

# An Add-on CNN based Model for the Detection of Tuberculosis using Chest X-ray Images

Roopa N K<sup>1</sup>, Mamatha G S<sup>2</sup>

Research Scholar, Dept. of Computer Science & Engineering, Sri Siddhartha Institute of Technology, Tumakuru, India<sup>1</sup>  
Dept. of Information Science and Engineering, RV College of Engineering, Bangalore, India<sup>2</sup>

**Abstract**—Machine Learning has been potentially contributing towards smart diagnosis in the medical domain for more than a decade with a target towards achieving higher accuracy in detection and classification. However, from the perspective of medical image processing, the contribution of machine learning towards segmentation is not been much to find in recent times. The proposed study considers a use case of Tuberculosis detection and classification from chest x-rays where a unique machine learning approach of Convolution Neural Network is adopted for segmentation of lung images from CXR. A computational framework is developed that performs segmentation, feature extraction, detection, and classification. The proposed system's study outcome is analyzed with and without segmentation over existing machine learning models to exhibit 99.85% accuracy, which is the highest score to date in contrast to existing approaches found in the literature. The study outcome based on the comparative analysis exhibits the effectiveness of the proposed system.

**Keywords**—Chest X-Ray; machine learning; convolution neural network; segmentation; detection; classification

## I. INTRODUCTION

Tuberculosis (TB) is a disease caused due to the bacteria *Mycobacterium tuberculosis* [1]. Most often, this affects the lungs and causes tissue damage in them, while its common symptom is found to be cough. These bacteria can spread through the air; hence they will spread from person to person when the infected person coughs, sneezes, or spits [2]. According to WHO, one-fourth of the world's population is already infected with these bacteria, but they are not ill and cannot transmit it further [3]. People infected with TB have a 5-10% lifetime risk of falling ill due to it. When the person falls ill due to TB, it shows immediate symptoms like cough, blood in the cough, high fever, chest pain, and in some cases, mental illness. In order to treat this, finding out the extent of damage caused to the lungs is most important. Early diagnosis is crucial for treatment.

TB is also most commonly confused with lung cancer due to all similar symptoms. When doctors have such confusion, they will go for a TB skin test and blood test; however, such a test doesn't exhibit the criticality of lung damage. Hence, Chest X-Ray (CXR) is commonly adopted for the identification of stages of the criticality of TB [4]. This is where the domain of medical image processing comes into play, where various algorithms and mechanisms are constructed in order to find all the indicative symptoms of abnormalities in the lung region when the subject is infected by TB. Some of the possible problems in this stage of acquisition of CXR image are that

there are higher possibilities of artifacts in that acquired CXR image [5]. This could be due to lower or fluctuating illumination conditions, absence of some significant region of lungs, movement of the subject during taking X-rays, etc. The conclusive remarks to state that a subject is confirmed of TB completely depends upon a manual analysis of the doctor. Because of this manual assessment, it's humanly impossible for any doctor to diagnose the disease for a large number of patients per day. Therefore, there is a serious need to make this identification system smarter and more intelligent, which can diagnose the disease without human intervention, which is manually not possible for humans. This is where artificial intelligence and machine learning come into the picture [6]. Currently, much research is being carried out toward adopting machine learning in diagnosing critical diseases [7]-[9]. There are also some significant studies where common diseases and diseases associated with COVID-19 have been assessed in recent times [10].

However, there are not many recent studies investigating TB. It has also been noticed that detection and classification completely depend upon the efficiency of processing input and extracted features. One such technique is known as segmentation, which differentiates the background image from the foreground image, offering more clearance to the system or a physician to make a conclusive decision about the diagnosis.

This paper introduces a novel computational model to perform segmentation and thereby contribute towards detecting and classifying CXR in TB patients. The contribution of the study is i) an augmented U-Net model is presented to carry out segmentation, ii) a computationally efficient training model is introduced, iii) a simplified preprocessing is carried out to leverage accuracy, and iv) accomplishing higher accuracy of classification till date. The organization of the manuscript is as follows: Discussion of existing literature where Machine learning is used for identification and classification is carried out in Section II followed by exclusive highlights of the research problem in Section III; Section IV showcases the research methodology, while Section V discusses system implementation. Section VI discusses the results obtained from the study. Finally, Section VII makes conclusive remarks about the study's contribution.

## II. RELATED WORK

This Section discusses the different mechanisms carried out in the existing system toward analyzing CXR images. The prime emphasis is given to the latest publication with different techniques of detection and classification of CXR. Although

the proposed study targets the detection and classification of Tuberculosis, this section mainly studies all potential approaches toward screening any form of significant deformities in CXR. This is meant to assist more information that could be assistive towards the detection and classification of Tuberculosis too.

The recent work carried out by Lin et al. [11] has presented a detection of COVID-19 symptoms right from CXR images using an adaptive attention network using ResNet. Although this work aims to extract contextual information for lesion detection in CXR, its limitation is that it is quite an iterative process in the course of learning the network. Another recent work by Wu et al. [12] has used a unique classification technique using the fractional order of convolution approach. The method also makes use of a radial Bayesian network in order to identify complex structures. However, the limitation of the model is that it involves a higher number of layered operations to perform classification. A study towards segmentation has been reported by Eslami et al. [13] that uses an adversarial network with conditional generative attributes where a network is constructed for all the pixels. Although the model claims support for multitasking applications, its limitation remains in complex architecture design over many pixels. A study towards adopting an adversarial network is also witnessed in work carried out by An et al. [14]. The study has developed a multi-appearance model for carrying out the extraction of significant features associated with COVID-19.

Further, the model uses a design of an adaptive network for a multi-scale adversarial domain for targeting better accuracy. The study's limitation is using a complex structure of an adaptive network irrespective of an accuracy of 98.83%. Londono et al. [15] present a unique evaluation-based model where a Convolution Neural Network (CNN) is used for training the dataset of CXR with COVID-19 symptoms. The study also performs preprocessing of data to deal with variability issues during analysis. The limitation of the study is lowered accuracy score of 91.5% while ignoring the process of masking the images. The problems associated with detecting lesion in CXR is carried out by Li et al. [16] where an amplitude modulation scheme has been presented to extract deformation features within CXR with deformable convolution. The study also uses the loss function of regression in order to carry out optimization; however, a lowered precision score of 0.914 is obtained as a limitation. A unique form of machine learning approach called a deep zoom neural network is presented by Wang et al. [17] that targets optimizing the training process using CXR images for thoracic diseases. The study has used U-Net in order to carry out segmentation followed by using attention heatmap, the regions of lesions are extracted. The limitation of the study is that it is more inclined toward detection and possesses fewer conditional constraints for supporting classification. The study of tuberculosis detection is presented by Rahman et al. [18] using deep CNN for classification. The study's limitation is its accuracy of 98.6% without much emphasis on the masking process. Further study towards segmentation of CXR images is carried out by Munawar et al. [19], focusing on using the generative adversarial network. This study carries out segmentation masking while the model training is carried out

using multiple discriminators. The limitation of the study is its lower accuracy, irrespective of using the sophisticated adversarial network. It is also observed that most of the existing system focuses more or less on detection or classification; however, not much emphasis is offered to outlier detection, which is essential to confirm its accuracy. A study in such a direction is carried out by Kim et al. [20] using artificial intelligence. The study also uses a Recurrent Neural network to perform learning operations. The limitation of the model is the inclusion of higher convolution layers; still, extraction of potential features doesn't carry any objective function.

Paluru et al. [21] have presented a study of chest CT images, especially focusing on the segmentation process using a unique form of CNN called as Anam-Net. The study claims to include lower parameters to show its lightweight features. The limitation of the model is that it still has a dependency on high-end resources in order to make it functional. Synthesis is one of the essential operations for analyzing CXR, while a study in this direction is carried out by Salehinejad et al. [22]. The study has used CNN, whose outcome shows the supportability of five classes of CXR. The limitation of this model is that it lacks potential preprocessing and planning while taking input for real- images. Catala et al. [23] have investigated CXR images for the identification of Pneumonia with more emphasis on the dataset. Work carried out by Zhang et al. [24] has carried out a segmentation process for CXR using deep learning, which is capable of generating information about infection over the lungs. The limitation of the study is its lowered accuracy score of 95.9%. The work carried out by Zhou et al. [25] used ResNet and Support Vector Machine (SVM) in order to detect abnormalities in CXR images. The study also uses image regrouping, where encoders of deep networks extract the features. The limitation of the study is its lowered accuracy of 93% only. The work carried out by Zaidi et al. [26] has used a tailored make CNN model to perform lung segmentation. The limitation of the study is its inclusion of a higher number of iterations in order to achieve below-average accuracy. Similarly, there are various other works in the same problem domain carried out by Wu et al. [27], Yan et al. [28], Fan et al. [29], and Lian et al. [30]. Therefore, various detection and classification mechanisms exist in current times towards CXR images with reported claims of accuracy outcome and limitation. The next section highlights identified research problems.

### III. RESEARCH PROBLEM

After reviewing the existing system to analyze CXR images, the following are the open-end research problems being identified:

- Existing models for the detection of lung abnormalities have mainly found lung anatomy, which is preferentially applicable to low-level image processing. However, such methods often produce inferior segmentation when certain areas of the lung are malformed or missing from CXR.
- Machine learning techniques are mainly used for classification, mainly for feature extraction. CNN is one such dominant machine learning model. However, existing studies using this technique were generally

found to be computationally slow and expensive due to higher iterations and resource involvement. It also does not help extract information on proper thresholding, which could be a potential problem during training operations. Other than that, there is not much emphasis on consistency and scalability since the uncertainty increases when the segmented regions change to different resolutions. Beyond that, not many techniques have been used to classify specific regions of the lung and its connected organs. From this perspective, research on segmentation is rather sparse.

- Existing studies have adopted the distortion model for the investigation of CXR. They were primarily used for the segmentation of lung fields in CXR; however, their performance is not up to the mark for large networks or high numbers of training images. There is less evidence to be verified at the same time, which reduces applicability in a practical world scenario. Furthermore, the mechanism to initialize the lung model during Tuberculosis has also been found to be manual and error-prone.
- There is no denying the fact that adopting CNN or other machine learning approaches gives good classification performance. However, the participation of machine learning approaches to perform autonomous segmentation is much lower than in existing systems. Without a proper segmentation considering all possible constraints of the CXR images, an improved form of identification and classification cannot be made. Furthermore, most machine learning approaches applied to CXR images have achieved low accuracy scores, while there is still scope to optimize them. Therefore, existing studies are found to be more inclined to implement complex architectures on machine learning rather than trying simplified modeling of it.

Therefore, the problem statement of the proposed study is "Optimized usage of machine learning towards identification of abnormalities in CXR images with more focus on segmentation approach is quite a challenging task".

#### IV. PROPOSED SYSTEM

The research work reported in this paper aims to develop an efficient and robust computational model that can accurately identify and classify TB disease using chest X-ray (CXR) images. As it has been identified based on review of literature that there is wide adoption of CNN for medical image analysis such as detecting tuberculosis from chest X-ray images. However, due to the complexity of CXR images, which include detailed information about the shoulder bones, rib cage, and outer body of the person, direct usage of CNN on these images may not yield very accurate results. In this regard, the proposed work presents a highly integrated system with more optimized computing operation in automating the task of TB diagnosis. The schematic architecture of the proposed system is illustrated in Fig. 1.

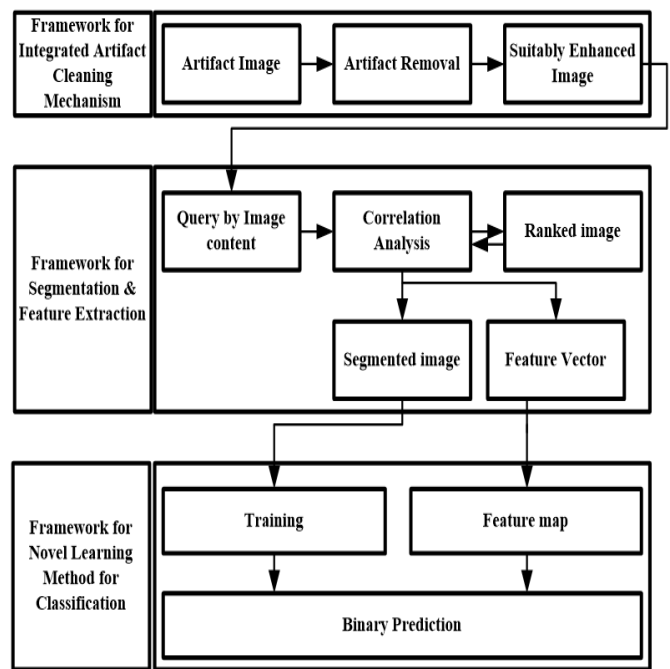


Fig. 1. Block based Architecture of the proposed system.

The proposed system aims to accurately diagnose TB by using a combination of exploratory data analysis, image segmentation, and machine learning. Firstly, the system executes exploratory data analysis to understand the data and determine the preprocessing requirements to make the input data suitable for learning models. This includes data cleaning, artifact removal normalization, and feature engineering.

Secondly, the system performs image segmentation to extract the region of interest from the input CXR image. This is done using a customized and enhanced version of the UNet Model, which accurately locates the lungs and masks out residual regions. This segmentation process reduces the surface of computational complexity, improves the feature extraction process, and increases accuracy by extracting only the region of interest. The generated masks from the proposed UNet learning model accurately represent the boundaries of the relevant structures in the images.

Finally, the generated masks serve as input to a CNN classifier to diagnose or classify TB. The CNN is trained on a large and diverse dataset of CXR images with and without TB, and is designed to accurately classify the input image as TB positive or TB negative.

Overall, this system has the potential to significantly improve the accuracy and efficiency of TB diagnosis, which is crucial for effective treatment and management of the disease. The performance of proposed system will be thoroughly evaluated on various metrics, and compared with existing methods in the literature.

##### A. Introduction to CXR based Diagnosis of TB

To understand the research flow, it is essential to comprehend the working mechanism of chest X-ray (CXR) and how it can be used to diagnose tuberculosis (TB). It is important to note that there are several techniques available to

diagnose TB in patients, which range from invasive to non-invasive methods. Radiological diagnosis is considered a non-invasive method as it does not require any surgical tools to be inserted inside the patient's body, thereby posing a lower risk. However, CXR diagnosis is not always reliable as it requires an expert pulmonologist to interpret the results accurately. Fig. 2 to 4 illustrate how an expert pulmonologist makes a diagnosis using CXR.



Fig. 2. Illustration of normal lungs.



Fig. 3. Illustration of TB lungs.



Fig. 4. Illustration of damaged lung tissues due to TB.

In CXR, an image of the chest is captured using an X-ray machine, which produces a black and white image of the chest cavity. The image contains detailed information about the structures inside the chest, including the lungs, heart, and ribs. A trained pulmonologist can use this image to identify abnormalities such as nodules, masses, and infiltrates that may indicate the presence of TB. The process of CXR diagnosis involves a thorough examination of the image by an expert pulmonologist who looks for specific patterns and abnormalities that may indicate TB.

The presence of TB in the lungs can cause damage to lung tissue, as shown in Fig. 3. However, it can be difficult for a layperson to distinguish tissue damage caused by TB from damage to the muscles of the lungs. To diagnose TB, doctors typically perform a TB skin test or a TB blood test. The TB skin test involves injecting a small amount of tuberculin under the skin and observing whether a blister forms at the injection site after 72 hours. However, this test is painful and only indicates the presence of TB bacteria on the skin. Similarly, the TB blood test only indicates the presence of TB bacteria in the blood. Neither of these tests can reveal the extent of lung damage caused by TB, which is critical for accurate diagnosis. The most definitive way to assess lung damage is through a lung biopsy, but this invasive procedure carries a risk of secondary lung infections. To avoid this risk, expert doctors rely on two non-invasive methods: listening to the sound of the lymph glands with a stethoscope while the patient breathes, or using chest X-rays to diagnose tissue damage in the lungs.

The manual process requires specialized knowledge and experience, which can make it challenging to diagnose TB accurately. Therefore, an effective computational model is developed that can accurately diagnose TB using CXR images.

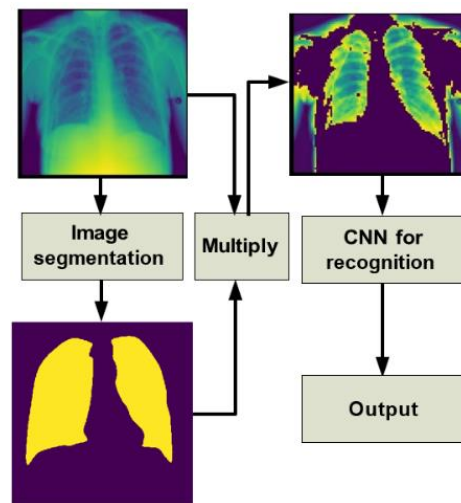


Fig. 5. Outline of the proposed computational model for TB diagnosis.

The illustration of computational model is shown in Fig. 5, which uses image processing techniques and machine learning algorithms to automate the process and improve the accuracy of TB diagnosis. The proposed computational model, as illustrated in Fig. 5, consists of three main components: a CXR image, its segmentation using a customized UNet model, and TB recognition using a CNN with the obtained mask. The

target accuracy for proposed system is a minimum of 90%, because an accuracy of 90% is the bare minimum in the medical system to be acceptable to the reliability of a biopsy. Additionally, our proposed method is non-invasive, providing a desirable alternative to invasive biopsy for TB diagnosis. One of the major advantages of our model is the adoption of novel segmentation techniques using machine learning, which makes the entire process autonomous and independent of human intervention. This feature not only reduces the possibility of errors due to human intervention but also saves time and resources.

## V. SYSTEM IMPLEMENTATION

As mentioned in the previous section, the proposed system aims at identifying and classifying tuberculosis from chest radiographs (CXR). This section discusses the implementation procedure adopted in the proposed system design system. The entire discussion is carried out in systematic manner following preprocessing operation and Add on CNN based TB detection. Here Add on means applying proposed customized UNet model for precise mask generation for supervised learning.

### A. Pre-processing

The first step of the research methodology is to collect a large dataset of CXR images that includes both normal and TB-infected images. The dataset considered is diverse enough to cover normal and TB infections images. In the next step, the CXR dataset is then subjected to extensive preprocessing operation to enhance the image quality and remove any artifacts. This implementation of preprocessing phase includes checking size of image and converting to grayscale, resizing, and removing any artifacts or noise, and then splits them into training and testing datasets using an 80:20 ratio. The preprocessing over input CXR images is done using a Python library OpenCV's `cv2.resize()` function. For example, the signature used for resizing image to  $512 \times 512$  is shown as follows:

$$\text{Resized\_CXR} = \text{cv2.resize}(\text{CXR}, (512, 512)).$$

The enhancement of the each CXR images in the dataset are carried out using adaptive histogram-based equalization approach. The implementation step is discussed as follows:

- 1) Let an input CXR image be denoted by  $I(x, y)$ , where  $x$  and  $y$  represent the spatial coordinates of the image.
- 2) Divide the input image  $I(x, y)$  into non-overlapping tiles of size  $N \times N$ .
- 3) For each tile, compute the histogram  $H(i, j, k)$ , where  $i$  and  $j$  represent the pixel coordinates within the tile and  $k$  represents the intensity levels ranging in 0 to  $L - 1$ , where  $L$  is the number of gray levels.
- 4) Apply the Contrast Enhancement Function (CEF) on each tile, which maps the original pixel values to new enhanced values, denoted by  $E(x, y)$ . The CEF is numerically expressed as follows:

$$E(i, j) = \frac{\text{round}(\frac{\text{CDF}(U(i, j)) - \text{CDF}_{\min}}{(1 - \text{CDF}_{\min}) \times (L - 1)})}{(1 - \text{CDF}_{\min}) \times (L - 1)} \quad (1)$$

where CDF is the Cumulative Distribution Function,  $\text{CDF}_{\min}$  is the minimum CDF value in the tile, and round function rounds the values to nearest integer.

- 5) Replace the pixels within each tile with the enhanced values  $E(x, y)$ .
- 6) Reconstruct the output image by combining the enhanced tiles.

The output image is denoted by  $I'(x, y)$ , as enhanced version of the input CXR image  $I(x, y)$ . Similar operation is executed for  $\forall \text{CXR} \in \text{Dataset}$ . Afterwards, dataset is split into training and testing sets with an 80:20 ratio using Python's `train_test_split()` function from the scikit-learn library. Overall, this phase applies a basic pre-processing scheme to make CXR images suitable for segmentation process.

### B. Segmentation and Mask Generation

In this phase, a customized UNet model (i.e., UNet with bi-ConvGRU model) is developed and trained to segment the lung region from the input CXR images. Once the lung region is segmented, a mask of the region of interest (ROI) is generated to focus on the segmented lungs only. The output of the proposed UNet model is a binary mask that shows the segmented lung region.

The customization in UNET is done by integrating it with bi-ConvGRU layer to each of the encoder and decoder blocks in the UNET model. Basically, the convolutional layers in the encoder and decoder are replaced with bi-ConvGRU which consist of a set of updates and reset gates that control the flow of information through the network, and a hidden state that stores the current state of the network. The bi-convGRU layers learn to selectively update and forget information based on the input image and the previous state of the network. This layers also have attention mechanisms that allow the network to focus on specific parts of the input image when making predictions. The attention mechanisms use learned weights to selectively weight the input image features at different spatial locations. The mathematical model for the U-Net architecture with bi-ConvGRU can be described as follows:

**Input:** An enhanced image of size  $W \times H \times C$ , where  $W$  is width,  $H$  is height, and  $C$  is the number of channels.

**Output:** A segmentation map or a mask of size  $W \times H$ , where each pixel represents the class of the corresponding pixel in the input image.

**Encoder:** The input image is passed through a series of convolutional layers with filters of increasing size. Here, the encoder takes the input  $X$  and generates a set of feature maps  $E$  such that:

$$E = \{E_1, E_2, E_3, E_4, E_5\} \quad (2)$$

Each convolutional layer is followed by batch normalization and activation function (ReLU). The output feature maps are downsampled using max-pooling such that:

$$E_i = \text{Pool}(\text{ReLU}(\text{Conv}(E_{i-1}))) \quad (3)$$

The output of the  $i$ th encoder block is denoted as  $U_i$ , where  $i$  denotes the depth of the block. The output feature map  $U_i$  has a spatial size of  $W_i \times H_i$  and contains  $C_i$  channels.

**Decoder:** The decoder upsamples the feature maps from the encoder, restoring the original image size. The  $i$ th decoder

block takes the output from the  $i$ th encoder block and concatenates it with the feature map from the corresponding level of the encoder, which helps to capture more detailed information. The output of the  $i$ th decoder block is denoted as  $D_i$ , where  $i$  denotes the depth of the block. The output feature map  $D_i$  has a spatial size of  $W_i \times H_i$  and contains  $C_i$  channels.

**Bi-ConvGRU:** The bi-ConvGRU processes the feature maps from the encoder and decoder. It enhances the network's ability to capture long-range dependencies. The bi-ConvGRU is implemented as follows:

- A convolutional layer with  $1 \times 1$  kernel size is applied to the feature maps from the encoder and decoder to reduce the number of channels.
- The output feature maps are then passed through two layers of bi-directional ConvGRU, which processes the feature maps in both forward and backward directions.
- The output of the bi-ConvGRU is then upsampled to the original image size.

The output of the bi-ConvGRU is denoted as  $G$  and has a spatial size of  $W \times H$  and a single channel.

**Final layer:** The output of the bi-ConvGRU is passed through a convolutional layer with  $1 \times 1$  kernel size to obtain the final segmentation map. Fig. 6 shows schematic architecture of the implemented customized UNet model.

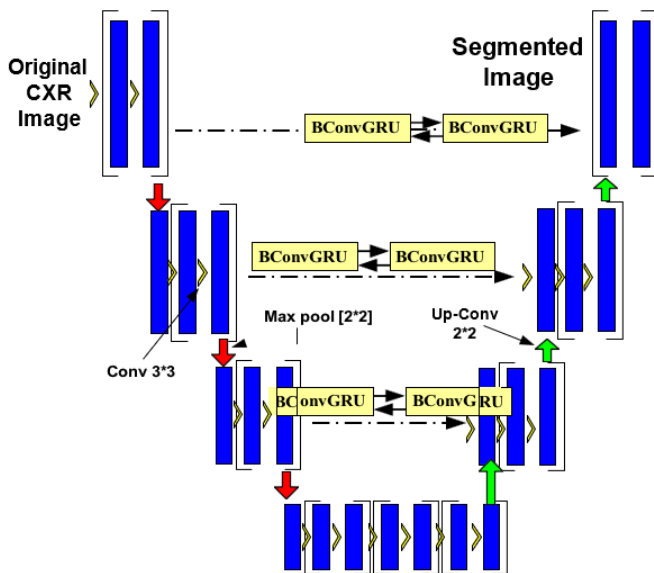


Fig. 6. Proposed customized U-Net model for segmentation and mask generation.

For  $\forall$  pixel at location  $(i, j)$  in the feature map at layer  $l$ , the encoding process applies a convolution operation to the corresponding receptive field in the previous feature map at layer  $l - 1$ , adds a bias term, and applies an activation function to obtain the output feature map at layer  $l$ . This process is repeated for all pixels in the feature map to obtain the full output feature map at layer  $l$ , expressed as follows:

$$h_{i,j}^l = f(W_l \times h_{i,j}^{l-1} + b_l) \quad (4)$$

where:

- $h_{i,j}^{l-1}$  represents the feature map at layer  $l - 1$  of the previous convolutional layer,
- $h_{i,j}^l$  represents the feature map at layer  $l$ ,
- $W_l$  is the weight matrix for the convolutional operation at layer  $l$ ,
- $b_l$  is the bias vector at layer  $l$ , and
- $f(\cdot)$  is the activation function applied element-wise to the output of the convolution operation.

Specifically, the customized UNet is used to learn the mapping between the input image and the output segmentation masks. The learned model parameters can then be used to minimize the energy function, which consists of a data fitting term and a regularizing term, to generate the optimized segmentation. Eq. (5) represents the image segmentation function with the addition of transform domain analysis to enhance its performance. It includes a term that measures the curvature of the segmentation contours in the transform domain, which helps to improve the smoothness of the segmentation.

$$\operatorname{argmin}_u |k| + \mu \int_c |\Delta u|^2 dx + \int (u - f)^2 dx \quad (5)$$

In Eq. (5), the variable  $f$  represents the original image pixel values,  $u$  represents the average pixel strength in a segment,  $\Delta u$  represents the variance of pixel intensities in a segment, and  $k$  represents a constant that is optimized by Bi-Conv-GRU.

### C. CNN based TB Diagnosis

In the previous section, the image segmentation is performed using a customized UNet model, which generates a binary mask of the regions of interest (ROI) in the input image. The trained UNet model is then used to generate masks from other CXR datasets by processing the input CXR images through the trained UNet model. The output of the UNet model is a binary mask that shows the segmented lung region. The generated mask is then used to train a CNN model for TB detection. The input to the CNN model is the generated mask, and the output is the binary classification of TB or non-TB. The CNN model is trained using the preprocessed training dataset, and its performance is evaluated using the preprocessed validation dataset. The TB diagnosis model involved implementation of the CNN architecture which includes convolutional layers, pooling layers, and fully connected layers with activation functions such as ReLU and sigmoid. The loss function used could be binary cross-entropy, and the optimization algorithm is used as Adam optimizer. Mathematically, this process can be represented as follows:

Let  $I$  be the input CXR image, and  $M$  be the binary mask generated by the UNet model. The trained UNet model can be represented as a function  $F$ , which takes the input image  $I$  and generates the binary mask  $M$  such that:

$$M = F(I) \quad (6)$$

The binary mask  $M$  is then fed into a CNN model, which takes it as input and generates a binary classification output, indicating whether the image contains TB or not. The CNN

model can be represented as a function  $G$ , which takes the binary mask  $M$  as input and generates the binary classification output  $y$ :

$$y = G(M) \tag{7}$$

Thus, the entire process of image segmentation and mask generation using UNet, and TB detection using a CNN can be represented as a composite function  $H$ , which is the composition of  $F$  and  $G$ , given as follows:

$$H(I) = G(F(I)) \tag{8}$$

To train the CNN model, we use the binary mask  $M$  generated by the UNet model as the input, and its corresponding label (0 or 1, indicating whether the image contains TB or not) as the output. The CNN model is trained using a binary cross-entropy loss function, given as follows:

$$L(y_{true}, y_{pred}) = -\alpha + (\beta \times \gamma) \tag{9}$$

where:

$$\alpha = (y_{true} \times \log(y_{pred}))$$

$$\beta = (1 - y_{true})$$

$$\gamma = \log(1 - y_{pred})$$

In Eq. (9), the variable  $y_{true}$  is the true label,  $y_{pred}$  is the predicted output from the CNN model, and  $\log$  represents the natural logarithm. The training process involves minimizing the loss function over a set of training images and their corresponding labels, using an optimizer, Adam. The negative sign in the beginning of the right-hand side of the equation is used to indicate that we want to minimize the loss function. During training, the objective is to minimize the difference between the true labels and the predicted labels. By adding a negative sign to the equation, we can use optimization algorithms that are designed to minimize a function, rather than maximize it. In other words, by minimizing the negative of the log likelihood loss, we are maximizing the likelihood of the predicted labels given the true labels. Once the CNN model is trained, it can be used to classify new CXR images as either TB positive or negative, by first generating the binary mask using the UNet model and then passing it through the CNN model. The proposed learning model can be used as a tool for early diagnosis of damage caused by TB in the lungs. To be more specific, this device will help in the detection of tuberculous pneumonia or tuberculous pneumothorax.

## VI. EXPERIMENTAL ANALYSIS

The design and development of the proposed system is done using Python executed in Anaconda distribution. The proposed work encompasses two main contributions Viz. i) development of novel customized UNet learning model which is trained to generate binary mask as segmented lung region from the given chest X-ray image; ii) implementing and training CNN with generated mask to diagnose Tuberculosis from the given mask generated. Both models are integrated and their response are synchronized to carry out detection of Tuberculosis. This section presents the outcome and performance analysis of the proposed learning models.

### A. Performance Analysis of Segmentation Model

The dataset used to train the proposed customized UNet model for image segmentation is highlighted in Table I.

TABLE I. DATASET USED FOR TRAINING UNET FOR IMAGE SEGMENTATION

Dataset	Total images	Training images	Testing ratio
Lung image segmentation [31]	704	563	141

As shown in Table I, the dataset obtained from Kaggle which consists of total 704 chest X-rays with corresponding masks. The dataset is further split in training and testing set with a ratio of 80:20. The parameters used to train the models is also highlighted in Table II.

TABLE II. TRAINING PARAMETERS USED FOR IMAGE SEGMENTATION

No. Epochs	Batch size	Loss function	Optimizer
50	32	Binary crossentropy	Adam

The training of the proposed augmented UNet model for generating mask in form of segmented image is carried out with total 563 images. The model is trained for 50 epochs, followed by batch size equals 32. In addition, binary cross entropy and Adam is used as loss function and optimized to calibrate learning of the model. The trained UNet model is further applied to different chest-Xray to generate mask.

The interpretation of the Fig. 7 reveals that the predicted mask is mostly the same as the original mask with only minor differences, it indicates that the UNET-based segmentation model is performing well on the given CXR image. The smooth edge border in the predicted mask could be due to the model's ability to capture fine details and edges in the image. A good performance is achieved by the customized based segmentation model indicating that it is able to accurately generate mask with lung areas. This would be useful in a TB detection system, as it can provide a good scope for an accurate localization of the disease for further analysis and diagnosis.

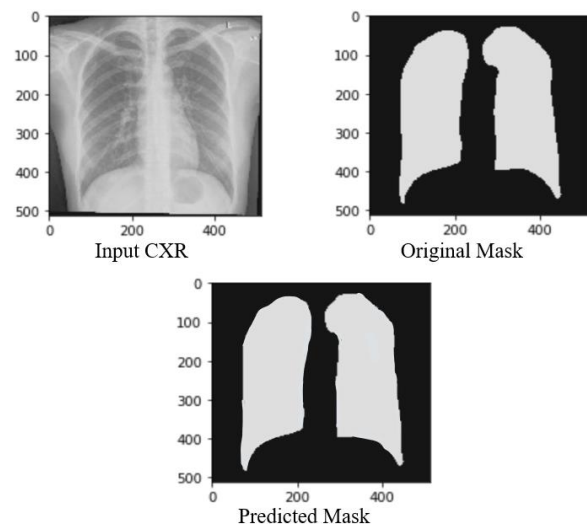


Fig. 7. Depiction of mask predicted by the proposed customized U-Net model with input CXR and original mask.

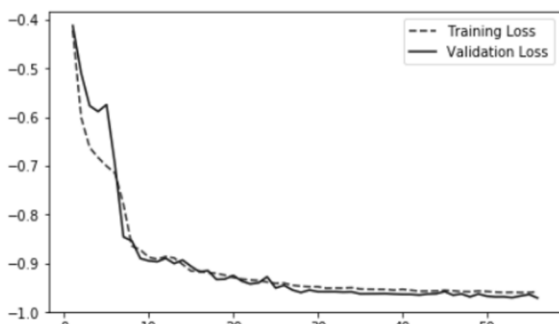


Fig. 8. Model training and validation loss curve analysis.

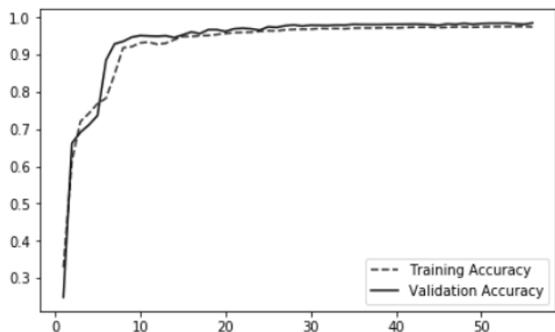


Fig. 9. Model training and validation accuracy curve analysis.

The training and validation loss curve in Fig. 8 shows the evolution of the loss (or cost) function over time as the model is trained on the training data. It can be seen from the graph trend that the loss decreases over time or epochs as the model learns to make better predictions. Similarly, the graph trend of training and validation accuracy curve shown in Fig. 9 accuracy curves follow a stable and linear trend towards 0.9696 accuracy on 50 epochs suggests that the model has converged and is performing well on the given task. Although, there is initial fluctuations in the training and validation accuracy curves which are common and can be attributed to the random initialization of weights in the model and the stochastic nature of the optimization algorithm. As the training progresses, the model learns more features from the data and the accuracy of the model on both the training and validation data gets improved. This also indicates that the model is not overfitting to the training data and generalizes well to new, unseen data.

**B. Performance Analysis of CNN based TB Detection**

The dataset used for generating mask using trained UNet model and training CNN model for detecting TB is highlighted in Table III.

A synthetic dataset is used in the phase of disease (TB) detection. Many chest X-rays images including normal and Tuberculosis are collected from different sources to build a dataset. The dataset consists of 7000 images, where 3500 images are belongs to normal CXR and remaining 3500 CXR images belongs to TB. The images are processed with the trained UNet model to generate a mask for all the input images towards performing efficient supervised learning for disease detection. After generating mask, dataset is split into training

and testing set with ratio of 80:20. Table IV shows the training parameters used for training the CNN for TB detection.

The classification model is trained over 5600 masks for 50 epoch, and batch size considered equals to 32. Binary cross entropy loss function and Adam optimizer is used to improve the learning of the CNN learning model. The performance of trained CNN is assessed with testing dataset. The analysis is conducted using confusion matrix and classification performance indicators such as accuracy, precision, recall rate and F1-score. The accuracy is a general metric that measures the overall performance of the model, while precision and recall are more specific metrics that measure the model's ability to correctly classify positive samples. The F1-score is a balanced metric that considers both precision and recall. The confusion matrix for TB detection model is shown in Figure 10.

The confusion matrix shown in Fig. 10 reveals that the trained CNN model using the generated masks from the proposed customized UNet model has performed well in detecting both normal and TB images. The confusion matrix shows that out of 700 normal images, 693 have been correctly predicted as normal, which gives a precision of 1.00 and a recall of 0.99. Similarly, out of 700 TB images, all have been correctly predicted as TB, giving a precision of 0.99, recall and F1-score of 1.00 as shown in Table V.

TABLE III. DATASET USED FOR GENERATING MASK AND TRAINING CNN FOR TB DETECTION

Dataset	Total images	Training images	Testing ratio
TB lungs classification dataset	7000	5600	1400

TABLE IV. TRAINING PARAMETERS FOR TB DETECTION USING CNN

No. Epochs	Batch size	Loss function	Optimizer
50	32	Binary crossentropy	Adam

TABLE V. QUANTIFIED OUTCOMES FOR TB DETECTION

	Precision	Recall	F1-score	Support
<b>Normal Lung</b>	0.99	0.99	0.99	700
<b>TB Lung</b>	1.00	1.00	1.00	700
<b>micro avg</b>	0.99	0.99	0.99	1400
<b>macro avg</b>	1.00	0.99	0.99	1400
<b>weighted avg</b>	1.00	0.99	0.99	1400

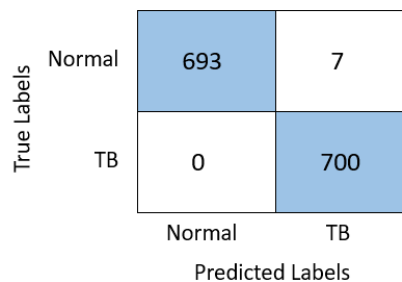


Fig. 10. Confusion plot for assessing performance of trained CNN for TB detection.



The quantified values shown in Table V represent the performance metrics of a binary classification model for the classes "Normal" and "TB". The precision for "Normal CXR" is 1.00, meaning that all the samples predicted as "Normal" were correctly classified, and for "TB" it is 0.99, indicating that out of all samples predicted as "TB", 99% were actually "TB". The recall for "Normal" is 0.99, meaning that out of all the actual "Normal" samples, the model correctly identified 99%, and for "TB" it is 1.00, indicating that the model correctly identified all the actual "TB" samples.

The micro avg. represents the performance metrics computed globally by counting the total true positives, false negatives, and false positives, and the macro avg represents the average of the performance metrics calculated for each class. The weighted avg. takes the weighted average of the performance metrics, where the weights are the number of samples in each class. Overall, the model has high precision, recall, and F1-score for both classes, indicating a good performance in the binary classification task.

### C. Extensive Analysis with Different Versions of CNN

This section presents an extensive analysis of the proposed work by evaluation proposed system with different version of existing CNN models. Moreover, entire analysis is carried in two scenarios viz. i) performance analysis with proposed segmentation scheme using customized UNet and ii) performance analysis without segmentation (i.e., models are trained directly on the input images with any pre-processing and segmentation).

For an effective analysis, the proposed system is subjected to a comparative analysis. However, this is a different form of comparative analysis, where the outcome of proposed system (performance parameters with segmentation) is further subjected to 7 version of neural network viz. Resnet 18 [32], Resnet 50 [33], Resnet101 [34], VGG19 [35], Densenet201 [36], Squeezenet [37], and Mobilenet [38].

The prime justification behind this mode of comparative analysis is to assess whether the performance of proposed classification model with segmentation using augmented U-Net model can be further improved upon using above mentioned seven variants of CNN model with different layers. The idea is to perform an assessment with respect to standard performance parameters of Accuracy, Precision, Sensitivity, F1-Score, and Specificity using dual combination of with proposed system (i.e., with segmentation) and without proposed system (i.e., without segmentation).

As shown in Fig. 11, the accuracy of learning models with proposed segmentation scheme ranges between 99-100, whereas that of without segmentation ranges between 95-97 for different types of CNNs architecture or model. Nearly similar trend is also observed for precision (Fig. 12), sensitivity (Fig. 13), F1 score (Fig. 14) and specificity (Fig. 15).

The F1 score with segmentation is found in range of 97-98 whereas that of without segmentation is found between 95 and 96 mainly. Out of different variants of CNN, the outcome is found to be better for Squeezenet and Mobilenet for accuracy followed by Resnet 18 in Fig. 11. The outcome of precision from Fig. 12 showcases Resnet50, Resnet101, and Squeezenet

to be performing well compared to other variants of CNN model. The outcome of sensitivity from Fig. 13 highlights better performance of Densenet201 followed by Mobilenet and Resnet101. Similar performance trend can be also seen with respect to higher F1 score from Fig. 14 compared to others versions of CNN. The performance of Resnet50, Resnet101, Densenet201, and Squeezenet offers nearly similar performance for specificity shown in Fig. 15.

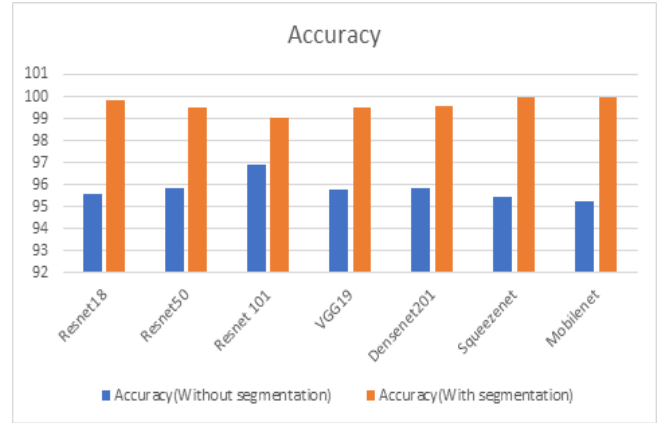


Fig. 11. Comparative analysis of accuracy (%).

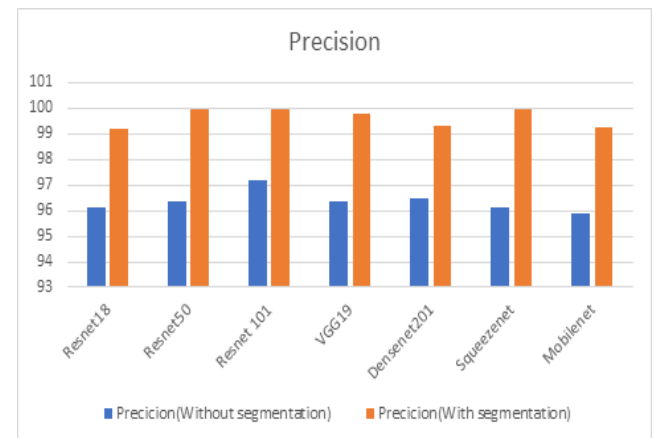


Fig. 12. Comparative analysis of precision (%).

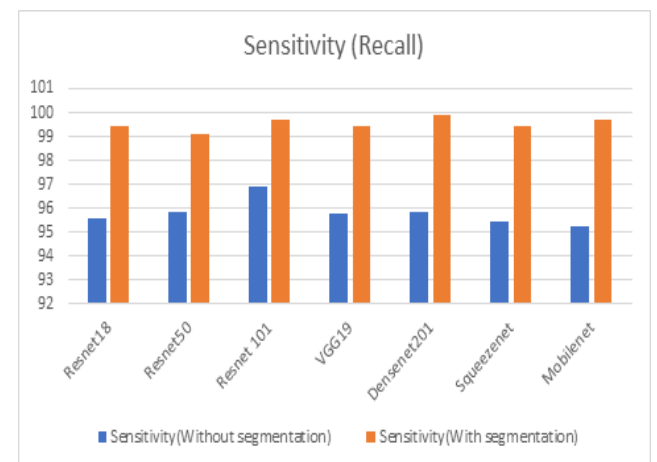


Fig. 13. Comparative analysis of sensitivity (%).

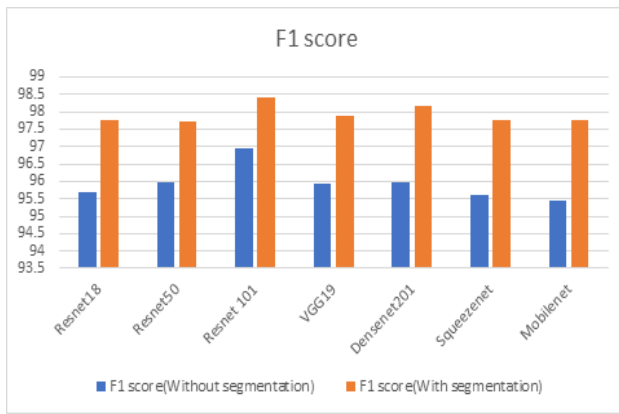


Fig. 14. Comparative analysis of f1 score (%).

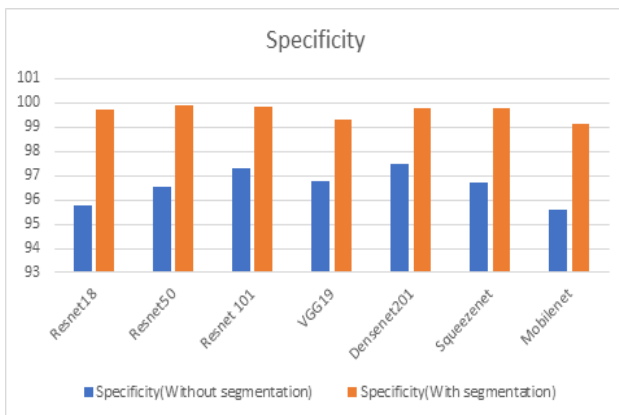


Fig. 15. Comparative analysis of specificity (%).

Based on the analysis and performance statistics, it can be concluded that the proposed scheme has provided a better performance with strong potential in TB diagnosis. The customized UNet model is able to accurately capture the important features in the input CXR images and generate precise masks, which are then used to train the CNN model. This results in a more robust and accurate CNN model, which is able to accurately classify the CXR images as normal or TB. Therefore, the customized UNet model acts as an important pre-processing step, which enhances the performance of the CNN model. Overall, the combination of exploratory data analysis, image segmentation, and appropriate training of CNN provided a more accurate and efficient approach to TB diagnosis, which is crucial for effective treatment and management of the disease.

## VII. CONCLUSION

In this paper, the research work has suggested a novel automated disease diagnosis scheme based on the Add-on CNN learning model which benefited using augmented U-Net based segmentation algorithm. The implementation of the proposed system follows a systematic procedure which included exploratory data analysis, image segmentation, and CNN to accurately diagnose TB. The proposed system effectively preprocesses the input data, locates the lungs, and extracts the region of interest from the CXR images, using customized UNet model which is augmented with Bi-ConvGRU layers to carry out precise and optimized form of

segmentation. This procedure greatly improves the accuracy and efficiency of the subsequent classification process. The generated masks from the proposed UNet Model accurately represent the boundaries of the relevant structures in the images, which improves the feature extraction process and increases the accuracy of the CNN classifier. Our evaluation on various metrics with extensive analysis demonstrates the effectiveness and potential of the proposed system in TB diagnosis. In the future, we plan to further improve the proposed system using self-exploration like reinforcement learning and apply it to larger datasets and different types of medical images. Also, the future study focuses on security aspect of medical imaging in telemedicine application.

## REFERENCES

- [1] Y. Shah, *Molecular Epidemiology Study of Mycobacterium Tuberculosis Complex*, IntechOpen, ISBN: 9781839680991, 1839680997, 2021.
- [2] B.R. Bloom, K.K. Holmes, P. Jha, S. Bertozzi, *Major Infectious Diseases*, ISBN: 9781464805240, 1464805245, 2017.
- [3] World Health Organization, *Global Tuberculosis Report 2019*, ISBN: 9789241565714, 9241565713, 2019.
- [4] G.B. Migliori, M. C. Raviglione, *Essential Tuberculosis*, ISBN: 9783030667030, 3030667030, 2021.
- [5] S. Bhalla, S. Martínez-Jiménez, *Chest Imaging*, Oxford University Press, ISBN: 9780199858064, 0199858063, 2019.
- [6] K.C. Santosh, N. Das, S. Ghosh, *Deep Learning Models for Medical Imaging*, ISBN: 9780128236505, 0128236507, Elsevier Science, 2021.
- [7] J. Latif, C. Xiao, S. Tu, S. U. Rehman, A. Imran and A. Bilal, "Implementation and Use of Disease Diagnosis Systems for Electronic Medical Records Based on Machine Learning: A Complete Review," in *IEEE Access*, vol. 8, pp. 150489-150513, 2020, doi: 10.1109/ACCESS.2020.3016782.
- [8] M. R. Ahmed, Y. Zhang, Z. Feng, B. Lo, O. T. Inan and H. Liao, "Neuroimaging and Machine Learning for Dementia Diagnosis: Recent Advancements and Future Prospects," in *IEEE Reviews in Biomedical Engineering*, vol. 12, pp. 19-33, 2019, doi: 10.1109/RBME.2018.2886237.
- [9] G. Joo, Y. Song, H. Im and J. Park, "Clinical Implication of Machine Learning in Predicting the Occurrence of Cardiovascular Disease Using Big Data (Nationwide Cohort Data in Korea)," in *IEEE Access*, vol. 8, pp. 157643-157653, 2020, doi: 10.1109/ACCESS.2020.3015757.
- [10] G. Joo, Y. Song, H. Im and J. Park, "Clinical Implication of Machine Learning in Predicting the Occurrence of Cardiovascular Disease Using Big Data (Nationwide Cohort Data in Korea)," in *IEEE Access*, vol. 8, pp. 157643-157653, 2020, doi: 10.1109/ACCESS.2020.3015757.
- [11] Z. Lin et al., "AANet: Adaptive Attention Network for COVID-19 Detection From Chest X-Ray Images," in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 11, pp. 4781-4792, Nov. 2021, doi: 10.1109/TNNLS.2021.3114747.
- [12] J. -X. Wu, P. -Y. Chen, C. -M. Li, Y. -C. Kuo, N. -S. Pai and C. -H. Lin, "Multilayer Fractional-Order Machine Vision Classifier for Rapid Typical Lung Diseases Screening on Digital Chest X-Ray Images," in *IEEE Access*, vol. 8, pp. 105886-105902, 2020, doi: 10.1109/ACCESS.2020.3000186.
- [13] M. Eslami, S. Tabarestani, S. Albarqouni, E. Adeli, N. Navab and M. Adjouadi, "Image-to-Images Translation for Multitask Organ Segmentation and Bone Suppression in Chest X-Ray Radiography," in *IEEE Transactions on Medical Imaging*, vol. 39, no. 7, pp. 2553-2565, July 2020, doi: 10.1109/TMI.2020.2974159.
- [14] J. An, Q. Cai, Z. Qu and Z. Gao, "COVID-19 Screening in Chest X-Ray Images Using Lung Region Priors," in *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 11, pp. 4119-4127, Nov. 2021, doi: 10.1109/JBHI.2021.3104629.
- [15] J. D. Arias-Londoño, J. A. Gómez-García, L. Moro-Velázquez and J. I. Godino-Llorente, "Artificial Intelligence Applied to Chest X-Ray

- Images for the Automatic Detection of COVID-19. A Thoughtful Evaluation Approach," in IEEE Access, vol. 8, pp. 226811-226827, 2020, doi: 10.1109/ACCESS.2020.3044858.
- [16] C. Li, D. Zhang, S. Du and Z. Tian, "Deformation and Refined Features Based Lesion Detection on Chest X-Ray," in IEEE Access, vol. 8, pp. 14675-14689, 2020, doi: 10.1109/ACCESS.2020.2963926.
- [17] K. Wang, X. Zhang, S. Huang, F. Chen, X. Zhang and L. Huangfu, "Learning to Recognize Thoracic Disease in Chest X-Rays With Knowledge-Guided Deep Zoom Neural Networks," in IEEE Access, vol. 8, pp. 159790-159805, 2020, doi: 10.1109/ACCESS.2020.3020579.
- [18] T. Rahman et al., "Reliable Tuberculosis Detection Using Chest X-Ray With Deep Learning, Segmentation and Visualization," in IEEE Access, vol. 8, pp. 191586-191601, 2020, doi: 10.1109/ACCESS.2020.3031384.
- [19] F. Munawar, S. Azmat, T. Iqbal, C. Grönlund and H. Ali, "Segmentation of Lungs in Chest X-Ray Image Using Generative Adversarial Networks," in IEEE Access, vol. 8, pp. 153535-153545, 2020, doi: 10.1109/ACCESS.2020.3017915.
- [20] C. -M. Kim, E. J. Hong and R. C. Park, "Chest X-Ray Outlier Detection Model Using Dimension Reduction and Edge Detection," in IEEE Access, vol. 9, pp. 86096-86106, 2021, doi: 10.1109/ACCESS.2021.3086103.
- [21] N. Paluru et al., "Anam-Net: Anamorphic Depth Embedding-Based Lightweight CNN for Segmentation of Anomalies in COVID-19 Chest CT Images," in IEEE Transactions on Neural Networks and Learning Systems, vol. 32, no. 3, pp. 932-946, March 2021, doi: 10.1109/TNNLS.2021.3054746.
- [22] H. Salehinejad, E. Colak, T. Dowdell, J. Barfett and S. Valaee, "Synthesizing Chest X-Ray Pathology for Training Deep Convolutional Neural Networks," in IEEE Transactions on Medical Imaging, vol. 38, no. 5, pp. 1197-1206, May 2019, doi: 10.1109/TMI.2018.2881415.
- [23] O. D. T. Catalá et al., "Bias Analysis on Public X-Ray Image Datasets of Pneumonia and COVID-19 Patients," in IEEE Access, vol. 9, pp. 42370-42383, 2021, doi: 10.1109/ACCESS.2021.3065456.
- [24] P. Zhang, Y. Zhong, Y. Deng, X. Tang and X. Li, "Drr4covid: Learning Automated COVID-19 Infection Segmentation From Digitally Reconstructed Radiographs," in IEEE Access, vol. 8, pp. 207736-207757, 2020, doi: 10.1109/ACCESS.2020.3038279.
- [25] C. Zhou, J. Song, S. Zhou, Z. Zhang and J. Xing, "COVID-19 Detection Based on Image Regrouping and Resnet-SVM Using Chest X-Ray Images," in IEEE Access, vol. 9, pp. 81902-81912, 2021, doi: 10.1109/ACCESS.2021.3086229.
- [26] S. Z. Y. Zaidi, M. U. Akram, A. Jameel and N. S. Alghamdi, "Lung Segmentation-Based Pulmonary Disease Classification Using Deep Neural Networks," in IEEE Access, vol. 9, pp. 125202-125214, 2021, doi: 10.1109/ACCESS.2021.3110904.
- [27] Y. -H. Wu et al., "JCS: An Explainable COVID-19 Diagnosis System by Joint Classification and Segmentation," in IEEE Transactions on Image Processing, vol. 30, pp. 3113-3126, 2021, doi: 10.1109/TIP.2021.3058783.
- [28] F. Yan, X. Huang, Y. Yao, M. Lu and M. Li, "Combining LSTM and DenseNet for Automatic Annotation and Classification of Chest X-Ray Images," in IEEE Access, vol. 7, pp. 74181-74189, 2019, doi: 10.1109/ACCESS.2019.2920397.
- [29] D. -P. Fan et al., "Inf-Net: Automatic COVID-19 Lung Infection Segmentation From CT Images," in IEEE Transactions on Medical Imaging, vol. 39, no. 8, pp. 2626-2637, Aug. 2020, doi: 10.1109/TMI.2020.2996645.
- [30] J. Lian et al., "A Structure-Aware Relation Network for Thoracic Diseases Detection and Segmentation," in IEEE Transactions on Medical Imaging, vol. 40, no. 8, pp. 2042-2052, Aug. 2021, doi: 10.1109/TMI.2021.3070847.
- [31] Kaggle: Your Home for Data Science. (n.d.). <https://www.kaggle.com/nikhilpandey360/lung-segmentation-from-chest-x-ray-dataset>.
- [32] X. Ou et al., "Moving Object Detection Method via ResNet-18 With Encoder-Decoder Structure in Complex Scenes," in IEEE Access, vol. 7, pp. 108152-108160, 2019, doi: 10.1109/ACCESS.2019.2931922.
- [33] X. Yu, C. Kang, D. S. Guttery, S. Kadry, Y. Chen and Y. -D. Zhang, "ResNet-SCDA-50 for Breast Abnormality Classification," in IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 18, no. 1, pp. 94-102, 1 Jan.-Feb. 2021, doi: 10.1109/TCBB.2020.2986544.
- [34] Zhang, Q. A novel ResNet101 model based on dense dilated convolution for image classification. SN Appl. Sci. 4, 9 (2022). <https://doi.org/10.1007/s42452-021-04897-7>.
- [35] A. F. Ibrahim, S. P. Ristiawanto, C. Setianingsih and B. Irawan, "Micro-Expression Recognition Using VGG19 Convolutional Neural Network Architecture And Random Forest," 2021 4th International Symposium on Agents, Multi-Agent Systems and Robotics (ISAMSR), 2021, pp. 150- 156, doi: 10.1109/ISAMSR53229.2021.9567872.
- [36] K. Adam, I. I. Mohamed and Y. Ibrahim, "A Selective Mitigation Technique of Soft Errors for DNN Models Used in Healthcare Applications: DenseNet201 Case Study," in IEEE Access, vol. 9, pp. 65803-65823, 2021, doi: 10.1109/ACCESS.2021.3076716.
- [37] B. Qiang et al., "SqueezeNet and Fusion Network-Based Accurate Fast Fully Convolutional Network for Hand Detection and Gesture Recognition," in IEEE Access, vol. 9, pp. 77661-77674, 2021, doi: 10.1109/ACCESS.2021.3079337.
- [38] K. Kadam, S. Ahirrao, K. Kotecha and S. Sahu, "Detection and Localization of Multiple Image Splicing Using MobileNet V1," in IEEE Access, vol. 9, pp. 162499-162519, 2021, doi: 10.1109/ACCESS.2021.3130342.