

MULTIMODAL FEATURE AND TRANSFER LEARNING IN DEEP ENSEMBLE MODEL FOR LUNG DISEASE PREDICTION

N.Dhivya ^{#1}, Dr.P.Sharmila ^{#2}

Ph.D Scholar ^{#1}, Professor ^{#2}

dhivyarubi@gmail.com, mavish2009@gmail.com

^{#1}PG and Research Department of Computer Science

Navarasam Arts and Science College for Women, Arachalur, Erode.

^{#2}Department of Computer Science,

KPR College of Arts Science and Research, Arasur, Coimbatore.

ABSTRACT

Lung disease is common throughout the world. These include chronic obstructive pulmonary disease, pneumonia, asthma, tuberculosis, fibrosis, etc. Timely diagnosis of lung disease is essential. Many image processing and machine learning (ML) models have been developed to make more accurate with less time and effort. But still, the degraded accuracy performance was resulted in these systems. In recent days, Deep Learning (DL) model plays an important role in classifying the lung diseases for the early prediction process. Amongst, an efficient Convolutional Neural Network (CNN)-based lung disease detection system is developed with additional layers to classify the segregated lung sections into various lung diseases using Chest X-Ray (CXR) images. However, this model results in epistemic uncertainty which effects the performance of DL models employed for lung disease diagnosis. Hence, in this paper, a multi-modal approach called Ensemble Deep Lung Disease Predictor (EDepLDP) framework is proposed to solve the epistemic uncertainty issue and develops a reliable solution for rapid detection of various diseases using CXR and Computerized Tomography (CT) images. Initially, the collected images are segmented using U-Net model to get enhanced lung Region of Interest (ROIs). Then, InceptionResNetV2 and Xception are used to hierarchically extract informative features from segmented CXR images and discriminative features from segmented CT images respectively. The extracted deep features are fed into the softmax layer of conGRU-LSTM to perform the classification task. Moreover, the TL model is developed to learn the weight for the InceptionResNetV2, Xception and conGRU-LSTM which is obtained from the pre-trained Efficient-Net model. Also, the domain adaptation strategy is a subset of TL model which mainly addresses the situation where different but related datasets for a common learning task. This adaption strategy reduces the domain shift or data distribution using Maximum Mean Discrepancy (MMD) for the efficient classification of various lung diseases. The test outcomes reveal that the EDepLDP model accomplishes an overall accuracy of 92% and 92% on the collected CXR and CT images which is contrasted with the classical CNN models.

Keywords: Chest X-Rays, Computerized Tomography, Epistemic Uncertainty, Transfer Learning, Normalized Feature Inputs

1. INTRODUCTION

Lungs play a vital role in the human system, which performs expansion and relaxation to bring in oxygen and take out carbon dioxide. Lung diseases are respiratory diseases that affect the various organs and tissues associated with breathing, leading to airway diseases, lung tissue diseases, and lung circulation diseases [1, 2]. Some of the lung diseases like common cold and influenza cause mild discomfort and hindrance while others like lower respiratory infections (pneumonia), tuberculosis (TB), Chronic Obstructive Pulmonary Disease (COPD), lung cancer, COVID-19 are leading causes of respiratory morbidity and mortality among adults pneumonia, tuberculosis and lung cancer are life-threatening and cause severe acute respiratory problems [3]. The diagnosis of these lung diseases from clinical analysis are challenging and lacks sensitivity, preventing patients from making a prompt diagnosis.

As a result, healthcare professionals opted for the diagnostic imaging modalities like CT imaging, CXR, Positron Emission Tomography (PET), and Magnetic Resonance Imaging (MRI) [4, 5]. Even though these state-of-the-art imaging techniques yield precise, reliable, and impermeable results, but their accessibility and availability are constrained in underdeveloped countries due to the lack of a strong healthcare system and patient comfort. To solve these issues, Computer-Aided Diagnostic (CAD) systems is developed which efficiently automate the medical image interpretation process and minimize the human effort [6, 7]. CAD systems can help radiologists by doing the trivial processing and presenting the information in a meaningful way so that, the radiologist can make more accurate decisions by spending less amount of time and energy [8]. However, CAD diagnostics are inaccurate because the imaging films of each patient have different characteristics and anatomic structures like body fat or distorted bones and finds difficult to analyses vast amounts of medical image data.

In recent days, DL models plays an important role in the medical field especially for medical image analysis [9, 10]. In DL model, there are mainly three steps involved in DL model like image preprocessing, training and classification for lung disease prediction. Lung disease detection generally deals with classifying an image into healthy lungs or disease-infected lungs. The lung disease classifier sometimes known as a model which is obtained via training. Training is the process in which a neural network learns to recognize a class of images. Using DL, it is possible to train a model that can classify images into their respective class labels. Therefore, to apply DL for lung disease detection, the first step is to gather images of lungs with the disease to be classified. The second step is to train the neural network until it is able to recognize the diseases. The final step is to classify new images. The classified new images unseen by the model before are shown to the model and the model predicts the class of those images [11].

By utilizing the idea of DL models, various literature have been suggested for the automated identification of pulmonary diseases by utilizing various imaging modalities. For instance, automated DL -based lung disorder diagnosis model [12] for CXR scans is developed. In this model, the healthy and infected CXR classification is performed on the entire CRX image. For the infected CXR images, the classification model is trained on the segmented lung region. Finally, these segmented lung region images are used to train the model for the classification of pulmonary diseases. The gathering of the manual lung masks and using them to segment the lung region on the CXR. Finally, the DL architecture are exploited by proposing a custom CNN architecture with additional layers and modified hyperparameters to meet the required results. The input CXR is examined for healthy or infected at the surface level and the infected images

are further processed for class level label classification. However, this model suffers from epistemic uncertainty which has a major impact on the performance of DL models used to diagnose lung diseases. The ambiguity in CXR images must be successfully resolved by increasing the architecture of DL and the learning of multimodal features.

Hence, in this paper, EDepLDP model is proposed to solve the epistemic uncertainty issue and develops a reliable solution for rapid detection of various diseases using CXR and CT images. Initially, image segmentation is performed using the U-Net model to obtain an enhanced lung ROIs. Then, the InceptionResNetV2 model is applied to extract informative features from segmented CXR images and Xception are used to extract discriminative features from segmented CT images. For the classification task, the retrieved deep features are given as input to the softmax layer of conGRU-LSTM. This conGRU-LSTM network is more suitable for learning the sequence patterns and provides effective classification task. Moreover, the TL model is developed to learn the weight for the InceptionResNetV2, Xception and conGRU-LSTM which is obtained from the pre-trained Efficient-Net model. Also, the domain adaptation strategy is a subset of TL model which mainly addresses the situation where different but related datasets for a common learning task. This adaption strategy reduces the domain shift or data distribution using Maximum Mean Discrepancy (MMD) for the efficient classification of various lung diseases.

The remaining sections of this manuscript are prepared as follows: Section II studies the work related to this research. Section III explains the proposed EDepLDP model for lung disease classification. Section IV illustrates its model's performance effectiveness. Section V summarizes the whole work and suggests future enhancement.

2. LITERATURE SURVEY

Bharati et al. [13] introduced a novel hybrid DL model by integrating VGG, data augmentation and Spatial Transformer Network (STN) with CNN called VDSNet to recognize lung disorders from CXR scans. The new model was applied to NIH CXR dataset collected from Kaggle repository. Full and sample versions of the dataset were considered. For both full and sample datasets, VDSNet outperforms existing methods in terms of a number of metrics including precision, recall and validation accuracy. However, the validated accuracy was slightly reduced while processing with full version of dataset.

Hashmi et al. [14] presented a new method depending on the weighted classifier for recognizing pneumonia from CXR scans. In this method, the weighted predictions from the classical DL structures were combined. This approach was a supervised learning approach in which the network predicts the result based on the quality of the dataset used. Also, TL model was utilized to modify the DL and increase detection accuracy. Partial data augmentation techniques were employed to increase the training dataset in a balanced way to improve the model's efficiency. Conversely, the amount of inaccurately labeled images was not limited.

Wang et al. [15] developed a parallel Channel Attention feature Fusion (PCAF) scheme called MCFF-Net according to the characteristics of CXR images. Based on this module, a new CNN structure is proposed to classify CXR images in order to diagnose and detect COVID-19 cases. Also, the recognition efficiency was enhanced by adopting 3 classifiers such as fully connected (FC), GAP-FC, and Conv1-GAP. MCFF-Net was used to perform a 4-class classification experiment on a dataset containing four types of image of COVID-19, normal, bacterial

pneumonia, and viral pneumonia, with excellent performance. However, this model provides lower performances on larger dataset.

Vieira et al. [16] presented a lung disorder recognition system based on the image resizing scheme with the maximum window function, which preserves anatomical structures of the chest. First, the CXR scans were acquired. Then, a pre-processing methodology was developed which includes a new image resizing method with the maximum window function that preserves anatomical structures of the CXR images. An evaluation methodology was defined that comprises different CNN architectures combined with TL and fine-tuning of all layers of CNN to categorize the scans into different types of pneumonia. However, this model has a slower convergence rate.

Cao & Zhao [17] described an automatic lung segmentation algorithm on chest x-ray images based on fusion variational auto-encoder and three-terminal attention mechanism. In this model, a constructed VAE was introduced in every layer of the decoder-encoder to capture high-level semantic characteristics like the symmetrical correlation between the left and right thoraxes. Also, a 3-terminal attention strategy was applied, which utilizes the channel and spatial attention units to automatically localize and identify the target lung region. However, this model finds difficulty in segmenting the local areas with a large number of opacities.

Tasci et al. [18] introduced a Voting-based ensemble DL (VEDL) method focusing on image augmentation and preprocessing variations for tuberculosis detection. The ensemble deep learning method selects the best pipelines employing different preprocessing, augmentation alternatives, and CNN models for tuberculosis detection. The voting-based (i.e., soft voting and Bayesian optimization-based weighted voting) ensemble of various fine-tuned CNN models (i.e., Inception V3 and Xception) with the preprocessing (e.g., Contrast-limited adaptive histogram equalization (CLAHE)) and image data augmentation (translation, rotation, and scaling) variations was used in this process for efficient TB detection. But, it needs more memory for training.

Zhang et al. [19] designed a VGG framework with fewer layers for detecting pneumonia from CXR scans. First, the CXR scans were collected and pre-processed by the Dynamic Histogram Equalization (DHE) technique. Then, VGG-based CNN model was developed to extract the features from original images or previous feature maps which contained only six layers combining ReLU activation function, drop operation, and max-pooling layers. This model effectively captures the characteristics of CXR scans and learns them to recognize whether a person suffers from pneumonia or not. However, this model results in overfitting issues.

Li et al. [20] introduced an innovative approach for earlier identification of COPD using a graph convolution network (GCN). This approach leverages minimal chest CT image information with inadequate labeling. The primary goal is to generate a graph utilizing arbitrarily generated areas of interest (i.e., ROIs) from the separated lung parenchyma and then input it into the GCN model for COPD identification. However, this method's classification accuracy was lower compared to other classical approaches.

Junayed et al. [21] constructed an efficient End-to-End Deep Neural Network (E2E-DNN) for interstitial lung disease (ILD) recognition and classification using lung CT image patches. The collected data was pre-processed to remove noise and unstructured patterns. A new architecture was developed to classify captures of low-level textural characteristics of lung tissue. This architecture consists of various convolutional layers with various filters, ReLU activation,

batch normalization, max-pooling, and flatten, dense, and dropout layers. Categorical crossentropy was employed as a loss function for training purposes, and adam was also used to optimize the model. However, this model requires a considerable amount of time to train.

Ravi et al. [22] presented a multichannel EfficientNet deep learning-based stacking ensemble approach for lung disease detection using chest X-ray images. In the model, different multichannel EfficientNet models were used to extract the features from the collected dataset. The features from EfficientNet models are fused together. Next, the fused features are passed into more than one non-linear fully connected layer. Finally, the features passed into a stacked ensemble learning classifier for lung disease detection. The stacked ensemble learning classifier contains random forest and Support Vector Machine (SVM) in the first stage and logistic regression in the second stage for lung disease detection. However, this model result lower generalizability on unseen data samples from lung diseases.

Kim et al. [23] introduced a DL model called EfficientNet v2-M for multi-class lung diseases' classification on CXR images. The constructed EfficientNet v2-M model utilizes the pre-trained weights of ImageNet to classify lung diseases on CXR images to improve the efficiency and accuracy of CADs' diagnostic performance. Initially, the data were augmented to increase the number of samples and variance diversities. Then, they were directly inputted into an EfficientNet v2-M model to extract their meaningful features in identifying lung disease categories. But, this model loses a substantial amount of useful data, resulting in overfitting difficulties.

3. PROPOSED METHODOLOGY

In this section, EDepLDP framework is briefly illustrated. Initially, the CXR and CT images are gathered. Then, the gathered images are segmented using U-Net architecture. The segmented images are given as input to the InceptionResNetV2 and Xception are used to determine an informative features from segmented CXR images and discriminative features from segmented CT images respectively. The extracted deep features are fed into the softmax layer of conGRU-LSTM to perform the classification task. By utilizing the transfer learning (TL) model, acquired the learned weight for InceptionResNetV2, Xception, and conGRU-LSTM from a previously-trained Efficient-Net for the efficient classification of different lung diseases. The figure 1 depicts the complete structure of the proposed model.

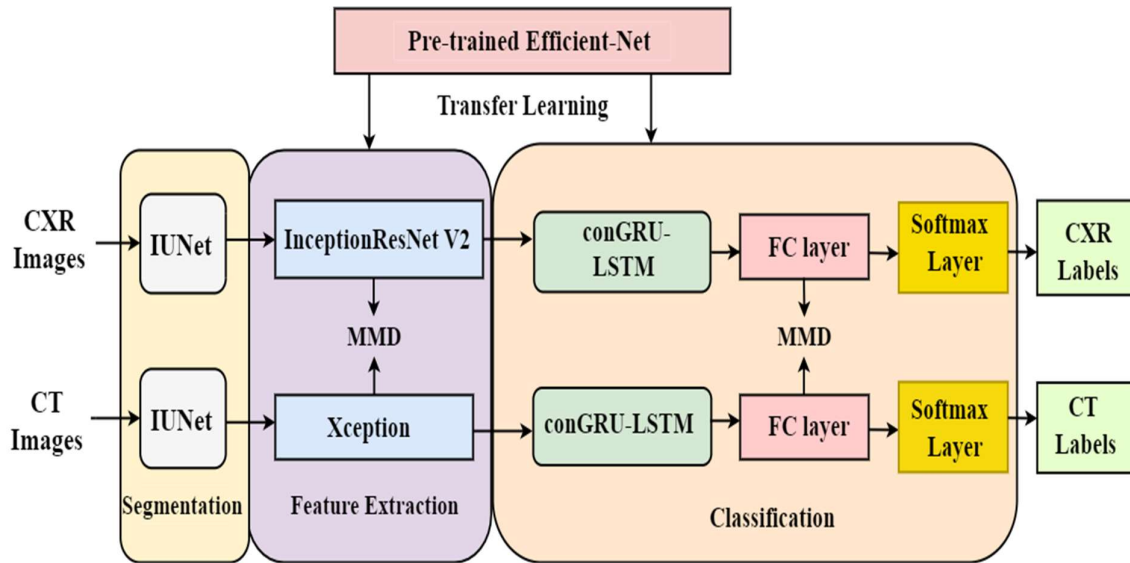



Figure 1 Block Structure of the Proposed Model





3.1 Dataset Description

In this study, two benchmark databases are considered, including:

1. CXR data: CXR dataset [24] consists of 112,120 frontal-view X-ray images of 30,805 distinct patients with fourteen disease image labels text-mined from the corresponding radiological reports using natural language processing. Each image may have multiple labels. This database includes fourteen common thoracic pathologies including Atelectasis, Consolidation, Infiltration, Pneumothorax, Edema, Emphysema, Fibrosis, Effusion, Pneumonia, Pleural thickness, Cardiomegaly, Nodule, Mass, and Hernia. The information of COVID-19 and Non-Covid is obtained from [25], which contains 25 images of COVID-19 and 75 images of non-Covid chest X-ray cases. It is essential to emphasize that these Non-Covid (normal) situations might consist of other unhealthy conditions, such as bacterial or viral infections, chronic obstructive pulmonary disease, and even a combination of two or more. For the experimental purposes, only five pathologies i.e., atelectasis, infiltration, pneumonia along with COVID-19 and Non-Covid are considered in this framework. The total CXR images obtained from the given dataset for the performance evaluation are listed in table 1.

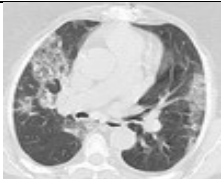
Table 1 Observed CXR images


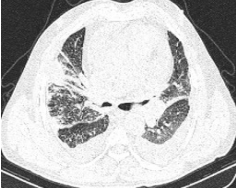

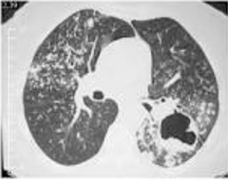
Lung Disease Categorizes	CXR Image	Number of images Considered
Covid		1345

Normal		1345
Pneumonia		1443
Atelectasis		290
Infiltrate		270

2. CT data: As like CXR data, the CT images also opted for five diseases categories like atelectasis, infiltration, pneumonia along with COVID-19 and Non-Covid. In this study, different open public portals are analyzed to collect the CT data for the experiment purposes. The lung atelectasis images is taken form [26], The Covid and Non-Covid (Normal) images are gathered from [27], the viral pneumonia are taken from [28] and infiltration is obtained from [29]. The total CT images obtained from the given dataset for the performance evaluation are listed in table 2.

Table 2 Observed CT images

Lung Disease Categorizes	CT Image	Number of images Considered
Covid		1002

Normal		984
Pneumonia		1762
Atelectasis		310
Infiltrate		260

3.2 U-Net based Segmentation task on CXR and CT images

In this model, U-Net architecture model is used for the segmentation purposes, it composed of contraction (encoder) and expansion (decoder) paths. Where the contraction path is built up with repeated 3×3 convolution operations that are followed by the Relu Activation function and 2×2 max pooling operation for downsampling with the stride of 2. The number of feature channels are doubled by each downsampling operation. At the expansion path up-convolution operation is applied that utilizes 2×2 transposed convolutions to increase the resolution. Moreover, the details that were lost in the contraction path after the downsampling are also recovered by using skipping operations which allows the network to pass the feature maps from the encoder to the decoder. The figure 2 depicts the U-Net structure for segmenting CXR and CT images.

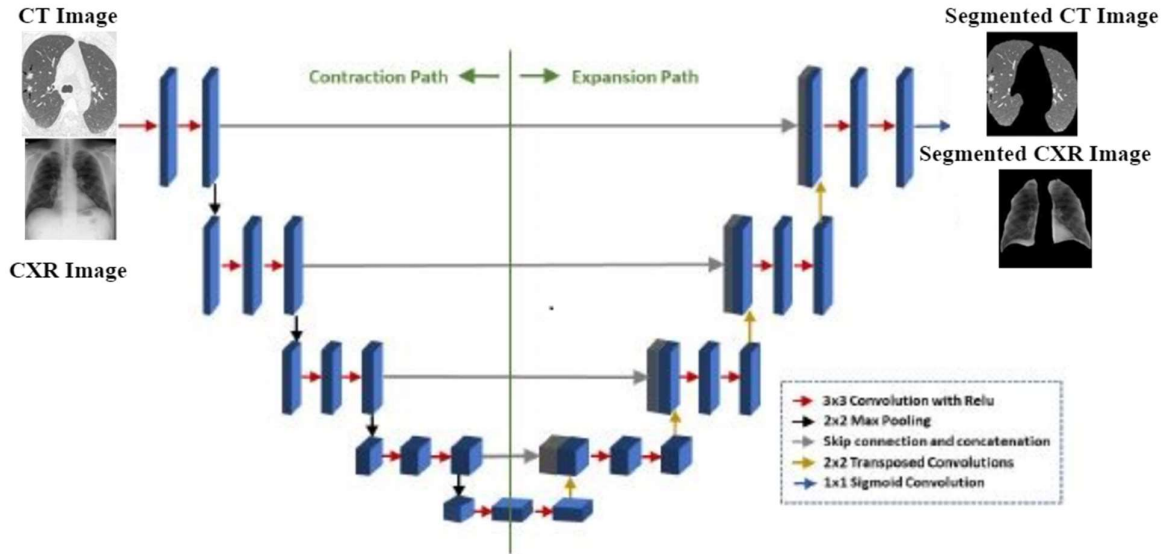


Figure 2 U-Net structure based segmented CXR and CT images

U-Net structure is trained on the manual lung masks of the CXR and CT image dataset to enhance the lung ROIs. Since the performance of a CNN architecture while training is highly dependent upon the initialization of the learnable parameters. So, the weights of U-Net architecture are trained on the collected dataset and obtained as initial weights in order to produce the optimal segmentation results on CXR and CT images. These initial weights are then multiplied with the input data with an additional bias value is depicted in Eq. (1)

$$A_j^g = \sum_{h=1}^H w_{jk}^h u_h + y_j^k, \quad h = 1, 2, \dots, H \quad (1)$$

In above Eq. (1), w^k represents the weights of the trained U-Net architecture, u depicts the input variable, y^k is the value add as a bias variable. The super and subscripts, h , and j donate hidden layer of the model, the number of input variables, and the neurons respectively. The resultant A_j^g is given as input to the activation function which is responsible for activating or deactivating a neuron.

The model is further optimized using binary cross-entropy (BCE) loss function that calculates the loss between the segmented masks and the manual reference masks gathered for each sample image i.e., CXR and CT which is formulated as in equation 2.

$$BCE = -\frac{1}{N} \sum_{j=1}^n w_j \log w_j' + (1 - w_j) \log (1 - w_j') \quad (2)$$

From Eq. (2), N represents the total number of images used for training, w_j donates the manual reference masks for each sample image i , and w_j' donates the segmented masks (image) generated by the model.

3.3 Feature Extraction by InceptionResNetV2 and XceptionV3

In this model, the segmented images from the U-Net model are given as input to the different CNN models like InceptionResNetV2 and XceptionV3 are used to hierarchically extract informative and discriminative features from the CXR and CT images respectively. The feature extraction processes through passing the segmented CXR images in every layers of the InceptionResNetV2 is illustrated in Table 3 and for CT image in layers of the XceptionV3 structure is illustrated in Table 4.

Table 3 InceptionResNetV2 structure for segmented CXR images

Layer	Patch Size	Input Size
Conv	$224 \times 224 \times 3$	3×3
Conv	$111 \times 111 \times 32$	3×3
Filter contact	$109 \times 109 \times 64$	$3 \times 3 \text{ pool} + 3 \times 3 \text{ conv}$
Filter contact	$54 \times 54 \times 160$	$1 \times 1 \text{ conv}, 3 \times 3 \text{ conv} + 1 \times 1 \text{ conv}, 7 \times 1 \text{ conv}, 1 \times 7 \text{ conv}, 3 \times 3 \text{ conv}$
Filter contact	$52 \times 52 \times 128$	$3 \times 3 \text{ conv} + \text{max pool}$
Inception-ResNet-A $\times 10$	$26 \times 26 \times 256$	—
Reduction-A	$26 \times 26 \times 256$	—
Inception-ResNet-B $\times 20$	$13 \times 13 \times 768$	—
Reduction-B	$13 \times 13 \times 768$	—
Inception-ResNet-C $\times 10$	$6 \times 6 \times 1534$	—
Average pooling	$6 \times 6 \times 1534$	6×6
Dropout	$1 \times 1 \times 1534$	$Keep = 0.5$
Fully Connected (FC)	1534	1534×1000
FC	1000	Logits
Softmax	500	<i>Classifier (3 classes)</i>

The InceptionResNetV2 is the combination of residual connection and the Inception architecture. This structure takes the advantages of residual network and retains the unique characteristics of the multi-convolutional core of the Inception network and these residual connections are implicit approaches for training very deep architectures. Generally, Inception-ResNetV2 architecture represented in all layers are established before the FC layer. The Inception-ResNetV2 model contains three basic types of inception modules like Inception-ResNet-A, Inception-ResNet-B, and Inception-ResNet-C. These modules are responsible for both reducing the number of parameters small Conv layers (e.g., $1 \times 7, 7 \times 1$) and generating the discriminatory features. Each module is self-possessed of several Conv and pool layers.

Inception-ResNetV2 also contains two types of reduction modules, which are responsible for reducing the image size. In this framework, Inception-ResNetV2 model has a default input size $299 \times 299 \times 3$; thus, it is resized to 224×224 during training. The structure of inceptionResNetV2 is illustrated in table 1, which is used to extract the informative features from the CXR images.

Table 4 Xception structure for segmented CT images

Layer	Outcome dimension	Patch size/Stride
Convolutional layer (<i>conv</i>)	$149 \times 149 \times 32$	$3 \times 3 \times 32/2$
	$147 \times 147 \times 64$	$3 \times 3 \times 64$
Separable convolution	$147 \times 147 \times 128$	$3 \times 3 \times 128/1$
Max pooling	$74 \times 74 \times 128$	$3 \times 3/2$
Separable convolution	$74 \times 74 \times 256$	$3 \times 3 \times 256/1$
Max pooling	$37 \times 37 \times 256$	$3 \times 3/2$
Separable convolution	$37 \times 37 \times 728$	$3 \times 3 \times 728/1$
Max pooling	$19 \times 19 \times 728$	$3 \times 3/2$
Separable convolution	$19 \times 19 \times 728$	$3 \times 3 \times 728/1$
Separable convolution	$19 \times 19 \times 728$	$3 \times 3 \times 728/1$
	$19 \times 19 \times 1024$	$3 \times 3 \times 1024/1$
Max pooling	$10 \times 10 \times 1024$	$3 \times 3/2$
Separable convolution	$10 \times 10 \times 1536$	$3 \times 3 \times 1536/1$
	$10 \times 10 \times 2048$	$3 \times 3 \times 2048/1$
Mean pooling	$1 \times 1 \times 2048$	
FC	2×2048	Logits
Softmax	1000	Classifier

The Xception structure is extended extended form of the Inception layout in which the basic Inception units have been substituted with the adapted depthwise discrete convolutions. It is the pointwise convolution preceded by the depthwise convolution. Additionally, this Xception architecture contains the residual/skip links analogous to the ResNet architecture. It may be partitioned into 3 units: input stream, core stream and output stream. The core stream unit is continued 8 times. In table 2, the pre-trained Xception structure applied in this methodology is

depicted. By using this Xception structure, the discriminative features are extracted from the segmented CT images.

3.4 Transferring learned weights using pretrained EfficientNet with MMD

In this model, the pretrained EfficientNet models is used as a transfer learning model with the aim to achieve better performances for lung disease classification using CXR and CT images. The EfficientNet group consists of 8 models between EfficientNet B0 and EfficientNet B7, as the model number grows, the number of calculated parameters does not increase much, while accuracy increases noticeably. The main building block for EfficientNet is the inverted bottleneck MBConv, it is a blocks consist of a layer that first expands and then compresses the channels, so direct connections are used between bottlenecks that connect much fewer channels than expansion layers. The EfficientNet model efficiently improves the resolution of image data (i.e., segmented CXR and CT images) by uniformly scaling depth, width, and resolution while scaling down the model. The first step in the compound scaling method is to search for a grid to find the relationship between the different scaling dimensions of the baseline network under a fixed resource constraint. In this way, a suitable scaling factor for depth, width and resolution dimensions is determined. These coefficients are then applied to scale the baseline network to the desired target network [30]. The width, depth and resolution of the network are balanced by calculating the appropriate composite ratio coefficient ϑ which is formulated in Eq. (3),

$$\begin{aligned}
 \text{Depth } d &= \alpha^\vartheta \\
 \text{Width } w &= \beta^\vartheta \\
 \text{Resolution } r &= \gamma^\vartheta
 \end{aligned}
 \tag{3}$$

$\alpha \geq 1, \beta \geq 1, \gamma \geq 1$

Where $\alpha \geq 1, \beta \geq 1, \gamma \geq 1$ can be used for scaling network width, depth and resolution coefficient, given value ϕ can be used to determine the amount of effective resources extension model, constant α, β, γ is used to allocate these resources to the network depth, width and resolution of three dimensions. Since the cost of computing in convolution networks is largely due to convolution operations and scaling convolution network as given in Eq. (3) increases the Floating Point Operations Per Second (FLOPS) of the network by approximately $(\alpha, \beta, \gamma)^\vartheta$ in total [30]. The structure of EfficientNet is given in table 5.

Table 5 EfficientNet structure

Stage i	Layer	Resolution	Channels	Layers
1	(Conv 3×3)	224×224	32	1
2	MBConv1, $k3 \times 3$	112×112	16	1
3	MBConv6, $k3 \times 3$	112×112	24	2
4	MBConv6, $k5 \times 5$	56×56	40	2
5	MBConv6, $k3 \times 3$	28×28	80	3
6	MBConv6, $k5 \times 5$	28×28	112	3

7	<i>MBCConv6, k5 × 5</i>	14 × 14	192	4
8	<i>MBCConv6, k3 × 3</i>	7 × 7	320	1
9	<i>Conv1; Pooling; FC</i>	7 × 7	1280	1

In the EDepLDP framework, the pre-trained Efficient-Net model is introduced for transferring the learned weights to the feature extraction models like InceptionResNetV2, XceptionV3. The pretrained Efficient-Net model models has rich features and contain deeper depth or width or more input image resolution which effectively fine-tunes the weights of the model on medical image classification, reduce the training time, provides faster convergence rate, and achieve optimal performances in detecting patient’s data samples of CXR and CT as either lung disease or normal. Since, EfficientNet model prefers to learn domain specific features on top layers, the main challenge is to reduce the shift between two domain (Source and target) distributions on these layers. The framework develops a domain adaptation learning method in Efficient Net model to conduct transfer learning. EfficientNet combines domain adaptation and feature learning in a training process, so that the features of domain invariance can be predicted. Then, the proposed transfer learning-based model trained by source domain site data can be used to assist in predicting target site data without degradation of the prediction performance due to domain drift.

While transferring the learned weights, it is necessary to measure the distribution distance between source domain sites and target domain sites, and select the source domain sites closest to the target domain sites. The maximum mean discrepancy (MMD) in the regenerative kernel Hilbert space is an effective method for estimating the distance between two distributions. Based on two distributed samples, the average difference between two samples corresponding to f can be obtained by subtracting the function mean of different samples, and MMD is the maximum value of the average difference. For the convenience of calculation, the square form of MMD is generally adopted. The process of using MMD to estimate the difference between two domains is constructed as follows.

The source domain site data in a given source domain is denoted as:

$$D_s = (x_1, x_2, \dots, x_m) \tag{4}$$

where x represents the source domain site data and m represents the source domain site data number. The target site data in the target domain is denoted as:

$$D_t = (y_1, y_2, \dots, y_n) \tag{5}$$

Where y represents the target domain site data and n represents the target domain site data number. The nonlinear mapping function in the Hilbert space (H) of the regenerative kernel is denoted as μ . Then the squared form of MMD is defined as follows:

$$MMD_H^2(D_S, D_T) = \left\| \frac{1}{n} \sum_{i=1}^m \mu(x_i) - \frac{1}{n} \sum_{i=1}^n \mu(y_i) \right\|^2 \tag{6}$$

The difference in distribution between two domains is the distance between the two data distributions. The smaller the MMD value, the closer the two domains are. MMD is mainly used in transfer learning to select the source domain site that is most suitable for migration to the target domain site by calculating the similarity between the source domain and the target domain based on MMD.

3.5 Transfer Learning Based Classification Model

The extracted deep features from the InceptionResNetV2, Xception and Efficient-Net are fed into the conGRU-LSTM model to accurately classify the various lung disease. In the proposed conGRU-LSTM model, transfer learning is used for knowledge exchange between features and classes relation among X-ray and CT images and improve the target task of lung disease classification. The transformed features conGRU-LSTM model have been trained in the source domain, while their parameter will be frozen during the transfer learning. The conGRU-LSTM replaces the matrix multiplication with a convolution operation for each gate in the LSTM cell. In this way, it captures the underlying spatial features by performing convolution operations in multidimensional data.

In constructive conGRU-LSTM, LSTM architecture includes three gate the input, forget and output while the conGRU has only two gates layers: reset gate r_t and an update gate u_t . The update gate checks the memory of the earlier cell to stay active and the reset gate is used to combine input sequence of next cell with preceding cell memory. However, LSTM is a bit different in some ways: firstly, the conGRU cell consists of two gates as a substitute LSTM are three. Secondly, the input and forget gate in LSTM are merged to update gate and for hidden state reset gate are directly applied. The constructive steps involved in conGRU-LSTM is listed below.

$$i_t = \sigma(W_{ix} * x_t + W_{ih} * h_{t-1} + W_{ic} \circ c_{t-1} + b_i) \quad (7)$$

$$f_t = \sigma(W_{fx} * x_t + W_{fh} * h_{t-1} + W_{fc} \circ c_{t-1} + b_f) \quad (8)$$

$$o_t = \sigma(W_{ox} * x_t + W_{oh} * h_{t-1} + W_{oc} \circ c_{t-1} + b_o) \quad (9)$$

$$g_t = \tanh(W_{gx} * x_t + W_{gh} * h_{t-1} + b_g) \quad (10)$$

$$c_t = f_t \tanh c_{t-1} + i_t \circ g_t \quad (11)$$

$$h_t = o_t \circ \tanh (c_t) \quad (12)$$

The Eq. (7) and (8) shows that is passed as an input to the first layer of conGRU (u_t), whereas u_t and h_{t-1} are multiplied to weight and this information is forwarded to reset gate (r_t).

$$u_t = \sigma(W_{ux} * c_t + W_{uh} * h_{t-1} + b_u) \quad (13)$$

$$r_t = \sigma(W_{rx} * c_t + W_{rh} * h_{t-1} + b_r) \quad (14)$$

$$z_t = \tanh(W_{zx} * x_t + W_{zh} * h_{t-1} + b_z) \quad (15)$$

$$p_t = u_t \circ h_{t-1} + (1 - u_t) \circ z_t \quad (16)$$

From above Eq. (7-16), t denotes the time step in conGRU-LSTM, x_t , h_{t-1} and c_{t-1} denotes the input data at time t , h_t denotes the final state; c_t denotes the state of the storage (memory) cell; i_t, f_t, o_t are the input gate, forget gate and output gate of LSTM respectively. u_t and r_t are the gates of GRU. The hidden layer of conGRU-LSTM model are g_t, z_t and p_t . The weights of conGRU is W_{ux}, W_{rx} and W_{zx} ; the weight of LSTM is depicted as $W_{ix}, W_{fx}, W_{ox}, W_{gx}$; The biases of conGRU-LSTM are $b_i, b_f, b_o, b_g, b_u, b_r$ and b_z , \circ, σ and \tanh denote the convolution operation, element multiplication, Sigmoid function and \tanh function.

Assume the mapping function $M_{CGL}^i(.)$ denote the i^{th} layer of the ConGRU-LSTM (cGL) layers, and then the representation features of source and target subjected CXR and CT image data can be formulated as:

$$M_S^{cGL} = m_{cGL}^N(\dots M_{cGL}^1(\chi_S)) \quad (17)$$

$$M_T^{cGL} = m_{cGL}^N(\dots M_{cGL}^1(\chi_T)) \quad (18)$$

M_S^{CGL} and M_T^{CGL} are the final feature representation of the two image domain through ConGRU-LSTM layers. The output M_S^{CGL} and M_T^{CGL} through conGRU-LSTM layers are then input their discriminative and informative features obtained through InceptionResNetV2, Xception and Efficient-Net. The MMD is also employed in this classification assignment to enforce the extracted features constraints during transfer learning. By embedding the learned representations output by feature extraction layers from two domains into a reproducing kernel Hilbert space, MMD eliminates the domain discrepancy with the help of adaptation network, and, the squared formulation of MMD is computed in Eq. (6). Since, the transfer learning is also applied in ConGRU-LSTM classification part, the constructed MMD in Eq. (6) is re-modified according to the ConGRU-LSTM layers.

Assume, M_{2d}^i denote the i^{th} layer of ConGRU-LSTM for two domain $2d$, then the transfer loss L_{MMD} from the feature extraction-MMD can be computed as

$$D_i^S = m_{2d}^i((m_{2d}^{i-1}(\dots m_{2d}^1(M_S^{CGL}))) \tag{18}$$

$$D_i^T = m_{2d}^i((m_{2d}^{i-1}(\dots m_{2d}^1(M_T^{CGL}))) \tag{19}$$

$$L_{MMD} = \sum_{i=1}^N d_{MMD}^2(D_i^S, D_i^T) \tag{20}$$

From Eq. (20), it is proved that the source and target domain in ConGRU-LSTM shares the same parameters of $M_{CGL}^i(.)$ and the parameters are independent in $M_{2d}^i(.)$ to classify the extracted CXR and CT images. Finally, the reshaped output features are fed into softmax layer to classify different types of lung diseases. Thus, the constructed framework is trained and utilized to resolve the epistemic uncertainty issue and develops a reliable solution for rapid detection of various diseases using CXR and CT images.

4. RESULT AND DISCUSSION

In this section, the efficiency of the EDepLDP model is examined by implementing it in MATLAB 2019a using the CXR and CT images which is discussed in Section 3.1. For the experimental purposes, 70% of images are taken for training and the remaining 30% are taken for testing of the EDepLDP model from each categories of lung diseases using both CXR and CT images. Further a comparative analysis is carried out to understand the improvement of the EDepLDP model contrasted to the existing models including VEDL model [17], VGG model [18], GCN model [19], CNN model [12], E2E-DNN model [20], EfficientNet v2-M model [23]. The evaluation metrics used to measure the success of the proposed and existing models are briefly illustrated below.

4. 1. Accuracy

It is the fraction of accurately categorized examples in each category of lung diseases over the sum amount of examples considered.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+F} \tag{21}$$

In above Eq. (21), TP (True Positive) defines the outcome that the model accurately classifies the lung diseases types as itself e.g., pneumonia is classified as pneumonia. TN (True Negative) defines the outcome that the model the classifier accurately classifies the Covid-19 as Covid-19. FP (False Positive) defines the outcome that the model inaccurately classifies the lung diseases (Atelectasis\Covid-19\Infiltrate\Normal\Pneumonia) as Atelectasis. FN (False Negative) defines the outcome that the model inaccurately classifies the Normal as Infiltrate.

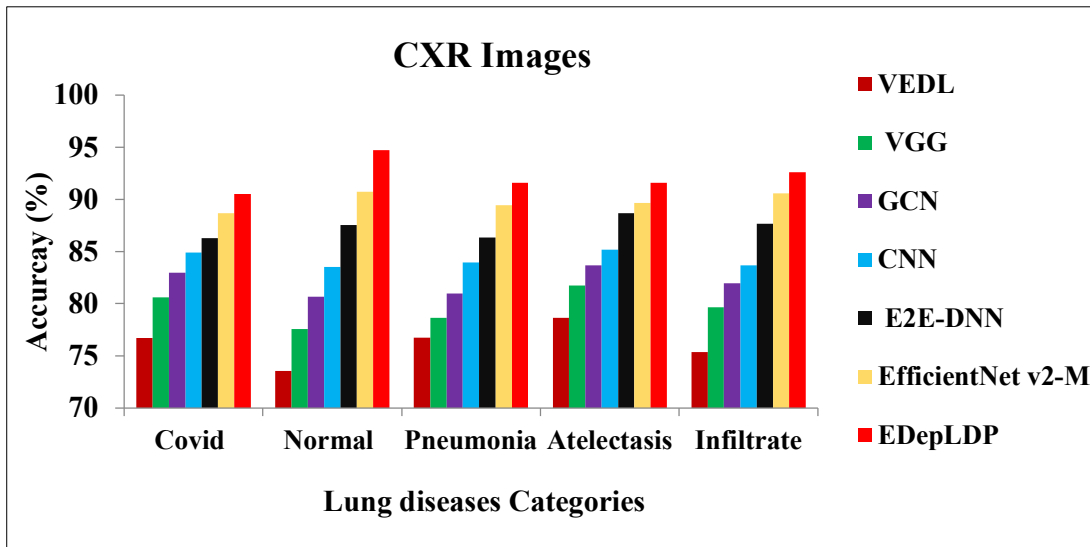


Figure 3. Accuracy Comparison for lung disease category using CXR images

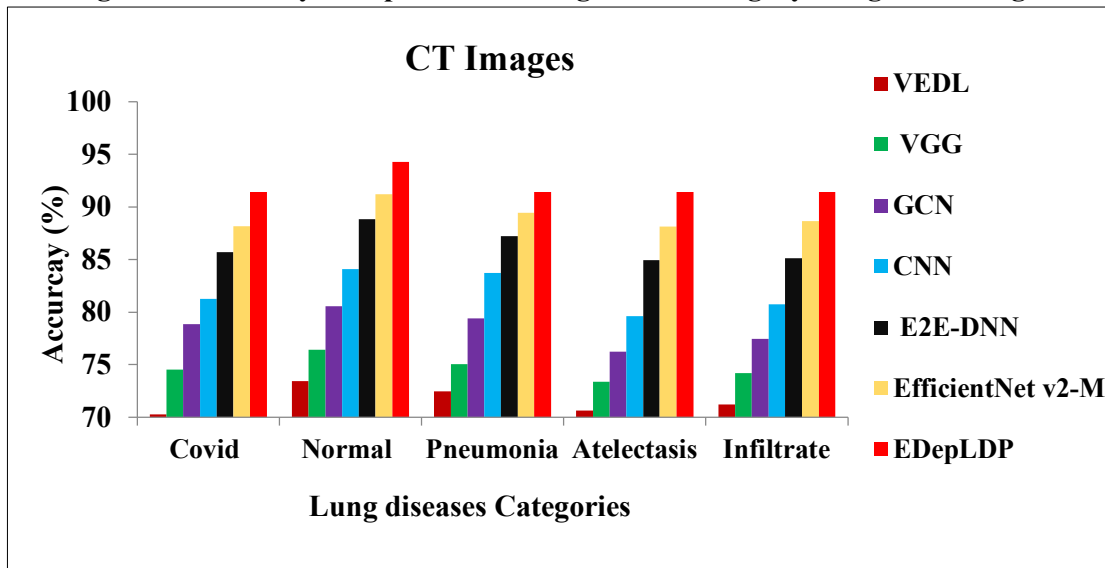


Figure 4. Accuracy Comparison for lung disease category using CT images

Figure 3 and 4 displays the accuracy (in %) achieved by CNN, VEDL, VGG, GCN, E2E-DNN, EfficientNet v2-M, and EDepLDP models for diagnosing various lung disease categories such as Covid-19, Normal, Pneumonia, Atelectasis and Infiltrate using CT and CXR images respectively. It is determined that the accuracy of EDepLDP for each category of lung diseases is superior to that of other classification models because it enhances the number of learning examples for each category of diseases from CXR and CT images. For example, in the categorization of atelectasis, the accuracy of EDepLDP is 22.90% and 29.44% greater than CNN, 16.27% and 24.61% greater than VEDL, 13.03% and 19.94% greater than VGG, 10.71% and 14.84% greater than GCN, 5.66% and 7.63% greater than E2E-DNN, 2.27% and 3.72% greater than EfficientNet v2-M for CXR and CT images respectively. This demonstrates that the suggested EDepLDP model is efficient to categorize lung disease with greater accuracy than all other existing models.

4.2 Precision

It is the percentage of accurately categorized examples of lung diseases categories at TP and FP rates.

$$Precision = Precision = \frac{TP}{TP+F} \tag{22}$$

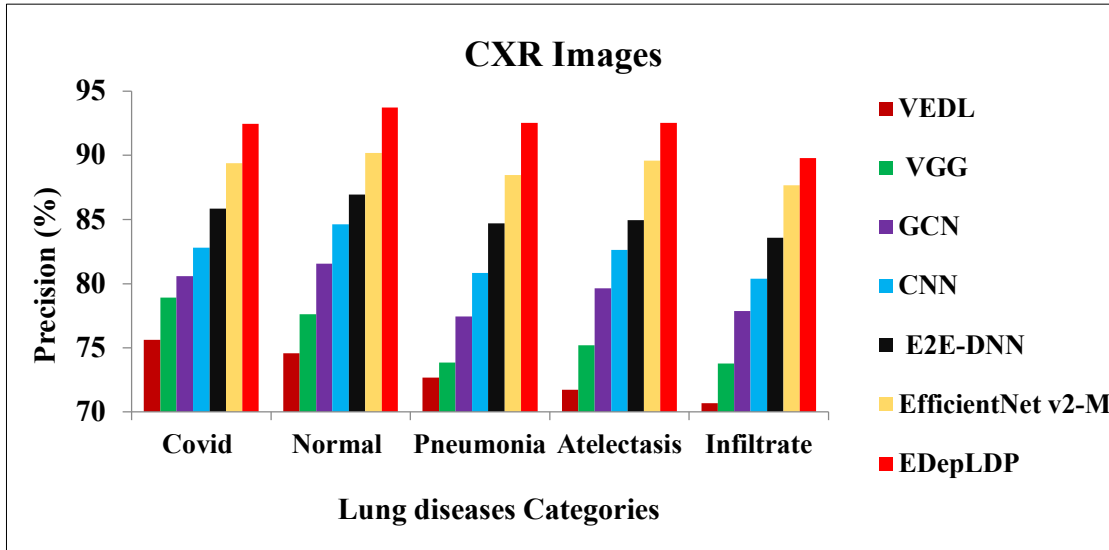


Figure 5. Precision Comparison for lung disease category using CXR images

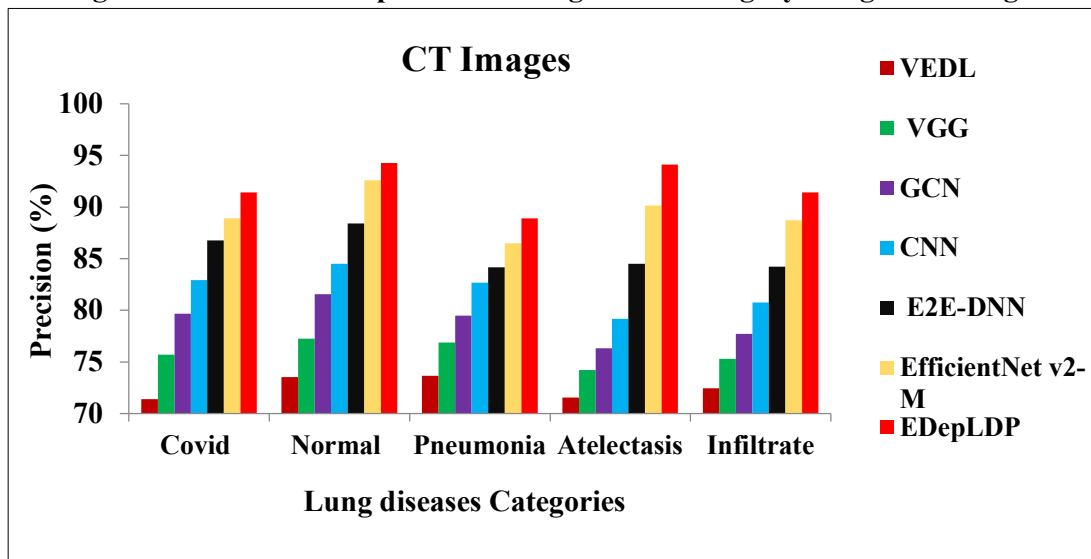


Figure 6. Precision Comparison for lung disease category using CT images

Figure 5 and 6 illustrates the precision (in %) obtained by CNN, VEDL, VGG, GCN, E2E-DNN, EfficientNet v2-M, and EDepLDP models for diagnosing various lung disease categories such as Covid-19, Normal, Pneumonia, Atelectasis and Infiltrate using CT and CXR images respectively. It is determined that the precision of EDepLDP for each category of lung diseases is superior to that of other classification models for each category of diseases from CXR and CT images. For example, in the case of Covid classification, the precision of EDepLDP is 22.27% and 28.11% greater than CNN, 17.18% and 20.78% greater than VEDL, 14.76% and 14.75% greater than VGG, 11.68% and 10.26% greater than GCN, 7.69% and 5.37% greater than E2E-DNN, 3.45 % and 2.87% greater than EfficientNet v2-M for CXR and

CT images respectively. This demonstrates that the suggested EDepLDP model is efficient to categorize lung disease with greater precision than all other existing models.

4.3 Recall

It is the ratio of exactly classified categories of lung diseases cases at TP and FN rates.

$$Recall = \frac{TP}{TP + FN} \tag{23}$$

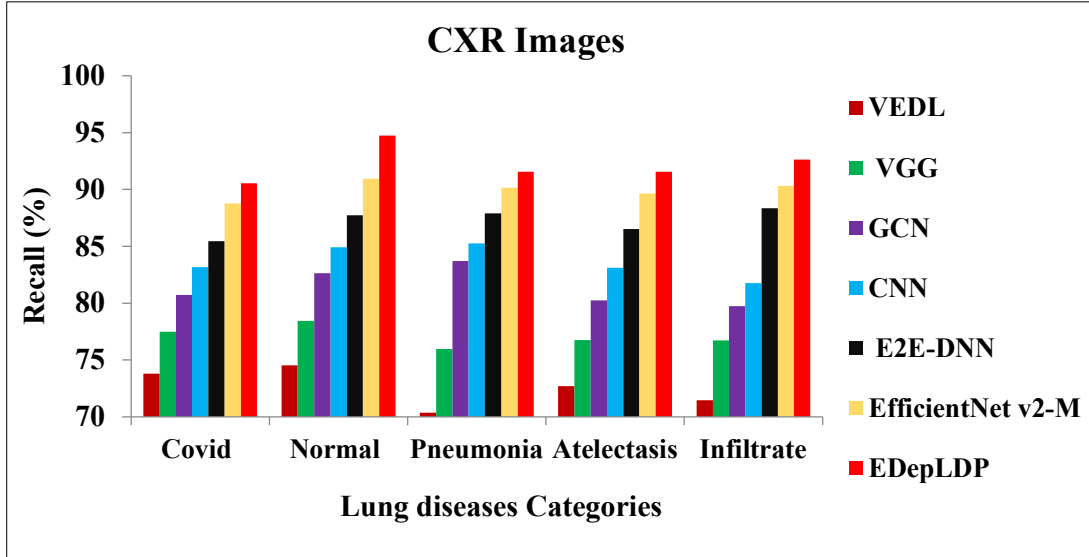


Figure 7. Recall Comparison for lung disease category using CXR images

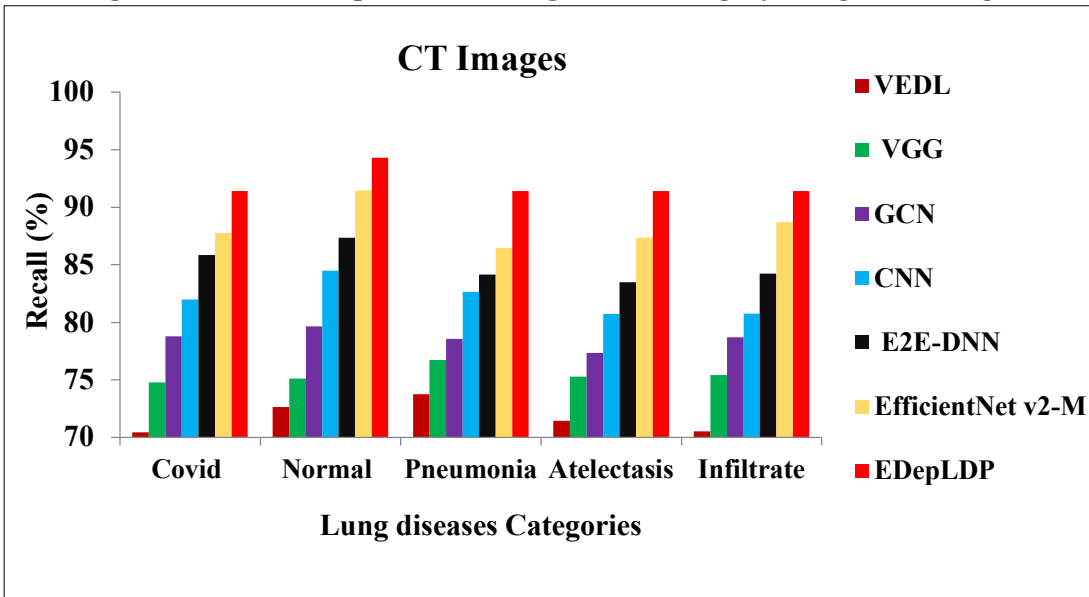


Figure 8. Recall Comparison for lung disease category using CT images

Figure 7 and 8 depicted the recall (in %) obtained by CNN, VEDL, VGG, GCN, E2E-DNN, EfficientNet v2-M, and EDepLDP models for diagnosing various lung disease categories such as Covid-19, Normal, Pneumonia, Atelectasis and Infiltrate using CT and CXR images respectively. It is determined that the recall of EDepLDP for each category of lung diseases is superior to that of other classification models for each category of diseases from CXR and CT images. For example, in the case of pneumonia classification, the recall of EDepLDP is 30.11% and 23.97% greater than CNN, 20.55% and 19.16% greater than VEDL, 9.39 and 16.38%

greater than VGG, 7.38% and 10.59% greater than GCN, 4.16% and 8.65% greater than E2E-DNN, 1.55% and 5.74% greater than EfficientNet v2-M for CXR and CT images respectively. Thus, it proves that the proposed models can increase the recall of classifying the lung diseases types compared to all other existing models.

4.4 F1-Score

It is defined as the weighted mean of precision and recall, with '1' being the highest and '0' being the lowest.

$$F1 - score = 2 \times \frac{Precision \cdot Recall}{Precision + Recca} \tag{24}$$

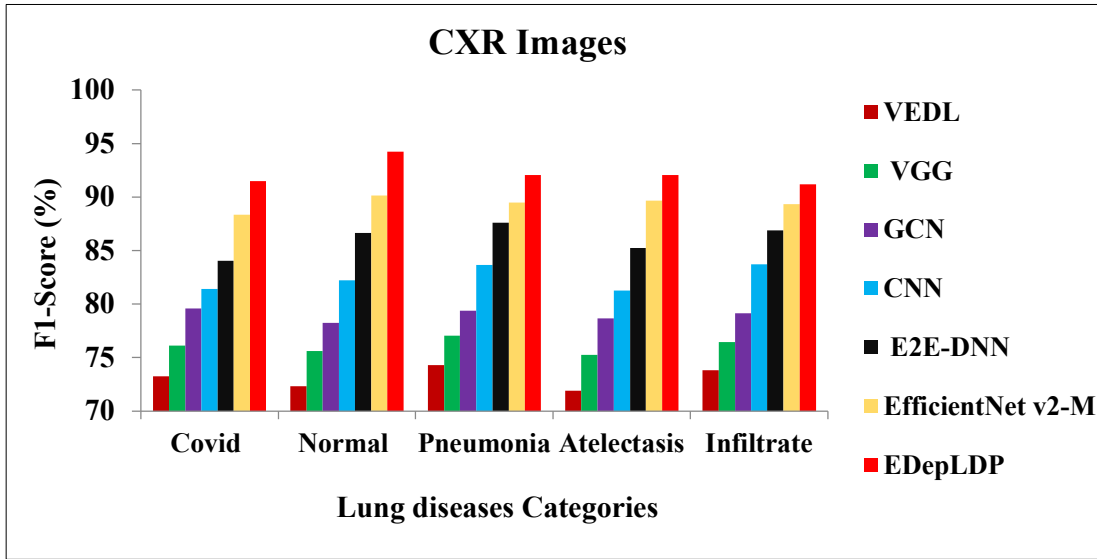


Figure 9. F1-score Comparison for lung disease category using CT images

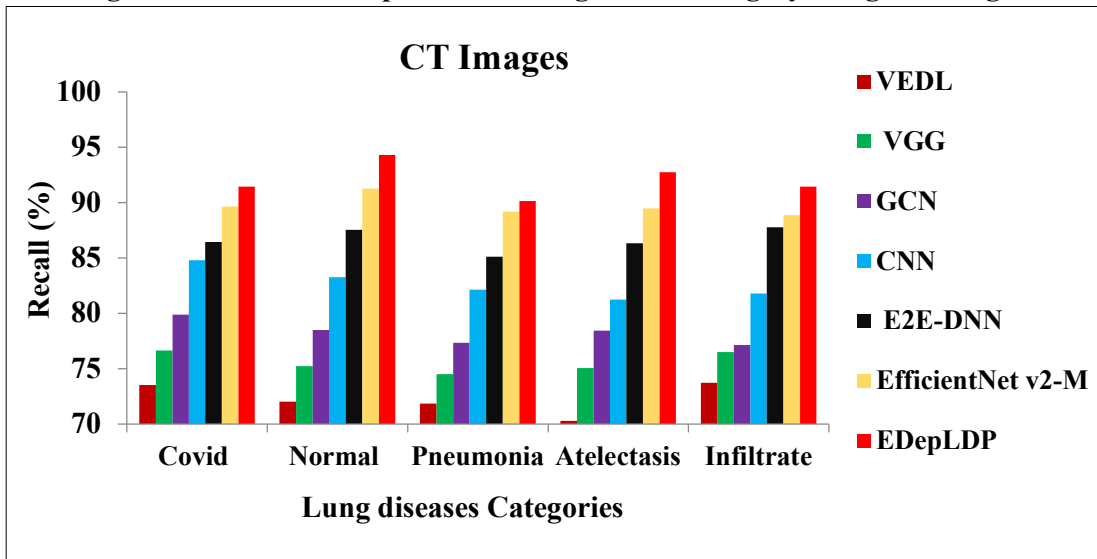


Figure 10. Comparison for lung disease category using CT images

Figure 9 and 10 provides the F1-score (in %) obtained by CNN, VEDL, VGG, GCN, E2E-DNN, EfficientNet v2-M, and EDepLDP models for diagnosing various lung disease categories such as Covid-19, Normal, Pneumonia, Atelectasis and Infiltrate using CT and CXR images respectively. It is determined that the F1-score of EDepLDP for each category of lung diseases is superior to that of other classification models for each category of diseases from

CXR and CT images. For example, in the case of normal classification, the F1-score of EDepLDP is 30.27% and 30.90 greater than CNN, 24.62% and 25.34% greater than VEDL, 9.39 and 20.09% greater than VGG, 14.62% and 13.25% greater than GCN, 8.77% and 7.68% greater than E2E-DNN, 4.56% and 3.29% greater than EfficientNet v2-M for CXR and CT images respectively. Thus, it proves that the proposed models can increase the F1-score of classifying the lung diseases types compared to all other existing models.

5. CONCLUSION

Lung diseases are one of the common causes of death worldwide and the number is increasing as new lethal viruses and infections are reported to infect the thoracic region e.g. Pneumonia Covid-19. Several CXR and CT based CAD systems are developed for timely diagnosis of lungs diseases. Most of these disease classification systems are not robust enough to provide an accurate classification results which also results in uncertainty. To overcome the limitations of existing approaches, this paper develops a DL model for rapid lung disease diagnosis using CXR and CT images. Initially, image segmentation is performed using the U-Net model to obtain an enhanced lung ROIs. Then, InceptionResNetV2 and Xception are used to extract hierarchically informative and discriminative features from segmented images. Finally, the retrieved deep features are sent into the softmax layer of conGRU-LSTM. Utilizing the TL model, learnt weight for InceptionResNetV2, Xception, and conGRU-LSTM was gained from an pre-trained Efficient-Net for the efficient classification of various lung disorders. Compared to conventional CNN models, the EDepLDP model achieves an accuracy of % and % on the gathered CXR and CT images as demonstrated by the test results.

REFERENCES

1. Cruz, A. A. (2007). Global surveillance, prevention and control of chronic respiratory diseases: a comprehensive approach. World Health Organization.
2. Marciniuk, D. D., & Schraufnagel, D. E. (2017). The global impact of respiratory disease. European Respiratory Society.
3. De Marco, R., Accordini, S., Cerveri, I., Corsico, A., Sunyer, J., Neukirch, F., ... & Burney, P. (2004). An international survey of chronic obstructive pulmonary disease in young adults according to GOLD stages. *Thorax*, 59(2), 120-125.
4. Rengier, F., Melzig, C., Derlin, T., Marra, A. M., & Vogel-Claussen, J. (2019). Advanced imaging in pulmonary hypertension: emerging techniques and applications. *The International Journal of Cardiovascular Imaging*, 35, 1407-1420.
5. Farhat, H., Sakr, G. E., & Kilany, R. (2020). Deep learning applications in pulmonary medical imaging: recent updates and insights on COVID-19. *Machine vision and applications*, 31, 1-42.
6. El-Baz, A., Beache, G. M., Gimel'farb, G., Suzuki, K., Okada, K., Elnakib, A., ... & Abdollahi, B. (2013). Computer-aided diagnosis systems for lung cancer: challenges and methodologies. *International journal of biomedical imaging*, 2013.
7. Shariaty, F., & Mousavi, M. (2019). Application of CAD systems for the automatic detection of lung nodules. *Informatics in Medicine Unlocked*, 15, 100173.
8. Zakirov, A. N., Kuleev, R. F., Timoshenko, A. S., & Vladimirov, A. V. (2015). Advanced approaches to computer-aided detection of thoracic diseases on chest X-rays. *Appl Math Sci*, 9(88), 4361-4369.

9. Wu, C., Luo, C., Xiong, N., Zhang, W., & Kim, T. H. (2018). A greedy deep learning method for medical disease analysis. *IEEE Access*, 6, 20021-20030.
10. Kieu, S. T. H., Bade, A., Hijazi, M. H. A., & Kolivand, H. (2020). A survey of deep learning for lung disease detection on medical images: state-of-the-art, taxonomy, issues and future directions. *Journal of imaging*, 6(12), 131.
11. Ma, J., Song, Y., Tian, X., Hua, Y., Zhang, R., & Wu, J. (2020). Survey on deep learning for pulmonary medical imaging. *Frontiers of medicine*, 14, 450-469.
12. Zaidi, S. Z. Y., Akram, M. U., Jameel, A., & Alghamdi, N. S. (2021). Lung segmentation-based pulmonary disease classification using deep neural networks. *IEEE Access*, 9, 125202-125214.
13. Bharati, S., Podder, P., & Mondal, M. R. H. (2020). Hybrid deep learning for detecting lung diseases from X-ray images. *Informatics in Medicine Unlocked*, 20, 1-14.
14. Hashmi, M. F., Katiyar, S., Keskar, A. G., Bokde, N. D., & Geem, Z. W. (2020). Efficient pneumonia detection in chest x-ray images using deep transfer learning. *Diagnostics*, 10(6), 417.
15. Wang, W., Li, Y., Li, J., Zhang, P., & Wang, X. (2021). Detecting COVID-19 in chest x-ray images via MCFF-Net. *Computational Intelligence and Neuroscience*, 2021, 1-8.
16. Vieira, P., Sousa, O., Magalhães, D., Rabêlo, R., & Silva, R. (2021). Detecting pulmonary diseases using deep features in X-ray images. *Pattern Recognition*, 119, 1-13.
17. Cao, F., & Zhao, H. (2021). Automatic lung segmentation algorithm on chest x-ray images based on fusion variational auto-encoder and three-terminal attention mechanism. *Symmetry*, 13(5), 1-15.
18. Tasci, E., Uluturk, C., & Ugur, A. (2021). A voting-based ensemble deep learning method focusing on image augmentation and preprocessing variations for tuberculosis detection. *Neural Computing and Applications*, 33(22), 15541-15555.
19. Zhang, D., Ren, F., Li, Y., Na, L., & Ma, Y. (2021). Pneumonia detection from chest X-ray images based on convolutional neural network. *Electronics*, 10(13), 1512.
20. Li, Z., Huang, K., Liu, L., & Zhang, Z. (2022). Early detection of COPD based on graph convolutional network and small and weakly labeled data. *Medical & Biological Engineering & Computing*, 60(8), 2321-2333.
21. Junayed, M. S., Jeny, A. A., Islam, M. B., Ahmed, I., & Shah, A. S. (2022). An efficient end-to-end deep neural network for interstitial lung disease recognition and classification. *Turkish Journal of Electrical Engineering and Computer Sciences*, 30(4), 1235-1250.
22. Ravi, V., Acharya, V., & Alazab, M. (2022). A multichannel EfficientNet deep learning-based stacking ensemble approach for lung disease detection using chest X-ray images. *Cluster Computing*, 1-23.
23. Kim, S., Rim, B., Choi, S., Lee, A., Min, S., & Hong, M. (2022). Deep learning in multi-class lung diseases' classification on chest X-ray images. *Diagnostics*, 12(4), 915.
24. Xiaosong Wang, Yifan Peng, Le Lu, Zhiyong Lu, Mohammadhadi Bagheri, Ronald M. Summers. ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases, *IEEE CVPR*, pp. 3462-3471, 2017.
25. <https://www.kaggle.com/paultimo/thymooney/chestxray-pneumonia>
26. <https://radiopaedia.org/articles/lung-atelectasis?lang=us>

27. <https://www.kaggle.com/datasets/mehradaria/covid19-lung-ct-scans>
28. <https://radiopaedia.org/articles/viral-respiratory-tract-infection>
29. <https://radiopaedia.org/playlists/41156?lang=us>
30. Tan, M., Chen, B., Pang, R., Vasudevan, V., Sandler, M., Howard, A., & Le, Q. V. (2019). Mnasnet: Platform-aware neural architecture search for mobile. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 2820-2828).