Kütahya Dumlupınar University
Institute of Graduate Studies

Journal of Scientific
Reports-A
E-ISSN: 2687-6167
Number 52, March 2023

RESEARCH ARTICLE

DETECTION OF PNEUMONIA FROM X-RAY IMAGES USING DEEP LEARNING TECHNIQUES

Halit BAKIR[1*], Semih OKTAY[2], Timuçin Emre TABARU[3,4,5]

[1*]Computer Engineering Department, Sivas University of Science and Technology, Sivas, halit.bakir@sivas.edu.tr, ORCID: 0000-0003-3327-2822
[2]Sivas University of Science and Technology, Defense Technologies Department, Sivas, oktaysemih@outlook.com, ORCID: 0000-0002-7426-5584
[3]Department of Electrical-Electronic Engineering, Sivas University of Science and Technology, Sivas, etabaru@sivas.edu.tr, ORCID: 0000-0002-1373-3620
[4]UNAM-Institute of Materials Science and Nanotechnology, Bilkent University, Ankara 06800, Turkey
[5]Department of Clinical Engineering Research and Application Center, Erciyes University, Kayseri 38039, Turkey

Receive Date:15.12.2022 *Accepted Date: 28.03.2023*

ABSTRACT

X-ray images is one of the most common utilities used by health care specialists for detecting healthy problems in patients' chest. In this work, deep learning techniques have been adopted for diagnosing and detecting of lung diseases. First, an experimental study has been conducted for selecting the best artificial neural network ANN model that can be used for lung X-Ray image classification. The obtained best model has been used for classifying the lung X-Ray images into three classes (Multi class classification) namely bacterial pneumonia, viral pneumonia, and healthy lung. After that, three well-known CNN architectures, namely ResNet, Inception, and MobileNet have been adopted and used as a feature extractor for the selected best ANN model. Moreover, the above-mentioned ANN model (both with and without the features extraction phase) has been used for classifying the lung X-Ray images as healthy and pneumonia lungs (Binary classification). As a result of the study, the proposed ANN model with ResNet feature extraction phase gave the highest classification accuracy rate of 81.67% when multi-class classification has been conducted on the lung X-Ray dataset. On the other hand, the proposed ANN model with MobileNet feature extraction phase gave the highest accuracy rate of 95.67% when a binary classification has been conducted on the X-Ray image dataset.

Keywords: *X-Ray, Artificial Neural Networks, Pneumonia; Pre-Trained Models, Deep Learning*.

1. INTRODUCTION

Today, X-Ray images are an effective tool used in the health sector for detecting many health problems. One of the most important advantages of X-ray images is that it helps to analyse the lungs by imaging pneumonia in them. The COVID-19 epidemic, which has exploded in recent years, has

JSR A
Journal of Scientific Reports

shown that one of undoubtedly the most important issues in this field is the detection of the disease earlier and at a lower cost. However, it is difficult to detect pneumonia on X-Ray images, as well as to decide whether the detected pneumonia is bacterial or viral. It is a problematic situation that takes a long time for radiologists and remains in the middle of great hesitations about making the correct diagnosis. On X-Ray images, the part of the pneumonia is usually not conspicuously visible when viewed. It may even resemble images of other diagnoses and may overlap at some points. At the same time, it may appear as benign abnormalities that do not damage the lungs on the images, leading to a misdiagnosis. However, what is desired in the health sector is to make the most accurate diagnosis in a fast time. Deep learning techniques that emerged with the development of technology have been included in our lives and have become applicable in almost every sector. Deep learning is a machine learning technique that is effectively used in many problems and applications these days. In other words, deep learning is a branch of machine learning used in multiple domains such as natural language processing, computer vision, and voice recognition. The most striking feature of deep learning compared to previously used techniques is that it does not explicitly require a feature extraction step. The deep learning model receives unprocessed input to determine which parts of the model are valuable and will affect the result. Then it maps the predicted output to the desired output. Features are automatically extracted by the deep learning model without any manual intervention from the outside. Although the deep learning models that have been used and created recently within this technology require high levels of processing performance, this situation is no longer considered a problem with the increase in processor power in the 21st century. Deep learning models are a combination of traditional neural networks with an increasing number of hidden layers. Deep learning techniques can automatically reveal image features that contain the image's most distinctive information. There is no human intervention or manual feature determination in the extraction of features in these types of architectures. These techniques have recently been widely used in image recognition and image classification processes [5]. Deep learning techniques have been used to serve disease detection purposes as well. Multiple models have been proposed using this technology and adopted in various domains and obtained aver high results. In fact, these methods have become able to reduce the workload of radiologists and doctors working in the health sector and early detect the type of disease accurately. In this context, the use of deep learning techniques to detect and classify the disease from chest X-Ray images can be shown as an important development in terms of service to human health. Many research, articles, and applications have been conducted on this subject, and the detection accuracy results have increased and the detection time have decreased day by day. However, deep learning methods are still considered as an area for improvement as they give different accuracy results depending on the dataset and hyperparameters used in the models. Therefore, different methods, models and hyperparameters can be adjusted experimentally, and the best results can be obtained by performing performance analyse.

2. RELATED WORKS

In Ayan et al.'s study, two well-known CNN architectures, namely Xception and VGG16, have been used for classifying lung X-Ray images as normal and pneumonia cases [1]. At the same time, the performance of both architectures has been compared using different metrics. As a result of the research, it has been stated that the Xception architecture was better at detecting pneumonia cases, while the VGG16 architecture was more successful in detecting normal cases. Rajpurkar et al. [2], a deep CNN model called CheXNet has been proposed to detect pneumonia from lung X-Ray images. It is stated that the proposed model is trained with one of the largest publicly available datasets, called ChestX-Ray14. As a result, it was observed that the CheXNet model obtained F1 score, exceeding the average radiologist's performance. In [3], [4], a dataset composed of 5856 computer tomography (CT) images and X-ray images has been used. The dataset includes 4273 pneumonia images and 1583 healthy images. Multiple well-known CNN architectures namely VGG16, Inception_V3, VGG19, Inception_ResNet_V2, DenseNet201, MobileNet_V2, Resnet50, and Xception were adopted to classify the used dataset's images. As a result of the study, while CNN architectures such as MobileNet_V2, Resnet50, and Inception_Resnet_V2 showed high performance, VGG16, VGG19, Inception_V3, Xception, and DenseNet201 CNN architectures were partially successful. The model that gave the highest accuracy rate was Resnet50 with an accuracy of 96.61%. In [4], 6 different models have been created for the detection of pneumonia from X-ray images. Two of these models consist of two and three convolution layers and have been named model 1 and model 2 in the study, while the others are the pre-trained VGG16, ResNet50, VGG19, and Inception-v3 models. At the end of the study, performance comparisons of these models have been conducted, and in the results, the model 1 and model 2 obtained a validation accuracy reached 85.26% and 92.31%, respectively. The accuracy of the other models i.e. VGG16, ResNet50, VGG19, and Inception-v3 were 87.28%, 77.56%, 88.46%, and 70.99%, respectively. In the study of Kamrul et al. [5], a VGG16 CNN architecture has been used to detect a viral infection called COVID-19. It has been stated that, this architecture can detect pneumonia with a high accuracy reached 91.69% and sensitivity of 95.92%. In [6], a two-class classification model has been proposed to detect pneumonia. Instead of using different trained architectures as in other studies, it has been shown that creating a CNN network from scratch gives an effective accuracy result. The results obtained from the proposed model showed that the test accuracy was 95.31%, and the validation accuracy was 93.73%. In [7], four different models' architectures have been compared to design a model that gives better results than the models in the literature. Two of these models are the pre-trained i.e. ResNet152V2 and MobileNetV2 models, while one of the other two models is a CNN-based model and the remaining model is the LSTM-CNN-based model. As a result of the research, the most successful results have been obtained using the ResNet152V2 with 99.22% of accuracy, 99.43% of Precision and 99.44% of F1 score rate. Also, it has been stated that approximately 91% of successful results have been obtained from the other adopted models. In [8], it has been aimed to define a model to identify and localize the pneumonia from the chest X-Ray images. The proposed model is based on Mask-RCNN, a deep neural network that combines global and local features for pixel-by-pixel segmentation. It was observed that the presence of pneumonia was not evident in the images, and the distinction between the two conditions was rather vague. In addition, it has been inferred that the larger image may be more useful for deeper information. However, it is stated that the computational cost will also increase exponentially when dealing with large image. Finally, data augmentation, dropout and L2 regularization techniques have been used in order to prevent the overfitting problem. In [9], it has been aimed to propose a model that could classify the data set consisting of four different X-Ray image classes, including coronavirus X-Ray images. To this end, seven CNN architectures namely VGG16, DenseNet201, VGG19, Inception_ResNet_V2, Resnet50, Inception_V3, and MobileNet_V2 have been adopted and compared in order to find the best model. It has been concluded that the Inception_ResNet_V2 model gave a better result than the other models with an accuracy rate of 92.18%. In [10], a CNN-based model called CVDNet model has been proposed for faster detection of patients with COVID-19 virus from a dataset containing X-Ray images of viral pneumonia, COVID-19, and healthy people. The proposed CVDNet model has been trained using a small dataset containing 2905 X-Ray images, and the model's accuracy rate was 96.69%. In [11], five CNN architectures, namely ResNet50, ResNet152, ResNet101, Inception_ResNetV2, and InceptionV3 have been tested to select the best model for detecting COVID-19 cases from three X-Ray datasets. By investigating the results of the study, it is seen that the ResNet50 pre-trained architecture gives the highest accuracy among the tested five architectures for the used three different datasets. In [12], a model that can detect and classify COVID-19 cases has been proposed. In the proposed model, an CNN-based architecture called DarkCovidNet, inspired by the DarkNet architecture, has been adopted. Less layers and filters have been used in the proposed architecture compared to the original DarkNet architecture. It has been stated that the proposed architecture was able to perform multi-class and binary classification tasks with an accuracy of 87.02% and 98.08%, respectively. In [13], the use of ResNet model variants were proposed for differentiating COVID-19 pneumonia from bacterial, other viral pneumonia, and healthy cases in X-Ray. First, X-Ray images have been classified into three different classes using the ResNet50 architecture: healthy cases, bacterial and viral pneumonia. Later, all cases classified as viral pneumonia have been differentiated as COVID-19 induced pneumonia and other viral pneumonia using the ResNet101 architecture. As a result of the research, a success rate of 93.01% has been achieved in the first stage of the model, while a very high success rate of 97.22% has been obtained in the second stage. In [14], two different deep learning and machine learning-based models have been proposed for conducting multi-class and binary classification over the used the used X-Ray image dataset. Also, SMOTE algorithm, which is one of the oversampling methods used to make the image distribution in the classes equal and to overcome the imbalance data problems, has been used for balancing the dataset. Some metrics such as accuracy, precision, recall, and F1 score were used to evaluate the proposed model. As a result, a 95% of accuracy rate for binary classification has been achieved from the proposed two models. For multiclass classification, the average values for precision, recall and F1 score of the CNN and Ensemble models were 80%, 78%, 78% and 77%, 75%, and 75%, respectively. Nayak et al. [15] proposed a deep learning-assisted automatic COVID-19 detecting method using lung X-Ray images. They used eight different pre-trained CNN network models and compared the accuracy results of the models to find the best result. Some well-known CNN pre-trained models such as SqueezeNet, ResNet, GoogleNet, etc. have been adopted for conducting classification tasks in the study. As a result of the research, ResNet-34 outperformed other architectures with 98.33% of accuracy. In the study of Apostolopoulos et al., [16] a dataset composed of 3905 X-ray images has been used. First, the images have been divided into 2 classes namely COVID-19 and non-COVID-19, then, the same dataset has been divided into seven classes namely COVID-19, enema, effusion, emphysema, fibrosis, pneumonia, and normal. After that, MobileNet v2 model has been used as a feature extractor, and three different experimental studies have been carried out. As a result, the highest accuracy rate was 99.18% which obtained by applying binary classification, and 87.66% obtained by conducting multi-class classification. In the study of Bhardwaj et al. [17], chest X-ray image dataset includes 2161 COVID-19, 2022 pneumonia, and 5863 healthy people has been used. A pre-processing phase has been conducted including image normalization and contrast enhancement for improving the quality of the images. The authors adopted multiple well-known CNN architectures for classifying the images in the dataset. As a result of the experiments, high accuracy rates reached 98.33% obtained by applying binary classification and 92.36% obtained by conducting multi-classification have been observed. In [18], some CNN pre-trained model has been adopted as a feature extraction phase for a deep learning model used for diagnosis of cataract disease in the eye.

To our literature review, all studies conducted till now have used pre-trained models to diagnose chest diseases from X-Ray images. Classical neural networks (ANN) have not been tested for this issue in any of the reviewed previous studies. Therefore, in this work, an artificial neural network model were proposed and adopted to detect chest diseases based on X-ray images. Also, some pre-trained CNN models were used as feature extractor for the proposed ANN model to improve its results. Particularly, this work contains the following contributions:

- An artificial neural network model was proposed for classifying the lung images as a binary and multi-class classification.
- Multiple experiments have been conducted in order to select the best structure for the proposed ANN model.
- Three well-known CNN architectures, including ResNet, Inception, and MobileNet, have been adopted to be used as a feature extractor for the proposed ANN model.
- The proposed ANN architecture has been tested in terms both of binary and multi-class classification and the classification accuracy reached 81.67% and 95.67% for multi-class classification and binary classification respectively.

3. MATERIAL AND METHOD

3.1. Used Dataset

In this study, chest X-Ray images obtained from the popular database Kaggle's "Chest X-Ray Images (Pneumonia) with new class" radiography database [19] has been used. This dataset contains three different classes of chest X-Ray images including people with bacterial pneumonia, people with viral pneumonia (pneumonia), and healthy (normal) people. There is a total of 4479 lung X-Ray images in the dataset including 1493 bacterial pneumonia images, 1493 viral pneumonia images, and 1493 normal images. The dataset has been divided into 80% as a training dataset, 10% as a validation dataset, and 10% as a testing dataset. Figure 1 shows some sample images from the used dataset.

3.2. Method

In this study, performance analysis for deep learning techniques has been performed and different models have been tested in order to select the best model that can classify chest X-Ray images as viral pneumonia, bacterial pneumonia, or healthy with high performance and as low computational overhead as possible. An artificial neural network model has been proposed to classify chest X-Ray images in the used dataset. In addition, three pre-trained CNN models, namely ResNet, Inception, and MobileNet, have been used as a feature extractor to improve the results of the proposed ANN model.

3.2.1. Artificial neural network (ANN)

Artificial neural networks are systems consisting of processing elements connected to each other with different weight coefficients, which have been proposed to mimic the information exchange between the nerve cells in the human nervous system. Feed forward-back propagation ANN model is the most used method among the artificial neural network (ANN) techniques, where, this model can work with the logic of backward propagation of errors. Generally, ANN models are composed of five main components namely an input layer, weight coefficients, activation function, hidden layers, and output layer. A neural network contains three-layer is shown schematically in Figure 2 [20].

3.2.2. Convolutional neural network

Convolutional neural network (CNN) is a type of NN that is commonly applied in computer vision [21] and natural language processing [22] domains. Generally, CNN can contains three different type of layers, namely input layer, one or more hidden layer, and finally an output layer. The hidden layer section can contain five different layer types including convolutional layers, activation function layers, pool layers, fully connected layers and normalization layers. The convolution operation conducted using multiple filters can be used for extracting features (feature map) from the dataset, which can be used as input for the next layers. The Pooling layer, also known as down-sampling layer, is used to reduce the size of feature maps such that the total computational time of the model can be reduced. MaxPooling and average pooling are the most popular used pooling operations. If we compare it with other classification algorithms and look at its advantages, CNN requires much less pre-processing and can give more successful results as the number of samples in the training dataset increases.

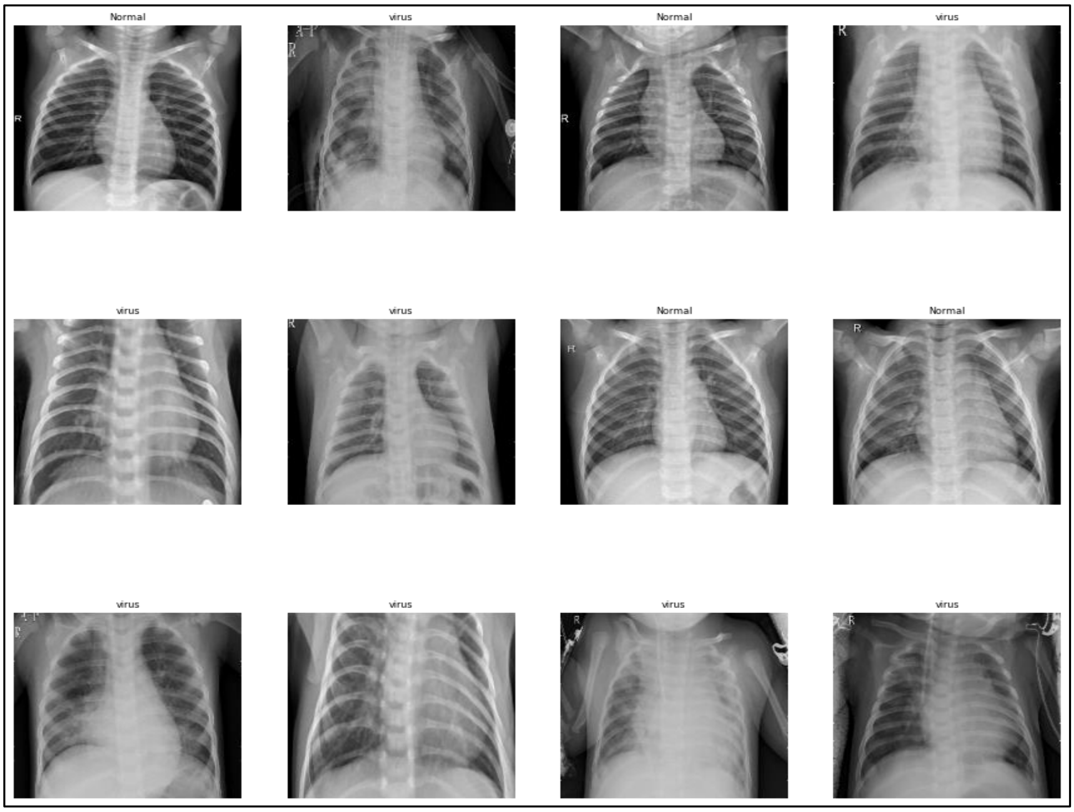


Figure 1. Some examples of X-Ray images used.

JSR A
Journal of Scientific Reports

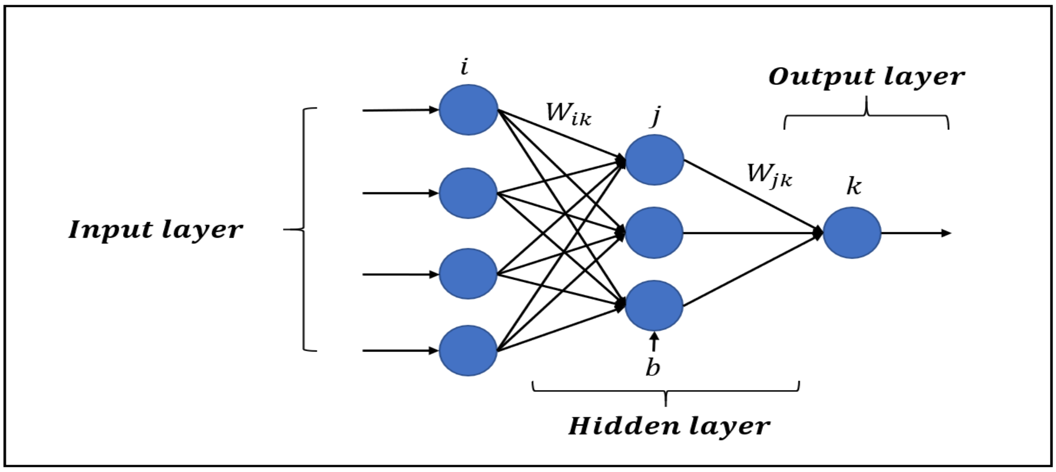


Figure 2. General representation of artificial neural network.

3.2.3. ResNet model

ResNet 2015 [23], short for Residual Network, is a CNN architecture introduced to solve the most complex problems. Some additional layers have been proposed and added in order to improve the performance of the deep neural networks. Particularly, skip connection has been proposed in order to skip un-useful or not used layers in order to train a very deep structure without any overfitting problem. The reason behind adding more layers is for those layers to learn more and more complex features. Figure 3 shows the difference between the block used in the ResNet architecture and the block used in the normal CNN architecture.

JSR A
Journal of Scientific Reports

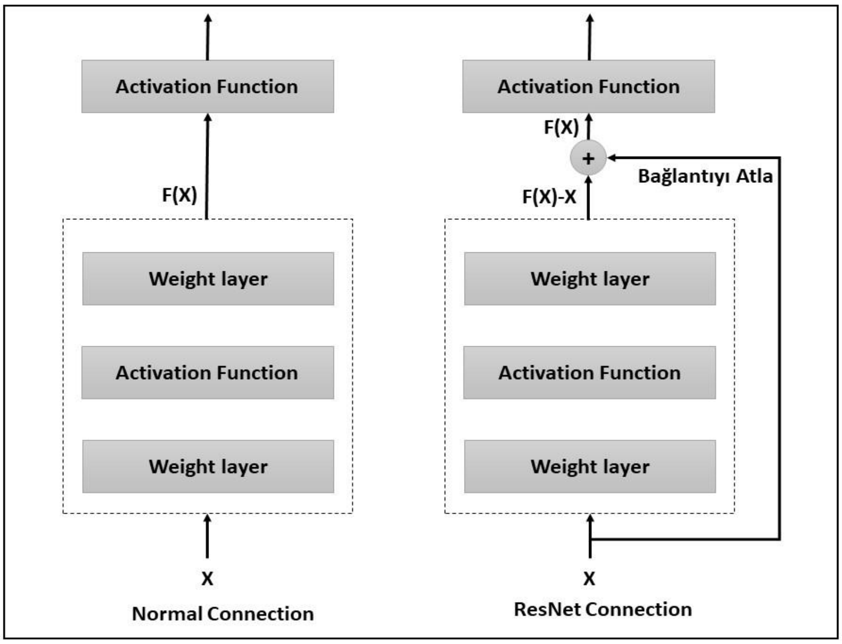


Figure 3. The difference between the block used in the ResNet architecture and the block used in the normal CNN architecture.

3.2.4. Inception net

Inception Network was once considered a cutting-edge deep learning architecture to solve image recognition and detection problems and was proposed in the 2014 article titled as "Going deeper with convolutions" [24]. The article proposes a new type of architecture called GoogleNet or Inception Network. It is basically a CNN architecture with a depth of 27 initial layers or filter blocks. Each initial layer is composed of a combination of convolutional layers (i.e. a layer containing a various filter sizes) with a single output vector that forms the input of the next layers. Figure 4 shows the filter logic used in the Inception architecture.

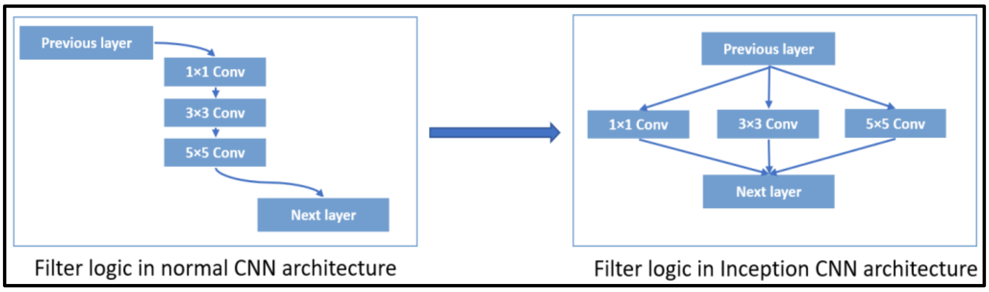


Figure 4. Filter logic used in the Inception architecture.

3.2.5. MobileNet model

MobileNet V3 is a combination of hardware-aware architectures initially proposed to be used in limited recourse platforms such as mobile phones but it got more popular, and it is used in almost all platforms nowadays [25]. MobileNet architecture is based on different types of convolutional layers different from the classical one called Depthwise Separable convolution. The convolutional layers used in the MobileNet structure are composed of two steps namely Depthwise convolution and Pointwise convolution. Figure 5 shows the difference between the convolution operation used in the MobileNet architecture and the normal convolution operation.

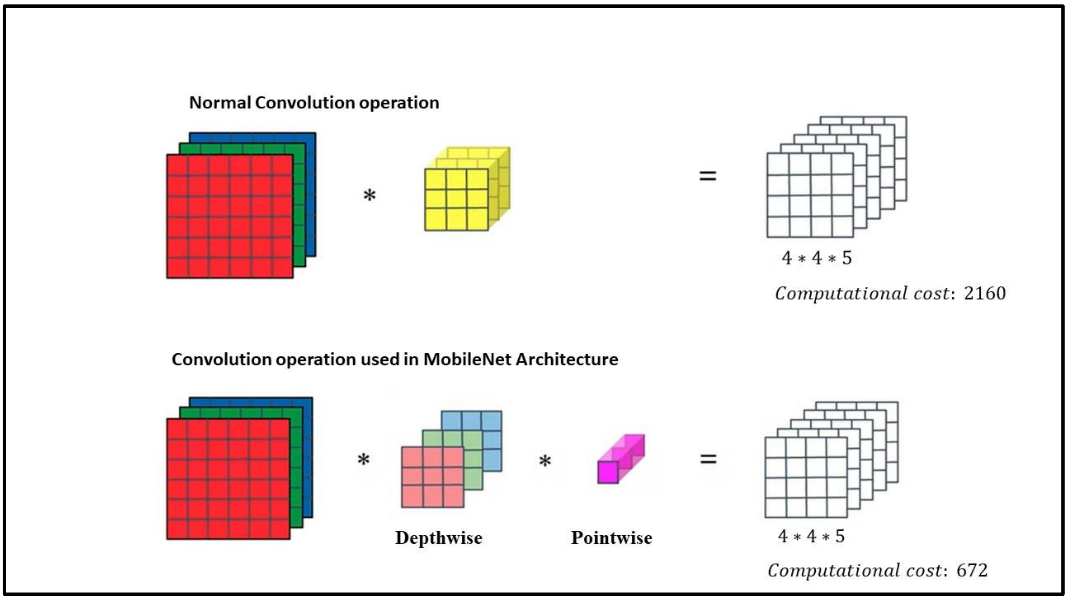


Figure 5. The difference between the convolution operation used in the MobileNet architecture and the normal convolution operation.

3.2.6. Proposed model

In this study, multiple neural network architectures, multiple CNN architectures and pre-trained CNN models have been tested and multiple experiments have been conducted, it has been observed that the neural network architecture can give better results for this problem. Therefore, in this study, a classical artificial neural network-based model has been adopted, and multiple pre-trained CNN models have been tested as a feature extraction phase for the proposed ANN model. First, we started with a shallow neural network architecture. Afterward, the number of layers used in the model and other hyperparameters have been changed and the model that gave the best result has been selected. Particularly, different values for regularization techniques such as L1 and L3, kernel initializers, activation functions, loss functions, batch normalization have been applied and tested in order to increase the performance and reduce the overfitting problem. In addition to this, all other hyperparameters have been manipulated with different values, and at the end, the model that gave the most suitable and the best results for the research problem has been selected.

First, the images in the dataset have been resized to 180 x 180 size before being used for training the proposed ANN model. Then, the pixel values in the image have been normalized to be between 0 and 1. Afterward, the used dataset has been divided into 80% for training, 10% for testing, and 10% for validating the proposed model. In the next step, the two values for "batch size" has been tested i.e. 32 and 64, but when the batch size was 64, the model gave a better result, so the batch size of 64 has been used. Also, the data samples in each batch have been configured to be randomly selected when entered into the model. The same logic has been applied to the validation dataset. Therefore, the model was validated using different data samples in each epoch. This means that the model is trained using different data and validated using different data in each epoch. This prevents the model from falling into overfitting to some extent. In the experiments conducted to obtain the best performance, it was aimed to determine the number of layers that gave the best performance and the number of neurons in each layer. As a result of multiple experiments, when we reached a neural network model consisting of 7 layers as in Table 1, the best results have been obtained.

The “ReLu” activation function has been used in the hidden layers and the “softmax” activation function has been used in the output layer since the addressed problem is multi-class classification problem. “Adam” has been used as an optimizer and the “learning rate” value has been selected as 0.001 after trying a range of values. Moreover, "sparse categorical crossentropy" has been used as a loss function. Also, "he_uniform" has been used as a kernel initializer to optimize the performance of the model.

Table 1. Structure of the proposed ANN model.

Layers Number	Neurons Number
1	512
2	256
3	128
4	128
5	64
6	32
7	3

After selecting the best hyperparameter values of the ANN model that gives the best results, it is suggested to use some well-known pre-trained CNN architectures as a feature extraction phase to further improve the results of the proposed model. In particular, three well-known pre-trained architectures, namely ResNet, Inception, and MobileNet, have been tested. Figure 6 shows the final version of the proposed model.

4. EXPERIMENTAL RESULTS

The models proposed in this study has been implemented using Jupyter notebook and python programming language on a laptop with Intel(R) Core (TM) i7-9750H CPU @ 2.60GHz 2.59 GHz processor, NVIDIA GeForce GTX 1650 graphics card and 8 GB of memory. Multiple Python libraries such as Tensorflow, Tensorflow_hub, and Keras have been adopted to implement the proposed model.

First of all, some pre-processing operations have been applied to the images in the dataset, for example, all the images in the dataset have been resized to 180×180 and their pixel values have been rescaled to be between 0 and 1. We have conducted multiple experiments to select the best model that can achieve this task with the best score. For example, a different number of dense layers such as 2, 3, 4, 5, and 6 have been tested, and it is concluded that the model with 6 layers can give the best results. Also, we have tested 32, 64, 128,256, and 512 values for the number of units in each dense layer, 32 and 64 values for batch size, and 0.1, 0.01, and 0.001 values for the learning rate, uniform, normal, glorot_uniform, he_normal, he_uniform values for the weight initializer, and tanh and relu for the activation function used in the model's layers. As mentioned previously mentioned the model gave the best results contains six dense layers with relu activation function and he_uniform weight initializer in each layer.

JSR A
Journal of Scientific Reports

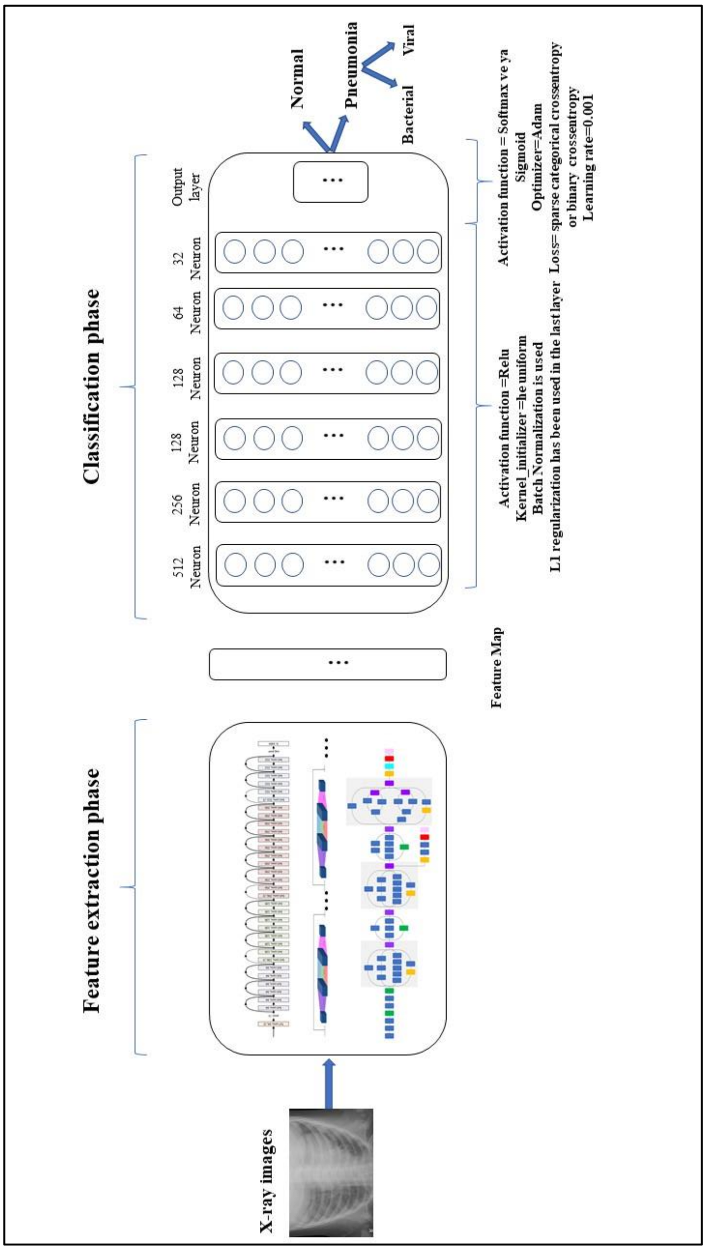


Figure 6. The proposed neural network model.

Also, we choose to use the Adam optimizer with a learning rate of 0.001 and 50 epochs to be used during training of the ANN model. Furthermore, the softmax or Sigmoid activation function has been adopted in the output layer for conducting the classification task depending on whether the task is a binary or multi-class classification. As a result of various trials, it has been observed that the model was exposed to a certain level of overfitting as shown clearly in the top section of figure 7. a. It can be noted from the figure that while the training loss decreased in accordance with the epoch number, it was not the case for the validation loss, where the validation loss increased with the increase in the epoch number. Also, we can note from the bottom section in figure 7. a, that while the training accuracy of the model increased in accordance with the increase in the number of epochs the validation accuracy was almost constant. No doubt it can be said that there is an overfitting problem in the model. So, various regularization techniques have been tested to overcome this problem. First, the "Batch Normalization" regularization technique has been applied between each hidden layer, but sufficient results could not be obtained. Afterward, the "dropout" technique has been applied, but this technique had a negative effect on the model's performance. At the same time, since the loss value of the model got close to zero after a certain number of epochs, the training of the model slowed down and almost stopped. In order to solve this problem, the "L1" regularization technique has been tested. When we look at the results after the application of L1 regularization, we noted that the accuracy rates obtained from the experiments performed on the test dataset improved as can be seen in Table 2. Also, figure 7. a show the changes in the loss and accuracy values during training and validating of the model before using the L1 regularization technique. It can be clearly seen from the figure that while the training loss decreases as the epoch progresses, the validation loss increases as the epoch progresses. Figure 7.b shows the changes in the loss and accuracy values during training and validating the model after using the L1 regularization technique. It can be observed from Figure 7.b that this problem has disappeared to a good extent. Particularly, by looking at the figure we can observe that each of the training loss and the validation loss decreased in accordance with the increase in epoch number. Also, the training accuracy and validation accuracy increased in accordance with the increase in the number of epochs. So, we could mitigate the overfitting problem to some extent. The highest accuracy rate obtained by applying the pure artificial neural network model for multi-class classification was 78.33%.

After, that we proposed using some well-known CNN architectures as a feature extraction phase to improve the results of the proposed neural network model. To this end, three pre-trained CNN models, namely ResNet, Inception, and MobileNet, have been used. The feature maps extracted using these models have been used to train the proposed neural network model.

First, the ResNet model were used as a feature extractor for the proposed ANN model. Since this model does not face the problem of overfitting like other CNN architectures, thanks to its skip connection logic, the "L1" regularization technique has not been applied. After training this model, the accuracy rate on the test dataset reached 81.67%. Then, the Inception structure were used as a feature extractor for the proposed neural network model. It has been seen that the accuracy rate of the trained model on the test dataset dropped down to 75% and negatively affected the original model.

Finally, the MobileNet architecture were used as a feature extractor for the proposed artificial neural network model. In this experiment, the performance of the model reached 78.67% and a slight improvement has been noted over the original model's performance.

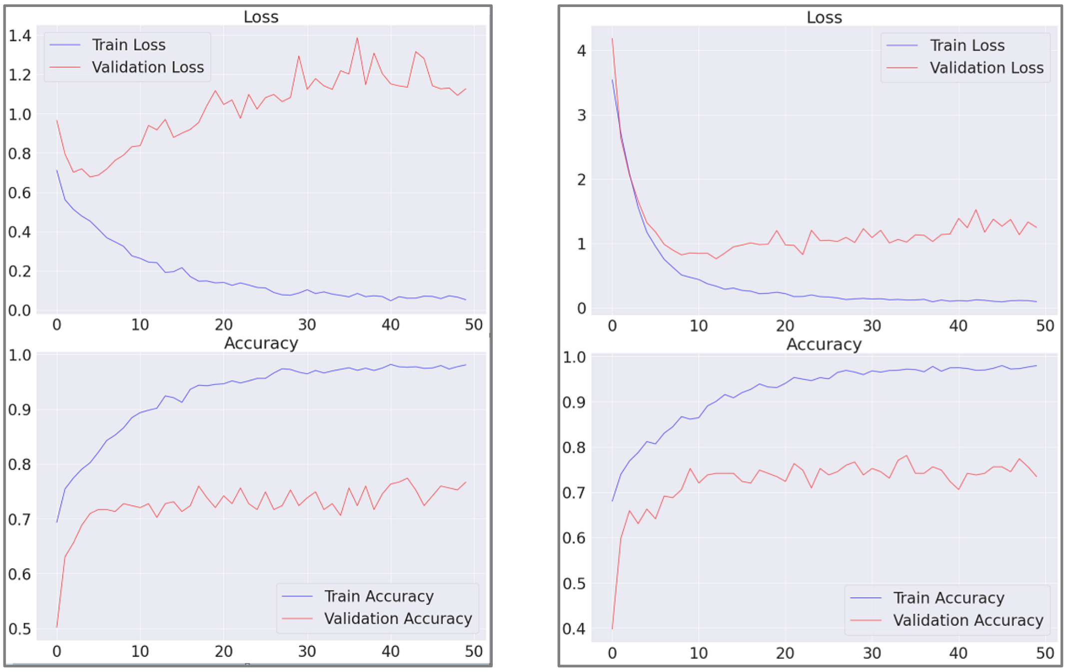
As can be seen in Table 3, the ResNet-ANN model obtained the best accuracy value between the applied four scenarios, i.e. pure ANN model, Inception-ANN model, MobileNet-ANN model, and ResNet-ANN model. Also, the Precision, Recall, and f1 score values of the proposed ResNet-ANN model are shown in Table 4.

However, when we look at the results, the expected success has not been achieved even in the ResNet-ANN model, which gave the best results. So, we investigated the confusion matrix and the classification reports of the proposed models. As can be seen from the classification reports illustrated in Table 4 and the confusion matrix illustrated in Figure 8, the normal class has been detected with very high accuracy, and it was noticed that the highest error rate occurred between bacterial and viral classes. This is because the images in bacterial pneumonia and viral pneumonia classes have so many common features and similarities. Normally, these two classes have been considered as a single class in the previous studies in the literature [26]–[28]. Also, the high number of correctly classified images in the healthy class and the high number of misclassifications in other classes (which can be observed from the confusion matrix) support this situation. For example, 26 chest images from the first class (with Bacterial pneumonia class) have been wrongly perceived as second class (Viral pneumonia class), and 18 chest images from the second class have been wrongly perceived as first class. Also, by looking at the classification report, we can note that the Precision, Recall, and Accuracy metrics in the zeroth class (normal chest images) were very high, while these metrics were low for the other two classes. Therefore, we decided to test the performance of the proposed models after combining the two abnormal classes to be one class (applying binary classification for classifying the X-ray images as diseased and healthy.

Table 2. The effect of L1 regularization technique on the proposed ANN model.

L1 Regularization Technique	Accuracy Rate (%)
NONE	76,33
ACTIVE	78,33

JSR A
Journal of Scientific Reports



a. Before L1 regularization

b. After L1 regularization

Figure 7. Loss and accuracy curves of the training and validation values of the model before and after the use of the L1 regularization.

Table 3. Comparison of the proposed models' results in multi-class classification.

Used Model	Accuracy Rate (%)
Artificial Neural Networks (ANN)	78,33
ResNet-ANN	**81,67**
Inception-ANN	75
MobileNet-ANN	78,67

Table 4. Classification report of the ResNet-ANN model used for multi-class classification.

Class	Precision	Recall	F1-Score
0	91	99	95
1	79	70	74
2	75	76	75

JSR A
Journal of Scientific Reports

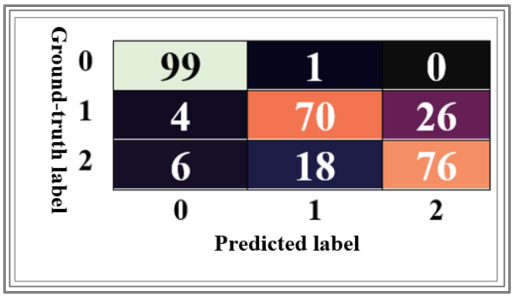


Figure 8. The confusion matrix of the ResNet-ANN model used for multi-class classification.

After that, we proposed using the same proposed architectures i.e. pure ANN model, Inception-ANN model, MobileNet-ANN, and ResNet-ANN model for conducting binary classification. To this end, we changed the output layer in the model i.e. Softmax layer into Sigmoid layer, and the "sparse_categorical_crossentropy" loss function into "binary_crossentropy".
When the results obtained after conducting the binary classification process have been investigated it is observed that great success has been achieved by all the proposed architectures. In this case study, the proposed MobileNet-ANN model has achieved the best results with a classification accuracy of 95.67%.

Also, the classification report and confusion matrix of the MobileNet-ANN model, which gives the highest result in the binary classification dataset, are shown in Table 6 and Figure 9. As seen in the classification report and confusion matrix, the misclassification between the classes is very low. Particularly, the F1 scores of the proposed MobileNet-ANN architecture reached 97 and 94% for pneumonia and normal classes respectively. Also, it can be concluded from the confusion matrix that only 4 data samples from the normal class and 9 samples from the pneumonia class have been wrongly predicted by the model.

Table 5. Comparison between the results of the proposed models when they were used for binary classification.

Used Model	Accuracy Rate (%)
Artificial Neural Networks (ANN)	92,67
ResNet-ANN	94,33
Inception-ANN	94,33
MobileNet-ANN	**95,67**

Table 6. Classification report of the MobileNet-ANN model used for binary classification.

Class	Precision	Recall	F1-Score
0	91	96	94
1	98	95	97

JSR A
Journal of Scientific Reports

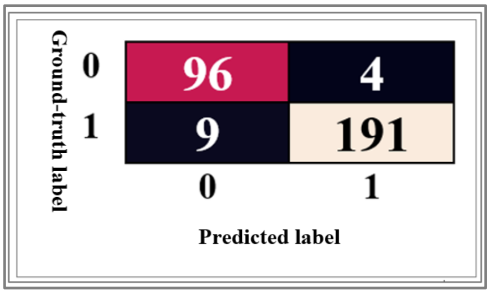


Figure 9. The confusion matrix of the MobileNet-ANN model used for binary classification.

5. COMPARISON STUDY

In this section, the results obtained using the proposed ANN model have been compared with some well-known and pre-trained CNN models. Particularly, in order to prove the efficiency of the proposed model, the pre-trained models that we used as the feature extraction phase in our model, namely ResNet, MobileNet, and Inception, have been tested as End-to-End models. In other words, the used pre-trained ResNet, MobileNet, and Inception models have been tested in terms of classifying X-ray images into two classes as standalone models. In addition, the results obtained in this work have been compared with the results of some previously conducted studies. using similar datasets. Table 7 shows the conducted comparison study. Also, figure 10 illustrates a comparison between the proposed models in terms of the obtained classification accuracy.

Table 7. Comparison of the obtained results.

Study	**Year**	**Used Model**	**Classification Accuracy**
DARICI et al. [14]	2020	CNN model	95%
Kaushik et al. [23]	2020	CNN model	92.31%
Mabrouk et al. [24]	2022	Ensemble CNN model	93.91%
Sharma et al. [25]	2020	CNN model	90.68%
Tested End-to-End MobileNet	-	Pre-trained MobileNet	88.33%
Tested End-to-End ResNet	-	Pre-trained ResNet	91.33%
Tested End-to-End Inception V3	-	Pre-trained Inception V3	91.33%
ResNet-ANN	**-**	**Pre-trained ResNet + ANN**	**94,33**
Inception-ANN	**-**	**Pre-trained Inception V3 + ANN**	**94,33**
MobileNet-ANN	**-**	**Pre-trained MobileNet + ANN**	**95,67**

6. DISCUSSION AND CONCLUSION

Lung X-Ray images can be used as a first-line procedure to detect and classify healthy persons, patients with viral pneumonia, and patients with bacterial pneumonia. At the same time, the similarity between the features of the lung X-Ray images with viral pneumonia and the image with bacterial pneumonia makes it difficult for radiologists and physicians to diagnose the disease. In this study, an artificial neural network-based model has been proposed to diagnose pneumonia lung images. The proposed ANN model has been built empirically so that a range of values for each hyperparameter has been tested and the structure that can give the best results has been chosen. To this end, multiple values have been tested for multiple hyperparameters including the number of dense layers, the unit number in each layer, activation functions, the weight initializer that should be used in each layer, the batch size used during fed the data to the model, and the learning rate that should be used for training the model. First, the proposed ANN model has been tested as an end-to-end model in order to evaluate its efficiency in detecting pneumonia from X-ray images. This model has been named as Pure ANN model. After that, we proposed to use some well-known CNN architectures as a feature extraction phase to improve the performance of the proposed pure ANN model. To this end, three different pre-trained CNN architectures namely ResNet, Inception, and MobileNet, have been adopted as a feature extraction phase for the proposed pure ANN model. Therefore, four different deep learning models, namely pure ANN, ResNet-ANN, MobileNet-ANN, and Inception-ANN, have been proposed for detecting pneumonia from X-Ray images. The proposed four models have been used in two different scenarios. In the first scenario, the proposed models have been adopted for classifying the lung X-ray images into three classes (multi-class classification) namely normal lung, lung with viral pneumonia, and lung with bacterial pneumonia. In the second scenario, the proposed models have been adopted for classifying the X-ray images as pneumonia-free lung (or normal lung) and lung with pneumonia. As a result, it was concluded that the deep learning techniques can reliably distinguish lung X-Ray images of patients with bacterial and viral pneumonia from lung X-Ray images of other healthy people. In fact, the deep learning model that gives the best results has been obtained by hybridizing the MobileNet well-known model with a fine-tuned ANN model, and its accuracy rate reached 95.67%, as seen in Figure 10. However, since the X-Ray images of patients with viral and bacterial pneumonia are very similar, all the models used were partially successful in distinguishing between the viral and bacterial pneumonia images. Particularly, when the X-ray images have been classified into three classes using the proposed model the ResNet-ANN gave the best results with an accuracy rate of 81.67%. At the end of the conducted study, and as proof of concepts for the efficiency of the proposed models, we tested the used three pre-trained models as end-to-end models and the obtained results have been compared with the results obtained when the same models have been used as feature extractor for the proposed ANN model. We concluded that it is more efficient to use these models as aid models for some other models rather than using them as end-to-end models, where their accuracy did not exceed 91.33% compared to 94.33 and 95.67% that was obtained by using these models as a feature extraction phase for the proposed model. Therefore, it can be concluded that this model can be used as X-Ray image-based diseases detector in order to ease the work of health sector specialists.

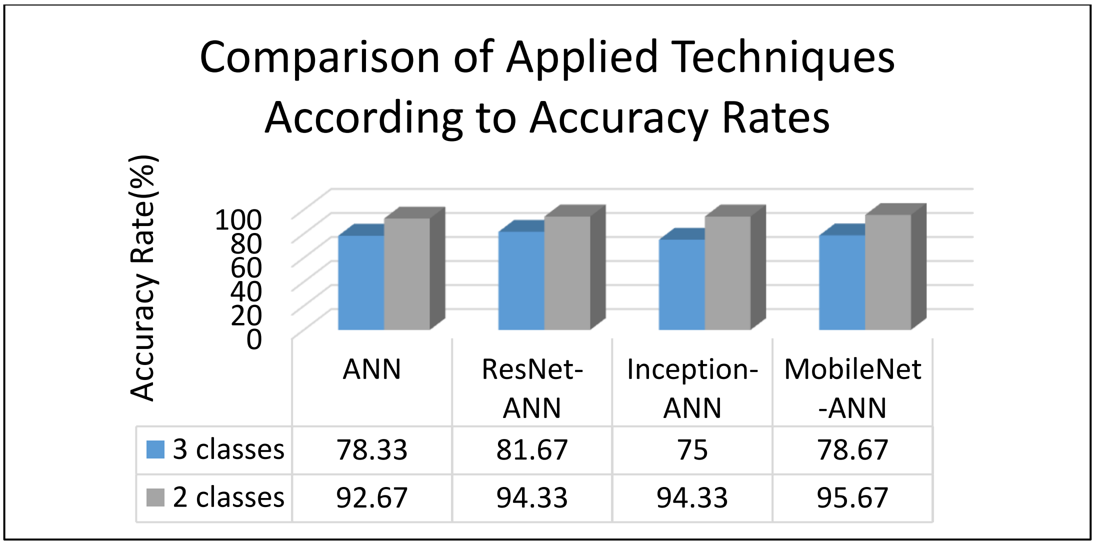


Figure 10. Comparison of proposed models According to Accuracy Rates.

In future studies, deeper and more fine-tuned structures can be conducted using a better level of processor capabilities and larger datasets. In addition, hyperparameters in the model can be fine-tuned using optimization methods such as Grid Search and Bayesian optimization algorithms. Also, the deep learning techniques such as well-known CNN architectures, ANN structures, and maybe auto-encoder structures can be tuned and used as a feature extractors for classical machine learning algorithms.

ACKNOWLEDGEMENT

The authors would like to thank Google Colaboratory for allowing to run the model effectively on the cloud and for free.

REFERANCES

[1] Ayan, R. and Ünver, H. M., (2019), “Diagnosis of pneumonia from chest X-ray images using deep learning,” in 2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT), pp. 1–5.

[2] Rajpurkar et al., (2017), “Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning,” arXiv preprint arXiv:1711.05225.

[3] El Asnaoui, K.,Chawki, Y. and Idri, A., (2021), “Automated methods for detection and classification pneumonia based on x-ray images using deep learning,” in Artificial intelligence and blockchain for future cybersecurity applications, Springer, pp. 257–284.

[4] Jain, R.., Nagrath, R., Kataria, G., Kaushik, V. S. and Hemanth, D. J., (2020), "Pneumonia detection in chest X-ray images using convolutional neural networks and transfer learning," Measurement, vol. 165, p. 108046.

[5] Hasan et al., (2021), "Deep learning approaches for detecting pneumonia in COVID-19 patients by analyzing chest X-ray images," Math Probl Eng, vol. 2021, pp. 1–8.

[6] Stephen, O., Sain, M., Maduh, U. J. and Jeong, D-U., (2019), "An efficient deep learning approach to pneumonia classification in healthcare," J Healthc Eng, vol. 2019.

[7] Elshennawy, N.M. and Ibrahim, D. M., (2020), "Deep-pneumonia framework using deep learning models based on chest X-ray images," Diagnostics, vol. 10, no. 9, p. 649.

[8] Jaiswal, A. K., Tiwari, P., Kumar, S., Gupta, D., Khanna, A. and Rodrigues, J. J. P. C., (2019) "Identifying pneumonia in chest X-rays: A deep learning approach," Measurement, vol. 145, pp. 511–518.

[9] El Asnaoui, K. and Chawki, Y., (2021), "Using X-ray images and deep learning for automated detection of coronavirus disease," J Biomol Struct Dyn, vol. 39, no. 10, pp. 3615–3626.

[10] Ouchicha, C., Ammor, O. and Meknassi, M., (2020), "CVDNet: A novel deep learning architecture for detection of coronavirus (Covid-19) from chest x-ray images," Chaos Solitons Fractals, vol. 140, p. 110245.

[11] Narin, A., Kaya, C., and Pamuk, Z., (2021), "Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks," Pattern Analysis and Applications, vol. 24, pp. 1207–1220.

[12] Ozturk, T., Talo, M., Yildirim, E. A., Baloglu, U. B., Yildirim, O. and Acharya, U. R., (2020), "Automated detection of COVID-19 cases using deep neural networks with X-ray images," Comput Biol Med, vol. 121, p. 103792.

[13] Jain, G., Mittal, D., Thakur, D. and Mittal, M. K., (2020), "A deep learning approach to detect Covid-19 coronavirus with X-Ray images," Biocybern Biomed Eng, vol. 40, no. 4, pp. 1391–1405.

[14] Darici, M. B., Dokur, Z. and Olmez, T., (2020) "Pneumonia detection and classification using deep learning on chest x-ray images," International Journal of Intelligent Systems and Applications in Engineering, vol. 8, no. 4, pp. 177–183.

[15] Nayak, S. R., Nayak, D. R., Sinha, U., Arora, V. and Pachori, R. B., (2021), "Application of deep learning techniques for detection of COVID-19 cases using chest X-ray images: A comprehensive study," Biomed Signal Process Control, vol. 64, p. 102365.

[16] Apostolopoulos, I. D., Aznaouridis, S. I. and Tzani, M. A., (2020), "Extracting possibly representative COVID-19 biomarkers from X-ray images with deep learning approach and image data related to pulmonary diseases," J Med Biol Eng, vol. 40, pp. 462–469.

[17] Bhardwaj P. and Kaur, A., (2021), "A novel and efficient deep learning approach for COVID-19 detection using X-ray imaging modality," Int J Imaging Syst Technol, vol. 31, no. 4, pp. 1775–1791.

[18] Bakır, H. and Yılmaz, Ş., "Using Transfer Learning Technique as a Feature Extraction Phase for Diagnosis of Cataract Disease in the Eye," International Journal of Sivas University of Science and Technology, vol. 1, no. 1, pp. 17–33.

[19] Kermany et al., (2018), "Identifying medical diagnoses and treatable diseases by image-based deep learning," Cell, vol. 172, no. 5, pp. 1122–1131.

[20] Khobragade, S., Tiwari, A., Patil, C. Y., and Narke, V., (2016), "Automatic detection of major lung diseases using chest radiographs and classification by feed-forward artificial neural network," in 2016 IEEE 1st international conference on power electronics, intelligent control and energy systems (ICPEICES), pp. 1–5.

[21] Hassaballah M., and Awad, A. I., (2020), Deep learning in computer vision: principles and applications. CRC Press.

[22] Otter, D. W., Medina, J. R. and Kalita, J. K., (2020), "A survey of the usages of deep learning for natural language processing," IEEE Trans Neural Netw Learn Syst, vol. 32, no. 2, pp. 604–624.

[23] He, K., Zhang, X., Ren, S. and Sun, J., (2016) "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778.

[24] Szegedy et al., (2015), "Going deeper with convolutions," in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1–9.

[25] Howard et al., (2019), "Searching for mobilenetv3," in Proceedings of the IEEE/CVF international conference on computer vision, pp. 1314–1324.

[26] Sirish Kaushik, V., Nayyar, A., Kataria, G. and Jain, R., (2020), "Pneumonia detection using convolutional neural networks (CNNs)," in Proceedings of First International Conference on Computing, Communications, and Cyber-Security (IC4S 2019), pp. 471–483.

[27] Sharma, H. , Jain, J. S., Bansal, P., and Gupta, S., (2020), "Feature extraction and classification of chest x-ray images using cnn to detect pneumonia," in 2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence), pp. 227–231.

[28] Al Mubarok, A. F., Dominique, J. A. M. and Thias, A. H., (2019), "Pneumonia detection with deep convolutional architecture," in 2019 International conference of artificial intelligence and information technology (ICAIIT), pp. 486–489.