A multichannel EfficientNet deep learning-based stacking ensemble approach for lung disease detection using chest X-ray images

Vinayakumar Ravi1 • Vasundhara Acharya2 • Mamoun Alazab3

Received: 30 January 2022 / Revised: 21 May 2022 / Accepted: 17 June 2022 / Published online: 19 July 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract
This paper proposes a multichannel deep learning approach for lung disease detection using chest X-rays. The multichannel models used in this work are EfficientNetB0, EfficientNetB1, and EfficientNetB2 pretrained models. The features from EfficientNet models are fused together. Next, the fused features are passed into more than one non-linear fully connected layer. Finally, the features passed into a stacked ensemble learning classifier for lung disease detection. The stacked ensemble learning classifier contains random forest and SVM in the first stage and logistic regression in the second stage for lung disease detection. The performance of the proposed method is studied in detail for more than one lung disease such as pneumonia, Tuberculosis (TB), and COVID-19. The performances of the proposed method for lung disease detection using chest X-rays compared with similar methods with the aim to show that the method is robust and has the capability to achieve better performances. In all the experiments on lung disease, the proposed method showed better performance and outperformed similar lung disease existing methods. This indicates that the proposed method is robust and generalizable on unseen chest X-rays data samples. To ensure that the features learnt by the proposed method is optimal, t-SNE feature visualization was shown on all three lung disease models. Overall, the proposed method has shown 98% detection accuracy for pediatric pneumonia lung disease, 99% detection accuracy for TB lung disease, and 98% detection accuracy for COVID-19 lung disease. The proposed method can be used as a tool for point-of-care diagnosis by healthcare radiologists. Journal instruction requires a city for affiliations; however, this is missing in affiliation 3. Please verify if the provided city is correct and amend if necessary.

Keywords Lung disease • Pneumonia • COVID-19 • Tuberculosis • Deep learning • Transfer learning • Multichannel • Stacking • Chest X-ray

1 Introduction
Climate change is a public health danger on a par with cigarette smoking. Increased atmospheric carbon dioxide levels in the atmosphere have already significantly increased air pollution, endangering respiratory health. According to Science, an increase in average global temperature of two degrees had posed much danger to humans. People with Asthma, Chronic obstructive pulmonary disease (COPD), and lung cancer are more prone to climate change. COPD was the third most significant cause of mortality globally in 2019, accounting for around 3.23 million fatalities [1]. A study has shown that heat waves caused due to climate change result in an increase of respiratory diseases among children [2]. Some studies also relate the cyclone to rise in respiratory diseases. Climate changes also causes change in the plant flowering season leading to extended pollen seasons and in turn leading to more human exposure. Pollinosis patients suffer from asthma caused by the thunderstorms occurring during the pollen seasons [3]. Several studies have demonstrated effects of ozone over respiratory symptoms, including shortness of breath, lower respiratory tract infections and
acute and transient decreases in lung function [4]. The extent to which air pollution contributes to the development of such complicated respiratory illnesses is still a point of contention. However, a pollution reduction would aid in reducing its effects on the health of patients. Lung disease, also known as respiratory disease, is a common and significant cause of illness and death around the world. Some of the lung diseases are pneumonia, COVID-19, Tuberculosis, etc. Please confirm the corresponding author is correctly identified and amend if necessary. Corresponding author is Vinayakumar Ravi.

Pneumonia is a lung illness caused by bacteria or viruses in one or both lungs. It is a life-threatening illness ranked the eighth most common cause of death in the United States [5]. Pneumonia claimed the lives of 2.5 million people worldwide in 2019. Studies show that pneumonia is the most significant cause of death among children under the age of five [6]. Pneumonia can be widely classified into two types: Bacterial and Viral. Bacterial pneumonia has more dangerous symptoms than viral pneumonia and needs antibiotic therapy for recovery. Scientists have identified that one of the main reasons for pneumonia is the increasing levels of air pollution. Studies show that in most countries, the disease is overlooked in the case of elderly patients and is untreated until it reaches a deathly point. Researchers feel that there is a pressing need to develop precise computer-aided diagnosis systems to assist in diagnosing the disease at an early stage, especially in the case of children. There are many tests in the market to diagnose pneumonia, such as chest X-rays (CXR) and chest magnetic resonance imaging (MRI) [7]. However, chest X-rays have proven to be more valuable than chest MRI as they take less time than MRI, and also, the machines to perform the MRI are often expensive, which might not be available in underdeveloped countries. Deep learning techniques have been continuously utilized in this field as they can efficiently perform the segmentation and classification of the diseases with a better percentage of accuracy in prediction than the typical radiologist [8, 9]. Even though they cannot replace the doctor, they have proven helpful as a secondary decision tool.

Tuberculosis (TB) is another detrimental variant of lung infections that has devastating impacts on humankind. The Mycobacterium TB bacteria cause TB. The lungs are the main target area of this disease, but they can also affect other body parts. TB is a contagious disease, i.e., when people infected with TB cough or sneeze, they transmit the disease-causing bacteria in the air. Only a small quantity of these germs are enough to victimize a healthy person. Although scientific discoveries and research have been helping to curb the growing influence of TB, the meager annual medical progress rate in this sector has been unsuccessful in bringing a drastic drop in TB-affected patients. According to the Global TB Report, 2020, generated by the WHO [10], approximately 10 million people were affected by TB worldwide in 2019. Additionally, HIV/AIDS and TB form a deadly combination. The HIV infection significantly minimizes the strength of the immunity system of an individual, which serves as a favorable condition for an HIV-positive patient to contract TB. Out of the 1.4 million deaths caused by TB in 2019, more than 200 thousand patients were HIV positive. The statistics collected from 2018-2018 show how 58 million lives were saved by early diagnosis of TB [11]. Researchers feel that the time taken to detect and diagnose TB plays a vital role in mitigating the TB’s spread and reducing the death rates.

Wuhan in China, at the end of 2019, first saw the outbreak of coronavirus disease 2019 (COVID-19) epidemic. It was caused by the severe acute respiratory syndrome coronavirus 2, or SARS-CoV-2. It created history by becoming the first pandemic caused by coronaviruses. Doctors and researchers worldwide are working together to find the process of damage caused by the virus, which remains a mystery. The symptoms of the disease are seen to vary from person to person. Some mild symptoms include shortness of breath, fever, and chills. Some significant illnesses include breathing troubles, inability to wake, and severe chest pains. 5,609,678 people have died so far from the coronavirus COVID-19 outbreak [12]. The disease outbreak has created chaos by increasing the number of hospitalizations and thereby overburdening the medical facilities. Studies have shown that understanding physiological principles involved in respiration and gas exchange is necessary to diagnose and detect respiratory illnesses correctly. COVID-19 is now diagnosed with the SARS-CoV-2 viral RNA test [13]. Other laboratory procedures such as blood tests to determine the blood cell count and immunological testing have proven to be the alternatives in diagnosing the disease. CT scans also proved to be helpful in the detection of the disease. However, in some patient cases, the lungs did not show changes or imaging alterations, making it challenging to capture. The early diagnosis and classification of this disease into severe and non-severe categories is vital to prioritize the cases that need hospitalization and mitigate the number of deaths. Early diagnosis of lung disease is important. Literature survey shows that various image processing, machine learning, and deep learning methods are employed for lung disease detection using chest X-rays.

A chest X-ray which creates an image of the heart, lungs, and bones, is one of the most reliable approaches to detect diseases that affect the lungs and heart. CT scan creates numerous computerized views obtained from different angles to generate a detailed view of the body. CT scans can generate 3D images of the chest, unlike chest X-rays, which can produce 2D images. CT scans have proven to be helpful for doctors to identify diseases.
precisely. However, CT scans expose the patients to radiation, and also the contrast dye utilized affects the patients with kidney problems. They are also expensive, making them non-affordable in most underdeveloped countries. On the other hand, chest X-rays use a tiny dose of ionizing radiation, and it is fast and easy. Radiologists are responsible for interpreting the chest X-rays and CT scans and reporting the findings. However, this approach is error-prone due to its subjective nature, unavailability of experts, and increased number of patients. To overcome these issues, machine learning algorithms were employed to learn the essential features from the medical images, which would aid in diagnosing the disease [14]. In the case of image classification problems, these algorithms would involve a feature extraction stage followed by a classifier algorithm that would perform the classification. Recent advances in the machine learning field and the development of convolutional neural networks have made deep learning a very desirable technique to solve most artificial intelligence problems. Its superior performance and ability to produce precise results without traditional feature engineering have made it a suitable candidate for the classification of chest X-ray images. Deep learning algorithms are now quite good at simulating human judgment using pattern recognition and have proven to be a handy secondary decision tool for doctors and radiologists [15].

Transfer learning is a very popular approach in computer vision that is very useful when the dataset is scarce [16]. To perform a new task, pretrained models that have solved similar classification problems are employed. These pretrained models act as a starting point for another model designed to perform a new task. The new model may use the entire model or parts of it. The main aim of utilizing the deep learning approach is to determine the correct weights by employing multiple forward and backward propagations. Weights generated by utilizing the pretrained models trained on massive datasets such as ImageNet have proven to be very useful in lung disease detection and classification. The weights obtained from the pretrained models have reduced training costs for the new model as it has to learn only a few last layer weights. Deep learning models trained via the transfer learning approach have shown great potential in diagnosing lung diseases with the ability to classify various etiologies compared to models trained from scratch [17]. Most of the existing works are based on a ImageNet database pretrained model for lung disease classification using chest X-ray. Most of the studies are related to a single lung disease and in addition most of the studies are not in detail. In the existing studies models, the features extracted from ImageNet database pretrained models are flattened or global average pooling operation was employed with the aim to reduce the dimension. Later, classification is done using a fully connected layer. Also, there is no guarantee that the final feature representation learnt by the classification layer is optimal. Each of the pretrained models have their own capability to extract its own unique features during finetuning. The current work considers the fusion of features from more than one EfficientNet pretrained model for lung disease detection. A detailed evaluation of the proposed method is done on more than one lung disease. This work makes the following contribution for lung disease detection using chest X-rays

- The proposed model leverages more than one pretrained model for feature extraction, specifically called a multichannel pretrained model. The multichannel pretrained models used in this work are EfficientNetB0, EfficientNetB1, and EfficientNetB2.
- The features from EfficientNet models are fused together.
- The fused features are passed into more than one non-linear fully connected layer. Finally, the features passed into the stacked ensemble learning classifier for lung disease classification.
- The performances of the proposed method is studied in detail for more than one lung disease such as pneumonia, TB, and COVID-19.
- t-SNE feature visualization was employed to ensure that the feature representation learned by the model is optimal and meaningful for lung diseases.
- Detailed investigation and analysis of all the deep learning-based pretrained models are shown on more than one lung disease dataset to show that the proposed method is generalizable.
- The performances of the proposed method for lung disease detection using chest x-rays compared with the similar methods to show that the method is robust and has the capability to achieve better performances.
- Classification of chest x-ray into Normal or Sick but not TB, or TB in Tuberculosis.

The rest of the sections are organized as follows; Sect. 2 discusses detailed literature surveys on lung disease detection, mainly for pneumonia, TB, and COVID-19. The proposed method details are included in Sect. 3. Section 4 includes detailed information on the lung disease datasets. Statistical metrics information is included in Sect. 5. Experimental analysis, results, and its discussions are included in Sect. 6. Finally, the conclusion and feature works are placed in Sect. 7.

2 Literature survey

The detailed surveys on lung disease detection are studied and reported in [14, 15, 18, 19]. There are many machine learning and deep learning approaches along with image processing employed for lung disease detection and
classification. Overall, ImageNet-based fine-tuned models showed better performances.

The literature has several studies on deep learning to identify lung illness. Pneumonia is one of the lung disease types that has led to many deaths among children. Computer-assisted diagnostics methods have shown promise in terms of increasing diagnostic accuracy and have reduced the time consumed in analyzing the chest X-rays by the doctors.

Siddiqi [20] proposed an 18-layer deep sequential convolutional neural network to classify a chest X-ray into normal or pneumonia. The model achieved a classification accuracy of 94.39%. However, the model achieved a low specificity value. A CNN model to classify chest X-rays into two categories (normal and pneumonia) was proposed in [21]. The model proposed by the authors achieved a decent accuracy of 95.3%, but the model did not incorporate any of the regularization techniques to overcome the issue of overfitting. The two-dimensional discrete wavelet transform approach to extract features from X-rays to differentiate a normal X-ray from pneumonia proposed in [22] achieved an accuracy of 97.11. However, the authors have not compared their approach with deep learning methods. Stephen et al. [23] trained CNN from scratch with data augmentation techniques to diagnose pneumonia and achieved an accuracy of 95.31%. Nevertheless, the authors did not demonstrate the model’s performance in comparison to the models that used transfer learning. The multi-layer perceptron and the CNN model proposed in [24] successfully detected and classified pneumonia. However, the technique did not handle the case of overfitting. Nahid et al. [25] diversified the set of discriminative features by combining image processing techniques and a two-channel CNN. The model was successful in diagnosing pneumonia with an accuracy of 97.92%. However, the crop operation employed to eliminate the unwanted regions would be tedious if a manual approach is utilized. CNN model with Grad-CAM to highlight the diseased regions and achieved an accuracy of 95.31%. Nevertheless, the authors did not use any augmentation techniques to achieve diverse training samples, leading to similar training samples.

Due to a shortage of training data, transfer learning is a regularly used approach for constructing medical imaging models. Rahman et al. [27] pretrained AlexNet, ResNet18, DenseNet201, and SqueezeNet to differentiate between bacterial, viral, and normal chest X-rays. Their approach produced an accuracy of 98%. A comparison of various deep learning-based CNNs was shown in [28]. The authors determined that fine-tuned versions of ResNet50, MobileNet_V2, and Inception_ResNet_V2 showed satisfactory performance in classifying the chest X-rays into pneumonia or normal. Xception Network pre-trained on ImageNet weights was proposed in [29] to classify the chest X-rays into normal and pneumonia. Mahajan et al. [30] used CheXNet weights to train the DenseNet and achieved a decent accuracy in detecting pneumonia. However, the model was not able to detect the subtypes of pneumonia. Inception V3 model with image processing techniques have been utilized in [31] to detect pediatric pneumonia from chest X-ray images. Similarly, Hasan et al. [32] combined image processing techniques with VGG16 and VGG19 models and achieved an accuracy of 96.2 in detecting pneumonia. A novel 21 layer convolutional neural network called PneumoniaNet was proposed in [33]. Its robustness was tested with noisy chest X-ray images. The model achieved a decent accuracy of 95.83%. However, the model still had scope for robustness. An ensemble of AlexNet, DenseNet121, InceptionV3, resNet18, and GoogLeNet neural networks with transfer learned weights were utilized to recognize pneumonia in [34].

Tuberculosis (TB) is a transmissible illness that is one of the world’s top 10 causes of mortality. Deep learning has become one of the most used approaches, especially for medical image analysis. Sun et al. [35] utilized CNN to extract discriminative features from Computerized Tomography (CT) scans and employed a Recurrent Neural Network to classify the type of TB. However, the model could not perform well due to noisy images in the training set. A pretrained ResNet50 model was used to classify TB into Multi-Drug Resistant and drug sensitive TB [36]. The model achieved an accuracy of 61.78%. However, the authors could have focused more on the abnormal portions of the lungs to improve accuracy. Che et al. [37] combined Mixup for data enhancement and ShuffleNet V2 for the classification of TB. The model achieved a decent AUC of 79.1. Nevertheless, the authors could have handled the imbalance in the dataset to achieve better performance scores.

Graph cut algorithm for segmentation and the binary classifier for classification was proposed in [38] to detect TB. The model was tested on two datasets, and it achieved an accuracy of 78.3 and 84% on the two datasets, respectively. Hooda et al. [39] proposed a CNN with seven convolutional layers and three fully connected layers to classify X-ray images into normal and abnormal. A decent accuracy of 94.73% was achieved with Adam’s optimizer. An ensemble of AlexNet and GoogLeNet was proposed in [40] to differentiate between pulmonary TB and normal cases. The model successfully achieved the goal with an AUC score of 0.99. However, the model is yet to be tested on images with subtle opacities. A CNN optimized for the task of TB detection along with Grad-CAM was proposed in [41].

GoogleNet, ResNet, and VGG were employed to extract features, and SVM was utilized to classify the features to
achieve the detection of TB in [42]. The model proposed achieved a decent accuracy, but the voting mechanism used by the models was a simple one. Similarly, Samuel et al. [43] proposed pretrained Inception V3 for feature extraction and SVM for the classification of TB. The model achieved a decent accuracy of 95.05%, but the model still had scope for improvement in the sensitivity score. Image enhancement techniques and pretrained ResNet50 and EfficientNet were employed in [44] to achieve the task of TB classification. The model was successful in detecting TB with an accuracy of 94.8%. However, the authors are yet to explore recent improvements in image processing techniques that can help improve the performance of deep learning models. Nine different models were used to classify TB from a vast dataset, and ChexNet was chosen as the best model for the task in [45]. A series of image processing techniques accompanied with feature extraction and crow search-based deep convolutional neural network (FC-SVNN) for classification of infection level of TB was proposed in [46]. The model achieved an accuracy of 93.5.

Guo et al. [11] proposed an approach to diagnosing TB by fine-tuning the CNN models using an artificial bee colony algorithm and later implementing an ensemble of those models to obtain the classification results. The approach attained an accuracy of 98% Gabor, Gist, histogram of oriented gradients (HOG), and pyramid histogram of oriented gradients (PHOG) features were extracted in [47] to differentiate TB from non-TB cases. SVM was employed as the classifier to classify the instances. The methodology obtained an accuracy of 92%. However, the authors are yet to solve the issue in cases where the cancer is misclassified as TB due to similar radiological patterns. Liu et al. [18] solved the most pressing issue of lack of dataset by creating a large dataset of TB cases with bounding box annotations and evaluation metrics that helped the researchers employ CNNs to achieve the TB diagnosis. They also summarized their results after running pretrained image classification models and TB area detection models. The proposed dataset has proven to be very useful for researchers seeking a career in this field.

The COVID-19 pandemic continues to pose several challenges to medical systems worldwide, and the ability to make quick clinical choices is critical [48]. Predictive machine learning algorithms that analyze medical images and estimate risk are pretty valuable.

Goel et al. [49] proposed an Optimized Convolutional Neural network (OptCoNet) to diagnose COVID-19 from chest X-ray images. Grey Wolf Optimization algorithm was used for optimization. The proposed approach attained an accuracy of 97.78%. However, the controlling parameters would have to be chosen carefully when the feature set increases. InceptionV3, Xception, and ResNetXt models were utilized to classify the chest X-ray images into normal and covid-19 [50]. The Xception model outperformed the other models by achieving an accuracy of 97.97%. Similarly, Apostolopoulos et al. [51] proposed VGG19, MobileNet v2, Inception, Xception, and Inception ResNet v2 to classify chest X-rays into common bacterial pneumonia, confirmed Covid-19 disease, and normal cases. The best accuracy achieved was 96.78%. However, the authors are yet to research the existence of biomarkers related to the Covid-19 disease. To identify COVID-19 from chest X-rays and CT images, [52] used an Inception Residual Recurrent Convolutional Neural Network with Transfer Learning (TL) technique and the NABLA-N network model for segmenting the areas affected by COVID-19. The approach attained an accuracy of 84.67% for X-ray images and 98.78% for CT images. Minaee et al. [53] detected COVID-19 from chest X-rays by employing ResNet18, ResNet50, SqueezeNet, and DenseNet-121. All the network models attained a favorable sensitivity rate of 98%. However, the authors still need to estimate the efficiency of the models with a larger dataset. Patch-based CNN with saliency maps to assist in diagnosing COVID-19 disease was proposed in [54]. The model trained well with smaller datasets. Kumar et al. [55] extracted discriminative features from the chest X-rays using ResNet52 to detect COVID-19. They employed SMOTE to handle the imbalance in the dataset and utilized Random Forest and XGBoost algorithms for classification. Their methodology obtained an accuracy of 97.3% with Random Forest and 97.7% with XGBoost. Pretrained AlexNet architecture to extract features that would aid in differentiating the chest X-rays into COVID-19, pneumonia, and Healthy categories was proposed in [56]. The features were later classified using the SVM algorithm. The model attained a good accuracy of 99.18%. However, the choice of optimal parameters for the SVM algorithm will affect the system’s performance.

Wang et al. [57] achieved pneumonia detection and lesion type classification by employing a novel prior-attention residual learning strategy with two 3-D ResNets. The approach demonstrated promising results with an accuracy of 93.3%. An approach to detect COVID-19 from CT images using weak labels was proposed in [58]. They utilized an attention-based deep 3D multiple instance learning (AD3D-MIL) and assigned the scans to patient-level labels. The model achieved an accuracy of 97.9%. Ouyang [59] proposed an online attention module with a 3D convolutional network (CNN) to focus on lung infection regions to differentiate pneumonia from COVID-19 disease. Their methodology attained an accuracy of 87.5%. However, the correlation between the localizations generated by the attention module and the imaging signs used in the clinical diagnosis is yet to be determined. Kang et al.
[60] employed multi-view representation learning and latent representation learning to obtain a precise classification between COVID-19 and community-acquired pneumonia. The model attained an accuracy of 95.5%. However, their approach could not identify normal cases. Chandra et al. [61] employed five supervised classification algorithms and computed the majority of the votes to obtain the aggregated results. The approach obtained an accuracy of 93.41%. However, a deeper comparison between the conventional algorithms and deep learning approaches is yet to be performed. The performance of 16 pretrained CNNs in classifying CT scans into COVID-19 and non-COVID-19 was proposed in [62]. The authors identified that DenseNet-201 outperformed the rest of the models by achieving the highest accuracy of 84.7%. Pham [63] employed AlexNet, GoogleNet, and SqueezeNet to classify COVID-19 from chest X-rays. The author identified that the models could generate acceptable results without the data augmentation with fine-tuning. Overall summary of the studies for pneumonia classification, TB classification, and COVID-19 classification using chest X-rays is shown in Tables 1, 2 and 3 respectively.

The above literature survey shows that most of the existing methods based on ImageNet pretrained models. Though the existing works shows a detailed analysis results, most of them are based on a single pretrained model and most of the studies are specifically showed the performances of the proposed method for only one lung disease. In addition, the existing methods used flattening or global average pooling in the process of final layer of feature extraction from ImageNet pretrained models. In this work, multichannel EfficientNet ImageNet based pretrained models are finetuned and stacking ensembling approach is proposed for lung disease detection. More than one EfficientNet model in this proposed work along with feature fusion and stacking ensembling approach helps to learn better feature representation which leads to higher performances for lung disease detection. The proposed method’s performance is shown on more one lung diseases such as pneumonia, TB, and COVID-19. In addition to Normal and TB classes, the TB lung disease dataset contains Sick but not TB patient CXR samples.

### Table 1: Related works in the literature to diagnose pneumonia

| Sl. No. | Dataset | Methodology | Accuracy | Reference |
|---------|---------|-------------|----------|-----------|
| 1       | OCT and chest X-ray images for classification | 18 Layer CNN model | 93.75 | [20] |
| 2       | OCT and chest X-ray images for classification | CNN model | 95.3 | [21] |
| 3       | OCT and chest X-ray images for classification | 2D Wavelet Transform and Random Forest | 97.11 | [22] |
| 4       | OCT and chest X-ray images for classification | CNN without Transfer Learning | 95.31 | [23] |
| 5       | Chest X-ray | Multilayer Perceptron and CNN | 94.4 | [24] |
| 6       | OCT and chest X-ray images for classification | Image Sharpening and customized CNN | 97.92 | [25] |
| 7       | OCT and Chest X-ray images for classification | CNN and Grad-CAM | 99.3 | [26] |
| 8       | Chest X-ray images (pneumonia) | AlexNet, ResNet18, DenseNet201 and SqueezeNet | 98 | [27] |
| 9       | OCT and chest X-ray images for classification | VGG16, VGG19, DenseNet201, Inception_ResNet_V2, Inception_V3, Resnet50, MobileNet_V2 and Xception | 96.61 | [28] |
| 10      | OCT and chest X-ray images for classification | Xception Net | 96.2 (AUC) | [29] |
| 11      | Mendeley data | DenseNet | 88.78 | [30] |
| 12      | OCT and chest X-ray images for classification | Inception V3 | 90.1 | [31] |
| 13      | Mendeley OCT and chest X-ray | VGG16, VGG19, Image Processing techniques | 96.2 | [32] |
| 14      | OCT and chest X-ray images for classification | PneumoniaNet | 95.83 | [33] |
| 15      | Guangzhou Women and Children’s Medical Center dataset | Ensemble of transfer learning using AlexNet, DenseNet121, InceptionV3,resNet18 and GoogLeNet neural networks | 96.4 | [34] |
| 16      | Chest X-ray images for classification | EfficientNet-B0, EfficientNet-B1 EfficientNet-B2 | 98 | Proposed approach |
3 Proposed multichannel EfficientNet deep learning-based stacking approach for lung disease detection

The proposed multichannel EfficientNet deep learning-based stacking approach for lung disease detection is shown in Fig. 1. The proposed framework contains the following

**Input layer** In the input layer, the model contains the chest X-rays of Normal and lung disease. The pixel values in the chest X-rays are normalized to be between 0 and 1.

**Hidden layer** The hidden layer is responsible for optimal feature extraction to accurately detect the optimal features to classify the lung disease. The hidden layer contains multichannel EfficientNet models such as EfficientNet-B0, EfficientNet-B1, and EfficientNet-B2. The properties of these models are tabulated in Table 4. The EfficientNet models are trained on ImageNet databases. It is one of the bigger datasets most commonly used for benchmarking models for image classification. This database contains 1,000 classes and the samples are annotated by humans. The pretrained models have rich features that can classify an image into its corresponding categories. With the aim to increase the performance of the models in ImageNet database, researchers have done various works by introducing CNN architectures. These architectures contain deeper depth or width or more input image resolution. In addition, though these pretrained models are based on natural images, fine-tuning the weights of the model on medical image classification shows better performances. Thus, in this work, the EfficientNet pretrained models were used as a transfer learning model with the aim to achieve better performances for lung disease classification using chest CXR images. This type of learning approach for lung disease classification using CXR images can reduce the training time, faster convergence rate, and achieve optimal performances in detecting patients’ data samples of chest CXR as either lung disease or normal. During backpropagation, the weights of the hidden layer are fine-tuned. Binary cross entropy loss function is used in this work and it is defined as follows

\[
\text{loss}(e_{\text{lung}}, p_{\text{lung}}) = -\frac{1}{N} \sum_{j=1}^{N} e_{\text{lung}} \log p_{\text{lung}} \\
+ [1 - e_{\text{lung}}] \log [1 - p_{\text{lung}}]
\]

where \(e_{\text{lung}}\) denotes expected class label and \(p_{\text{lung}}\) denoted predicted class label.
The input dimension included in the Table 4 indicates the shape of the chest X-ray. The EfficientNet model requires a fixed input dimension and there may be a chance that information loss can happen with this approach. Since we didn’t have access to clinical experts, we haven’t done any study on information loss during CXR image resize. This is very important in the healthcare and medical domain and this will be another direction of future work of the proposed work.

EfficientNet [64], as the name suggests, has shown high computational efficiency and has achieved top metrics of 84.4% accuracy on the ImageNet database. Model scaling is a technique that involves scaling the model in terms of depth, resolution, and width of the model to enhance its performance. Researchers have identified that depthwise scaling is the most famous among all the scaling techniques. One such example is the scaling of ResNet from ResNet18 to ResNet200. Studies have shown that traditional scaling helps to improve the performance of the models. However, after a certain point, traditional manual scaling degrades the performance. To solve this issue, EfficientNet was developed in [64]. It employed the compound scaling technique that used the strategy of scaling all three attributes together instead of scaling just one attribute. The compound scaling technique enhanced the performance of the models. The studies showed that the technique could be incorporated into any CNN, but the final model’s performance would depend on the baseline structure. Motivated by this fact, EfficientNet architecture was built using a multi-objective neural architecture search. It could optimize both accuracy and floating-point operations. The base model was EfficientNet-B0, based on

### Table 3 Related works in the literature to diagnose COVID-19

| Sl. No. | Dataset | Methodology | Accuracy (in percent) | Reference |
|--------|---------|-------------|-----------------------|-----------|
| 1      | Chest Imaging,SIRM COVID-19 Database, COVID-19 image data collection, COVID-19 Chest X-ray Dataset , Chest X-ray images (pneumonia) | Optimized Convolutional Neural network and Grey Wolf Optimization algorithm | 97.78 | [49] |
| 2      | Chest X-ray (Covid-19 & pneumonia) | Inception V3, Xception, and ResNeXt | 97.97 | [50] |
| 3      | COVID-19 image data,Radiological Society of North America (RSNA), Radiopaedia,Italian Society of Medical and Interventional Radiology (SIRM) | VGG19, MobileNet v2, Inception, Xception and Inception ResNet v2 | 96.78 | [51] |
| 4      | Kaggle Chest x-ray images and LUNA16 | Residual Recurrent Convolutional Neural Network and NABLA-N network | 98.78 (with CT images) | [52] |
| 5      | COVID-Xray-5k | ResNet18, ResNet50, SqueezeNet, and DenseNet-121 | 98 (sensitivity) | [53] |
| 6      | Japanese Society of Radiological Technology (JSTR) & NLM | Patch Based CNN with Saliency Map | 88.9 | [54] |
| 7      | Chest X-ray(pneumonia) & Italy Dataset | ResNet52, Random Forest and XGBoost algorithm | 9770.00% | [55] |
| 8      | COVID-19 Radiography database, COVID-chestxray-dataset,Chest X-Ray Images (pneumonia) &COVID-dataset | Pretrained AlexNet and SVM algorithm | 99.18 | [56] |
| 9      | Dataset from various hospitals | 3D ResNets with a prior-attention strategy | 93.3 | [57] |
| 10     | Shandong Province Hospital | Attention-based deep 3D multiple instance learning | 97.9 | [58] |
| 11     | Dataset from various hospitals | Online attention module with a 3D convolutional network (CNN) | 87.5 | [59] |
| 12     | Dataset from various hospitals | Multi-view representation learning and latent representation learning | 95.5 | [60] |
| 13     | COVID-Chestxray set, Montgomery set and NIH Chest X-ray14 set | Decision Tree, SVM, k-nearest neighbor, naive Bayes, ANN | 93.41 | [61] |
| 14     | COVID-19 CT database | DenseNet-201 | 84.7 | [62] |
| 15     | COVID-19 Radiography Database, COVID-19 Chest X-Ray Dataset Initiative and IEEE8023/Covid Chest X-Ray Dataset | AlexNet, GoogleNet, and SqueezeNet | 96 | [63] |
| 16     | Chest X-ray Images for Classification | EfficientNet-B0, EfficientNet-B1, EfficientNet-B2 | 98 | Proposed approach |
which a family of EfficientNets from B1-B7 were built that achieved top-1% accuracy on the ImageNet dataset. The performance of eight different scales CNN architectures on ImageNet dataset was shown in [64]. The baseline architecture EfficientNet-B0 had 5.3 million parameters and took 224*224 images as the input. The EfficientNet-B7 model had 66 million parameters and took 600*600 images as the input. In the proposed work, the ability of the compound scaling technique is leveraged to obtain precise results.

Global average pooling The EfficientNet model includes a series of convolution and pooling layers, followed by more than one fully connected layer. The fully connected layers contain the most number of parameters and thus carry the risk of overfitting and hindering the model’s generalization ability. In addition, these layers are heavily dependent on dropout regularization to reduce overfitting. To overcome this, we replace the topmost fully connected layer in the CNN with a Global average pooling (GAP) layer. GAP layer is integrated after the EfficientNet model that learns excellent localization. Though the problem is lung disease classification, integrating GAP layer learns important features of the infected region of the chest CXR sample. This type of learning capability helps the EfficientNet model to achieve better performance for lung disease detection.

### Table 4 Properties of EfficientNet models

| Model       | Size | Input dimension | Parameters   |
|-------------|------|-----------------|--------------|
| EfficientNet-B0 | 75   | 240 x 240       | 4,050,845    |
| EfficientNet-B1 | 31   | 260 x 260       | 6,576,513    |
| EfficientNet-B2 | 36   | 300 x 300       | 7,769,971    |

Fig. 1 Multichannel EfficientNet deep learning-based stacking approach for lung disease detection
disease classification. In addition, it reduces overfitting and can be used to reduce the spatial dimension of tensor.

**Feature fusion and classification** The features from EfficientNet-B0, EfficientNet-B1, and EfficientNet-B2 models are extracted. The dimension of these features of EfficientNet-B0, EfficientNet-B1, and EfficientNet-B2 models are 1280, 1280, and 1408 respectively. Since there was no improvement in accuracy after B2 model, the other EfficientNet models are not considered for lung disease detection. The features of EfficientNet-B0, EfficientNet-B1, and EfficientNet-B2 models are concatenated and the total dimension of the resulting feature vector is 3968. Since the dimension of the features was big, the features passed into more than one fully connected layer that helps to learn the optimal features required for lung disease classification in the output layer. In between the fully connected layer dropout (0.001) and batch normalization was included. Dropout helps to avoid overfitting during training process of lung disease classification. Batch normalization increases the speed of the learning process and stabilizes the whole learning process by applying a transformation that maintains the mean output close to 0 and the output standard deviation close to 1. The model contains the 3 fully connected layers in between the feature fusion and output or lung disease classification layer. In a fully connected layer, the neurons at the present layer connect to all the neurons of the next layer. The first fully connected layer contains 4000 neurons, second fully connected layer contains 1000 neurons, and third fully connected layer contains 64 neurons. These three layers help to map the features into a higher dimensional space in which the most important features are extracted to accurately classify the chest X-rays of lung diseases. The output layer contains a fully connected layer with one input neuron sigmoid activation function. Further to enhance the performance, the final output layer is replaced by stacking ensemble classifiers. Stacking is an ensemble technique that combines heterogeneous classifiers to estimate and correct their biases [46]. The member or base-level classifiers are trained using different learning algorithms and combined using a meta-level classifier. Stacking ensemble learning is a two-stage approach, the first stage contains SVM and Random forest and the second stage classifies the chest X-rays as either Normal or Lung disease using the predictions of the first stage. This helps to classify the data samples into either Normal or lung disease. The step by step approach of the proposed approach for lung disease detection using multichannel EfficientNet deep learning-based stacking approach is shown in Fig. 1. The proposed model takes chest X-ray samples as input and outputs a value as either Normal or Lung disease.

### 4 Description of lung disease datasets

In this work, the performance of the proposed method is evaluated on 3 types of lung diseases. They are pneumonia, TB, and COVID-19. The detailed statics of all 3 lung disease data are included in Table 5.

**Pediatric pneumonia** The pediatric pneumonia CXR database collected in Guangzhou Women and Children’s Medical Center [16]. Though there are many CXR datasets available for pneumonia classification, literature survey shows that there is only one benchmark dataset i.e. Guangzhou Women and Children’s Medical Center publicly available for research for pediatric pneumonia classification. It contains CXR images of retrospective cohorts of pediatric patients of one to five years old. During the dataset preparation, low-quality and unreadable CXR images were removed and diagnosis was performed by two doctors and after that the validation was done by another expert to avoid misclassification. The database has train and test and both contain CXR image samples for normal and pneumonia patients. Both the train and test datasets are completely unseen and have different distributions.

**Tuberculosis** There are 5 well-known benchmark datasets available for TB. Except [18] all other 4 datasets contain less number chest X-ray samples for Normal and TB. Thus, in this work a dataset developed by Liu et al. used [18]. The dataset was prepared by 5-10 radiologists and to avoid errors in labeling, double checking was done by another set of expert radiologists. Most of the existing datasets contain only chest X-ray samples for Normal and TB patients. But this dataset provides samples for additional class such as Sick but not TB.

**COVID-19** There are many chest X-rays datasets publicly available for COVID-19. However, most of the datasets are very small. In this work, the database was taken from Mendeley [65]. This dataset contain data samples for both chest X-ray and CT scan. But, in this work, only chest X-ray dataset is used.

| Lung disease      | Class     | Training | Testing | Total |
|-------------------|-----------|----------|---------|-------|
| pediatric pneumonia | Normal    | 930      | 2732    | 3662  |
|                   | pneumonia | 419      | 1151    | 1570  |
| Tuberculosis       | Normal    | 2589     | 1211    | 3800  |
|                   | Tuberculosis | 2724   | 1076    | 3800  |
|                   | Sick      | 567      | 233     | 800   |
| COVID-19           | Normal    | 2695     | 1981    | 4676  |
|                   | COVID-19 | 1153     | 851     | 2004  |
The chest X-rays of Normal and pediatric pneumonia is shown in Fig. 2. First two images are for Normal patients and next two images are from pediatric pneumonia patients. COVID-19 and Normal patients samples are shown in Fig. 3. In this figure, the first two images are for Normal patients and next two images are for COVID-19. In Fig. 4, the first row contains chest X-rays for Normal patients, second row contains chest X-rays of Sick but not TB patients, and third row contains chest X-rays of TB. It can be seen from the figures that the chest X-rays cannot be easily distinguished from the Normal and lung disease patient by radiologist. There may be possibility that even expert radiologists can make mistakes and to avoid this, in this work, a deep learning-based approach was employed to automatically classify the chest X-ray as Normal and lung diseases such COVID-19, pediatric pneumonia, and TB. The developed tool can assist radiologists to accurately detect lung disease and this can act like an early diagnosis tool for point-of-care diagnosis. This type of tool will be very helpful in developing and underdeveloped countries where the people don’t have access to expert radiologists. Overall, this type of developed tool will be more effective to accurately identify lung diseases.

5 Statistical metrics

To evaluate the performance of the proposed approach for lung disease detection, the following statistical metrics are used in this research study.

\[
\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} 
\]  

(1)

Accuracy is defined as the classifier’s ability to mark all healthy chest X-ray samples as healthy and all lung disease chest x-ray samples as lung disease.

\[
\text{Precision} = \frac{TP}{FP + TP} 
\]  

(2)

Precision is defined as a measure of the ability of a classifier to not mark a lung disease chest x-ray sample as healthy.

\[
\text{Recall} = \frac{TP}{TP + FN} 
\]  

(3)

Recall is defined as a measure of the ability of a classifier to mark all healthy chest X-ray samples as healthy.

\[
F1 - \text{score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} 
\]  

(4)

F1-score is the weighted average of precision and recall.

In the above statistical metrics, \(TP, TN, FP,\) and \(FN\) define true positive, true negative, false positive, and false negative respectively. These values are obtained from a confusion matrix. A confusion matrix (also called an error matrix) is a table that is used to describe the performance of a classifier. It provides the number of true positives, true negatives, false positives, and false negatives.

- True positive (TP) : lung disease chest x-ray identified as lung disease chest x-ray.
- True negative (TN) : healthy chest x-ray identified as healthy chest x-ray.
- False positive (FP) : healthy chest x-ray misclassified as lung disease chest x-ray.
- False negative (FN) : lung disease chest x-ray misclassified as healthy chest x-ray.

In this work, Precision, Recall, and F1-Score performances of the proposed method for lung disease detection is reported in both macro and weighted. In the case of macro-average, all classes have an equal contribution to the final average. For example, macro-averaged recall is performed by first computing the recall of each class and then taking
an average of all recalls. In the case of weighted-average, each class has a contribution to the final average that is weighted by its size. Both macro and weighted metrics compute the precision, recall, and F1-score for each class in the dataset and return the average.

6 Experiments, results and discussions

All the models were implemented using TensorFlow\textsuperscript{1} as back end with Keras\textsuperscript{2} front end library and scikit-learn\textsuperscript{3} was used for implementing machine learning algorithms.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{fig3}
\caption{Chest X-rays of Normal and COVID-19 patients (Left to right, first two images are for Normal chest X-rays and next two images are for COVID-19 chest X-rays)}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{fig4}
\caption{Chest X-rays of Normal (Row 1), Sick but not TB (Row 2), and TB (Row 3)}
\end{figure}
The experiments for all the models were run on Kaggle NVidia K80 GPUs\(^4\).

In this work, CNN-based pretrained models of ImageNet are fine-tuned for various lung disease datasets such as pediatric pneumonia, TB, and COVID-19. The models are given below

- Lung-M1
- Lung-M2
- Lung-M3
- Lung-M4
- Lung-M5
- Lung-M6
- Lung-M7

Model accuracy represents the percentage of accurate predictions done by the model. Model loss is the penalty for a bad prediction. Lower loss indicates better prediction. The model accuracy and loss were computed and plotted for various models on different datasets in the proposed work. Figure 5 represents the accuracy of the various models obtained on the pneumonia training dataset. The Lung-M7 (proposed) model achieved the highest training accuracy (98\%). Figure 6 represents the various model loss. The Lung-M7 model achieved lower loss when compared to other models indicating better prediction capability. The accuracy and model loss were also plotted on the Tuberculosis dataset and COVID-19 dataset. Figure 7 represents the accuracy of various models on the Tuberculosis dataset. As the figure represents, the Lung-M7 model outperformed the other models with an accuracy of 99\%. The model loss after running the models with the Tuberculosis dataset is represented in the Fig. 8. The Lung-M7 model achieved the lowest loss even with this training set. Figure 9 represents the model accuracy achieved by various models on the COVID-19 dataset. The Lung-M7 achieved the highest accuracy of 98\%. The model loss is plotted in Fig. 10. The Lung-M7 model attained lower loss when compared to the Lung-M1 model. The model’s accuracy and loss achieved by the Lung-M7 model in all the different datasets proved its efficacy.

Lung-M4, Lung-M5, and Lung-M6 models are based on EfficientNet-B0, EfficientNet-B1, and EfficientNet-B2 respectively. The proposed approach uses multichannel EfficientNet models i.e. EfficientNet-B0, EfficientNet-B1, and EfficientNet-B2. The existing methods are Lung-M1 [16], Lung-M2 [18], and Lung-M3 [50] and these methods are based on VGG-16, ResNet-50, and InceptionV3 respectively. Lung-M7 is the proposed method for lung disease classification. Parameter values for optimizer, learning rate, batch size and epochs are set based on hyperparameter tuning. Various trails of experiments were run and found adam, 0.0001, 64, and 15 value for optimizer, learning rate, batch size, and epochs respectively. To identify the best parameter for optimizer, two trails of experiments were run for adam and SGD till 10 epochs. The experiments with adam performed better than SGD. When the experiments were run for more than 10 epochs, the model has attained improvement in performance such as increase in training accuracy and decrease in training loss till 15 epochs. When we further run the experiments till 20 epochs, the model didn’t show any performance improvement. So to avoid overfitting, the experiments were stopped at epochs 15. Next, two trails of experiments were run for learning rates 0.0001, 0.001, 0.01, and 0.1. The experiments with 0.0001 were performed better compared to other learning rates. Further, to identify the optimal value for batch size, two trails of experiments were run with batch size 16, 32, 63, 96, and 128. The model with 64 batch size performed better than other values. During training, the data samples were shuffled and initially the models were initialized with ImageNet-based model weights. Further these weights are fine tuned across 15 epochs. The training accuracy and training loss for proposed multichannel EfficientNet model is shown in Figs. 5 and 6 for pediatric pneumonia, Figs. 7 and 8 for TB, and Figs. 9 and 10 COVID-19 respectively. The proposed approach, Lung-M7 has shown improvement in training accuracy and training loss across 15 epochs for all the lung diseases as shown in figures. Except Lung-M1, most of the models achieved 90\% training accuracy within 5 epochs and after showed slightly better training accuracy till 15 epochs. However, the Lung-M1, took more than 10 epochs to show more than 90\% training accuracy. Most importantly, all these models showed similar training accuracy.

---

\(^1\) https://www.tensorflow.org/.

\(^2\) https://keras.io/.

\(^3\) https://scikit-learn.org/.

\(^4\) https://www.kaggle.com/.
and loss on all three lung disease datasets. This may be due to the reason that all the datasets are chest x-rays of lung diseases.

Next, the trained models of Lung-M1, Lung-M2, Lung-M3, Lung-M4, Lung-M5, Lung-M6, and Lung-M7 are evaluated on the test datasets of pediatric pneumonia and the result is tabulated in Table 6. For all the models, the performances are reported in terms of accuracy, weighted and macro precision, weighted and macro recall, and weighted and macro F1-Score. The proposed model has shown 98% for pediatric pneumonia which is 1%, 2%, 3% better than the EfficientNet-B2, EfficientNet-B1, and EfficientNet-B0 respectively. This collectively learns the best features of EfficientNet-B0, EfficientNet-B1, and EfficientNet-B2. In addition, the proposed approach performed better than the existing models such as Lung-M1, Lung-M2, and Lung-M3. The accuracy of Lung-M1, Lung-M2, and Lung-M3 for pediatric pneumonia classification are 91%, 92%, and 93% respectively. Overall the proposed multichannel EfficientNet approach performed better than the single EfficientNet models and other ImageNet-based models such as VGG-16, InceptionV3, and ResNet50. Similar to accuracy the proposed model showed better
performances in terms of macro and weighted precision, macro and weighted recall, and macro and weighted F1-Score compared to the single EfficientNet models and existing ImageNet based models such as ResNet-50, InceptionV3, and VGG-16.

The model performance for TB is reported in Table 7. Lung-M7 showed 99% accuracy which is better than other lung disease detection models. Lung-M4, Lung-M5, and Lung-M6 accuracy for TB detection was 93%, 94%, and 94% respectively. The existing approaches such as Lung-M1, Lung-M2, and Lung-M3 accuracy was 72%, 77%, and 91% respectively. Overall, the proposed approach performs better than the single EfficientNet models and other existing models based on InceptionV3, ResNet50, and VGG16. Similar to pediatric pneumonia classification, all the models showed similar performance on TB classification. One important factor in TB classification is, the dataset has three classes instead of two classes. This indicates that the proposed model in this work has the capability to achieve better performances for more than two class classification. The model performance in terms of macro and weighted precision, macro and weighted recall, and macro and weighted F1-score was good for the proposed method compared to the existing models based on InceptionV3, ResNet50, and VGG16. Also, the multichannel EfficientNet model performed better than the single channel EfficientNet models.

COVID-19 lung disease classification results are reported in Table 8. The proposed approach has shown 98% accuracy which is 10%, 7%, 4% higher than the Lung-M4, Lung-M5, and Lung-M6. Lung-M6 performance is better than the Lung-M5 and Lung-M4, also these models performed better than the existing approaches based on InceptionV3, ResNet50, and VGG16. The VGG-16,
ResNet50, and InceptionV3 models accuracy is 74%, 81% and 86% respectively. Similar to the pediatric pneumonia lung disease classification, the models have shown similar performances for COVID-19 detection. The proposed method has shown better performances in terms of macro and weighted precision, macro and weighted recall, and macro and weighted F1-Score than the InceptionV3, VGG-16, ResNet50 and single channel EfficientNet models.

The detailed results of Lung-M7 for pediatric pneumonia, TB, and COVID-19 are reported in Tables 9, 10, 11 respectively. The model has shown more than 90% performances in terms of accuracy, precision, recall, and F1-score for all the classes in pediatric pneumonia, TB, and COVID-19. Thus the model can detect different lung diseases such as COVID-19, pediatric pneumonia and TB. Though the proposed model optimized for pediatric pneumonia classification, it achieved similar performances on other two lung diseases such as COVID-19 and TB. Mainly, the proposed model showed similar performances on all the classes of TB. This indicates that the model is robust and generalizable. Confusion matrix is a matrix plotted to understand the classification capability of models on the test data. The confusion matrix for the proposed model for Lung disease detection is shown in Fig. 11.

| Models               | Accuracy | Type   | Precision | Recall | F1-score |
|----------------------|----------|--------|-----------|--------|----------|
| Lung-M1              | 0.74     | Macro  | 0.81      | 0.69   | 0.69     |
|                      |          |        | 0.79      | 0.74   | 0.71     |
| Lung-M2              | 0.81     | Macro  | 0.84      | 0.79   | 0.80     |
|                      |          |        | 0.83      | 0.81   | 0.81     |
| Lung-M3              | 0.86     | Macro  | 0.86      | 0.86   | 0.86     |
|                      |          |        | 0.87      | 0.86   | 0.86     |
| Lung-M4              | 0.88     | Macro  | 0.91      | 0.86   | 0.87     |
|                      |          |        | 0.90      | 0.88   | 0.88     |
| Lung-M5              | 0.91     | Macro  | 0.91      | 0.92   | 0.91     |
|                      |          |        | 0.92      | 0.91   | 0.91     |
| Lung-M6              | 0.94     | Macro  | 0.94      | 0.93   | 0.93     |
|                      |          |        | 0.94      | 0.94   | 0.94     |
| Lung-M7 (Proposed approach) | 0.98 | Macro  | 0.98      | 0.98   | 0.98     |
|                      |          |        | 0.98      | 0.98   | 0.98     |

Table 9 Class-wise results of proposed method, i.e. Lung-M7 on pneumonia lung disease dataset

| Class              | Precision | Recall | F1-score |
|--------------------|-----------|--------|----------|
| Normal             | 0.93      | 0.99   | 0.96     |
| pediatric pneumonia| 1.00      | 0.97   | 0.99     |
| Macro average      | 0.97      | 0.98   | 0.97     |
| Weighted average   | 0.98      | 0.98   | 0.98     |

Table 10 Class-wise results of proposed method, i.e. Lung-M7 on Tuberculosis lung disease dataset

| Class              | Precision | Recall | F1-score |
|--------------------|-----------|--------|----------|
| Normal             | 0.99      | 1.00   | 0.99     |
| Sick but non-TB    | 0.98      | 1.00   | 0.99     |
| TB                 | 1.00      | 0.92   | 0.96     |
| Macro average      | 0.99      | 0.97   | 0.98     |
| Weighted average   | 0.99      | 0.99   | 0.99     |

Table 11 Class-wise results of proposed method, i.e. Lung-M7 on COVID-19 lung disease dataset

| Class              | Precision | Recall | F1-score |
|--------------------|-----------|--------|----------|
| Normal             | 0.97      | 0.98   | 0.98     |
| COVID-19           | 0.99      | 0.98   | 0.98     |
| Macro average      | 0.98      | 0.98   | 0.98     |
| Weighted average   | 0.98      | 0.98   | 0.98     |

Figure 11a shows the confusion matrix obtained on the Pneumonia test set. 5 samples of pneumonia as Normal and 29 samples of pneumonia as Normal. This indicates that the model has less misclassification for the pneumonia class than the Normal. The prediction accuracy was 97.83%. Figure 11b represents the confusion matrix obtained on TB test set. The proposed model misclassified 5 samples of Healthy into Sick but not TB, 1 sample of Sick but non-TB into Healthy and TB and 7 samples of TB into Healthy, and 12 samples of TB into Sick but non-TB. Overall, the model learned to detect all three classes but showed more misclassification for TB. The prediction accuracy achieved was 98.97%. Further investigation is
required to understand the reason behind the misclassification, and improvement to the proposed method has to be made to avoid misclassification. The large number of misclassifications of TB may be due to the dataset being highly imbalanced. The dataset has 0.25% data compared to Normal and Sick but not-TB classes. The Fig. 11c represents the confusion matrix achieved on COVID-19 test dataset. The proposed model misclassified 15 samples of COVID-19 into Normal and 25 classes of Normal into COVID-19. Overall, the proposed model has large misclassification for Normal in COVID-19 and pediatric pneumonia and TB in Tuberculosis lung disease classification. The model achieved a prediction accuracy of 98%. The model performed well on all three datasets. However, further enhancement of the proposed method is required to avoid misclassifications, and in addition, a detailed evaluation study is needed to identify the reason behind
misclassifications. These are future directions of the current work.

6.1 t-SNE feature visualization

The deep learning models are complex in nature and they are considered to be black-box. Recent days, interpretation and explainable deep learning plays an important role in deep learning. t-SNE can be effectively used to represent the hidden layer and it is one of the most commonly used approaches in recent days. t-SNE has several important parameters and they are are n_components, perplexity, learning rate, iterations, and embedding initialization. In this work, n_components is set to 2, perplexity is set to 40.0, learning rate is set to 150.0, and embedding initialization is set to PCA. No parameter tuning is done for these parameters and it is very important to achieve optimal performances. This work makes an attempt to see how the distribution of the features from the hidden layers of the proposed approach for lung disease classification. In this work, the penultimate layer features of the proposed model are extracted and the feature dimension was 64. This was passed into t-SNE. It implicitly uses PCA to reduce the dimensionality of features into 2 dimensions. These 2 dimensions are two principal components which are shown in Fig. 12 X axis and Y axis respectively. This type of visualization approach allowed us to verify the lung disease and Normal CXR data samples in separate clusters. The t-SNE feature visualization for pediatric pneumonia, TB and COVID-19 is shown in Fig. 12 respectively. Figure 12a represents the plot obtained on the Pneumonia dataset. The t-SNE clustered Normal and Pneumonia classes without overlapping (only a few misclassifications). Figure 12b represents the plot obtained on the Tuberculosis dataset. The algorithm clustered the healthy and sick but not TB and TB classes. Figure 12c represents the plot obtained on the COVID-19 dataset. Though the Normal and lung disease classes have formed a distinct cluster, both have still shown overlapping regions, mainly in COVID-19. This may be because there may be a possibility that the Normal patients show some symptoms of pneumonia but not COVID-19. This can be avoided by adding sufficient data during the lung disease model training for COVID-19 classification. This indicates that further study is required to analyze the misclassified samples and further development is needed to minimize misclassified CXR samples of lung diseases. This type of work can be considered one of the significant directions of future works. In addition to penultimate layer feature visualization, other hidden layers can be considered and studied in detail about the feature importance in detecting lung diseases. Because there is no guarantee that the penultimate layer features are optimal. This type of study can further enhance the detection rate of lung diseases of the proposed method. Kindly provide access date for the reference [12]. Jan, 2022

6.2 Comparison with existing methods

The results of proposed method based on multichannel EfficientNet models such as EfficientNet-B0, EfficientNet-B1, and EfficientNet-B2 fine tuned model and other existing method performances in terms of accuracy for pediatric pneumonia, TB, and COVID-19 is tabulated in Table 12. The existing methods Lian et al. [16], Liu et al. [18], Jain et al. [50] are based on VGG16, ResNet50 and InceptionV3 respectively. Also, all the existing models are not based on multichannel and the features are not fused together. In addition, all these existing methods employ a fully connected layer instead of stacking ensemble approaches for classification. The stacked ensemble approach is a two-stage approach, in which the first stage contains random forest and SVM classifier and in the next stage logistic regression is used for classification that takes prediction as input from the classifiers of the first stage. The classifiers used in the stacked ensemble approach contains parameters and optimal performance implicitly relies on these parameters. By hyperparameter tuning, we set n_components = 70 and criterion = gini in random forest and kernel=rbf and c=50 in SVM. In logistic regression, sigmoid activation function was used that results 0 or 1, where 0 indicates Normal and 1 indicates lung disease. In addition to these parameters, the classifier of stacked ensemble approach contains other parameters. However, default values of sklearn machine learning library was used in this work for additional parameters of random forest, SVM and logistic regression. For all lung disease classification, the proposed showed better accuracy compared to the existing methods. Most importantly, the proposed method accuracy is 5%, 8%, and 12% higher than the existing methods. The existing method based on InceptionV3 performed better than ResNet50 and VGG16. In addition, the existing methods are not robust and generalizable to other lung diseases. But, the current work includes a detailed study and its evaluation on more than one lung diseases and in addition a detailed evaluation of experiments are shown in this work. Overall, the proposed method has attained better accuracy compared to the three existing related methods for lung disease classification such as pediatric pneumonia, TB, and COVID-19.

7 Conclusion and future works

This work has proposed a framework to classify lung diseases from chest X-rays. The framework leverages multichannel EfficientNet deep learning-based stacking
ensemble approach. EfficientNet-B0, EfficientNet-B1, and EfficientNet-B2 models serve as feature extractor and followed by the features of EfficientNet models are concatenated. The concatenated features are passed into more one fully connected layer to learn the optimal features and a stacked ensemble approach is used for classification. It uses random forest and support vector machine in the first stage for prediction and followed logistic regression in the second stage for classification. The proposed model has performed better than the EfficientNet models such as EfficientNet-B0, EfficientNet-B1, and EfficientNet-B2 for pediatric pneumonia, COVID-19, TB lung disease

---

**Table 12** Comparison of results of proposed method and other existing methods for lung diseases classification

| Lung disease       | Accuracy |
|--------------------|----------|
|                    | Proposed approach | Liang et al. [16] | Liu et al. [18] | Jain et al. [50] |
| pediatric pneumonia| 98       | 91         | 92          | 93          |
| Tuberculosis       | 99       | 72         | 77          | 91          |
| COVID-19           | 98       | 74         | 81          | 86          |

---

**Fig. 12** Penultimate layer feature visualization using t-SNE
classification. In addition, the proposed approach outperformed the existing approaches in all 3 lung diseases. This indicates that the proposed method is robust and generalizable on unseen data samples from lung diseases. However, a detailed investigation and analysis of the proposed has to be done on the dataset which was collected in real-time. This will be considered as one of the future works. In addition, deep learning and machine learning-based models are not robust against adversarial environments. The robustness of the proposed method for lung disease classification has to be done in an adversarial environment and this type of work is very important in the current smart healthcare environment. In the current work, the features extracted from EfficientNet-B0, EfficientNet-B1, and EfficientNet-B2 models are concatenated and instead an attention-based feature fusion approach can be employed to select the optimal features. This type of method can enhance the performances for lung disease classification and in addition decreases the computational complexity of the model by reducing the number of features. The proposed model just classifies the lung disease instead of region of infected lung disease identification in chest X-rays can be useful for medical experts. In addition, infected region detection of chest x-rays can enhance the performances in accurately classifying lung disease and in addition decrease the computational complexity of the proposed EfficientNet-based deep learning approach for lung disease classification. Instead of simple concatenation as a feature fusion, other feature fusion can be employed and this type of method can increase the lung disease detection performances. In the current work, the CXR images were resized and there is no justification towards that the medical values are not devalorizing because of resizing. This type of study is remained as future work.

There are common features exist among the three lung diseases from the experimental studies. Involving clinical researchers and validating the proposed model performance on a real-time patient dataset can be further done to know the robustness and generalizability of the proposed approach towards lung disease detection. Additional lung disease analysis can be easily added to the proposed work as the current work is based on multichannel or multidimensional deep learning. Please note that the references are renumbered to ensure sequential order of citations. Kindly check and confirm the change. CorrectReference [66] was provided in the reference list; however, this was not mentioned or cited in the manuscript. As a rule, if a citation is present in the text, then it should be present in the list. Please provide the location of where to insert the reference citation in the main body text. Kindly ensure that all references are cited in ascending numerical order. Remove reference [66]

**Funding** Not applicable.

**Data availability** Data used in this study are available from the first author upon request.

**Code availability** The Programming codes implemented in this study are available from the first author upon request.

**Declarations**

**Conflict of interest** The authors declare no conflict of interest.

**Ethical approval** None.

**References**

1. CDC—data and statistics—chronic obstructive pulmonary disease (COPD). (2021, June 14). https://www.cdc.gov/copd/data.html. Accessed 16 Jan 2022
2. Xu, Z., Sheffield, P.E., Hu, W., Su, H., Yu, W., Qi, X., Tong, S.: Climate change and children’s health-A call for research on what works to protect children. Int. J. Environ. Res. Public Health 9(9), 3298–3316 (2012)
3. D’Amato, G., Cecchi, L., Liccardi, G.: Thunderstorm-related asthma: not only grass pollen and spores. J. Allergy Clin. Immunol. 121(2), 537–538 (2008)
4. Kelly, F.J., Fussell, J.C.: Air pollution and airway disease. Clin. Exp. Allergy 41(8), 1059–1071 (2011)
5. Olusupo, O., Xiao, H., Brown, J.D.: Relative clinical and cost burden of community-acquired pneumonia hospitalizations in older adults in the United States-a cross-sectional analysis. Vaccines 6(3), 59 (2018)
6. Rudan, I., Tomaskovic, L., Boschi-Pinto, C., Campbell, H.: Global estimate of the incidence of clinical pneumonia among children under five years of age. Bull. World Health Org. 82, 895–903 (2004)
7. Mahomed, N., van Ginneken, B., Philipsen, R.H., Melendez, J., Moore, D.P., Moodley, H., Madhi, S.A.: Computer-aided diagnosis for World Health Organization-defined chest radiograph primary-endpoint pneumonia in children. Pediatr. Radiol. 50(4), 482–491 (2020)
8. Hosny, A., Parmar, C., Quackenbush, J., Schwartz, L.H., Aerts, H.J.: Artificial intelligence in radiology. Nat. Rev. Cancer 18(8), 500–510 (2018)
9. Hassaballah, M., Awad, A.I. (eds.): Deep Learning in Computer Vision: Principles and Applications. CRC Press, Boca Raton (2020)
10. WHO, G.: Global tuberculosis report 2020. Glob. Tuberc. Rep. (2020)
11. Guo, R., Passi, K., Jain, C.K.: Tuberculosis diagnostics and localization in chest X-rays via deep learning models. Front. Artif. Intell. 3, 74 (2020)
12. Who Coronavirus (COVID-19) Dashboard. https://covid19.who.int/. Accessed Jan 2022
13. Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Cao, B.: Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 395(10223), 497–506 (2020)
14. Kieu, S.T.H., Bade, A., Hijazi, M.H.A., Kolivand, H.: A survey of deep learning for lung disease detection on medical images: state-of-the-art, taxonomy, issues and future directions. J. Imaging 6(12), 131 (2020)

© Springer
15. Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., Sánchez, C.J.: A survey on deep learning in medical image analysis. Med. Image Anal. 42, 60–88 (2017)
16. Liang, G., Zheng, L.: A transfer learning method with deep residual network for pediatric pneumonia diagnosis. Comput. Methods Prog. Biomed. 187, 104964 (2020)
17. Moses, D.A.: Deep learning applied to automatic disease detection using chest X-rays. J. Med. Imaging Radiat. Oncol. 65(5), 498–517 (2021)
18. Liu, Y., Wu, Y.H., Ban, Y., Wang, H., Cheng, M.M.: Rethinking computer-aided tuberculosis diagnosis. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2646–2655 (2020)
19. Roberts, M., Driggs, D., Thorpe, M., Gilbey, J., Yeung, M., Ursprung, S., Schönlieb, C.B.: Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans. Nat. Mach. Intell. 3(3), 199–217 (2021)
20. Siddiqi, R.: Automated pneumonia diagnosis using a customized sequential convolutional neural network. In: Proceedings of the 2019 3rd International Conference on Deep Learning Technologies, pp. 64–70. (2019)
21. Saraiva, A.A., Ferreira, N.M.F., de Sousa, L.L., Costa, N.J.C., Sousa, J.V.M., Santos, D.B.S., Soares, A.: Classification of images of childhood pneumonia using convolutional neural networks. In: BIOIMAGING, pp. 112–119 (2019)
22. Akgundogdu, A.: Detection of pneumonia in chest X-ray images by using 2D discrete wavelet feature extraction with random forest. Int. J. Imaging Syst. Technol. 31(1), 82–93 (2021)
23. Stephen, O., Sain, M., Madhu, U.I., Jeong, D.U.: An efficient deep learning approach to pneumonia classification in healthcare. J. Healthc. Eng. (2019). https://doi.org/10.1155/2019/4180949
24. Saraiva, A.A., Santos, D.B.S., Costa, N.J.C., Sousa, J.V.M., Santos, D.B.S., Soares, A.: Classification of images of childhood pneumonia using convolutional neural networks. In: BIOIMAGING, pp. 112–119 (2019)
25. Nahid, A.A., Sidker, N., Bairagi, A.K., Razzaque, M., Masud, M., Kozhani, Z.A., Mahmud, M.A.: A novel method to identify pneumonia through analyzing chest radiographs employing a multichannel convolutional neural network. Sensors 20(12), 3482 (2020)
26. Nguyen, H., Huynh, H., Tran, T., Huynh, H.: Explanation of the convolutional network classifying chest X-ray images supporting pneumonia diagnosis, EAI Endorsed Trans. Context-Aware Syst. Appl. 7(21), e3 (2020)
27. Rahman, T., Chowdhury, M.E., Chakravorty, A., Islam, K.R., Islam, K.F., Mazhar, R., Chowdhury, M.E.: Reliable tuberculosis detection using chest X-ray with deep learning approach. IEEE Access 8, 217897–217907 (2020)
28. Mann, A., Rangarajan, V., Devi, D., Prabhu, K.: Automated detection of tuberculosis using chest X-ray image. In: 2019 International Conference on Advances in Computing, Communication and Automation (ICACCA), pp. 1–6. IEEE (2019)
29. Chikara, P., Singh, P., Gupta, P., Bhatia, T.: Deep convolutional neural network with transfer learning for detecting pneumonia on chest X-rays. In: Advances in Bioinformatics, Multimedia, and Electronics Circuits and Signals, pp. 155–168. Springer, Singapore (2020)
Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
Mamoun Alazab received his PhD degree in Computer Science from the Federation University of Australia, School of Science, Information Technology and Engineering. He is an Associate Professor in the College of Engineering, IT and Environment at Charles Darwin University, Australia. He is a Cyber Security researcher and practitioner with industry and academic experience. Alazab’s research is multidisciplinary that focuses on Cyber Security and digital forensics of computer systems including current and emerging issues in the cyber environment like cyber-physical systems and internet of things with a focus on cyber crime detection and prevention. He has more than 100 research papers. He delivered many invited and keynote speeches, 22 events in 2018 alone. He convened and chaired more than 50 conferences and workshops. He works closely with government and industry on many projects. He is an editor on multiple editorial boards including Associate Editor of IEEE Access (2017 Impact Factor 3.5), Editor of the Security and Communication Networks Journal (2016 Impact Factor: 1.067) and Book Review Section Editor: Journal of Digital Forensics, Security and Law (JDFSL). He is a Senior Member of the IEEE.