



An Improving Lung Disease Detection by Combining Ensemble Deep Learning and Maximum Mean Discrepancy Transfer Learning

Thiruvankadasamy Kousiga^{1*} Palanisamy Nithya²

¹Department of Computer Science, PSG College of Arts and Science, Coimbatore - 641014, Tamil Nadu, India

²Department of Networking and Mobile Application,
PSG College of Arts and Science, Coimbatore - 641014, Tamil Nadu, India

* Corresponding author's Email: kousiga2318phd@gmail.com

Abstract: In accordance to the World Health Organization (WHO), various pulmonary diseases cause thousands of deaths annually. The early diagnosis is required to lessen the mortality rate. For this reason, A Convolutional Neural Network (CNN)-based Lung Disease (LD) detection system is developed to classify segregated lung sections into various pulmonary diseases types. However, epistemic uncertainty in the scanned images affecting the performance of detection classifiers. Hence, in this paper, a multi-modal approach is proposed to solve the epistemic uncertainty issue and provides a reliable solution for rapid detection of various LD types from CXR images. In this method, CT images are additionally used to improve model's performance as it contains detailed information that might be exploited to provide efficient results. Initially, the collected images are segmented using U-Net model to get enhanced lung Region of Interest (ROIs). Then ResNet50, DenseNet121, InceptionResNetV2 and XceptionV3 are used to hierarchically extract informative and discriminative features from collected CXR and CT images. The retrieved deep features are fed into the Ensemble-Convolutional Long Short Term Memory with Extreme Machine Learning (E-conLSTM-ELM) to minimize the computational time and increase the accuracy. Moreover, Transfer Learning (TL) model is employed to learn the weight of the E-conLSTM-ELM to exchange the knowledge between features and classes relation among CXR and CT images. Also, the domain adaptation approach is a variant of TL model that relies on employing similar datasets for a shared learning problem. This adaption strategy reduces the domain shift (data dispersion) using Maximum Mean Discrepancy (MMD). The shared semantic features from CT images through TL improve the in-depth learning of softmax layer to classify different LD types. The proposed work is simply named as Convolutional LD Scan (CovLscan) framework. The test outcomes reveal that the CovLscan model accomplishes an overall accuracy of 95.46% and 96.15% on the collected ChestX-ray8 and NIH-CXR datasets, which is higher than the existing models like Automated Hierarchical Deep Learning based LD Diagnosis (AHDL-LDD), EfficientNet version2-Medium (EfficientNet v2-M), Lung diseases prediction Network22 (LungNet22), Chest tract disorder prediction using Dilated Convolutional Network (CDCNet) and Auction-Based Optimization Algorithm-CNN (ABOA-CNN).

Keywords: Pulmonary disease, Chest X-ray, Computed tomography, Deep learning, Epistemic uncertainty.

1. Introduction

Pulmonary illnesses, often known as respiratory disorders, have a significant effect on the bronchi and other pulmonary tissues [1]. Examples include pneumonia, TB and Coronavirus Disease 2019 (COVID-19). According to the Federation of Global Pulmonary Communities, 334 million people worldwide have asthma, and 1.4 million die each year

from TB [2]. The COVID-19 pandemic affected every country in the globe, infecting millions and causing damage on medical facilities [3]. Respiratory disorders are the leading cause of death globally. Early detection is critical for enhancing long-term life expectancy and increasing the chances of rehabilitation [4].

In the past years, Pulmonary disorders are diagnosed using various medical imaging

technologies such as skin tests, biopsy, sputum sample tests, Positron Emission Tomography (PET), Magnetic Resonance Imaging (MRI), CXR and CT scans [5]. Among these, CXR are prominent medical imaging technology that allow for rapid superior quality assessments of the lung parenchyma and adjacent structures to aid in early diagnosis [6]. However, their accessibility and availability are limited in impoverished nations due to lack of a robust medical system and patient comfort [7]. Automated disease recognition and categorization systems are needed to program CXR analysis, reducing patient effort and improving overall health outcomes.

DL is an emerging field used to diagnose various LD categories, aiding healthcare providers in making accurate medical decisions [8]. DL models are managed by multi-channel neural networks, improving their image categorization and accuracy in LD detection [9]. This approach has led to numerous investigations on medical imaging detections for early LD detection and diagnosis. For instances, AHDL-LDD model [10] was developed using CXR scans. This model utilized lung mask annotations to enhance lung ROIs in a CXR database. U-Net structure was used to extract the customized masks to differentiate the normal and infection levels. A modified CNN structure was used to categorize segregated lung areas. The model's performance in diagnosing pulmonary diseases was significantly impacted by epistemic uncertainty.

In order to solve this, CovLscan is developed for efficient LDs classification using CXR images. Also, CT images are employed to improve the model's performance, because CT images includes detailed information that can be used to produce efficient results. Initially, both CXR and CT images are segmented using U-Net model to obtain improved lung ROIs. In this framework, various pre-trained CNN models like ResNet50, DenseNet121, InceptionResNetV2 and XceptionV3 are hierarchically extract informative and discriminative features from CXR and CT images. The conLSTM is used as classification model which gradually receives normalized feature inputs and fed into softMax for the classification task. The integration of con-LSTM and ELM i.e., E-conLSTM-ELM is designed to minimize computation time and optimize precision. The Deep feature from pre-trained CNN is fed into ELM, allowing the model to be optimized for accuracy without the need of a traditional Fully Connected (FC) layer. Moreover, TL is utilized in E-conLSTM-ELM for enhancing the task of LD classification by facilitating knowledge exchange between features and classes in CXR and CT images.

Deep domain adaptation is a type of TL that handles diverse datasets for the learning utilizing MMD to reduce domain shift. Additionally, the obtained CT labels will be fed into TL to learn different levels of extensive details and perspectives with shared semantic features from CT images for increasing the prediction result. Finally, the softmax layer of conLSTM is used for the prediction task. This model efficiently eliminates the epistemic uncertainty issue and provides a reliable solution for rapid detection of various LD types using CXR images

The manuscript's remaining portions are prepared as, Section II examines the work relating to this study. Section III explains the proposed CovLscan model for LD classification. Section IV illustrates its model's performance effectiveness. Section V summarizes the whole work and suggests future enhancement.

2. Literature survey

The Class Activation Region Influence Maximization Conditional Generative Adversarial Network (CARIM-cGAN) was created [11] using CXR data for LD classification. But, it struggles to separate high-level features and stochastic disparities affecting categorization accuracy.

A CNN with Contrast Adaptive Histogram Equalization (CLAHE) image scaling approach using the maximum window function was presented [12] using CXR data for LDs detection. However, this model provides lower accuracy results on larger dataset.

A multilayer EfficientNet-based stacked ensemble technique was employed [13] for LD identification utilizing CXR images. But, imbalanced sample distribution in the dataset might affect the accuracy results.

An EfficientNet v2-M was suggested [14] for categorizing the LD on CXR images. But, this model loses a substantial amount of useful data resulting in overfitting of classifier which reduces accuracy and recall.

LungNet22 was developed [15] which utilizes pre-trained CNN models to classify and predict the LD using CXR data. But, precision and recall score was lower on employing images with poor illumination and backgrounds.

A VGG19 model was constructed [16] for multi-label LD classification using CXR images. But, few examples failed to provide appropriate accuracy prediction outcomes due to poor image quality.

A Two Dimensional (2D) - CNN model and min-max scaling was developed [17] for the automated LD detection using CXR images. But, large class

variance in training models hinders accuracy and F1-Score results.

A hybrid DL model (RVCNet) was presented [18] using CXR images of multiple classes for predicting LDs. On the other hand, when the data was increased, the performance accuracy decreased.

A multi-classification method CDCNet was developed [19] which integrates residual network and dilated convolution to classify LD types using CXR images. However, the models hyper-parameter was not fine-tuned properly lowering the accuracy rate.

An Auction-Based Optimization Algorithm (ABOA)-CNN model was devised [20] using CXR images for the pulmonary disease prediction. But, this model failed to identify the optimal features subset which lowers the accuracy rate.

3. Proposed methodology

In this section, the suggested CovLscan model is shown. Fig. 1 pipeline of the suggested model. Table 1 lists the notations used in this study.

3.1 U-Net based segmentation

In this model, the U-Net structure is used for segmentation and it consists of contraction (encoder) and expansion (decoder) routes. For down-sampling, the compression approach uses recurrent 3x3 convolution functions, ReLU Stimulation and 2x2 max pooling, each of which increases the total number of feature layers.

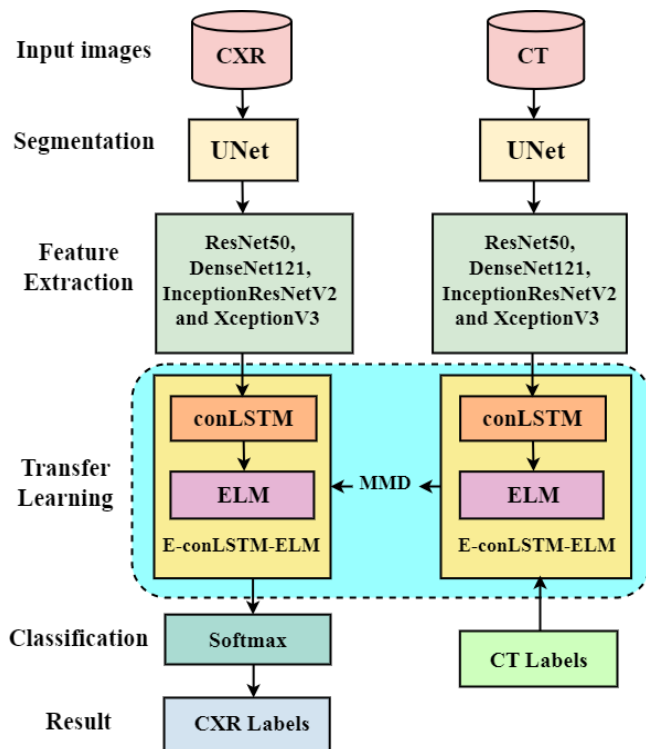


Figure. 1 Entire Pipeline of the Proposed Study

Table 1. Lists of notations

Notations	Description
W^h	Weights Of Trained U-Net
b^k	Bias Variable
h	Hidden Layer
k	Number of Input Variables
q	Number of Neurons
$\mathcal{A}_q^r$	Activation Function
n	Total Number of Images
W_q	Reference Masks for Each Image
W'_q	Segmented Masks
C_t	Cell State
I_t	Input Gate
F_t	Forget Gate
O_t	Output Cell
$\otimes$	Convolution operator
$\times$	Hadamard product
V	Distinct Training Images
$j(X)$	Activation Operation
$O_\ell(X)$	Output Vector of SLFNN
i_z	Input Weight respect to z^{th} Hidden Layer
b_z	Bias Weight respect to z^{th} Hidden Layer
α_z	Output Weight respect to z^{th} Hidden Layer
h	Output matrix of hidden layer
Q	Target Output
α	Output Weights
$h^\dagger$	Moore-Penrose Operation
J_y	Input Map Selection
$F(.)$	Activation Function
b_y^c	Training Bias
$\mathcal{K}_{zy}^c$	Variable Kernels
x	Input Activation
$F(X)$	Output Activation Node
$down(.)$	Down-Sampling
X_y^{c-1}	Local features from previous layers
X_y^c	Output activation of subsequent layers in conLSTM
D_S	Source Domain Data
m	Source Site Value
D_T	Target Domain Data
μ	Nonlinear Mapping Function in HS
$M^i(.)$	i^{th} layer of the E-ConLSTM-ELM
χ_s	Source Feature Representation
χ_T	Target Feature Representation
M_{2d}^i	i^{th} layer of E-ConLSTM-ELM for II Domain
L_{MMD}	Transfer Loss of MMD

The expansion route employs up-convolution with 2×2 transposed convolutions for precision, while skipping algorithms recover data lost during shrinkage channel down-sampling, facilitating feature map transfer. A U-Net structure is created using CXR and CT image datasets to enhance lung ROIs, with training efficiency determined by the configuration of adaptable parameters. The U-Net structure's weights are learned from a database and used to optimize segmentation results on CXR and CT images, with extra bias projections depicted in Eq. (1).

$$\mathcal{A}_q^h = \sum_{k=1}^K W_{qk}^h a_k + b_q^h, k = 1, 2, \dots, K \quad (1)$$

In Eq. (1), W^h denotes the trained U-Net weights, a symbolizes the input parameter, b^k is the value added as a bias. The model's hidden layer, input parameters, and neurons are denoted as h, k and q . An activation mechanism takes the outcome $\mathcal{A}_q^r$ as input and activates or deactivates a neuron. The Binary Cross-Entropy (BCE) loss operation is used in this model to estimate the difference between segmented masks and perceptual typical masks for every single image of CXR and CT, as stated in Eq. (2).

$$BCE = -\frac{1}{n} \sum_{q=1}^n W_m \log W'_q + (1 - W_q) \log (1 - W'_q) \quad (2)$$

In Eq. (2), n denotes the complete image numbers used for training, W_q provides subjective standard masks for each CT and CXR image and W'_q contributes segmented masks created by the model.

3.2 Feature extraction using ensemble pre-trained CNN models

In this model, the pre-trained CNN models ResNet50, DenseNet121, Inception-ResNet-V2 and Xception are employed to extract discriminative and informative features from the segmented CXR and CT images which is illustrated below.

ResNet50: ResNet50, a ResNet variation or Residual Network, is made up of 48 convolutional layers, one MaxPool, and one average pool layer. Every convolution unit consists of three convolution layers, as does each recognition component. ResNet-50 contains almost 23 million variables that can be directed.

DenseNet121: DenseNet121's input is a constant 224×224 RGB image. DenseNet121 consists of 121 layers, each with about 8 million variables. It is divided into DenseBlocks, with each unit having the same feature map size but different filter frequencies. The spaces between the blocks are known as a transitional layer, and they do batch normalization during downsampling. Finally, a pooling network with softmax stimulation is employed to categorize.

Inception-ResNet-V2: It has 164 layers and a picture input size of 299×299 . Its fundamental building block is the Residual Inception Block, which employs a 1×1 convolution filter diversification layer to increase the filter boundary density. Batch standardization is implemented on the highest standard layers. The design features multiple-sized convolutional filters with residual associations to reduce deterioration caused by deep networks and shorten training times.

Xception: It is a modification of the Inception structure that replaces the Inception elements with depth-wise independent convolutions. Xception beat the standard InceptionV3 on the ImageNet dataset, achieving greater Top-1 and Top-5 reliability. The amount of variables in Xception is approximately 23 million.

3.3 Transfer learning and E-conLSTM-ELM for classification

The extracted deep features from the ResNet50, DenseNet121, InceptionResNetV2 and Xception are fed into the E-conLSTM-ELM model to accurately classify the various LD. In the proposed E-conLSTM-ELM model, TL is used for knowledge exchange between features and classes relation among CXR and CT images and improve the target task of LD classification. The E-conLSTM-ELM model learned generated features in the domain and their variables were preserved throughout the TL. The model substitutes matrix multiplication with convolution computations for every gate in the LSTM cell, allowing it to capture fundamental spatial properties in multidimensional information.

The most significant component of the functional ConvLSTM structure is the cell state C_t which is utilized to store the data. If the input gate I_t is triggered, the input value is saved, but if the forget gate F_t is activated, the prior state c_{t-1} is discarded. Furthermore, the output cell O_t determines whether the current cell state c_t is turned into the ultimate hidden state h_t . In this approach, a basic LSTM model works. However, in the ConvLSTM layer, the inputs $X_1, X_2, X, \dots, X_n$, the cell states $C_1, C_2, C_3, \dots, C_n$, the hidden states

$H_1, H_2, H_3, \dots, H_n$ and the gates I_t, C and O_t are all termed as 3D tensors.

In order to illustrate the ConvLSTM layer, view the inputs and gates as vectors in a grid-like form in linear dimension. The ConvLSTM layer predicts the future state of a cell by gathering the inputs and final iteration of the local entities of the specific unit. The constructive steps involved in con-LSTM is listed below from Eqs. (3) to (7),

$$F_t = \sigma(W_{XF} \otimes X_t + W_{HF} \otimes H_{t-1} + W_{CF} \otimes C_{t-1} + b_f) \quad (3)$$

$$I_t = \sigma(W_{XI} \otimes X_t + W_{HI} * H_{t-1} + W_{CI} \otimes C_{t-1} + b_i) \quad (4)$$

$$O_t = \sigma(W_{XO} \otimes X_t + W_{HO} \otimes H_{t-1} + W_{CO} \otimes C_{t-1} + b_o) \quad (5)$$

$$C_t = F_t \times C_{t-1} + I_t \times \tanh(W_{XC} \otimes X_t + W_{HC} \otimes H_{t-1} + b_c) \quad (6)$$

$$H_t = O_t \times \tanh(C_t) \quad (7)$$

Where ' $\otimes$ ' indicates convolution, ' $\times$ ' indicates Hadamard product, W_{CF}, W_{CI}, W_{CO} and W_{HC} and b_f, b_i, b_o and b_c are the weight matrices and bias vectors will be updated in each update process. The ELM is combined with conLSTM to enhance the classification accuracy and eliminate the computational time. In the given V distinct training images samples from conLSTM is given in Eq. (8),

$$V = (X_z, T_z) | X_z \in G^e, T_z \in G^e, z = 1, \dots, n \quad (8)$$

The output operation of ELM based on Single Layer Feedforward Neural Networks (SLFNN) for updating the h hidden units and $j(X)$ as an initiation operation which is signified as follows,

$$O_\ell(X) = \sum_{z=1}^h \alpha_z J(i_z, b_z, x_y), y = 1, 2, 3 \dots n \quad (9)$$

In the preceding Eq. (9) $O_\ell(X)$ represents the resultant vector of SLFNN in reference to the input occurrences. The learning factors i_z and b_z will be calculated randomly at the hidden layers. The input weight vector i_z is connected with the z^{th} hidden layer and input block. b_z represents the bias of the z^{th} - hidden unit. The stimulation mechanism of the ELM layer in conLSTM is $J(i_z, b_z, x_y)$ where α_z connects the resultant weight to the z^{th} - hidden and output terminals. The Eq. (10) is compactly given as

$$h\alpha = Q \quad (10)$$

Where, h represents the generated vector of hidden layer Eq. (11), Q and α denotes the desired outcome and its corresponding output weights respectively represented in Eqs. (12) and (13)

$$h = \begin{bmatrix} J(i_1, b_1, x_1) & \cdots & J(i_h, b_h, x_1) \\ \vdots & \ddots & \vdots \\ J(i_1, b_1, x_V) & \cdots & J(i_h, b_h, x_N) \end{bmatrix}_{V \times \ell} \quad (11)$$

$$\alpha = \begin{bmatrix} \alpha_1^\delta \\ \vdots \\ \alpha_1^\delta \end{bmatrix}_{\ell \times V} \quad (12)$$

$$Q = \begin{bmatrix} T_1^\delta \\ \vdots \\ T_V^\delta \end{bmatrix}_{\ell \times e} \quad (13)$$

However, ELM selects the hidden node variables (e.g., i_z, b_z) arbitrarily and reduces the cost operation ($O_\ell(X) - Q$). From a linear algebraic perspective, Eq. (14) resembles the quadratic calculation with the outcome weights α can be determined statistically using a minimal-squares approach.

$$\alpha = h^\dagger Q \quad (14)$$

In Eq. (14), $h^\dagger$ employs the Moore-Penrose modified opposite of matrix h and matrix $Q = [q_1, q_2, \dots, q_n]^Q$ to determine the resultant weights, preserving training period by eliminating repetitive variable modifications with appropriate training variables such as learning speed and repetitions.

The outermost layer of conLSTM is used to generate image integration using ELM, which can be expressed as image vectors. The $4 \times 4 \times 512$ output map from conLSTM is flattened into a $1 \times 1 \times 512$ image vector that is sent to the ELM, allowing the model's efficiency to be optimized despite using the typical FC layer. The layer allocation of CovLscan is 408 layers of E-conLSTM-ELM, 1 layer of flattening and 8 levels of ELM. The E-conLSTM-ELM design connects the LSTM convolution and memory layers with Rectified Linear Unit (ReLU) stimulation and max-pooling layers in a sequential manner, as shown in Eq. (15).

$$X_y^C = F \left(\sum_{z \in J_y} X \frac{C-1}{z} \otimes \mathcal{K}_{zy}^C + b_y^C \right) \quad (15)$$

The resultant feature map is computed using Eq. (12), which $\frac{C-1}{z} i$ includes the local attributes from

preceding layer. The elements J_y , $F(\cdot)b_y^C$ and $\mathcal{K}_{zy}^C$ defines the input map decisions, stimulation operation, training bias and parametric kernels. The quadratic ReLU procedure is utilized to enable the CNN layers which is defines in Eq. (16) which improves training efficacy.

$$F(X) = \max(0, X) \quad (16)$$

Where X is the the node's input stimulation and $F(X)$ represents the resultant stimulation. A pooling layer is employed to minimize overfitting while minimizing computing nodes and its computation strain as illustrated in Eq. (17).

$$X_y^C = \text{down}(X_y^{C-1}) \quad (17)$$

In Eq. (17), $\text{down}(\cdot)$ demonstrates down-sampling, X_y^{C-1} indicates local attributes from prior layers and X_y^C denotes the output stimulation of following layers accompanied with conLSTM. To transfer learning weights, compute the variation between original and target domain sites and select the closest source domain position. This strategy improves the reliability of conLSTM model. MMD is used to divide Kernel Hilbert Space (KHS) to determine mean discrepancy by deducting the mean operation of each sample. The mean discrepancy between two instances can be calculated by subtracting the mean operation of each sample, often using the square shape for efficiency. Eq. (18) represents the actual domain data in a specific source realm.

$$D_S = (a_1, a_2, \dots, a_m) \quad (18)$$

Where a and m denotes the actual domain site data and its number correspondingly. Eq. (19) represents the expected source information within the target domain.

$$D_T = (y_1, y_2, \dots, y_n) \quad (19)$$

Where y indicates the desired range and n symbolizes data integer. The negative projection functions in the KHS of the regeneration kernel are known as μ . Eq. (20) describes the square representation of MMD.

$$MMD_{HS}^2(D_S, D_T) = \left\| \frac{1}{n} \sum_{i=1}^m \mu(a_i) - \frac{1}{n} \sum_{i=1}^n \mu(y_i) \right\|^2 \quad (20)$$

The variations in dispersion determines the proximity among two data allocations, with adjacent domains having a lower MMD value. MMD is employed in TL to choose the best appropriate source domain site for transferring to the destination domain based on resemblance. The mapping operation $M^i(\cdot)$ represents the i^{th} layer of E-ConLSTM-ELM. CXR and CT imaging data are used to create depictions aspects of the source (χ_s) and target (χ_T) which is formulated in Eqs. (21) and (22),

$$M_S = m^N(\dots M^1(\chi_s)) \quad (21)$$

$$M_T = m^N(\dots M^1(\chi_T)) \quad (22)$$

M_S and M_T are the resultant feature representations of the two image domains obtained by E-ConLSTM-ELM layer. The MMD is also employed in this classification assignment to enforce the extracted features constraints during TL task. Since, the TL is also applied in E-ConLSTM-ELM classification part, the constructed MMD in Eq. (23).

$$D_i^S = m_{2d}^i \left(m_{2d}^{i-1} \left(\dots m_{2d}^1(M_S) \right) \right) \quad (23)$$

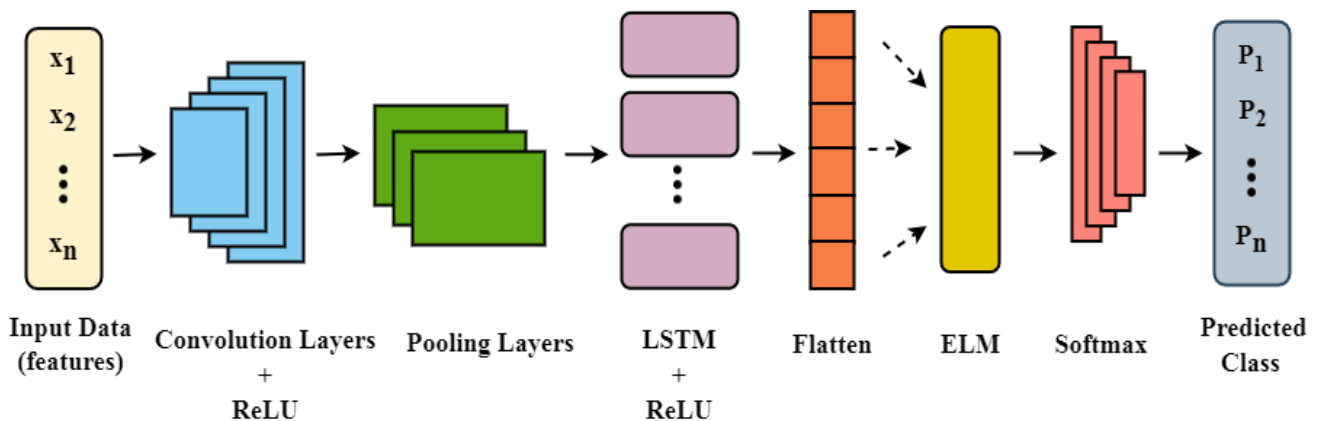


Figure. 2 Block Structure of E-conLSTM-ELM architecture

Eq. (23) is re-modified according to the layers and is represented in Eq. (24). Assume, M_{2d}^i denote the i^{th} layer of E-ConLSTM-ELM for second domain $2d$, then the transfer loss L_{MMD} from the feature extraction of MMD can be computed as represented in Eq. (24).

$$D_i^T = m_{2d}^i \left(m_{2d}^{i-1} \left(\dots m_{2d}^1(M_T) \right) \right) \quad (24)$$

It is shown that the original and desired domains in conLSTM shares the identical attributes of $M^i(.)$ and $M_{2d}^i(.)$ in which the parameters are independent to classify the extracted CXR and CT is in Eq. (25)

$$L_{MMD} = \sum_{i=1}^N d_{MMD}^2(D_i^S, D_i^T) \quad (25)$$

Finally, the reshaped output features are fed into softmax layer to classify different types of LD. Thus, the constructed framework resolve the epistemic uncertainty issue for immediate detection of various diseases utilizing CXR images. Fig. 2 show the E-ConLSTM-ELM layout.

4. Results and discussion

4.1 Dataset description

ChestX-ray8 [21] contains 108,948 frontal-view X-ray images of 32,717 distinct patients gathered from 1992 to 2015 with the text-mined eight frequent condition labels extracted from the text radiography reports using NLP algorithms. This dataset consists of eight classes like Atelectasis, Cardiomegaly, Effusion, Infiltration, Mass, Nodule, Pneumonia and Pneumothorax.

NIH Chest X-Ray [22] dataset is comprised of 112,120 X-ray images with disease labels from 30,805 unique patients. This dataset constitutes of 14 classes like Atelectasis, Cardiomegaly, Effusion, Infiltration, Mass, Nodule, Pneumonia, Pneumothorax, Consolidation, Edema, Emphysema, Fibrosis, Pleural_Thickening and Hernia. Additionally, the information of COVID-19 and Non-Covid (normal) is obtained from [25], which contains 6432 CXR images are taken along with these two datasets.

Totally, five classes (Covid-19, Pneumonia, Normal, Infiltrate and Atelectasis) are listed for the experimental purposes. In order to improve the pulmonary disease detection from CXR images for proposed model trained models generated from CT images are utilized. CT images are collected from various sources [24-28]. The same classes utilized in

CXR dataset are taken from the CT images. For the evaluation, only CXR datasets are used.

4.2 Experimental setup and performance evaluation

The implementation of both proposed and existing model is executed on a system with MATLAB 2019B using the datasets illustrated in section 4.1. The collected datasets are individually divided into 70% for training and 30% for testing. Table 3 depicts the parameter configuration of both existing and proposed model. A comparative study is presented between CovLscan and existing pulmonary diseases models like AHDL-LDD [10], Efficient Net v2-M [14], LungNet22 [15], CDCNet [19] and ABOA-CNN [20]. These proposed and current models are assessed using accuracy, precision and recall concisely described below.

Accuracy: It is the ratio of properly classified instances for every LD classes to the total number of instances evaluated.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (26)$$

In above Eq. (26), True Positive (TP) indicates the model correctly labels LD categories, for example, Infiltrate is categorized as Infiltrate, while True Negative (TN) indicates the predictor incorrectly identifies classes, for example Infiltrate as other categories. False Positive (FP) indicates the algorithm correctly classified other categories as others (other than Infiltrate) and False Negative (FN) indicates the model incorrectly predict others as Infiltrate. The same definition is applicable for all categories of diseases. The average values are calculated finally for all classes.

Precision: It represents the proportion of correctly classified instances of LD classes at TP and FP incidences. It is shown in Eq. (27),

$$Precision = \frac{TP}{TP+FP} \quad (27)$$

Recall: It is the proportion of precisely identified classified LD instances at TP and FN occurrences in Eq. (28),

$$Recall = \frac{TP}{TP + FN} \quad (28)$$

F1-Score: It is stated as the cumulative mean of precision and recall as in Eq. (28), where '1' is the highest and '0' is the lowest potential number.

Table 3. Parameter Configuration for Proposed and Existing Models

Model	Parameters	Range
AHDL-LDD [10]	No. of Convolutional (Conv) layers	3
	Stride	2
	Optimizer	Adam
	Activation Function	ReLU
	Batch Size	64
	No. of Epochs	200
	Weight Decay	0.0001
	Loss Function	BCE
	Learning Rate	0.0001
Efficient Net v2-M [14]	Input Layer	4
	Efficientnetv2-m layer	19
	Optimizer	Adam
	Activation Function	Sigmoid
	Dropout rate	0.4
	Batch Size	8
	No. of Epochs	50
	Loss Function	Categorical CE (CCE)
LungNet 22 [15]	No. of Conv layers	3
	No. of. Maxpooling Layer	2
	Stride	1
	Optimizer	Adam
	Activation Function	ReLU
	Batch Size	128
	No. of Epochs	300
	Loss Function	CCE
	Learning Rate	0.000001
CDCNet [19]	No. of. Conv Layer	2
	No. of Dense Layer	2
	Optimizer	SGD
	Activation Function	ReLU
	Batch Size	64
	No. of Epochs	500
	Loss Function	Mean Square Error (MSE)
	Learning Rate	0.1
ABOA-CNN [20]	Input layer	3
	Convolution Kernels	3
	Filters	2
	Optimizer	SGD
	Activation Function	Tanh
	Batch Size	120
	No. of Epochs	350
	Loss Function	Cross entropy
	Learning Rate	0.01
Proposed model	Number (No). of. U-Net Layers	23 Convolutional Layers with contracting and expansive layers
	Stride	2
	No. of. Layers -	121

DenseNet;	
No. of. Layers - InceptionResNetV2	164
No. of. Layers - Xception	71
E-conLSTM-ELM layers	
Input layer (feature)	2
Input dimension	192×192×3
Conv Layer	64
Kernel size	3
Activation Function	ReLU (64)
Batch Normalization	64
Average + Max Pooling Layers	15 (8+7)
Number of LSTM layer hidden unit	98
Time Step	6
Output dimension	4 ×4× 512
Flatten	1
Flatten Input	192×192×3
Flatten output	4×4×512
ELM Input Layer	2
ELM Hidden Layer	4
ELM Output Layer	2
ELM Input	1×1×512
ELM Output	5
Optimizer	Adam
Dropout rate	0.6
Activation Function	ReLU
Batch Size	64
No. of Epochs	100
Weight Decay	0.0005
Loss Function	BCE
Learning Rate	0.001

$$F1 - score = 2 \times \frac{Precision \cdot Recall}{Precision + Recall} \quad (29)$$

Area Under Curve (AUC): The AUC Score ranges from 0 to 1 by drawing the Receiver Operating Characteristic (ROC) curve, which compares the TP rate (TPR) to FP Rate (FPR) for every possible cut-off point of a diagnostic test.

Figs. 3 and 4 displays the accuracy (in %) achieved by AHDL-LDD, EfficientNet v2-M, LungNet22, CDCNet ABOA-CNN and CovLscan for diagnosing various LD categories like Covid-19, pneumonia, normal, infiltrate and atelectasis. The investigation shows that CovLscan excels than other models on two datasets. For example, in the pneumonia categorization, accuracy of CovLscan is 15.76%, 12.72%, 8.33%, 5.49% and 2.09% (for ChestX-ray8); 15.83%, 12.55%, 7.84%, 5.11% and 2.86% (for NIH-CXR) which is greater than other existing models respectively.

Figs. 5 and 6 shows the precision (in %) attained by proposed and existing models.

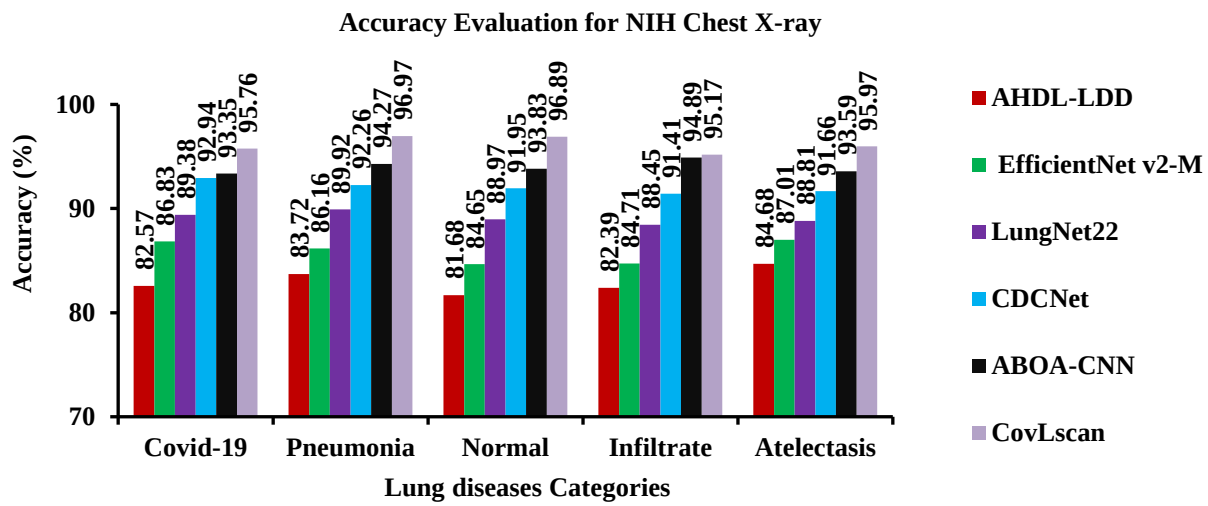


Figure. 4 Accuracy Comparison of LD category prediction models for NIH Chest X-ray

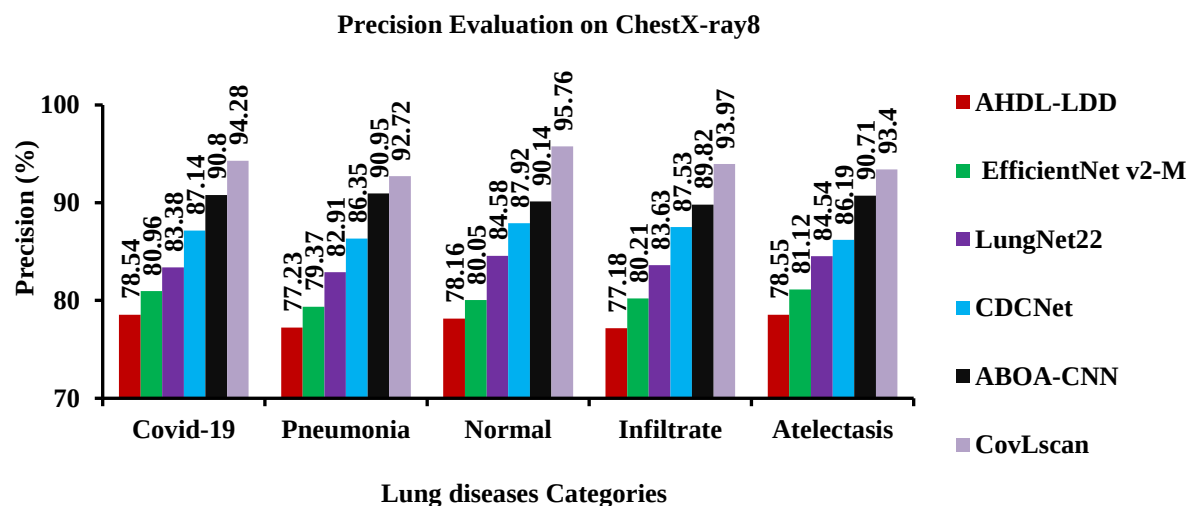


Figure. 5 Precision Comparison of LD category prediction models for ChestX-ray8

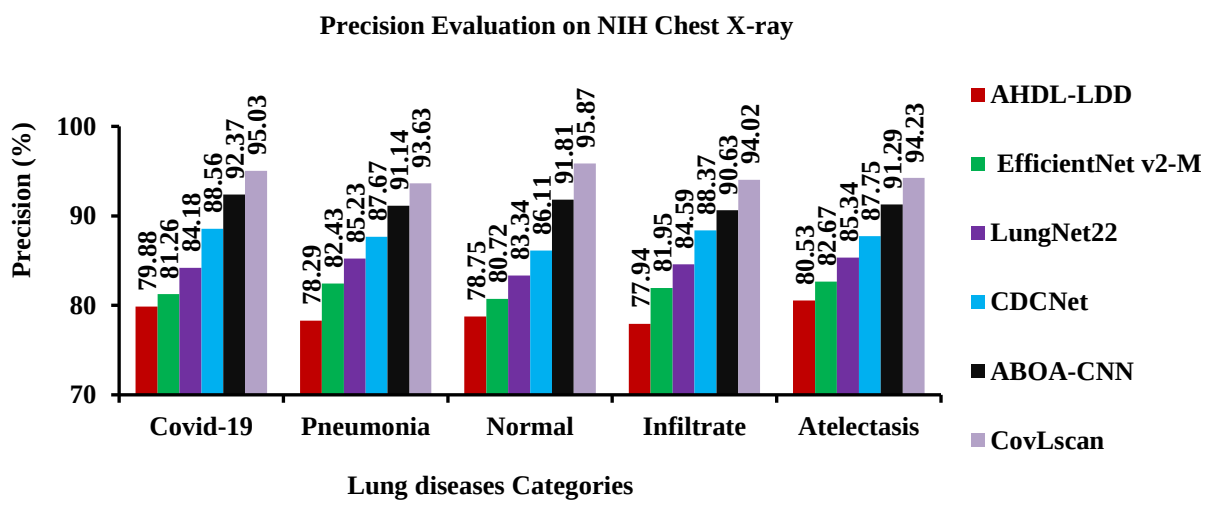


Figure. 6 Precision Comparison of LD category prediction models for NIH Chest X-ray

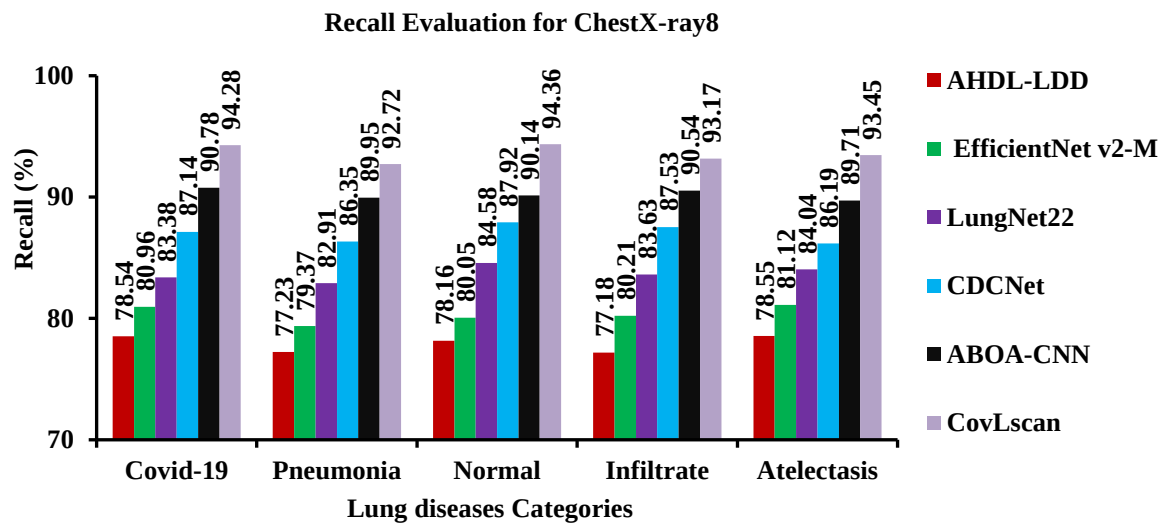


Figure. 7 Recall Comparison of LD category prediction models for ChestX-ray8

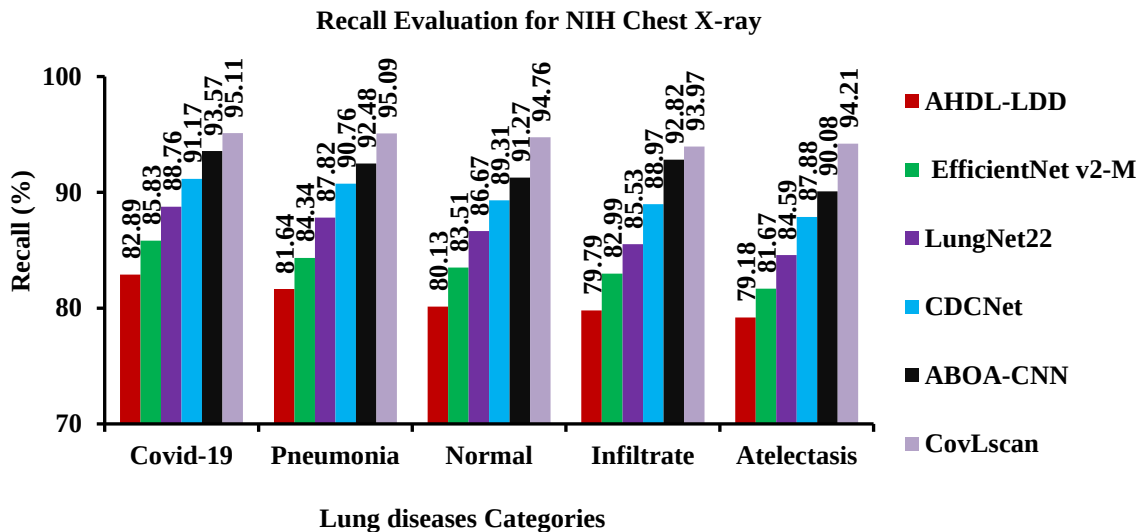


Figure. 8 Recall Comparison of LD category prediction models for NIH Chest X-ray

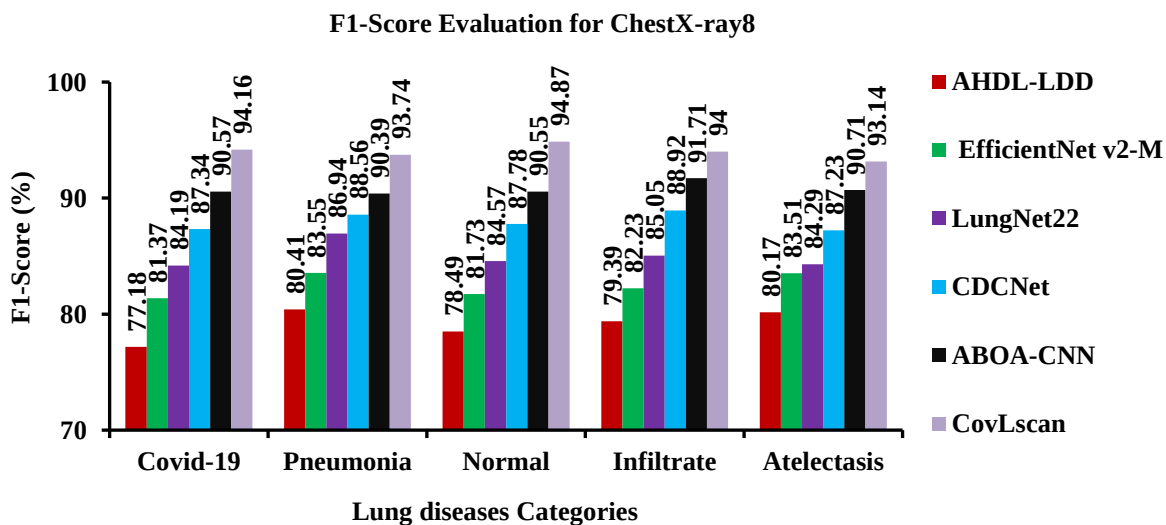


Figure. 9 F1-Score Comparison of LD category prediction models for ChestX-ray8

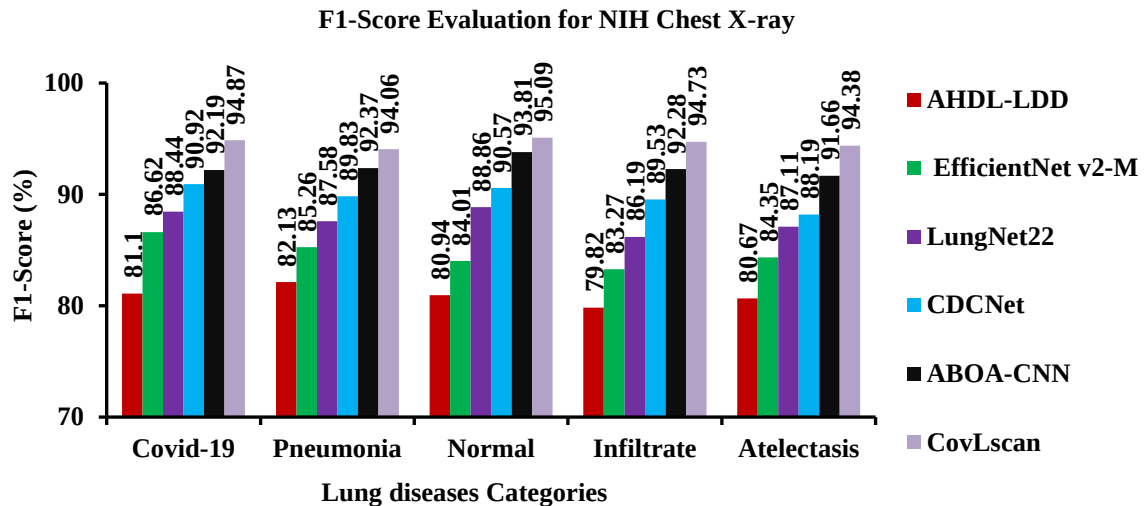


Figure. 10 F1-Score Comparison of LD category prediction models for NIH Chest X-ray

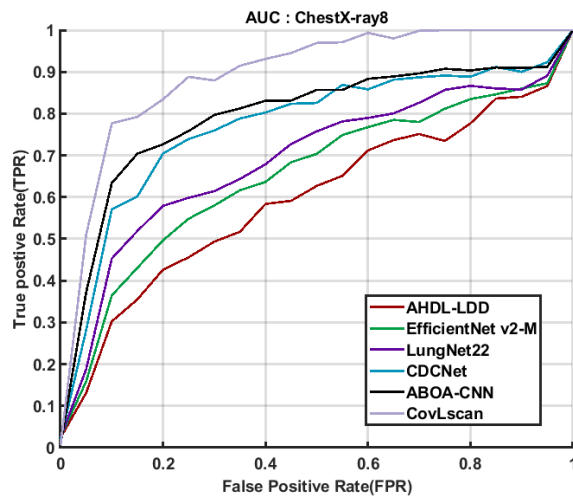


Figure. 12 AUC Comparison of LD category prediction models for ChestX-ray8

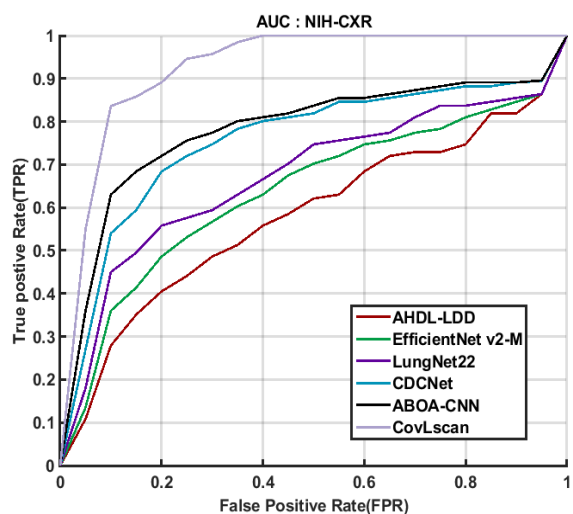


Figure. 12 AUC Comparison of LD category prediction models for NIH Chest X-ray

This analysis obtains that the proposed CovLscan model outperforms other models in predicting each LD categories using two CXR datasets. For example, in the case of Covid-19 classification, the precision of CovLscan is 20.04% and 18.97% greater than AHDL-LDD, 16.45% and 16.95% higher than EfficientNet v2-M; 13.07% and 12.89% greater than LungNet22; 8.19% and 7.31% greater than CDCNet; 3.83% and 2.88% higher than ABOA-CNN models for ChestX-ray8 and NIH-CXR respectively.

Figs. 7 and 8 depicted the recall (in %) obtained by existing models. It is determined that the recall of CovLscan for each LD category is superior to that of other classification models. For example, in the case of infiltrate classification, the recall of CovLscan is 20.72%, 16.16%, 11.41%, 6.44% and 2.90% (for ChestX-ray8); 17.77%, 13.23%, 9.87%, 5.62% and 1.24% (for NIH-CXR) is higher than AHDL-LDD, EfficientNet v2-M, LungNet22, CDCNet ABOA-CNN respectively.

Figs. 9 and 10 provides the F1-score (in %) obtained by existing models for diagnosing various LDs categories using two datasets correspondingly. It is determined that the F1-score of CovLscan for each LD category is superior than other models. For example, in the case of normal classification, F1-score of CovLscan is 20.87% and 17.48% greater than AHDL-LDD, 16.08% and 13.19% greater than Efficient Net v2-M, 12.18% and 7.01% greater than LungNet22, 8.08% and 4.99% greater than CDCNet, 4.77% and 1.36% greater than ABOA-CNN for ChestX-ray8 and NIH-CXR respectively.

Figs. 11 and 12 provides AUC obtained by existing models for diagnosing various LDs categories using two datasets correspondingly. It indicates that the AUC of CovLscan is 0.94% and % greater than AHD-LDD, 0.83% and 0.79% greater than EfficientNet v2-M, 0.72% and 0.61% greater than LungNet22, 0.61% and 0.54% greater than CDCNet, 0.5% and 0.47% greater than ABOA-CNN for ChestX-ray8 and NIH-CXR respectively.

In the literature, AHD-LDD [12], EfficientNet v2-M [16], LungNet22 [17], CDCNet [21] and ABOA-CNN [22] model have utilized NIH-CXR dataset for the evaluation. In this model, ChestX-ray8 have considered for the performance task. Hence, this work evaluates proposed and existing models on both ChestX-ray8 and NIH-CXR datasets by using the parameters as per Table 3. From the above comparison, it is proved that the proposed CovLscan model obtains efficient results on both ChestX-ray8 and NIH-CXR datasets for the classification lung cancer and its categories.

5. Conclusion

In this article, CovLscan model is created to reduce the ambiguity concerns and improve the LD categorization efficiency. This method segments the collected images using U-Net model and pre-trained CNN models for feature extraction. The con-LSTM is used for the classification task. The ELM is applied in conLSTM to reduce computational time and improve accuracy. TL model is used to exchange knowledge from CT features and classes to CXR feature learning. The domain adaptation strategy reduces domain shift using MMD for efficient classification. CovLscan model achieves an overall accuracy of 95.46% and 96.15% on ChestX-ray8 and NIH-CXR datasets, which is greater than AHD-LDD, EfficientNetv2-M, LungNet22, CDCNet and ABOA-CNN. In future, Multi-Scale Generative Adversarial Network (MS-GAN) model will be developed to improve the CovLscan by increasing labeled LD images and generating high-quality target images from source modal images facilitating efficient LD prediction.

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

Conceptualization, methodology, software, validation, Kousiga; formal analysis, investigation, Nithya; resources, data curation, writing—original

draft preparation, Kousiga; writing—review and editing, Kousiga; visualization, supervision, Nithya;

References

- [1] D. D. Marciniuk, and D. E. Schraufnagel, "The global impact of respiratory disease", *European Respiratory Society*, Vol. 161, No. 5, pp. 1153-1154, 2017.
- [2] S. Safiri, K. Carson-Chahhoud, M. Noori, S. A. Nejadghaderi, M. J. Sullman, J. A. Heris, K. Ansarin, M. A. Mansournia, G. S Collins, A. A. Kolahi and J. S. Kaufman, "Burden of chronic obstructive pulmonary disease and its attributable risk factors in 204 countries and territories, 1990-2019: results from the Global Burden of Disease Study 2019", *BMJ*, Vol. 373, pp. 1-13, 2021.
- [3] N. S. Hanafi, D. Agarwal, S. Chippagiri, E. A. Brakema, H. Pinnock, A. Sheikh, S. Liew, C. Ng, R. Isaac, K. Chinna, L. P.Wong, N. Hussein, A.I.A Bakar, Y. Pang, S. Juvekar and E. M. Khoo, "Chronic respiratory disease surveys in adults in low-and middle-income countries: A systematic scoping review of methodological approaches and outcomes", *Journal of Global Health*, Vol. 11, No. 1, pp. 1-11, 2021.
- [4] T. Jung and N. Vij, "Early diagnosis and real-time monitoring of regional lung function changes to prevent chronic obstructive pulmonary disease progression to severe emphysema", *Journal of Clinical Medicine*, Vol. 10, No. 24, pp. 1-18, 2021.
- [5] P. L. Saha, S. A. Nadeem and A. P. Comellas, "A survey on artificial intelligence in pulmonary imaging", *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, Vol. 13, No. 5, pp. 1-36, 2023.
- [6] J. Ma, Y. Song, X. Tian, Y. Hua, R. Zhang and J. Wu, "Survey on deep learning for pulmonary medical imaging", *Frontiers of medicine*, Vol. 14, No. 4, pp. 450-469, 2020.
- [7] E. Benmalek, J. Elmhamdi and A. Jilbab, "Comparing CT scan and chest X-ray imaging for COVID-19 diagnosis", *Biomedical Engineering Advances*, Vol. 1, pp. 1-7, 2021.
- [8] N. Arunkumar and G. Ramirez, "Optimal deep learning model for classification of lung cancer on CT images", *Future Generation Computer Systems*, Vol. 92, No. 2, pp. 374-382, 2019.
- [9] A. Asuntha and A. Srinivasan, "Deep learning for lung Cancer detection and classification", *Multimedia Tools and Applications*, Vol. 79, No. 6, pp. 7731-7762, 2020.

- [10] S. Z. Y. Zaidi, M. U. Akram, A. Jameel and N. S. Alghamdi, "Lung segmentation-based pulmonary disease classification using deep neural networks", *IEEE Access*, Vol. 9, No. 1, pp. 125202-125214, 2021.
- [11] K. Ann, Y. Jang, H. Shim and H. J. Chang, "Multi-Scale Conditional Generative Adversarial Network for Small-Sized Lung Nodules Using Class Activation Region Influence Maximization", *IEEE Access*, Vol. 9, No. 1, pp. 139426-139437, 2021.
- [12] P. Vieira, O. Sousa, D. Magalhães, R. Rabêlo and R. Silva, "Detecting pulmonary diseases using deep features in X-ray images", *Pattern Recognition*, Vol. 119, pp. 1-13, 2021.
- [13] V. Ravi, V. Acharya and M. Alazab, "A multichannel EfficientNet deep learning-based stacking ensemble approach for lung disease detection using chest X-ray images", *Cluster Computing*, Vol. 26, No. 2, pp. 1181-1203, 2022.
- [14] S. Kim, B. Rim, S. Choi, A. Lee, S. Min and M. Hong, "Deep learning in multi-class lung diseases classification on chest X-ray images", *Diagnostics*, Vol. 12, No. 4, pp. 1-24, 2022.
- [15] F. J. M. Shamrat, S. Azam, A. Karim, R. Islam, Z. Tasnim, P. Ghosh and F. De Boer, "LungNet22: A Fine-Tuned Model for Multiclass Classification and Prediction of Lung Disease Using X-ray Images", *Journal of Personalized Medicine*, Vol. 12, No. 5, pp. 1-29, 2022.
- [16] G. M. M. Alshmrani, Q. Ni, R. Jiang, H. Pervaiz and N. M. Elshennawy, "A deep learning architecture for multi-class lung diseases classification using chest X-ray (CXR) images", *Alexandria Engineering Journal*, Vol. 64, No. 3, pp. 923-935, 2023.
- [17] A. M. Q. Farhan and S. Yang, "Automatic lung disease classification from the chest X-ray images using hybrid deep learning algorithm", *Multimedia Tools and Applications*, Vol. 82, No. 25, pp. 38561-38587, 2023.
- [18] F. B. Alam, P. Podder and M. R. H. Mondal, "RVCNet: A hybrid deep neural network framework for the diagnosis of lung diseases", *Plos one*, Vol. 18, No. 12, pp. 1-24, 2023.
- [19] H. Malik, T. Anees, M. Din, and A. Naeem, "CDC_Net: Multi-classification convolutional neural network model for detection of COVID-19, pneumothorax, pneumonia, lung Cancer, and tuberculosis using chest X-rays", *Multimedia Tools and Applications*, Vol. 82, No. 9, pp. 13855-13880, 2023.
- [20] B. Annamalai, P. Saravanan, and I. Varadharajan, "ABOA-CNN: auction-based optimization algorithm with convolutional neural network for pulmonary disease prediction", *Neural Computing and Applications*, Vol. 35, No. 10, pp. 7463-7474, 2023.
- [21] <https://opendatalab.com/OpenDataLab/ChestX-ray8>
- [22] <https://www.kaggle.com/datasets/nih-chest-xrays/data>
- [23] <https://www.kaggle.com/datasets/prashant268/chest-xray-covid19-pneumonia>
- [24] <https://radiopaedia.org/articles/lung-atelectasis?lang=us>
- [25] <https://www.kaggle.com/datasets/mehradaria/covid19-lung-ct-scans>
- [26] <https://radiopaedia.org/articles/viral-respiratory-tract-infection>
- [27] <https://radiopaedia.org/playlists/41156?lang=us>