Shahrood University of Technology

Journal of Artificial Intelligence and Data Mining (JAIDM)
Journal homepage: http://jad.shahroodut.ac.ir

Research paper

Diagnosis and Classification of Tuberculosis Chest X-ray Images of Children Less Than 15 years at Mbarara Regional Referral Hospital Using Deep Learning

Simon Kawuma[1*], Elias Kumbakumba[2], Mabirizi Vicent[3], Deborah Nanjebe[4], Kenneth Mworozi[5], Adolf Oyesigye Mukama[6], Lydia Kyasimire[7]

1 Software Engineering, Mbarara University of Science and Technology, Mbarara, Uganda.
2 Department of Paediatrics, Mbarara University of Science and Technology, Mbarara, Uganda.
3 Department of Information Technology, Kabale University, Kabale, Uganda.
4 Epicentre Mbarara Research Centre, Mbarara, Uganda.
5 Department of radiology, Mbarara University of Science and Technology, Mbarara, Uganda.
6 Medical Laboratory Science, Epicentre Mbarara Research Centre, Mbarara, Uganda.
7 Department of Paediatrics, Mbarara University of Science and Technology, Mbarara, Uganda.

Article Info

Article History:
Received 01 April 2024
Revised 17 April 2024
Accepted 29 March 2024

DOI:10.22044/JADM.2024.14270.2530

Keywords:
Artificial Intelligence, models, Deep learning, tuberculosis, chest X-ray, convolution neural network

**Corresponding author: simon.kawuma@must.ac.ug (S. Kawuma).*

Abstract

Tuberculosis (TB) is an underestimated cause of death in children, with only 45% of cases correctly diagnosed and reported. It is estimated that 1.12 million TB cases occurred among newborns, children, and adolescents aged less or equal 14 years. In Uganda, TB prevalence is 8.5% in children and 16.7% in adolescents. Treatment and diagnosing TB is challenging and its high mortality rate is due to many lacks in the diagnosis of this illness especially among children. As a strategy to curb TB mortality rate in children, there exists a need to improve and expedite the screening for TB among children. Chest X-ray (CXR) is commonly used in TB burdened countries like Uganda to diagnose TB patients but interpretation of the patient's radiograph needs skilled radiologists who are few. To this end, this research aims to close the TB mortality gap in children by applying AI, primarily deep learning techniques, to detect TB in children. The study created five models, one from scratch and four pre-trained Transfer Learning (TL) and were trained and verified using digital CXR radiograph images of children who visit the TB clinic at Mbarara Regional Referral Hospital. The model classifies clinical images of patients into normal or Tuberculosis. TL models; VGG16, VGG19, Inception V3, and ResNet50 outperformed scratch model with validation accuracy of 79.91%, 69.21%, 53.0%, 51.09% and 50.01% respectively. We hope that once the deep learning models are implemented and adopted by the radiologist, it will reduce the time spent by radiologist while analysing CXR images.

1. Introduction

Tuberculosis (TB) is a fatal infectious illness caused by the bacillus Mycobacterium tuberculosis, and spread when infected people expel the bacteria into air, for example by coughing. TB was the highest cause of death from a single infectious agent until the coronavirus epidemic. In 2019, an estimated 10 million people contracted tuberculosis (TB), with 1.2 million TB fatalities among HIV-negative persons and an additional 208, 000 deaths among HIV-positive people, despite the fact that TB is preventable and curable. Adults represented for 88% of all TB patients, with children aged ≤15 years accounting for 12%[1]. TB is diagnosed in only 45% of children and at least 1.12 million TB cases in children [2, 3].

The World Health Organization (WHO) reports that the prevalence of TB is highest in South-East Asia (44%), Africa (25%), and the Western Pacific (18%), with lower percentages in the Eastern Mediterranean (8.2%), the Americas (2.9%), and Europe (2.5%). Indonesia (8.5%), China (8.4%), the Philippines (6.0%), Pakistan (5.7%), Nigeria (4.4%), Bangladesh (3.6%), and South Africa (3.6%) have high TB prevalence [1].

The TB burden situation in Uganda is similar, and the country is globally ranked among the top TB/HIV countries, with the largest estimated number of incidence TB cases among HIV-positive people. In 2021, the TB incidence rate was 200 per 100,000 people [4]. In Uganda, TB prevalence is 8.5% in children and 16.7% in adolescents [5].

The United Nations General Assembly meeting in September 2018 set the objective of detecting and treating 40 million Tuberculosis patients by 2022 [6]. One strategy for achieving this goal was to improve and speed up the screening and triage process for active tuberculosis. Although progress is being made, it is expected that the world will not abolish tuberculosis by 2035, as envisioned in the end TB strategy. Inadequacies in identifying people with TB, particularly children, are among the biggest barriers to eradicating the disease [1].

In comparison with other infectious, early diagnosis of tuberculosis is extremely challenging, and numerous diagnostic tests are required to discover TB [7]. Sputum smear microscopy, fast molecular tests, and culture tests are all ways for diagnosing active TB and drug-resistant TB; however, these methods are expensive and not easily available or applicable in low-resourced regions. Because of its relatively high sensitivity, the WHO recommends postero-anterior Chest X-ray (CXR) as one of the key modalities for TB detection and screening [8, 9]. Increased use of CXR could be critical in end TB initiatives, as it has been proven to identify 40-79% of TB patients who do not have any classical TB symptoms [8, 10].

Chest radiographs can be a significant auxiliary diagnostic tool in clinically demanding settings, especially for critically ill patients who require prompt care [11, 12]. However, there are significant intra- and inter-observer differences in CXR interpretation, resulting in TB misdiagnosis, and CXR interpretation is a time-consuming and subjective process. Furthermore, there are similarities in the radiologic patterns of tuberculosis and other lung disorders that can lead to misdiagnosis [13]. Although CXR is a rather effective and economical approach for TB screening and triaging, its use is limited in many high TB-burden countries due to a shortage of qualified radiologists [14] and availability of X-ray machines in the community where patients seek care [15].

Computer-Aided Detection (CAD) technology offer solution to this problem and can be utilized in remote locations to assist radiologist during TB screening [16, 17]. Deep learning AI algorithms have proven extraordinary capabilities and are increasingly being acknowledged as a promising alternative to assist high TB burden countries with few radiologists in screening TB patients. CAD tools employ AI to evaluate radiological images in order to diagnose anomalies and alleviate the shortage of radiologists, particularly in poor nations [18, 19]. CAD tools have been widely utilized to diagnose other diseases such as Alzheimer's, breast cancer, prostate cancer, lung cancer, and others [20-23]. The CAD technology provides an automatic and consistent interpretation in seconds to minutes, which could compensate for the limited number of qualified human readers while also reducing radiologists' workload, and reduce human errors [14]. Convolutional neural networks (CNNs) and deep neural networks (DNNs) are two of the most modern deep learning techniques that can be more accurate than humans for image recognition [16, 24].

Although tuberculosis is curable and preventable, child and adolescent TB is usually overlooked epidemic with a staggering disease burden. Treatment and diagnosing TB is difficulty and its high mortality rate is due to many gaps in the diagnosis of this illness especially among children. As a strategy to curb TB mortality rate in children, there exists a need to improve and expedite the screening for TB among children. CXR are commonly used in TB burden countries like Uganda to diagnose TB patients but interpretation of the patients' radiograph needs skilled radiologist who are few. Also, the X-ray instruments can only be found in referral hospitals rather than at community level where all patient seek care. To this end, this research study seeks to address this gap of TB children mortality by using AI specifically deep learning techniques to diagnose TB in children. The study developed deep learning models which were

trained and validated with digital CXR radiograph images of children who visit the TB clinic at Mbarara Regional Referralhospital.

The remainder of this paper is organised as follow; section 2 presents background information on models we adopted in this study. Section 3 presents literature summary on existing research studies. Section 4 presents research methodology whereas section 5 presents model performance on training and testing dataset. Section 6 presents results and discussion. Section 7 concludes the paper and outlines some avenues for future work. Finally, sections 8, 9 and 10 present acknowledgement, authors' conflict of interest and list of references respectively.

2. Deep convolutional neural networks-based Transfer Learning

As discussed earlier, in this study we used a total of five deep CNNs; VGG16, VGG19, Inception V3, and ResNet50 in their pre-trained initial weights for TL. Together with our proposed scratch model, these models were trained, validated and tested for classifying chest X-ray images into non-TB normal and TB cases. In this section, therefore, we expound more on these networks.

The Visual Geometry Group (VGG) has two different variants; VGG16 and VGG19 based on the number of the layers in the network [25]. These models are known for their deep architecture consisting of multiple convolutional layers making them suitable for capturing complex features in an image. During training phase, the VGG16 and VGG19 models adjust its internal weights to minimize the difference between prediction and the actual labels of the training image to allow the model learn the features that are most relevant for TB detection. This process is known as backpropagation [26].

The Inception V3 is a convolutional neural network architecture from the Inception family. It consists of 48 layers of symmetric and asymmetric building blocks; pooling, convolutional and auxiliary classifiers [27]. Inception V3 is commonly known for its inception modules which maximise computational resources through parallel processing of different filter sizes within the same layer [28]. In this study, Inception V3 was utilised to improve the computational efficiency of the proposed scratch model in TB chest X-ray analysis performance without compromising model accuracy.

Like VGG, Residual Network (ResNet) has got several variants including ResNet18, ResNet50, ResNet101 and ResNet152 [25]. Overtime ResNet has been applied in the medical research and a lot of success has been registered in biomedical image classification. In practice, ResNet50 consists of a series of convolutional layers, followed by batch normalisation and ReLu action functions. ResNet50 is renowned for its deep residual learning framework, which facilitates training of deep neural networks. In the context of this study, ResNet50 was employed to further enhance detection and classification of TB-related abnormalities in chest X-ray images leveraging its ability to learn complex patterns and features.

3. Literature review

Overtime, diagnosis of pulmonary infections for children under 15 years has been difficult, and patients who have been misdiagnosed have had poor treatments. According to a report by Dr. Barnes at the USC school of medicine in Log Angels in his study on 105 patients, clinical and radiographic characteristics of lung infections may occasionally be difficult to identify from one other [29, 30].

Singh *et al*. [31] reported a case of a 16-year-old girl who was experiencing lethargy, a nonproductive cough, considerable loss of weight, joint condition and nocturnal fever. Primary treatment for fever and cough was administered and an HIV test was done to narrow down the differential diagnosis to pneumocystis pneumonia, viral pneumonia or miliary tuberculosis. Although the patient did not survive, the condition was determined to be tuberculosis according to the results of nucleic acid amplification. Unluckily, the young girl passed away due to the incorrect diagnosis.

To complement clinic diagnosis of pulmonary infection, artificial intelligence powered systems, models and algorithms have been developed to reduce waiting time and diagnosis results. Methods like minimum distance classifiers have been utilized to detect tuberculosis using chest x-ray images [32]. Poornimadevi and Sulochana [33] employed registration based segmentation techniques in their automated system for tuberculosis detection. Registration based segmentation typically rely on similarity metrics such as feature-based metrics to measure similarity between images [34]. These metrics may not always capture the complex anatomical variations presented in the medical images hence affecting model accuracy. A similar system had also been developed by [35].

Parveen and Sathik applied unsupervised fuzzy c-means classification (FCM) techniques to develop an algorithm for detecting pneumonia infections [36]. The fuzzy c-means (FCM) techniques used in this case doesn't explicitly consider the spatial relationship between pixels (treats each pixel

independently) which are crucial in image processing tasks like segmentation and object detection [37]. Besides, FCM is sensitive to noise. Often images contain noise, which significantly impact clustering results of FCM especially where noise is not well separated from the actual image content hence affecting model accuracy [38].

Sharma *et al.* [39] adopted cutting edge technique to detect pneumonia clouds using chest X-ray images. In this study, Otsu thresholding technique was used to separate healthy part of the lungs from pneumonia infected cloudy regions. To compute detection results, the ratio area between healthy and infected regions of the lungs is calculated. Although study findings reveals that researchers were able to separate infected from healthy region of lungs, the Otsu thresholding technique used is sensitive to uncontrolled lighting condition or variations in illumination across an image [40]. Where the lighting conditions are not uniform or if there are large variations in illumination across the image, the background and foreground classes' intensity distribution may overlap, leading to suboptimal thresholding results [40].

Rahman *et al.* [25] proposed a method for detecting TB from chest X-ray images using image processing technique, data augmentation, deep learning classification techniques and image segmentation. In their study, a total of nine deep CNNs were used. The overall model accuracy and F-1 score was 96.8% and 98.54% respectively. Despite the higher detection accuracy (98.45%), visualization technique used in this case to confirm CNNs learning from segmented lung images can be difficult to understand the significance of features at different layers especially when the CNN becomes deeper [41]. The hyper parameters in visualization techniques such as activation maximization can significantly impact the results, leading to different interpretation of the network's learning [42].

Pasa et al. [43] presented a deep CNN with the shortcut connection in comparison to an automated diagnosis with the localization of tuberculosis using two identical datasets. The highest performance accuracy was 92.5%. Whereas the model achieved commendable detection accuracy, application of deep learning in diagnosing TB among children bellow 15 years of age remains elusive.

Besides the limitations discussed in each case above, the dataset used in each case contains images collected from adults (18 years and above) despite the increasing cases of tuberculosis among children below 15 years of age. Thus, it is upon this background that we proposed models for automated diagnosis and classification of tuberculosis among children below 15 years of age using chest X-ray images.

4. Research Methodology

The method steps discussed in this section are presented in Figure 1.

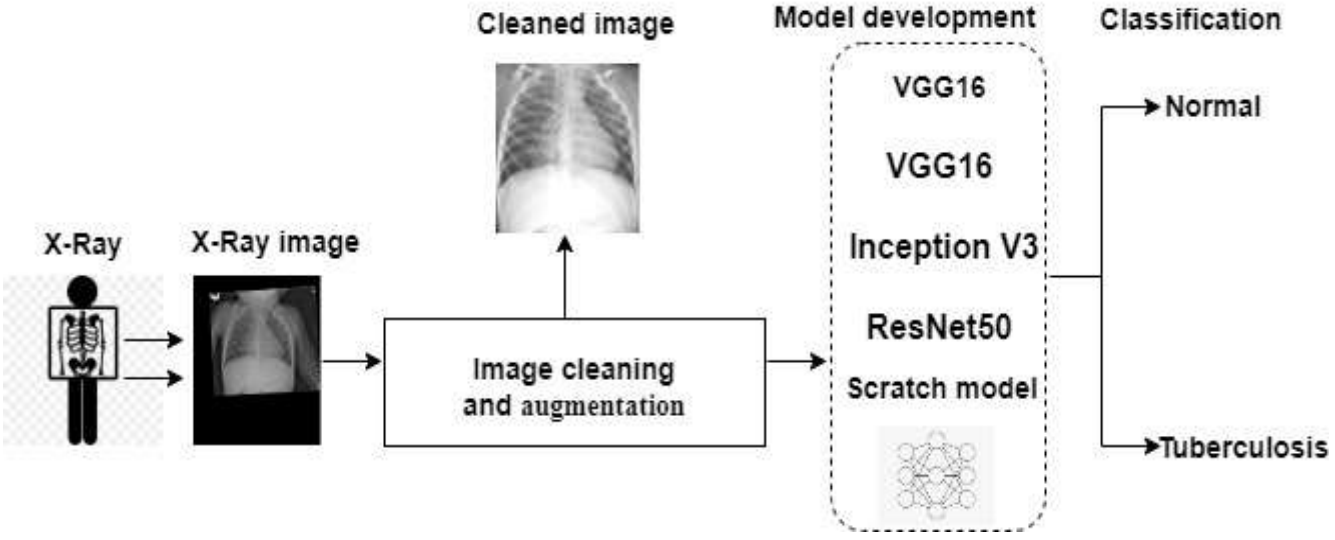

Figure 1. Overview of the methodology step.

4.1 Dataset acquisition

This is a retrospective study involving model development by employing TL to diagnose TB in children. The study used postero-anterior digital chest radiographs of children. The digital x-rays were collected from three recently concluded studies that

were looking at TB diagnosis among children≤15 at the paediatric wards of Mbarara Regional Referral Hospital and the TB-Cohort, Contact and LAM studies carried out at Epicentre, Mbarara University of Science and Technology in Uganda. This dataset consists of 476 CXR images, of which 275 are normal and 201 have TB. All files in this dataset were DICOM files with file extension (dcm).

4.2 Data pre-processing

As earlier mentioned, the study used DICOM images and these image files contain the patient information ID, date of birth, age, sex, pixel data, pixel depth and other information about diagnosis. However, all this information is not needed by the deep learning models and for our scope, we considered pixel data (real image) and the pixel depth (bits encoding) which are required by deep learning model. Also, to protect the patient's privacy from public, we had to remove the metadata of the images such that we do not risk the data to be exposed by mistake to anyone or public. For that reason, we only extracted pixel array of the CXR DICOM images and converted the images to Portable Network Graphics (PNG) format ready to be fed to the models, putting in consideration that we don't lose the quality and data of the image. To extract, pixel array, from DICOM image, we used both the **pydicom** and **NumPy** libraries to read the dicom images and then extracted the pixel data from them and then convert pixel data to type float to avoid overflow or underflow losses when rescaling the image. Before using the images, we rescaled its pixels and put their values between 0 and 255. We then converted the rescaled array to 8 bits unsigned integer (unit8) before we can create the image using the **Pillow** library. Lastly, we used the function save from pillow library to save images as PNG.

4.3 Data augmentation

Because the size of the input photos varied, the CXR images in this investigation were reduced to 256 x 256 pixels. Since the dataset has few images, the study adopted data augmentation techniques to improve model robustness and classification accuracy of deep learning models. Furthermore, data augmentation increases the samples in the datasets for training the models. One method for combating overfitting is data augmentation. Overfitting occurs when a model learns patterns that do not generalize to new data, i.e., the model begins to use irrelevant information to make predictions. Furthermore, to avoid overfitting, this study used Dropout layer to prevent a layer from seeing twice the exact same pattern. We used a zoom range of 0.3, horizontal and vertical flips both set to true, a width shift range of 0.2, a height shift range of 0.2, and a validation split of 0.3 for image augmentation.

4.4 Model Development

In this research study, we created a convolutional neural network (CNN) from scratch (ConvNet) by combining Convolutional layer, Pooling layer, and fully connected layer as discussed below;

The Convolutional layer is a fundamental component in the CNN architecture which is composed of filters for extracting features from the input image [44]. The feature extraction accuracy was affected by the undetected noise during data cleaning. To overcome this limitation, Rectified Linear Unit (ReLu) was used to activate the nodes that receive input signals and perform calculations.

The Pooling layer commonly known as "down sample layer" is responsible for reducing the spatial dimensions of the input data in terms of width and height while retaining the most important information of the input image [44]. Besides, the pooling layer also helped in reducing overfitting during model training and testing by aggregating information from the local regions of the input image hence the developed model was able to focus on most relevant information and ignored minor variations.

Fully connected layer also known as "dense layer" completes the linkage between model layers by connecting each neuron to every other neuron in the adjacent layers. The connection between neurons normalises activations in the network hence improving model training and performance.

The created model is made up of three blocks. In the first block, we used three convolution layers, each with 32 filters and a ReLu function. Following that is a max pooling layer with 64 filters, 3 × 3 in size, and a pool size of (2, 2) for down sampling. The third block is similar to the second block except that 128 filters are employed. To account for overfitting, we used a dropout layer with a 0.2 threshold for each block. To make the results more generalizable, the threshold value was adjusted to 0.5. Finally, a dense layer was used to reduce the vector height from 64 to 2 elements.

Furthermore, since our dataset is small, and yet model development from scratch needs large dataset, this research study adopted VGG16, VGG19, Inception V3 and ResNet TL models to evaluate the performance and accuracy of the model developed from scratch (ConvNet). The developed models classified the images into two categories namely normal or TB as shown in Figure 2.

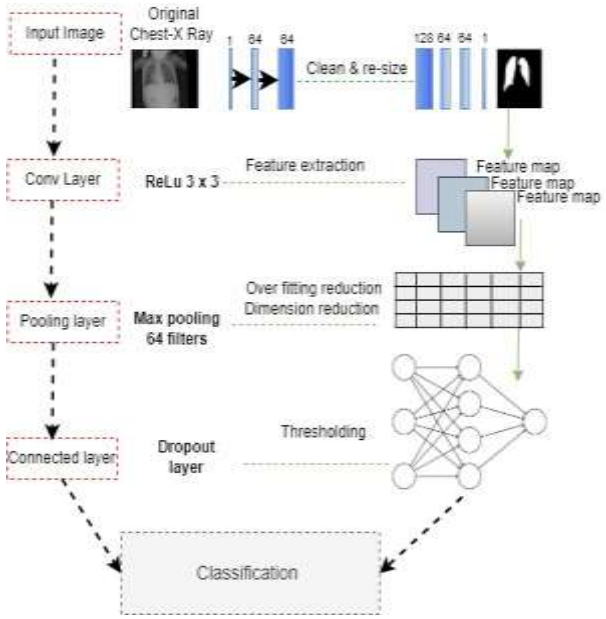

Figure 2. ConvNet model architecture.

5. Model performance on training and testing dataset

To assess model performance, we used precision, accuracy, recall and F1-score and these were used on five pre-trained models including VGG16, VGG19, Inception V 3, ResNet50 and Scratch model (ConvNet). The metrics are defined as below;

Precision is the quality of a positive prediction made by the model i.e., the number of normal images that the model correctly classifies as true positive or true negative denoted by the equation:

$$P = \frac{Tp}{Tp+Fp} \quad (1)$$

Where;

T_p denotes true positive, T_n denotes true negative, F_p denotes false positive, and F_n denotes false negative.

Accuracy is defined as the number of classifications successfully predicted by a model divided by the total number of predictions produced, as expressed by the equation:

$$A = \frac{Tp+Tn}{Tp+Fp+Tn+Fn} \quad (2)$$

Recall which is the percentage of data samples that the model correctly identified as belonging to the class of interest. In this case identifying TB infected images was the class of interest and is denoted by the equation:

$$R = \frac{Tp}{Tp+Fn} \quad (3)$$

F1-score is a metric that computes how many times the developed model makes correct prediction across the entire dataset and is denoted by the equation:

$$\text{F1-score} = 2\left(\frac{PxR}{P+R}\right) \quad (4)$$

The proposed models were trained and tested on a dataset that contained total of 476 including 275 normal and 201 infected with TB.

6. Results and Discussion

Figures 3-7, as well as tables 1 and 2, depict model performance on both training and testing datasets for VGG16, VGG19, Inception V3, ResNet50 and Scratch model (ConvNet) respectively.

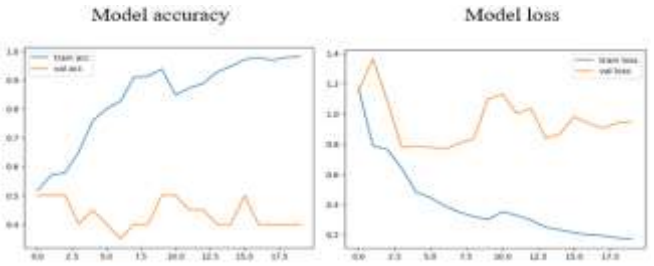

Figure 3. Training accuracy and loss for VGG16 model.

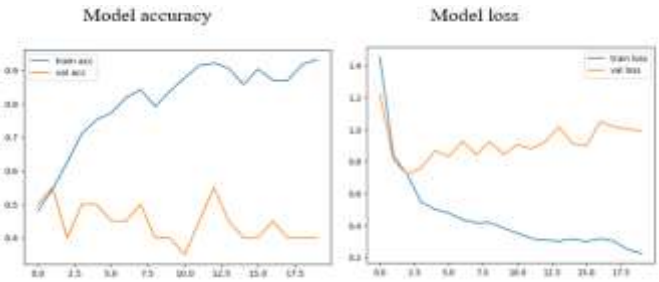

Figure 4. Training accuracy and loss for VGG19 model.

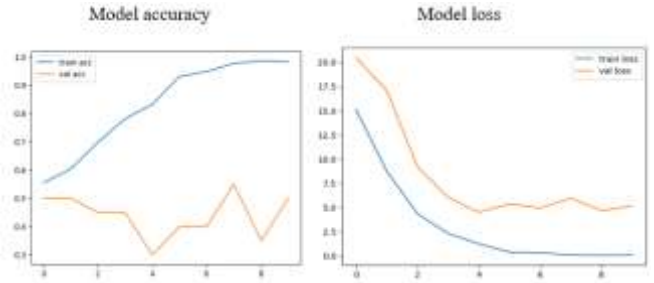

Figure 5. Training accuracy and loss for Inception V 3.

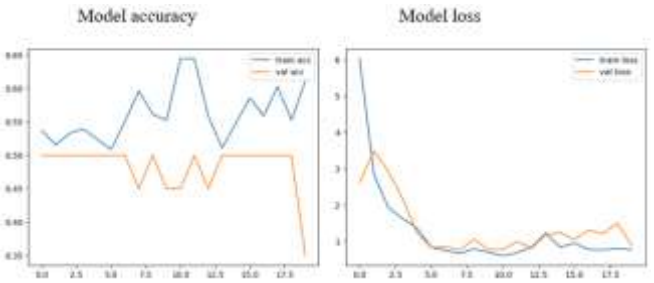

Figure 6. Training accuracy and loss for ResNet50 model.

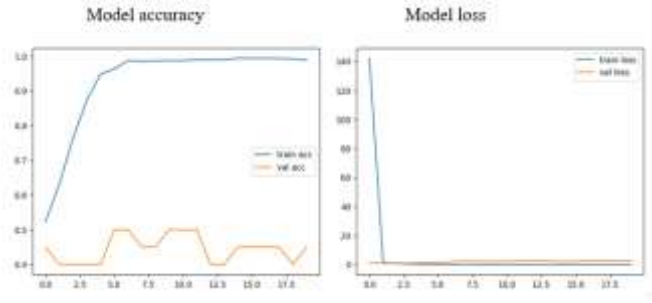

Figure 7. Training accuracy and loss for ConvNet model.

Table 1 outlines the performance metrics of each proposed model for TB detection on training dataset. As shown in Table 1, VGG16 achieved 96.23% with 0.2101 training loss, VGG19 model achieved 90.01% with a training loss of 0.2914. Inception V3 achieved a higher training accuracy of 97.33% with a training loss of 0.1372. Amongst TL models, ResNet50 had a least training accuracy with 63.81% and a training loss of 0.4922. The model developed from scratch (ConvNet) achieved 88.23% accuracy and with training loss of 0.3921.

Table 1. Training model performance.

Model	Precision	Recall	Training Accuracy	F1-Score	Training Loss
VGG16	0.9623	0.9542	0.9623	0.9582	0.2101
VGG19	0.9001	0.9001	0.9001	0.9001	0.2914
Ince V3	0.9733	0.9633	0.9733	0.9683	0.1372
ConvNet	0.8823	0.8752	0.8823	0.8785	0.3921
ResNet50	0.6381	0.6381	0.6381	0.6381	0.4922

As compared to training dataset, model performance on testing dataset and this is resulting from the limited dataset used for testing the model. Thus, we believe if the model is subjected to bigger dataset, performance accuracy would improve. Model performance on testing dataset is presented in Table 2. The VGG16 model achieved better performance compared to other models with 79.91% validation accuracy followed by VGG19 at 69.53%, Inception V3 at 53% whereas ResNet50 had the least validation accuracy of 51.09%. The scratch model validation accuracy was 50.01%.

Table 2. Validation model performance.

Model	Precision	Recall	Validation Accuracy	F1-Score	Validation Loss
VGG16	0.7991	0.7891	0.7991	0.7941	0.6912
VGG19	0.6921	0.6921	0.6921	0.6920	0.4251
Ince V3	0.5300	0.5300	0.5300	0.5300	0.4913
ConvNet	0.5001	0.5001	0.5001	0.5001	0.5821
ResNet50	0.5109	0.5109	0.5109	0.5109	0.7891

There have been numerous research conducted utilizing deep learning techniques to detect TB; for example, authors in [45] utilized CNNs to diagnose tuberculosis using chest x-ray images. The model achieved 88.76% detection accuracy. Hooda et al. utilized deep learning techniques to classify radiographs into tuberculosis and non-tuberculosis. Their model achieved 82.09% detection accuracy [46]. Also, Nguyen et al. carried-out a study to assess the performance of DenseNet model in diagnosing TB using CXR images obtained from Shenzhen and Chine. The performance accuracy was 94% [47]. Although these models achieved a better performance on both training and testing data sets, they were validated from scratch. Our model has been validated from both scratch and TL concept using VGG16, VGG19, Inception V3 and ResNet50 hence reducing the training time and improving model performance using a smaller dataset.

7. Model performance comparison

In this section we present a comparison of the TL models and the scratch model developed in terms of computation complexity, number of layers in each case, and the number of training iterations (epochs) used in each model. Table 3 presents comparison of the TL models used and the scratch model developed. In this comparison we consider the following;

7.1. Computation complexity:

In this study we refer computation complexity to the computation resources (time and space) required by the model to perform tasks including training, prediction and evaluation. The proposed model was developed using Python running Jupiter note book environment on an i7 processor with 16GB RAM equipped with RTX 4070Ti 6GB AMD GPUs from NVIDIA. *Time complexity* refers to the amount of memory required by the model in terms of the amount of input images to be processed whereas *space complexity* is the number of elementary objects required by the model to store during execution. Time and space complexity were computed by analyzing following metrics; best (θ), average (Ω) and worst (O) case scenarios of each model using equations 5, 6 and 7 respectively.

$\theta = f(n) \geq cg(n)$, for $n \geq n0$ …….5

$\Omega = c'g(n) \leq f(n) \leq c''g(n)$, for $n \geq n0$ ….6

$O = f(n) \leq cg(n)$, for $n \geq n0$ 7

Where n0 is the initial number (positive integer) for data input, n is increase in n0, and c is a constant value.

7.2. Epochs:

An epoch is a complete pass through the entire training dataset during the training process. During each epoch, the model updates its parameters based on the gradient of the loss function with respect to the parameters. After completing a single epoch, the model is said to have made one iteration over the entire dataset and has potentially updated its parameters to better fit the dataset. Since the dataset was complex, i.e. it contained DICOM files, we performed multiple epochs based on the model layers to converge attain a desired level of model performance.

7.3. Number of layers:

In the context of deep learning, model layers refer to individual building blocks that make up the architecture of the model. These layers are stacked together to form a neural network. Each layer performs a special task on the input (as discussed earlier in section 4.4) data and transforms it into a meaningful way to produce desired results.

Table 3. Model performance comparison based on computation complexity, epochs and number of layers.

Model	Computation complexity			Number of layers	Epochs(E)
	θ(ms/E)	Ω(ms/E)	O(ms/E)		
VGG16	134400	194400	248400	16	100
VGG19	128400	186600	238800	19	62
Ince V3	122400	178800	239400	48	30
ConvNet	146400	199200	261600	32	26
ResNet50	158400	218400	285600	50	24

8. Conclusion

Our study presents both a model developed from scratch ConvNet and the adopted VGG16, VGG19, Inception V3, and ResNet50 TL approaches were used to classify TB from the chest radiographs for children under 15 years at Mbarara regional referral hospital in Uganda. The ConvNet model achieved 50.01% validation accuracy. Among the pre-trained models, VGG16 had a better validation performance of 79.91% accuracy followed VGG19 with 69.21%, whereas Inception V3 scored 53% and ResNet50 yielded the least performance with 51.09% accuracy. Although data augmentation techniques were used on both sub-datasets, the resulting testing dataset (100) images were not enough to yield better results. Using a larger dataset can improve the accuracy and robustness of our suggested models, which is something that should be addressed in future research as more images are collected.

Acknowledgements

We would like to thank Mbarara Regional Referral Hospital and Epicentre Mbarara Uganda for providing the dataset we used in this study. We would also like to extend our appreciation for the seed funding from the faculty of Medicine at Mbarara University of Science and Technology.

References

[1] J. Chakaya et al., "Global Tuberculosis Report 2020–Reflections on the Global TB burden, treatment and prevention efforts," vol. 113, pp. S7-S12, 2021.

[2] A. M. Mandalakas et al., "Tuberculosis among children and adolescents at HIV treatment centers in sub-Saharan Africa," vol. 26, no. 12, p. 2933, 2020.

[3] P. J. Dodd, C. M. Yuen, C. Sismanidis, J. A. Seddon, and H. E. J. T. L. G. H. Jenkins, "The global burden of tuberculosis mortality in children: a mathematical modelling study," vol. 5, no. 9, pp. e898-e906, 2017.

[4] W. Tumuhimbise, A. J. G. I. R. Musiimenta, and Applications, "Barriers and Motivators of Private Hospitals' Engagement in Tuberculosis Care in Uganda," vol. 1, no. 4, pp. 279-290, 2021.

[5] C. Marquez *et al.*, "The age-specific burden and household and school-based predictors of child and adolescent tuberculosis infection in rural Uganda," vol. 15, no. 1, p. e0228102, 2020.

[6] U. J. N. Y. U. N. Assembly, "Political declaration of the High-Level Meeting of the General Assembly on the Fight Against Tuberculosis: resolution/adopted by the General Assembly," 2018.

[7] S. Sathitratanacheewin, P. Sunanta, and K. J. H. Pongpirul, "Deep learning for automated classification of tuberculosis-related chest X-Ray: dataset distribution shift limits diagnostic performance generalizability," vol. 6, no. 8, p. e04614, 2020.

[8] A. H. van't Hoog *et al.*, "Screening strategies for tuberculosis prevalence surveys: the value of chest radiography and symptoms," vol. 7, no. 7, p. e38691, 2012.

[9] E. Showkatian, M. Salehi, H. Ghaffari, R. Reiazi, and N. J. P. J. o. R. Sadighi, "Deep learning-based automatic detection of tuberculosis disease in chest X-ray images," vol. 87, no. 1, pp. 118-124, 2022.

[10] W. H. Organization, *Global tuberculosis report 2013*. World Health Organization, 2013.

[11] S. Modi et al., "Performance of clinical screening algorithms for tuberculosis intensified case finding among

people living with HIV in Western Kenya," vol. 11, no. 12, p. e0167685, 2016.

[12] P. Rajpurkar et al., "CheXaid: deep learning assistance for physician diagnosis of tuberculosis using chest x-rays in patients with HIV," vol. 3, no. 1, pp. 1-8, 2020.

[13] F. A. Khan *et al.*, "Computer-aided reading of tuberculosis chest radiography: moving the research agenda forward to inform policy," vol. 50, ed: Eur Respiratory Soc, 2017.

[14] Z. Z. Qin et al., "A new resource on artificial intelligence powered computer automated detection software products for tuberculosis programmes and implementers," vol. 127, p. 102049, 2021.

[15] T. Pande, M. Pai, F. A. Khan, and C. M. J. E. R. J. Denkinger, "Use of chest radiography in the 22 highest tuberculosis burden countries," vol. 46, no. 6, pp. 1816-1819, 2015.

[16] M. K. Puttagunta and S. Ravi, "Detection of Tuberculosis based on Deep Learning based methods," in *Journal of Physics: Conference Series*, 2021, vol. 1767, no. 1, p. 012004: IOP Publishing.

[17] R. Hooda, A. Mittal, and S. J. C. M. I. Sofat, "Tuberculosis detection from chest radiographs: a comprehensive survey on computer-aided diagnosis techniques," vol. 14, no. 4, pp. 506-520, 2018.

[18] C. Qin, D. Yao, Y. Shi, and Z. J. B. e. o. Song, "Computer-aided detection in chest radiography based on artificial intelligence: a survey," vol. 17, no. 1, pp. 1-23, 2018.

[19] U. Raghavendra, U. R. Acharya, and H. J. E. n. Adeli, "Artificial intelligence techniques for automated diagnosis of neurological disorders," vol. 82, no. 1-3, pp. 41-64, 2019.

[20] M. Adel et al., "Alzheimer's disease computer-aided diagnosis on positron emission tomography brain images using image processing techniques," in *Computer Methods and Programs in Biomedical Signal and Image Processing*: IntechOpen, 2019, p. 13.

[21] M. Firmino, G. Angelo, H. Morais, M. R. Dantas, and R. J. B. e. o. Valentim, "Computer-aided detection (CADe) and diagnosis (CADx) system for lung cancer with likelihood of malignancy," vol. 15, no. 1, pp. 1-17, 2016.

[22] A. A.-A. Valliani, D. Ranti, E. K. J. N. Oermann, and therapy, "Deep learning and neurology: a systematic review," vol. 8, no. 2, pp. 351-365, 2019.

[23] I. Bakkouri, K. J. M. T. Afdel, and Applications, "Computer-aided diagnosis (CAD) system based on multi-layer feature fusion network for skin lesion recognition in dermoscopy images," vol. 79, no. 29, pp. 20483-20518, 2020.

[24] Z.-P. Jiang, Y.-Y. Liu, Z.-E. Shao, and K.-W. J. A. S. Huang, "An Improved VGG16 Model for Pneumonia Image Classification," vol. 11, no. 23, p. 11185, 2021.

[25] T. Rahman *et al.*, "Reliable tuberculosis detection using chest X-ray with deep learning, segmentation and visualization," vol. 8, pp. 191586-191601, 2020.

[26] A. Khellal, H. Ma, and Q. Fei, "Convolutional neural network features comparison between back-propagation and extreme learning machine," in *2018 37th Chinese Control Conference (CCC)*, 2018, pp. 9629-9634. HIV treatment centers in sub-Saharan Africa," vol. 26, p. 2933, 2020.

[27] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818-2826.

[28] I. Delibasoglu and M. J. J. o. A. R. S. Cetin, "Improved U-Nets with inception blocks for building detection," vol. 14, pp. 044512-044512, 2020.

[29] P. F. Barnes, M. A. Steele, S. M. Young, and L. A. J. C. Vachon, "Tuberculosis in patients with human immunodeficiency virus infection: how often does it mimic Pneumocystis carinii pneumonia?," vol. 102, pp. 428-432, 1992.

[30] L. M. Pinto, A. C. Shah, K. D. Shah, and Z. F. J. R. M. C. Udwadia, "Pulmonary tuberculosis masquerading as community acquired pneumonia," vol. 4, no. 3, pp. 138-140, 2011.

[31] K. Singh, S. Hyatali, S. Giddings, K. Singh, and N. J. C. r. i. c. c. Bhagwandass, "Miliary tuberculosis presenting with ARDS and shock: a case report and challenges in current management and diagnosis," vol. 2017, 2017.

[32] R. N. Rohmah, A. Susanto, I. Soesanti, and M. Tjokronagoro, "Computer Aided Diagnosis for lung tuberculosis identification based on thoracic X-ray," in *2013 International Conference on Information Technology and Electrical Engineering (ICITEE)*, 2013, pp. 73-78.

[33] C. Poornimadevi and H. Sulochana, "Automatic detection of pulmonary tuberculosis using image processing techniques," in *2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, 2016, pp. 798-802.

[34] K. Kuppala, S. Banda, T. R. J. I. J. o. I. Barige, and D. Fusion, "An overview of deep learning methods for image registration with focus on feature-based approaches," vol. 11, pp. 113-135, 2020.

[35] S. Fatima, S. I. A. Shah, M. Z. J. I. J. o. C. Samad, and E. Engineering, "Automated tuberculosis detection and analysis using CXR's images," vol. 10, pp. 284-290, 2018.

[36] N. Parveen, M. M. J. J. o. X.-r. S. Sathik, and Technology, "Detection of pneumonia in chest X-ray images," vol. 19, pp. 423-428, 2011.

[37] J. Xu, T. Zhao, G. Feng, M. Ni, and S. J. I. J. o. F. S. Ou, "A fuzzy C-means clustering algorithm based on spatial context model for image segmentation," vol. 23, pp. 816-832, 2021.

[38] S. J. E. S. w. A. Askari, "Fuzzy C-Means clustering algorithm for data with unequal cluster sizes and contaminated with noise and outliers: Review and development," vol. 165, p. 113856, 2021.

[39] A. Sharma, D. Raju, and S. Ranjan, "Detection of pneumonia clouds in chest X-ray using image processing

approach," in *2017 Nirma University International Conference on Engineering (NUiCONE)*, 2017, pp. 1-4.

[40] M. Vicent, K. Simon, and S. J. I. I. P. Yonasi, "An algorithm to detect overlapping red blood cells for sickle cell disease diagnosis," vol. 16, pp. 1669-1677, 2022.

[41] E. Mohamed, K. Sirlantzis, and G. J. D. Howells, "A review of visualisation-as-explanation techniques for convolutional neural networks and their evaluation," vol. 73, p. 102239, 2022.

[42] F. Hohman, H. Park, C. Robinson, D. H. P. J. I. t. o. v. Chau, and c. graphics, "S ummit: Scaling deep learning interpretability by visualizing activation and attribution summarizations," vol. 26, pp. 1096-1106, 2019.

[43] F. Pasa, V. Golkov, F. Pfeiffer, D. Cremers, and D. J. S. r. Pfeiffer, "Efficient deep network architectures for fast chest X-ray tuberculosis screening and visualization," vol. 9, p. 6268, 2019.

[44] K. Simon, M. Vicent, K. Addah, D. Bamutura, B. Atwiine, D. Nanjebe, and A. O. Mukama, "Comparison of Deep Learning Techniques in Detection of Sickle Cell," 2023.

[45] Q. H. Nguyen, B. P. Nguyen, S. D. Dao, B. Unnikrishnan, R. Dhingra, S. R. Ravichandran, S. Satpathy, P. N. Raja, and M. C. Chua, "Deep learning models for tuberculosis detection from chest X-ray images," in *2019 26th international conference on telecommunications (ICT)*, 2019, pp. 381-385.

[46] L. G. C. Evalgelista and E. B. Guedes, "Computer-aided tuberculosis detection from chest X-ray images with convolutional neural networks," in *Anais do XV Encontro Nacional de Inteligência Artificial e Computacional*, 2018, pp. 518-527.

[47] R. Hooda, S. Sofat, S. Kaur, A. Mittal, and F. Meriaudeau, "Deep-learning: A potential method for tuberculosis detection using chest radiography," in *2017 IEEE international conference on signal and image processing applications (ICSIPA)*, 2017, pp. 497-502.