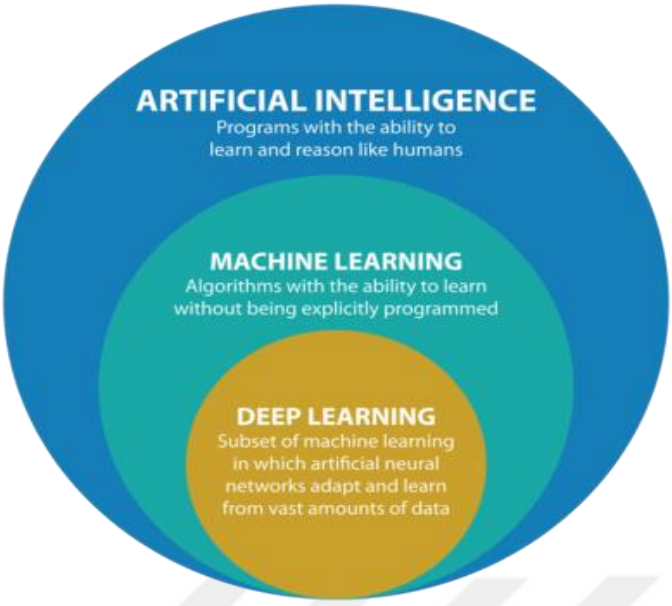

Figure 2.3. Artificial Intelligence, Machine Learning and Deep Learning

Figure 2.3 shows how AI, ML, and DL 10how10 hand in hand. DL is a branch of ML which need a large amount of data to learn and present solutions and likewise ML is a branch of AI which enable Ai programs to be able to learn and reason like human without being explicitly programmed.

2.3.1. Deep Learning

Deep learning (DL) is a branch of AI and machine learning methods based on artificial neural networks (Figure 3). A DL architecture can simply be referred to as an Artificial Neural Network (ANN) with two or more hidden layers with the purpose of enhancing the prediction accuracy greater than the traditional ML models [30], it is characterized by a deep stack of computation. DL consists of multiple layers (shallow and deep layers) which referred to as hidden layers. The nonlinear mapping relationship with multi-classification is achieved by the increase in layers of the network which in turn increases the network depth and cost, to provide data representation and feature extraction at different layers. Unlike traditional artificial neural networks, DL uses many hidden layers. In a conventional Deep Neural Network (DNN), a weighted and bias-corrected input value is passed through a non-linear activation function to derive an output[37]. This implies, that the objective of training a DNN is to optimize the weights of the network so that the loss function is minimized[30], [38].

The “Deep” in DL refers to several connected, and continuous layers of representation, usually derived from a “neural network” model, that acts as a filter in each layer to extract useful information from unwanted information (feature extraction). While “learning” refers to the process that improves a model performance by executing a certain process, where the main aim of learning is to enhance performance. DL can learn and gather knowledge from experience, without the need

for humans to specify the knowledge required by the computers. It makes flexible decisions according to the situation by learning a large amount of data and automatically extracting common features. The learning object of DL is also data. The difference from traditional machine learning is that DL requires a large amount of data, called Big Data to get good accuracy[39][12], which have now attracted a lot of attention[14] as it can help healthcare workers detect different TB disease (e.g., pulmonary TB, latent TB, and multidrug-resistant TB) and classify accordingly. Generally, this requires qualified expert analysis of the disease, whereas DL automatically identifies and analyzes the disease without human intervention.

Furthermore, DL incorporates the feature engineering step in traditional ML into a learning step which used to be a big obstacle (instead of extracting features manually, deep learning requires only a set of data with 11how11 preprocessing when necessary, and then learn the informative representations itself). This implies that the workload of feature engineering has shifted from humans to computers, allowing beginners in machine learning to effectively use deep learning for research or applications, especially in disease detection through CXR imaging. Various DL methods have been used in CAD system for medical analysis.

2.3.2. Convolution Neural Network (CNN) and its Architectures Under Study

A Convolutional Neural Network (CNN) is considered as one of the widely adopted Deep Neural Networks, which is basically made up of fully connected Artificial Neural Networks with several layers arranged in a style to take an input and give an output that links to the next layer in the network, which is known as “feed-forward” up until the last and final output classification layer.

CNN is a supervised deep learning architecture, that consist of three layers namely; convolutional layers, pooling layers, and fully connected layers, that have the ability to extract and learn useful features on its own [30], and have a standard structure [22], which are optionally followed by techniques to reduce overfitting and underfitting 11how11he purpose of optimizing accuracy and sensitivity (sensitivity is considered very useful in the field of healthcare) such as; batch normalization, dropout, augmentation, segmentation. This kind of deep convolutional neural network (DCNN) is among the most powerful DL tools used in the field of Artificial Intelligence because of its ability to process a huge amount of data through its multi-layer feature. It is also known for its quality of using convolution layers to replace common matrix multiplication functions in at least one or multiple layers of the network in order to delicately 11how nonlinear diversified functions [40]. CNN’s have been deployed in several practical applications such as; pattern recognition systems, digital image processing and analysis, voice and emotion recognition systems et al. The results achieved from these applications were disruptive in the field of image analysis. Hence has given the architecture a great reputation for yielding high performance in real-

life applications like healthcare systems, where it is used in medical data analytics and diagnostics among other things.

The neural networks in CNN possess a particular behavior through practice and instructions called learning. In this deep learning technique, the learning process is often carried out using convoluted architectures. In CNN architectures, numerical arrays (known as feature maps) are obtained from the input and output of every step in the network, where all feature maps at the output point shows specific feature(s) extracted through a filter on the input of that layer. Moreover, these resulting features in higher phases are generally in a smaller form than the regular shape in a much lower stage, showcasing more practical features extracted [41].

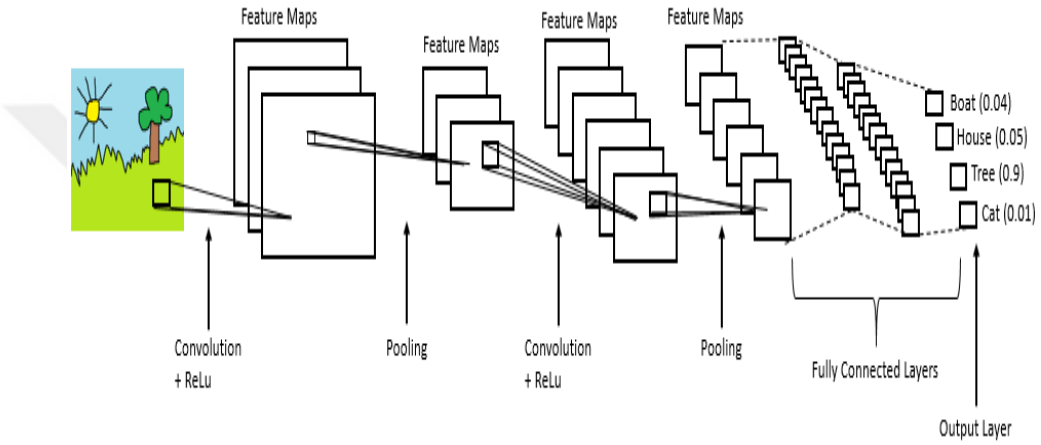

Figure 2.4. Convolutional Neural Network (CNN) Architecture [42]

igure 2.4 demonstrates CNN architecture with two convolutional layers. Each convolutional layer is followed by a pooling or subsampling layer, feature maps of each layer is then fed into the next layer. The feature map of the last pooling layer is fed into two fully connected layer which connect all neurons in one layer to all neurons in another layer and to a final output layer to classify result in either binary or multi-class.

We have several types of CNN pretrained networks in practice today with varying architectures. However, the architectures follow the same universal structure of applying convolutional layers to the input data to generate feature maps. Below are some of the commonly employed DCNN architectures that are used for feature extraction to handle several DL tasks, these includes; AlexNet [21], GoogLeNet [22], ResNet [43], VGG [44], Inception[45], MobileNet [46] etc.

2.3.2.1 AlexNet

AlexNet CNN architecture was developed in 2012 by some ML researchers Alex Krizhevsky and his team. These researchers had interests in the fields of AI, ML, and DL. AlexNet breakthrough came after it emerged winner at the ImageNet competition in 2012 (ILSVRC) [47], which helped in convincing the AI community to greatly adopt deep learning techniques for

application purposes on computer vision tasks twenty years later after researchers lost faith in DL with many layers [21], as a pre-trained model that has gained a rich feature representation to classify a variety of images ranging from various applications knowing its unusual for training a large amount of data with traditional ML techniques. The basic AlexNet structure is similar to that of the LeNet-5, but substantially bigger, with 60 million parameters. AlexNet consists of eight layers with weights; the first five layers are convolutional layers, and the remaining three layers are fully connected layers with some pooling layers, with input size of (224 x 224 x 3) [48].

2.3.2.2 GoogLeNet

GoogLeNet is a type of CNN that was developed by a team of researchers at Google in 2014 they introduced the Inception network. It is a pre-trained CNN based network that consists of 22 deep layers and 5 pooling layers with an input size of (224 x 224 x 3). The network was trained on the ImageNet dataset which classifies images into 1000 output categories. The GoogLeNet (Inception Network) is commonly used because it learned various attribute rendering techniques for a variety of images, as the Inception layers are repeated many times leading to the 22-layers of the network [49] [22]. The network is known as Inception architecture, which find optimal local sparse structure in a convolutional network which can be estimated by applying dense layer components. GoogLeNet won the first position in the 2014 classification and detection competition on the ImageNet challenge ILSVRC [47]. The inception model comprises of a cardinal unit that is referred to as an Inception cell. This unit is used in carrying out a sequence of convolutions at different scales, which then collects the results in a single place. In this model, 1 by 1 convolution is used to reduce the input channel distance from top to bottom so that computations are saved. Hence, for every cell on the input, the network learns series of 1×1, 3×3, and 5×5 filters that are used in extracting features at distinct weights from the input parameters [49]. And hence comes with a linear layer for easy adaption of the network to other label sets.

2.3.2.3 VGG

VGG is a form of convolutional neural network that came into existence in the year 2014 and went ahead to clinch second position in classification problem and first in localization problem on the ImageNet challenge 2014. This CNN based network extends far back with a much deeper and easy to understand the architecture of convolutional structures that achieved better performance than AlexNet architecture with an increase in its architectural layers depth and using (3 × 3) convolution filters in each layer, after the breakthrough of AlexNet [21] in 2012 on the image analysis with DL. When the network first emerged, researchers and industry experts regarded it as the network with the most depth, however, the narratives have changed in recent times, with the evolution in DL and its techniques giving more room for the creation of more robust networks with hundreds of deep layers [50]. Several types of VGG are now in existence with different level of layers such as; VGG-11 which consists of 11 layers, VGG-16 which consists of 16 layers, and

VGG-19 consists of 19 layers. However, they all share the same basic architecture which is made up of 8 convolutional layers, 3 fully connected layers, ReLU activation function on each hidden layer, with a SoftMax function at the output layer similar to that of AlexNet in the study of K. Simonya, and A. Zisserman [44] and hence, has an input size of (224 x 224 x 3).

2.3.2.4 ResNet

Residual network (ResNet) is one of the several types of CNNs developed by Kaiming He et al [43], which gave the rise to the development of much deeper networks with several hundreds of layers unlike the likes of AlexNet et al [50]. The method of residual CNN model is basically adding alternate routes as a link after two layers each, this method is similar to the architectural representation in VGG architecture [44]. The network won the first position in the 2015 classification, detection, and localization competition on the ImageNet challenge with a depth of 152 layers 8 × deeper than the architectural design it was inspired on, which achieved achieves 3.57% error after applying the ensemble method [43]. In the era of AI, it is believed that residual networks are easier to train because they successfully result in lower training and test errors in most of its applications. A good example is the training of residual network of 1000 layers supervised by a team of researchers [51]. This process was not possible in practice before the emergence of ResNet, however, with this architecture it is practically possible to carry out several complex computer vision application tasks nowadays. ResNet happen to have a similar concept of alternate routes just like in LSTM methods. However, the concept of shortcut route or skip layers on ResNets assist in optimizing the function of the network to make them perform better than other regular networks [43]**,** [52]. There are several ResNet with various added layers such as ResNet18, ResNet34, and ResNet101 et al. Nonetheless, the basic structure remains the same all through

2.3.2.5 MobileNet

MobileNet is also a CNN architecture that is based on depth-wise separable convolutions as its key building block, this enables an accurate tradeoff between accuracy and latency (computational time) due to its global hyperparameters that can reduce the cost for developers, it produces a strong performance on other models built on ImageNet dataset. MobileNet is built with 28 layers with asynchronous gradient descent similar to Inception V3. It was basically designed for mobile and embedded vision applications with a wide range of applications in the field of computer vision. Its performance on object detection won the COCO challenge in 2016 [53], [23]. Sandler M. et al [54] build a new model based on MobileNetV1 which introduces an inverted residual and linear bottleneck called MobileNetV2 that improves the state-of-the-art accuracy of MobileNetV1 on various image recognition tasks. [55] in their work reduced the computational cost by trading off a reasonable amount of accuracy of the MobileNetV1 as it one of the setbacks of CNN models by using (3 x 3) depth-wise convolution and using batch normalization at each convolution layer**.**

3 MATERIALS

In this section we present the in-depth materials used in this whole study and also present the dataset used in some of the related studies, along with the performance metrics used for the evaluation of our models to obtain results.

Table 3.1. Dataset Description

Dataset	Positive Data	Negative Data	Total
Public Dataset	3500	3500	7000
Private Dataset	1342	3073	4415

Table 3.1 diagrammatically illustrate our primary study involves two separate datasets; (I) processed public dataset that include NLM (National library of medicine USA), Belarus, and RSNA (Radiological Society of North America) from T. Rahman et al Kaggle repository, and (II) A private deidentified dataset in compliant with the Health Insurance and Accountability Act from Nigeria (West Africa). From the table the public dataset consists a total of 7000 CXR images with 3500 as negative and 3500 as positives. Meanwhile our private dataset consists a total of 4415 unprocessed CXR images with different pixels values with 3073 as negative and 1342 as positives as shown in Table 3.1. The two separate datasets were divided into three sets; (I) used for training, (II) validation and (III) testing of the architectures. The images were divided randomly in the ratios 80:10:10, with 80% in training set as it requires more images for training. Thus, the training set for the public data contains 5600 images out of 7000 images, and validation and testing contain 700 each. Whereas, on the private data and 3530 training images out of 4415 images, and validation and testing contain 441 each. Hence data augmentation was performed on the positive data of the private dataset only to balance the dataset CXR images. The validation set was used for selecting the proper values of hyper-parameters while training and the testing set were only used for model evaluation to calculate the performance of the architecture.

For our private data collection; (I) A mobile X- ray unit was used, provided by US government, nomenclature X-ray portable apparatus (NSN 6525-01-523-1989), (II) A fixed stationed X-ray unit Philip model of 150kv capacity, (III) A Kodak digital image processing unit (CR140) were used.

Meanwhile a computer with an Intel core i7-3632QM CPU @2.20GHz processor and 16GB RAM, alongside with private obtained GPUs from google Colab pro+ were used for data preprocessing via TensorFlow (Keras) framework library and Python programming language for the classification of TB in our study.

3.1. Experimental Performance Criteria

To have a better understanding of the performance metrics based on DL in the detection of TB using different CNN method we carefully show an overview of the performance criteria in this section.

Note; Out of various performance metrics in CAD system for medical analysis, Sensitivity and Specificity are very important. It is also important to understand false negative and false positive, as we intend to make the model predicts whether or not the result of the diagnosis is positive or otherwise, based on the image of the disease.

I. **Accuracy**: The Classification accuracy of a given classifier is referred to as the ability of the classifier to predict the class label of new or unseen data correctly. It is the measure of the correct number of predictions made by the model divided by the total number of predictions as in (Eq.1), which implies we calculate the accuracy by dividing the number of truly classified instances by the number of instances in the test phase.

Accuracy is measured as; $\boldsymbol{ACC = \frac{TP + TN}{TP + TN + FP + FN}}$ **(1)**

Accuracy is very useful when the target classes are well balanced, and hence does not produce a good result with unbalanced classes

II. **Precision:** This is the ability of a classification model to identify only the relevant class. Which is the number of true positives divided by the summation of true positives number and the number of the false positives;

Precision is measures as; $\boldsymbol{Pr = \frac{TP}{TP+FP}}$ **(2)**

III. **Sensitivity**: Sensitivity is also known as "recall" or "true positive rate (TPR)" in performance metrics. It is the ability of a model to correctly predict only the positive cases in a dataset. In our study, this implies the number of TB cases correctly computed. It is the number of true positives divided by the summation of true positives and false negative numbers as in (Eq.3).

Sensitivity is measured as; $\boldsymbol{SEN = \frac{TP}{TP+FN}}$ **(3)**

Sensitivity tells us what proportion of the positive class got correctly classified. A simple illustration is to determine what proportion of the actual sick people were correctly detected by the model.

Note: The Trade-off between sensitivity and precision is that sensitivity expresses the ability to find all relevant instances in a dataset, precision expresses the ability to find all relevant instances in a dataset, precision expresses the proportion of data points our model says was relevant versus what were actually relevant.

IV. **F1 Score**: This is a way we can actually combine and report back the relationship of precision and recall. It is measured using the harmonic mean of precision and recall taking both metrics into account in Eq.4;

F-Score is measured as; $\boldsymbol{F1 - Score = 2 * \frac{Precision \times Recall}{Precision + Recall}}$ **(4)**

The harmonic mean is used instead of a simple average in other to punish extreme values. This means in a problem that a classifier that has a perfect precision of 1.0 and the sensitivity is horrible e.g., at 0.0, this implies if we are just to take a simple average it would be 0.5 but when we apply the 1.0 precision and sensitivity of 0.0 to the harmonic mean then the F1 score will be 0.

V. **False Negative Rate**; False Negative Rate (FNR) also known as "miss rate" or "fall out". illustrate the proportion of the positive class that was incorrectly classified by the model. The goal is to correctly classify the positive class rather than a negative class in a classification problem, which implies; a higher TPR and a lower FNR are desired

False Negative Rate is measured as; $\boldsymbol{FNR = \frac{FN}{TP+FN}}$**=1- TPR** **(5)**

VI. **Specificity**: Specificity is also known as the true negative. It's the ability of a model to predict how accurate the overall negative result is, and hence illustrate how much of the predicted negative result was right. It is denoted as (Eq.3). In medical imaging analysis, this means illustrating the proportion of people without a disease been correctly identified by a model.

Specificity is measured as; $\boldsymbol{SPE = \frac{TN}{TN+FP}}$ **(6)**

Confusion Metrics: Also known as table of confusion. The main point behind the confusion metrics and various metrics is that they are all fundamental ways of comparing the predicted values versus the true values, however what constitute a "good metrics" will really depend on the specific situation.

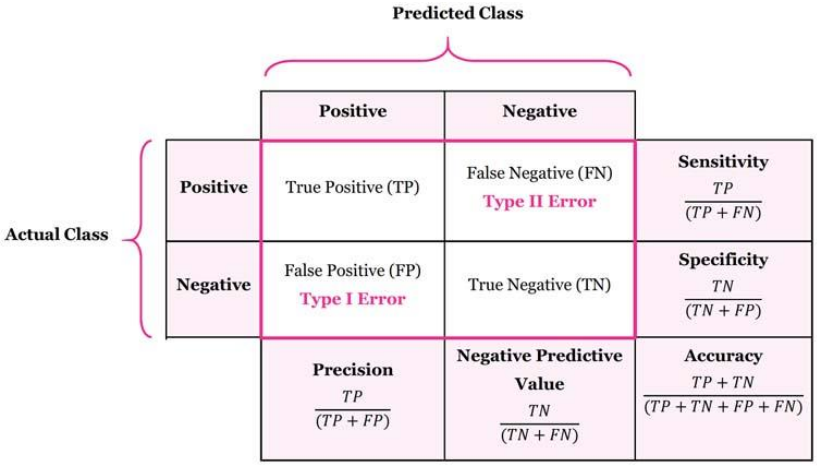

Figure 3.1. Confusion metrics chart

Figure 3.1 of our study, carefully describe the predicted values of the confusion matrix as either positive and negative, and the actual values as True and False.

I. True positive (TP): Here disease is positive and the model predicted positive (correct identification of result)

II. True negative (TN): Here disease is negative and the model predicted negative (correct identification of result).

III. False positive (FP): Here disease is negative but our model predicted positive (Type 1 Error, incorrect identification of result).

IV. False negatives (FN): Here disease is positive but the model predicted negative (Type 2 Error, incorrect identification of result)

It is very important to understand false negatives and false positives to understand the review on the performance of various CNN architectures. CAD system might come with some difficulty in trying to diagnose patients whether or not they have a disease base on the image of the disease, here it implies, the ultimate goal is to try our best to reduce FN (if not a patient might have a disease but the model predicts the patient is free from the disease) which is an extremely dangerous position to put a patient. Therefore, DL techniques in the health system try to trade-off precision vs sensitivity in order to help minimize FN. Furthermore, FP is not great also because a model predicts a patient doesn't have a disease after a diagnosis, but then the disease is present in the patient, whereas in real-world when its FP there are other more biopsy to perform to see if that person really has the disease, but FN will walk away without a further diagnosis, thinking the disease is negative, meanwhile it is positive.

4 METHOD

The methodology of this study was carried out in five different stages as listed below and shown schematically in figure 6.

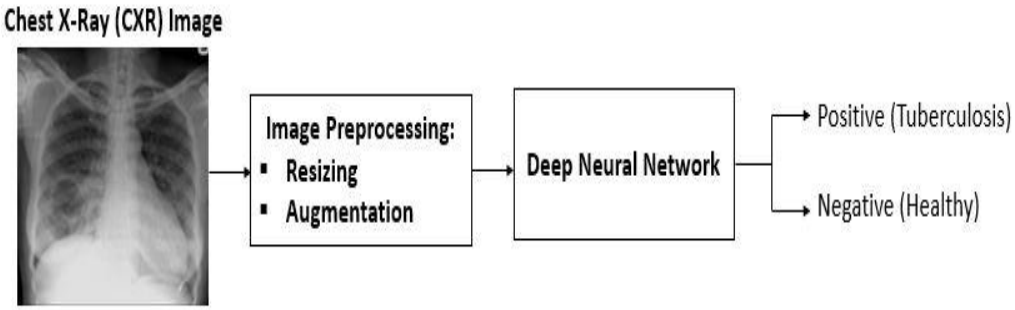

Figure 4.1. Schematic flow of the Study

The schematic of the whole study is illustrated in figure 4.1, which illustrate the flow of our method. The first step is the data gathering, then the second step which is the processing or our data (CXR-images), the third step of our schematic flow is passing the processed data into our models, and the final stage the result is obtained.

I. The public dataset was used for the binary classification of TB positive and negative cases using several pretrained DCNNs models as shown in Table 5.1 with the help of Figure 4.1 design,

II. The private dataset was also used for binary classification of TB positive and negative cases using the same pretrained models from the first method and design flow of Figure 4.1, as shown in Table 5.2 respectively.

III. The pretrained model's configuration were tweaked to form a custom model by training some layers and leaving out some layers, or by adding/reducing dense layers to classify the TB positive and negative cases on our two separated dataset.

IV. Feature extraction using pretrained models was used on the public dataset before piping it to two separate classical method, namely; Support vector machine (SVM), and random forest (RF), for binary classification of TB

V. Feature extraction using pretrained models was used on the private dataset before piping it to two separate classical method, namely; SVM, and RF, for binary classification of TB.

4.1. Preprocessing

The sizes of the private dataset CXR images varied with no actual standard, hence the datasets were preprocessed to resize the CXR images 224*224 to obtain a standard for the dataset, whereas most of the DCCNs models accepts 224*224 except for NasNet that accept only 331*331 as input shape, hence we also converted the images from grey scaled to Blue, Green, Red (BGR) to obtain three channels since pretrained models accept only 3 channels , and further store it into a NumPy array, which we further normalized it to fit into the different DCNNs models used. Whereas the public dataset has standard but the same preprocessing technique carried out on the private dataset as stated above were carried out to obtain a similar dimension of images and to carry out the same comparison.

4.2. Augmentation

Table 4.1. Dataset with Augmentation

Dataset	Total	Positive Data	Negative Data	Train	Test	Valid
Public	7000	3500	3500	5600	700	700
Private	6000	3000	3000	4800	600	600

From Table 4.1. to improve our classification performance knowing DL requires a huge amount of data and the amount of dataset available in our study is limited and not balanced, we used different data augmentation techniques to increase the number of our positive data to 3000, and gave away some negative images, rather than collecting new data which proved to be a problem, and the augmentation technique will help to avoid overfitting. This technique in turn expands the diversity of data available for training models and helped us in having a balance data for both positive and negative classification. The built-in Keras framework for data augmentation was used. However, in this study, different image augmentation techniques (re-scaling, and fill mode) were utilized to generate more images, as shown in Figure 4.2. In this work, our fill mode was set to reflect for convenience.

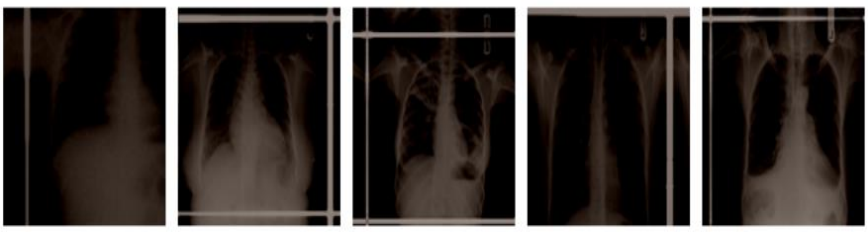

Figure 4.2. Image Augmentation

Figure 4.2 shoes the different images obtained from our data augmentation.

5 RESULTS AND DISCUSSIONS

This section presents the results of our study and other various studies related to our work. We present different tables that shows the binary classification performance of the study based on different performance metrics from section 3.1 on various DCNN architectures under review for the detection of TB using CAD systems from various studies ranging from 2015-2022. Hence, we presented our results in three different tables, and in the Table 5.4 we showed the comparison of our work with the works of the related literatures.

Table 5.1. Performance of Different Models for the Classification of TB with Public Data

DCNNs Models	Accuracy	Weighted Average				
		Sensitivity	Specificity	Precision	Recall	F1-Score
VGG 16	99	99	99	99	99	99
VGG 19	99	99	99	99	99	99
ResNet152	82	78	87	82	82	82
RestNet50V2	97	99	95	97	97	97
ResNet152V2	98	98	97	98	98	98
ResNet101V2	99	97	100	99	99	99
MobileNetV1	**100**	**99**	**100**	**100**	**100**	**100**
MobileNetV2	99	100	97	99	99	99
InceptionV3	97	99	95	97	97	97
InceptionRNet	98	97	99	98	98	98

In Table 5.1 we showed results of the binary classification of our public dataset using different DCCNs models as stated in method two of section 4, hence MobileNetV1 appears to be the best result out of the ten (10) other DCCNs method used followed by VGG16 and VGG19. Weighted average was the standard set to collate our result from the performance metrics used.

Table 5.2 Performance of Different Models for the Classification of TB with Unbalance Private Data,

Model	Train Accuracy	Test Accuracy	Weighted average		
			Precision	recall	f1
VGG 16	78	78	81	78	75
ResNet50V2	85	84	86	84	82
ResNet152V2	86	85	86	85	83
ResNet101V2	85	85	87	85	83
MobileNetV2	71	71	80	71	61
MobileNetV1	85	85	87	85	83
InceptionV3	85	87	87	87	86
InceptionRNet	84	83	84	83	81

In Table 5.2, we showed the results of our unbalanced private dataset using the same DCCNs models and the same techniques used in the first method as in section 4, hence MobileNetV1 and RestNet152V2 where the models with the best results, and VGG16 happens to be the worst result, for the metrics we used for evaluation we used the weighted average as our result likewise in the previous table.

Table 5.3. Performance of Different Models for the Classification of TB with Balance Private Data

DCNNs Models	Accuracy	Weighted Average				
		Sensitivity	Specificity	Precision	Recall	F1-Score
VGG 16	**94**	**94**	**94**	**94**	**94**	**94**
VGG 19	92	92	92	92	92	92
ResNet152	84	86	82	85	84	84
RestNet50V2	92	94	89	84	84	92
ResNet152V2	92	92	91	92	92	92
ResNet101V2	92	90	94	92	92	92
MobileNetV1	**94**	**92**	**96**	**94**	**94**	**94**
MobileNetV2	90	88	91	90	90	90
InceptionV3	90	89	91	90	90	90
InceptionRNetV2	90	89	90	89	89	89

In Table 5.3, we showed the results of our private dataset using the same DCCNs models and the same techniques used in the first method as in section 4, hence MobileNetV1 and VGG16 where the models with the best results, for the metrics we used for evaluation we used the weighted average as our result likewise in the previous table.

Table 5.4. Classical Methods

Classical Models			Accuracy	Weighted Average				
				Sensitivity	Specificity	Precision	Recall	F1-Score
SVM	VGG16	Private Data	93	94	93	93	93	93
RF			91	93	89	91	91	91
SVM		Public Data	100	100	100	100	100	100
RF			97	98	96	97	97	97
SVM	MNET1	Public Data	99	99	99	99	99	99
RF			95	97	92	95	95	95
SVM		Private Data	89	88	90	89	89	89
RF			91	93	89	91	91	91

Table 5.4 illustrate the classical model we used on our dataset as stated in method 4, and 5 of section 4 of our study. We used two different pretrained models as our feature extractors (VGG16 and MNET1 model), and later pipped them into two different classical ML method (SVM, and Random Forest) to obtain result for our binary classification using our public dataset and our balanced private dataset after several pre-processing.

Table 5.5. Result of Various Studies under Review using CNN for Diagnosing (TB) in Terms of; Accuracy, Precision, Sensitivity and Specificity

Paper	Dataset	Architecture	Description	ACC	PRE	SEN	F1	SPE
F.Pasa et al [20]	MC	Custom (7 layers)	There was no Ensemble method performed	79.0	-	-	-	-
	SZ			84.4	-	-	-	-
	Combine			86.2	-	-	-	-
T. Rahman et al[27]	NLM Belarus RSNA	Ensemble of 9 CNN Models	with lungs segmentation and Aug	99.9	99.91	99.9	99.9	99.52
		Ensemble of 9 CNN Models	without lungs segmentation and Aug	97.07	97.34	97.07	97.14	97.36
		ResNet18	On Original CXR	95.57	96.12	95.57	95.71	95.80
		ResNet50		95.86	96.39	95.86	95.99	96.54
		ResNet101		96.88	97.19	96.88	96.96	97.32
		ChexNet		97.07	97.34	97.07	97.14	97.36
		InceptionV3		95.95	96.40	96.00	96.20	96.89
		VGG19		95.81	96.38	95.81	95.95	96.76
		DenseNet201		95.83	96.47	95.84	95.99	97.46
		SqueezeNet		95.45	96.14	95.45	95.62	96.69
		MobileNetV1		95.26	95.9	95.26	95.43	95.62
		ResNet18	With Segmentation (Segmented lung)	99.86	99.85	99.85	99.85	99.63
		ResNet50		99.88	99.88	99.88	99.88	99.52
		ResNet101		99.79	99.79	99.79	99.78	99.39
		ChexNet		99.69	99.69	99.69	99.69	98.68
		InceptionV3		99.83	99.83	99.83	99.83	99.62
		VGG19		99.88	99.88	99.88	99.88	99.41
		DenseNet201		99.90	99.91	99.90	99.90	99.52
		SqueezeNet		99.67	99.67	99.66	99.66	98.56
		MobileNetV1		99.76	99.76	99.76	99.76	99.15

R. Hooda et al [31]	NLM (SZ, MC), Belarus JSRT	3-CNN	ensemble with data Aug	88.24	-	88.42	-	88.00
		AlexNet	Data Aug	83.53	-	-	-	-
		GoogLeNet		80.59	-	-	-	-
		ResNet		84.12	-	-	-	-
Y. Xiong et al [24]	PUH 246	Custom CNN (TB-AI)		-	-	97.94	-	83.65
S. Hwang et al. [28]	KIT 10848	Custom CNN (AlexNet +1 Layer)	Performance comparison without transfer learning	77.3	-	-	-	-
	MC 138			78.8	-	-	-	-
	SZ 662			75.8	-	-	-	-
	KIT 10848		Performance comparison with transfer learning	90.2	-	-	-	-
	MC 138			90.5	-	-	-	-
	SZ 662			90.3	-	-	-	-
P. Lakhani et al [29]	MC SZ B TJ UH	GoogLeNet-TA	Pre-trained with Aug	95.3	-	92.0	-	98.7
		AlexNet -TA		94.3	-	92.0	-	94.7
		Ensemble	Pre-trained with Aug	96.0	-	97.3	-	94.7
		Ensemble	Pre-trained and Radiologist AUG	98.7	-	97.3	-	100
		GoogLeNet--U	Untrained	-	-	-	-	-
		AlexNet -U		-	-	-	-	-
		GoogLeNet-UA	Untrained with Aug	-	-	-	-	-
		AlexNet -UA		-	-	-	-	-
P. Rajpurkar et al [19]	GF Jooste Hospital SA	C-DL algorithm (With clinicians)	No pre-processing	78.0	-	67.0	-	87.0
		C-DL algorithm (Without Clinicians)		61.0	-	70.0	-	52.0
E.J Hwang et al [56]	SNUH	Custom CNN (DLAD) SEN CUT OFF		96.2	-	95.2	-	100
	BMC			94.5	-	94.3	-	95.7
	KUHG			100	-	100	-	91.4
	DEMC			100	-	100	-	98.0
	MC			100	-	100	-	93.8
	Shenzhen			95.3	-	94.7	-	91.1
	SNUH	Custom 27Layers CNN (DLAD) Specificity Cut Off		75.0	-	84.3	-	100
	BMC			75.9	-	90.0	-	100
	KUHG			80.6	-	99.0	-	100
	DEMC			71.9	-	98.6	-	100
	MC			71.9	-	84.6	-	100

	Shenzhen			77.1	-	84.1	-	99.1
	Combine Dataset	without DLAD	Board certified radiologists	79.7	-	90.6	-	94.8
			Thoracic radiologists	87.0	-	95.2	-	93.0
		Custom 27Layers CNN (DLAD)	Board certified radiologists	84.9	-	93.0	-	95.4
			Thoracic radiologists	89.7	-	96.4	-	93.6
M. H. A. Hijazi et al [30]		InceptionV3	Pre-processing	85.0	-	80.7	-	90.2
		VGG16		85.6	-	81.7	-	90.3
		Custom		90.0	-	88.6	-	91.7
		Ensemble		91.0	-	89.6	-	90.7
M.F Alcantara et al[3]	PIH (4701)	GoogLeNet Binary classification	No pre-processing	89.6	-	-	-	-
		GoogLeNet Multi classification		62.07	-	-	-	-
S. K. T. Hwa [33]	SZ MC	VGG16	Original images	88.64	-	86.36	-	90.91
		InceptionV3		79.55	-	77.27	-	81.82
		VGG16	Canny Edge Image	88.64	-	84.09	-	93.18
		InceptionV3		76.14	-	77.27	-	75.00
		Ensemble of ABC	Canny Edge image	92.05	-	88.64	-	95.45
		Ensemble of ABCD		89.7	-	90.91	-	88.64
U.K Lopes et al [13]	MC	GoogLeNet	Simple Features	78.2	-	-	-	-
		RESNET		74.6	-	-	-	-
		VGG		75.3	-	-	-	-
	SHZ	GoogLeNet		82.2	-	-	-	-
		RESNET		83.4	-	-	-	-
		VGG		82.8	-	-	-	-
	MC	GoogLeNet	Bag of features	78.2	-	-	-	-
		RESNET		81.0	-	-	-	-
		VGG		80.3	-	-	-	-
	SHZ	GoogLeNet		83.4	-	-	-	-
		RESNET		79.9	-	-	-	-
		VGG		83.5	-	-	-	-
	MC	Ensemble (3 CNN)	Ensemble Using Simple Features	76.0	-	-	-	-
			Ensemble Using Bag of features	82.6	-	-	-	-
	SHZ	Ensemble (3 CNN)	Ensemble Using Simple Features	84.6	-	-	-	-

			Ensemble Using Bag of features	84.6	-	-	-	-
M.F Alcantara et al, [57]	PIH (4701)	Custom (GoogLeNet)	binary	93.4	-	-	-	-
			Multi-class	-	96.05	-	-	-
Our Study Method One	Public Data	VGG16	Best Performing Models	99	99	99	99	99
		MobileNetV1		100	100	99	100	100
Our Study Method Two	Private Data (Nigeria)	VGG16	Best Performing Models	94	94	94	94	94
		MobileNetV1		94	94	92	94	96

Table 5.5 shows the result of various studies that were used for comparison with our tow separate studies.

In our entire study, the CXR images were binary classified (positive and negative). For each DCNNs model used, the learning rate set to 0.0001 and epoch set to 150 respectively. Activation function was set to SoftMax and categorical cross-entropy were all employed to train the model. Table 5.1 shows the results obtained from the training while using public data (weighted average of the training result was considered for all the models in the method). Whereas Table 5.3 shows the result obtained from the training while using our private dataset (weighted average of the training result from the classification table was considered for all the models in the method),

This study has several limitations such as; Some CXR-images of our private data are skewed and having no standard due to the machines been used in collecting and storing the images and without having a radiologist or expert to look at the data after preprocessing, some of the images poses a large scale of black area which we cropped some part of that because we believe it can be a problem. Nonetheless, we believe that our results could be helpful future prospective researches for the detection of TB in Nigeria and the African region at large. Furthermore, with regard to the data used for testing, validation, and training, due to human error and the fact that all images were obtained from one center, ignite the possibility that some patients CXR images found in the test set might also be found in the training set even after a very careful scrutinizing of the data. Additional images were also excluded due to size or standard set while preprocessing the data. In Table 5.2.

In our first method using pretrained models to train our public data, we obtained a state of the art performance as in T. Rahman et al [32], meanwhile in our second method our first experiment was carried out on the original private dataset that is unbalance with no much standards and we obtain a poor result as presented in Table 5.2. The images of the negative set where higher than that of the positive set as seen from Table 3.1, which after splitting the data into train, test, and validation, we had 307 negative images and 134 positive images a total of 441 images in our validation and test folders respectively. Hence further preprocessing techniques such as augmentation were carried out to enhance our binary classification performance knowing the

dataset is new and have never been processed. In our augmentation we gave out some images of the negative dataset to manually down sample it to three thousand (3000) and we carried data augmentation on only the positive data to three thousand (3000), to balance both positive dataset and negative dataset before performing train, test, validation split with 80% for training, 10% for testing and validation respectively. After several experiment to enhance our result, we obtain a good performance as presented in Table 5.3.

A general comparison with other studies was achieved while using our private data as it showed great promise, and very competitive, which implies that the result obtained in our study was excellent due to the dataset that was at our disposal on the task set. Based on figure 4.1 which systematically explains the workflow of our models on our datasets in detecting TB, Figure 5.4 illustrate the confusion matrix of our private data binary classification which shows that our model wrongly classified eleven (11) patients to be TB negative when they were actually TB positive patients, and predicted twenty-five (25) patients to be TB positive while they were actually negative with a sensitivity and specificity of 92%, and 96% respectively meanwhile all other predictions were correct this result was obtained after series of pre-processing techniques, trial and error methods to obtain a better result. Initially our data was not balanced and the CXR-images where having different pixel size, which led to having a low result compared to after the data was balanced (having the same number of positive CXR-images and as well negative CXR-images to be 3000). In figure 5.5 which shows a confusion matrix of the same model as in figure 5.4, shows how the model performed which wrongly classified seventy (70) patients to be TB negative when they were actually TB positive, and also wrongly classified three (3) patients to be positive while they were actually negative. While in figure 5.1 binary classification, which happened to be our best performing model while using our public dataset, it wrongly predicted one (1) patient to be TB negative when they were actually TB positive patients, and also predicted two (2) patients to be TB positive when they were actually TB negative which can be term as a state-of-the-art performance.

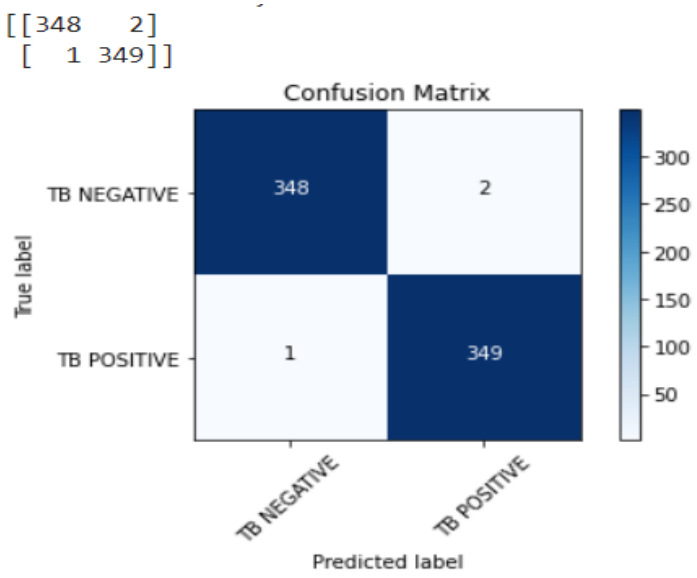

Figure 5.1. Confusion Matrix of our MobileNet V1 Model with Public Data

Figure 5.1 shows the confusion matrix of our public data using mobileNet V1 which appears to be a very good result which only three images where wrongly classified and with six-hundred ninety-seven images correctly classified.

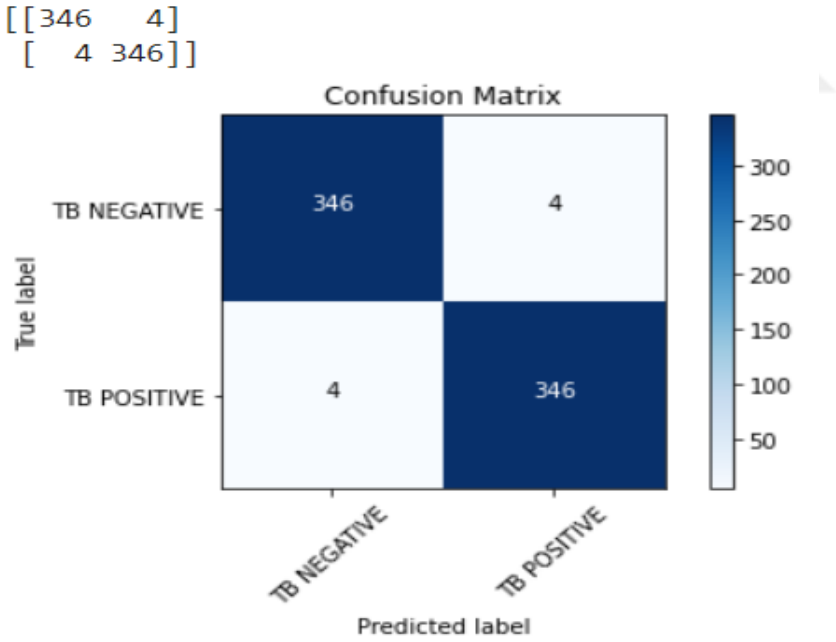

Figure 5.2**.** Confusion Matrix of our VGG16 with Public Data

Figure 5.2 represent the confusion matrix of our public data using VGG16 which appears to be a good result considering eighth images wrongly classified out of seven hundred images, with six-hundred ninety-two images correctly classified.

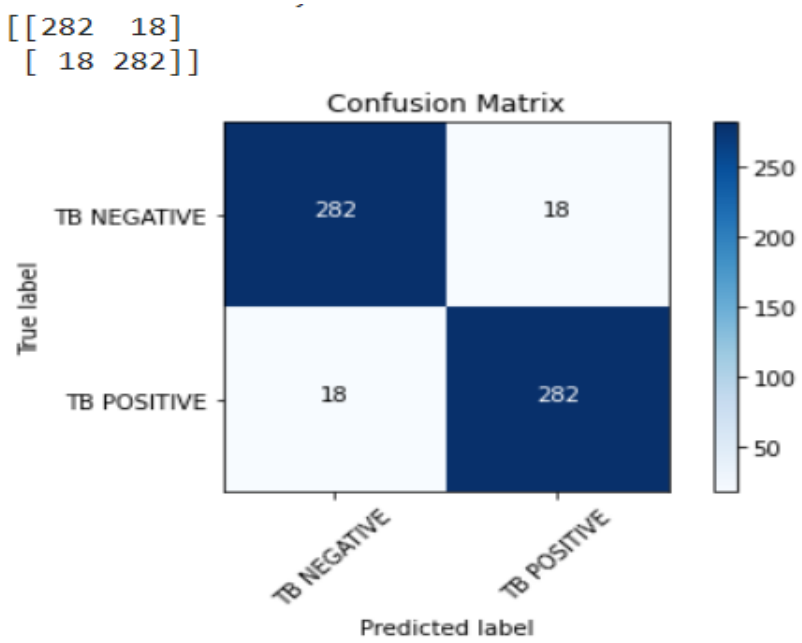

Figure 5.3. Confusion Matrix of our VGG16 with Balanced Private Data

Figure 5.3 shows the confusion matrix of our private data using VGG16 which appears to be a fair result which wrongly classified thirty-six images out of seven hundred CXR-images after pre-processing, having in mind it's an entirely new and a private data.

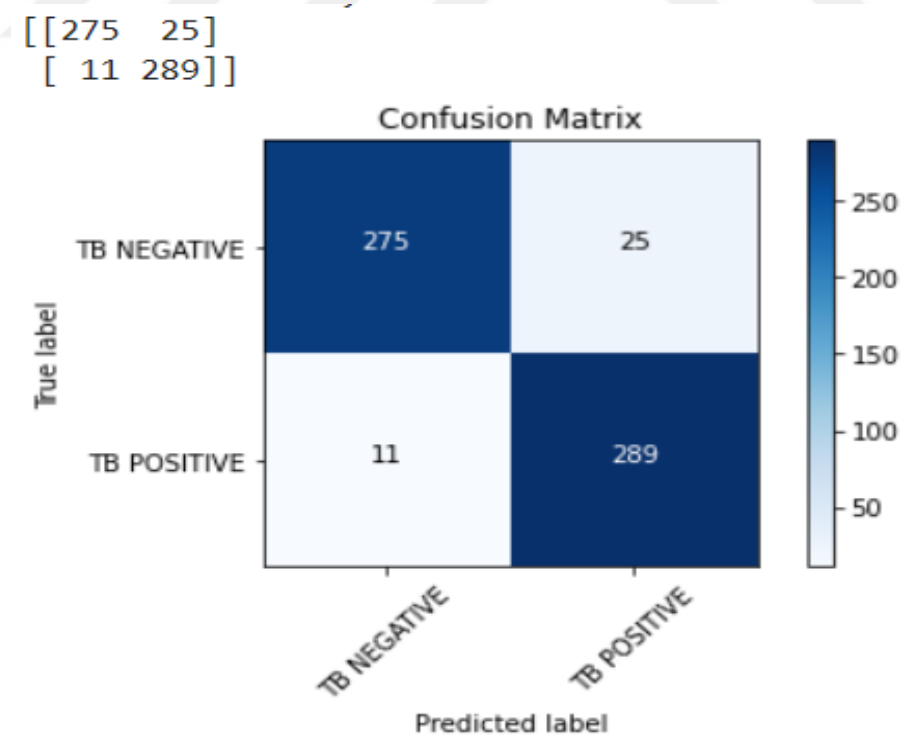

Figure 5.4. Confusion Matrix of our MobileNet V1 Model with Balance Private Data

Figure 5.4 shows the confusion matrix of our private data using MobileNet V1 which appears to be a fair result wrongly classifying thirty-six CXR-images out of seven hundred CXR-images after several pre-processing having in mind it's also an entirely new and a private data.

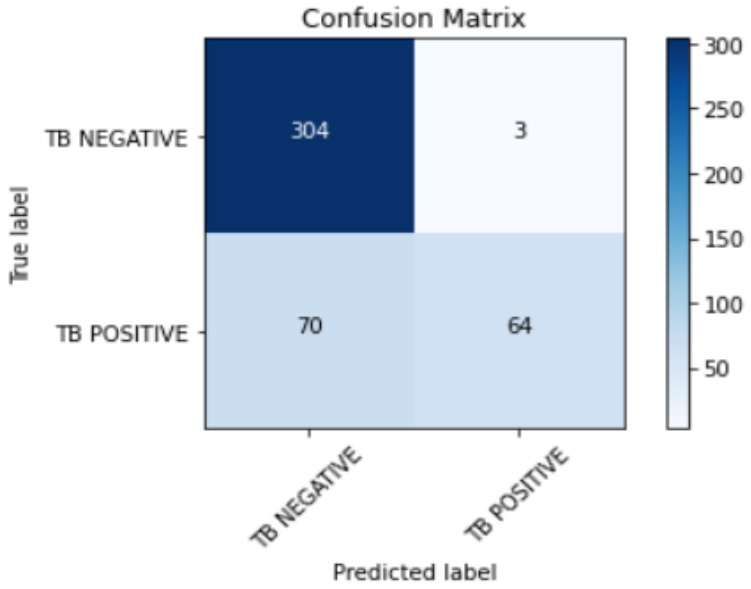

Figure 5.5 Confusion Matrix of our MobileNet V1 Model with Unbalance Private Data

Figure 5.5 shows the confusion matrix of our unbalanced private data using MobileNet V1, wrongly classifying seventy-three CXR-images out of seven hundred CXR-images. The confusion matrix also showed that the model classified sixty-four CXR-images as true positives and three hundred and four images as true negative. This is a result that clearly can't be accepted in medical field because it will cause so many casualties.

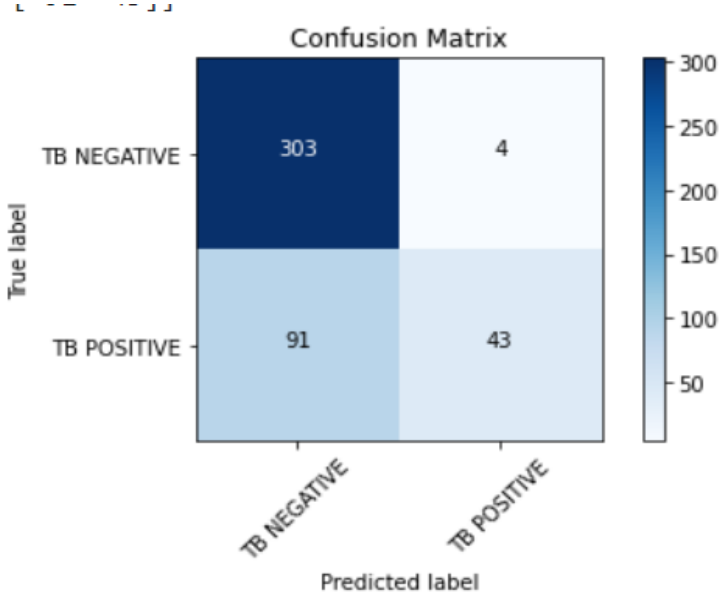

Figure 5.6 SVM Linear Unbalance Private Data

Figure 5.6 shows the confusion matrix of our unbalanced private data using classical machine learning method (SVM), the model wrongly classified ninety-five CXR-images out of seven hundred CXR-images. It also shows that the confusion matrix forty-three CXR-images as true positives and three hundred and three images as true negative, clearly showing the unbalance nature of the model.

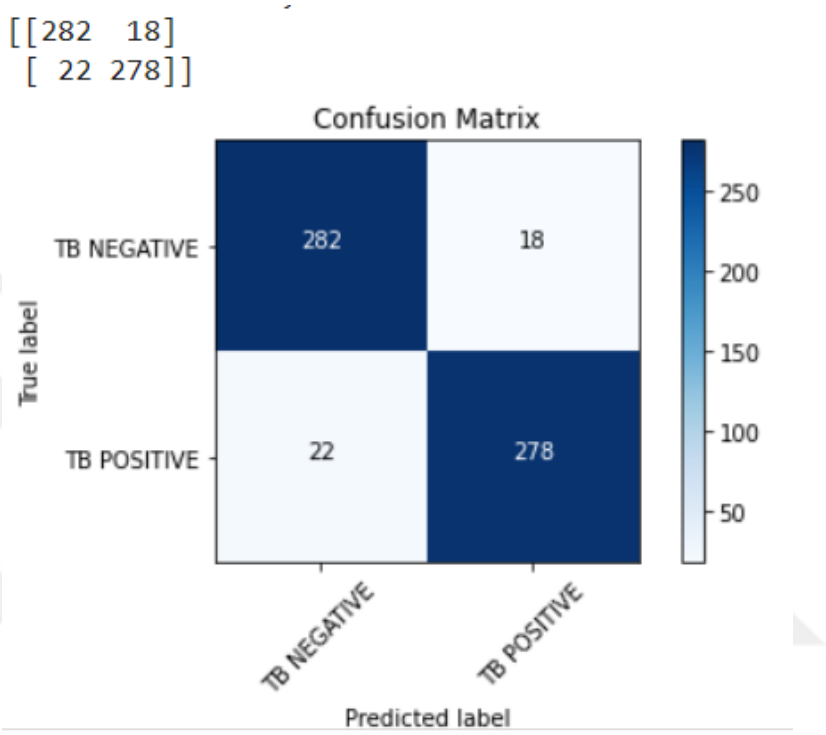

Figure 5.7. Confusion Matrix of SVM Linear on balance Private Data (Classical ML)

Figure 5.7 shows the confusion matrix of our private data using classical ML method (SVM) on our balanced data, wrongly classifying forty CXR-images out of seven hundred CXR-images after several pre-processing.

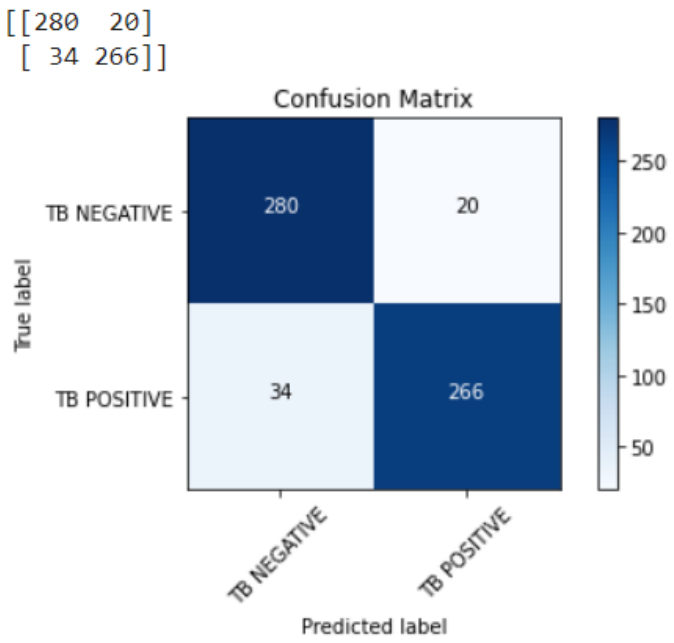

Figure 5.8. Confusion Matrix of Random Forest Classifier on Balanced Private Data (Classical ML)

Figure 5.8 shows the confusion matrix of our private data using classical ML method (RF) on our balanced data, wrongly classifying fifty-four CXR-images out of seven hundred CXR-images.

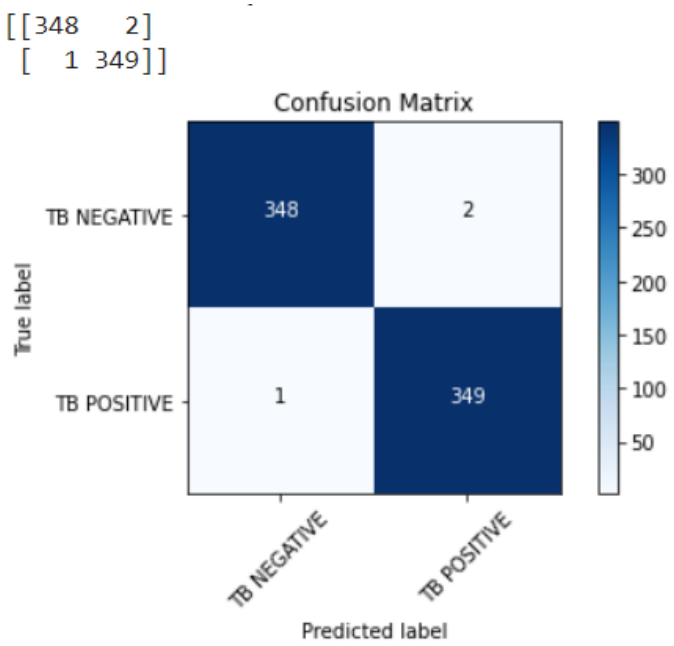

Figure 5.9. SVM PUBLIC

Figure 5.9 represent the confusion matrix of our public data using SVM which appears to be a good result considering three images are wrongly classified out of seven hundred images.

6 CONCLUSİON

In conclusion, a pretrained deep convolutional neural network have been proposed to effectively detect tuberculosis from CXR images. A number of pretrained deep convolutional neural networks were applied to the public and private CXR images. These pretrained models used in our experiment includes: VGG16, VGG19, ResNet152, ResNet50V2, ResNet101V2, ResNet152, MobileNetV1, MobileNetV2, InceptionV3, and InceptionRNet. The result of this experiment shows that MobileNetV1 achieved the best performance when applied on the public dataset, this is because its good for vision problems, due to its ability to trade-off part of its performance and reduce computational speed by using depth-wise separable convolutions as its building block and the use of shortcut at the residual connection layers transforming lower-level concept to lower level descriptors. Meanwhile, for our private dataset, VGG16 and MobileNetV1 achieved the best performance. The performance of the deep learning model applied on the private dataset was observed to be less than that of the public dataset due to different factors, such as; demography of data collection, number of data obtained from the different regions, type of machines used, personal, and many more. However, the results in this proposed study demonstrate the effectiveness of DCCNs in predicting tuberculosis disease from CXR images. In our future work we will work hand in hand with clinicians to collate a more quality dataset and with high tech X-ray machines and computers to reduce the black area in the X-ray images, also define the different stages of the disease, hence involve some expert physicians and radiologist to observe the dataset and the result obtained, which will definitely increase the performance of our study with a significant percentage to help as an agent in rolling out TB in Nigeria, and Africa as a whole.

REFERENCES

[1] S. Desmon, "Taking a Deep Dive into Why Nigeria's TB Rates are So High," 2018. https://ccp.jhu.edu/2018/10/22/nigeria-tb-rates-high-tuberculosis/ (accessed Jan. 20, 2021).

[2] WHO, "The End TB Strategy Global strategy and targets for tuberculosis prevention, care and control after 2015," 2015. https://www.who.int/tb/strategy/End_TB_Strategy.pdf.

[3] M. F. Alcantara *et al.*, "Improving tuberculosis diagnostics using deep learning and mobile health technologies among resource-poor communities in Perú," *Smart Heal.*, vol. 1–2, no. March, pp. 66–76, 2017, doi: 10.1016/j.smhl.2017.04.003.

[4] WHO, "WHO Global TB report 2016," 2016. https://www.who.int/tb/publications/global_report/gtbr2016_executive_summary.pdf.

[5] A. D. Harries *et al.*, "Deaths from tuberculosis in sub-Saharan African countries with a high prevalence of HIV-1," *Lancet*, vol. 357, no. 9267, pp. 1519–1523, 2001, doi: 10.1016/S0140-6736(00)04639-0.

[6] A. CDC, "Tuberculosis," *Tuberculosis in africa*. https://africacdc.org/disease/tuberculosis/#:~:text=TB is also the leading,%25) from the African region (accessed Dec. 24, 2020).

[7] WHO, "World TB Day," 2018. https://www.afro.who.int/health-topics/tuberculosis-tb.

[8] WHO, "World TB Day." https://www.who.int/campaigns/world-tb-day (accessed Mar. 24, 2022).

[9] A. Vassall and M. Gidado, "Post-2015 Development Agenda: Nigeria Perspectives - Tuberculosis," pp. 1–40, 2015.

[10] O. Chukwu and B. Okoeguale, "FIRST National TB Prevalence Survey 2012, Nigeria," Nigeria, 2012. [Online]. Available: https://www.who.int/tb/publications/NigeriaReport_WEB_NEW.pdf.

[11] WHO, "Chest Radiography in Tuberculosis," *WHO Libr. Cat. Data*, p. 44, 2016, [Online]. Available: http://www.who.int/about/licensing/copyright_form%0Ahttp://www.who.int/about/licensing/copyright_form).

[12] S. Wang and R. M. Summers, "Machine learning and radiology," *Med. Image Anal.*, vol. 16, no. 5, pp. 933–951, 2012, doi: 10.1016/j.media.2012.02.005.

[13] U. K. Lopes and J. F. Valiati, "Pre-trained convolutional neural networks as feature extractors for tuberculosis detection," *Comput. Biol. Med.*, vol. 89, pp. 135–143, 2017, doi: 10.1016/j.compbiomed.2017.08.001.

[14] B. Chen, L. Zhang, H. Chen, K. Liang, and X. Chen, "A novel extended Kalman filter with support vector machine based method for the automatic diagnosis and segmentation of brain tumors," *Comput. Methods Programs Biomed.*, no. xxxx, p. 105797, 2020, doi: 10.1016/j.cmpb.2020.105797.

[15] E. H. Houssein, M. M. Emam, A. A. Ali, and P. N. Suganthan, "Deep and machine learning techniques for medical imaging-based breast cancer: A comprehensive review," *Expert Syst. Appl.*, no. April, p. 114161, 2020, doi: 10.1016/j.eswa.2020.114161.

[16] T. Ozturk, M. Talo, E. A. Yildirim, U. B. Baloglu, O. Yildirim, and U. Rajendra Acharya, "Automated detection of COVID-19 cases using deep neural networks with X-ray images," *Comput. Biol. Med.*, vol. 121, no. April, p. 103792, 2020, doi: 10.1016/j.compbiomed.2020.103792.

[17] C. R. Olsen, R. J. Mentz, K. J. Anstrom, D. Page, and P. A. Patel, "Clinical applications of machine learning in the diagnosis, classification, and prediction of heart failure: Machine learning in heart failure," *Am. Heart J.*, vol. 229, pp. 1–17, 2020, doi: 10.1016/j.ahj.2020.07.009.

[18] J. Chaki, S. Thillai Ganesh, S. K. Cidham, and S. Ananda Theertan, "Machine learning and artificial intelligence based Diabetes Mellitus detection and self-management: A systematic review," *J. King Saud Univ. - Comput. Inf. Sci.*, no. xxxx, 2020, doi: 10.1016/j.jksuci.2020.06.013.

[19] P. Rajpurkar *et al.*, "Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists," *PLoS Med.*, vol. 15, no. 11, pp. 1–17, 2018, doi:

10.1371/journal.pmed.1002686.

[20] F. Pasa, V. Golkov, F. Pfeiffer, D. Cremers, and D. Pfeiffer, "Efficient Deep Network Architectures for Fast Chest X-Ray Tuberculosis Screening and Visualization," *Sci. Rep.*, vol. 9, no. 1, pp. 2–10, 2019, doi: 10.1038/s41598-019-42557-4.

[21] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386.

[22] C. Szegedy *et al.*, "Going deeper with convolutions," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2015, pp. 1–9, doi: 10.1109/CVPR.2015.7298594.

[23] A. Howard *et al.*, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," 2017.

[24] Y. Xiong, X. Ba, A. Hou, K. Zhang, L. Chen, and T. Li, "Automatic detection of mycobacterium tuberculosis using artificial intelligence," *J. Thorac. Dis.*, vol. 10, no. 3, pp. 1936–1940, 2018, doi: 10.21037/jtd.2018.01.91.

[25] E. J. Hwang *et al.*, "Development and Validation of a Deep Learning-based Automatic Detection Algorithm for Active Pulmonary Tuberculosis on Chest Radiographs," *Clin. Infect. Dis.*, vol. 69, no. 5, pp. 739–747, 2019, doi: 10.1093/cid/ciy967.

[26] P. Rajpurkar *et al.*, "CheXaid_ deep learning assistance for physician diagnosis of tuberculosis using chest x-rays in patients with HIV," *npj Digit. Med.*, vol. 3, no. 1, pp. 1–8, 2020, doi: 10.1038/s41746-020-00322-2.

[27] T. Rahman *et al.*, "Reliable tuberculosis detection using chest x-ray with deep learning, segmentation and visualization," *arXiv*, no. November, 2020, doi: 10.1109/access.2020.3031384.

[28] S. Hwang, H.-E. Kim, J. Jeong, and H.-J. Kim, "A novel approach for tuberculosis screening based on deep convolutional neural networks," *Med. Imaging 2016 Comput. Diagnosis*, vol. 9785, p. 97852W, 2016, doi: 10.1117/12.2216198.

[29] P. Lakhani and B. Sundaram, "THORACIC IMAGING: Deep Learning at Chest Radiography Lakhani and Sundaram," *Radiology*, vol. 284, no. 2, pp. 574–582, 2017, [Online]. Available: http://pubs.rsna.org.ezp-prod1.hul.harvard.edu/doi/pdf/10.1148/radiol.2017162326.

[30] M. H. A. Hijazi, L. Q. Yang, R. Alfred, H. Mahdin, and R. Yaakob, "Ensemble deep learning for tuberculosis detection," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 17, no. 2, pp. 1014–1020, 2019, doi: 10.11591/ijeecs.v17.i2.pp1014-1020.

[31] R. Hooda, A. Mittal, and S. Sofat, "Automated TB classification using ensemble of deep architectures," *Multimed. Tools Appl.*, vol. 78, no. 22, pp. 31515–31532, 2019, doi: 10.1007/s11042-019-07984-5.

[32] T. Rahman *et al.*, "Reliable tuberculosis detection using chest x-ray with deep learning, segmentation and visualization," *arXiv*, no. August, 2020, doi: 10.1109/access.2020.3031384.

[33] S. K. T. Hwa, M. H. A. Hijazi, A. Bade, R. Yaakob, and M. S. Jeffree, "Ensemble deep learning for tuberculosis detection using chest X-ray and canny edge detected images," *IAES Int. J. Artif. Intell.*, vol. 8, no. 4, pp. 429–435, 2019, doi: 10.11591/ijai.v8.i4.pp429-435.

[34] "First choice," 2022. http://firstchoice-testing.com/services/occupational-health/.

[35] "CXR Image," *HealthEngine*. https://healthinfo.healthengine.com.au/chest-x-ray (accessed Jun. 22, 1BC).

[36] T. Fujioka *et al.*, "The Utility of Deep Learning in Breast Ultrasonic Imaging: A Review," *Diagnostics*, vol. 10, no. 12, p. 1055, 2020, doi: 10.3390/diagnostics10121055.

[37] J. Schmidhuber, "Deep Learning in neural networks: An overview," *Neural Networks*, vol. 61. pp. 85–117, 2015, doi: 10.1016/j.neunet.2014.09.003.

[38] L. Deng and D. Yu, "Deep learning: Methods and applications," *Found. Trends Signal Process.*, vol. 7, no. 3–4, pp. 197–387, 2013, doi: 10.1561/2000000039.

[39] X. W. Chen and X. Lin, "Big data deep learning: Challenges and perspectives," *IEEE Access*, vol. 2, pp. 514–525, 2014, doi: 10.1109/ACCESS.2014.2325029.

[40] I. Namatēvs, “Deep Convolutional Neural Networks: Structure, Feature Extraction and Training,” *Inf. Technol. Manag. Sci.*, vol. 20, no. 1, pp. 40–47, 2018, doi: 10.1515/itms-2017-0007.

[41] Y. Lecun, L. Bottou, Y. Bengio, and P. Ha, “Gradient-Based Learning Applied to Document Recognition,” *Proc. IEEE*, no. November, pp. 1–46, 1998.

[42] Prabhu, “Understanding of Convolutional Neural Network (CNN) Deep-Learning.” https://medium.com/@RaghavPrabhu/understanding-of-convolutional-neural-network-cnn-deep-learning-99760835f148.

[43] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-Decem, pp. 770–778, 2016, doi: 10.1109/CVPR.2016.90.

[44] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.*, pp. 1–14, 2015.

[45] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the Inception Architecture for Computer Vision,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-Decem, pp. 2818–2826, 2016, doi: 10.1109/CVPR.2016.308.

[46] A. G. Howard *et al.*, “MobileNets: Efficient convolutional neural networks for mobile vision applications,” *arXiv*, 2017.

[47] O. Russakovsky *et al.*, “ImageNet Large Scale Visual Recognition Challenge,” *Int. J. Comput. Vis.*, vol. 115, 2014, doi: 10.1007/s11263-015-0816-y.

[48] A. A. Almisreb, N. Jamil, and N. M. Din, “Utilizing AlexNet Deep Transfer Learning for Ear Recognition,” in *2018 Fourth International Conference on Information Retrieval and Knowledge Management (CAMP)*, Mar. 2018, pp. 1–5, doi: 10.1109/INFRKM.2018.8464769.

[49] M. Cicero *et al.*, “Training and Validating a Deep Convolutional Neural Network for Computer-Aided Detection and Classification of Abnormalities on Frontal Chest Radiographs,” *Invest. Radiol.*, vol. 52, no. 5, pp. 281–287, 2017, doi: 10.1097/RLI.0000000000000341.

[50] J. Jordan, “Common architectures in convolutional neural networks,” 2018. https://www.jeremyjordan.me/convnet-architectures/ (accessed Jan. 15, 2020).

[51] Sasha Targ, D. Almeida, and K. LymanEnlitic, “RESNET IN RESNET: Genaralizing Residual Architectures.” p. 7, 2016, doi: arXiv:1603.08029v1.

[52] H. Sak, A. Senior, and F. Beaufays, “Long short-term memory recurrent neural network architectures for large scale acoustic modeling,” *In INTERSPEECH-2014*, pp. 338-342., 2014.

[53] J. Huang *et al.*, “Speed/accuracy trade-offs for modern convolutional object detectors,” *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, vol. 2017-Janua, pp. 3296–3305, 2017, doi: 10.1109/CVPR.2017.351.

[54] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, “MobileNetV2: Inverted Residuals and Linear Bottlenecks,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 4510–4520, 2018, doi: 10.1109/CVPR.2018.00474.

[55] A. T. Sahlol, M. A. Elaziz, A. T. Jamal, R. Damaševičius, and O. F. Hassan, “A novel method for detection of tuberculosis in chest radiographs using artificial ecosystem-based optimisation of deep neural network features,” *Symmetry (Basel).*, vol. 12, no. 7, 2020, doi: 10.3390/sym12071146.

[56] E. J. Hwang *et al.*, “Development and Validation of a Deep Learning-based Automatic Detection Algorithm for Active Pulmonary Tuberculosis on Chest Radiographs,” *Clin. Infect. Dis.*, vol. 69, no. 5, pp. 739–747, 2019, doi: 10.1093/cid/ciy967.

[57] M. F. Alcantara *et al.*, “eRx – A technological advance to speed-up TB diagnostics,” *Smart Heal.*, vol. 16, no. September 2019, pp. 1–11, 2020, doi: 10.1016/j.smhl.2020.100117.

CURRICULUM VITAE

Muhammad Zaharaddeen Abubakar

[redacted]