

# Lung-GANs: Unsupervised Representation Learning for Lung Disease Classification Using Chest CT and X-Ray Images

Pooja Yadav , Neeraj Menon , Vinayakumar Ravi , and Sowmya Vishvanathan 

**Abstract**—Lung diseases are a tremendous challenge to the health and life of people globally, accounting for 5 out of 30 most common causes of death. Early diagnosis is crucial to help in faster recovery and improve long-term survival rates. Deep learning techniques offer a great promise for automated, fast, and reliable detection of lung diseases from medical images. Specifically, convolutional neural networks have accomplished encouraging results in disease detection. In spite of that, the performance of such supervised models depends heavily on the availability of large labeled data, the collection of which is an expensive and tedious task, specially for a novel disease. Therefore, in this article, we propose a deep unsupervised framework to classify lung diseases from chest CT and X-ray images. Our framework introduces multiple-layer generative adversarial networks called Lung-GANs that learn interpretable representations of lung disease images using only unlabeled data. We use the lung features learned by the model to train a support vector machine and a stacking classifier. We demonstrate through experiments that the proposed method outperforms the current state-of-the-art unsupervised models in lung disease classification. Our model obtained an accuracy of 94%–99.5% on all the six large-scale publicly available lung disease datasets used in this study. Hence, the proposed framework will simplify lung disease detection by reducing the time for diagnosis and increasing the convenience of diagnostics.

**Index Terms**—COVID-19, CT scan, generative adversarial networks, lung disease, pediatric pneumonia, pneumonia, tuberculosis, unsupervised representation learning, X-ray.

## I. INTRODUCTION

LUNG diseases (or respiratory diseases) are among the principal causes of death and disability worldwide. The most common lung diseases include tuberculosis, pneumonia, and the recent Coronavirus Disease 2019 (COVID-19). As stated in the Global Impact of Respiratory Disease report by the Forum

Manuscript received 31 May 2021; revised 14 July 2021; accepted 5 August 2021. Date of publication 30 August 2021; date of current version 16 June 2023. Review of this manuscript was arranged by Department Editor N. Gerdts. (*Corresponding author: Pooja Yadav*)

Pooja Yadav and Neeraj Menon are with the VMware, Bangalore 560076, India (e-mail: shanmukhiyadav@gmail.com; neerajmenons@gmail.com).

Vinayakumar Ravi is with the Center for Artificial Intelligence, Prince Mohammad Bin Fahd University, Khobar 34754, Saudi Arabia (e-mail: vravi@pmu.edu.sa).

Sowmya Vishvanathan is with the Center for Computational Engineering and Networking (CEN), Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore 641112, India (e-mail: v\_sowmya@cb.amrita.edu).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TEM.2021.3103334>.

Digital Object Identifier 10.1109/TEM.2021.3103334

of International Respiratory Societies,<sup>1</sup> more than 10 million people suffer and 1.4 million die from tuberculosis (TB) every year. Pneumonia also claims millions of lives annually and is the largest cause of mortality in children under the age of five. Since late December 2019, the COVID-19 pandemic has engendered over 3.51 million deaths globally, making it one of the most lethal pandemics in history. It is evident that lung diseases impose an immense health burden worldwide.

Respiratory diseases account for 5 out of 30 most common causes of death; therefore, their prevention, control, and cure are paramount [1]. Early diagnosis is needed to help in faster recovery and improve survival rates. Traditionally, chest radiographs (X-rays) and computed tomography (CT) scans are used for lung disease detection. Chest X-rays are inexpensive, easy to operate, widely available, and more time-efficient than CT. They contain an enormous amount of information about the patient's health. Despite these advantages, interpreting them remains a challenge. Even experienced radiologists have trouble distinguishing between similar lesions or detecting obscure nodules. Moreover, manual screening of lung diseases is time-consuming, labor-intensive, and suffers from inter-observer and intra-observer differences [2]. In addition, the shortage of radiologists and the growing number of lung-infected patients may delay diagnosis and treatment. To solve these problems, especially under the ongoing crisis caused by the COVID-19 pandemic, it has become crucial to develop an effective computer-aided diagnosis (CAD) system for lung disease screening that reduces the time for diagnosis and increases the convenience of diagnostics [3]. This motivated us to present a solution that would play a pivotal role in diagnosing lung diseases by enabling experienced radiologists to reduce their workload and assist the less experienced physicians in accurate and timely decision-making.

Recent developments in deep learning (DL) have shown vast potential in the automatic identification and classification of patterns in medical images [4]. Specifically, convolutional neural networks (CNNs) have achieved remarkable performance in disease detection, including lung diseases. However, these models are heavily dependent on large amounts of labeled training data or fine-tuning of millions of parameters from pretrained CNNs. Moreover, the quantity of annotated data available in the medical imaging domain is limited, more so for a new disease. Although

<sup>1</sup>[Online]. Available: [https://www.who.int/gard/publications/The\\_Global\\_Impact\\_of\\_Respiratory\\_Disease.pdf](https://www.who.int/gard/publications/The_Global_Impact_of_Respiratory_Disease.pdf)

auto-annotation platforms are available for lung diseases, the efficiency and accuracy of these systems are less [5], and they are still in the initial stage of research. These issues limit the performance of these supervised learning models on new data. Conversely, unsupervised representational learning algorithms match the performance of supervised models using only unlabeled data. Autoencoder is one such unsupervised learning algorithm in which the encoder takes input data and outputs a compressed representation of the input. The decoder then uses this representation to reconstruct the input data. However, these models perform poorly in image generation as the reconstructed images are often blurry and lower in quality due to compression. Generative adversarial networks (GANs) [6] is another unsupervised technique that generates unseen image samples without supervision using a two-player min–max game and extends a training dataset artificially. Recently, a variant of GANs called the deep convolutional GANs achieved promising performance in image synthesis tasks, proving that GANs capture the complex distributions of data and learn good representations of images.

Therefore, in this research, we propose a deep unsupervised framework using a generative adversarial networks model (Lung-GANs) to learn representations of lung disease images using only unlabeled data. The framework would assist clinicians in automated, fast, and reliable detection of lung diseases.

The major contributions of our study are as follows.

- This article, to our knowledge, is the first one to propose a GAN-based single framework to classify unsupervised lung disease images.
- We use the trained discriminator for lung classification tasks, showing significant performance improvement over current leading unsupervised techniques.
- We show detailed investigations and experiments on the proposed method using six large-scale publicly available chest X-ray and CT scan datasets.
- We visualize the features learnt by Lung-GANs to empirically evaluate its effectiveness in classifying lung diseases.

Datasets and code are available at <https://github.com/yadavpa1/Lung-GAN>.

The rest of this article is structured as follows. Section II presents a review of the literature on automated lung disease diagnosis. Section III describes the proposed DL framework along with its implementation details. Datasets descriptions are provided in Section IV. Section V provides the results and its related discussion. Conclusion is presented at the end of the article.

## II. LITERATURE SURVEY

Over the past decades, several CAD systems have been developed to assist doctors in reading medical images and be of potential clinical use [7]. However, developing a reliable CAD system is challenging. The emergence of powerful graphics processing units (GPUs) and DL methods such as CNNs have opened up newer opportunities to build such a system. Abiyev and Ma’aitah [8] employed a backpropagation neural network, a competitive neural network, and a CNN for the diagnosis of chest diseases. Tang *et al.* [9] also used deep CNN architectures like AlexNet, GoogLeNet, VGGNet-16, ResNet-50, and DenseNet

to perform normal versus abnormal chest X-ray classification. Bhandary *et al.* [10] suggested a DL framework to identify lung cancer and pneumonia in CT and X-ray images. Initially, they performed classification using a modified AlexNet with SVM classifier. Later, an ensemble-feature technique is used to improve the classification performance. A DL framework called VDSNet was proposed by Bharati *et al.* [11] that uses CNN to detect lung diseases from X-ray images. Refs. [12] and [13] present a comprehensive overview of various DL-based detection systems for lung diseases available in the literature.

Numerous attempts have been made to detect pneumonia using DL methods. Saraiva *et al.* [14] used CNNs for pneumonia versus normal classification and a k-fold cross-validation for its evaluation. Chouhan *et al.* [15] adopted a transfer learning approach for classifying pneumonia from X-rays. They ensemble five pretrained models to obtain a better performance in pneumonia recognition. Rahman *et al.* [16] also used a transfer learning procedure using CNNs to detect pneumonia. Liang and Zheng [17] too utilized transfer learning and developed a network architecture with residual structures to prevent overfitting. They also introduced dilated convolutions during training to minimize the loss of feature space information. Akgundogdu [18] adopted a two-stage approach for detecting pneumonia in chest X-rays, where the first stage involved feature extraction using a 2-D discrete wavelet transform. And the second stage involved classification using the random forest algorithm. Dey *et al.* [19] proposed a modified VGG19 architecture and an ensemble feature scheme to diagnose pneumonia from chest X-rays.

Since the outbreak of COVID-19, several CAD systems have been proposed for COVID-19 screening using medical images. For instance, Mei *et al.* [20] proposed a CNN-based AI model that combines CT data and clinical information to diagnose COVID-19. Studies such as [21], [22] utilized pre-trained CNNs for COVID-19 diagnosis and later optimized their hyperparameters to improve performance. Apostolopoulos *et al.* [23] trained Mobile Net from scratch to evaluate the significance of features extracted for classification. A semi-supervised framework based on graphs was introduced by Aviles-Rivero *et al.* [24] for COVID-19 analysis using chest X-rays. Wang *et al.* [25] also used chest X-rays to differentiate COVID-19 from pneumonia and localize the lesions. Amyar *et al.* [26] suggested a multitask DL model to spot COVID-19 cases as well as segment COVID-19 lesions from chest CT. They used one decoder for image reconstruction, one decoder for segmentation, a multilayer perceptron for classification, and a common encoder for all the aforementioned tasks. Wang *et al.* [27] integrated two 3-D ResNets into one model with a prior attention mechanism for COVID-19 detection in volumetric CT images. Xu *et al.* [28] applied a deep-CNN to chest CT images for extracting features, which were combined with other modality data through feature late fusion to perform disease detection. Chen *et al.* [29] employed a contrastive learning approach to learn feature representation of chest CT images. Their method is data-efficient, unlike other approaches mentioned above. According to Roberts *et al.* [30], none of the current ML literature methods for COVID-19 diagnosis using CT and X-ray imaging

is of potential clinical use because of their involved biases and flaws in methodologies.

DL-based approaches for TB classification using chest X-rays have achieved significant results in computer-aided diagnostics. Jaeger *et al.* [31] developed an automated system that performs lung segmentation, followed by texture, edge, and shape feature computation, and finally, classification using a pretrained binary classifier. Lakhani and Sundaram [32] evaluated the efficiency of two DCNNs, AlexNet and GoogLeNet, in detecting TB from chest X-rays. Lopes and Valiati [33] and Rahman *et al.* [34] used pretrained CNNs to classify chest X-rays into TB and non-TB. Chandra *et al.* [35] presented a hierarchical feature extraction and classification scheme to detect normal and abnormal X-ray images. Bayesian-based CNN (B-CNN) proposed by Abideen *et al.* [36] used a Bayesian-based CNN for TB identification. Rajaraman and Antani [37] used an ensemble of modality-specific DL models to reduce overfitting, leading to improved accuracy in TB detection. Munadi *et al.* [38] assessed the impact of enhancing images on the performance of DL techniques in TB classification. They demonstrated that the use of image enhancement systems to preprocess TB images would allow the DL models to perform better.

The above-detailed literature survey on lung diseases, such as pneumonia, COVID-19, and tuberculosis, shows that most existing CAD systems use DL-based supervised approaches that require labeled data. Many of these supervised models surpass expert performance in CT and X-ray classification. However, the development of robust DL-based CT and X-ray data analysis requires extensive labeled data, and getting enough data in the medical domain is often difficult, costly, and demands many experts. To address these limitations, we propose an unsupervised representation learning approach to learn optimal features from CT and X-rays without labels. Moreover, most existing studies lack a detailed performance analysis on different lung diseases, which is vital as a proposed system for detection and classification should be generalizable across similar diseases. Thus, we present a detailed experimental analysis of our method on various lung disease datasets.

### III. PROPOSED METHODOLOGY

GANs are most straightforward to apply in cases where both models are multilayer perceptrons. However, since the datasets used in this work consist of complex images like chest CT and X-ray, a deeper network to obtain accurate image representations is required. Therefore, we used the CNN architecture for both our models. The generator is trained to learn a mapping from a random noise vector  $z$  obeying a prior distribution  $p_z(z)$  to output images  $x = G_r(z; \theta^{(G_r)})$  from the data distribution  $p_{\text{data}}(x)$ . Its adversary, the discriminator is trained to differentiate between the real data and the fake data from generator distribution  $p_g$ . It is reused for extracting features during classification. The additional training data provided by the generative model helps it in learning better image representations.

#### A. Objective

The discriminator  $D_r$  is trained to estimate the probability, denoted as  $D_r(x; \theta^{(D_r)})$ , that a sample belongs to real training

data rather than the generator  $G_r$ . In other words, the goal of discriminator is to

- 1) maximize  $D_r(x)$  for all real training images; and
- 2) minimize  $D_r(x)$  for all fake images drawn from the generator.

To summarize, the objective of  $D_r$  is to maximize

$$\mathbb{E}_{x \sim p_{\text{data}}(x)} \log D_r(x) + \mathbb{E}_{z \sim p_z(z)} [\log(1 - D_r(G_r(z)))] \quad (1)$$

which is called as the minimax loss function and is used to estimate the discriminator loss during its training.

The generator  $G_r$  is simultaneously trained to maximize  $D_r(G_r(z))$  or, in other words, to minimize  $\log(1 - D_r(G_r(z)))$ . We call it a perceptual loss which is defined as follows:

$$l_{\text{perceptual}} = \mathbb{E}_{z \sim p_z(z)} [\log(1 - D_r(G_r(z)))] \quad (2)$$

To summarize, the objective of a GAN can be expressed as

$$\begin{aligned} \min_{G_r} \max_{D_r} V(D_r, G_r) &= \mathbb{E}_{x \sim p_{\text{data}}(x)} \log D_r(x) \\ &+ \mathbb{E}_{z \sim p_z(z)} [\log(1 - D_r(G_r(z)))] \end{aligned} \quad (3)$$

For  $G_r$  to produce outputs resembling the real training data, we trained it to equal the expected value of the features present in the multifeature layer of the discriminator. The feature-matching loss of generator is defined as follows:

$$l_{\text{feature\_match}} = \| \mathbb{E}_{x \sim p_{\text{data}}(x)} f(x) - \mathbb{E}_{z \sim p_z(z)} f(G_r(z)) \|_2^2 \quad (4)$$

where  $f(x)$  denotes activations on the multifeature layer of  $D_r$ .

Combining (2) and (4), our overall objective for training the generator is to minimize the final loss, i.e.,

$$l_{\text{final}} = l_{\text{perceptual}} + l_{\text{feature\_match}} \quad (5)$$

That is, sum of perceptual and feature matching loss,  $l_{\text{final}}$  is used to estimate loss during the generator training.

#### B. GAN Architecture

We propose an enhanced version of existing architecture by Lin *et al.* [39]. The existing architecture can only output images of size  $256 \times 256$ . In comparison, our proposed architecture can output lung disease images with a size of  $512 \times 512$  by adding an extra transpose convolutional layer in the generator and a convolutional layer in the discriminator. This improvement lets our method learn multiscale spatial information from higher resolution images and, as a result, achieve better results than other methods. The proposed architecture is shown in Fig. 1.

*1) Generator Architecture:* The generator takes in a 100-dimensional latent noise vector  $z$  drawn from a uniform distribution and outputs a single  $512 \times 512 \times 3$  image. The latent vector is passed through a dense layer to give activations that are reshaped into a  $4 \times 4 \times 1024$  tensor. This tensor is further passed through a series of seven transpose convolutional layers (also called deconvolutional layers) to upsample  $4 \times 4 \times 1024$  feature maps, first to  $8 \times 8 \times 512$ , then  $16 \times 16 \times 256$ , then  $32 \times 32 \times 128$ , then  $64 \times 64 \times 64$ , then  $128 \times 128 \times 32$ , then  $256 \times 256 \times 16$ , and finally to  $512 \times 512 \times 3$ . Each deconvolutional layer (except the last one) has batch normalization as well as an activation layer after it. The model uses a decay factor of

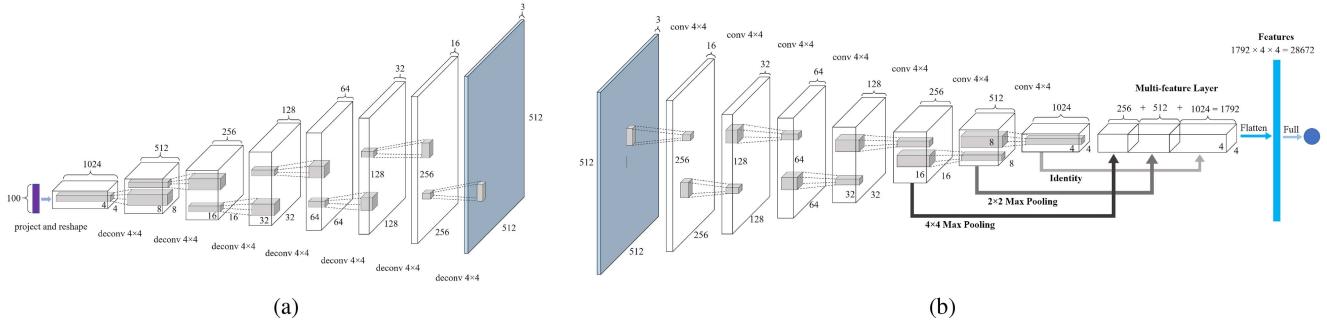


Fig. 1. Network architecture of the proposed Lung-GANs. (a) Generator ( $G_r$ ) takes as input a 100-dimensional uniformly distributed noise vector  $z$  and outputs a  $515 \times 512$  RGB image. (b) Discriminator ( $D_r$ ) extracts hidden features from the  $512 \times 512$  input image and learns to classify between real and synthesized images.

0.9, a hyperbolic tangent (tanh) activation in the output layer and ReLU activation for all other layers. The kernel size and stride are set to 4 and 2, respectively, to avoid the uneven overlaps caused when the latter is indivisible by the former [40]. The total number of parameters in the generator is approximately 12 million. Fig. 1(a) shows the layered architecture of the generator.

2) *Discriminator Architecture*: The discriminator has a CNN architecture that takes in both real and synthesized images of shape  $512 \times 512 \times 3$  as input to output a probability value indicating the probability of an input image being real rather than fake. Each convolutional layer has a batch normalization layer (except the first) and an activation layer after it. The strided convolutions in the model facilitate its learning of its own spatial downsampling. The input is downsampled from  $512 \times 512 \times 3$  to  $256 \times 256 \times 16$ , then  $128 \times 128 \times 32$ , then  $64 \times 64 \times 64$ , then  $32 \times 32 \times 128$ , then  $16 \times 16 \times 256$ , then  $8 \times 8 \times 512$ , and finally to  $4 \times 4 \times 1024$ . Further,  $4 \times 4$  max pooling is performed on the last third convolutional layer and  $2 \times 2$  max-pooling on the last second layer. The identity function is performed on the last convolutional layer to construct  $4 \times 4$  sized feature maps. These maps are concatenated to form a multifeature layer, flattened, and fed into a single sigmoid output. The model uses a kernel size and stride of  $5 \times 5$  and two, respectively. The activation function used is LeakyReLU with a 0.2 slope. The total parameters are approximately 17 million. The model architecture is given in Fig. 1(b).

The features in the multifeature layer of the model serve two purposes: i) classification task ii) in calculating the feature matching loss (4) during the training of generator to evaluate the feature similarity between real and fake images in the multifeature layer.

3) *Training and Implementation Details*: The proposed architecture is trained and tested using a DL and reinforcement learning (RL) library called TensorLayer.<sup>2</sup> Adaptive moment estimation (Adam) optimizer is used as it requires less memory space and is computationally efficient. The hyperparameters used for training are *momentum* = 0.5, *learning\_rate* = 0.0002, and *batch\_size* = 64. The network has approximately 29 million parameters. Both generator and discriminator models were trained by stochastic gradient descent (SGD)

for 100 epochs, but the training was stopped early based on the validation performance. All training and testing processes were performed on a Nvidia Tesla k80 GPU with 12 GB memory.

### C. Classifier Architecture

Even though various machine learning models are available for classification tasks, no single model can be claimed to be superior to others. However, combining multiple such models is an excellent alternative to obtain even better classification accuracies. The main idea is to create an ensemble of accurate and diverse classifiers such that the combination surpasses all the individual classifiers. To achieve this, we employed the scheme of stacked generalization (or stacking) on the features present in the multifeature layer of the trained discriminator for lung disease classification. Stacked generalization is a method that combines heterogeneous classifiers to reduce their biases or minimize their generalization error rates [41].

We created an ensemble of classifiers in which the base learners are trained using random forest and linear support vector classification (SVC). The outputs of the base learners are stacked together and used as input to a meta-classifier. The meta-classifier is generated using logistic regression following a cross-validation like process. The input features are standardized (with zero mean and unit variance) prior to SVC training, which was then performed with *tolerance* = 0.0001, regularization parameter  $C$  = 1.0, *max\_iter* = 1000, and *random\_state* = 42. The parameters used for random forest training are *n\_estimators* = 10 and *random\_state* = 42 and that for logistic regression include *tolerance* = 0.0001, *max\_iter* = 100, and  $C$  = 1.0.

In order to obtain an ensemble that outperforms all its member classifiers, the members must be accurate and diverse, i.e., they make errors at distinct instances. To satisfy this requirement, we choose linear SVC and random forest as our base classifiers. Both these algorithms use different methods to represent the knowledge, and thus the hypothesis space is explored from different perspectives. As a result, when their predictions are combined, the resultant classifier achieves better accuracy than each individual classifier. The classifier architecture is given in Fig. 2.

<sup>2</sup><https://tensorlayer.readthedocs.io/en/latest/>.

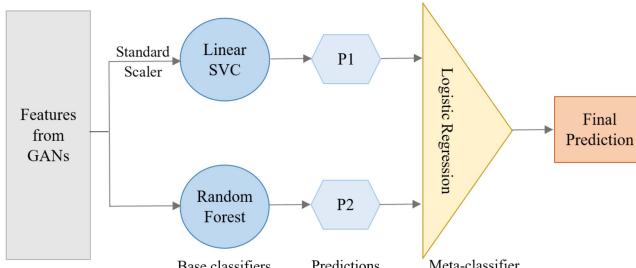


Fig. 2. Stacking Classifier Architecture.

TABLE I  
SUMMARY OF DATASETS

Dataset	Class	Train Set	Test Set	Total
<b>TB vs. Healthy</b>	TB	560	240	800
	Healthy	560	240	800
<b>Healthy vs. Sick</b>	Healthy	1,400	600	2000
	Sick	1,400	600	2000
<b>Pneumonia vs. Normal</b>	Pneumonia	700	300	1,000
	Normal	700	300	1,000
<b>COVID-19 vs. Pneumonia</b>	COVID-19	700	300	1,000
	Pneumonia	700	300	1,000
<b>COVID-19 vs. Non-COVID (CT)</b>	COVID-19	2,160	1,629	3,789
	Non-COVID	1,050	789	1,839
<b>COVID-19 vs. Non-COVID (X-ray)</b>	COVID-19	1,504	1,128	2,632
	Non-COVID	1,850	1,390	3,240

#### IV. DATASETS

This section details the descriptions of the datasets used in this work. The image preprocessing steps taken include color mode conversion, resizing, and normalizing. Since the images in our datasets are in a pixel range of [0255] and have different color modes and sizes, each image is normalized to  $[-1,1]$  range, converted to RGB, and resized to  $512 \times 512$  pixels. We used 70% of the dataset for training and the rest 30% as a testing set. Table I summarizes the class-wise distribution of all datasets. Fig. 3 shows chest X-rays and CT images from all the datasets class-wise. The Python Imaging Library is used for all image processing tasks. More detailed descriptions of datasets are as follows.

##### A. Tuberculosis (TB) vs. Healthy Dataset

We used the tuberculosis X-ray (TBX11 K) dataset provided by Liu *et al.* [42] to perform TB classification. It includes four categories: latent TB, active TB, unhealthy but non-TB, and healthy X-ray images. Each image in the dataset has a resolution of  $512 \times 512$ . We randomly selected 800 images of class label healthy and 800 images of class label TB. The final dataset consists of 1600 X-ray images.

##### B. Healthy vs. Sick Dataset

This dataset is composed of 4000 X-ray images chosen from the TBX11 K dataset [42]. It includes 2000 images belonging to the class label healthy and 2000 images belonging to the

class label sick. All images in the dataset have a resolution of  $512 \times 512$ .

##### C. Pneumonia vs. Normal Dataset

We utilized the dataset provided by Daniel *et al.* [43], which contains 5856 images of frontal chest X-rays of varying resolutions. More precisely, there are 1583 normal case and 4273 pneumonia case images. The radiographic images belong to pediatric patients of 1–5 years old from a Medical Center in Guangzhou. We prepared our dataset with 2000 X-ray images by randomly selecting 1000 normal case and 1000 pneumonia case images.

##### D. COVID-19 vs. Pneumonia Dataset

We used the database provided in [44], [45]. It contains 3616 COVID-19 and 1345 viral pneumonia images. We randomly selected 1000 images of viral pneumonia and 1000 images of COVID-19 to generate a dataset with 2000 X-ray images.

##### E. COVID-19 vs. Non-COVID CT Scan Dataset

We formed this dataset by extracting all CT images from [46] to perform COVID-19 vs. Non-COVID classification. It contains 7555 images, of which 5427 are COVID-19 images, and 2628 are non-COVID images.

##### F. COVID-19 vs. Non-COVID X-Ray Dataset

We generated this dataset using the chest X-ray images contained in [46], which originally consists of a total of 17 099 X-ray and CT images. Our dataset comprises 9544 images of X-ray, of which 4044 are COVID-19, and 5500 are non-COVID images.

## V. EXPERIMENTAL RESULTS

This section validates the effectiveness of our architecture on lung disease classification. We applied the proposed model as a feature extractor and evaluated the performance of a regularized linear L2-SVM classifier and a stacking classifier model (Fig. 2) fitted on top of these features.

##### A. Performance Metrics

The performance of the proposed and existing approaches were evaluated using a confusion matrix. The true positive (TP), true negative (TN), false positive (FP), and false negative (FN) values were taken from the confusion matrix and four performance metrics were computed as follows:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}}. \quad (6)$$

Accuracy is the ability of a classifier to mark all positive cases as positive and all negative cases as negative.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}. \quad (7)$$

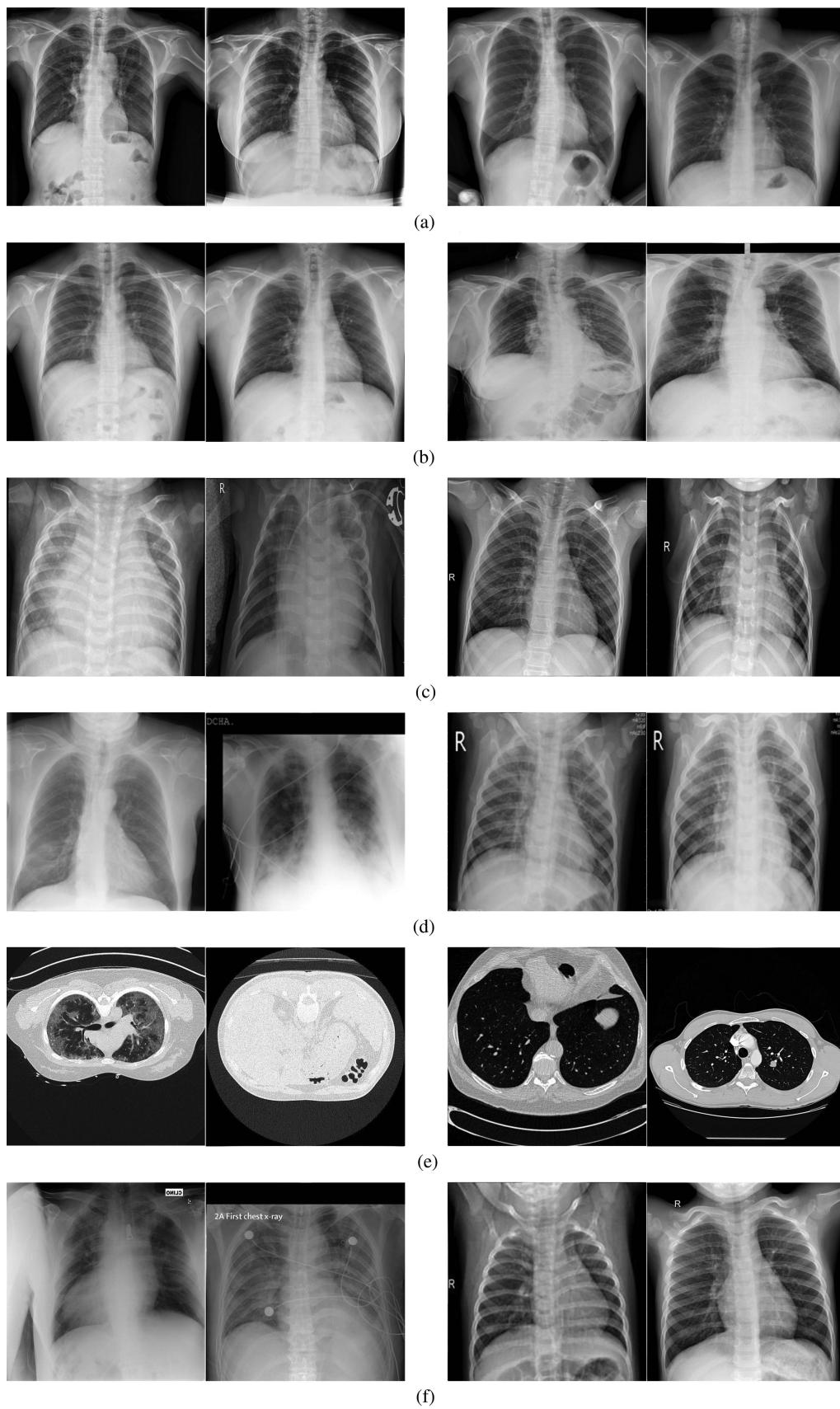


Fig. 3. Class-wise chest X-rays and CT scans from lung disease datasets. (a) TB vs. healthy dataset. (b) Healthy vs. sick dataset. (c) Pneumonia vs. normal dataset. (d) COVID-19 vs. pneumonia dataset. (e) COVID-19 vs. non-COVID CT scan dataset. (f) COVID-19 vs. non-COVID X-ray dataset.

Precision is a measure of the ability of a classifier to not mark a negative case as positive.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \quad (8)$$

Recall is a measure of the ability of a classifier to mark all positive cases as positive.

$$\text{F1-score} = \frac{\text{TP}}{\text{TP} + \frac{1}{2}(\text{FN} + \text{FP})}. \quad (9)$$

F1-score is the weighted average of precision and recall, and assesses the overall performance of the classifier efficiently.

Here, TP refers to the number of cases correctly identified as lung disease patients. On the other hand, FP represents the number of cases misclassified as patients. TN gives the number of cases rightly identified as healthy and FN refers the number of cases wrongly classified as healthy.

In addition to the above metrics, the macro-average and weighted average are also calculated. Both macro metric and weighted metric compute the precision, recall, and F1-score for each class in the lung disease dataset and return the average, but the latter considers the proportion for each class in the dataset while the former does not.

We also use the receiver operating characteristic (ROC) metric to evaluate the classifier's output quality. A ROC curve is a graphical plot that features FP rate on the horizontal axis, and TP rate on the vertical axis. Let TP rate and FP rate be denoted by TPR and FPR, respectively. The curve shows a trade-off between sensitivity (or TPR) and sensitivity or (1-FPR). The formulas of TPR and FPR are as follows:

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \quad (10)$$

TPR refers to the rate in which the positive cases are correctly diagnosed as lung disease patients.

$$\text{FPR} = \frac{\text{FP}}{\text{TN} + \text{FP}}. \quad (11)$$

FPR refers to the rate in which negative cases are incorrectly identified as lung disease patients.

Area under ROC curve (AUC) is a measure of the ability of a classifier to differentiate between classes and is used as a summary of the ROC curve. AUC ranges in value from 0 to 1. Higher AUC indicates better performance of the classifier model.

We also plot the precision–recall curve to study the output of classifier. It shows the trade-off between precision and recall for different threshold. Average precision (AP) is a metric that summarizes a precision–recall curve and is formulated as follows:

$$\text{AP} = \sum_n (\text{Recall}_n - \text{Recall}_{n-1}) \text{Precision}_n \quad (12)$$

where  $\text{Precision}_n$  is the precision and  $\text{Recall}_n$  is the recall at  $n$ th threshold.

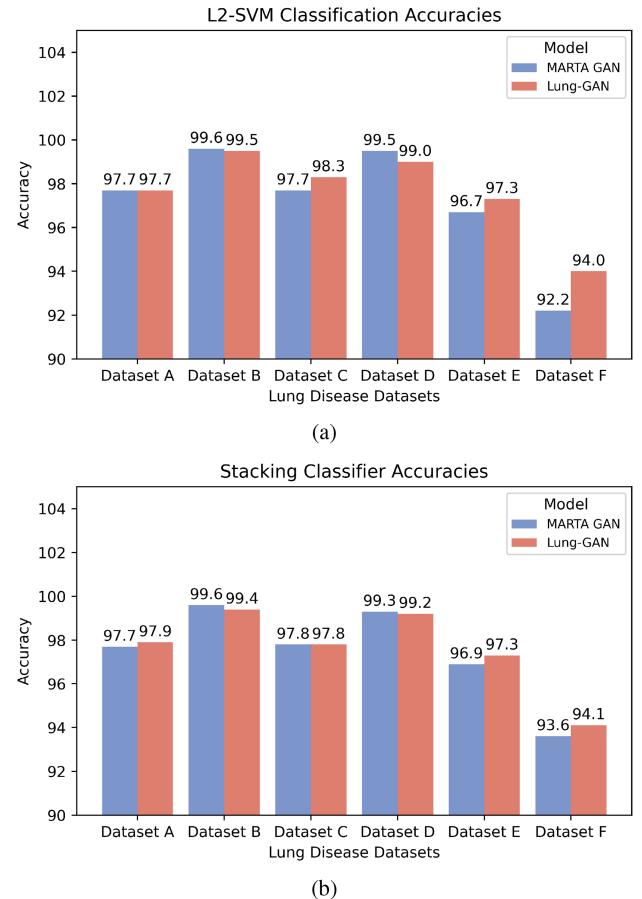


Fig. 4. Classification results using Lung-GAN features on six lung disease datasets. (a) L2-SVM classifier. (b) Stacking classifier.

### B. Performance Analysis

To assess the quality of representations learned by Lung-GANs, we extracted features in the multifeature layer in the discriminator and trained both L2-SVM and stacking classifier on top of these features. The classification performances of these classifiers are detailed in Tables II and III.

Fig. 4 summarizes the results of two GAN architectures, MARTA GAN, and Lung-GAN using L2-SVM [Fig. 4(a)] and the stacking classifier [Fig. 4(b)] models on six lung disease datasets. In most cases, the stacking classifier outputs better accuracies than SVM. In other words, the stacking classifier predominantly provides better generalizations than SVM. Therefore, stacking classifier is chosen. Fig. 4 shows that MARTA GANs have slightly better accuracy than Lung-GANs on Datasets B and D and an equal accuracy on Dataset C. However, on Datasets A, E, and F, Lung-GANs output is higher than that of MARTA GANs. MARTA GANs can produce  $256 \times 256$  resolution images, while Lung-GANs can generate higher quality lung disease images with  $512 \times 512$  resolution using an additional layer in the generator network, which helps produce features that enhance the model performance.

The confusion matrices depicting stacking classifier's performance on each of the six lung disease datasets are shown in Fig. 5. The performance results are described below in detail.

TABLE II  
PERFORMANCE METRICS ACHIEVED WITH LUNG-GANs USING L2-SVM CLASSIFIER

Dataset	Class	Precision	Recall	F1-score	Accuracy (%)
<b>Tuberculosis (TB) vs. Healthy</b>	TB	1.00	0.96	0.98	<b>97.7</b>
	Healthy	0.96	1.00	0.98	
	Macro	0.98	0.98	0.98	
	Weighted	0.98	0.98	0.98	
<b>Healthy vs. Sick</b>	Healthy	0.99	1.00	1.00	<b>99.5</b>
	Sick	1.00	0.99	0.99	
	Macro	1.00	0.99	0.99	
	Weighted	1.00	0.99	0.99	
<b>Pneumonia vs. Normal</b>	Pneumonia	0.99	0.98	0.98	<b>98.3</b>
	Normal	0.98	0.99	0.98	
	Macro	0.98	0.98	0.98	
	Weighted	0.98	0.98	0.98	
<b>COVID-19 vs. Pneumonia</b>	COVID-19	0.99	0.99	0.99	<b>99.0</b>
	Pneumonia	0.99	0.99	0.99	
	Macro	0.99	0.99	0.99	
	Weighted	0.99	0.99	0.99	
<b>COVID-19 vs. Non-COVID (CT)</b>	COVID-19	0.97	0.95	0.96	<b>97.3</b>
	Non-COVID	0.98	0.98	0.98	
	Macro	0.97	0.97	0.97	
	Weighted	0.97	0.97	0.97	
<b>COVID-19 vs. Non-COVID (X-ray)</b>	COVID-19	0.95	0.95	0.95	<b>94.0</b>
	Non-COVID	0.93	0.93	0.93	
	Macro	0.94	0.94	0.94	
	Weighted	0.94	0.94	0.94	

TABLE III  
PERFORMANCE METRICS ACHIEVED WITH LUNG-GANs USING STACKING CLASSIFIER

Dataset	Class	Precision	Recall	F1-score	Accuracy (%)
<b>Tuberculosis (TB) vs. Healthy</b>	TB	0.97	0.99	0.98	<b>97.9</b>
	Healthy	0.97	0.97	0.98	
	Macro	0.98	0.98	0.98	
	Weighted	0.98	0.98	0.98	
<b>Healthy vs. Sick</b>	Healthy	0.99	0.99	0.99	<b>99.4</b>
	Sick	0.99	0.99	0.99	
	Macro	0.99	0.99	0.99	
	Weighted	0.99	0.99	0.99	
<b>Pneumonia vs. Normal</b>	Pneumonia	0.97	0.98	0.98	<b>97.8</b>
	Normal	0.98	0.97	0.98	
	Macro	0.98	0.98	0.98	
	Weighted	0.98	0.98	0.98	
<b>COVID-19 vs. Pneumonia</b>	COVID-19	1.00	0.99	0.99	<b>99.2</b>
	Pneumonia	0.99	1.00	0.99	
	Macro	0.99	0.99	0.99	
	Weighted	0.99	0.99	0.99	
<b>COVID-19 vs. Non-COVID (CT)</b>	COVID-19	0.97	0.95	0.96	<b>97.3</b>
	Non-COVID	0.98	0.98	0.98	
	Macro	0.97	0.97	0.97	
	Weighted	0.97	0.97	0.97	
<b>COVID-19 vs. Non-COVID (X-ray)</b>	COVID-19	0.94	0.95	0.95	<b>94.1</b>
	Non-COVID	0.94	0.93	0.93	
	Macro	0.94	0.94	0.94	
	Weighted	0.94	0.94	0.94	

1) *Tuberculosis (TB) vs. Healthy Dataset*: The confusion matrix in Fig. 5(a) shows that the classifier is able to correctly predict 232 out of 240 TB images as TB and 238 out of 240 healthy images as healthy. This results in an overall classification accuracy of 97.9%. The recorded precision, recall, and F1-score values are also high.

2) *Healthy vs. Sick Dataset*: Our model yields a high accuracy of 99.4% on the healthy vs. sick dataset. Fig. 5(b) shows that the classifier predicts 596 out of 600 sick images and 597

out of 600 healthy images accurately. The value of all metrics for each class label as well as on the macro and weighted average metric is 0.99.

3) *Pneumonia vs. Normal Dataset*: It can be seen in Fig. 5(c) that only 5 (out of 300) pneumonia images are misclassified as normal (FNs) and only 8 (out of 300) normal images are misclassified as pneumonia (FPs). The overall classification accuracy obtained on this dataset is 97.8%. Both precision and recall values are high.

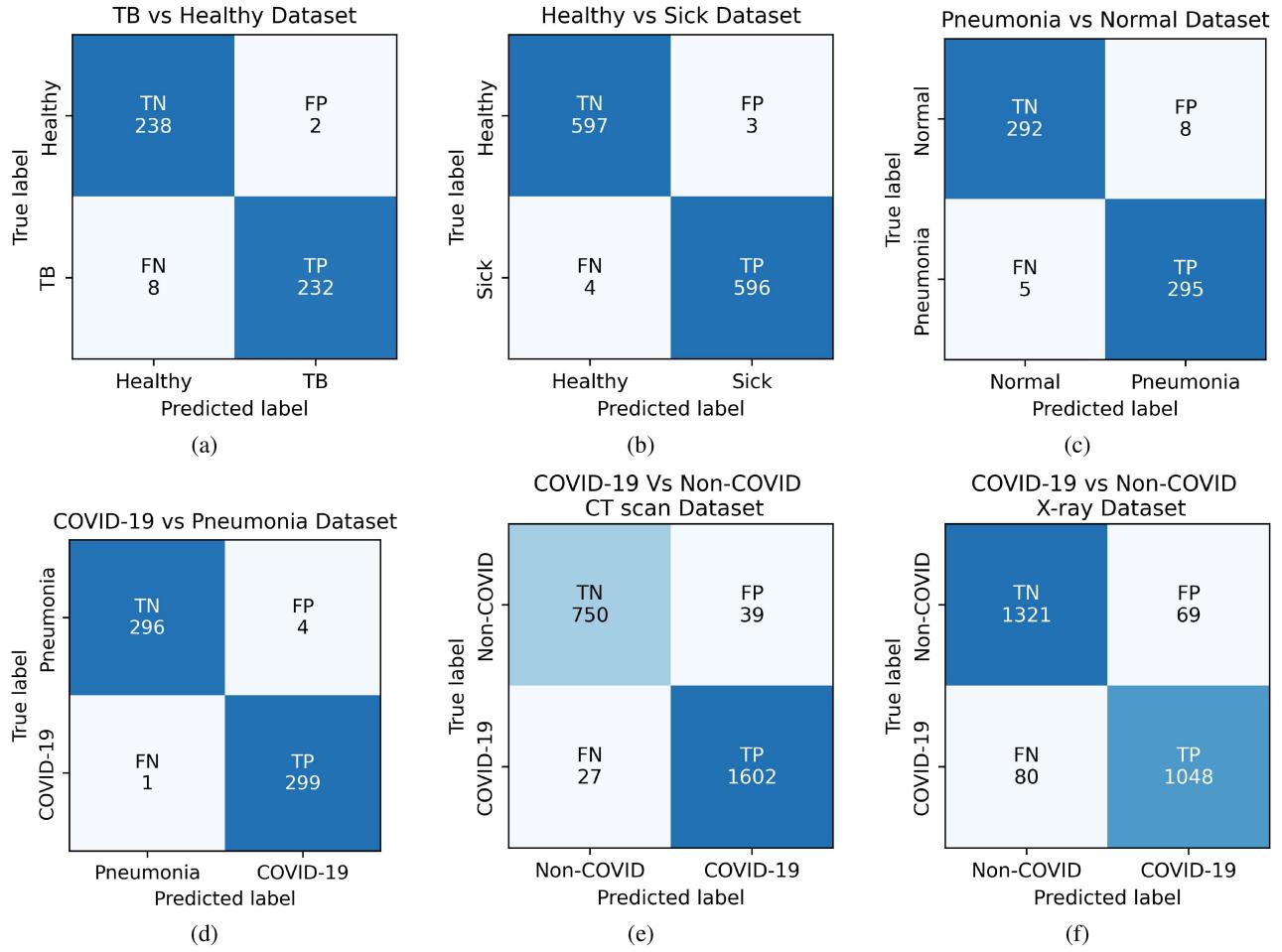


Fig. 5. Confusion matrices showing stacking classifier performance on six lung disease datasets using Lung-GANs.

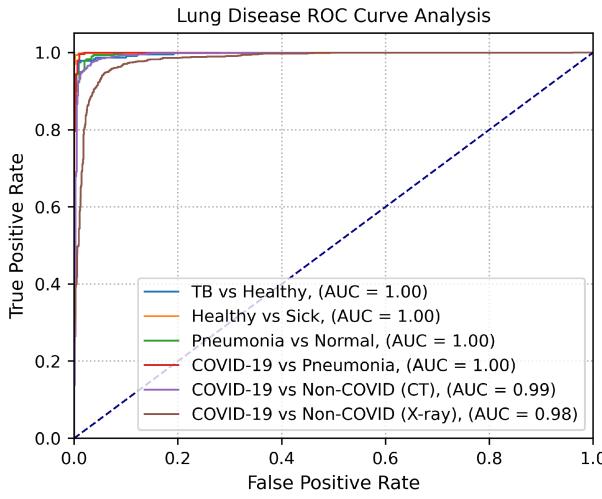


Fig. 6. Lung disease ROC curve analysis for six lung disease datasets.

4) *COVID-19 vs. Pneumonia Dataset*: The model outputs a 99.2% accuracy on the COVID-19 vs. pneumonia dataset. The confusion matrix in Fig. 5(d) shows that the classifier is able to accurately predict 299 out of 300 COVID-19 images and 296 out of 300 pneumonia images. The precision is 1 for class label

COVID-19 and 0.99 for pneumonia. Similarly, recall for label COVID-19 is 0.99 and 1 for pneumonia.

5) *COVID-19 vs. Non-COVID CT Scan Dataset*: The confusion matrix for this dataset plotted in Fig. 5(e) shows that 27 out of 1602 images are misclassified as non-COVID (FNs) and

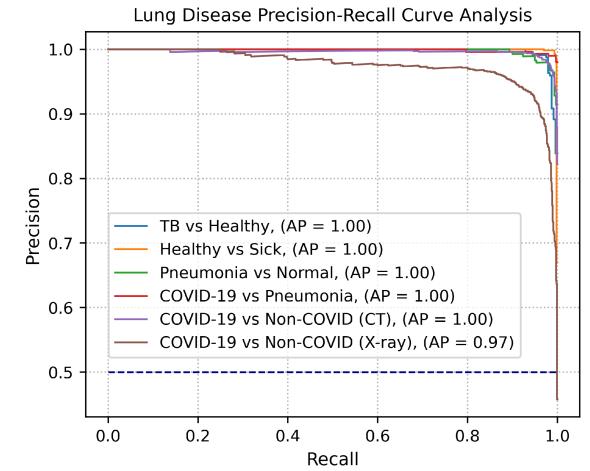


Fig. 7. Lung disease precision-recall curve analysis for six lung disease datasets.

TABLE IV  
COMPARISON OF OUR METHOD WITH EXISTING METHODS FOR LUNG DISEASE CLASSIFICATION

Dataset	Proposed	DCEC [47]	IDEC [48]	Convolutional Autoencoder with Stacking Classifier
<b>Tuberculosis (TB) vs. Healthy</b>	<b>97.9</b>	79.3	80.7	91.0
	TP = 232 , TN = 238, FP = 2 , FN = 8	TP = 192 , TN = 189, FP = 51, FN = 48	TP = 195 , TN = 190, FP = 50, FN = 45	TP = 222 , TN = 215, FP = 25, FN = 18
<b>Healthy vs. Sick</b>	<b>99.4</b>	83.7	84.6	96.9
	TP = 596, TN = 597, FP = 3, FN = 4	TP = 504, TN = 501, FP = 99, FN = 96	TP = 510, TN = 506, FP = 94, FN = 90	TP = 576, TN = 587, FP = 13, FN = 24
<b>Pneumonia vs. Normal</b>	<b>97.8</b>	82.2	85.5	91.1
	TP = 295, TN = 292, FP = 8 , FN = 5	TP = 248, TN = 245, FP = 55, FN = 52	TP = 259, TN = 254, FP = 46, FN = 41	TP = 275, TN = 272, FP = 28, FN = 25
<b>COVID-19 vs. Pneumonia</b>	<b>99.2</b>	84.5	87.3	94.8
	TP = 299, TN = 296, FP = 4 , FN = 1	TP = 256, TN = 251, FP = 49, FN = 44	TP = 269, TN = 255, FP = 45, FN = 31	TP = 289, TN = 280, FP = 20, FN = 11
<b>COVID-19 vs. Non-COVID (CT)</b>	<b>97.3</b>	92.5	93.4	96.3
	TP = 1,602, TN = 750, FP = 39, FN = 27	TP = 1,559, TN = 680, FP = 109, FN = 70	TP = 1,569, TN = 690, FP = 99, FN = 60	TP = 1,590, TN = 740, FP = 49, FN = 39
<b>COVID-19 vs. Non-COVID (X-ray)</b>	<b>94.1</b>	89.5	90.0	92.4
	TP = 1,048, TN = 1,321, FP = 69, FN = 80	TP = 975, TN = 1,280, FP = 110, FN = 153	TP = 985, TN = 1,282, FP = 108, FN = 143	TP = 1,028, TN = 1,301, FP = 89, FN = 100
Average Accuracy	<b>97.6</b>	85.3	86.9	93.8
Average Sensitivity	<b>0.98</b>	0.86	0.88	0.94

39 out of 750 images are misclassified as COVID-19 (FPs). The classification accuracy obtained on this dataset is 97.3%.

6) *COVID-19 vs. Non-COVID X-Ray Dataset*: From the confusion matrix plotted in Fig. 5(f), we can see that 80 out of 1128 images are incorrectly classified as non-COVID. Similarly, 69 out of 1390 non-COVID images are classified as COVID-19. Among all the six datasets, the model outputs the least accuracy of 94.1% on this dataset.

Both COVID-19 vs. Non-COVID CT scan and X-ray datasets are imbalanced, i.e., the number of samples for each class label is not equal or balanced. For example, in the training set of the CT scan dataset, there are 1110 more COVID-19 images than non-COVID images. For the X-ray dataset, there are 346 fewer COVID-19 images compared to the non-COVID images. That is, the CT scan dataset is biased toward the COVID-19 class, and the X-ray dataset is skewed toward the non-COVID class. Regardless, our method obtained approximately equal values on the precision, recall, and F1-score metrics, showing that the model does not overfit the majority class in both datasets.

The ROC curve and precision-recall curve plots showing all datasets are presented in Figs. 6 and 7. A classifier that gives curves closer to the top-left corner indicates superior performance. In view of that, Fig. 6 confirms the efficacy of the features learnt by Lung-GANs. The AUC score obtained for the first four datasets is 1.00, 0.99 for COVID-19 vs. non-COVID CT dataset, and 0.98 for COVID-19 vs. non-COVID X-ray dataset. Similarly, the AP score from the precision-recall curve is 1.00 for the first 5 datasets and 0.97 for the last dataset.

Overall, our model yields significantly high accuracies on all the datasets, with the lowest accuracy being 94.1%. This suggests that the proposed architecture learns meaningful image

representations that help in enhancing the classifier performance on lung disease classification.

We also performed several experiments by decreasing the number of layers in the proposed architecture. We observed that reducing the number of layers lowered the classification performance. On the other hand, increasing the count of layers in the model would increase the number of network parameters, training time, and make it massively large only for an ever-so-slight improvement in the classification performance.

### C. Comparison With Existing Methods

Table IV analyzes the classification performance of our proposed Lung-GANs and three other existing unsupervised methods. Lung-GANs achieved the highest classification accuracy on all datasets used in this work. In particular, it outperformed the IDEC [48] (an unsupervised algorithm that jointly learns deep feature representations and performs clustering) by 12.3% on average taken over all the six datasets. In addition, it produced 10.7% more accuracy than DCEC [47], an improved unsupervised deep clustering algorithm that incorporates convolutional neural networks. We also compared our method to a convolutional autoencoder (two encoder and decoder layers) that is used to extract features. The extracted features are classified using a stacking classifier, which involves SVM and random forest. We can see that our method achieved 97.6% average accuracy compared to the 93.8% average accuracy obtained by the autoencoder. This is because the encoder's compression of an input image before it is reconstructed by the decoder leads to a loss of some information. Such information is not lost in the case of adversarial networks.

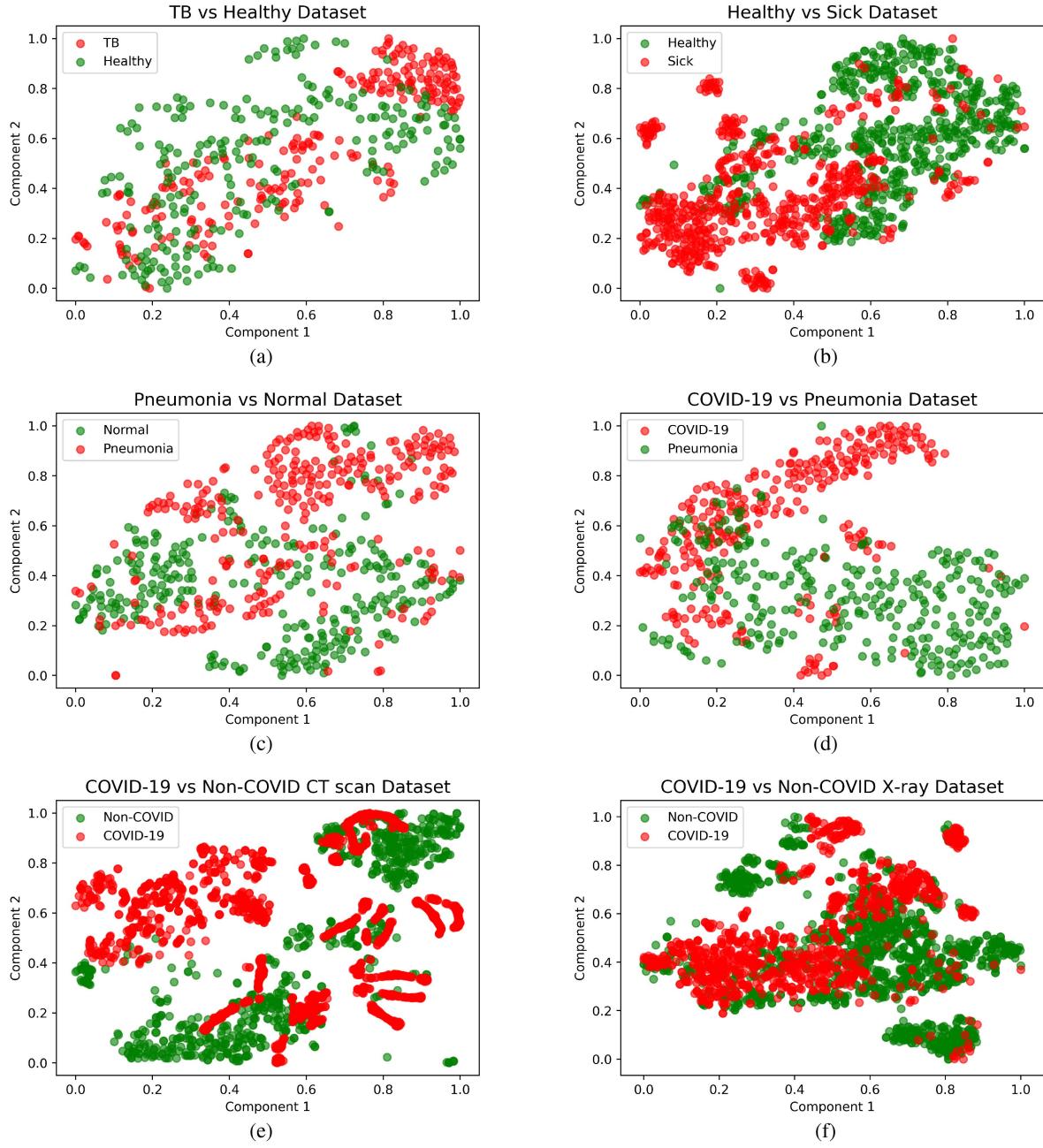


Fig. 8. 2-D feature visualization of Lung-GAN image representations on six lung disease datasets.

The merit of Lung-GANs over IDEC, DCEC, and convolutional autoencoders lie in its ability to learn better interpretable image representations even from challenging lung disease datasets. This capability allows our method to attain consistently superior performance than the rest of the methods in all metrics on all the six lung disease datasets. Most importantly, our method reports better sensitivity, meaning that the probability of any lung disease patient being classified as “diseased” using our method is higher than the compared methods. However, one limitation of Lung-GANs with respect to IDEC and DCEC is that it does not jointly accomplish feature extraction and classification tasks but follows a two-stage process.

#### D. Feature Visualization

We visualized the global image representations encoded in the Lung-GAN’s features of the lung disease datasets using t-distributed stochastic neighbor embedding, commonly known as the t-SNE technique [49]. t-SNE maps the high-dimensional data into a low-dimensional space and attempts to find hidden structures in the data, identifying natural clusters and smooth nonlinear variations along the dimensions. We computed the features and then used the t-SNE algorithm to find a faithful representation of these features in a 2-D plane. Fig. 8 shows the 2-D visualization of these features for six lung disease datasets. The parameters used for tuning t-SNE are perplexity

$= 30$ , learning rate  $= 200$ , early exaggeration  $= 12$ , number of iterations  $= 100$ , and the method used is Barnes-Hut.

Consider the TB vs. healthy visualization in Fig. 8(a) where the red data points are TB image features, and the green data points are healthy image features. The red and green data points are well-separated but overlap in a few regions. Similar is the case in Fig. 8(c) that depicts pneumonia image features vs. normal image features. Fig. 8(e) also shows a similar separation of COVID-19 and non-COVID image features. Fig. 8(f) has the maximum overlap between the features of COVID-19 and non-COVID X-ray images, which explains the least accuracy of our model on that particular dataset. On the other hand, Fig. 8(d) shows reasonably good segregation between COVID-19 and pneumonia image features. The model performs the best in separating healthy from sick as depicted in Fig. 8(b). We can see that the healthy and sick image features are grouped in their own clusters. There are almost very few overlaps in this figure, which justifies the 99.4% accuracy of the model on the respective classification task. Overall, the t-SNE plots in Fig. 8 help in finding structures within the features learned by the Lung-GANs and draw meaningful insights into what is going on.

The features learned by our method on all the lung-disease datasets can be visualized and understood as described in [50]. But presently, our work does involve any radiologists or medical experts to validate the visualization results. They have to be included to discuss and obtain more specific details about the similar features extracted from the lung diseases.

## VI. ADVANTAGES AND LIMITATIONS OF LUNG-GANS

In this article, we developed an effectual and accurate CAD system and demonstrated its feasibility in detecting lung diseases such as pneumonia, tuberculosis, and COVID-19 from chest X-ray and CT images. Our CAD system has some advantages. First, it offers higher sensitivity in lung diseases and better performances than other leading unsupervised techniques. Second, it enables rapid identification of COVID-19 patients, which can help combat the current pandemic and reduce the pressure on health-care systems. Third, it has achieved promising results in differentiating COVID-19 from similar types of pneumonia. This ability helps avoid misclassifications; for example, a patient with regular pneumonia symptoms might get wrongly classified as a COVID-19 patient. Finally, the proposed system can easily be extended for the computer-aided detection of similar emerging pandemics while requiring only unlabeled data.

Our study also has some limitations which need to be addressed. First, even though the robustness of the proposed GANs is seen on many datasets of similar lung diseases, the datasets are not real-time patient samples. Second, the framework does not have a proper strategy to handle misclassified samples. Third, our GAN architecture converts all input images into  $512 \times 512$  dimensions, which might cause information loss in higher dimension images. Further study is required to avoid such loss, such as using a multiscale learning framework [51]. Finally, the similarity of the features of diseases can be studied in detail by including Siamese neural networks, which would help understand the relationship/similarity between features of

similar diseases. There could be criteria or some threshold factors based on the features learned by the network to decide whether the patient has COVID-19, pneumonia, etc.

## VII. CONCLUSION

Lung diseases are a severe matter of concern all over the world. Early diagnosis is crucial to help in faster recovery and improve long-term survival rates. In light of this, we present a DL-based unsupervised framework called Lung-GANs to classify lung diseases from chest X-ray and CT images. Our framework introduces a multiple-layer GAN architecture to learn image representations of various lung diseases such as pneumonia, tuberculosis, and COVID-19. These representations are then utilized in a stacking classifier involving random forest and linear SVC to classify lung diseases. This article, to our knowledge, is the first one to propose a GAN-based single framework to classify unsupervised lung disease images.

The major strength of Lung-GANs is that it does not require an enormous amount of labeled data to provide better generalization and can learn interpretable representations even from challenging X-ray and CT datasets. Moreover, the framework is not confined to any particular lung disease and can be extended to other similar lung diseases. It performed remarkably well on TB vs. healthy, pneumonia vs. normal, COVID-19 vs. pneumonia, COVID-19 vs. non-COVID classification, and yielded significantly high accuracies (94%–99.5%) on all the six different lung disease datasets. We demonstrate through experiments that our method outperforms the current leading unsupervised models in lung disease classification. It reported higher accuracy (up to 99.5%) and better sensitivity compared to the existing methods on six different lung disease datasets. In conclusion, Lung-GANs provide a noteworthy improvement in computer-aided diagnosis of lung diseases.

Presently, the GAN and stacking classifier stages of our framework are disconnected. In the future, we plan to connect the two stages by proposing a new loss function that improves the classification performance further. In addition, we intend to evaluate our framework on multiclassification tasks and enhance it to handle severely imbalanced datasets better.

## ACKNOWLEDGMENT

The authors would like to sincerely thank all the reviewers and editors for their valuable guidance and suggestions to enhance the quality and depth of the manuscript.

## REFERENCES

- [1] T. Lallukka, A. Millear, A. Pain, M. Cortinovis, and G. Giussani, “GBD 2015 mortality and causes of death collaborators. global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980–2015: A systematic analysis for the global burden of disease study 2015,” *Lancet*, vol. 389, no. 10064, p. E1, 2017.
- [2] L. H. Garland, “On the scientific evaluation of diagnostic procedures: Presidential address thirty-fourth annual meeting of the radiological society of north america,” *Radiology*, vol. 52, no. 3, pp. 309–328, 1949.
- [3] V. Ravi, H. Narasimhan, C. Chakraborty, and T. D. Pham, “Deep learning-based meta-classifier approach for COVID-19 classification using CT scan and chest X-ray images,” *Multimedia Syst.*, pp. 1–15, 2021, doi: [10.1007/s00530-021-00826-1](https://doi.org/10.1007/s00530-021-00826-1).
- [4] D. Shen, G. Wu, and H.-I. Suk, “Deep learning in medical image analysis,” *Annu. Rev. Biomed. Eng.*, vol. 19, pp. 221–248, 2017.

- [5] A. Arunachalam, V. Ravi, V. Acharya, and T. D. Pham, "Toward data-model-agnostic autonomous machine-generated data labeling and annotation platform: COVID-19 autoannotation use case," *IEEE Trans. Eng. Manage.*, pp. 1–12, 2021, doi: [10.1109/TEM.2021.3094544](https://doi.org/10.1109/TEM.2021.3094544).
- [6] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," *Adv. Neural Inf. Proc. Syst.*, vol. 27, 2014.
- [7] B. Van Ginneken, B. T. H. Romeny, and M. A. Viergever, "Computer-aided diagnosis in chest radiography: A survey," *IEEE Trans. Med. Imag.*, vol. 20, no. 12, pp. 1228–1241, Dec. 2001.
- [8] R. H. Abiyev and M. K. S. Ma'aithah, "Deep convolutional neural networks for chest diseases detection," *J. Healthcare Eng.*, vol. 2018, pp. 1–11, 2018.
- [9] Y.-X. Tang *et al.*, "Automated abnormality classification of chest radiographs using deep convolutional neural networks," *NPJ Digit. Med.*, vol. 3, no. 1, pp. 1–8, 2020.
- [10] A. Bhandary *et al.*, "Deep-learning framework to detect lung abnormality—A study with chest X-ray and lung CT scan images," *Pattern Recognit. Lett.*, vol. 129, pp. 271–278, 2020.
- [11] S. Bharati, P. Podder, and M. R. H. Mondal, "Hybrid deep learning for detecting lung diseases from x-ray images," *Inform. Med. Unlocked*, vol. 20, 2020, Art. no. 100391.
- [12] J. Ma, Y. Song, X. Tian, Y. Hua, R. Zhang, and J. Wu, "Survey on deep learning for pulmonary medical imaging," *Frontiers Med.*, vol. 14, no. 4, pp. 450–469, 2020.
- [13] S. T. H. Kieu, A. Bade, M. H. A. Hijazi, and H. Kolivand, "A survey of deep learning for lung disease detection on medical images: State-of-the-art, taxonomy, issues and future directions," *J. Imag.*, vol. 6, no. 12, p. 131, 2020.
- [14] A. A. Saraiva, N. M. F. Ferreira, L. L. de Sousa, N. J. C. Costa, J. V. M. Sousa, D. Santos, A. Valente, and S. Soares, "Classification of images of childhood pneumonia using convolutional neural networks," *BIOIMAGING*, 2019, pp. 112–119, doi: [10.5220/0007404301120119](https://doi.org/10.5220/0007404301120119).
- [15] V. Chouhan *et al.*, "A novel transfer learning based approach for pneumonia detection in chest X-ray images," *Appl. Sci.*, vol. 10, no. 2, p. 559, 2020.
- [16] T. Rahman *et al.*, "Transfer learning with deep convolutional neural network (CNN) for pneumonia detection using chest X-ray," *Appl. Sci.*, vol. 10, no. 9, p. 3233, 2020.
- [17] G. Liang and L. Zheng, "A transfer learning method with deep residual network for pediatric pneumonia diagnosis," *Comput. Methods Programs Biomed.*, vol. 187, 2020, Art. no. 104964.
- [18] A. Akgundogdu, "Detection of pneumonia in chest X-ray images by using 2D discrete wavelet feature extraction with random forest," *Int. J. Imag. Syst. Technol.*, vol. 31, no. 1, pp. 82–93, 2021.
- [19] N. Dey, Y.-D. Zhang, V. Rajinikanth, R. Pugalenth, and N. S. M. Raja, "Customized VGG19 architecture for pneumonia detection in chest X-rays," *Pattern Recognit. Lett.*, vol. 143, pp. 67–74, 2021.
- [20] X. Mei *et al.*, "Artificial intelligence-enabled rapid diagnosis of patients with Covid-19," *Nature Med.*, vol. 26, no. 8, pp. 1224–1228, 2020.
- [21] D. Ezzat, A. E. Hassani, and H. A. Ella, "An optimized deep learning architecture for the diagnosis of COVID-19 disease based on gravitational search optimization," *Appl. Soft Comput.*, vol. 98, p. 106742, 2020.
- [22] R. Zhang *et al.*, "COVID19XrayNet: A two-step transfer learning model for the COVID-19 detecting problem based on a limited number of chest X-ray images," *Interdiscipl. Sci.: Comput. Life Sci.*, vol. 12, no. 4, pp. 555–565, 2020.
- [23] I. D. Apostolopoulos, S. I. Aznaouridis, and M. A. Tzani, "Extracting possibly representative COVID-19 biomarkers from X-ray images with deep learning approach and image data related to pulmonary diseases," *J. Med. Biol. Eng.*, vol. 40, pp. 462–469, 2020.
- [24] A. I. Aviles-Rivero, P. Sellars, C.-B. Schönlieb, and N. Papadakis, "Graphx covid: Explainable deep graph diffusion pseudo-labelling for identifying COVID-19 on chest X-rays," 2021, arXiv preprint [arXiv:2010.00378v2](https://arxiv.org/abs/2010.00378v2).
- [25] Z. Wang *et al.*, "Automatically discriminating and localizing Covid-19 from community-acquired pneumonia on chest X-rays," *Pattern Recognit.*, vol. 110, 2021, Art. no. 107613.
- [26] A. Amyar, R. Modzelewski, H. Li, and S. Ruan, "Multi-task deep learning based CT imaging analysis for COVID-19 pneumonia: Classification and segmentation," *Comput. Biol. Med.*, vol. 126, 2020, Art. no. 104037.
- [27] J. Wang *et al.*, "Prior-attention residual learning for more discriminative COVID-19 screening in CT images," *IEEE Trans. Med. Imag.*, vol. 39, no. 8, pp. 2572–2583, Aug. 2020.
- [28] M. Xu *et al.*, "Accurately differentiating between patients with COVID-19, patients with other viral infections, and healthy individuals: Multimodal late fusion learning approach," *J. Med. Internet Res.*, vol. 23, no. 1, 2021, Art. no. e25535.
- [29] X. Chen, L. Yao, T. Zhou, J. Dong, and Y. Zhang, "Momentum contrastive learning for few-shot COVID-19 diagnosis from chest CT images," *Pattern Recognit.*, vol. 113, 2021, Art. no. 107826.
- [30] M. Roberts *et al.*, "Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans," *Nature Mach. Intell.*, vol. 3, no. 3, pp. 199–217, 2021.
- [31] S. Jaeger *et al.*, "Automatic tuberculosis screening using chest radiographs," *IEEE Trans. Med. Imag.*, vol. 33, no. 2, pp. 233–245, Feb. 2014.
- [32] P. Lakhani and B. Sundaram, "Deep learning at chest radiography: Automated classification of pulmonary tuberculosis by using convolutional neural networks," *Radiology*, vol. 284, no. 2, pp. 574–582, 2017.
- [33] U. Lopes and J. F. Valiati, "Pre-trained convolutional neural networks as feature extractors for tuberculosis detection," *Comput. Biol. Med.*, vol. 89, pp. 135–143, 2017.
- [34] T. Rahman *et al.*, "Reliable tuberculosis detection using chest X-ray with deep learning, segmentation and visualization," *IEEE Access*, vol. 8, pp. 191 586–191601, 2020.
- [35] T. B. Chandra, K. Verma, B. K. Singh, D. Jain, and S. S. Netam, "Automatic detection of tuberculosis related abnormalities in chest X-ray images using hierarchical feature extraction scheme," *Expert Syst. Appl.*, vol. 158, 2020, Art. no. 113514.
- [36] Z. U. Abideen *et al.*, "Uncertainty assisted robust tuberculosis identification with Bayesian convolutional neural networks," *IEEE Access*, vol. 8, pp. 22 812–22825, 2020.
- [37] S. Rajaraman and S. K. Antani, "Modality-specific deep learning model ensembles toward improving tb detection in chest radiographs," *IEEE Access*, vol. 8, pp. 27 318–27326, 2020.
- [38] K. Munadi, K. Muchtar, N. Maulina, and B. Pradhan, "Image enhancement for tuberculosis detection using deep learning," *IEEE Access*, vol. 8, pp. 217 897–217907, 2020.
- [39] D. Lin, K. Fu, Y. Wang, G. Xu, and X. Sun, "Marta GANs: Unsupervised representation learning for remote sensing image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 11, pp. 2092–2096, Nov. 2017.
- [40] A. Odena, V. Dumoulin, and C. Olah, "Deconvolution and checkerboard artifacts," *Distill.*, vol. 1, no. 10, p. e 3, 2016.
- [41] D. H. Wolpert, "Stacked generalization," *Neural Netw.*, vol. 5, no. 2, pp. 241–259, 1992.
- [42] Y. Liu, Y.-H. Wu, Y. Ban, H. Wang, and M.-M. Cheng, "Rethinking computer-aided tuberculosis diagnosis," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 2646–2655.
- [43] D. Kermany, K. Zhang, M. Goldbaum *et al.*, "Labeled optical coherence tomography (OCT) and chest X-ray images for classification," *Mendeley Data*, vol. 2, no. 2, 2018.
- [44] M. E. Chowdhury *et al.*, "Can AI help in screening viral and Covid-19 pneumonia?," *IEEE Access*, vol. 8, pp. 132 665–132676, 2020.
- [45] T. Rahman *et al.*, "Exploring the effect of image enhancement techniques on Covid-19 detection using chest X-rays images," *Comput. Biol. Med.*, 2021, Art. no. 104319.
- [46] W. El-Shafai, "Extensive Covid-19 X-ray and CT chest images dataset," 2020. [Online]. Available: <https://data.mendeley.com/datasets/8h65ywd2jr/3>
- [47] X. Guo, X. Liu, E. Zhu, and J. Yin, "Deep clustering with convolutional autoencoders," in *Proc. Int. Conf. Neural Inf. Process.* Springer, 2017, pp. 373–382.
- [48] X. Guo, L. Gao, X. Liu, and J. Yin, "Improved deep embedded clustering with local structure preservation," in *Proc. Int. Joint Conf. Artif. Intell.*, 2017, pp. 1753–1759.
- [49] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 86, pp. 2579–2605, 2008. [Online]. Available: <http://jmlr.org/papers/v9/vandermaaten08a.html>
- [50] D. Bau, J.-Y. Zhu, H. Strobelt, B. Zhou, J. B. Tenenbaum, W. T. Freeman, and A. Torralba, "Gan dissection: Visualizing and understanding generative adversarial networks," in *Proc. Int. Conf. Learn. Representations (ICLR)*, 2019.
- [51] S. Sriram, R. Vinayakumar, V. Sowmya, M. Alazab, and K. Soman, "Multi-scale learning based malware variant detection using spatial pyramid pooling network," in *Proc. IEEE INFOCOM Conf. Comput. Commun. Workshops*, 2020, pp. 740–745.