Arabian Journal for Science and Engineering (2022) 47:2123–2139
https://doi.org/10.1007/s13369-021-06127-z

RESEARCH ARTICLE-COMPUTER ENGINEERING AND COMPUTER SCIENCE

Diagnosis of Pediatric Pneumonia with Ensemble of Deep Convolutional Neural Networks in Chest X-Ray Images

Enes Ayan[1] · **Bergen Karabulut**[1] · **Halil Murat Ünver**[1]

Received: 6 May 2021 / Accepted: 17 August 2021 / Published online: 12 September 2021

Abstract
Pneumonia is a fatal disease that appears in the lungs and is caused by viral or bacterial infection. Diagnosis of pneumonia in chest X-ray images can be difficult and error-prone because of its similarity with other infections in the lungs. The aim of this study is to develop a computer-aided pneumonia detection system to facilitate the diagnosis decision process. Therefore, a convolutional neural network (CNN) ensemble method was proposed for the automatic diagnosis of pneumonia which is seen in children. In this context, seven well-known CNN models (VGG-16, VGG-19, ResNet-50, Inception-V3, Xception, MobileNet, and SqueezeNet) pre-trained on the ImageNet dataset were trained with the appropriate transfer learning and fine-tuning strategies on the chest X-ray dataset. Among the seven different models, the three most successful ones were selected for the ensemble method. The final results were obtained by combining the predictions of CNN models with the ensemble method during the test. In addition, a CNN model was trained from scratch, and the results of this model were compared with the proposed ensemble method. The proposed ensemble method achieved remarkable results with an AUC of 95.21 and a sensitivity of 97.76 on the test data. Also, the proposed ensemble method achieved classification accuracy of 90.71 in chest X-ray images as normal, viral pneumonia, and bacterial pneumonia.

Keywords Deep learning · Convolutional neural networks · Pneumonia · Transfer learning · Medical image analysis

1 Introduction

Pneumonia is an acute infection that occurs in the lungs and is one of the leading causes of child mortality worldwide. According to reports published in 2006 [1], 2011 [2], and 2018 [3], pneumonia was ranked first among the causes of child mortality under the age of five-year-old children, with 19%, 18%, and 16%, respectively. The late diagnosis of the disease due to limited resources makes pneumonia the most dangerous disease that causes childhood deaths especially in low and middle-income countries [4]. Moreover, pneumonia also poses a risk to people over 65 years of age and those with prior health problems [5]. The World Health Organization (WHO) describes some symptoms of pneumonia as the presence of fast breathing, chest indrawing, cough, cold and difficulty breathing [6]. If pneumonia is detected in its early stage, antibiotic therapy can be initiated to effectively treat pneumonia [6]. The most suitable method for the diagnosis of pneumonia is radiography (chest X-rays) [7]. However, medical images are still analyzed by human radiologists with speed, fatigue and experience constraints. Raising a qualified radiologist requires large financial costs and takes years [8]. In addition, there are not enough radiologists in rural areas of low-income countries. Moreover, diagnosis of pneumonia in chest X-ray is a difficult task even for well-trained radiologists because of similarities between pneumonia and other diseases. For example, if bacterial or viral pneumonia cases are misdiagnosed by experts, this brings with the wrong treatment process, which can result in the death of the patient [9]. In addition to all this, the new Coronavirus (Covid-19) is quite similar to pneumonia and it is difficult to differentiate them from each other [10]. For all these reasons, there is a high demand to develop computer-aided systems (CAD) to facilitate the diagnosis of pneumonia [11]. In recent years, deep learning-based CAD approaches have gained popularity in solving medical problems. The main advantage of these CAD approaches is making rapid inference and the ability to perform complex cognitive tasks that

✉ Enes Ayan
enesayan@kku.edu.tr

[1] Department of Computer Engineering, Faculty of Engineering and Architecture, Kirikkale University, Yahsihan, Kirikkale, Turkey

require special expertise [12]. Deep learning-based CAD systems, especially convolutional neural networks (CNNs) inspired by the mammalian visual cortex have the potential to evaluate and learn a large number of attributes by themselves, including those that were not previously considered by radiologists [13]. CNNs have been used successfully in various medical imaging problems such as skin lesion segmentation [14] and skin cancer classification from dermoscopic images [15], detection of diabetic retinopathy from retinal fundus images [16], early lung cancer detection in CT images [17], detection of arrhythmia from electrocardiograms [18], classification of pulmonary tuberculosis in radiographs [19], and hemorrhage detection from computed tomography scans [12].

In this study, instead of training or designing a CNN model from scratch, the recommended methodology is based on the idea of training well-known and successful CNN architectures using appropriate transfer learning and fine-tuning strategies, and then combining the strengths of these models. For this purpose, VGG-16, VGG-19 [20], ResNet-50 [21], Inception-V3 [22], MobileNet [23], SqueezeNet [24], and Xception [25] CNN models were trained for detection of pneumonia in chest X-ray images. As a result of the training, it was observed that each CNN model has a different generalization abilities on the dataset. Based on this observation, the three most successful CNN models, ResNet-50, Xception and MobileNet were selected for the ensemble method. The classification results obtained by the selected CNN models were combined using a probability-based ensemble method for achieving the final classification output. As a result of this ensemble methodology, satisfying classification results were obtained. We also developed and trained a new CNN model to validate the effectiveness of the proposed ensemble method. In addition, there are few studies in the literature that perform normal, viral pneumonia, and bacterial pneumonia classification in chest X-ray images. Most studies focus on performing normal and pneumonia classification. Therefore, we also re-trained selected CNN models for classification of test data as normal, viral pneumonia, and bacterial pneumonia. Another issue investigated in this study is the effect of class-based balanced data usage on the classification performance of CNN networks. According to the test results, it was observed that CNNs trained with balanced data achieved more successful results in terms of classification performance. The main contributions of this work are summarized below.

- We proposed an ensemble method using various convolutional neural network models' forecasts to strengthen the final classification decision.
- Effective transfer learning and fine-tuning strategies were analyzed instead of training a convolutional neural network from scratch.
- A novel CNN model was trained from scratch and the achieved results were compared with the classification performance of the proposed ensemble method.
- We observed the class distribution effect on the classifiers while classifying chest X-ray images

The rest of the paper is organized as follows: Sect. 2 provides review of related works. Section 3 presents the dataset, materials and methods used in the study. In Sect. 4, experimental setup and evaluation metrics are explained. In Sect. 5, the classification performance of the method is given. In Sect. 6 the proposed method is compared with the successful classification methods in the literature. The final section includes the conclusions of the study.

2 Related Works

Automated diagnosis of pneumonia in chest X-ray images is one of the most remarkable developments in recent times. There are many studies in which researchers have tried to diagnose pneumonia using deep CNN models. Rajpurkar et al. [26] developed a 121-layer CNN model named CheXNet. They trained CheXNet with 100,000 chest X-ray images with 14 different diseases. The proposed model was tested with 420 chest X-ray images and the test results were compared with those of expert radiologists. As a result, it was seen that the deep learning-based CNN model exceeded the average performance of radiologists in detecting pneumonia. Kermany et al. [27] utilized transfer learning for training a CNN model performing the detection of pneumonia in chest X-ray images. Rajaraman et al. [28] used a CNN-based system for classifying chest X-rays as normal vs pneumonia, bacterial vs viral pneumonia and normal, bacterial vs viral pneumonia. They trained CNN models with the region of interest areas (ROI) that include only the lungs instead of the whole image. Stephen et al. [29] proposed a CNN model. Unlike other methods based solely on transfer learning or traditional handcrafted techniques, they trained the CNN model from scratch to extract attributes from a given chest X-ray image to achieve remarkable classification performance, and used it to determine if a person was infected with pneumonia or not. Liang and Zheng [30] performed pneumonia detection with a CNN model architecture using residual connections and dilated convolution methods. They also discovered the transfer learning effect on CNN models when classifying chest X-ray images.

Siddiqi [31] proposed a sequential CNN model with 18 layers that performs automatic pneumonia diagnosis. Chouhan et al. [32] used five pre-trained CNN models on ImageNet for the detection of pneumonia by using transfer

learning and ensemble methodology. Gu et al. [33] proposed a two-stage method to identify bacterial and viral pneumonia. The proposed method consists of lung region identification using a fully convolutional network (FCN) and pneumonia category classification with a deep convolutional neural network (DCNN). Rahman et al. [34] performed the detection of pneumonia using four pre-trained CNN models on ImageNet with the transfer learning method. They classified chest radiography images with three different classification schemes as normal vs pneumonia, bacterial vs viral pneumonia, and normal, bacterial vs viral pneumonia. Togacar et al. [35] used three known CNN models for the feature extraction phase of the pneumonia classification problem. They trained each model separately using the same data and obtained 1000 features from the last fully connected layers of each CNN. 1000 features reduced use of the minimum redundancy maximum relevance (mRMR) feature selection method and obtained features given as an input to machine learning classification algorithms for the pneumonia classification problem. Mittal et al. [36], proposed a CapsNet CNN model using multi-layered capsules for diagnosis of pneumonia in chest X-ray images.

3 Materials and Methods

In this study, our main aim is to develop an effective pneumonia detection system in chest X-ray images by using different deep convolutional neural networks. Our methodology is summarized in Fig. 1. This consists of data augmentation, transfer learning and fine-tuning strategies, feature extraction, and probability-based ensemble classification. All steps are explained in more detail under the relevant title. The dataset used in the study is introduced first before moving on to the proposed deep learning-based method. Understanding the desired output versus the given input image may make it easier to understand the method we proposed in the study.

3.1 Dataset

The dataset used in this study was provided by Kermany and Goldbaum [37] based on a chest X-ray scan database from pediatric patients from one to five years of age at the Guangzhou Women and Children's Medical Center. The dataset consists of 5,856 chest X-ray images in total. In the training subset of the dataset, there are chest X-ray images from 5,232 patients, 3,883 of which are labeled as pneumonia and 1,349 as normal. In the test subset of the dataset, there are chest X-ray images from 624 patients, 390 of which are labeled as pneumonia and 234 as normal. Also, there are two different types of agents, bacterial and viral, in pneumonia patients. Table 1 shows the numerical distribution of normal, bacterial and virus-derived samples in the dataset. In addition, some images in the dataset labeled as pneumonia or normal are given as an example in Fig. 2. The quality of the images in this dataset differs due to various factors such

Table 1 Summary of the chest X-ray dataset

Class	Train	Validation	Test
Normal	1349	234	234
Pneumonia Viral	1345	148	148
Pneumonia Bacterial	2538	242	242
Total	5232	624	624

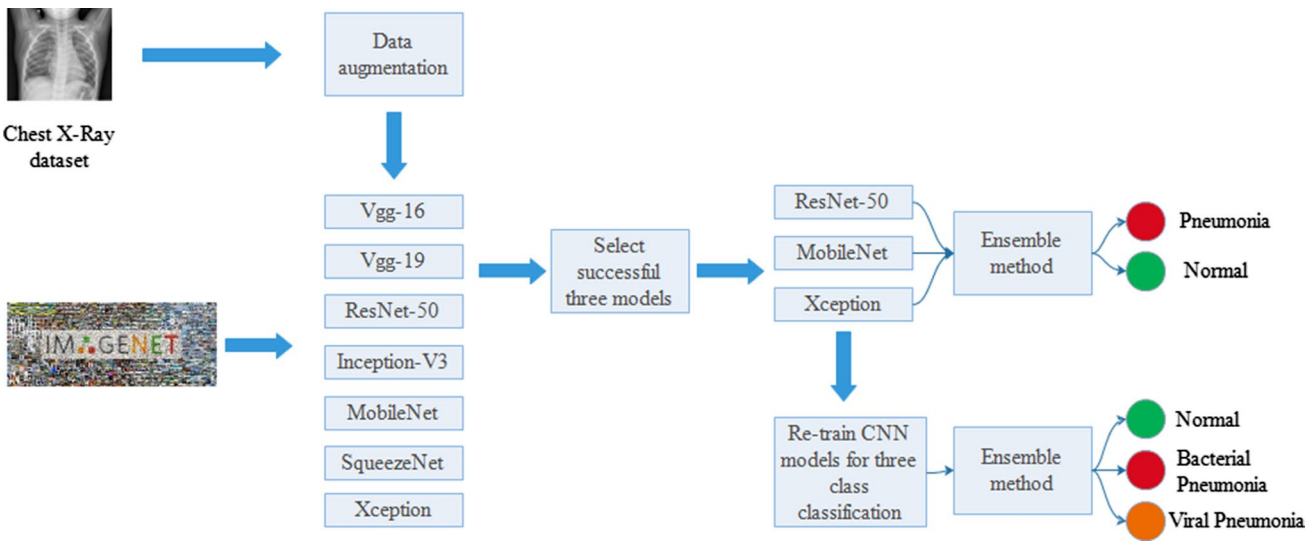

Fig. 1 Summary of the proposed method

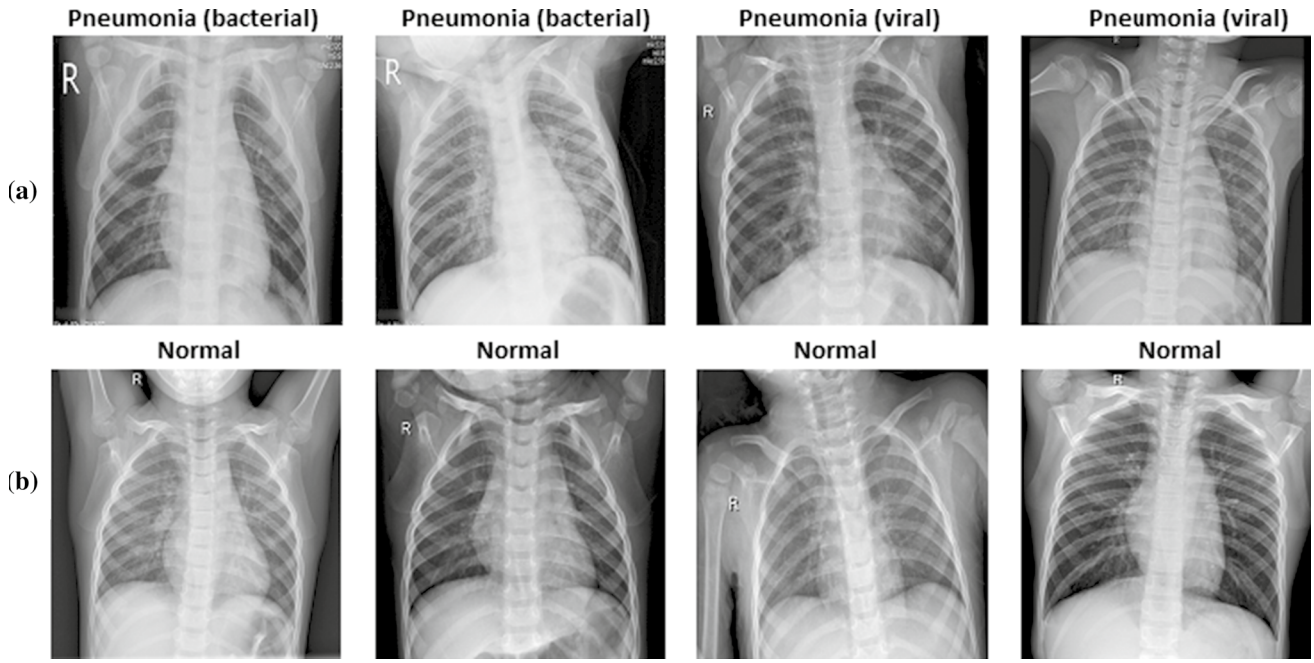

Fig. 2 Images from chest X-ray dataset: **a** pneumonia labeled images **b** normal labeled images

as different scanning devices, operator's working habits, and shooting positions.

3.2 Data Augmentation and Transfer Learning-Fine Tuning

For improving the generalization ability of CNN models, there is a need for massive data during the training phase [38]. However, usually not enough data can be obtained for some computer vision problems. Especially, obtaining and labeling data for medical problems is a laborious and time-consuming task. Fortunately, different methods have been developed to overcome this problem. Data augmentation, which is one of this methods, improves the model's generalization ability, prevents overfitting, and increases the model accuracy [38–40]. In this study, the data augmentation method was used before and during the training phase. When the dataset was examined before the training, it was seen that the number of normal-labeled images was 2,534 less than that of pneumonia labeled images. In order to reduce this gap, image processing methods were applied to the normal chest X-ray images. 1,349 chest X-ray images in the training dataset were randomly augmented using rotation (range of angle + 10, − 10), zooming (range, 0.8–1.2), and flipping (horizontal), and 2,534 new augmented chest X-ray images were obtained. In this way, the class-based balance was achieved in the train part of the dataset. Moreover, while training selected models, real-time data augmentation (shifting, rotating, zooming, and flipping) methods were used for avoiding overfitting.

Another important method that is used to improve the performance of deep neural networks is transfer learning. The main philosophy of transfer learning is based on the use of knowledge learned for a particular task to solve another similar problem [41]. Recently, very few researchers have trained a CNN network from scratch. Instead, they have used CNN's filters pre-trained on ImageNet [42] data, usually consisting of 1.2 million images and 1,000 classes [43]. In this way, the training time is decreased considerably [44].

There are different approaches to use of transfer learning in deep neural networks such as using feature extractor, fine-tuning and pre-trained models [45]. Choosing one of these approaches for the training a network varies depending on the problem and whether your data has similar features to ImageNet data. In the first layers of a CNN network, general features appear that are independent of data, such as edges, corners, and colors while toward the next layers, more data-specific features emerge like textures. Therefore, first-layer kernels can be used for different datasets and different problems. In the fine-tuning process, some convolutional layers from the beginning of the model are frozen during the training and weights of the frozen layers are not updated. In this study, a two-stage transfer learning strategy was used. In the first stage, seven well-known CNN models were trained with different transfer learning and fine-tuning strategies for classifying chest X-ray images as pneumonia vs normal. Before starting training, the weights of the convolutional layers

were taken from pre-trained ImageNet weights. The classifier parts of the networks were trained from scratch. There are different hyperparameters while training a CNN model with a fine-tuning method. These hyperparameters are the number of frozen convolutional layers, the number of fully connected layers, the number of dropouts used between the fully connected layers, the optimization algorithm, the learning rate, and epoch size. In the study, we determined optimal hyper-parameters by the trial-and-error strategy given in Fig. 3. At the end of the trials, we saved the most successful hyper-parameters and model weights. According to the test results, the three most successful CNN models were selected for the ensemble method among the seven CNN models.

The selected CNN models were tested using the ensemble method for classifying chest X-rays as pneumonia or normal. These CNN models are ResNet-50, Xception and MobileNet. In the first model, ResNet-50, configuration, the first 100 layers were frozen. After the convolution layers, global average pooling layer was used instead of a flattened layer. In this way, the loss of spatial relationships in the image was prevented and the parameter size was reduced. After the global average pooling layer, two fully connected layers with 512 neurons and one output layer with two neurons were added to the model (512, 512, 2). In the Xception model configuration, the first 60 layers were frozen, and the remaining layers were trained. After the convolution layers, global average pooling layer was used, followed by three fully connected layers (512, 512, 2). In the MobileNet configuration, the first 20 layers were frozen, global average pooling layer was used in the output of the convolutional layer, and fully connected layers, 512, 512, 2, were used as classifier. Before the fully connected layers in every CNN model, dropout layers were used to avoid overfitting by 0.5%. Also, batch normalization layers were used to overcome the internal covariate shift problem. In addition to all these, L2 regularization method was used in fully connected layer by 0.001 to prevent overfitting. All models were trained with balanced and imbalanced chest X-ray datasets. The selected models accept different resolution images, for example, ResNet-50 and MobileNet accept $224 \times 224 \times 3$ RGB image as input, while Xception accepts $299 \times 299 \times 3$ RGB image as input. In Fig. 4, the fine-tuning configurations of the selected successful CNN models are given. In addition, information about fine-tuning strategies is presented in Table 2.

In the second transfer learning strategy, the previously selected three CNN models which were trained with chest X-rays were trained for classifying chest X-rays as normal, bacterial pneumonia, or viral pneumonia. For this purpose, CNN models previously trained to classify normal or pneumonia were re-trained using the last training weights for classifying normal, bacterial pneumonia, or viral pneumonia. Only the output layers of the selected CNN models were changed as three and they were trained using the same hyper-parameters and fine-tuning strategies for 25 epochs with the chest X-ray dataset.

3.3 Convolutional Neural Networks

CNNs are developed to learn spatial hierarchies of features from data using a backpropagation algorithm. They have key components such as convolution, pooling, and fully connected (FC) layers [46]. In the convolution layer, a fixed size filter (3×3, 5×5, 7×7) is applied to the image for extracting distinctive features. After the convolution operation, an activation map is created for every filter. Previous layer activation maps are used as input for the next layer filters. The pooling layer reduces image size while keeping image features. Thus, a decrease is revealed in the model parameters, and calculation costs are reduced. The FC layers are the output of CNNs, and they use features extracted by convolutional layers for classification. Within the scope the study, seven well-known CNN models were trained to diagnose pneumonia in chest X-ray images. According to the test results, the three most successful

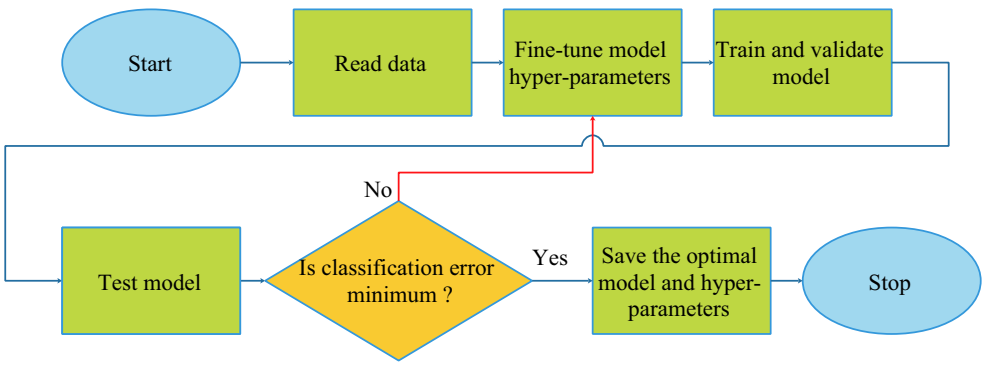

Fig. 3 Determining optimal hyperparameters

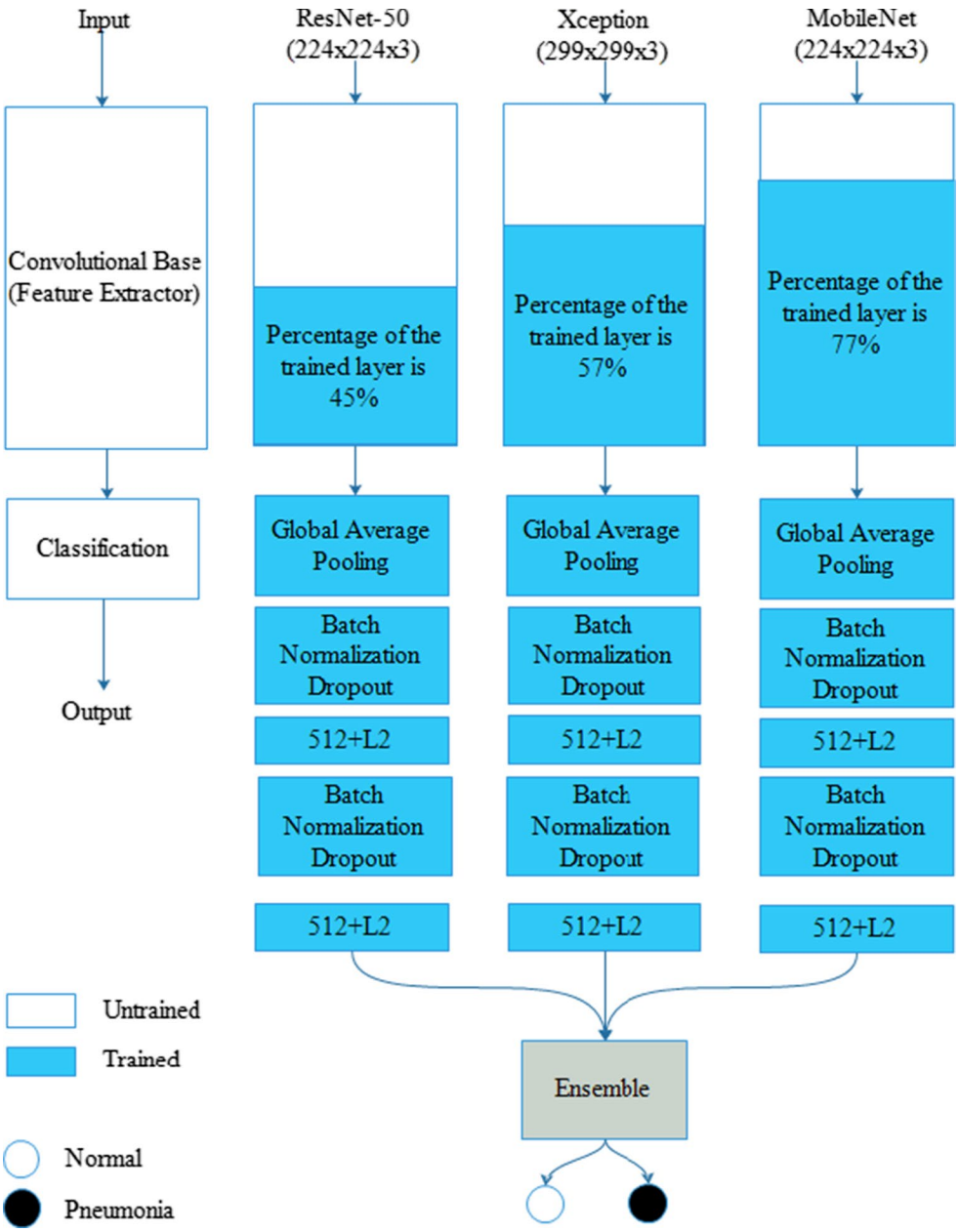

Fig. 4 Transfer learning and fine-tuning strategies for ResNet-50, Xception and MobileNet models

Table 2 The number of parameters and layers belonging to the CNN models

Models	Total number of parameters	Total trained parameters	Total untrained parameters	Total number of layers	Total number of trained layers
ResNet-50	24,912,770	20,769,794	4,142,976	183	83
Xception	22,184,490	17,252,138	4,932,352	140	80
MobileNet	4,023,490	3,968,130	55,360	90	70

CNN models were selected for the ensemble method. General information about the selected successful CNN architectures is given under the following subtitles.

3.3.1 ResNet

ResNet was introduced in 2015. In the same year, it won the Large-Scale Visual Recognition Challenge (ILSVRC)

2015 challenge due to its excellent performance in the image classification category. In deep neural networks, during the backpropagation, calculated gradients decrease through the layers and they become smaller and smaller when they reach the bottom of the network. Therefore, the weights of the initial layers either update very slowly or remain the same. This situation affects the training negatively and is known as the vanishing gradient problem. He et al. solved this problem using residual connections between the layers. In traditional neural networks, each layer is fed by the previous layer. However, in a network with residual connections, each layer is fed 2 or 3 hops away from the next layer. Through this method, the vanishing gradient problem is solved in deeper networks. There are different ResNet architectures such as ResNet-18, ResNet-34, ResNet-50, ResNet-101 and ResNet-152. In this study, ResNet-50 architecture was used for classification of the chest X-rays.

3.3.2 Xception

Xception was designed by François Chollet in 2017, inspired by the Inception V3 architecture. The main contribution of the Xception model is that it uses the depthwise separable convolution process with a little change. A depthwise separable convolution process consists of depthwise convolution followed by a pointwise convolution. In Xception architecture, depthwise separable convolution is reversed, feature maps are narrowed by applying a 1×1 filter, and the results are combined by applying a separate 3×3 filter. The Xception model has 36 convolutional layers to extract features, fully connected layers, and a logical classification layer at the output. Each of the convolutional layers, consisting of 14 modules in total, has linear residual connections.

3.3.3 MobileNet

The MobileNet was proposed by Andrew G. Howard and colleagues in 2017, to create a light but effective model that can be used in mobile or embedded systems. MobileNet architecture has reduced the number of parameters, calculation capacity, and model complexity by making use of the depthwise separable convolution process. Through this, it has been made available on mobile and embedded devices that have hardware limitations. MobileNet consists of 28 layers and after every depthwise separable convolution, batch normalization and ReLU layers are used.

3.3.4 Proposed CNN Model

We designed a novel CNN architecture (Pneumonia Net, PNet) in light of the recent developments in CNN architectures for detecting pneumonia in chest X-ray images. The PNet model is consisting of five convolution blocks and one output block, while designing the PNet we utilized depthwise separable convolutions (DWSC) [22], residual connections [21], squeeze and exciton blocks (SE) [47], and dilated convolutions [30]. In the first block of model, we used a normal convolution operation using 3×3 filters with a 2-dilation rate to increase the receptive field. After convolution operations, batch normalization and average pooling layers were used. In the second block, we used DWSCs to reduce computational cost [25]. After DWSCs, one batch normalization and one average pooling layers were used, respectively. Then we utilized the SE block with a residual connection for extracting useful feature channels and ignoring useless information. The other convolutional blocks are the same as block two. Only in the last convolutional block, average pooling wasn't used, and the block was terminated with a global average pooling layer. There are two Dropout layer, one batch normalization layer and two dense layers in the output block of the model. The detailed architecture is presented in Fig. 5. The PNet was trained from scratch with balanced and unbalanced datasets. In the training phase, we set number of epochs to 50, batch size 32, learning rate 1e-4. Adam was used for optimizing categorical cross-entropy loss function. The softmax activation function was used at the last dense layer. The proposed model has 2,877,313 total parameters.

4 Experimental Setup

In the study, seven well-known CNN models and a novel CNN model were trained for classifying chest X-ray images. During the training of well-known models, different transfer learning and fine-tuning strategies were tried, and configurations ensuring successful results were used in the testing phase. At the training stage, a batch size of 32, learning rate of 1e-4 were determined. We trained models with different epoch sizes but after 25 epochs models started to overfit. Adam was used as the optimization function to minimize categorical cross-entropy loss function. The softmax activation function was used in the last layer for classification. Early stopping was utilized to overcome overfitting of models. All experiments were performed on a workstation with a 64 GB RAM and Nvidia 1080 Ti graphics card with Ubuntu 14.04 operating system. In the software development step, Python programming language and Keras deep learning library were used. Among the seven CNN models, the three most successful ones were selected for the ensemble method.

4.1 Ensemble Methodology

In general, gathering expert opinions in solving a problem increases the accuracy and reliability of prediction. There

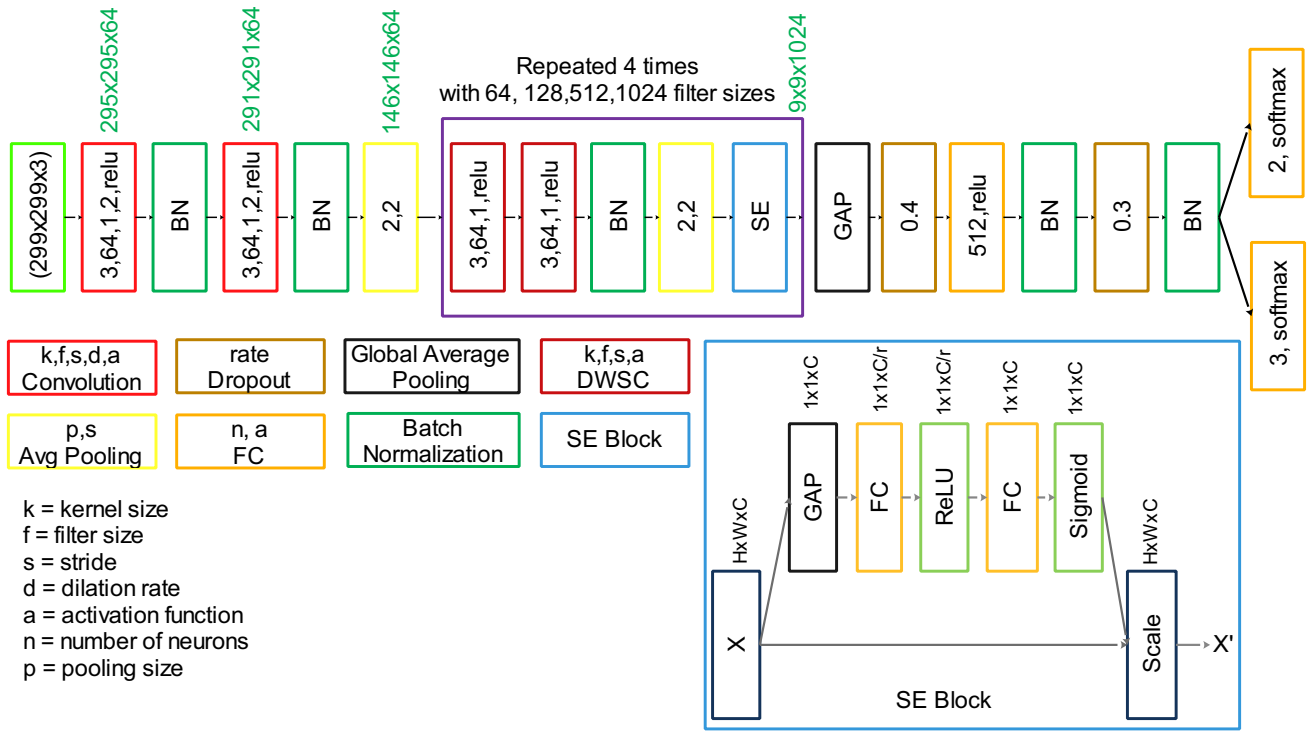

Fig.5 Details of PNet architecture

are two types of ensemble strategy used in CNN models in the literature.

In the first, various CNN models are used to extract features from the image. The extracted features are combined and used for classification tasks with different machine learning algorithms such as k-nearest neighbors, support vector machines, decision trees, logistic regression, naive bayes, etc. [35] This method has some constraints such as two different training processes and complex algorithms. In the second method, model predictions are combined using a mathematical model. This method is simple, fast and reliable [48]. The advantage of this method is that the ensemble system correctly classifies the data as a result of the voting done by the correct prediction of other models.

For this purpose, an ensemble method was used to increase classification performance in this study. As a result of the tests conducted after the training phase, it was seen that the classification performance of each network was different. Some networks achieved better results in detecting normal labeled samples, while others did better in detecting pneumonia labeled samples. In order to achieve the highest possible accuracy, a CNN ensemble method was designed, and a final decision was obtained by evaluating the decision of each network.

The proposed ensemble method is based on probabilistic voting. We have considered the output of each CNN model to determine the confidence values of each class. A CNN is considered as a function $f : x \rightarrow R^n$ that calculates n confidence values for each sample $i = 1, \ldots, n$ as $k_i \epsilon R$. For a new image, a previously unseen x CNN model outputs new k_i values where $k_i \epsilon [0, 1]$ and $\sum_{i=1}^{n} k_i = 1$. In our classification problem k_1, k_2, k_3 indicates the confidence scores of a CNN for each class, as normal (c_1), pneumonia (c_2), or bacterial pneumonia (c_1), normal (c_2), and viral pneumonia (c_3). A CNN determines the class of a previously unseen image x using maximum likelihood function:

$$x \rightarrow c_i, \quad if\ k_i = \max(f(x)) \tag{1}$$

In the classification scenario we used, each CNN model has two confidence outputs in the first step, and three confidence outputs in the second step. So, we calculated $k_{ij}^{'}$ confidence scores for each model ($k_{ij}^{'} \in [0, 1], i = 1, \ldots, n,\ j = 1, \ldots, m$). After that, we summed class-based $k_{ij}^{'}$ scores of each model and divided them by the number of ensemble models (m). In our case we have three CNN models and $m = 3$. To determine the final prediction, a maximum likelihood function was used. The formula for this voting system is given in the below equation:

$$k_i'' = \frac{\sum_{j=1}^{m} k_{ij}^{'}}{m},\ i = 1, \ldots, n \tag{2}$$

$$\text{where } k'' = \left(k_1'', \ldots, k_n''\right) \quad x \to c_i, \quad if\ k_i'' = \max\left(k''\right) \tag{3}$$

4.2 Evaluation Criteria

In the study, CNN models were trained on the training data, the model hyperparameters were fine-tuned on the validation data and the performance evaluation was done on the test data. This approach is named train-validation-test design. There are various evaluation metrics for measuring the performance of a classification system. The classification performance of the proposed method in this study was evaluated using the accuracy (Acc), sensitivity (Sen), specificity (recall) (Spe), precision, f1-score, the area under curve (AUC), and receiver operating characteristic (ROC) curve criteria. Especially, AUC and ROC are reliable criteria in datasets that have an imbalanced distribution. AUC represents the area under the ROC curve.

Table 3 Classification results (%) of the proposed method and CNN models on balanced data

Models	Acc	Sen	Spe	AUC
VGG-16	92.94	94.35	90.59	92.48
VGG-19	91.02	90.05	89.31	90.68
ResNet-50	**94.42**	**95.89**	**91.45**	**93.67**
InceptionV3	93.91	96.66	89.31	92.99
Xception	**95.03**	**97.43**	**91.02**	**94.23**
MobileNet	**94.87**	**96.92**	**91.45**	**94.19**
SqueezeNet	94.07	94.61	93.88	93.89
PNet	93.91	96.41	89.74	93.08
Ensemble	**95.83**	**97.76**	**92.73**	**95.21**

Bold values represent selected best model results, and proposed model results

5 Results

The main aim of this study is to develop a methodology for diagnosing pneumonia correctly in chest X-ray images. For this purpose, eight different CNN models were trained with balanced and imbalanced datasets. According to the test results, on balanced data, ResNet-50, Xception, MobileNet, and PNet achieved AUC of 93.67, 94.23, 94.19, 93.08, respectively. On the other hand, the proposed ensemble method achieved the highest AUC of 95.21. Other evaluation criteria, namely, Acc, Sen, and Spe, were calculated

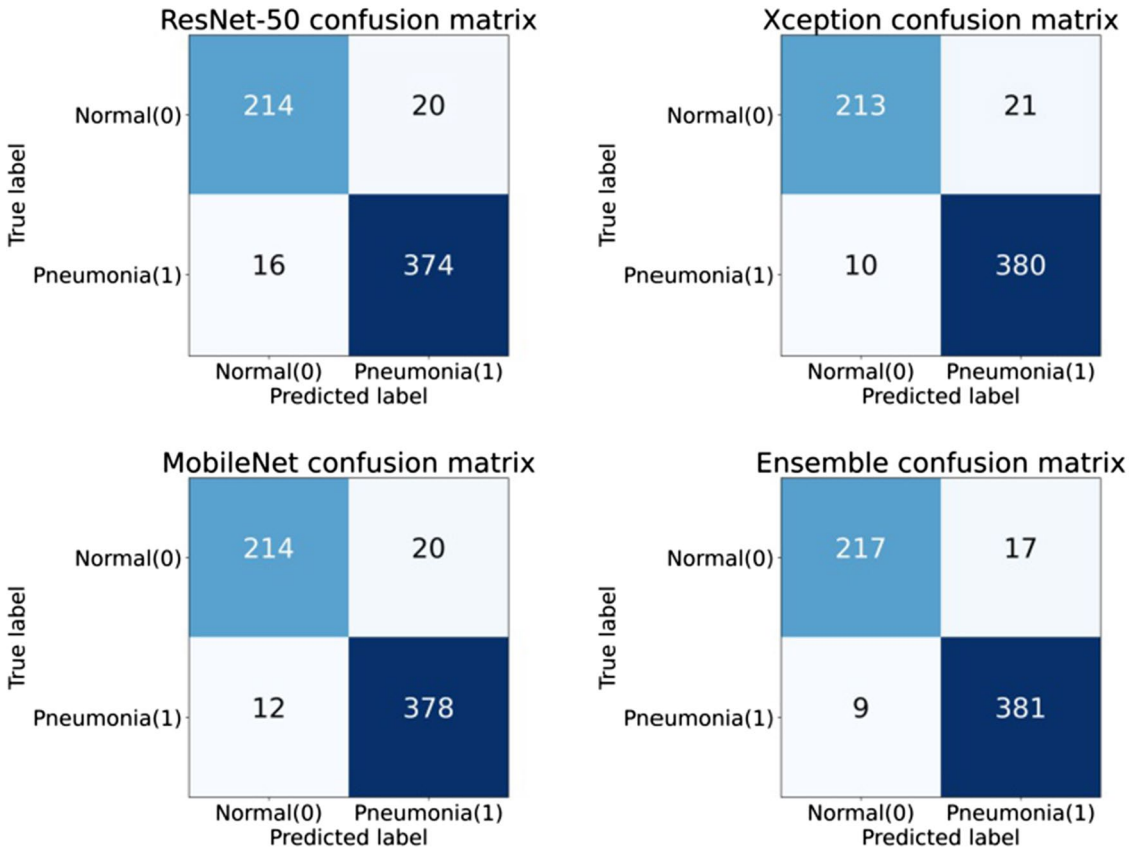

Fig. 6 Confusion matrices ResNet-50, Xception, MobileNet, and Ensemble method

Table 4 Class-based precision, recall and f1-score results (%) of ensemble method with balanced dataset

Class	Precision	Recall	f1-score
Normal	96.02	92.74	94.35
Pneumonia	95.73	97.69	96.70

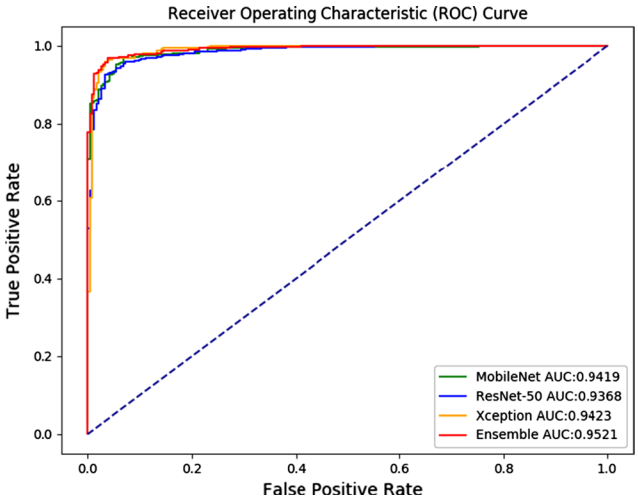

Fig. 7 ROC curves of ResNet-50, Xception, MobileNet, and ensemble method

Table 5 Classification results (%) of the proposed method and CNN models on imbalanced data

Models	Acc	Sen	Spe	AUC
Vgg-16	**92.94**	**94.35**	**90.59**	**92.48**
Vgg-19	**92.46**	**95.64**	**97.17**	**91.41**
Xception	91.98	97.73	82.90	90.17
ResNet-50	88.46	98.97	70.94	84.96
InceptionV3	90.86	94.61	84.61	89.62
MobileNet	**92.46**	**97.69**	**83.76**	**90.73**
SqueezeNet	89.42	89.74	88.88	89.32
PNet	91.98	97.17	83.33	90.26
Ensemble	**93.26**	**97.17**	**86.75**	**91.65**

Bold values indicated the 4 models that were trained with imbalanced data and got the best results in the text

and are given in Table 3. Also, the confusion matrices of the selected CNN models and proposed ensemble method are shown in Fig. 6.

In addition, we calculated class-based precision, recall and f1-score and the results are given in Table 4. Also, the ROC curves of the proposed method and selected CNN models are given in Fig. 7.

Table 6 Three class-based classification results (%) on the test data with the proposed ensemble method

Classes	Precision	Recall	f1-score	Test samples
Normal	95.56	91.88	93.60	234
Bacterial Pneumonia	88.78	97.93	92.58	242
Viral Pneumonia	88.37	77.03	82.31	148

We also trained eight CNN models with imbalanced data and examined test results by the evaluation criteria. As in the balanced dataset, we selected the three most successful models for the ensemble method. VGG-16, VGG-19, and MobileNet outputs were ensembled, and results are given in Table 5. Considering the classification results of the models using balanced and imbalanced data, it was seen that better results were obtained using balanced data. Therefore, we re-trained ResNet-50, Xception, and MobileNet models using their pre-trained weights (chest X-ray) for classification into three categories as normal, viral pneumonia, and bacterial pneumonia. Also, we applied the proposed ensemble method on pre-trained CNN models and the obtained results are given in Table 6. In addition, the confusion matrices of each CNN model and ensemble method are given in Fig. 8 along with the classification accuracy rates. We also shared PNet classification performance at three categories as normal, viral pneumonia, and bacterial pneumonia in Table 7.

6 Discussion

In this section, the proposed methodology is analyzed systematically, and its positive and negative aspects are discussed compared to other methods in the literature. The obtained results in this study were compared with other studies that achieved successful results in the literature. This comparison is given in Table 8.

Liang et al. proposed a new CNN model in their study. Instead of using pre-trained ImageNet weights, they trained their CNN model from scratch by ChestXray14 dataset. The used dataset contains 112,120 frontal chest X-ray images labeled with up to 14 different chest diseases. The CNN model trained with ChestXray14 was re-trained for the pneumonia classification problem using pre-trained weights. Thus, instead of ImageNet weights, which are less relevant to pneumonia data, ChestXray14 weights were used to solve the pneumonia classification problem. They also trained different CNN models such as VGG-16, DenseNet-121, Inception V3, and Xception using ImageNet weights with a pneumonia dataset to compare the success of the proposed transfer learning strategy. The proposed methodology in this study achieved better results in terms of Acc, Sen, Spe 95.83, 97.67, and 92.3, respectively than the CNN model

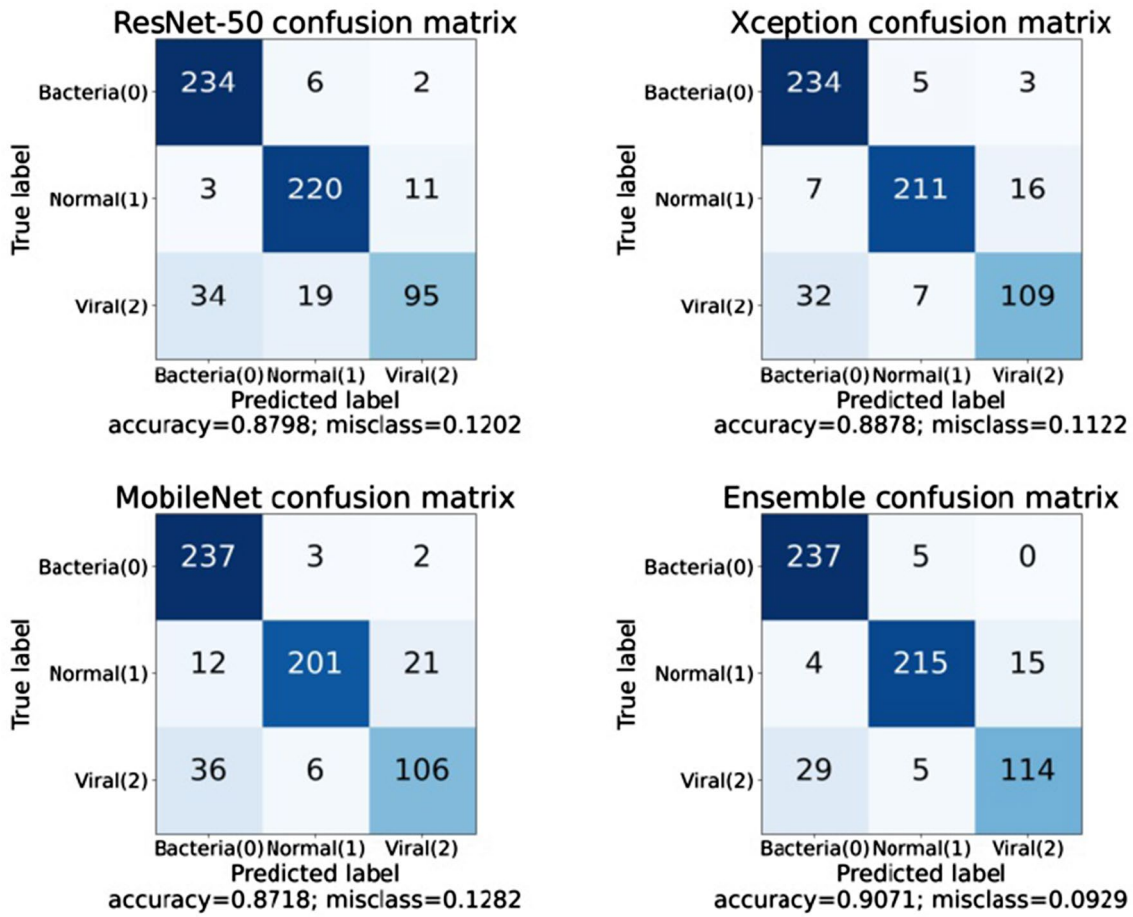

Fig. 8 Confusion matrices ResNet-50, Xception, MobileNet, and Ensemble method

Table 7 Three class-based classification results (%) on the test data with the PNet

Classes	Precision	Recall	f1-score	Test samples
Normal	89.86	91.88	90.91	234
Bacterial Pneumonia	89.45	94.63	91.97	242
Viral Pneumonia	86.05	75.00	80.14	148

proposed by Liang and Zheng. In addition, instead of the long training time in Liang and Zheng's study, it has been seen that pre-trained CNNs on ImageNet can achieve successful results with proper transfer learning and fine-tuning operations. When the confusion matrices given in Fig. 9 are examined, it is seen that the proposed methodology has achieved more successful results in both (a) balanced data and (b) unbalanced data than the study proposed by Liang and Zheng. When comparing our results with those of Kermany et al., our method obtained higher results in terms of Acc 95.83, Sen 97.76, and Spe 92.73, but the AUC value of their method is higher than ours. Siddiqi proposed an 18-layer sequential CNN model to solve the pneumonia detection problem. He used the original test set that consists of 624 chest X-ray images to evaluate the model. The proposed method in our study achieved more successful results in Acc 95.83 and Spe 92.73 criteria, while his Spe 99.0 criterion is higher than in our method. The disadvantage of the proposed model is the weak ability to detect normal case images. Chouan et al. employed five pre-trained models, namely, AlexNet, DenseNet-121, ResNet-18, Inception-V3 and GoogLeNet for pneumonia detection. They utilized transfer learning and fine-tuning when training models with chest X-ray images. They analyzed the results and proposed a voting ensemble model. Their ensemble method achieved higher results in Acc, Sen, AUC, 96.39, 99.6, 99.34, respectively. They did not share Spe criteria in the study. Rajaraman et al. trained a customized VGG-16 model using only extracted ROI lung areas instead of the whole image. Their model achieved Acc of 96.2, Spe of 99.5, and AUC of 99.0. The sensitive criterion was not shared in the study. They also classified images as normal, bacterial pneumonia, and viral pneumonia, and achieved Acc of 91.80. Their results

Table 8 Comparison of the proposed method with the literature studies (%)

References	Classes	Acc	Sen	Spe	AUC
Liang and Zheng [30]	Normal vs Pneumonia	90.5	96.7	80.3	95.3
Kermany and Goldblum [37]	Normal vs Pneumonia	92.80	93.2	90.11	96.8
Siddiqi [31]	Normal vs Pneumonia	94.39	99.0	86.0	–
Chouhan et al. [32]	Normal vs Pneumonia	96.39	99.6	–	99.34
Rajaraman et al. [28]	Normal vs Pneumonia	96.2	99.5	–	99.0
Stephen et al. [29]	Normal vs Pneumonia	93.73	–	–	–
Togacar et al. [35]	Normal vs Pneumonia	99.41	99.61	99.22	99.0
Mittal et al. [36]	Normal vs Pneumonia	95.90			
Gu et al. [33]	Bacterial Pneumonia vs Viral Pneumonia	80.40	77.55	92.67	82.34
Proposed method	Normal vs Pneumonia	95.83	97.76	92.73	95.21

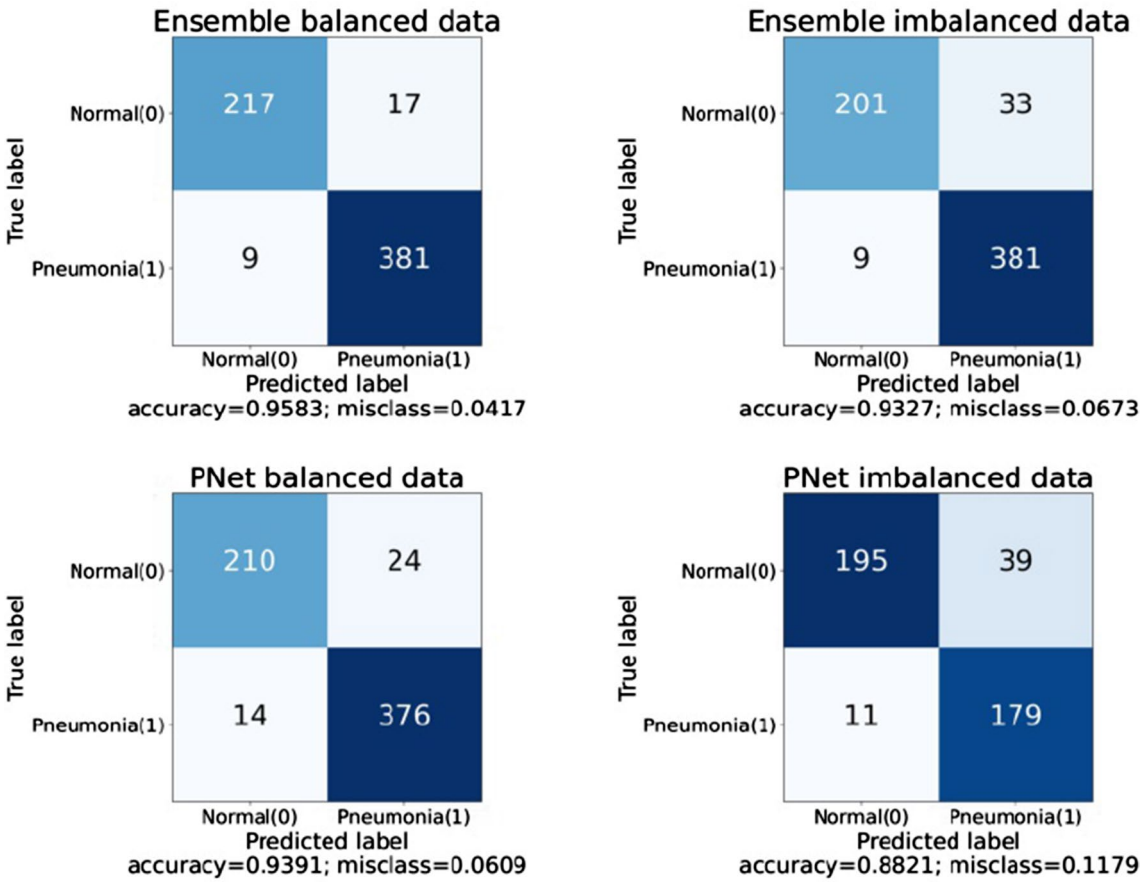

Fig. 9 Confusion matrices: ensemble networks trained by balanced data, ensemble networks trained by imbalanced data, PNet trained by balanced data and trained by imbalanced data

are slightly more successful than ours, but their methodology has a complex ROI extraction process. Stephen et al. designed a CNN model for classifying pneumonia from chest X-ray images. Their model achieved validation Acc of 93.78. No other evaluation criteria were shared in the study. Togacar et al. trained three pre-trained CNN models, namely, AlexNet, VGG-16 and VGG-19 to extract features from chest X-rays. The obtained 1,000 features from the last fully connected layers of each CNN were reduced to 100 using the mRMR feature selection method. The reduced features were given as an input to machine learning classification algorithms (decision tree, k-nearest neighbors, linear discriminant analysis, linear regression, and support vector machine) for pneumonia classification. Also, they

augmented the dataset for avoiding imbalance problems among the classes and divided the dataset as a 70% train set and 30% test set. According to the test results, the linear discriminant analysis method achieved the best result with Acc of 99.41 among the other methods. However, they used a different test sets from the original test set distribution (624) and they did not give detailed information in the study. Also, their CNN models were prone to overfitting and were not well trained on data. The proposed methodology in their study achieved more successful results than ours, but their methodology is complex, time-consuming, and requires more resources. Mittal et al. designed two CNN models named as the integration of convolutions with capsules (ICC) and the ensemble of convolutions with capsules (ECC), utilizing multi-layered capsules. Their proposed E4CC model achieved Acc of 96.36. No other evaluation criteria were shared in their study. Gu et al. proposed a two-stage method for classifying pneumonia as viral and bacterial. In the first step, they segmented left and right lung regions using an FCN model. In the second step, a DCNN model was utilized to classify lung regions. The proposed method obtained Acc of 80.40. While the proposed method remains behind some studies in the literature, it has achieved more successful results than other studies.

6.1 Examination of the Proposed Methodology

One of the aims of the proposed methodology in the study is to investigate the different generalization abilities of various CNN models on input data. This generalization ability depends on various factors such as number and type of convolution layers, type of pooling layers, activations functions, etc. The first 20 feature maps of the first convolution layer and the first 20 feature maps of the last convolution

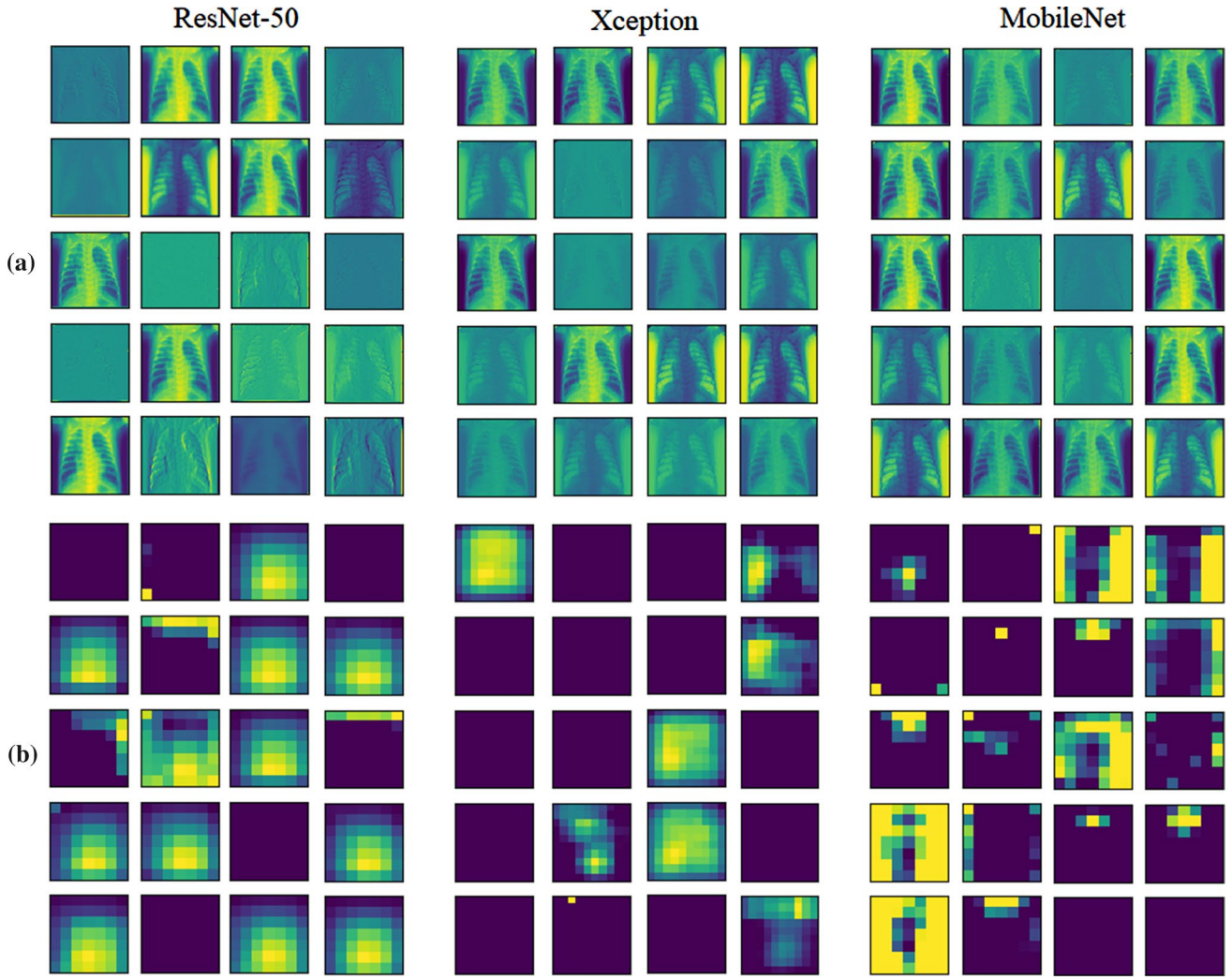

Fig. 10. 20 feature maps of the first and last convolutional layers of the selected CNN models: **a** after first convolutional layer, and **b** after last convolutional layer

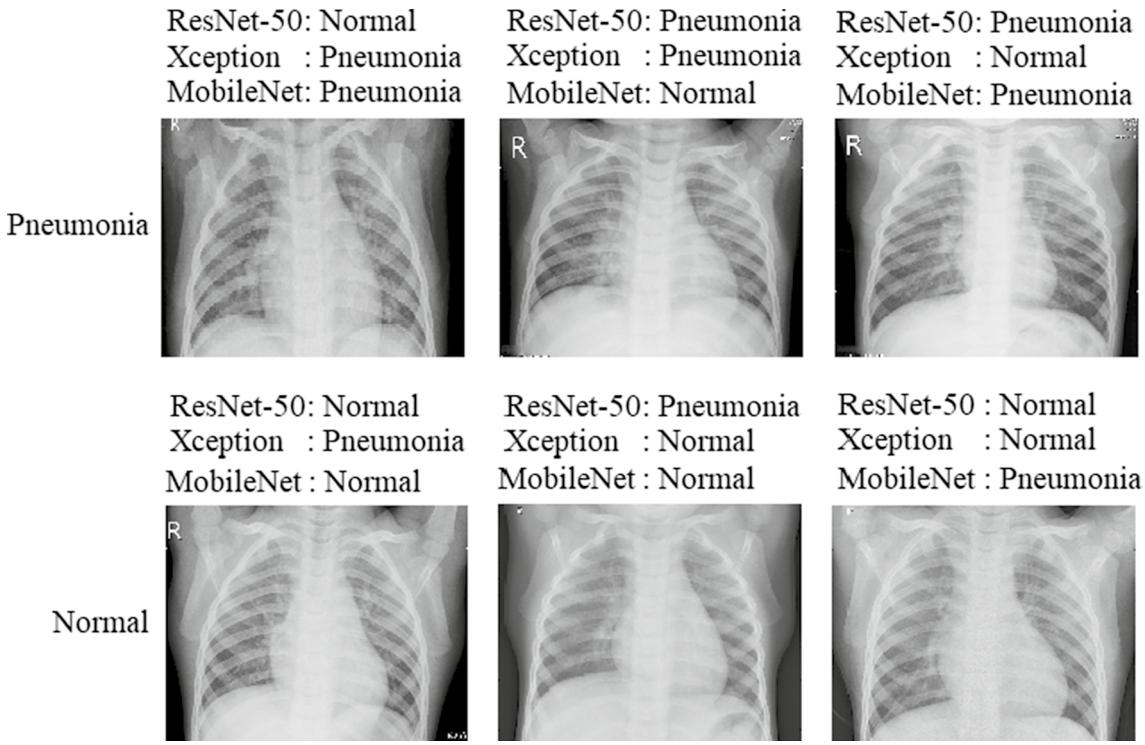

Fig. 11 Correctly diagnosed samples by virtue of the voting system

layer from the three different CNN models using the same input image are given in Fig. 10. When the feature maps of the last convolutional layer are examined, it is seen that the models obtain different features from the same input image. From this point, we proposed an ensemble method between the successful CNN models. As a result of using the ensemble method, we improved classification accuracy. In Fig. 11, pneumonia labeled and normal labeled images are given which are correctly identified thanks to the ensemble method. In addition, it is clearly seen that in Fig. 6 there is a decrease at FN value in the confusion matrix of the proposed method. Also, we classified chest X-ray images as normal, viral pneumonia, or bacterial pneumonia using the proposed methodology. The confusion matrices for each CNN model and the proposed method are given in Fig. 8. When the classification results are examined, it is seen that ResNet-50, Xception, and MobileNet achieved Acc of 87.98, 88.78, 87.18 accuracies, respectively. On the other hand, the proposed ensemble method obtained classification accuracy of 90.71. As a result, the proposed method and transfer learning strategies achieved successful results in both two way and three-way classifications.

The pneumonia images misdiagnosed by the proposed methodology are given in Fig. 12. In addition, in this study, the effects of the balanced data used in the training of CNNs on the classification results were examined and the results are presented in Tables 3 and 5. As seen in the tables, the distribution of classes in the data is very important, because CNNs learn their own features from data. In the first results, the insufficient number of normal samples reduced the success of the networks in classifying normal-labeled images. It was clearly seen that the models trained with the dataset balanced with the data obtained by the synthetic data augmentation method achieved better classification results. In addition, when the confusion matrices given in Fig. 9 are examined, the decrease in the number of FP samples is clearly seen. This situation reveals that data balance in classes is important in classification problems.

We also designed and trained a novel CNN model PNet for pneumonia detection in chest X-rays. According to our test results, the proposed ensemble method was showed better performance than a scratch-trained CNN model. Figure 9 is showing a performance comparison of proposed ensemble method and PNet on balanced and imbalanced train data. In addition, designing a CNN model needs huge experiments and knowledge according to train a pre-trained CNN model. We also designed and trained a novel CNN model (PNet) for pneumonia detection in chest X-rays. According to our test results, the proposed ensemble method was shown better performance than a scratch-trained CNN model. Figure 9 is showing a performance comparison of the proposed ensemble method and PNet. In addition, designing a CNN model needs huge experiments and domain knowledge according to train a pre-trained CNN model with transfer learning. Also,

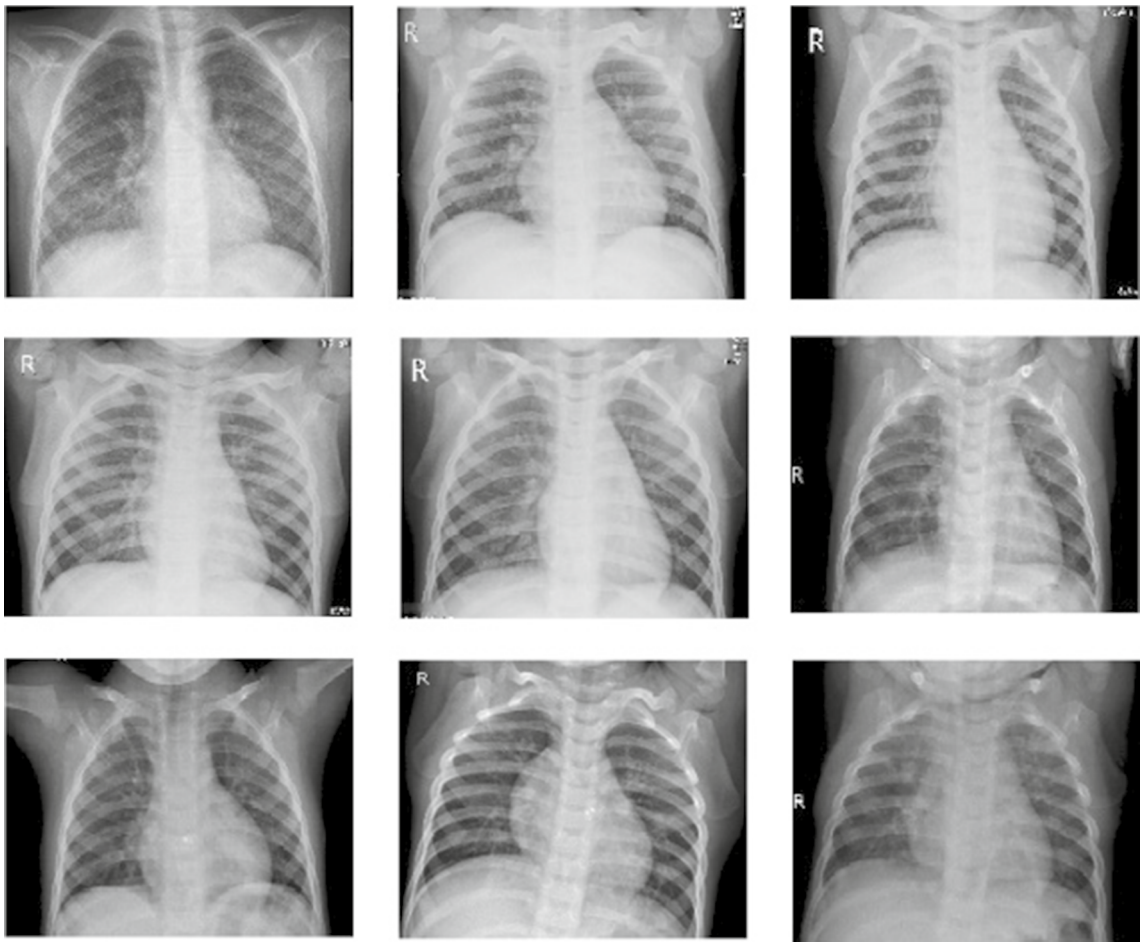

Fig. 12 Pneumonia labeled samples misdiagnosed as normal by the ensemble system

CNN models trained from scratch need more data, more training time, and more epochs to gain a better generalization ability on input data.

On the other hand, there are two limitations of our study. The first one is determining hyperparameters of pre-trained CNN models, while using them for own problem with transfer learning and fine-tuning method. Determining a suitable pre-trained CNN model for a related problem, fully connected layer size and, the number of freezing layers are important steps in transfer learning. Most of the researchers determine these parameters with the trial–error method or their experiences. As a result, determining transfer learning parameters can reveal long trial-and-error processes. The second limitation is related to bias and variance terms in ensemble learning. High variance and low bias are important dual to ensemble CNN models. In this study, we also tried group of 5 and 7 CNN models for the proposed ensemble method but the classification results were not satisfactory than group of 3.

7 Conclusions

In this study, a CNN ensemble method has been proposed that automatically classifies normal and pneumonia cases or normal, bacterial pneumonia and viral pneumonia cases in chest X-ray images. In this context, seven different CNN models VGG-16, VGG-19, ResNet-50, Inception-V3, Xception, MobileNet and, SqueezeNet were trained on the same dataset with optimal transfer learning and fine-tuning strategies. Among the trained CNN models, the three most successful ones ResNet-50, Xception and MobileNet were selected for the CNN ensemble method. The final output was obtained with this ensemble method during the test phase. ImageNet weights were used in the feature extraction phases of the models and a different number of layers was not trained in each model. A global average pooling layer was added to the end of the convolutional layers, thereby preventing the loss of spatial features in the image. In the classifier part of the models, fully connected layers were used. As a result, it was discovered that each CNN network has a different feature extraction abilities and it was observed

that the classification performance increases as a result of ensembling these abilities. On the other hand, the effect of class distribution on the classification performance was also examined and it was seen that CNNs achieve more successful results on balanced data. Further, a CNN model (PNet) was designed and trained from scratch using the same dataset. The proposed ensemble method achieved better results than PNet.

Consequently, the proposed CNN ensemble method achieved a satisfying classification performance on the chest X-ray dataset. The proposed ensemble method can be used for solving different medical problems. In future work, we plan to develop a weighted ensemble methodology based on the classification accuracy of the CNN model.

Declarations

Conflict of interest Authors declare no conflict of interest.

References

1. Wardlaw, T.M.; Johansson, E.W.; Hodge, M.J.: Pneumonia: The Forgotten Killer of Children. Unicef (2006)
2. You, D.; Jones, G.; Wardlaw, T.: Levels & Trends in Child Mortality: Report 2011. Estimates Developed by the UN Inter-Agency Group for Child Mortality Estimation. United Nations Children's Fund, New York (2011)
3. Hug, L.; Sharrow, D.; Zhong, K.; You, D.; Unicef;Organization, W.H.; Group, W.B.: Levels & Trends in Child Mortality: Report 2018, Estimates Developed by the the UN Inter-agency Group for Child Mortality Estimation. United Nations Children's Fund (2018)
4. McAllister, D.A.; Liu, L.; Shi, T.; Chu, Y.; Reed, C.; Burrows, J.; Adeloye, D.; Rudan, I.; Black, R.E.; Campbell, H.: Global, regional, and national estimates of pneumonia morbidity and mortality in children younger than 5 years between 2000 and 2015: a systematic analysis. Lancet Glob Health. **7**, e47–e57 (2019)
5. WHO: Priority diseases and reasons for inclusion. In: Chapter 6. 22-Pneumonia (2014)
6. Drake, D.E.; Cohen, A.; Cohn, J.: National hospital antibiotic timing measures for pneumonia and antibiotic overuse. Qual Manag Health Care. **16**, 113–122 (2007)
7. WHO: Standardization of Interpretation of Chest Radiographs for the Diagnosis of Pneumonia in Children. World Health Organization, Geneva (2001).
8. Ker, J.; Wang, L.; Rao, J.; Lim, T.: Deep learning applications in medical image analysis. IEEE Access. **6**, 9375–9389 (2017)
9. Neuman, M.I.; Lee, E.Y.; Bixby, S.; Diperna, S.; Hellinger, J.; Markowitz, R.; Servaes, S.; Monuteaux, M.C.; Shah, S.S.: Variability in the interpretation of chest radiographs for the diagnosis of pneumonia in children. J. Hosp. Med. **7**, 294–298 (2012)
10. Loey, M.; Smarandache, F.; M Khalifa, N.E.: Within the Lack of Chest COVID-19 X-ray dataset: a novel detection model based on GAN and deep transfer learning. Symmetry. **12**, 651 (2020)
11. Shen, D.; Wu, G.; Suk, H.-I.: Deep learning in medical image analysis. Annu Rev Biomed Eng. **19**, 221–248 (2017)
12. Grewal, M.; Srivastava, M.M.; Kumar, P.; Varadarajan, S.: Radnet: radiologist level accuracy using deep learning for hemorrhage detection in ct scans. In: IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), pp. 281–284: IEEE (2018)
13. Mazurowski, M.A.; Buda, M.; Saha, A.; Bashir, M.R.: Deep learning in radiology: An overview of the concepts and a survey of the state of the art with focus on MRI. J. Magn. Reson Imaging. **49**, 939–954 (2019)
14. Ünver, H.M.; Ayan, E.: Skin lesion segmentation in dermoscopic images with combination of YOLO and grabcut algorithm. Diagnostics. **9**, 72 (2019)
15. Esteva, A.; Kuprel, B.; Novoa, R.A.; Ko, J.; Swetter, S.M.; Blau, H.M.; Thrun, S.: Dermatologist-level classification of skin cancer with deep neural networks. Nature **542**, 115 (2017)
16. Gulshan, V.; Peng, L.; Coram, M.; Stumpe, M.C.; Wu, D.; Narayanaswamy, A.; Venugopalan, S.; Widner, K.; Madams, T.; Cuadros, J.: Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA **316**, 2402–2410 (2016)
17. Huang, P.; Park, S.; Yan, R.; Lee, J.; Chu, L.C.; Lin, C.T.; Hussien, A.; Rathmell, J.; Thomas, B.; Chen, C.: Added value of computer-aided CT image features for early lung cancer diagnosis with small pulmonary nodules: a matched case-control study. Radiology **286**, 286–295 (2017)
18. Rajpurkar, P.; Hannun, A.Y.; Haghpanahi, M.; Bourn, C.; Ng, A.Y.: Cardiologist-level arrhythmia detection with convolutional neural networks (2017). arXiv preprint https://arxiv.org/abs/1707.01836)
19. Lakhani, P.; Sundaram, B.: Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. Radiology **284**, 574–582 (2017)
20. Simonyan, K.; Zisserman, A.: Very deep convolutional networks for large-scale image recognition (2014). arXiv preprint https://arxiv.org/abs/1409.1556
21. He, K.; Zhang, X.; Ren, S.; Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
22. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z.: Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2818–2826 (2016)
23. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H.: Mobilenets: efficient convolutional neural networks for mobile vision applications (2017). arXiv preprint https://arxiv.org/abs/1704.04861)
24. Iandola, F.N.; Han, S.; Moskewicz, M.W.; Ashraf, K.; Dally, W.J., Keutzer, K.: SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size (2016). arXiv preprint https://arxiv.org/abs/1602.07360)
25. Chollet, F.: Xception: Deep learning with depthwise separable convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1251–1258 (2017)
26. Rajpurkar, P.; Irvin, J.; Zhu, K.; Yang, B.; Mehta, H.; Duan, T.; Ding, D.; Bagul, A.; Langlotz, C.; Shpanskaya, K.: Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning (2017). arXiv preprint https://arxiv.org/abs/1711.05225
27. Kermany, D.S.; Goldbaum, M.; Cai, W.; Valentim, C.C.; Liang, H.; Baxter, S.L.; McKeown, A.; Yang, G.; Wu, X.; Yan, F.: Identifying medical diagnoses and treatable diseases by image-based deep learning. Cell. 172, 1122–1131. e9 (2018)
28. Rajaraman, S.; Candemir, S.; Kim, I.; Thoma, G.; Antani, S.: Visualization and interpretation of convolutional neural network predictions in detecting pneumonia in pediatric chest radiographs. Appl Sci. **8**, 1715 (2018)
29. Stephen, O.; Sain, M.; Maduh, U.J.; Jeong, D.-U.: An efficient deep learning approach to pneumonia classification in healthcare. J. Healthc Eng. **2019** (2019)
30. Liang, G.; Zheng, L.: A transfer learning method with deep residual network for pediatric pneumonia diagnosis. Comput. Method Prog. Bio.104964 (2019)

31. 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)
32. Chouhan, V.; Singh, S.K.; Khamparia, A.; Gupta, D.; Tiwari, P.; Moreira, C.; Damaševičius, R.; de Albuquerque, V.H.C.: A novel transfer learning based approach for pneumonia detection in chest X-ray images. Appl Sci. **10**, 559 (2020)
33. Gu, X.; Pan, L.; Liang, H.; Yang, R.: Classification of bacterial and viral childhood pneumonia using deep learning in chest radiography. In Proceedings of the 3rd International Conference on Multimedia and Image Processing, pp. 88–93 (2018)
34. Rahman, T.; Chowdhury, M.E.; Khandakar, A.; Islam, K.R.; Islam, K.F.; Mahbub, Z.B.; Kadir, M.A.; Kashem, S.: Transfer Learning with Deep Convolutional Neural Network (CNN) for Pneumonia Detection using Chest X-ray. Appl Sci. **10**, 3233 (2020)
35. Toğaçar, M.; Ergen, B.; Cömert, Z.; Özyurt, F.: A deep feature learning model for pneumonia detection applying a combination of mRMR feature selection and machine learning models. IRBM. **41**, 212–222 (2020)
36. Mittal, A.; Kumar, D.; Mittal, M.; Saba, T.; Abunadi, I.; Rehman, A.; Roy, S.: Detecting pneumonia using convolutions and dynamic capsule routing for chest X-ray Images. Sensors. **20**, 1068 (2020)
37. Kermany, D.; Goldbaum, M.: Labeled optical coherence tomography (OCT) and Chest X-Ray images for classification. Mendeley Data. **2** (2018)
38. Shorten, C.; Khoshgoftaar, T.M.: A survey on image data augmentation for deep learning. J. Big Data. **6**, 60 (2019)
39. Ayan, E.; Ünver, H.M.: Data augmentation importance for classification of skin lesions via deep learning. In: Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT), IEEE, pp. 1–4 (2018)
40. Perez L.; Wang, J.: The effectiveness of data augmentation in image classification using deep learning (2017). arXiv preprint https://arxiv.org/abs/1712.04621)
41. Weiss, K.; Khoshgoftaar, T.M.; Wang, D.: A survey of transfer learning. J. Big Data. **3**, 9 (2016)
42. Deng, J.; Dong, W.; Socher, R.; Li, L.-J.; Li, K., Fei-Fei, L.: Imagenet: a large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255, IEEE (2009)
43. Voulodimos, A.; Doulamis, N.; Doulamis, A.; Protopapadakis, E.: Deep learning for computer vision: a brief review. Comput. Intel. Neurosci. **2018**, 14 (2018)
44. Rawat, W.; Wang, Z.: Deep convolutional neural networks for image classification: a comprehensive review. Neural Comput. **29**, 2352–2449 (2017)
45. Yosinski, J.;Clune, J.;Bengio, Y., and Lipson, H.: How transferable are features in deep neural networks? In: Advances in Neural Information Processing Systems, pp. 3320–3328 (2014)
46. Koushik, J.: Understanding convolutional neural networks (2016). arXiv preprint https://arxiv.org/abs/1605.09081)
47. Hu, J.;Shen, L.; Sun, G.: Squeeze-and-excitation networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7132–7141 (2018)
48. Harangi, B.: Skin lesion classification with ensembles of deep convolutional neural networks. J. Biomed. Inform. **86**, 25–32 (2018)