

High-precision multiclass classification of lung disease through customized MobileNetV2 from chest X-ray images

FM Javed Mehedi Shamrat^a, Sami Azam^{b, **}, Asif Karim^{b, *}, Kawsar Ahmed^{c, d}, Francis M. Bui^c, Friso De Boer^b

^a Department of Software Engineering, Daffodil International University, Birulia, 1216, Dhaka, Bangladesh

^b Faculty of Science and Technology, Charles Darwin University, Casuarina, NT 0909, Australia

^c Department of Electrical and Computer Engineering, University of Saskatchewan, Saskatoon, SK S7N 5A9, Canada

^d Group of Bio-photomatiX, Department of Information and Communication Technology, Mawlana Bhashani Science and Technology University, Tangail, 1902, Bangladesh

ARTICLE INFO

Keywords:

MobileLungNetV2
MobileNetV2
CNN
Lung disease
Deep learning
Multiclass classification

ABSTRACT

In this study, multiple lung diseases are diagnosed with the help of the Neural Network algorithm. Specifically, Emphysema, Infiltration, Mass, Pleural Thickening, Pneumonia, Pneumothorax, Atelectasis, Edema, Effusion, Hernia, Cardiomegaly, Pulmonary Fibrosis, Nodule, and Consolidation, are studied from the ChestX-ray14 dataset. A proposed fine-tuned MobileLungNetV2 model is employed for analysis. Initially, pre-processing is done on the X-ray images from the dataset using CLAHE to increase image contrast. Additionally, a Gaussian Filter, to denoise images, and data augmentation methods are used. The pre-processed images are fed into several transfer learning models; such as InceptionV3, AlexNet, DenseNet121, VGG19, and MobileNetV2. Among these models, MobileNetV2 performed with the highest accuracy of 91.6% in overall classifying lesions on Chest X-ray Images. This model is then fine-tuned to optimise the MobileLungNetV2 model. On the pre-processed data, the fine-tuned model, MobileLungNetV2, achieves an extraordinary classification accuracy of 96.97%. Using a confusion matrix for all the classes, it is determined that the model has an overall high precision, recall, and specificity scores of 96.71%, 96.83% and 99.78% respectively. The study employs the Grad-cam output to determine the heatmap of disease detection. The proposed model shows promising results in classifying multiple lesions on Chest X-ray images.

1. Introduction

Lung disease affects a lot of people in various ways and is one of the major causes of death around the world. It has been demonstrated that prior lung diseases, such as emphysema, chronic bronchitis, pulmonary fibrosis, and pneumonia, are associated with an increased risk of lung cancer, even in non-smokers [1,2]. The probability of contracting lung disease is quite high, particularly in developing and low-middle income regions, where millions of people endure poverty and air pollution. According to the WHO, approximately 4 million early deaths occur yearly due to domestic air pollution-related illnesses, such as asthma and pneumonia [3]. Accordingly, it is essential to employ an efficient diagnostic method to assist with the early detection of lung lesions [4,5].

As early lung disease identification and treatment are crucial to

optimal outcomes, early examination and diagnosis may reduce the life-threatening aspect of lung diseases and enrich the quality of life for individuals already afflicted. Chest X-rays (CXRs) are a common method used to identify lung diseases, evaluate the severity, and detect possible complications [10]. Several recent studies [11–14] demonstrate the effectiveness of lung segmentation approaches for automatic CXR image processing. Chest X-rays [15–17] may show several concurrent anomalies. Capturing anomalies from a complicated thoracic background with just the human eye is very time-consuming, and the procedure is sensitive, as well as susceptible to user bias. Manual labelling by radiologists, requires major healthcare resources. Therefore, computer technologies may be employed to analyze chest radiographs as effectively as radiologists, in order to enhance workflow prioritization and clinical decision support, in large-scale projects and worldwide

* Corresponding author.

** Corresponding author.

E-mail addresses: sami.azam@cdu.edu.au (S. Azam), asif.karim@cdu.edu.au (A. Karim).

<https://doi.org/10.1016/j.combiomed.2023.106646>

Received 10 June 2022; Received in revised form 30 January 2023; Accepted 6 February 2023

Available online 10 February 2023

0010-4825/© 2023 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

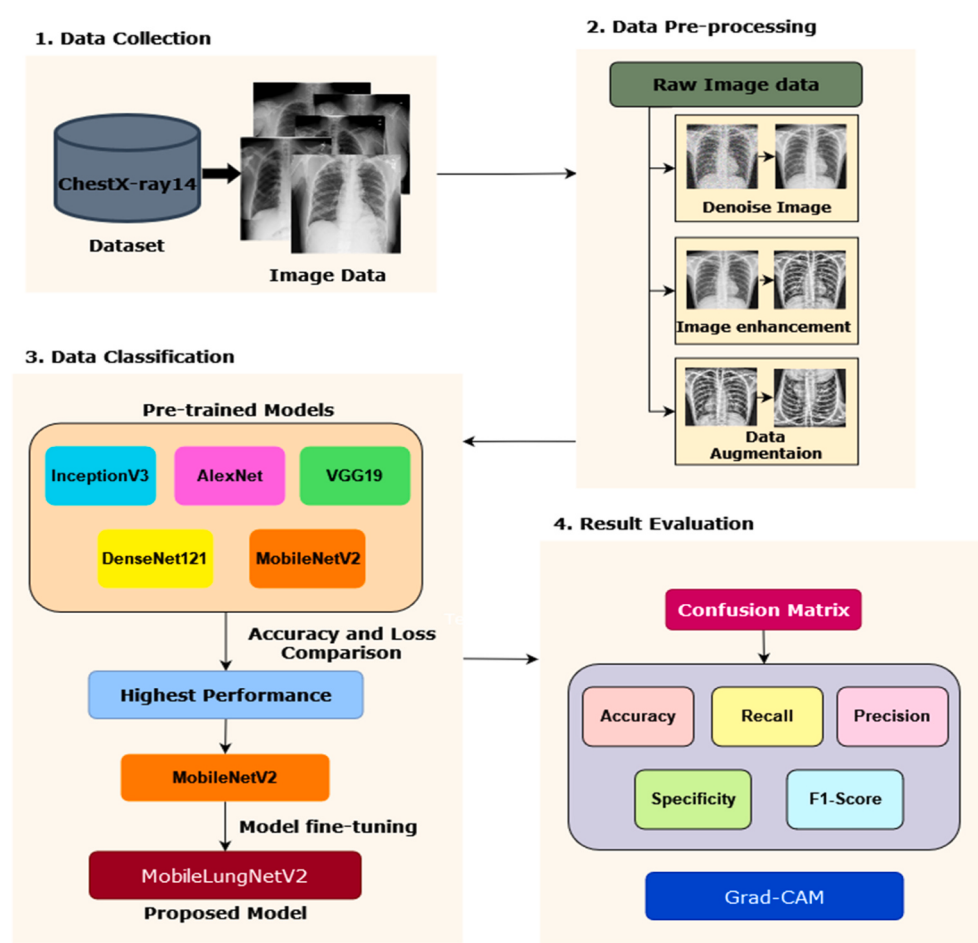


Fig. 1. Overview of the approach for multiclass lung lesion classification.

population health initiatives. Machine learning and deep learning can play a significant role in accurate clinical diagnosis [34–37].

Deep learning-based algorithms have achieved satisfactory performance in a number computer vision tasks [7–9], including image classification [18], medical diagnosis [19], scene identification [20], disease prediction [21], and healthcare analysis [22]. The rapid improvement of deep learning techniques has been facilitated by the creation of several annotated image datasets [23–26]. Characteristics indicated by these annotations have been crucial in overcoming obstacles in a variety of medical image analysis areas, including the identification of anatomical and pathological aspects in radiological scans. Deep learning approaches have been used for the detection and classification of a variety of diseases, such as lymph nodes, interstitial lung disease [27,28], cerebral microbleed detection [29], colon cancer classification [30], spinal radiological score prediction [31], automated pancreas segmentation [32], and pulmonary nodule detection [33].

Developing countries experience a shortage of skilled radiologists, particularly in rural regions. Additionally, detecting and classifying lung disease utilizing chest X-ray imaging is a difficult task for radiologists. Therefore, it is in the interest of researchers to create automated lung disease detection tools [78–80]. In these scenarios, a computer-aided diagnostic (CAD) system can be used to do large scale diagnosis of lung disease through examining CXR pictures. Significant improvements in computing power, as along with the availability of extremely large datasets labeled with chest X-rays, contributed to the accuracy of image classification. Several recent strategies have been presented to automatically diagnosis lung diseases in CXR images. In 2017, Wang et al. [81] presented the biggest publicly accessible chest X-ray dataset, titled "Chest X-ray 14," which included 14 of the most prevalent lung diseases.

Numerous studies [82–85] are conducted on this massive dataset. Wang et al. [86] suggested a unified weakly-supervised multi-label classification framework by taking into consideration multiple multi-label DCNN loss functions and distinct pooling algorithms. Since one chest radiograph may have many abnormal patterns, Yao et al. [87] created a method that further utilizes the statistical label correlations, resulting in enhanced performance. Likewise, Kumar et al. [88] used multi-label learning approaches and studied possible label dependencies. Rajpurkar et al. [89] constructed a deep learning model called CheXNet and made its optimization tractable by using batch normalization [90] and dense connections [91].

This research paper may help medical practitioners, as well as researchers, identify lung lesions using deep learning methodology. In this study, we implemented several pre-trained CNN models to classify lung lesions in Chest X-ray images. Of these models, we determine the model's highest performance for modification to achieve the proposed fine-tuned MobileLungNetV2 model with superior classification accuracy. To improve the performance of computer-aided diagnostic systems (CADs), we classify lung lesions in chest X-ray images using the proposed model. The primary contributions of this study are as follows.

- The proposed fine-tuned MobileLungNetV2 model is based on MobileNetV2, but with certain modifications made to achieve higher disease classification accuracy compared to pre-trained models.
- The proposed model has been developed to better extract image features and identify lung abnormalities.
- The proposed model outperformed both prior researches done on classification of lungs disease using deep learning models and the

implemented pre-trained models of the current study as demonstrated in Table 6.

To achieve and highlight the obtained result, we followed the processes summarized as follows.

- The raw chest X-ray images utilized in the investigation are obtained from the dataset ChestX-ray14. The images are pre-processed to improve the consistency of the data and reduce noise.
- A Gaussian Filter is employed to denoise the noisy images.
- CLAHE is applied to the dataset to achieve clear contrast images.
- Augmentation is performed to increase the number of data in lacking classes of the dataset.
- Five transfer-learning models: InceptionV3, AlexNet, DenseNet121, VGG19, and MobileNetV2, are assessed to identify the most accurate model.
- MobileNetV2 is modified in order to develop the proposed MobileNetV2 model.
- To evaluate performance, a confusion matrix is generated to compute accuracy, recall, precision, specificity, false positive rate, false negative rate as well as f1-score.
- Grad-CAM is produced for visual depiction of the classification outputs.
- To further evaluate the results, a performance comparison with prior research on the same dataset is provided.

The remainder of the paper is arranged as follows: The literature study is described in Section 2 to use deep learning to diagnose lung diseases. Section 3 gives the dataset description of the study. The proposed methodology and result analysis are presented in Sections 4 and 5, respectively. Consequently, Sections 6 and 7 provide the discussion and conclusion of the study, respectively. An overview of the study is shown in Fig. 1.

2. Literature review

Machine learning and deep learning techniques are widely used in CADs. The diagnosis of lung disease has been the focus of numerous studies in the past. It is clear from analyzing those research articles that the majority of researchers have utilized machine learning and deep learning algorithms on X-ray images to predict the disease with high accuracy and efficiency so that they can create an appropriate diagnostic system. For instance, a two-stream collaborative network with lung segmentation, (TSCN) has been used by Chen et al. to categorize multi-label CXR images with 0.823 mean Area Under Curve (AUC) value [40]. The authors used U-Net to train an efficient lung segmentation tool. They then aggregated the contextual information using a feature fusion approach. Another study [41] proposed DualCheXNet, a unique twofold asymmetric features extraction network for multi-class pulmonary disease classification in CXRs. The proposed technique supports two distinct feature fusion processes, namely feature-level fusion (FLF) and decision-level fusion (DLF) which correspond to the complimentary feature learning of DualCheXNet. The study achieves 0.823 AUC values as well.

Pan et al. [42] aimed to analyze and evaluate the usefulness of optimal CNN for abnormality diagnosis in chest radiographs. DenseNet & MobileNetV2 CNN algorithm were applied for classifying chest X-rays as normal or abnormal, as well as for predicting the occurrence of 14 distinct pathological abnormalities. MobileNetV2 outperformed DenseNet in the study with 0.900 and 0.893 AUC values, respectively. Other authors [43] attempted to combine the effectiveness of CNN for extracting visual features from the dataset with the effectiveness of task transformation approaches for multiple label classification, utilizing problem transformation approaches such as Binary Relevance, Label Powersets, and Classifier Chains that gains 0.804, 0.811 and 0.794 AUC values, respectively. Another approach [44] is of the integration of

multiple features. Two distinct techniques were employed: a localization approach that focuses on pathological areas utilizing pre-trained DenseNet-121, and a classification strategy that integrates four types of features generated with Generalized Search Tree (GIST), Scale-Invariant Feature Transform (SIFT), Histograms of Oriented Gradient (HOG) and Local Binary Pattern (LBP), as well as convolutional network features. The study classifies the multiple diseases with average AUC value of 0.8097.

Gong et al. [45] used a deformable Gabor convolution (DGConv) which improves the interpretability of deep networks and allows complex spatial variations. To increase robustness for complex objects, the features are trained at deformable sampling points using adaptive Gabor convolutions. The DGConv layer replaces conventional convolutional layers and is readily taught with end-to-end and gains 0.8501 AUC value. Wang et al. [46] present an adaptive sampling strategy that continuously analyzes the model's performance while training and automatically increases the weight of classes with low performance. Data augmentation is done by arbitrarily repeating its data samples and the resulting dataset is shuffled and divided into batches of equal size which are input into the model. The model was tuned using a stochastic gradient descent (SGD) technique. The model's performance has shown an average AUC value of 0.082. ResNet34 and DenseNet121 were two of the network topologies evaluated in Ref. [47]. The study evaluated image dimensions varying from 32×32 to 600×600 pixels. 80% of the samples were utilized for training and 20% for validation. The study shows the AUC ratio of $86.7\% \pm 1.2$ and $80.7\% \pm 1.5$ for thoracic mass and pulmonary nodule detection respectively.

Baltruschat et al. [48] investigated the ResNet-50 architecture in order to get a better understanding of the various techniques and their applicability to chest X-ray categorization. Through a systematic evaluation, they obtained an AUC value of around 0.800 utilizing 5-fold resampling and a multi-label loss function. Ho and Gwak [49] used a multi-task deep learning model to support visualizations used in saliency maps of the disease areas as well as for multiclass classification. A framework for self-training knowledge distillation (KD) was demonstrated to outperform both the well-established baseline training technique and conventional KD.

Albahi et al. [50] proposed a strategy for supplementing three deep CNN models with synthetic data to identify fourteen lung-related pathologies. The algorithms utilized were DenseNet121, ResNet152V2, and InceptionResNetV2. The proposed models were trained and tested for multiple class classification in order to detect anomalies in chest X-ray images. The Rozenber et al. study [51] was based on a unique loss function that is a continuous relaxation of a discrete conception of weak supervised learning. Additionally, the paper proposes a neural network design that compensates for both the patch dependency and shift invariance by applying Conditional Random Field layers and anti-aliasing filters. Bharati et al. [52] present a novel hybrid framework for deep learning termed VGG Data STN associated with CNN, VDSNet. This system combines CNN with VGG, and applied data augmentation, with a spatial transformer network (STN). In addition, Vanilla Gray, Hybrid CNN + VGG, Vanilla RGB and a reconfigured Capsule Network were used. The validation accuracy of the proposed VDSNet model was 73%.

The literature studied for the research successfully classify multiple lung lesions from the chestX-ray14 dataset. However, the sample used to train the dataset not large enough to create a strong model. Furthermore, the dataset is significantly unbalanced. As a result, the model is trained excessively for one class and insufficiently for another. Consequently, despite the fact that the models could detect multiple lung lesions successfully, their performance was ultimately inadequate when applied more broadly. In the proposed study, the models are trained and tested using 15 classes from the chestX-ray14 dataset. Data augmentation techniques are employed on the dataset to increase the number of data in underrepresented classes and reduced data from the over-represented classes. Thus, the implemented models were trained to

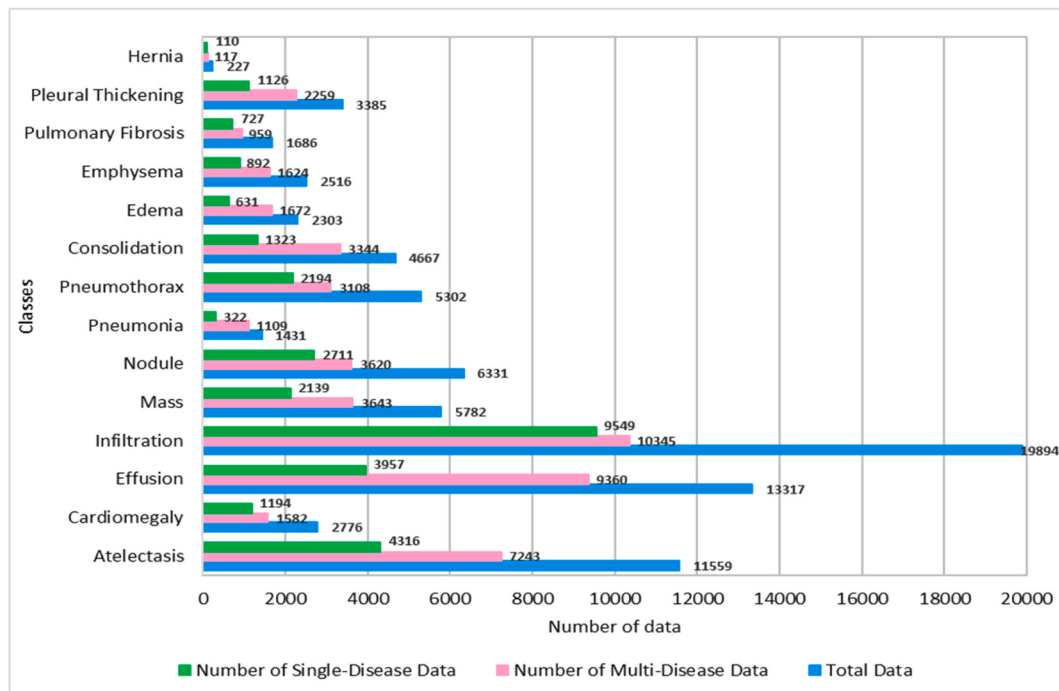


Fig. 2. Data distribution of all classes in the ChestX-ray14 dataset.

Table 1

Overview of research employing deep learning techniques and their performance in classifying lung lesions.

Ref	Dataset	Class	Imaging Type	Model	Performance
Chen et al. [40]	JSRT, Montgomery County, ChestX-ray14	2	CXR	TSCN	AUC = 0.823
Chen et al. [41]	ChestX-ray14	14	CXR	DualCheXNet	AUC = 0.823
Pan et al. [42]	ChestX-ray14, RIH-CXR	14	CXR	DenseNet	AUC = 0.893
				MobileNetV2	AUC = 0.900
Allaouzi et al. [43]	ChestX-ray14, CheXpert	14	CXR	Binary Relevance	AUC = 0.804
				Label Powersets	AUC = 0.811
				Classifier Chains	AUC = 0.794
Ho et al. [44]	ChestX-ray14	14	CXR	DenseNet-121	AUC = 0.8097
Gong et al. [45]	ChestX-ray14	14	CXR	DGFN	AUC = 0.8501
Wang et al. [46]	ChestX-ray14	14	CXR	Modified DenseNet121	AUC = 0.82
Sabottke et al. [47]	ChestX-ray14	14	CXR	ResNet34	AUC = 86.7% ± 1.2
				DenseNet121	AUC = 80.7% ± 1.5
Baltruschat et al. [48]	ChestX-ray14	14	CXR	ResNet-50	AUC = 0.800
Ho and Gwak [49]	ChestX-ray14	14	CXR	Self-train KD	AUC = 0.826
Albahli et al. [50]	ChestX-ray14	14	CXR	InceptionResNetV2	AUC = 0.80
Rozenber et al. [51]	ChestX-ray14, Kaggle RSNA Pneumonia Detection Challenge Dataset	14	CXR	Modified ResNet50	Accuracy = 0.58–0.95
Bharati et al. [52]	ChestX-ray14	15	CXR	VDSNet	Accuracy = 73%

achieve improved performance and increased robustness over existing models.

3. Dataset

The ChestX-ray14 dataset is one of the largest collections of anterior-view pulmonary X-ray scan data. For this research a total of 112,120 bilateral thoracic X-ray data were obtained from the dataset [53]. Wang et al. [38] released the first standard for identification and localization using a weakly supervised CNN architecture the ChestX-ray-8 dataset. Using Wang's published large data set ChestX-Ray14, Rajpurkar et al. [39] constructed the 121-layer CheXNet model and identified 14 distinct chest disorders. The ChestX-ray14 dataset contains data of 14 disease classes and a normal class from 30805 unique patients. These disease classes are Emphysema, Infiltration, Mass, Pleural Thickening, Pneumonia, Pneumothorax, Atelectasis, Edema, Effusion, Hernia, Cardiomegaly, Pulmonary Fibrosis, Nodule, and Consolidation. The normal class is labeled as "No finding". The datasets were labeled using Natural

Table 2

The properties of the raw dataset.

Properties	Data
Total number of classes	15
Total Data	112120
Normal Data	60412
Total Single-Disease Data	31191
Total Multi-Disease Data	49991
Gender (Male: Female)	56.49%: 43.51%

Language Processing on the radiological reports of the associated data. From the data labels, it is observed that the dataset contains images with multiple diseases (multi-disease) in a single X-ray image as well as images with a single-disease. Fig. 2 depicts the dispersion of data through all classes in the employed dataset. The dataset images are in a Portable Network Graph (PNG) format with the size of 1024 × 1024 pixels. The age variation of the patients is from 1 year to 95 years with both male

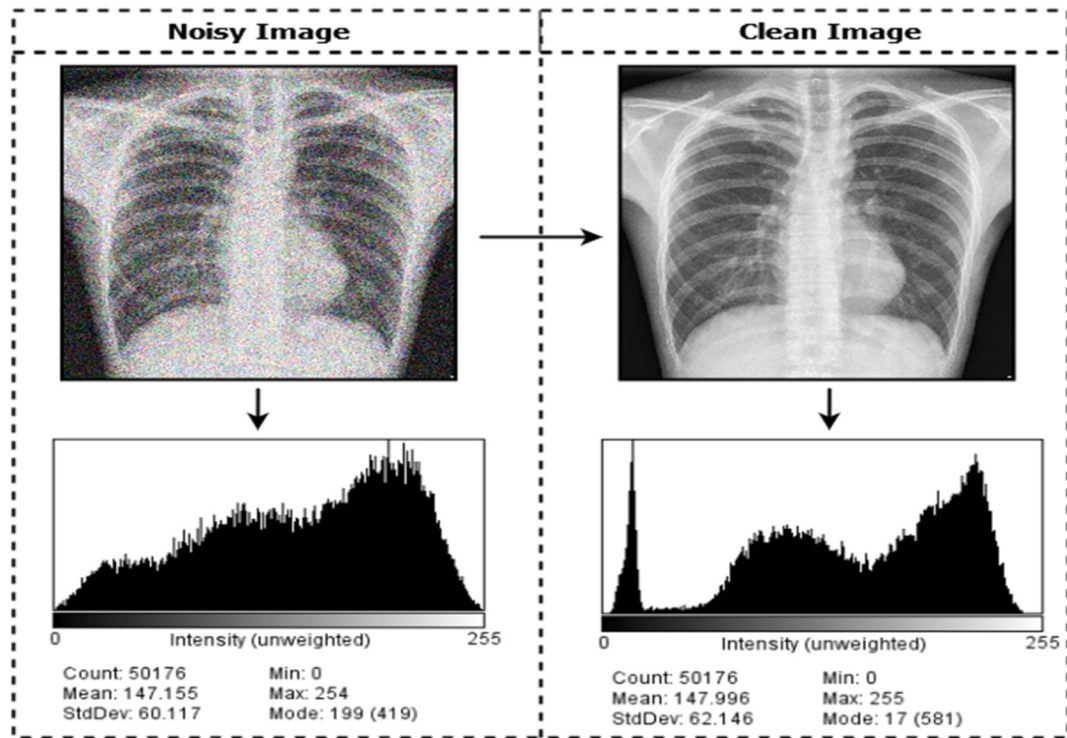


Fig. 3. The result of applying the Gaussian Filter to the images.

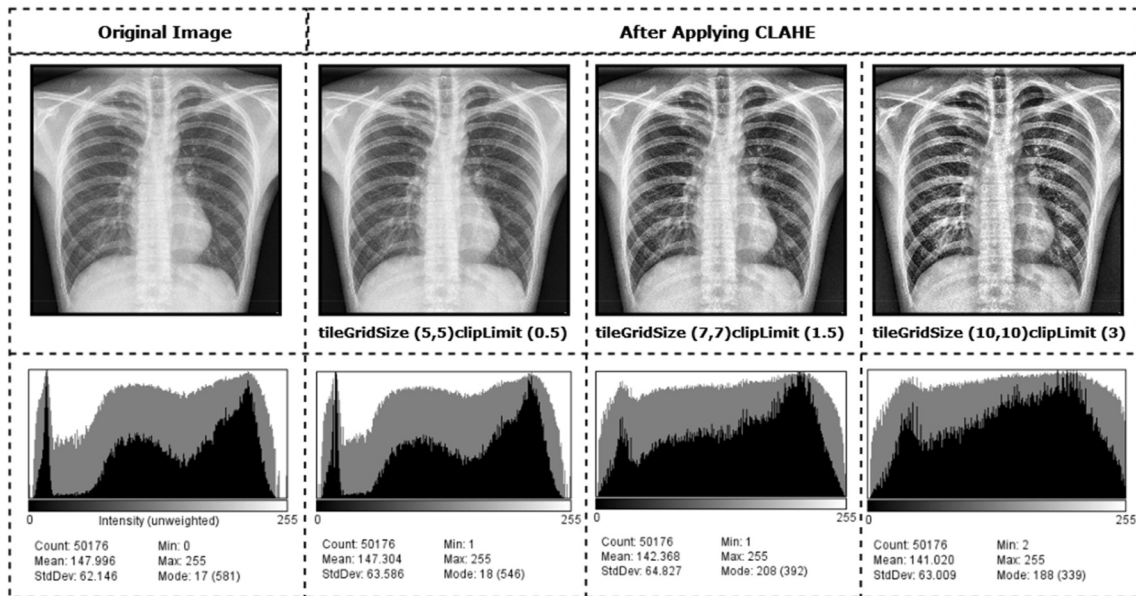


Fig. 4. The outcome of implementing the CLAHE enhancement method in fifteen categories: Atelectasis, Cardiomegaly, Effusion, Infiltration, Mass, Nodule, Pneumonia, Pneumothorax, Consolidation, Edema, Emphysema, Pulmonary Fibrosis, Pleural Thickening, Hernia, and No Finding, with different tileGridSize and clip limit combinations.

and female patients. Table 2 presents the detailed properties of the dataset.

3.1. Dataset pre-processing

Because the CNN technique employed for classification requires clean, improved, and balanced image data [53], image-preparation and image-balancing procedures are used to provide the model with a high-quality image. This section discusses the various image-processing

techniques. The CLAHE approach separates the images into contextual portions called tiles, calculates a histogram for each, and then approximates the output to a specified histogram distribution parameter.

3.1.1. Denoising the image (Gaussian Filter)

Gaussian blur, often referred to as Gaussian smoothing, is the outcome of filtering an image using a Gaussian function [61]. It is a frequent function in graphics software, which is often used to reduce visual noise. The visual result of this blurring approach is a smooth blur

Table 3
Data distribution for each class after augmentation.

Index	Class	data	Index	Class	data
1	Atelectasis	4316	9	Consolidation	3969
2	Cardiomegaly	2388	10	Edema	3786
3	Effusion	3957	11	Emphysema	3568
4	Infiltration	5000	12	Pulmonary Fibrosis	3635
5	Mass	2139	13	Pleural Thickening	2252
6	Nodule	2711	14	Hernia	1980
7	Pneumonia	2898	15	No Finding	5000
8	Pneumothorax	4388		Total Data	51987

resembling viewing the images through a transparent screen, which is distinguishable from the bokeh effect generated by an out-of-focus lens or the shadow of an object under normal light [62]. In computer vision methods, Gaussian smoothing is frequently employed as a pre-processing step to improve visual structures at various scales [63]. Fig. 3 displays the results of the deployment of the Gaussian Filter.

3.1.2. Image enhancement (CLAHE)

Contrast Limited Adaptive Histogram Equalization (CLAHE) is used to increase the contrast of a picture. CLAHE is a more advanced variant of Adaptive Histogram Equalization (AHE) [58]. CLAHE has been found to improve the contrast of low-contrast images [59]. CLAHE improves both the local contrast of medical imaging and its usability [60]. The CLAHE technique focuses on enhancing local contrast to overcome the constraints of global methods. The tile size and clip limit are critical hyper-parameters for this method. An incorrect choosing of hyper-parameters could have a big influence on the image quality. As we can observed from Fig. 4, several combinations such as tileGridSize (5, 5) with the clip limit (0.5), tileGridSize (7, 7) with the clip limit (1.5) and tileGridSize (10, 10) and the clip limit (3) parameters are investigated, and the optimal ones (tileGridSize (10, 10) and the clip limit (3)) are chosen. It is also observable that the features of the selected image are more prominent. The histogram indicates that the contrast of a CLAHE image is much greater than that of the initial source image. Fig. 4

depicts the output findings after executing the CLAHE algorithm.

3.1.3. Image augmentation

A deep learning model needs a large number of inputs to function with optimal efficiency. In this work, several data augmentation approaches are used to boost the enhanced data. By adding more distinct samples to the training datasets, data augmentation may improve the performance and outcomes of machine learning algorithms. If the dataset employed to train the model is sufficiently broad and varied, the approach is more effective and precise. Through the use of image enhancement methods, accuracy of the results is increased. Moreover, data augmentation methods are an effective method for diversifying datasets. Generally, to provide large-capacity learners with the more relevant training material, data augmentation techniques have been used to enhance the size of training sets. Nevertheless, a new trend is developing in the area of deep learning research in which samples are reinforced utilizing the test data augmentation technique [64–67]. The addition of test data may increase the stability of trained models [68–70]. Test data augmentation can therefore be used to improve the prediction performance of deep neural networks and open up fascinating new opportunities for medical image analysis [71–73]. Mirroring, rotating, zooming, flipping, and cropping are the most often used ways of augmenting data.

In this study, the dataset is adjusted using oversampling and undersampling techniques. First, a random undersampling approach is used to reduce the class (Infiltration, and No Finding) with excess data. This technique deletes data at random from the majority of classes, reducing the quantity of data per class to 5000. Then, the oversampling (data augmentation) approaches are used to increase the class with inadequate data (Emphysema, Infiltration, Mass, Pleural Thickening, Pneumonia, Pneumothorax, Atelectasis, Edema, Effusion, Hernia, Cardiomegaly, Pulmonary Fibrosis, Nodule, and Consolidation). Several augmentation procedures are used in this study: Rotate 90° right, Rotate 90° left, Rotate 45° Horizontal, Vertical flip, Rotate 45° Vertical, Translate (x, y (28.0, 13.0)) and Horizontal flip on image data that has been preprocessed. [Table 3](#) displays the result of the data augmentation.

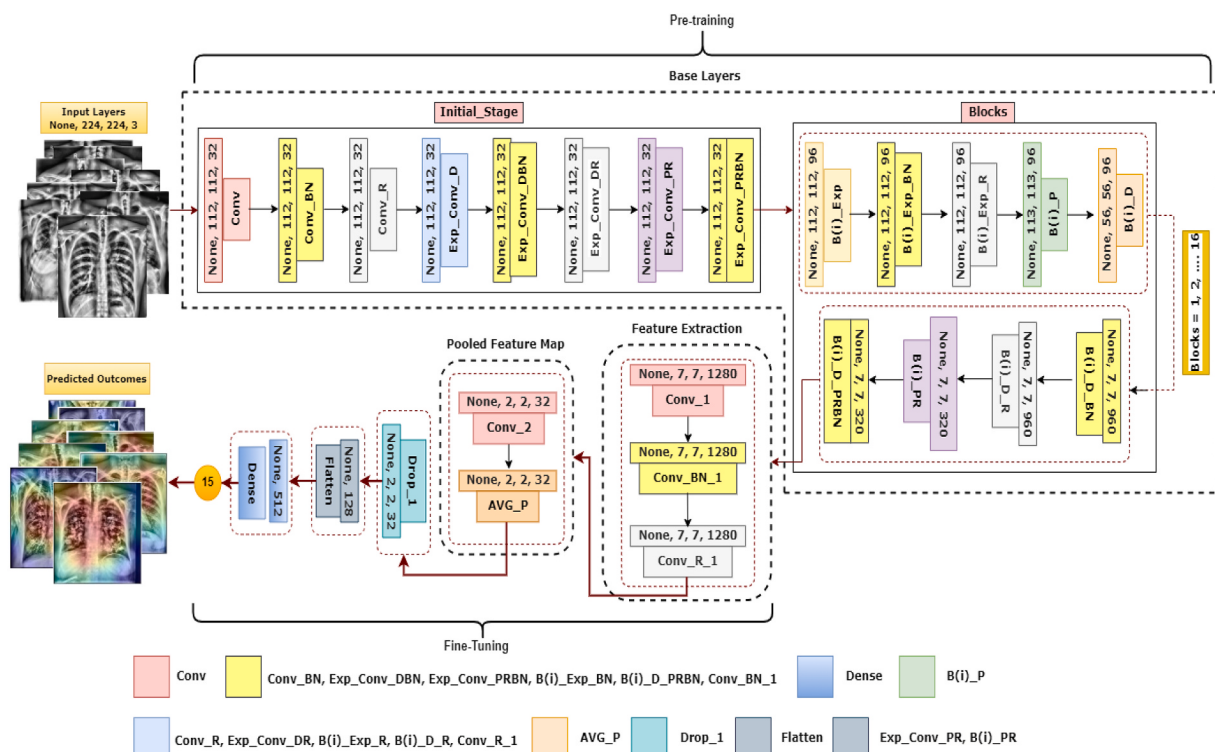


Fig. 5. The architecture of MobileLungNetV2, a fine-tuned MobileNetV2 architecture.

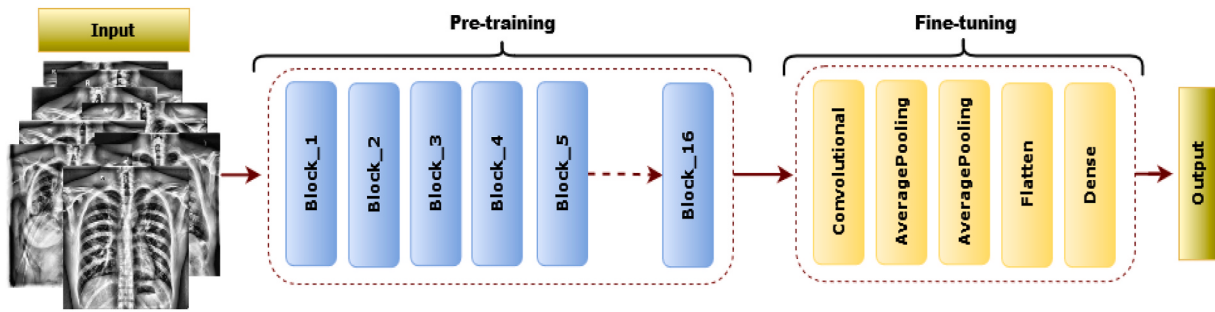


Fig. 6. The sketch map of fine-tuning strategy, transform a pre-trained MobileNetV2 to a fine-tuned MobileLungNetV2.

4. Proposed model

The principal purpose of this study is to employ transfer learning methods to obtain an appropriate classification accuracy on the NIH chest X-ray dataset. Five pre-trained algorithms were analyzed to find the optimum effective deep learning approach for the lung lesion classification task. These models are; InceptionV3 [75], AlexNet [6], DenseNet121 [76], VGG19 [77], and MobileNetV2 [55]. Among these pre-trained models MobileNetV2 showed the highest accuracy. To further improve to the classification results, MobileNetV2 was modified. The ablation study was performed to determine the different hyper-parameters. A custom fine-tuning transfer learning approach, named MobileLungNetV2, is designed and applied by adding multiple layers to the MobileNetV2 network to obtain the highest accuracy over the existing pre-trained models. This study aimed to develop a CNN-based method for lung lesion image classification. The presented system was developed in Python utilizing Keras framework [54], which is a TensorFlow-based platform. All tests were conducted using AMD Ryzen 7 (3900) CPU running at 3.90 GHz with 8 cores, 16 threads, and 64 GB of RAM.

The dataset includes 51,987 images categorized as; Emphysema, Infiltration, Mass, Pleural Thickening, Pneumonia, Pneumothorax, Atelectasis, Edema, Effusion, Hernia, Cardiomegaly, Pulmonary Fibrosis, Nodule, and Consolidation and No Finding. The dataset is split into 60% for training, 20% for validation, and 20% for testing.

4.1. MobileLungNetV2

The fine-tuned MobileNetV2 frameworks known as MobileLungNetV2 outperform five pre-trained model architectures in classification accuracy, shown in section 5. Consequently, the MobileLungNetV2 architecture is introduced and tested using the National Institutes of Health's chest X-ray dataset built on the MobileNetV2 architecture. Additionally, hyper-parameter tuning was conducted to increase the architecture durability in terms of lung lesion identification. Fig. 5 illustrates the model structure.

The pre-trained MobileNetV2 [55] introduces a module that includes an inverting residual structure. MobileNetV2 is designed from the bottom up using fully convolutional layers made of filters and residual bottleneck layers. MobileNetV2's structure begins with fully convolutional layers made up of 32 filters and 19 residual bottlenecks. It's split into two blocks, each one with 3 layers. These blocks begin and end with a 1×1 convolution layer comprising 32 filters. However, the 2nd block is a fully linked layer with a depth of one. The ReLU is used at several levels of the architecture. The distinction between the two blocks is in stride length, with block one having a stride length of 1 and block two having a stride length of 2.

Fig. 5 shows MobileLungNetV2 is constructed by linking layers following the 16 blocks of pre-trained MobileNetV2. The layers consist of convolutional layers, an AveragePooling2D layer, a dropout layer, and a flatten layer that is linked to a dense layer. The network's default input layer requires an image with a size of 224×224 , which is a gray scale image. The 16 blocks consist of 2 types of blocks in the model. One has a residual block stride of 1. The second is a block with a stride of 2 for shrinking. Each block has three layers. The first layer consists of 1×1 convolution with ReLU6. The second layer is the depth-wise convolution layer. The third layer is also another 1×1 convolution without any linearity. Similarly, up to 16 blocks are constructed with the same alignment.

After block sixteen, adding a convolution layer, an AveragePooling2D layer, a dropout layer, a flatten layer and a dense layer. The convolutional layer has the kernel size of 3×3 , the same padding and a ReLU activation function with 32 channels followed by a AveragePooling2D layer with a pool size of 2×2 . A dropout layer has been added with 32 channels to prevent the overfitting. A flatten layer has also been added, which is connected to a dense layer that generates output for the 15 classes. The RMSprop optimizer and SoftMax activation were employed in the final layer. Throughout the process, a frequency of learning rate of 0.001 is applied. Finally, we evaluated the f1-score, recall, precision, specificity and accuracy. In Fig. 6, the strategy of fine-tuning is depicted.

Table 4
Altering the different layers to assess the ablation experiment.

Case Study	Layer Identity	V_Acc (%)	V_Loss (%)	Ts_Acc (%)	Ts_Loss (%)	Outcome
1	AveragePooling2D	94.76	0.192	96.97	0.10	Identical Performance
	GlobalAveragePooling2D	90.73	0.32	91.65	0.29	Accuracy Dropped
	GlobalMaxPooling2D	89.84	0.38	90.76	0.30	Accuracy Dropped
	MaxPooling2D	91.25	0.29	91.98	0.28	Accuracy Dropped
2	Dropout	94.76	0.192	96.97	0.10	Identical Performance
	Flatten	90.81	0.36	90.90	0.35	Accuracy Dropped
	MaxPooling2D	88.72	0.49	89.10	0.43	Accuracy Dropped
	GlobalMaxPooling2D	90.49	0.36	90.98	0.30	Accuracy Dropped
3	Flatten	94.76	0.192	96.97	0.10	Identical Performance
	GlobalMaxPooling2D	88.36	0.54	89.16	0.44	Accuracy Dropped
	GlobalAveragePooling2D	87.73	0.69	88.50	0.51	Accuracy Dropped
	MaxPooling2D	90.18	0.30	90.98	0.28	Accuracy Dropped

Table 5

Shifting Optimizer and Learning rate for ablation study analysis.

Case Study	Optimizer	Learning Rate	V_Acc (%)	V_Loss (%)	Ts_Acc (%)	Ts_Loss (%)	Outcome
4	Adagrad	0.00000001	94.76	0.192	96.97	0.10	Identical Performance
		0.000001	90.55	0.38	90.60	0.35	Accuracy Dropped
		0.0000001	91.63	0.33	92.12	0.29	Accuracy Dropped
	RMSprop	0.00001	89.18	0.47	89.98	0.42	Accuracy Dropped
		0.00000001	92.36	0.28	92.96	0.25	Accuracy Dropped
		0.0000001	89.71	0.44	90.13	0.30	Accuracy Dropped
	SGD	0.00000001	90.86	0.35	91.02	0.33	Accuracy Dropped
		0.00001	88.62	0.56	89.11	0.49	Accuracy Dropped
		0.00000001	88.16	0.69	88.96	0.64	Accuracy Dropped
	Nadam	0.0000001	85.76	0.80	86.15	0.78	Accuracy Dropped
		0.00000001	86.94	0.75	87.23	0.71	Accuracy Dropped
		0.00001	80.23	0.83	80.97	0.81	Accuracy Dropped
		0.00000001	91.56	0.31	92.08	0.27	Accuracy Dropped
		0.0000001	87.32	0.69	88.77	0.67	Accuracy Dropped
		0.00000001	90.44	0.35	91.16	0.30	Accuracy Dropped
		0.00001	85.61	0.79	86.24	0.74	Accuracy Dropped

4.2. Ablation study of the MobileLungNetV2

As part of ablation research, four experiments were performed by changing different sections of the suggested MobileLungNetV2 architecture based on the fine-tuned MobileNetV2 framework. It is feasible to create a more durable architecture with improved classification accuracy by altering multiple aspects of the architecture. Accordingly, the AveragePooling2D layer, Dropout layer, Flatten layer, Loss function, Learning rate and Optimizer were all subjected to an ablation analysis.

In this analysis, the validation loss is denoted by 'V_Loss,' the validation accuracy by 'V_Acc,' the test loss by 'Ts_Loss,' and the test accuracy by 'Ts_Acc.' Table 4 illustrate the case study 1, 2, 3 respectively. Case study 1 shows that the AveragePooling2D layer archives the highest accuracy for the MobileLungNetV2 model with the V_Acc 94.76% and Ts_Acc 96.97%. Furthermore, the accuracy of the GlobalAveragePooling2D, GlobalMaxPooling2D, and MaxPooling2D layers drops significantly, with respective V_Acc values of 90.73%, 89.84%, and 91.25%, and Ts_Acc values of 91.65%, 90.76%, and 91.98%.

In case study 2, the Dropout layer has the best accuracy with V_Acc at 94.76% and Ts_Acc at 96.97%, for the MobileLungNetV2 model. Additionally, the accuracy of the Flatten, MaxPooling2D,

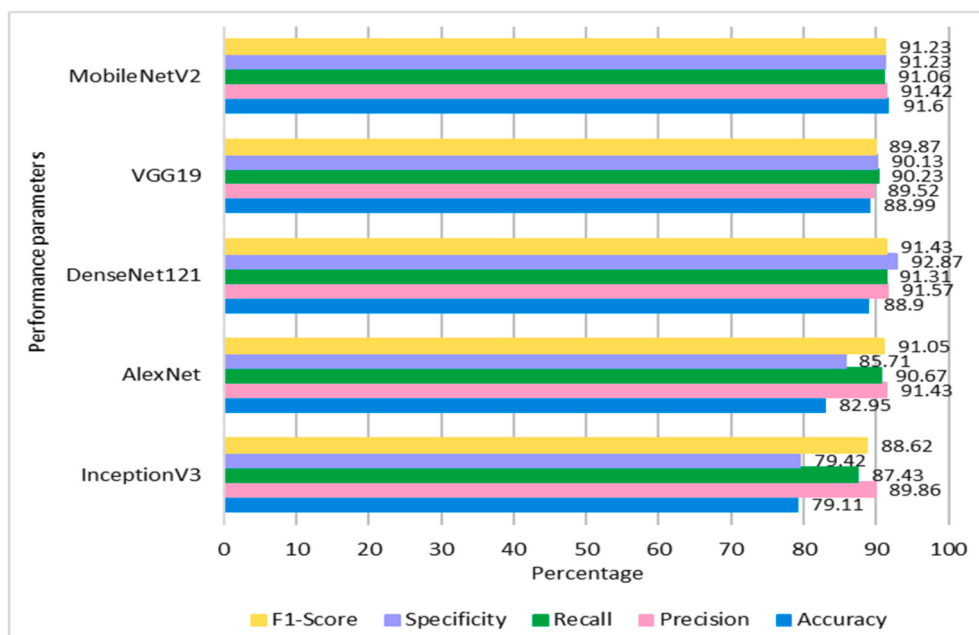
GlobalMaxPooling2D, and GlobalAveragePooling2D layers decreases significantly, with respective V_Acc values of 90.81%, 88.72%, 90.49%, and 89.65%, and Ts_Acc values of 90.90%, 89.10%, 90.98%, and 90.71%.

In case study 3, the Flatten layer has the highest accuracy for the MobileLungNetV2 model, with V_Acc at 94.76% and Ts_Acc at 96.97%. Additionally, the accuracy of the GlobalMaxPooling2D, GlobalAveragePooling2D, and MaxPooling2D layers decreases dramatically, with V_Acc values of 88.36%, 87.73%, and 90.18%, and Ts_Acc values of 89.16%, 88.50%, and 90.98%, respectively.

In case study 4, Table 5 depicted that the 'Adagrad' optimizer with a learning rate of 0.00000001 improved model performance with Ts_Acc values of 94.76% and Val_Acc values 96.97%. With such a learning rate of 0.00000001, the 'Adagrad' optimizer still had the lowest Ts_Loss of 0.10. Other optimizers, such as RMSprop, SGD, and Nadam, achieved our model's Ts_Acc values greater than 90%.

5. Results analysis

In the study, a total of 51987 chest x-ray images were used. The data used in each class is set out in Table 3. Each image is resized to 224 ×

**Fig. 7.** Performance measures of the transfer learning models.

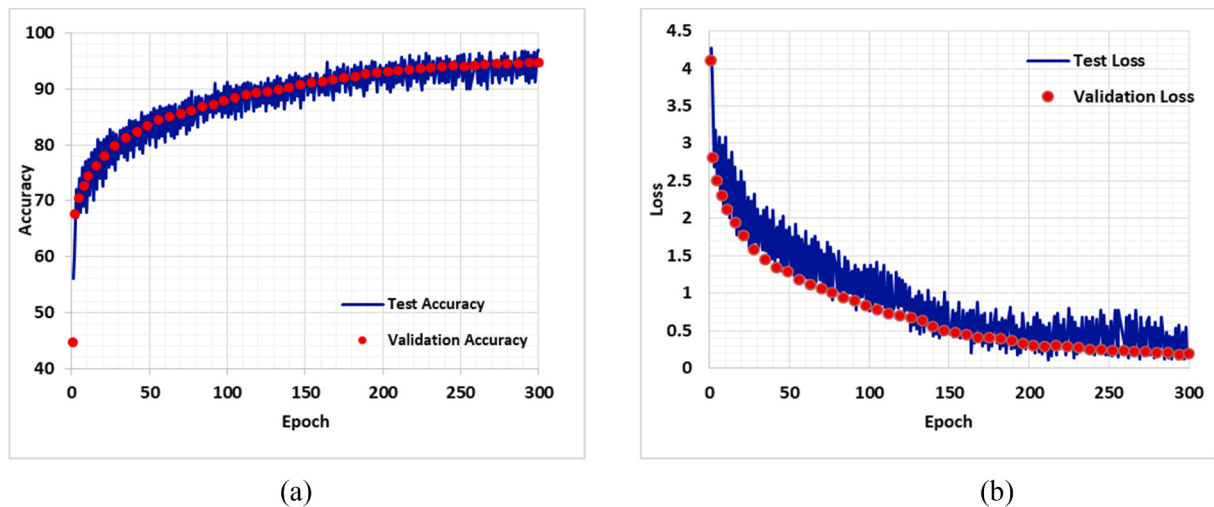


Fig. 8. The proposed model (MobileLungNetV2) (a) test and validation accuracy of 300 epochs, (b) test and validation loss of 300 epochs.

224 pixels. The dataset is divided into training, validation, and test sets at the ratio of 60:20:20. This study implemented five transfer learning models: MobileNetV2, VGG19, DenseNet121, AlexNet, and InceptionV3. The transfer learning models are assessed against the proposed MobileLungNetV2 model.

5.1. Evaluation matrix

To evaluate the transfer learning models, performance parameters such as Accuracy, Specificity, Recall, Precision, False Positive Rate (FPR), False Negative Rate (FNR) and F1-score are determined. The parameters are calculated using a confusion matrix generated for each individual model. Accuracy is calculated to determine the percentage of

correct predictions. Precision is calculated to determine the probability of positive classifications. Specificity determines the percentage of correctly predicted negative classifications from all performance parameters. Contrasting specificity, recall determines the percentage of correctly predicted positive classes. The F1-score is used to determine the balance between specificity and recall. The performance parameters are expressed in the following equations (1)–(7).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

Truth	Atelectasis -	0	0	8	1	0	5	0	5	0	6	0	7	0	4	827
	Cardiomegaly -	4	0	0	3	0	0	1	0	0	6	0	3	2	459	0
	Effusion -	0	2	4	0	5	0	0	0	7	4	0	0	768	0	1
	Infiltration -	0	1	0	3	0	7	2	0	0	5	0	978	0	0	4
	Mass -	0	0	2	0	0	2	0	0	0	8	416	0	0	0	0
	Nodule -	7	0	0	0	0	4	0	0	0	510	19	0	2	0	0
	Pneumonia -	0	0	0	2	0	0	1	1	565	0	0	5	0	6	0
	Pneumothorax -	0	2	0	0	3	0	1	861	6	0	2	0	3	0	0
	Consolidation -	4	0	3	0	0	7	764	0	0	0	3	5	0	0	8
	Edema -	0	2	0	5	0	737	4	0	0	0	2	0	3	4	0
	Emphysema -	1	1	3	0	696	0	0	3	2	0	0	5	0	2	0
	Pulmonary Fibrosis -	0	1	2	711	0	0	3	2	0	2	2	0	0	0	4
	Pleural Thickening -	3	0	431	5	0	3	2	0	4	0	0	1	1	0	0
	Hernia -	0	382	0	0	0	2	0	5	0	0	0	4	0	3	0
	No Finding -	977	0	3	4	0	7	1	0	0	5	0	0	3	0	0
		Predicted														
		No Finding -	Hernia -	Pleural Thickening -	Pulmonary Fibrosis -	Emphysema -	Edema -	Consolidation -	Pneumothorax -	Pneumonia -	Nodule -	Mass -	Infiltration -	Effusion -	Cardiomegaly -	Atelectasis -

Fig. 9. Confusion matrix for the MobileLungNetV2 model.

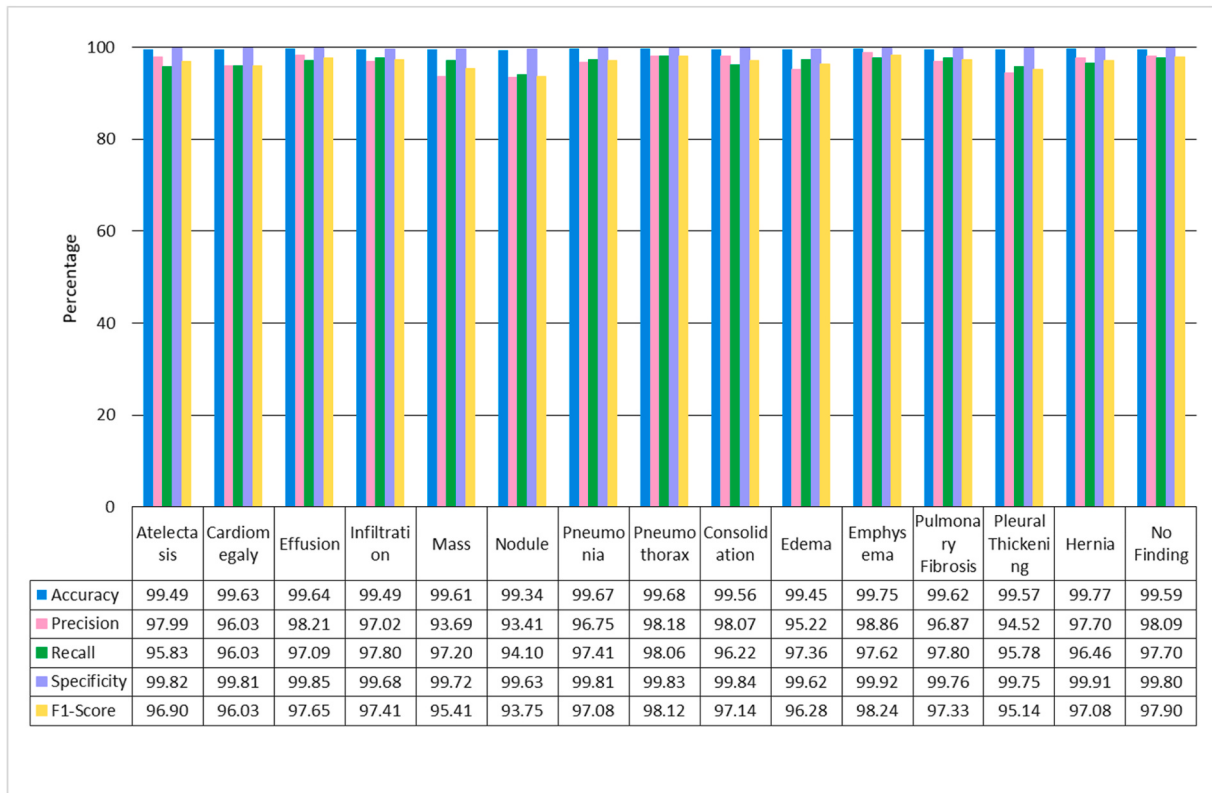


Fig. 10. Performance measures of each class.

$$\text{Sensitivity} / \text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (4)$$

$$F1 - \text{Score} = 2 \left(\frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \right) \quad (5)$$

$$FPR = \frac{FP}{TN + FP} \quad (6)$$

$$FNR = \frac{FN}{FN + TP} \quad (7)$$

Here, True Positive (TP) is identified as lung disease where the patients actually have lung disease. True Negative (TN) is predicted as the absence of lung disease where there is no actual existence of lung disease. False Positive (FP) predicts the presence of a lung disease that is not present. Similarly, False Negative (FN) predicts no lung disease where lung disease is present. These elements are then used to form the equations to generate the values of the performance parameters.

5.2. Performance of pre-trained transfer learning models

Initially, five transfer learning models were trained and tested. To determine the performance of the models, the performance parameters are computed for each model using the numbered equations (1)–(7). The performance measures outcomes are presented in Fig. 7. Overall, the model MobileNetV2 shows consistent performance with 91.6% accuracy. It achieved a precision of 91.42% and recall, specificity and F1-scores are 91.06%, 91.23% and 91.23% respectively. Although DenseNet121 has the high specificity score of 92.87%, it lacks accuracy with 88.9%. Similarly, the accuracy of the remaining models – VGG19, AlexNet, and InceptionV3 all have relatively low accuracy with 88.99%,

82.95% and 79.11% respectively.

5.3. Performance of proposed MobileLungNetV2

The transfer learning MobileNetV2 model shows the most promise compared to the other models used to predict lung lesions. However, the prediction accuracy of the MobileNetV2 model is not accurate enough because MobileNetV2 only has an accuracy of 91.6%, which means the incorrect prediction rate of the model is 8.4%. To diagnose any disease with this level of accuracy is unsatisfactory. We aim to build a model with the lowest possible chance of misdiagnosis. In order to achieve this, the accuracy should be as close to 100% as possible without being overfitted. Consequently, the paper proposes the fine-tuned MobileLungNetV2. The improved model is trained with 300 epochs. The training and validation accuracy for each epoch is recorded along with the loss value. From Fig. 8(a) and (b), it is observed that the accuracy of the model gradually increases with increasing epoch contrary to the loss value that gradually decreases. The final training accuracy of the model stands at 95.92% and the validation accuracy is at 94.76%. For the loss value, the lowest training loss value at 300th epoch is 0.15 and the lowest validation loss is 0.192.

The confusion matrix generated for the finetuned model is illustrated in Fig. 9. Here the matrix is constructed for 15 classes implementing the test dataset. The test set consists of 10397 inputs in total. Among them 10082 inputs were correctly classified by the model. The number of misclassified inputs is 310. In the matrix the vertically placed classes indicate the true or real value of a data and the horizontally placed classes indicates the prediction value of a data by the model.

The performance measures of the proposed model is calculated by applying the values of the confusion matrix to equations (1)–(7). Each of the measures is color coded and shown in Fig. 10. The accuracy and specificity of the model across all the classes are very high at over 99%. The other measures (precision, recall and f1-score), vary greatly between classes. The highest accuracy is 99.77% from class Hernia. Class

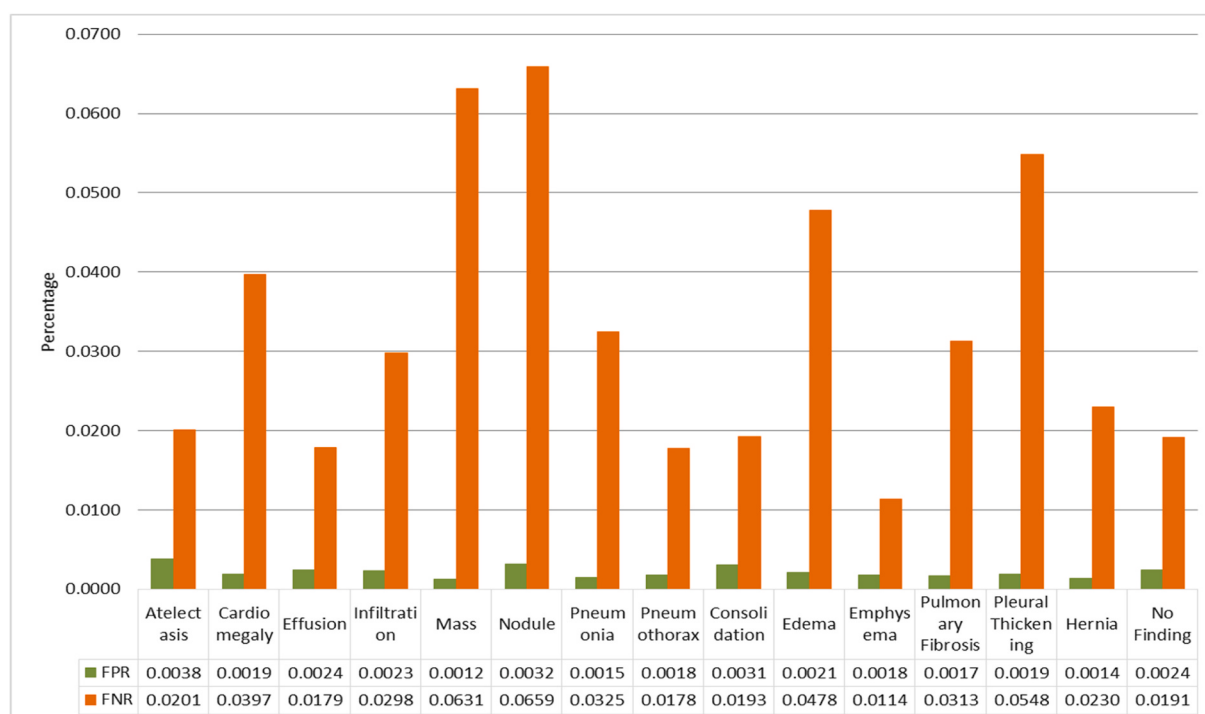


Fig. 11. False Positive Rate (FPR) and False Negative Rate (FNR) of each class.

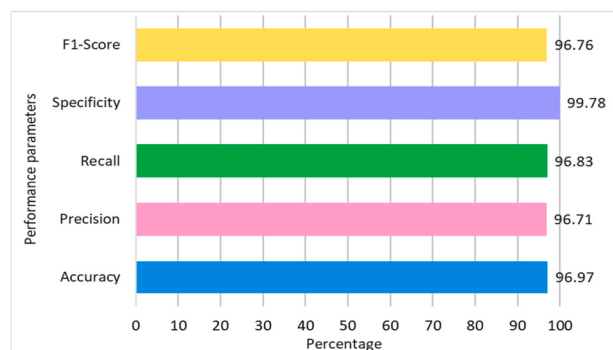


Fig. 12. Overall performance of the MobileLungNetV2 model.

Emphysema obtains the highest specificity, precision and f1-score with 99.92%, 98.86%, and 98.23% respectively. The highest recall recorder at 98.06% is from class Pneumothorax. The lowest accuracy, precision, recall and f1-score are obtained from class Nodule with 99.34%, 93.41%, 94.10% and 93.75% respectively. The lowest Specificity is obtained from class Edema with 99.62%.

To determine the false rates, the False Positive Rate (FPR) and False Negative Rate (FNR) of the proposed model are measured for each class. It is observed that the highest FPR is obtained from the Atelectasis class and the lowest rate is obtained from the Mass class with 0.0038% and 0.0012% respectively. Similarly, the highest FNR is 0.0659% for Nodule class and lowest is 0.0114% for Emphysema class. In Fig. 11, the values of FPR and FNR are recorded for each class. It shows that the FPR remains nearly same for all the classes, however the FNR varies drastically among all the classes of the dataset.

The overall performance of the model is presented in Fig. 12. The overall accuracy of the model is determined as 96.97%. Similarly the precision, recall and f1-score are 96.71%, 96.83% and 96.76% respectively. It is notable that the model has a high specificity score of 99.78%.

While numerous efforts have been made to enhance the applicability and expandability of deep learning, it is crucial to develop the

interpretability of deep convolutional neural models in medical imaging applications. Selvaraju et al. [74] illustrated the working of deep learning using a procedure dubbed Gradient Weighted Class Activation Mapping (Grad-CAM). Grad-CAM is effective with any heavily associated neural network and permits the algorithm's new information to be determined while executing prediction or classification operations. The input is a conventional X-ray image, and the suggested framework is applied as a detection strategy. Grad-CAM is applied to the last convolution layer immediately after the proposed model's label prediction. Fig. 13 demonstrates the visualization of heat maps on X-ray images using the proposed approach.

6. Discussion

The study aims to classify multiple lung lesions using convolutional neural networks. To achieve the best results, five transfer-learning models, and a fine-tuned model, MobileLungNetV2 based on the MobileNetV2 model are trained on the ChestX-ray14 dataset. Many previous research studies have made use of the ChestX-ray14 dataset as it is one of the biggest accessible datasets of this type. Therefore, previous studies using the ChestX-ray14 dataset have been analyzed in order to compare the accuracy of findings and techniques, as shown in Table 1. In the current study, the MobileNetV2 was chosen as the model to be adapted based on its higher accuracy among the five pre-trained models. To determine the performance of the models, several performance measures are computed and the performance of the models are determined class-wise for a detailed examination. In Table 6, a performance comparison of existing studies is considered with the proposed study on the same dataset. The AUC value is used as the main performance measure for the comparison. Here, it is observed that among the existing studies, the highest average AUC value is 0.850. The fine-tuned MobileLungNetV2 shows a higher AUC value of 0.923 showing the higher efficiency of the proposed model in determining the classification of the lung lesions.

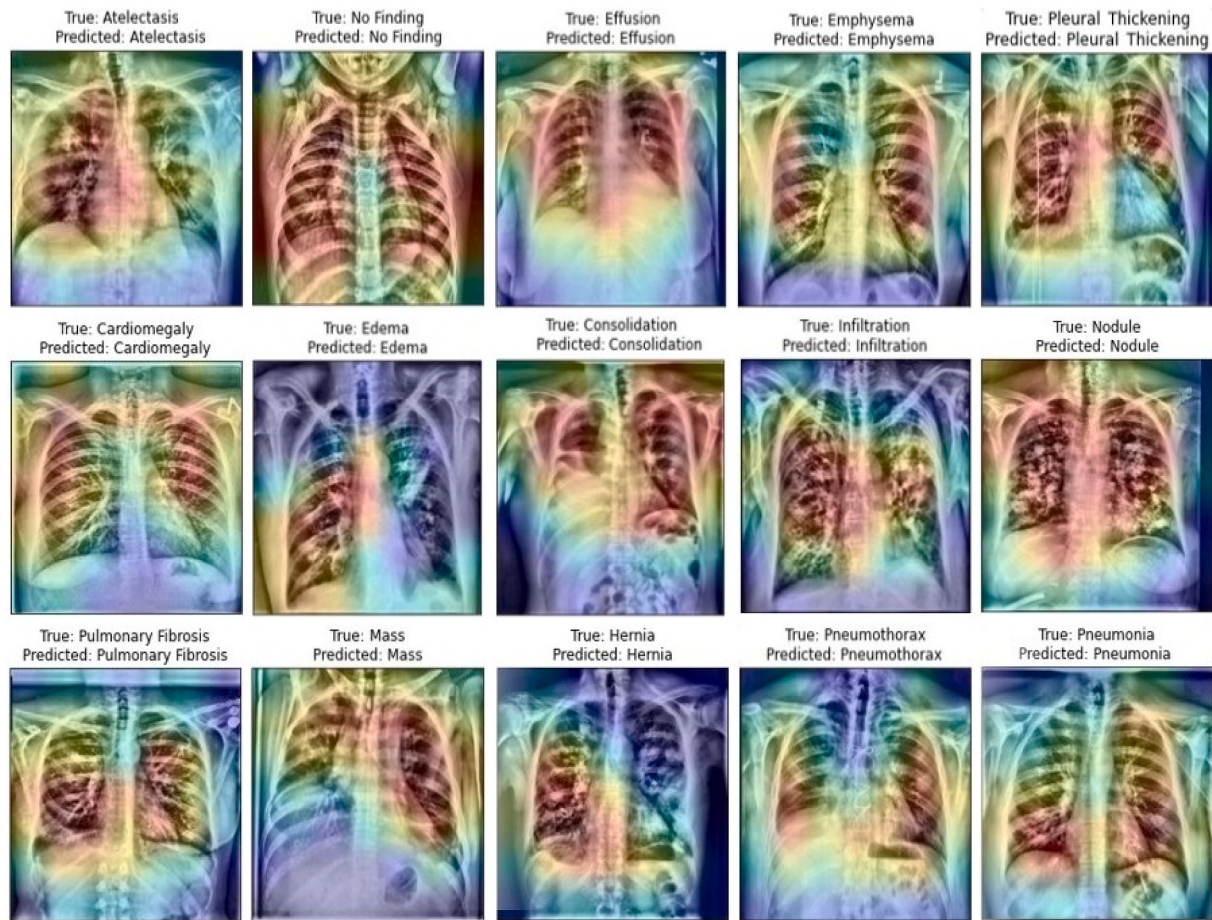


Fig. 13. Visualization of lung infections in X-ray images using Grad-CAM on MobileLungNetV2 model.

Table 6

Comparison of AUC value among the existing studies with the proposed study to determine performance using ChestX-Ray14 dataset.

Class/Models	Existing					Transfer learning models					Proposed
	[44]	[45]	[49]	[56]	[57]	InceptionV3	AlexNet	DenseNet121	VGG19	MobileNetV2	MobileLungNetV2
Atelectasis	0.795	0.817	0.751	0.801	0.767	0.711	0.784	0.851	0.802	0.798	0.936
Cardiomegaly	0.887	0.928	0.896	0.884	0.883	0.757	0.821	0.901	0.928	0.815	0.855
Effusion	0.875	0.875	0.882	0.872	0.828	0.777	0.769	0.796	0.777	0.778	0.931
Infiltration	0.703	0.745	0.705	0.702	0.709	0.783	0.698	0.844	0.741	0.864	0.912
Mass	0.835	0.880	0.844	0.822	0.821	0.875	0.744	0.792	0.786	0.935	0.883
Nodule	0.716	0.786	0.752	0.747	0.758	0.869	0.761	0.918	0.851	0.789	0.962
Pneumonia	0.742	0.779	0.763	0.733	0.731	0.843	0.772	0.809	0.903	0.822	0.904
Pneumothorax	0.863	0.893	0.878	0.865	0.846	0.893	0.879	0.756	0.798	0.919	0.967
Consolidation	0.786	0.809	0.798	0.796	0.745	0.786	0.853	0.767	0.846	0.858	0.889
Edema	0.892	0.892	0.870	0.891	0.835	0.801	0.827	0.73	0.811	0.914	0.958
Emphysema	0.875	0.939	0.918	0.894	0.895	0.748	0.734	0.776	0.756	0.887	0.935
Pulmonary Fibrosis	0.756	0.817	0.803	0.800	0.818	0.797	0.767	0.796	0.883	0.849	0.891
Pleural Thickening	0.774	0.814	0.779	0.786	0.761	0.834	0.859	0.791	0.878	0.852	0.872
Hernia	0.836	0.921	0.876	0.882	0.896	0.782	0.786	0.883	0.797	0.927	0.975
No Finding	–	–	–	–	–	0.908	0.897	0.877	0.792	0.941	0.989
Average	0.809	0.850	0.826	0.820	0.807	0.811	0.796	0.819	0.823	0.863	0.923

7. Conclusion

It is understood that lung disease is a major cause of mortality worldwide. This paper proposed a novel approach to detect lung lesions from X-ray imaging. An improved classification model, that is derived from the transfer learning model MobileNetV2, MobileLungNetV2, is suggested. In this study, 14 lung lesions and normal lung X-ray images are examined. Following image pre-processing, five transfer learning models are applied to the dataset. The models, InceptionV3, AlexNet,

DenseNet121, VGG19 and MobileNetV2 show overall classification accuracies of 79.11%, 82.95%, 88.9%, 88.99% and 91.6% respectively over the 15 classes of the dataset. Because MobileNetV2 outperforms the other models in terms of accuracy, the model is selected to be modified further to boost performance. The improved fine-tuned MobileLungNetV2 model is constructed using 16 blocks with new neural network layers and is trained on 300 epochs with hyper-parameters. The proposed model has an overall accuracy, and F1-score of 96.97% and 96.76%, respectively and the test loss of the model is as low as 0.15,

demonstrating that the model performs with much improved efficiency over other models. The proposed model achieves a higher classification accuracy compared to other pre-trained models. However, the study has limitations as it does not take into account the time complexity of the classification models. The time complexity of a program is the time required to execute it. As time complexity decreases, execution speeds improve. A model is determined to be highly efficient when it provides high classification accuracy with a low time complexity. Furthermore, training neural networks requires a substantial quantity of data. Though the dataset used in the study had a total of 51987 samples, an even larger dataset would increase the robustness of the model. A model's robustness depends on the change of the model's performance due to incorporating new data against training data. Additionally, the robustness of the model increases when the model is trained on multiple datasets. Some highly representative computational intelligence algorithms can be used for the classification task. For future work, algorithms such as like monarch butterfly optimization (MBO), earthworm optimization algorithm (EWA), elephant herding optimization (EHO), moth search (MS) algorithm, Slime mould algorithm (SMA), hunger games search (HGS), Runge Kutta optimizer (RUN), colony predation algorithm (CPA), and Harris Hawks optimization (HHO), can be employed for classification of pulmonary diseases. Additionally, the fine-tuned model can be applied to classify other datasets, containing different pulmonary diseases, as well as for other types of diseases that may be detectable using medical imaging.

Declaration of competing interest

The authors declare no conflicts of interests.

References

- [1] D.R. Brenner, J.R. McLaughlin, R.J. Hung, Previous lung diseases and lung cancer risk: a systematic review and meta-analysis, *PLoS One* 6 (2011), e17479.
- [2] F.M. Shamrat, S. Azam, A. Karim, R. Islam, Z. Tasnim, P. Ghosh, F. De Boer, Lungnet22: a fine-tuned model for multiclass classification and prediction of lung disease using X-ray images, *J. Personalized Med.* 12 (2022) 680.
- [3] O. Ruuskanen, E. Lahti, L.C. Jennings, D.R. Murdoch, Viral pneumonia, *Lancet* 377 (2011) 1264–1275.
- [4] D.S. Kermay, M. Goldbaum, W. Cai, C.C. Valentim, H. Liang, S.L. Baxter, A. McKeown, G. Yang, X. Wu, F. Yan, et al., Identifying medical diagnoses and treatable diseases by image-based deep learning, *Cell* 172 (5) (2018) 1122–1131.
- [5] Saleh Albahli, Efficient GAN-based Chest Radiographs (CXR) augmentation to diagnose coronavirus disease pneumonia, *Int. J. Med. Sci.* 17 (10) (2020) 1439.
- [6] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, *Commun. ACM* 60 (6) (2017) 84–90.
- [7] Y. Wang, X. Qiao, G.G. Wang, Architecture evolution of convolutional neural network using Monarch butterfly optimization, *J. Ambient Intell. Hum. Comput.* (2022) 1–15.
- [8] Z. Cui, F. Xue, X. Cai, Y. Cao, G.G. Wang, J. Chen, Detection of malicious code variants based on deep learning, *IEEE Trans. Inf. Inf.* 14 (7) (2018) 3187–3196.
- [9] G.G. Wang, M. Lu, Y.Q. Dong, X.J. Zhao, Self-adaptive extreme learning machine, *Neural Comput. Appl.* 27 (2) (2016) 291–303.
- [10] T. Franquet, Imaging of community-acquired pneumonia, *J. Thorac. Imag.* 33 (5) (2018) 282–294.
- [11] Y. Tang, Y. Tang, J. Xiao, R.M. Summers, Xlsor: A Robust and Accurate Lung Segmentor on Chest X-Rays Using Criss-Cross Attention and Customized Radiorealistic Abnormalities Generation, 2019 arXiv preprint arXiv:1904.09229.
- [12] J.C. Souza, J.O.B. Diniz, J.L. Ferreira, G.L.F. da Silva, A.C. Silva, A.C. de Paiva, An automatic method for lung segmentation and reconstruction in chest X-ray using deep neural networks, *Comput. Methods Progr. Biomed.* 177 (2019) 285–296.
- [13] S. Wang, Y. Yang, Z. Ma, X. Li, C. Pang, A.G. Hauptmann, Action recognition by exploring data distribution and feature correlation, in: 2012 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2012, pp. 1370–1377.
- [14] Y. Gordienko, P. Gang, J. Hui, W. Zeng, Y. Kochura, O. Alienin, O. Rokoviy, S. Stirenko, Deep learning with lung segmentation and bone shadow exclusion techniques for chest x-ray analysis of lung cancer, in: International Conference on Computer Science, Engineering and Education Applications, Springer, 2018, pp. 638–647.
- [15] N. Dey, S. Borra, A.S. Ashour, F. Shi, Machine Learning in Bio-Signal Analysis and Diagnostic Imaging, Academic Press, 2018.
- [16] B. Chen, J. Li, X. Guo, G. Lu, Dualchexnet: dual asymmetric feature learning for thoracic disease classification in chest x-rays, *Biomed. Signal Process Control* 53 (2019), 101554.
- [17] K. Santosh, S. Antani, D.S. Guru, N. Dey, Medical Imaging: Artificial Intelligence, Image Recognition, and Machine Learning Techniques, CRC Press, 2019.
- [18] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, in: Proceedings of the Advances in Neural Information Processing Systems, Lake Tahoe, NV, USA, 3–6 December 2012.
- [19] S. Akter, F.M. Shamrat, S. Chakraborty, A. Karim, S. Azam, Covid-19 detection using deep learning algorithm on chest X-ray images, *Biology* 10 (2021) 1174.
- [20] B. Zhou, A. Lapedriza, J. Xiao, A. Torralba, A. Oliva, Learning deep features for scene recognition using places database, in: Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 6–13 December 2014.
- [21] S. Xie, Z. Yu, Z. Lv, Multi-disease Prediction Based on Deep Learning: a Survey, CMES-computer Modeling in Engineering and Sciences, 2021.
- [22] L.S. Kondaka, M. Thenmozhi, K. Vijayakumar, R. Kohli, An intensive healthcare monitoring paradigm by using IoT based machine learning strategies, *Multimed. Tool. Appl.* 81 (26) (2022) 36891–36905.
- [23] Olga Russakovsky, Deng Jia, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, et al., Imagenet large scale visual recognition challenge, *Int. J. Comput. Vis.* 115 (3) (2015) 211–252.
- [24] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollr, L. Zitnick, Microsoft coco: common objects in context, in: Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014.
- [25] J. Johnson, A. Karpathy, L. Fei-Fei, DenseCap: fully convolutional localization networks for dense captioning, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016.
- [26] R. Krishna, Y. Zhu, O. Groth, J. Johnson, K. Hata, J. Kravitz, S. Chen, Y. Kalantidis, L.-J. Li, D.A. Shamma, et al., Visual Genome: connecting language and vision using crowdsourced dense image annotations, *Int. J. Comput. Vis.* 123 (2017) 32–73.
- [27] H.R. Roth, L. Lu, A. Seff, K.M. Cherry, J. Hoffman, S. Wang, J. Liu, E. Turkbey, R. M. Summers, A new 2.5D representation for lymph node detection using random sets of deep convolutional neural network observations, in: Proceedings of the International Conference on Medical Image Computing and Computer Assisted Intervention, Boston, MA, USA, 2014.
- [28] H. Shin, H. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, R. Summers, Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learnings, *IEEE Trans. Med. Imag.* 35 (2016) 1285–1298.
- [29] Q. Dou, H. Chen, L. Yu, L. Zhao, J. Qin, D. Wang, V. Mok, L. Shi, P. Heng, Automatic detection of cerebral microbleeds from mr images via 3d convolutional neural networks, *IEEE Trans. Med. Imag.* 35 (2016) 1182–1195.
- [30] Z. Tasnim, S. Chakraborty, F.M. Shamrat, A.N. Chowdhury, H.A. Nuha, A. Karim, S. B. Zahir, M.M. Billah, Deep learning predictive model for colon cancer patient using CNN-based classification, *Int. J. Adv. Comput. Sci. Appl.* 12 (2021).
- [31] A. Jamaludin, T. Kadir, A. Zisserman, SpineNet: automatically pinpointing classification evidence in spinal MRIs, in: Proceedings of the International Conference on Medical Image Computing and Computer Assisted Intervention, Athens, Greece, 2016.
- [32] H. Roth, L. Lu, A. Farag, H.-C. Shin, J. Liu, E.B. Turkbey, R.M. Summers, DeepOrgan: multi-level deep convolutional networks for automated pancreas segmentation, in: Proceedings of the International Conference on Medical Image Computing and Computer Assisted Intervention, Munich, Germany, 5–9 October 2015.
- [33] A. Setio, F. Ciompi, G. Litjens, P. Gerke, C. Jacobs, S. van Riel, M. Wille, M. Naqibullah, C. Snchez, B. van Ginneken, Pulmonary nodule detection in ct images: false positive reduction using multi-view convolutional networks, *IEEE Trans. Med. Imag.* 35 (2016) 1160–1169.
- [34] S. Bharati, P. Podder, M.R.H. Mondal, Hybrid deep learning for detecting lung diseases from X-ray images, *Inform. Med. Unlocked* 20 (2020), 100391.
- [35] K. Sriporn, C.F. Tsai, C.E. Tsai, P. Wang, Analyzing lung disease using highly effective deep learning techniques, *Health Care* 8 (No. 2) (2020, April) 107 (MDPI).
- [36] P. Rajpurkar, J. Irvin, K. Zhu, et al., CheXNet: radiologist-level pneumonia detection on chest x-rays with deep learning, *ArXiv 1711.05225 [preprint]*, <http://arxiv.org/abs/1711.05225>. (Accessed 2 February 2022). Posted 2017.
- [37] L. Yao, E. Pöhlitz, D. Dagunts, B. Covington, D. Bernard, K. Lyman, Learning to diagnose from scratch by exploiting dependencies among labels, *ArXiv 1710.10501 [preprint]*, <https://arxiv.org/abs/1710.10501>. (Accessed 2 February 2022). Posted 2017.
- [38] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, R.M. Summers, ChestX-ray8: hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017, pp. 2097–2106.
- [39] P. Rajpurkar, J. Irvin, R.L. Ball, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C.P. Langlotz, et al., Deep learning for chest radiograph diagnosis: a retrospective comparison of the CheXNeXt algorithm to practicing radiologists, *PLoS Med.* 15 (2018), e1002686.
- [40] B. Chen, Z. Zhang, J. Lin, Y. Chen, G. Lu, Two-stream collaborative network for multi-label chest X-ray Image classification with lung segmentation, *Pattern Recogn. Lett.* 135 (2020) 221–227.
- [41] B. Chen, J. Li, X. Guo, G. Lu, DualCheXNet: dual asymmetric feature learning for thoracic disease classification in chest X-rays, *Biomed. Signal Process Control* 53 (2019), 101554.
- [42] I. Pan, S. Agarwal, D. Merck, Generalizable inter-institutional classification of abnormal chest radiographs using efficient convolutional neural networks, *J. Digit. Imag.* 32 (5) (2019) 888–896.
- [43] I. Allouzi, M.B. Ahmed, A novel approach for multi-label chest X-ray classification of common thorax diseases, *IEEE Access* 7 (2019) 64279–64288.

- [44] T.K.K. Ho, J. Gwak, Multiple feature integration for classification of thoracic disease in chest radiography, *Appl. Sci.* 9 (19) (2019) 4130.
- [45] X. Gong, X. Xia, W. Zhu, B. Zhang, D. Doermann, L.A. Zhuo, Deformable gabor feature networks for biomedical image classification, in: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2021, pp. 4004–4012.
- [46] H. Wang, Y.Y. Yang, Y. Pan, P. Han, Z.X. Li, H.G. Huang, S.Z. Zhu, Detecting thoracic diseases via representation learning with adaptive sampling, *Neurocomputing* 406 (2020) 354–360.
- [47] C.F. Sabottke, B.M. Spieler, The effect of image resolution on deep learning in radiography, *Radiology: Artif. Intell.* 2 (1) (2020), e190015.
- [48] I.M. Baltruschat, H. Nickisch, M. Grass, T. Knopp, A. Saalbach, Comparison of deep learning approaches for multi-label chest X-ray classification, *Sci. Rep.* 9 (1) (2019) 1–10.
- [49] T.K.K. Ho, J. Gwak, Utilizing knowledge distillation in deep learning for classification of chest X-ray abnormalities, *IEEE Access* 8 (2020) 160749–160761.
- [50] S. Albahli, H.T. Rauf, A. Algosaibi, V.E. Balas, AI-driven deep CNN approach for multi-label pathology classification using chest X-Rays, *PeerJ Computer Science* 7 (2021) e495.
- [51] E. Rozenberg, D. Freedman, A.A. Bronstein, Learning to localize objects using limited annotation, with applications to thoracic diseases, *IEEE Access* 9 (2021) 67620–67633.
- [52] S. Bharati, P. Podder, M.R.H. Mondal, Hybrid deep learning for detecting lung diseases from X-ray images, *Inform. Med. Unlocked* 20 (2020), 100391.
- [53] Dataset: <https://www.kaggle.com/nih-chest-xrays/data>. (Accessed 25 February 2022).
- [54] Keras. Available online: <https://keras.io/>. (Accessed 25 February 2022).
- [55] MobileNetV2. Available Online: <https://iq.opengenus.org/mobilenetv2-architecture/>. (Accessed 26 February 2022).
- [56] M. Ivo, Baltruschat, hannes nickisch, michael grass, tobias knopp, and axel saalbach. Comparison of deep learning approaches for multi-label chest x-ray classification, *Sci. Rep.* 9 (6381) (2019).
- [57] S. Guendel, S. Grbic, B. Georgescu, S. Liu, A. Maier, D. Comaniciu, Learning to recognize abnormalities in chest X-rays with location-aware dense networks, in: *Proc. Iberoamer. Congr. Pattern Recognit.*, Springer, Cham, Switzerland, Nov. 2018, 757765.
- [58] T. Mahmood, J. Li, Y. Pei, F. Akhtar, An automated in-depth feature learning algorithm for breast abnormality prognosis and robust characterization from mammography images using deep transfer learning, *Biology* 10 (2021) 859.
- [59] K. Zuiderveld, Contrast Limited Adaptive Histogram Equalization, *ACM Digital Library*, New York, NY, USA, 1994.
- [60] P. Wang, J. Wang, Y. Li, P. Li, L. Li, M. Jiang, Automatic classification of breast cancer histopathological images based on deep feature fusion and enhanced routing, *Biomed. Signal Process Control* 65 (2021), 102341.
- [61] E.K. Hodson, D.R. Thayer, C. Franklin, Adaptive Gaussian filtering and local frequency estimates using local curvature analysis, *IEEE Trans. Acoust. Speech Signal Process.* ASSP-29 (Aug. 1981) 854–859.
- [62] M. Vaezi, B. Bavarian, Contrast-dependant spread filters, *Proc. SPIE: Image Processing Algorithms and Techniques* 1244 (1990) 100–107.
- [63] H. Jeong, C.I. Kim, Adaptive determination of filter scales for edge detection, *IEEE Trans. Pattern Anal. Mach. Intell. PAMI-14* (May 1992) 579–585.
- [64] G. Wang, W. Li, M. Aertsen, J. Deprest, S. Ourselin, T. Vercauteren, Aleatoric uncertainty estimation with test-time augmentation for medical image segmentation with convolutional neural networks, *Neurocomputing* 338 (2019) 34–45.
- [65] J. Nalepa, M. Marcinkiewicz, M. Kawulok, Data augmentation for brain-tumor segmentation: a review, *Front. Comput. Neurosci.* 13 (2019) 83.
- [66] N. Moshkov, B. Mathe, A. Kertesz-Farkas, R. Hollandi, P. Horvath, Test-time augmentation for deep learning-based cell segmentation on microscopy images, *Sci. Rep.* 10 (2020) 5068.
- [67] M. Amiri, R. Brooks, B. Behboodi, H. Rivaz, Two-stage ultrasound image segmentation using U-Net and test time augmentation, *Int. J. Comput. Assist. Radiol. Surg.* 15 (2020) 981–988.
- [68] D. Shanmugam, D. Blalock, G. Balakrishnan, J. Guttag, Better aggregation in test-time augmentation, in: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, Montreal, BC, Canada, 2021, pp. 1214–1223.
- [69] D. Jha, P.H. Smedsrud, D. Johansen, T. de Lange, H.D. Johansen, P. Halvorsen, M. A. Riegler, A comprehensive study on colorectal polyp segmentation with ResUNet ++, conditional random Field and test-time augmentation, *IEEE J. Biomed. Health Inform.* 25 (2021) 2029–2040.
- [70] I. Kandel, M. Castelli, Improving convolutional neural networks performance for image classification using test time augmentation: a case study using MURA dataset, *Health Inf. Sci. Syst.* 9 (2021) 33.
- [71] D. Hoar, P.Q. Lee, A. Guida, S. Patterson, C.V. Bowen, J. Merrimen, C. Wang, R. Rendon, S.D. Beyea, S.E. Clarke, Combined transfer learning and test-time augmentation improves convolutional neural network-based semantic segmentation of prostate cancer from multi-parametric MR images, *Comput. Methods Progr. Biomed.* 210 (2021), 106375.
- [72] C. Gonzalo-Martín, A. García-Pedrero, M. Lillo-Saavedra, Improving deep learning sorghum head detection through test time augmentation, *Comput. Electron. Agric.* 186 (2021), 106179.
- [73] S. Cohen, N. Dagan, N. Cohen-Inger, D. Ofer, L. Rokach, ICU survival prediction incorporating test-time augmentation to improve the accuracy of ensemble-based models, *IEEE Access* 9 (2021) 91584–91592.
- [74] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-cam: visual explanations from deep networks via gradient-based localization, in: *Proceedings of the 2017 IEEE International Conference on Computer Vision*, Venice, Italy, 22–29 October 2017, pp. 618–626.
- [75] InceptionV3. Available Online: <https://keras.io/api/applications/inceptionv3/>. (Accessed 26 February 2022).
- [76] DenseNet121. Available Online: <https://keras.io/api/applications/densenet/>. (Accessed 26 February 2022).
- [77] VGG19. Available Online: <https://keras.io/api/applications/vgg/>. (Accessed 26 February 2022).
- [78] U. Avni, H. Greenspan, E. Konen, M. Sharon, J. Goldberger, X-ray categorization and retrieval on the organ and pathology level, using patch-based visual words, *IEEE Trans. Med. Imag.* 30 (3) (2011) 733–746.
- [79] S. Jaeger, A. Karargyris, S. Candemir, L. Folio, J. Siegelman, F. Callaghan, Automatic tuberculosis screening using chest radiographs, *IEEE Trans. Med. Imag.* 33 (2) (2014) 233–245.
- [80] P. Pattarapisetwong, W. Chiracharit, Automatic lung segmentation in chest radiographs using shadow filter and multilevel thresholding, *Int Comput Sci Eng Conf* (2016) 1–6.
- [81] Xiaosong Wang, Yifan Peng, Le Lu, Zhiyong Lu, Mohammadhadi Bagheri, Ronald M. Summers, Chestx-ray8: hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 2097–2106.
- [82] Qingji Guan, Yaping Huang, Multi-label Chest X-Ray Image Classification via Category-wise Residual Attention Learning, *Pattern Recognition Letters*, 2018.
- [83] Li Yao, Eric Poblentz, Dmitry Dagunts, Covington Ben, Devon Bernard, Kevin Lyman, Learning to Diagnose from Scratch by Exploiting Dependencies Among Labels, 2017 arXiv preprint arXiv:1710.10501.
- [84] Zhe Li, Chong Wang, Mei Han, Xue Yuan, Wei Wei, Li-Jia Li, Li Fei-Fei, Thoracic disease identification and localization with limited supervision, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8290–8299.
- [85] Pranav Rajpurkar, Jeremy Irvin, Kaylie Zhu, Brandon Yang, Hershel Mehta, Tony Duan, Daisy Ding, Aarti Bagul, Langlotz Curtis, Katie Shpanskaya, et al., CheXnet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning, 2017 arXiv preprint arXiv:1711.05225.
- [86] X. Wang, et al., ChestX-Ray8: Hospital-Scale Chest X-Ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases, 2017, pp. 3462–3471.
- [87] L. Yao, et al., Learning to Diagnose from Scratch by Exploiting Dependencies Among Labels, 2017 arXiv preprint arXiv:1710.10501.
- [88] P. Kumar, M. Grewal, M.M. Srivastava, Boosted Cascaded Convnets for Multilabel Classification of Thoracic Diseases in Chest Radiographs, 2017 arXiv preprint arXiv:1711.08760.
- [89] P. Rajpurkar, et al., CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning, 2017 arXiv preprint arXiv:1711.05225.
- [90] S. Ioffe, C. Szegedy, Batch normalization: accelerating deep network training by reducing internal covariate shift, in: *International Conference on Machine Learning*, 2015.
- [91] G. Huang, et al., Densely connected convolutional networks, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017.