Turkish Journal of Science & Technology
19(1), 55-64, 2024

Research Paper
https://doi.org/10.55525/tjst.1222836

Automated Tuberculosis Classification with Chest X-Rays Using Deep Neural Networks -Case Study: Nigerian Public Health

Muhammad Zaharaddeen Abubakar[1], Mustafa Kaya[2*], Mustafa Eris[3], Muhammad Mansur Abubakar[4], Serkan Karakuş[5], Khalid Jibril Sani[6]

[1,2,3,6] Digital Forensic Engineering, Faculty of Technology, Firat University, Elazığ, Türkiye
[4] Computer Engineering, Faculty of Engineering, Firat University, Elazığ, Türkiye
[5] Software Engineering, Faculty of Engineering, Firat University, Elazığ, Türkiye
[*] mkaya@firat.edu.tr

(Geliş/Received: 26/12/2022; Kabul/Accepted: 13/02/2024)

Abstract: Tuberculosis, a contagious lung ailment, stands as a prominent global mortality factor. Its significant impact on public health in Nigeria necessitates comprehensive intervention strategies. Detecting, preventing, and treating this disease remains imperative. Chest X-ray (CXR) images hold a pivotal role among diagnostic tools. Recent strides in deep learning have notably improved medical image analysis. In this research, we harnessed publicly available and proprietary CXR image datasets to construct robust models. Leveraging pre-trained deep neural networks, we aimed to enhance tuberculosis detection. Impressively, our experimentation yielded remarkable outcomes. Notably, f1-scores of 98% and 86% were attained on the respective public and private datasets. These results underscore the potency of deep neural networks in effectively identifying tuberculosis from CXR images. The study emphasizes the promise of this technology in combating the disease's spread and impact.

Key words: Deep learning, Tuberculosis (TB), ResNet, MobileNet, Data augmentation.

Derin Sinir Ağlarını Kullanan Göğüs Röntgenleri ile Otomatik Tüberküloz Sınıflandırması Örnek Çalışma: Nijerya Halk Sağlığı

Öz: Bulaşıcı bir akciğer rahatsızlığı olan tüberküloz, önde gelen küresel ölüm faktörü olarak karşımıza çıkıyor. Nijerya'da halk sağlığı üzerindeki önemli etkisi, kapsamlı müdahale stratejilerini gerektirmektedir. Bu hastalığın tespit edilmesi, önlenmesi ve tedavi edilmesi hâlâ zorunludur. Tanı araçları arasında göğüs röntgeni (CXR) görüntüleri çok önemli bir role sahiptir. Derin öğrenmedeki son gelişmeler, tıbbi görüntü analizini önemli ölçüde iyileştirdi. Bu araştırmada, sağlam modeller oluşturmak için kamuya açık ve tescilli CXR görüntü veri kümelerinden yararlandık. Önceden eğitilmiş derin sinir ağlarından yararlanarak tüberküloz tespitini geliştirmeyi hedefledik. Etkileyici bir şekilde, deneylerimiz dikkate değer sonuçlar verdi. Özellikle, ilgili açık ve özel veri setlerinde %98 ve %86'lık f1 puanlarına ulaşıldı. Bu sonuçlar, CXR görüntülerinden tüberkülozun etkili bir şekilde tanımlanmasında derin sinir ağlarının gücünün altını çiziyor. Çalışma, bu teknolojinin hastalığın yaygın etkisiyle mücadelede umut vaat ettiğini belirliyor.

Anahtar kelimeler: Derin öğrenme, Tüberküloz, ResNet, MobileNet, Veri arttırma.

1. Introduction

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 that 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 being 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 of 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 – 7] .

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

---

[*] Corresponding author: mkaya@firat.edu.tr. [1] 0000-0002-0240-6827, [*2]0000-0002-0160-4469, [3]0000-0002-1757-8496, [4]0000-0001-8410-0057, [5]0000-0001-5639-1408, [6]0000-0003-2593-243X

in Africa (after South Africa, Ethiopia, and DR Congo) [8, 9]. Due to the rate of the disease in the region, TB is considered a national emergency problem in Nigeria.

In a study conducted in 2020, B. Odume et al. investigated the impact of the COVID-19 epidemic on the findings of tuberculosis cases in Nigeria. It was emphasized that after the global epidemic, the TB program urgently needs to adapt to the new normal and strengthen the patient-centered approach to TB care, adopt digital health technology, increase awareness, and increase community-based active TB case finding [10].

However, Various approaches to 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 problems [11, 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 the 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 are now widely adopted by the research communities [13]. These models have been applied in different fields of medicine such as; brain tumor [14], breast cancer [15], COVID-19 [16], Heart failure [17], Diabetes detection [18], and detection of thoracic pathologies [19].

However, several studies have used custom CNN models in their research to classify TB positive and negative from CXR images. F. Pasa et al [20] built a custom deep-learning CNN model with five (5) convolutional blocks, similar to AlexNet [21]. In their study, they scored an Accuracy of 86.2% and an AUC of 88.4% on the Maryland (MC), and Shenzhen (SZ) datasets combined. They believe their model is competitive even if comparison with other techniques is difficult, hence argued that saliency maps and gradient-weighted class activation mapping (Grad-CAMs) are useful tools for an in-depth visual explanation of TB, which they believe will provide an approximate visual diagnosis and might in-turn be a useful additional tool for clinicians.

Y. Xiong [22] also designed a custom CNN model called (TB-AI) which was originally trained on the Google CIFAR-10 dataset. The custom model was trained using 201 samples which were collected as test and used to examine Tb-AI. Meanwhile, further pre-processing techniques were carried out to have enough data to curb the problem of training a few datasets which might result in type one error and type two error even though they claim to have a human pathologist on the ground to analyze inconsistencies between their model and human. TB-AI scored 97.94% sensitivity score and a specificity of 83.65% score.

1.1. Related works

E. J. Hwang et al [23] developed a custom DCNN-based automatic detection (DLAD) model made up of 27 layers with 12 residual connections, trained via a semi-supervised localization approach and only a portion of the training data was annotated. The last layer of the model was made up of Image classification and lesion localization (segmentation of the lung), using 60,768 CXRs images collected from Seoul National University Hospital (SNUH), to classify TB, and further performance of their model was tested on 6 independents datasets to confirm the generalization performance. Certified radiologists labeled and annotated the CXR images and the performance of the model was further observed by other physicians and high cutoffs were set for Sensitivities and specificities, and scored 98% through in-house validation. Which demonstrated excellent performance in TB detection, outperforming physicians and other several studies such as P. Rajpurkar et al [24].

Meanwhile, some other studies used pre-trained CNNs with the concept TL to binary classify TB into positive and negative. T. Rahman et al [25] 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 9 pre-trained DL CNN techniques, with the help of TL, from their pre-trained initial weights, they trained, validated, and hence tested to obtain 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 [26] also believe DCNN for TB screening is a promising DL method for various visualization tasks since it enables 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 further used TL to binary detect TB infections on the three available public datasets, and scored the 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.

R. Hooda [27] presented an ensemble of three standard architectures used for binary image classification with AlexNet, GoogLeNet, and ResNet. where the architectures were all trained from scratch. The dataset is the combination of four public datasets which consist of 1.133 images that were pre-processed using augmentation to increase the number of images as in [25] and other similar studies. Their ensemble model achieves an accuracy of 88% and an AUC of 93%. They believe in a DL-based CAD system, segmenting lung region is not mandatory as opposed to [24].

M. H. A. Hijazi et al [28] also used the two publicly available datasets MC and SZ, in which they first pre-processed the CXRs images to retain only the Region of Interest (ROI). And also used two different pre-trained architectures; VGG 16 and InceptionV3, and further developed a custom CNN architecture with 15 layers to detect TB from the preprocessed CXR images, where the process took less time to complete training. And further performed an ensemble of the architectures, where it scored the best accuracy of 91.0%, sensitivity of 89.6%, and specificity of 90.7% than the individual architecture.

S. K. T. Hwa, M. H. A. Hijazi et al [29] also used the publicly available datasets MC and SZ similar to other literature, in which they first pre-processed the CXRs images to obtain edge feature using Canny Edge detector, which they believe CXR images with more unusual edges could increase the detection rate. They perform an ensemble of two pre-trained CNN architectures as in [25] and [28]. They believe sensitivity is considered more in medical image analysis; hence their work was to get a high sensitivity score.

Furthermore, some studies used private datasets, such as; P. Lakhani and B. Sundaram [30] that used four de-identified HIPAA-compliant datasets exempt from review by the institutional review board (which is believed to be one of the problems of medical datasets), They employed two different DCNNs which were AlexNet and GoogLeNet to binary classify the TB CXR images by training from scratch and by using pre-trained networks on ImageNet (where the pre-trained network performed better), and further perform augmentation with multiple preprocessing techniques which further increased accuracy. Ensembles were further performed on the best-performing algorithms using AlexNet and GoogLeNet architecture with an AUC score of 99%. For cases where the classifiers had type one error, an independent board-certified cardiothoracic radiologist interpreted the result without seeing the model result. Their study also shows that even with CAD systems there is a need for cardiothoracic radiologist experts to further give verdict on the result.

1.2. 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 ranging from custom DCNNs to binary classifying of TB.

In our study, we present the use of several pre-trained DCNN architectures to make binary classifications on our private dataset from the western Africa region, Nigeria in particular as most of the studies in the literature are from Europe and Asia, and we further compare our result using the same models on public 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, f1-score, specificity, and sensitivity at 94% respectively for the latter and 94% precision, 94% recall, 94% f1-score, 96% specificity, and 92% sensitivity for the former. Whereas for our public dataset, we achieved 99% accuracy, sensitivity, specificity, precision, recall, and f1-score respectively with VGG16 and VGG19, and 99% sensitivity, with 100% accuracy, specificity, precision, recall, and f1-score with MobileNetV1, as our best performing DCNN models while using the public dataset.

The proposed approach is presented in Section 2. The discussion and results are described in Section 3. The conclusion and future work are presented in Section 4.

2. Materials

This study involves two separate datasets; (I) a processed public dataset that includes 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 identified dataset in compliance with the Health Insurance and Accountability Act from Nigeria (West Africa). TB case images of the Nigerian people in private datasets consist of images obtained with different imaging devices and are generally of low resolution and quality. On the other hand, the images obtained from the public data set have higher resolution and can be used directly. Positive and negative image counts in public and private datasets are shown in Table 1. From Table 1. the public dataset consists of a total of 7000 CXR images with 3500 as negative and 3500 as positive. Meanwhile, our private dataset consists of a total of 4415 unprocessed CXR images with different pixel values with 3073 as negative and 1342 as positive as shown in Table 2. 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 the 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, 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 was only used for model evaluation to calculate the performance of the architecture.

Table 1. Class-wise image counts in the public and private datasets

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

For our private data collection; (I) A mobile X-ray unit was used, provided by the 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 privately 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. Method

The methodology of this study was carried out in three different stages as listed below and shown schematically in Figure 1.

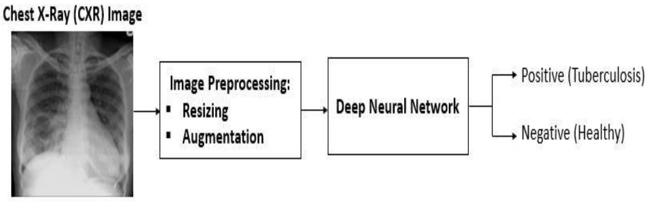


Figure 1. Schematic flow of the study

I. The public dataset was used for the binary classification of TB positive and negative cases using several trained DCNN models as shown in Table 4,
II. The private dataset was also used for the binary classification of TB positive and negative cases using the same pre-trained models from the first method as shown in Table 3. respectively.
III. The pre-trained model's configuration was 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 datasets.

3.1. 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 accept 224*224 except for NasNet which accept only 331*331 as input shape, hence we also converted the images from grayscaled to Blue, Green, Red (BGR) to obtain three channels since pre-trained 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 standards the same preprocessing techniques 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.

3.2. Preprocessing

Table 2. Class-wise image counts in the augmented dataset

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 2. 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 helps us have balanced 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, zoom, and fill mode) were utilized to generate more images, as shown in Figure 2. In this work, the original image was zoom range was set to 20% max, with fill mode set to reflect for convenience.

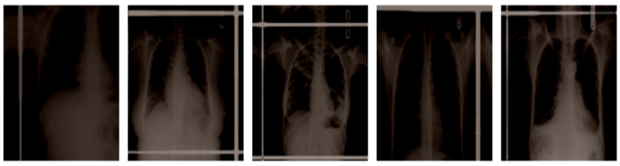


Figure 2. Sample images from the augmented dataset

3.3. Performance evaluation

Many different methods were used in the literature for the performance evaluation of the DCNN models. In our study, the performances of the classification methods were evaluated with sensitivity, specificity, precision, and F1 scores given in Equations 1-5 respectively. The binary classification of the negative and positive class of our models was hence schematically presented with the help of a confusion matrix to showcase the indicators; True Positive, (TP), True Negative (TN), False Positive (FP), and False Negative (FN) as result where the model correctly predicts the positive, negative class, and incorrectly predicts the positive, negative class, respectively.

$$\text{Accuracy} = (TP+TN)/(TP+FN) + (FP+TN) \tag{1}$$

$$\text{Sensitivity} = (TP)/(TP+FN) \tag{2}$$

$$\text{Specificity} = (TN)/(FP+FP) \tag{3}$$

$$\text{Precision} = (TP)/(TP+FP) \tag{4}$$

$$\text{F1 Score} = (2TP)/(2TP+FN+FP) \tag{5}$$

The F1 score is a critical evaluation metric commonly used in medical image classification tasks and other binary classification scenarios. It strikes a balance between precision (the ratio of true positive predictions to the total predicted positives) and recall (the ratio of true positive predictions to the total actual positives).

In medical image classification, such as detecting diseases like tuberculosis from chest X-rays, both precision and recall are of utmost importance. A high precision ensures that the positive predictions made by the model are highly reliable, reducing false positives. This is crucial in medical settings to avoid unnecessary follow-up tests or treatments. On the other hand, a high recall indicates that the model is adept at identifying the actual instances of the disease, reducing false negatives, and ensuring that no cases are overlooked.

The F1 score combines precision and recall into a single metric, providing a comprehensive measure of a model's performance. It is particularly useful when the class distribution is imbalanced, which is often the case in medical scenarios where diseases might be rare. A high F1 score implies a model that strikes a strong balance between accurate positive predictions and comprehensive disease detection.

In medical image classification, a high F1-score signifies that the model is both accurate in identifying the disease and comprehensive in capturing most cases, instilling confidence in its clinical applicability and potential to aid medical professionals in accurate diagnosis.

4. Results and discussions

Table 3. Performance of different models for the classification of TB with private data

Models	**Acc.**	**Weighted Average**				
		Sen.	**Spe.**	**Pre.**	**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

The motivation behind this study is to use a private data set from a highly affected region and to compare it with a more prepared public dataset from a different region, hence examining both regions' datasets by using different pre-trained and custom DCCN algorithms to assist physicians in the diagnosis of TB in affected populations. In our Study, the CXR images were binary classified (+ve and -ve). For each DCNN model used, the learning rate was set to 0.0001 and the epoch set to 150 respectively. The activation function was set to Softmax and categorical cross-entropy was employed to train the model. Table 3 shows the results obtained from the training while using public data (the weighted average of the training result was considered for all the models in the method). Whereas Table 4 shows the result obtained from the training while using our private dataset (the weighted average of the training result from the classification table was considered for all the models in the method).

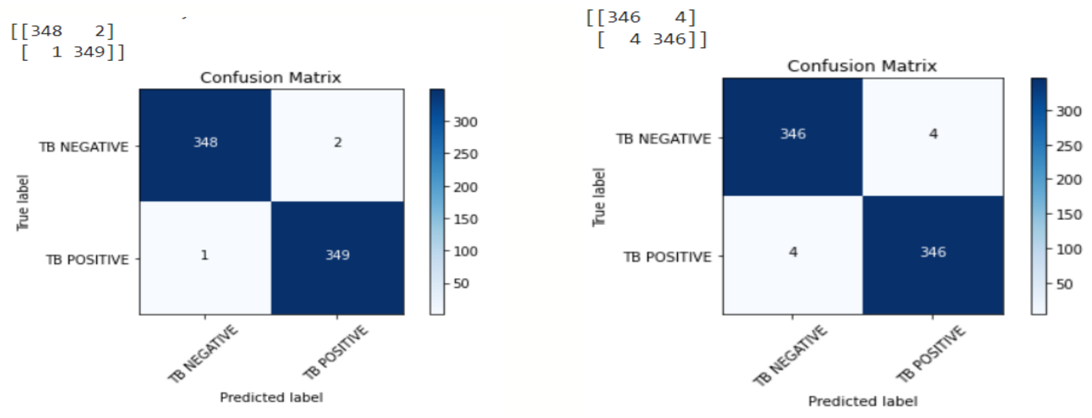
Table 4. Performance of different models for the classification of TB with public data

Models	**Acc.**	**Weighted Average**				
		Sen.	**Spe.**	**Pre.**	**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
InceptionResNet	98	97	99	98	98	98

This study has several limitations such as; Some CXR-images of our private data are skewed and have no standard due to the machines being 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 pose a large scale of the black area which we cropped some part of that because which we believe it can be a problem. Nonetheless, we believe that our results could establish a foundation for future prospective research for the detection of TB. Furthermore, regarding 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.

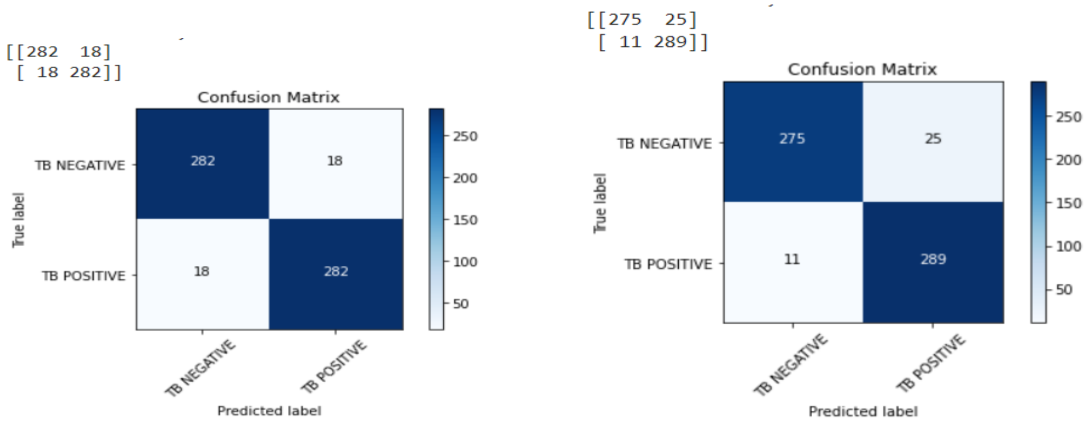
In our first method using pre-trained models to train our public data, we obtained a state-of-the-art performance as in T. Rahman et al [25], and a good result while using the second method for the binary classification.

A general comparison with other studies can be achieved while using our private data as it has shown great promise, and is competitive, which implies that our study performs well. Figure 3. using the confusion matrix systematically explains the performance of our models on our datasets in detecting TB, Figure 3. in our binary classification shows that our model wrongly classified eleven (11) patients to be TB-negative when they were TB-positive patients, and predicted twenty-five (25) patients to be TB positive while they were negative with a sensitivity and specificity of 92%, and 96% respectively meanwhile all other predictions were correct. In Figure 3. a) 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 TB positive patients and also predicted two (2) patients to be TB positive when they were actually TB negative. Then, respectively, in Figure 3. b) the public data set, which was trained with the VGG16 model, and in c) and d), the private data set was used with the VGG16 and MobileNet models, and confusion matrices were given.



a) MobileNet V1 Model with Public Data

b) VGG16 with Public Data



c) VGG16 with Private Data

d) MobileNet V1 Model with Private Data

Figure 3. Confusion Matrix of Optimal Models on Public and Private Datasets

5. Conclusion

In conclusion, a pre-trained deep convolutional neural network has been proposed to effectively detect tuberculosis from CXR images. Several pre-trained deep convolutional neural networks were applied to the public and private CXR images. These pre-trained models used in our experiment includes: VGG16, VGG19, ResNet152, ResNet50V2, ResNet101V2, ResNet152, MobileNetV1, MobileNetV2, InceptionV3, and InceptionResNet. The result of this experiment shows that MobileNetV1 achieved the best performance when applied to the public dataset. And 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 conclusion, this study demonstrated the effectiveness of deep neural networks (DNN) in tuberculosis detection from chest X-ray images. Our best model achieved impressive f1-scores of 98% and 86% on public and private datasets, respectively.

This underscores the robustness of DNNs in accurately identifying tuberculosis cases, showcasing their potential as a valuable tool for enhancing disease detection efforts. These findings carry significant implications for tuberculosis detection in Nigeria, a country where the disease is a substantial public health concern. The utilization of advanced image analysis techniques, as evidenced by our study, could significantly aid healthcare professionals in swiftly and accurately identifying tuberculosis cases. This could potentially lead to earlier interventions, reduced transmission rates, and improved patient outcomes.

From a forward-looking perspective, future research in this area could explore the integration of multi-modal data, such as combining X-ray images with clinical data or other diagnostic modalities. Furthermore, investigating the generalizability of the deep learning models across different populations and settings would enhance the applicability of these technologies on a broader scale. Additionally, efforts could be directed toward creating user-friendly tools for real-time decision support in clinical practice, further enhancing the impact of tuberculosis detection initiatives in resource-constrained environments.

References

[1] Desmon S. "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/ (Last accessed Aug. 29, 2023).

[2] WHO. "The End TB Strategy Global strategy and targets for tuberculosis prevention, care and control after 2015," 2015, https://apps.who.int/iris/rest/bitstreams/1271371/retrieve. (Last accessed Aug. 29, 2023).

[3] Alcantra MF, Liu C, Liu B, Brunette M, Zhang N, Sun T, Zhang P, Chen Q, Li Y, Albarracin CM, Peinado J, Garavito ES, Garcia LL, Curioso WH. "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. Global Tuberculosis report 2016, https://apps.who.int/iris/bitstream/handle/10665/250441/9789241565394-eng.pdf?sequence=1 (Last accessed Aug. 29, 2023)

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

[6] CDC A. "Tuberculosis," Tuberculosis in Africa. https://africacdc.org/disease/tuberculosis/#:~:text=TB, Africa CDC, (Last accessed Aug. 29, 2023).

[7] WHO. "World Tuberculosis Day," 2018. https://www.afro.who.int/health-topics/tuberculosis-tb. (Last accessed Aug. 29, 2023).

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

[9] Chukwu O and Okoeguale B. "FIRST National TB Prevalence Survey 2012, Nigeria," Nigeria, 2012. [Online]. Available: https://www.who.int/tb/publications/NigeriaReport_WEB_NEW.pdf. (Last accessed Aug. 29, 2023).

[10] Odume B, Falkun V, Chukwuogo O, Ogbudebe C, Useni S, Nwokoye N, Aniwada E, Olusola Faleye B, Okekearu I, Nongo D, Odusote T, and Lawanson A. Impact of COVID-19 on TB active case finding in Nigeria, Public Health Action. 2020 Dec 21; 10(4): 157–162. doi: 10.5588/pha.20.0037

[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). (Last accessed Aug. 29, 2023).

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

[13] Lopes UK and Valiati JF. "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] Chen B, Zhang L, Chen H, Liang K, and Chen X. "A novel extended Kalman filter with support vector machine based method for the automatic diagnosis and segmentation of brain tumors", Comput. Methods Programs Biomed. 2021 Mar;200:105797. doi: 10.1016/j.cmpb.2020.105797. PMID: 3331787

[15] Houssein EH, Emam MM, Ali AA, and Suganthan PN. "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] Ozturk T, Talo M, Yildirim EA, Baloglu UB, Yildirim O, and Rajendra Acharya U. "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] Olsen CR, Mentz RJ, Anstrom KJ, Page D, and Patel PA. "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] Chaki J, Thillai Ganesh S, Cidham SK, and Ananda Theertan S. "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] Rajpurkar P 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] Pasa F, Golkov V, Pfeiffer F, Cremers D, and Pfeiffer D. "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] Krizhevsky A, Sutskever I, and Hinton GE. "ImageNet Classification with Deep Convolutional Neural Networks,"

Commun. ACM, vol. 60, no. 6, pp. 84–90, June 2017. doi = 10.1145/3065386

[22] Xiong Y, Ba X, Hou A, Zhang K, Chen L, and Li T. “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.

[23] Hwang EJ *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.

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

[25] Rahman T *et al.*. “Reliable tuberculosis detection using chest x-ray with deep learning, segmentation, and visualization,” IEEE Access., August, 2020, doi: 10.1109/access.2020.3031384.

[26] Hwang S, Kim HE, Jeong J, and Kim HJ. “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.

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

[28] Hijazi MHA, Yang LQ, Alfred R, Mahdin H, and Yaakob R. “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.

[29] Hwa SKT, Hijazi MHA, Bade A, Yaakob R, and Jeffree MS. “Ensemble deep learning for tuberculosis detection using chest X-ray and canny edge detected images,” *Int. J. Artif. Intell.*, vol. 8, no. 4, pp. 429–435, 2019, doi: 10.11591/ijai.v8.i4.pp429-435.

[30] Lakhani P and Sundaram B. “Deep Learning at Chest Radiography: Automated Classification of Pulmonary Tuberculosis by Using Convolutional Neural Networks.” Radiology, vol. 284, no. 2, pp. 574–582, Aug., 2017. doi: 10.1148/radiol.2017162326. Epub 2017 Apr 24. PMID: 28436741.