Lung Tuberculosis Detection using Chest x-ray Images based on Deep Learning Approach

Addis Meshesha
Department of Electrical and Computer Engineering,
Wollo University,
Kombolcha, Ethiopia

Getasew Abeba
Department of Information Technology
Woldia, Ethiopia

Sefinew Getnet
Department of Electrical and Computer
Engineering, Woldia University
Woldia, Ethiopia

Nune Sreenivas
DEPARTMENT OF CSE, School of Electrical
Engineering and Computing (SoEEC)
ADAMA Science and Technology University,
Adama, Ethiopia

ABSTRACT

Due to the enormous expansion in human population across the globe in the modern period, automatic disease diagnosis has become essential for medical technology. To help radiologists and physicians diagnose and screen patients, as well as to improve diagnostic times and reduce mortality rates, a framework for automated image processing-based illness identification is essential. These days, lung tuberculosis poses a serious risk and is expanding around the world. Using automated detection, identification, and diagnosis systems can help to improve the pace at which diseases are diagnosed and prevent them from spreading over the world, which will help to alleviate this grave issue. The design and development of computer aided diagnosis method will facilitate lung tuberculosis diseases screening. In this work, we conducted an experiment based on adaptive histogram equalization and Gaussian pre-processing with thresholding, active contour model and morphological lung area segmentation approaches integrated with deep learning (Xception stacked with LSTM &ViT-LSTM) enables to feature extract and classify lung tuberculosis chest-xray accurately. The ViT-LSTM network architecture performs better with 93.4% accuracy in comparison with convolutional neural network-based model Xception stacked with LSTM recur-rent neural network with 88% accuracy for detecting lung tuberculosis diseases and also we performed a set of experiments, and the results indicate 41% improvement in average computational speed.

Keywords

X-ray, Deep learning, Xception ViT LSTM Thresholding, Morphological Active contour model

1. INTRODUCTION

The respiratory system is made up of all the parts that either directly or indirectly support breathing, which is a vital function for survival. The most vital organ in the human body, the lungs are in charge of giving blood its oxygen [1]. Millions of people die each year from the prevalent and extremely contagious lung disease known as lung tuberculosis (TB TBC). According to a 2018 World Health Organization (WHO) report, 1.4 million people worldwide—including 208,000 HIV-positive individuals—died from tuberculosis infections, which affected 10 million people worldwide. Early identification, categorization, and precise diagnosis are crucial for improving patient survival [2]. The most used radiographic procedure or diagnostic tool for tuberculosis in medical practice is the chest radiograph test [3][4][5], as compared to other screening mechanisms such as microbiological sputum smears, microscopy, skin testing, blood tests, and biosensor tests, among others. Traditionally, when a patient has a lot of medical images taken for them, the manually annotated method of lung TB chest radiograph testing is laborious, time-consuming, and highly subjective inaccuracy in medical imaging diagnosis. But because there are so many patients and there aren't enough qualified medical technologists or radiologists, there is a considerable chance of human error when analyzing chest x-ray for pulmonary tuberculosis in the many people whose diagnoses are still unreported. As a result, developing accurate CAD (Computer aided diagnosis) techniques needed to evaluating and analyzing for lung TB patient cases from chest x-ray images by reducing diagnosis or reading time. The four phases of a typical CAD approach are preprocessing, lung segmentation, Feature extraction and classification. In this method, ROI (region of interest) is image segmented to reduce search space and also significant features extracted manually annotation to outperform a feature vector based on pre-processed and lung segmented image [6][7].

However, to overcome this traditional manual annotation method and the existing proposed system we used innovate technical approaches to diagnosing case abnormality chest x-ray images from normal in medical application using CAD system. In the first phase of our proposed we used AHE, Gaussian filter and normalization pre-processing technique. Adaptive histogram equalization, Gaussian filter and normalization preprocessing technique (noise removal) tend to enhance or improve the quality of the image. In the lung region segmentation, we can reduce the search space by using thresholding, morphological and active contour model operator. The output result of lung segmentation integrated with feature extraction and classification by applying deep learning approach. Now a days, deep learning approach is fast growing field, has been performing exceptionally good in medical application and the approach uses complex layers to progressively extract high level feature and transformation from the raw input data image in order to obtain hierarchical representations learning [8][9].

In deep learning techniques, convolutional neural network (xception model) and vision transformer architecture are the state-of-the-art learning algorithm. Xception deep convolutional neural network architecture is mostly introduces a linear combination or stack of separable convolution layers or depth wise separable convolution layers with the adoption of residual connections [10]. CNNs are widely used in the machine learning Feld and are suitable for feature extraction in specific local regions. However, they are unable to capture the contextual relationship between image features in the global context. In contrast, the ViT applies an attention mechanism to understand the global relationships among features. These learning algorithms are very important model in our thesis to extract the feature of the whole input image (data) and used as over fitting reduction of the complex network and efficient, Captures information of shape, size, etc. It's also relatively simple, high accuracy rather than other deep learning. Recurrent neural networks (LSTM model) are used to obtain inferences about sequence to-sequence relationships and memorizes some past data. The feature of Xception network and vision transformers are extracted and taken as input for the LSTM recurrent neural network (used for classification purpose). LSTM also used to effectively address vanishing gradient problem by it can frees up the memory and important in the final classification purpose.

The system aims to assist medical technologist and improve and enhances the accuracy of clinical diagnosis. Structure of the lung is shown in Figure 1 [1].

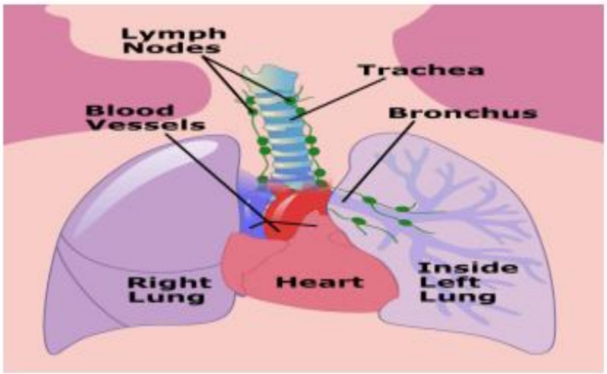


Figure 1. Structure of lung

1.1 Machine Learning

The scientific study of algorithms and statistical models that computer systems employ to successfully complete a job without the need for explicit instructions—instead relying on patterns and inference—is known as machine learning (ML). As a subset of artificial intelligence, it is recognized. To generate predictions or judgments without being expressly trained to do so, machine learning algorithms create a mathematical model using sample data, or "training data" [29]. The fields of artificial intelligence (AI) and machine learning (ML) have advanced quickly in recent years. Techniques of ML have played important role in medical field like medical image processing, computer-aided diagnosis, image interpretation, image fusion, image registration, image segmentation, image-guided therapy, image retrieval and analysis [30]. There are three types of machine learning algorithms [29][31]: supervised learning, unsupervised learning and semi supervised learning.

Supervised learning, the model is learned from the input and the expected output data. These are the most common way of learning. It uses labeled training data to learn the mapping function from the input variables to the output variables. Examples: Neural Networks (MLP), Logistic Regression, SVM, K-Nearest neighbors, Linear Regression, Decision Tree etc. Unsupervised learning, the model is learned only from the input data. This approach is particularly useful in practice since unlabeled data is abundant while labeled data is scarce and requires a lot of effort to collect. This approach gives input data to the algorithm and it learns and predict from experience, which is mostly through association, clustering and dimensionality reduction. It mostly learns the correlations among the input data to reconstruct it again. Examples: K-Means clustering, hierarchical clustering, mixture models, etc. Semi-Supervised learning, in this approach both kinds of data are used to train the model. The model is first pre-trained using unsupervised data and then improved with supervised data. When a neural network is to be used for classification; it first pre-trained layer by layer using unsupervised training algorithm. Then finally the network can be trained with a standard training algorithm, for classification or prediction. Deep learning (also known as deep structured learning or hierarchical learning) is part of a broader family of machine learning methods based on learning data representations, as opposed to task-specific algorithms. Learning can be supervised, semi-supervised or unsupervised [54]. In deep learning, a computer model learns to perform classification tasks directly from images, text, or sound. Deep learning models can achieve state-of-the-art accuracy, sometimes exceeding human-level performance. Deep learning Models are trained by using a large set of labeled data and neural network architectures that contain many layers. Thus, it plays a major role in computer vision and medical imaging. In fact, similar impact is happening in domains like text, voice, etc. Various types of deep learning algorithms are in use in researches like Convolutional Neural Networks (CNN), Deep Neural Networks (DNN), Deep Belief Network (DBN), Deep Auto-encoder (DA), Deep Boltzmann Machine (DBM), Deep Conventional Extreme Machine Learning (DC-ELM), Recurrent Neural Network (RNN) etc.[30].

1.2 Artificial Neural Networks

The human nervous system served as the conceptual and structural model for artificial neural networks. An artificial neural network (ANN) is made up of linked neurons that receive input, analyze it, and then send the results from the current layer to the one below. Every neuron in the network aggregates the input data, applies the activation function to the aggregated data, and then generates an output that might potentially propagate to the subsequent layer. ANNs are strong in the field of machine learning. However, numbers of neurons in deep neural network systems are still not comparable to the number of neurons in human. One of the most complex neural network architectures (i.e. GoogLeNet) has nearly 6.7 million parameters [32]. In 1958, Rosenblatt [33] generated one of the earliest types of artificial neural networks called as perceptron. It is one of the earliest neural networks based on human brain system. It consists of input layer that is directly connect to output layer and was good to classify linearly separable patterns. It is a simple mathematical model of a biological neuron, whose output can be given as:

$$F(x) = \begin{cases} 1 & \text{if, } w.x + b > 0 \\ 0 & \text{otherwise} \end{cases} \qquad (1)$$

Where F(x) is the output of the neuron, w is a vector of real-valued weights, w. x is the dot product of the weight vector w and the vector x, and b is the bias (i.e. a neuron added to each pre-output layer that stores the value of 1), bias units aren't connected to any previous layer and in this sense don't represent a true "activity". Perceptron is accepted as one of the first artificial neural networks to be produced.

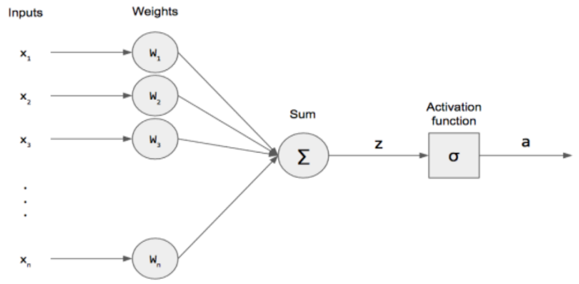


Figure 2. Perceptron algorithm

To solve complex patterns, neural networks with a layered architecture were introduced known as Deep Neural Networks (i.e. Input layer, output layer and one or more hidden layers) [30]. Multilayer perceptron (MLP) is one of feed forward ANN consists of at least three layers of nodes (Figure 3). The first layer is the input layer and the last layer is called the output layer. Middle layers are called the hidden layers. Due to its hidden layers, MLP can distinguish data that is not linearly separable. In a MLP system, the number of input and output nodes is determined according to the data. For example, in order to design a network architecture for handwritten digit recognition where numbers are stored in 28x28 size images, there will be 784 nodes (one input node for one pixel, 28x28 = 784) in the input layer and nodes (one node for the each number) in the output nodes. Figure 3 shows a simple 3- layer neural network architecture and a deep neural network (1-input, 3-hidden, 1-output).

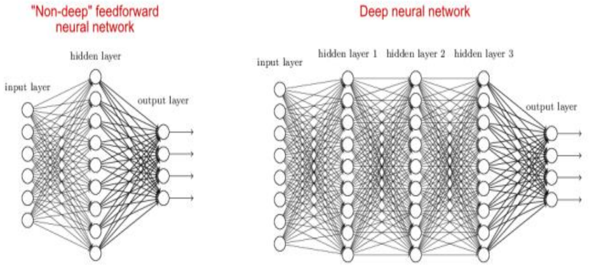


Figure 3. a simple and deep neural networks

Determining the number of hidden layers and the number of nodes in the hidden layer is a design issue during the training process. Too many nodes will make training longer and the network may lose its generalization ability. On the other hand, with too few nodes, the network uses too little information and may not solve the complex models. Activation functions are non-linear complex functional mappings between the inputs and response variable. They introduce non-linear properties to our Network. Their main purpose is to convert an input signal of a node in ANN to an output signal. That output signal now is used as an input in the next layer in the stack [34]. In artificial neural networks, the activation function of a node defines the output of that node given an input or set of inputs [29]. The three commonly used activation functions in neural network are sigmoid, rectified linear unit, Tanh etc. Sigmoid (Logistic) is between 0 and 1. It is easy to understand and apply but it has major reasons which have made it fall out of popularity: vanishing gradient problem, 0 < output < 1 makes optimization harder, sigmoids saturate and kill gradients, have slow convergence. Tanh (hyperbolic tangent function), its range is between -1 and 1 i.e., -1 < output < 1. Hence optimization is easier in this method. It is always preferred over sigmoid function. But still, it suffers from vanishing gradient problem. ReLU, it has become very popular in the past couple of years. It was recently proved that it had 6 times improvement in convergence from tanh function. Nowadays, almost all deep learning models use ReLU because it avoids and rectifies vanishing gradient problem [34].

1.3 Convolutional Neural Network (CNN)

Convolutional neural network (ConvNets or CNNs) is one of the main categories of deep learning algorithm to do images recognition, images classification, objects detections, face recognition etc. CNN image classification takes an input image, process it and classify it under certain categories (E.g., Normal or abnoral). CNNs take an input image as array of pixels and it depends on the image resolution. In CNN, raw data is represented as tensor. Tensor concept can be generalized as higher order matrices. Based on the image resolution, it will see H × W × D (H = Height, W = Width, D = Dimension). E.g., an image of 6 × 6 × 3 array of matrix of RGB (3 refers to RGB values).

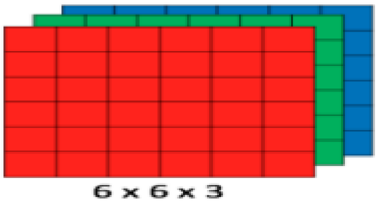


Figure 4. array of RGB Matrix

According to technical specifications, each input picture for deep learning CNN models is sent through a sequence of convolution layers with filters (Kernels), a pooling layer, fully connected layers (FC), and the softmax function in order to identify an item with probabilistic values between 0 and 1. Using filters, CNNs identify certain patterns; pooling layers then assist the model in ignoring duplicated input.

Figure 5 shows a complete flow of CNN to process an input image and classifies the objects based on values.

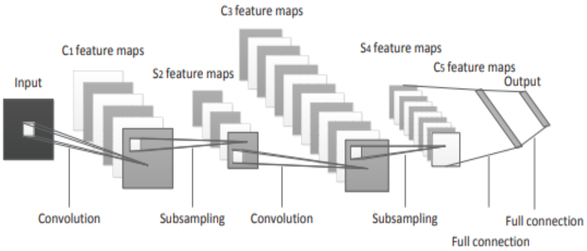


Figure 5: convolutional neural network architecture

1.4 Convolution Layer

CNNs use convolution as its initial layer to extract information from input images. Convolution uses tiny squares of input data to learn visual attributes, preserving the link between pixels. A filter or kernel and an image matrix are the two inputs used in this mathematical technique. The kernel is chosen as a three-dimensional structure if the input is a three-channel RGB picture. The kernel won't be able to extract enough features if it is too tiny, like 2 × 2. For instance, little kernels are unable to identify large, intricate patterns. On the other hand, bigger kernel size increases computation complexity. Generally, most of the time small kernel sizes such as 3 × 3 or 5 × 5 are used in the CNN training. Strides are the number of pixels shifts over the input matrix. When the stride is 1 then we move the filters to 1 pixel at a time, when the stride is 2 then we move the filters to 2 pixels at a time through the input matrix and so on. Consider a 5 × 5 input matrix whose image pixel values are 0, 1 and filter matrix 3 × 3 as shown in figure 6.

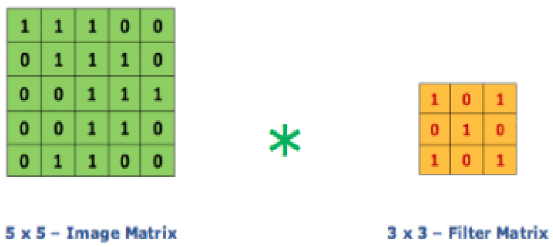


Figure 6. Image matrix and filter matrix

Then the convolution operation of 5 × 5 image matrix multiplies with 3 × 3 filter matrix with a stride 1 × 1 generates a feature map output as shown in figure 7.

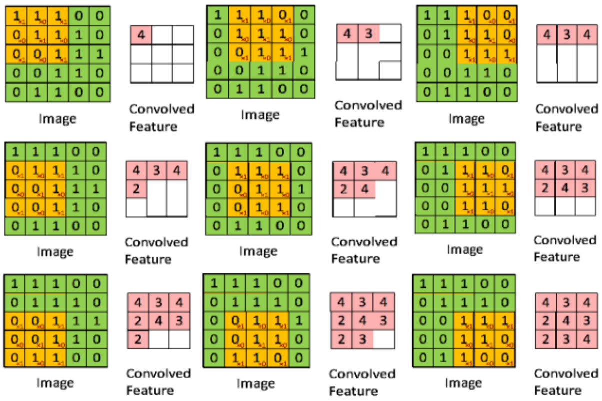


Figure 7: an output feature maps for 5 × 5 input convolved with a 3 × 3 kernel [35]

The number of kernels dictates how many distinct characteristics should be the emphasis of the design, making it a crucial decision. The network could overlook some patterns in the data if there aren't enough kernels. Conversely, there shouldn't be too many kernels for there to be redundant filters. For instance, if the kernel size is 3 × 3, employing 64 kernels will result in duplicate filters as 3 × 3 sizes are unable to produce 64 valid, distinct filters. Too many filters might cause memory problems in addition to duplicate filters as each convolved picture takes up memory on the computer. A feature map is produced by each kernel as it is slid over the input picture. These output images are concatenated after all kernels have produced their outputs. If the input image is 2-dimensional, its outputs will be 3-dimensional tensor. If the input image is 3-dimensional volumetric data, its outputs will be 4-dimensional tensor. This extra dimension comes from using many filters. Depth of the tensor and depth of the kernel must be same. For example, consider an input tensor of a size 30 × 30, and a kernel of a size 3 × 3. The resulting size of the convolution is 28 × 28. We may have more than one kernel, which are applied on the input tensor. As a result, size of the output tensor becomes K×28×28. For the next convolution operation, kernel dimension becomes K × N × M. In CNN model representations; this size is generally represented as K@N×M.

1.5 Pooling Layer

When the images are too big, the pooling layer reduces the number of parameters. Sub-sampling, also known as spatial pooling lowers each map's dimensionality while keeping the crucial details. There are three forms of spatial pooling: sum-pooling, max-pooling, and average-pooling. The biggest element from the corrected feature map is taken via max pooling. The average of the corrected feature map pieces is used in average pooling. Sum pooling involves adding each feature map piece. The window size and stride value are two crucial hyper-parameters for the pooling procedure.

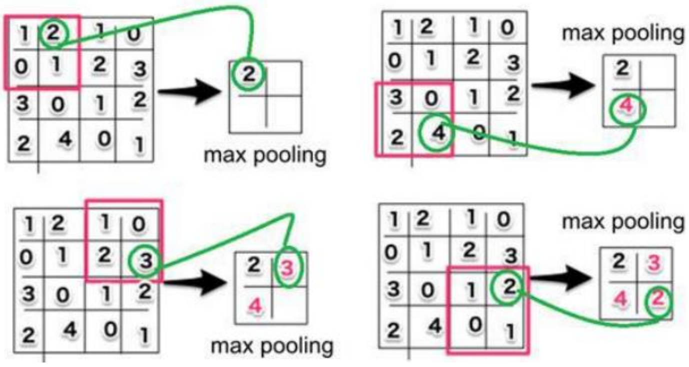


Figure 8. Example of Max-pooling

1.6 Fully Connected Layer

The layer we call as FC layer, we flattened our matrix into vector and feed it into a fully connected layer like neural network. Convolution and pooling layer generate rectangular shaped outputs. These outputs are converted to vector format so that they can be multiplied by the weight matrix. For example, if there are 64 feature map layers each of which has 5×5×3 voxels, in the fully connected layer these volumes are converted to a 4800×1 vector (5×5×3×64 = 4800). The layer before the fully connected layer represents high-level features. With the help of a fully connected layer, these high-level features can be multiplied by the weights of the hidden layers the remaining part of the system works like MLP do.

1.7 Recurrent neural network

The most common kind of artificial neural network is the recurrent neural network, which forms directed networks along temporal sequences through connections between computational or hierarchical nodes. For processing temporal information, the recurrent neural network model is a great fit [36]. The RNN's units are directly cyclically connected, allowing it to store its internal hidden state and aid in the modeling of dynamic temporal activity. An RNN's hidden states serve as a network's memory by storing data from earlier states. Three different kinds of neuron layers make up the basic architecture: input, hidden and output layer nodes. Recurrent neural networks allow the network to flow only in the direction of feedback, from input layers to output layers. One important aspect of RNNs is how the hidden layers are connected. The hidden layer is sent to the output layer, while the input layer nodes are linked to the other layer of hidden nodes. The node output data is sent back to the hidden layer node, and it may even contain information about nodes that are next to one another in the hidden layer. When managing sequential output prediction, where the current output depends on both the current input and the prior outputs, cyclical connections play a crucial role. A recurrent neural network makes use of the LSTM and GRU (gated Recurrent Unit) models. To manage the memorization process, both models are benefit to a gating mechanism. A GRU and an LSTM differ in that a GRU has GRU gates (update gates and reset), whereas an LSTM has three LSTM type gates (output, forget and input gates). Larger sequences and a larger dataset make LSTM more effective. However, GRU du to have a lower memory limit.

1.8 Motivation

Lung TB is the most frequent and severe infectious illness that affects the lungs of humans, and it is also the top cause of mortality worldwide among many other dangerous bacterial infections. Essentially, over 95% of lung tuberculosis cases worldwide spread quickly in many underdeveloped and rich nations [37]. Millions of people die each year from tuberculosis, a widespread and highly infectious lung illness, as a result of inadequate treatment and incorrect diagnosis. There is necessary to innovate with technical approaches to diagnosing case abnormality chest radiograph images from normal in medical application using CAD system. The advancement technique of deep learning with CAD has significant role to make examining the process better efficient and effective method. The radiologist benefit from the deep learning method with CAD for screening or detecting lung tuberculosis based on chest radiograph (CXR) images. A deep learning solution to interpret chest radiograph (CXR) for presence of lung tuberculosis in a cost-effective manner would expand to reach or address of early detection and highly proper treatment of lung tuberculosis diseases in low- and middle-income countries.

2. RELATED WORKS

This section contains a number of earlier studies that used various methodologies to investigate automated lung TB detection systems. A variety of methods have been developed by various researchers to use computer-aided diagnosis to identify pulmonary TB. Results from these thesis studies are also given and analyzed, along with the field of attention and the gap.

TB Detection in Chest Radiograph using CNN was proposed by Rahul Hooda and Ajay Mittal [11]. In this work, they have used two convolutional network architecture such as: VGGNet and AlexNet network architecture. They evaluate and verified their proposed method on two publicly available data acquisition and combined to form the final dataset. This data set includes Shenzhen and Montgomery chest x-ray (chest radiograph) set. Their proposed method was based on 3-dimensional analysis of the chest radiological image data. The MC Dataset comprises 138 CXR images consisting of 80 normal images and 58 abnormal (TB) images while Shenzhen dataset has 662 images consisting of 326 normal and 336 abnormal cases. Therefore, the data set comprises a total number of 800 images are collected. Among total number of 800 images, 560(70%) images were used for training and 240 (30%) for the remaining part of 240 images used for testing purpose. The proposed systems have achieved an accuracy of VGGNet and Alex Net is 81.6%, 80.4% respectively. From accuracy results, VGGNet is better accuracies value than Alex Net architecture. CNNs are based on feed-forward neural network architectures and automatic selection of features. The performance of extracted features in CNN network architecture depends on the depth of the architecture. The gap of this work is reduced computing system performance in terms of accuracy measures. The system performance of the proposed approach can be further enhanced by increasing the chestx-ray dataset.

Rohilla and Ajay Mittal [12] also further improved tuberculosis classification using CNN based on chest radiograph images. For performing the proposed system, images from different data set namely Montgomery, JSRT, Shenzhen, Belarus dataset have been used. In this work, a method for TB classification is proposed which uses deep learning architecture, ResNet. The ResNet architecture has been customized to perform classification of chest radiograph images in to two classes, that is TB Positive and TB negative image. In this architecture, best features are automatically extracted based on the training images and their outputs. The number of training image is increased by using data augmentation techniques. The experiment result is obtained on the data set of 1133 images, among 1133 images ,499 are TB positive and TB negative images. This model achieved on test accuracy of 84.12%.

Jaeger et al. [13] presented a Tuberculosis detection method in which intensity mask, lung model mask, and Log Gabor mask are used for lung segmentation. In this work, different shape and texture descriptors are used to find the pathological patterns in chest-radiograph mages. For each descriptor, histogram bins are used to represent its distribution and value of each histogram bin for every descriptor is considered as a feature. The dataset comprises 138 chest-radiograph images collected from publicly available USNLM database of Montgomery County. Linear support vector machine (SVM) is used as a classifier to classify the chest radiograph images into normal and TB positive classes. According to experiment results, the overall accuracy of detecting TB combining with all masks is 83.12%. Another work by these researchers in [14], also presented a similar automated method in which two separate feature sets namely object detection-based features and CBIR (Content-based Image Retrieval) based features are used, after segmenting lung boundary using graph cut segmentation method. Finally, SVM is used as the classifier to classify chest radiograph as normal and TB infected cases. Results are obtained using three datasets, of which two are used for training and one for testing the method. The performance of object detection and CBIR feature vectors is found to be 0.87 and 0.90 respectively in terms of AUC.

A Potential Method for automatic lung Tuberculosis Detection using Chest Radiograph was conducted by Rahul Hooda, Sanjeev Sofat and Simranpreet Kaur [15]. The proposed framework has been validated by the dataset of USNLM challenge using the service of health department at medical college, Shenzhen China and Montgomery County chest radiograph. They develop, simple deep learning which is convolutional neural network architecture, that has number of layers in between LeNet and Alex Net network architecture. The performance of the extracted feature in deep convolutional network depends on the depth of the network architecture. The experimental results are evaluated by training and testing the proposed architecture based on Montgomery County and Shenzhen data set. On the total of 800 chest radiographs images, which are used for the training purpose and 200 chest radiograph images are used for validation case. The last detection results are gained by fusing the probability prediction output of the network architecture. The average detection validation accuracy of the proposed system is 82.09%. Future work includes extending the developed method to classify chest radiograph images in to different Tuberculosis manifestations for which a larger dataset is required.

Betsy Antony and Nizar Banu P K [15], proposed Lung Tuberculosis detection based on chest radiograph images (xray images). In this work, the method consists four steps: filtering, segmentation, and extraction feature and classification stages. To remove unwanted noise from an image, median filtering technique is done at the first stage. For the next stage they combined two segmentation methods like watershed model and gray level thresholding model, and a fused image is generated which yields an accurate result. Features like area, major and minor axis, eccentricity feature, mean, standard deviation, skewness, and kurtosis are extracted from ROI of fused image. Finally, they have used the approach of three classifiers algorithm: KNN, SMO and Simple linear regression classifiers. The dataset comprises of a total number of 662 images available where 326 images are TB negative and 336 images are TB positive, collected from publicly available National Library of Medical medicine (NLM) data acquisition. The

Proposed systems have achieved with 80%, 75% and 79% accuracy by KNN, SMO and Simple linear regression classifiers respectively. From the accuracy results obtained we observed that KNN classifier performs maximum accuracy compared to the two classifiers algorithm. The research gap is mainly caused by two major impediments: Necessary to design and implement with huge amount of data set, reduced accuracy measures of computing system for accurate medical diagnosis by analyzing the chest radiograph images.

The work done in [16] proposed Framework of Predicting Drug Resistance of Lung Tuberculosis by Utilizing Radiological Images. The authors proposed a general framework to do the predicting of lung drug-resistant tuberculosis in radiological images. To solve the sample of predicting problem, with convolutional neural networks, they have used VGG16 network architecture as the basic model. They introduce VGG16 network architecture to examine drug resistance of lung tuberculosis and test proposed method based on imageCLEF2017 image acquisition. The total numbers of dataset are 230, including 134 dataset of tuberculosis drug sensitivity, and 96 datasets belonging to multi drug resistance by Utilizing Radiological Images. The authors proposed a general framework to do the predicting of lung drug-resistant tuberculosis in radiological images. To solve the sample of predicting problem, with convolutional neural networks, they have used VGG16 network architecture as the basic model. They introduce VGG16 network architecture to examine drug resistance of lung tuberculosis and test proposed method based on imageCLEF2017 image acquisition. The total numbers of dataset are 230, including 134 dataset of tuberculosis drug sensitivity, and 96 datasets belonging to multi drug resistance. Their proposed method was based on 2D and 3D analysis of the radiological image data. And they have tested the proposed methods on ImageCLEF2017 tuberculosis dataset, and obtained the accuracy of 64%. The performance of the model can be further improved by more testing radiological image, with huge number of data more features will be learned by the proposed model.

M.K Osman and M.Y Mashor [17] proposed Compact single hidden layer feed forward network for Mycobacterium tuberculosis detection using tissue slide images. The dataset comprises total numbers of available dataset are 1603, including 620 datasets of TB; Non-TB is 498, and 485 datasets belonging to overlapped tuberculosis case. From these datasets, and their proposed model used 603 samples for testing and 1000 samples for training purpose. The experimental result is indicating that their proposed model achieved accuracy of 75.46%.

According to [18], the authors attempted to develop Automatic classification of pulmonary (lung) tuberculosis and sarcoidosis using Random Forest algorithms. In this work, the researchers proposed based on the performance analysis of super vector model, Random Forest, Logistic regression, Naïve Bayes techniques for automated pulmonary tuberculosis and sarcoidosis classification system. The data set used in this paper /research work is selected from HIS database in the counts of pulmonary tuberculosis and sarcoidosis is 485 and 1990 images respectively. The system capable of detecting TB and sarcoidosis with 0.82%, 0.853%, 0.84% and 0.85% accuracy by super vector model, Random Forest, Logistic regression, Naïve Bayes classifier respectively. From this paper the performance experimental result work, we can observe that the Random Forest approach gives better accuracies results compared to other three classifiers models. In [19], the researchers design and developed classification of Lung tuberculosis with in SURF spatial pyramid features. In this paper, the authors have presented the use of local features of SURF (Speed-Up Robust Features) extracted from the segmented lung images using a grid window of various spacing, hence controlling the consistency of the SURF features. The data set used for testing and training was collected from U.S. National library of Medicine (USNLM) using the services of the health department at Montgomery County (MC), USA. It contains of 138 chest-radiograph images collected under MC's tuberculosis screening program. The dataset comprises 80 chest-radiograph images that are normal and remaining 58 chest-radiographs have TB manifestations. The paper elaborates the complete implementation and designing of Computer assisted (aided) diagnosis (CAD) system to facilitate tuberculosis screening and presents the performance analysis based on the available USNLM database. They have used super vector machine (SVM) classifier for the CAD system. This model was can achieved performance of an Area under /and the Receiver Operating Characteristic curve (ROC)metric of AUC 89%. The limitation of this paper/research work is that, small numbers of data images and used only for training and testing case.

Mostofa Ahsan and Rahul Gomes [20], presented Application of convolutional neural network based on transfer learning for lung tuberculosis detection. In this work, they have presented a Convolutional network architecture approach that can uses VGG16 Net for classifying CXR (chest radiograph) images to identify patients suffering from tuberculosis diseases. The average classification accuracy from both Shenzhen and Montgomery were approximately 80.4%. The authors also achieved the classification accuracy by using only Shenzhen data set (82.5%) and Montgomery having 78.3%. For this work, they have used a total 276 CXRs data images from Montgomery and 1324 from publically available Shenzhen datasets. The dataset was split/depart using 75 to 25 ratios where 75% used for training case and 25% for testing. Their proposed method was used 3-dimensional analysis of the chest-radiological (chest x-ray) image data. In Future work, would extend in the work include running the model on a system with a higher configuration so that the augmentation could be done on all images before training the model using VGG16 Net. They achieved good performance accuracy measures by training case of batch size of 16, there is a scope for further increase if the image augmentation would be applied for all CXRs (chest radiographs) images. Therefore, from this section that we reviewed the literature papers, we examine that a new significant approach is expected to obtain an improved performance in lung tuberculosis detection. We design and implement a system to improve lung tuberculosis detection and classification performance in terms accuracy by using deep learning approaches. Besides this, a new approach of preprocessing and lung segmentation were provided.

3. MATERIALS AND METHODS

In our proposed framework, we have used two publicly available datasets which are Montgomery and Shenzhen data set [21][38]. The Montgomery County Dataset comprises 138 CXR images consisting of 80 normal images and 58 abnormal (TB) images, and also Shenzhen dataset has 662 images consisting of 326 normal and 336 abnormal cases and JSRT(Japanese Radiological Technology Society) dataset. JSRT comprises There are 154 X-ray images have lung nodules and 93 normal Xray images in the dataset.

Therefore, the data set comprises a total number of 1047 images are collected. In this work, we proposed the design and testing of automatic tuberculosis detection with computer assisted diagnosis.

The proposed framework has the four main stages: image pre-processing, Lung region segmentation, feature extraction and classification.

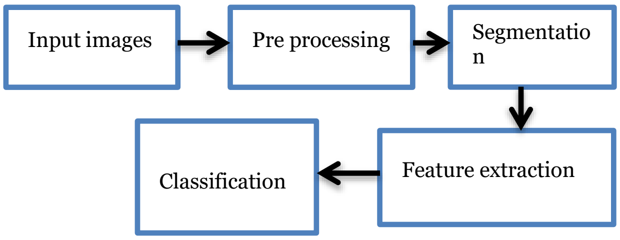

Input images
Pre processing
Segmentatio n
Feature extraction
Classification

Figure 9. Workflow of the proposed system

In the pre-processing stage, the pre-processing stage performs image resizing, AHE, Gaussian filter (noise removal) and normalization technique tends to enhance or improve the quality of the image. The second stage of our proposed the framework, Lung region segmentation; in the lung region segmentation we can reduce the search space by using thresholding, morphological and active contour model operators. The output obtained from those taken as an input enhanced prepressed image and were performed by thresholding, morphological and Active counter model which can helps to focus on the lung region of the obtained results. Output map from these lung segmentation methods is used to get a Region of Interest reduction. The output result of lung segmentation integrated with feature extraction and classification by applying deep learning approach. Feature extraction system was done by deep learning (deep Convolutional neural net-work) which is Xception model architecture, and Vision transformer deep learning algorithm. In the fourth stage of classification process, long short-term memory (LSTM) outputs the decision that whether image is TB positive (abnormal) or TB negative (normal).

3.1 . Preprocessing stage

In this section, the pre-processing stage performs by using image resizing, AHE, Gaussian filter and normalization technique. The resizing image can rescale lung TB image dimensions to 512x512 pixels format in size. It is a very important part in image processing technique, to enlarge and decrease the given medical image (dataset) size in pixel-by-pixel format in the given chest x-ray. The next stage is Adaptive histogram equalization and it's used to improve or enhance image contrast. Image contrast enhancement is significant role in medical image processing applications. The enhancement method is due to the fact that visual screening and examination of medical digital image is important in the diagnosis of the diseases. AHE is clearly different from the ordinary histogram equalization method because histogram equalization implies only single or one histogram form. However, AHE method generates various histograms. It helps to redistributes the value of intensity in data image. Adaptive histogram equalization can improves/enhances on this by transformation of each and every pixel with a function transformation based on derived form of neighborhood region. The standard histogram approach, which involves enclosing each pixel value in a square, is used to change each pixel.

In, AHE, the intensity of value each pixel of the transformed function is based on the ordinary histogram of square surrounding the pixel value in data image. Then, the data image was noise filtered using Gaussian. Gaussian filter is done after Adaptive histogram equalization [22]. Gaussian filter works the value of each pixel weighted average with other pixel in neighboring pixel. Finally, normalization is an important stage to preprocess each input pixels using different pre-processing method. Based on our data we have use the value between zero and one based on standard deviation bounds. By Partitioning each input parameter or pixel by its standard deviation distribution, the pixel values from pixel between values 0 and 1 [23]. Therefore, we applying normalization based on our medical image processing to make optimization basically simple and used to diminish the pixel value range from 0 –255 to 0 – 1.

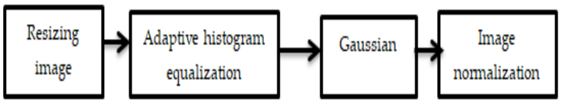
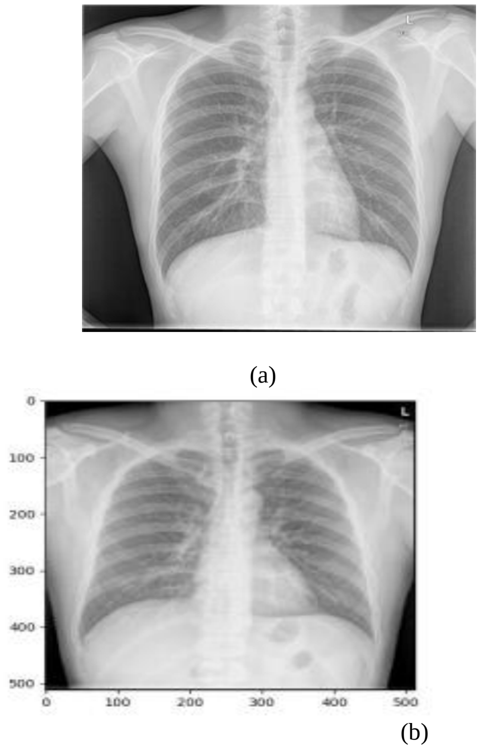

Resizing image
Adaptive histogram equalization
Gaussian
Image normalization

Figure 10. Work flow of preprocessing stage.



(a)

(b)

Figure11. (a) Original lung tuberculosis image (b) Resized image

3.2. Lung region segmentation

Lung region segmentation is a vital role and an essential process in image medical enhancement (preprocessing) procedure. Our aim is segmentation of lung image as region of interest (search space) extraction to simplify or modifying the level representation of a medical image into something that is more meaningful and easily to analyze properly. Abnormality and normality of medical lung images will be indicated according to lung image segmentation of accurate region of interest extraction. Different methods for lung segmentation are presented which includes pixel classification, rule-based methods, the Log Gabor mask, active shapes, the intensity mask and the lung model mask thresholding, morphological, active contour etc.

However, as compared to other technical methods, the segmented image obtained by thresholding and morphological operators has the advantage of quick processing, less storage space, and simplicity of manipulation. And also, active contour model better to segment for lung area and clear the border of the image accurately. In the following figure illustrates the work flow of our proposed lung image segmentation algorithms.

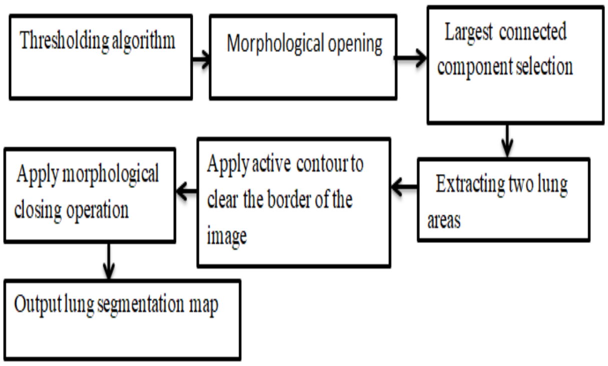

Thresholding algorithm
Morphological opening
Largest connected component selection
Extracting two lung areas
Apply active contour to clear the border of the image
Apply morphological closing operation
Output lung segmentation map

Figure 12. Lung region segmentation image procedure.

In the lung region segmentation image procedure, the first step is Image threshold algorithm threshold algorithm is a simple, computationally effective and efficient lung area segmentation of partitioning or splitting data image into a foreground region of pixels and background. This type image technique analysis is can isolate the targeted object by converting image gray scale into image binary. In the threshold operation of the algorithm, the equation one implies the value of x and y as an input which is basically the pixel values of images with in respect to x and y-axis is one should be the threshold value when we compare the value of inputs. When the input value pixel is greater than the value of threshold then it gives output value is set to one and it reveals color of white in gray medical images. The procedures that follow the image thresholding algorithm are.

$$y(x,y) = \begin{cases} 1 & s(x,y) < R \\ 0 & otherwise \end{cases} \qquad (2)$$

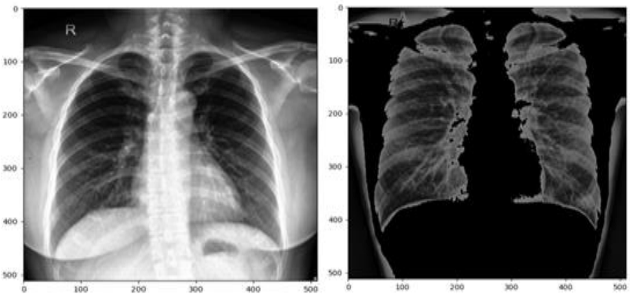
S (x, y) is an input image; y (x, y) is an output image and R is threshold value [24]. To find the thresholding value x, The threshold value x is for each pixel (x, y) is calculated by [25]:

$$R(x,y) = \frac{Max+Min}{2} \qquad (3)$$

Where min and max are the minimum and maximum gray level value. The thresh-old value in the lung region segmentation approach used to normalize the value ranges from -1024 to around 400 [16]. Anything that comes above the value 400, we cannot take into regard as those are the bones with in different radio density. The gray level value is calculated as [24].

Gray level value = R + 1024 (4)

The gray level value is: min = 0 and max = 1424. Therefore, the threshold value R (x, y) is 712. After thresholding algorithm, morphological opening is clearly pointed out. however, Morphology operational algorithm can be implying as a collecting of medical image processing technical approaches that makes the processes images by taking size and shapes into morphology of an image. The approach morphological operations considered by using the element structure on the top of an input medical image to create an image of similar size. In the operator, by considering the information, compare the value of pixels in the given information image with it's the region of neighbours into regard to approximate each and every pixel values in the return image. The morphological operations to give an opening (space) between regions that are connected through very thin holes, almost without affecting the original shape of the larger regions. Then, extracting the lung of the two regions and largest lung region of the component connected selection and also applying active contour to clear the border of the image. Finally, the morphological closing operation smoothes region of contours it in general fusing narrow breaks, fills small holes in the region, and fills gaps in the contour. As a result, it fills the small holes and gaps and in the section of contour objects boundaries.



(a) (b)

Figure 13. (a) Preprocessed image (b). Output lung segmentation map

4. FEATURE EXTRACTION

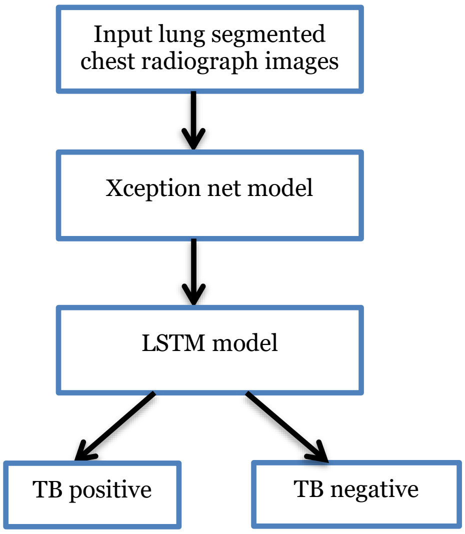
"Extreme Inception" is the acronym for the net architecture known as Xception. The network's foundation for feature extraction is made up of 36 convolutional layers of the Xception architecture. With the use of residual connections, it presents a linear stack, combination, or depth-wise separable convolution layer [10]. The most crucial layers of the Xception model architecture are the depth-wise separable convolutions. These have the ability to drastically cut down on computation and simulate network parameters. The network design factorizes into three different forms: 1x1 point convolution, 3x3 depth-wise convolution, and ordinary 3x3 convolution. The depth wise separable convolution (3x3) divides the calculation into depth wise convolution, which may apply one convolutional filter component per input channel, and pointwise convolution. The conventional 3x 3 convolutions out complete the computation in a single step for both spatial and channel wise aspects.

According to [26], the vision transformer (VIT) has achieved flawless performance on a variety of computer vision tasks. In order to achieve adequate image classification performance, the Vision Transformer (VIT) [27] splits images into patches and then utilizes a transformer to pattern the similarity among these patches as sequences. The organization of VIT may be summed up as follows: 1) Create patches out of the provided image. 2) Use flattened patches to create patch embeddings, which are lower-dimensional linear embeddings. 3) Include a positional embedding and a class token. 4) Apply the patch sequence to the transformer layer, and then obtain the label by using a class token. 5) Transfer the class token values to the Multi-Layer Perception (MLP) in order to obtain the output prediction. In order to get a lengthy feature vector representation of every patch, the patches are also fed into the linear projection layer. Furthermore, the series of embedded patches now includes position embedding. The transformer cannot hold the information if positional encoding is not applied. Finally, patch embedding's containing a positional encoding and a class token are fed into the transformer layer to produce the acquired representations of the class token. Therefore, the most crucial component of the VIT structure is the transformer encoder, which houses the Multi-Head Self-Attention (MHSA) block and the MLP block. The input layer of the encoded layer was combined with positional and patch embedding's. The inputs are normalized in the VIT architecture by the normalization layer prior to being supplied into the Multi-Headed Attention (MHA) block.

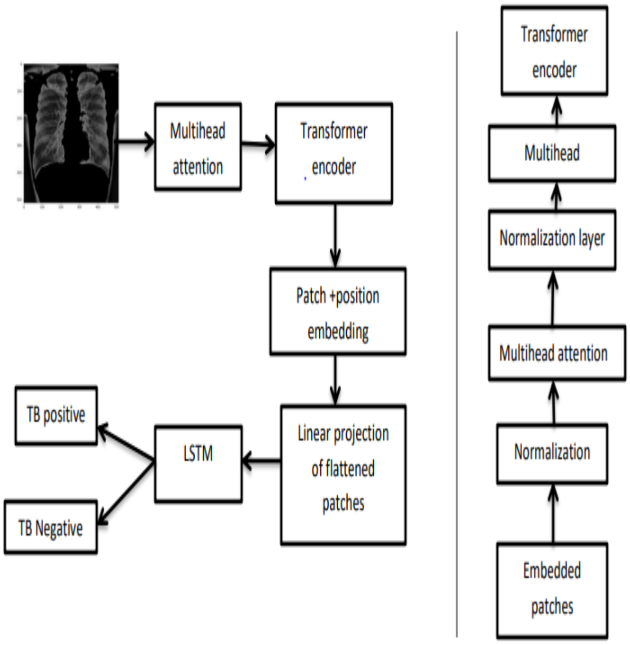
5. CLASSIFICATION STAGES

The feature vector from the last layer of the Xception and ViT (vision transformers) network are extracted and used as input vector for the LSTM network (used for classification stage). LSTM network is an improvement and cyclic connection of recurrent neural network with three type gates such as an input gate, output gate and forget gate. It learns the long terms dependencies in temporal direction with the aforementioned gates. LSTM is easier to optimize and address vanishing gradient problems and these gates enable the input features vector to propagate through the hidden layer node without effecting the output. LSTM also used to effectively address vanishing gradient problem by it frees up the memory locations and helpful in the final classification stages.

The following work flow shows proposed model architecture for network lung tuberculosis detection.



(a) Xception –LSTM architecture



(b)

Figure 14: (a) Xception –LSTM architecture (b) ViT – LSTM architecture

6. EXPERIMENTRESULTANDDISUSSION

In this section of the work, we will present and elaborates detailed description of experimental result analysis and implementing or designing of automatic lung tuberculosis detection with CAD using deep learning approaches via tools named as chest radiograph. The experimental data set splitting and discerption, experimental result evaluation criteria and experimental result analysis are discussed. We will also discuss and compared our experimental result works with the results of other previously thesis works based on the evaluation performance metrics. We will use four performance evaluation metrics like accuracy, Recall, f1-score and precision with confusion matrix on our approach to clearly point out effectiveness of proposed the models.

6.1 Data set splitting and description

In our experiment work, we have used three publicly available datasets which are Montgomery and Shenzhen data set. The Montgomery County Dataset comprises 138 CXR images consisting of 80 normal images and 58 abnormal (TB) images and also Shenzhen dataset has 662 images consisting of 326 normal , 336 abnormal cases and JSRT(Japanese Radiological Technology Society) dataset. JSRT comprises There are 154 X-ray images have lung nodules and 93 normal Xray images in the dataset.Therefore, the data set comprises a total number of 1047 images are collected. Among 1047 images, 790(80%) images were used for training and the other part (20%) for the remaining 257 images used for testing purpose.

6.2 Hyper parameters configurations

To develop and implement our proposed frame work, some model hyper parameter is evaluated or configured to achieve good experimental result. By using some of the model hyper parameter and their values basically based on our experimental result step, and is a very significant role in order to get an excellent experimental result work. A value of some of the model parameters will be able to evaluating by reviewing different research works, but for some other contributes it is significance role to undertake some model preliminary experiments results. The ultimate goal of testing different hyper parameters is to decrease the percentage of loss function whereas improving the performance values for our evaluation metrics. Some model hyper parameters are outlined below:

Activation function: activation function is the most important part in our designing implementation models which decides whether or not a neuron will be activated or not and transformed to the other layer of the network. This function simply meaning that it will able to decide whether the neuron's input to the other network is relevant or not in the process for classification. For this case, it is also the neurons /artificial neurons transformation which can converge the layer of network. The function used in normalizing the value of the output result between 0 to 1 or/and -1 to 1 (between-1 and 1). It plays important in the process of the network backpropagation. During back propagation, loss function gets updated, and activation function important the gradient decent curves to achieve to local minima.

Epochs: a group of samples which are passed through the training data set is epoch. Increase the number of epochs until the testing accuracy begins decreasing even when training case accuracy is increasing (overfitting). To compute the weight update for each input sample, but store these values during one pass through the training set which is named as an epoch. At the end of the epoch, all the contributions are added, and only the weights will be updated with the composite value. This method adapts the weights with a cumulative weight update, so it will follow the gradient more closely. Training case basically involves feeding training samples as input vectors through a neural network.

Optimizer: for updating weight model parameter we can use optimizer to reduce loss function. The optimizer is responsible of reduction of the objective function neural network. The choice of a best optimizer is very significant. A wide range of optimizer options are available to reduce loss function. Some optimizers such as Nestov, Momentum optimizer, Adagrad, RMSProp, and the list goes on. But the best is the Adam optimizer which stands for Adaptive Moment 60 Estimation [37]. Adam is a combination of sparse descent gradients and RMSprop. Therefore, our network was trained case with Adam optimizer parameter.

Learning rate: in our test model, learning rate is an important part of parameter for training process. During the training case, a hyper parameter that effective and efficiently controls the step size and makes the training process faster. However, selecting the value of the learning rate hyper parameter is sensitive. If the selected learning rate is too much large, then the local minimum may be overstepped constantly, resulting in oscillations and very slow convergence to the lower error rate case. If the selected learning rate value is too very low, the amount of iterations required may be too high amount update steps, resulting in poor performance. We use in our work; the model parameter set to learning rate value is 0.001.

Dropout rate: to reduce over fitting from the training data set based on our proposed models, we employed a recently-developed, very effective and efficient dropout regularization method. Dropout is referred to as an alternative regularization technique by decreasing the impact of any particular node on the output. And we integrated the parameter optimizer and learning rate value with regularization (dropout to 0.5). Loss function: The aim of training and testing different model parameter is used to decrease the percentage of loss function whereas enhancing the performance based our performance evaluation metrics. Now, the loss function used our work while training is binary Cross-Entropy Loss Function. The reason behind using a binary Cross-Entropy Loss Function is that we have binary output classes of the medical input images.

6.3 Experiment result evaluation

A confusion matrix is a table that is usually used to elaborate classification performance. In our implementation, Classification performance is evaluated or conducted based on precision, Recall, accuracy and F1-score. In our implementation, Classification performance is evaluated or conducted based on precision, Recall, accuracy and F1-score. We can compute the performance evaluation metrics in terms of the combining all the conditions of amount of false positives, number of true positives, and number of false and true negatives. True positive means the total amount of abnormal cases correctly classified, true negative signifies the total amount of normal cases perfectly classified, false positive describes the total amount of abnormal case wrongly detected/classified when they are clearly normal cases and false negative signifies the number of wrongly classified normal cases when they are clearly abnormal cases. The following is a definition of these parameters.

$$\text{Precision} = \left(\frac{TP}{TP+FP}\right) \qquad (5)$$

$$\text{Recall} = \left(\frac{TP}{TP+FN}\right) \qquad (6)$$

$$\text{F1\_score} = 2x\left(\frac{precisionxRecall}{precision+Recall}\right) \qquad (7)$$

$$\text{Accuracy} = \left(\frac{TP+TN}{TP+TN+FN+FP}\right) \qquad (8)$$

The obtained experimental results are explained and presented in the form of experiments so that the best classifying model can be detected for lung tuberculosis. These experiment results are evaluated on f1-score, precision accuracy, recall and comparison on measures of evaluation metrics.

Table 1. Classification performance based on precision, Recall, and accuracy metrics

Task	System	Precision (%)	Recall (%)	F1-score (%)	Accuracy (%)
Tuberculosis images	Xception-LSTM	92.67	87.34	89.43	88
	ViT-LSTM	91.81	91.71	92.77	93.4

The network architecture is trained using the different model hyper parameters and its classification performance is computed using training and testing dataset. The classification performance metrics obtained is shown in table 1. From the performance metrics it is clear that Xception stacked LSTM trained with model hyper parameter obtained in terms of accuracy (88%), F1-score (89.43%), Recall (87.34%), precision (92.67%) for lung tuberculosis detection. And also, the proposed model's (ViT-LSTM) accuracy, precision, recall, and F1-score were 93.4%, 91.81%, 91.71%, and 92.77%, respectively. The performance is based on the TB positive and TB negative classification of the classifier. In our thesis work we investigated Computer assisted or aided diagnosis (CAD) system for the binary classification of lung tuberculosis (TB positive and TB negative) using chest radiograph dataset based on the new deep learning method. We used the dataset learning rate of 0.001, and Adam optimizer with a training loop of epochs 30, 50, and 150. The following figure shows computational speed improvement of after and before segment lung tuberculosis chest radiograph images.

Table 2. Computational time of before and after lung segmented image based on parameter values

Tuberculosis image	Data acquisition	Number of epochs	Computational time (in seconds)
Before lung segment	350	30	703.25433
	550	50	4951.54489
	1047	150	27013.51342
After lung segmented	350	30	547.78613
	550	50	2371.73321
	1047	150	13782.54901

According to the above table, it is clearly observed that the average computational speed of the testing and training data is presented based on before segment and after segmented lung tuberculosis dataset. Before segmentation of the training data was employed, the computational time of the first 350 image is 703.25433 sec and after segmented the training data, the computational time of 350 lung TB image is 547.78613sec. When we train the data of before segmented image of 550 images, it results the computation time of 4951.54489 sec but after segmented image it results the computation time of 2371.73321sec.and also when we train the data set of before segmented image of 1047 images is 27013.51342 sec and after segmented image, it is 13782.54901sec. The results indicate 41% improvement in average computational speed.

Table 3. Comparisons result work

Authors	Method	Data set	Number of images	Accuracy
Rohilla [11]	Alex net and VGG net	Mixed Shenzhen& Montgomery	800	81.6%
R.Hooda[12]	Layer b/n LeNet and Alex net	Mixed Shenzhen& Montgomery	800	82.09%
Proposed model	**Xception-LSTM**	Mixed Shenzhen, Montgomery, JSRT dataset	1047	88%
	ViT-LSTM	Mixed Shenzhen, Montgomery, JSRT dataset	1047	93.4%

To evaluate the performance of our proposed system, we will discuss and compared our experimental result works with the results of other previously thesis works and based on performance metrics. Comparative result work of different deep learning models with the proposed exceptions combined LSTM models and ViT-LSTM elaborated from table 3. The performance of the evaluation or comparison results are based on accuracy performance metrics. The System or method in Rohilla and R. Hooda give the performance in terms of accuracy are 81.6%, 82.09%, respectively, whereas, the results obtained using the network architecture used here, give an accuracy of 88% and 93.4%.

When we compared with the previous published research work and existing methods, our proposed method achieves the better performance in terms of accuracy measures based on mixed Shenzhen and Montgomery lung tuberculosis data set. The research work proposed by [20], uses deep convolutional neural network (layer between LeNet and Alex Net) architecture. They have used architecture CNN with three fully connected layer, and seven convolutional layers. Their proposed approach based on three publically available dataset which are Montgomery (138 images) and Shenzhen (662), JSRT (247) data set. The method by [20], was not capable to reveals huge feature model parameter like, max pooling, global average pooling, separable convolution or depth wise separable convolution and transformers etc.

Therefore, the approach was not efficient and effective to extract their proposed frame work. The research work proposed by [19], uses deep Alex Net and VGG Net model architecture. They achieved the maximum accuracy of 81.6%. The method is not efficient and effective method. In our research work, we designed and implemented lung tuberculosis detection system by using computer assisted detection system of the new deep learning approach. The new deep learning techniques and approaches are Xception, Vision transformers and long short-term memory network architecture. The experiment result of our proposed approach is based different hyper parameter configuration and their values like activation function, learning rate, regularization (dropout) etc. The features generated by our Xception deep convolutional model are relatively good enough to be used in medical usage. From the above table, our comparable experimental result work shows that combined Xception and LSTM, ViT-LSTM and compared with other deep learning models. Our deep learning approaches have adequate performance results in terms of accuracy measures.

7. CONCLUSION AND FUTURE WORK

The main objective of our thesis work is to develop detection system for lung tuberculosis using deep learning approach. The approaches were performed by chest-xray images for lung tuberculosis detection and comparative experiment result analysis were evaluated with existing proposed frame work. The proposed frame work has four stages pre-processing, feature extraction, and classification. In pre-processing stage, we have used Adaptive histogram equalization and Gaussian filter technique. Adaptive histogram equalization (AHE) is done for image contrast and Gaussian filter is done for noise removal. The preprocessed image obtained as an input was subjected to thresholding, morphological analysis, and Active Counter model operator, facilitating the focus of the obtained findings on the lung area. The outcome from the lung area along with the feature extraction and classification was achieved by the application of deep learning techniques, namely Xception and ViT for feature extraction and LSTM network architecture for classification. The effectiveness of our suggested model, which uses an Xception stacked LSTM trained with a model hyper parameter, for the identification of pulmonary TB was measured in terms of accuracy (88%), F1-score (89.43%), recall (87.34%), and precision (92.67%). Additionally, the accuracy, precision, recall, and F1-score of the suggested model (ViT-LSTM) were 93.4%, 91.81%, 91.71%, and 92.77%, respectively. The performance is based on the TB positive and TB negative classification of the classifier. Using a lung TB image before and after segment, we performed a set of experiments, and the results indicate a 41% improvement in average computational speed. In future work, the research endeavors may employ the ensemble learning technique to enhance the detection accuracy of the model.

Contributions statement

Conceptualization, A.M; methodology, A.M, N.S,G.A and S.G.; software, A.M,G.A, S.G; validation, A.M., N.S; formal analysis, A.M.; investigation, G.A,S.G and N.S; writing—original draft preparation, A.M and S.G; writing— review and editing, A.M., G.A. and N.S. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement

Publicly available datasets were analysed in this study (https://lhncbc.nlm.nih.gov/LHC-downloads/dataset.html),

(https://paperswithcode.com/dataset/jsrt)

8. REFERENCES

[1] TUN; K.M.M. and KHAING, A.S. Implementation of lung cancer nodule feature extraction using digital image processing, 2014.

[2] https://www.who.int/news-room/fact-sheets/detail/tuberculosis

[3] S. Zhang,] Stefan Jaeger, Alexandros Karargyris, Sameer Antani, and George Thoma, September. Detecting Tuberculosis in Radiographs Using Combined Lung Masks.IEEE,2012.

[4] Jutturong Ckumdee, Somchai Santiwatanakul, Thongchai Kaewphinit, .Development of a rapid and sensitive DNA turbidity biosensor test for diagnosis of katG gene in isoniazid resistant Mycobacterium tuberculosis,2017.

[5] C. Leung, .Reexamining the role of radiography in tuberculosis case finding. The International Journal of Tuberculosis and Lung Disease, 2011.

[6] Rahul Hooda, Sanjeev Sofat, Simranpreet Kaur, Ajay Mittal, Fabrice M´eriaudeau,2017,September. Deep-learning: A Potential Method for Tuberculosis Detection using Chest Radiography

[7] Rahul Hooda , Ajay Mittal, 2018 ,Marcch. Automated Tuberculosis Classification of Chest Radiographs by Using Convolutional Neural Networks.

[8] Avinash Kumar, Sobhangi Sarkar and Chittaranjan Pradhan , Malaria Disease Detection Using CNN Technique with SGD, RMSpropand ADAM Optimizers, January,2020,

[9] Patil, M. J. S.”Deep learning in low resolution image recognition”. Vishwakarma Journal of Engineering Research, 1(2), 101-107,2017

[10] Chollet ,F.,Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1251-1258),2017.

[11] Anuj Rohilla, Rahul Hooda, Ajay Mittal,” TB Detection in Chest Radiograph Using Deep Learning Architecture,” Augest, 2017.

[12] Rahul Hooda, Ajay Mittal,”Automated Tuberculosis Classification of Chest Radiographs by Using Convolutional Neural Networks,” 2018, March.

[13] Stefan Jaeger, Alexandros Karargyris, Sameer Antani, and George Thoma,”Detecting Tuberculosis in Radiographs Using Combined Lung Masks,”2012.

[14] Jaeger, Stefan, Alexandros Karargyris, Sema Candemir, Les Folio, Jenifer Siegelman, Fiona Callaghan, Zhiyun Xue et al. Automatic tuberculosis screening using chest radiographs." IEEE transactions on medical, 2014.

[15] Rahul Hooda, Sanjeev Sofat, Simranpreet Kaur, Ajay Mittal, Fabrice M´eriaudeau,” Deeplearning: A Potential Method for Tuberculosis Detection using Chest Radiography. 2017, September, 2017.

[16] Mengchi Lu, Long Gao, and Xifeng Guo, Jianping Yin ,” Framework of Predicting Drug Resistance of Lung Tuberculosis by Utilizing Radiological Images,” 2018,March.

[17] M. K. Osman, Mohd Halim Mohd Noor, M. Y. Mashor, H. Jaafar, Compact Single Hidden Layer Feedforward Network for Mycobacterium Tuberculosis Detection.IEEE Xplore.

[18] Yuanli Wu, Hong Wang, Fei Wu. Automatic Classification of Pulmonary Tuberculosis and Sarcoidosis based on Random Forest, 2017

[19] Fares Hasan Obaid Alfadhli, Ali Afzalian Mand, Md. Shohel Sayeed, Kok Swee Sim, Mundher Al-Shabi ,. Classification of Tuberculosis with SURF Spatial Pyramid Features.2017.

[20] Fares Hasan Obaid Alfadhli, Ali Afzalian Mand, Md. Shohel Sayeed, Kok Swee Sim, Mundher Al-Shabi ,. Classification of Tuberculosis with SURF Spatial Pyramid Features.2017.

[21] (https://lhncbc.nlm.nih.gov/LHC-downloads/dataset.html).

[22] Yun, G.H., Oh, S.J. and Shin, S.C. Image Preprocessing Method in Radiographic Inspection for Automatic Detection of Ship Welding Defects. Applied Sciences, 12(1), p.123.2022

[23] Nikhil B. Image Data Pre-Processing for Neural Networks [online]. Available: https://becominghuman.ai/image-data-pre-processing-for-neural-networks.

[24] Eng. Michael Samir Labib Habib ” Msc thesis, Dept.Systems and A computer aided diagnosis system (CAD) for the detection of pulumnary nodules on CT scans”, Biomedical Eng., Cairo univ.,Giza, Egypt,2009.

[25] Senthilkumaran N and Vaithegi S” Image segmentation by using thresholding technique for medical images” , An International Journal (CSEIJ), Vol.6, No.1, 2016.

[26] K. Han, Y. Wang, H. Chen, X. Chen, J. Guo, Z. Liu, Y. Tang, A. Xiao, C. Xu, Y. Xu, et al., “A survey on vision transformer,” IEEE transactions on pattern analysis and machine intelligence, 2022.

[27] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, et al., “An image is worth 16x16 words: Transformers for image recognition at scale,” arXiv preprint arXiv:2010.11929, 2020.

[28] López, Y.P., Costa Filho, C.F.F., Aguilera, L.M.R. and Costa, M.G.F., October. Automatic classification of light field smear microscopy patches using Convolutional Neural Networks for identifying Mycobacterium Tuberculosis. In CHILEAN

Conference on Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON) (pp. 1-5). IEEE, 2017.

[29] https://en.wikipedia.org [online available] accessed 2019.

[30] Muhammad Imran Razzak, Saeeda Naz and Ahmad Zaib ‘’Deep Learning for Medical Image Processing: Overview, Challenges and Future ‘’

[31] Wogayehu Atilaw Mengesha , Menore Tekeba ,’’Lung Nodules Detection from Computed Tomography Scans Using Deep Belief Networks ‘’Addis Ababa University,October, 2018.

[32] C. Szegedy et al."Going deeper with convolutions," 2014. [Online Available] https://arxiv.org/pdf/1409.4842.pdf.

[33] F. Rosenblatt, "The Perceptron: A Probabilistic graphical model for information storage and organization in the brain," Psychological Review, vol. 65, no. 6, pp. 65-386, 1958

[34] Anish Sing Walia, ―Activation functions and its types-Types of Optimization Algorithms used in Neural Networks and Ways to Optimize Gradient Descent.2017. [Online Available] https://towardsdatascience.com/

[35] RaghavPrabhu, ―Understanding of Convolutional Neural Network (CNN)—Deep Learning, https://medium.com [online available] accessed 2019

[36] Abraham, A. Artificial neural networks. Handbook of measuring system design.2005.

[37] M. Arfan Jaffar, Riyadh, Saudi Arabia. Deep Learning based Computer Aided Diagnosis System for Breast Mammograms.

[38] Anon JSRT Database,Japanese Society of Radiological Technology.