



COVID-19 and pneumonia diagnosis from chest X-ray images using convolutional neural networks

Muhab Hariri¹ · Ercan Avşar^{2,3}

Received: 3 November 2022 / Revised: 15 January 2023 / Accepted: 25 February 2023 / Published online: 13 March 2023
© The Author(s), under exclusive licence to Springer-Verlag GmbH Austria, part of Springer Nature 2023

Abstract

X-ray is a useful imaging modality widely utilized for diagnosing COVID-19 virus that infected a high number of people all around the world. The manual examination of these X-ray images may cause problems especially when there is lack of medical staff. Usage of deep learning models is known to be helpful for automated diagnosis of COVID-19 from the X-ray images. However, the widely used convolutional neural network architectures typically have many layers causing them to be computationally expensive. To address these problems, this study aims to design a lightweight differential diagnosis model based on convolutional neural networks. The proposed model is designed to classify the X-ray images belonging to one of the four classes that are Healthy, COVID-19, viral pneumonia, and bacterial pneumonia. To evaluate the model performance, accuracy, precision, recall, and F1-Score were calculated. The performance of the proposed model was compared with those obtained by applying transfer learning to the widely used convolutional neural network models. The results showed that the proposed model with low number of computational layers outperforms the pre-trained benchmark models, achieving an accuracy value of 89.89% while the best pre-trained model (Efficient-Net B2) achieved accuracy of 85.7%. In conclusion, the proposed lightweight model achieved the best overall result in classifying lung diseases allowing it to be used on devices with limited computational power. On the other hand, all the models showed a poor precision on viral pneumonia class and confusion in distinguishing it from bacterial pneumonia class, thus a decrease in the overall accuracy.

Keywords Classification · Convolutional neural networks · COVID-19 · Deep learning · Transfer learning · Lung diseases

1 Introduction

The flare-up of the COVID-19 has increased the need for new effective and faster diagnostic methods than those manual diagnosis provided by the experts. The huge number of infected people and insufficient number of medical staff and health facilities in some countries increased the burden on the health system. On the other hand, the widespread usage of rapid diagnosis tools, which help in taking measurements and suggesting an appropriate treatment, is an evidence of both sieging effect of the pandemic and their usefulness in

mitigating the spread of virus. In recent years, the reliance on machine learning techniques in the medical field has increased dramatically. Roy et al. (2022) discussed the prospects of supervised machine learning (SML) in the healthcare sector, the challenges it faces, how to solve it and the opportunity for healthcare through AI and SML in the near future. In general, these techniques have proven to be effective in diagnosing the diseases with acceptable accuracy and high speed. Jaiswal et al. (2021) proposed an optimized technique for identification of blindness in retinal images using the deep learning models. Ensembles of convolutional neural networks (CNN) has shown to be an efficient tool for skin cancer detection (Al-Karawi 2022) while segmentation of skin diseases also possible with the methods based on CNN (Huang et al. 2020). Among the many other studies based on medical images, there are different applications such as detection and diagnosis of gastric cancer (Cao et al. 2019), breast cancer (Wang et al. 2019), brain tumors (Salçin 2019), pneumonia (Avsar 2021), lung diseases (Kabiraj 2022) and lung cancer (Gunjan et al. 2022; Agarwal 2021). The use

✉ Muhab Hariri
mohab369@gmail.com

¹ Electrical and Electronics Engineering Department, Çukurova University, 01330 Adana, Turkey

² National Institute of Aquatic Resources, Technical University Denmark, 9850 Hirtshals, Denmark

³ Computer Engineering Department, Dokuz Eylül University, 35390 İzmir, Turkey

of machine learning techniques in the medical field was not limited to diagnosing diseases only, but also included several domains such as segmenting the medical images (Pal et al. 2022; Rajinikanth et al. 2022) and use the segmented images for specific purposes like predicting the type of the fetal brain (pathological or neurotypical) and predicting the gestational age of the fetus (Gangopadhyay et al. 2022).

Symptoms of the COVID-19 vary from person to person; however, the most frequently reported symptoms include fatigue, coughing, and shortness of breath. The problem is that these symptoms may be associated with other similar illnesses such as pneumonia (Zayet et al. 2020). Reverse transcription polymerase chain reaction (RT-PCR) tests are currently one of the most popular and reliable methods to discover the presence or absence of this virus. However, these tests have many drawbacks. This method is slow as sometimes it takes 24 h for a result to appear. In addition, it puts medical staff at risk of catching the virus as a result of the physical contact with the patient. It is also expensive, thus inaccessible for poor countries. Therefore, a quick, reliable and cheap way to diagnose COVID-19 and pneumonia infections is necessary and help to take the appropriate actions. Chest radiography (chest X-ray) is also a commonly used method for diagnosing the lung diseases and detect COVID-19; however, this method has drawbacks as well. To diagnose the diseases by the X-ray images, experts are required to inspect the images. In addition, it can provide false results because of the similarity between chest X-ray images of people infected with COVID-19 and different types of pneumonia.

CNN is a popular machine learning method that is used to classify images and detect objects. In this work, a sequential CNN architecture is proposed to detect the X-ray images belonging to patients with COVID-19, Viral Pneumonia, and Bacterial Pneumonia. For benchmarking purposes, the classification performance of the proposed architecture was compared with those obtained by widely-used CNN models pretrained on ImageNet dataset. These benchmark models are MobilenetV2, InceptionResNetV2, ResNetV2, EfficientNet B2, EfficientNet B0, NasNetMobile, InceptionV3, VGG16 and VGG19. These models differ in terms of design, number of parameters and depth. These factors allow for a comparison of models interchangeably with the proposed model. In terms of the practical implications, the property of being lightweight allows the proposed model to be used on devices with limited processing capability. In other words, it opens the way to design and develop cheap auxiliary tools to detect lung diseases.

Many works have been conducted to diagnose COVID-19 and pneumonia diseases; however, most of these works merge viral and bacterial lung diseases into one category. This leads to less understanding about how CNNs perform in classifying these diseases separately and provides a limited

diagnosis scheme. In addition, the number of parameters is not discussed in the proposed models presented in these studies, so it is not clear how well these models can work in environments with low resources. Therefore, within the scope of this study, the answers to the following research questions are sought:

Q1: How successful is the proposed CNN model in detecting the lung diseases separately (i.e. COVID-19, Viral Pneumonia and Bacterial Pneumonia)?

Q2: Can a light model with low number of parameters, and thus low computation cost, performs well for this classification task?

As a result of the experiments performed, a CNN model is proposed to address the limitations of the existing studies. In particular, the contributions of this study are given in the list below.

- The proposed model has a lower number of convolutional layers and parameters than the benchmark models. Therefore, this is a lightweight model that requires relatively small amount of calculation at training and test phases.
- It may achieve better classification results. In particular, the proposed model is capable of distinguishing the COVID-19, viral and bacterial pneumonia cases with a high true detection rate.
- As a result of being lightweight, the proposed model does not require expensive and powerful hardware as it includes relatively a low number of parameters. This advantage makes it applicable to devices with low computational power such as edge devices and single board computers.

The remainder of this paper is organized as follows: In Sect. 2, the existing studies in the related literature are reviewed. In Sect. 3, the dataset, models and performance metrics are introduced. Section 5 and 6 represent the results and discussion. Finally, the paper is concluded in Sect. 6.

2 Literature review

Among various approaches for COVID-19 detection, chest X-ray images are widely used and hence there are many available studies in this context. Due to the automated feature extraction capability of convolutional neural networks (CNN), they are used for classification of unstructured data such as images. Therefore, there are numerous previous studies in which chest X-ray images were used together with CNN models for detection of COVID-19 infections. In some studies, the researchers aim to discriminate the X-ray images of positive COVID-19 cases from the healthy X-ray images. However, COVID-19 cases are very likely to be confused

with pneumonia infections which can be bacterial or viral, hence, there are other studies where the detection is considered as a three classes or a four classes problem to detect COVID-19 and pneumonia together.

The number of studies considering a binary problem to detect healthy and COVID-19 X-ray images is relatively high. For instance, Reynaldi et al. (2021) used CNN with the Resnet-101 model as an image recognition method to detect COVID-19. The authors used a dataset contains 2562 images categorized as positive COVID-19 (1281 images) and negative COVID-19 (1281 images). Contrast Limited Adaptive Histogram Equalization (CLAHE) preprocessing process was applied on dataset and the results showed that the model with CLAHE data achieved better result with accuracy of 99.61% compared with the raw data where the accuracy was 99.22%. In addition, Hemdan et al. (2003) used many deep convolutional neural network models (VGG19, DenseNet201, InceptionV3, ResNetV2, Inception-ResNetV2, Xception, and MobileNetV2) to classify X-ray images into positive or negative COVID-19. The authors used dataset of 50 chest X-ray images that includes 25 positive and 25 healthy cases. The results showed that VGG19 and DenseNet201 provides the highest classification performance with accuracy of 90%. Narin et al. (2021) used 5 pre-trained convolutional neural network models namely, ResNet50, ResNet101, ResNet152, InceptionV3 and InceptionResNetV2 to detect COVID-19. The dataset they used contains 7396 chest X-ray images classified as 341 COVID-19 images, 2800 Normal images, 2772 Bacterial Pneumonia and 1493 Viral Pneumonia. The dataset was divided into three binary-class datasets: dataset-1 contains COVID-19 and Normal classes, dataset-2 contains COVID-19 and Viral Pneumonia classes while dataset-3 contains COVID-19 and Bacterial Pneumonia classes. ResNet50 model achieved the best classification results with accuracy of 96.1%, 99.5% and 99.7% for dataset-1, dataset-2 and dataset-3, respectively. Ohata et al. (2020) used transfer learning models as features extractors to detect COVID-19. Transfer learning models that were used in this work are VGG16, VGG18, InceptionV3, InceptionResNetV2, ResNet50, NASNet-Large, NASNetMobile, Xception, MobileNet, DenseNet121, DenseNet169 and DenseNet201. These models were combined with many classifiers like k-Nearest Neighbor, Bayes, Random Forest, Multilayer Perceptron (MLP), and Support Vector Machine (SVM). The authors used two datasets where both of them have the same images for the COVID-19 class, but they have different images for the healthy class. The datasets are balanced and consist of 194 images for each class. The results showed that MobileNet model with the SVM classifier (linear kernel) achieved the best mean accuracy of 98.46% for one of the datasets while DenseNet201 model with MLP classifier was the best for another dataset with a mean accuracy of 95.64%.

In another work with binary class of images, Breve et al. (2011) performed a set of exhaustive classification experiments. In COVID-19 detection problem, they used 21 different CNN models that are VGG, ResNet, DenseNet, EfficientNet and their derivatives (i.e. DenseNet121, EfficientNetB1, ResNet152). In addition, ensembles of these CNN models were also employed. Their dataset contains 16,352 chest X-ray images where 2358 images are COVID-19 positive and 13,994 are COVID-19 negative. The negative data includes images with non-COVID-19 pneumonia. The results showed that DenseNet169 achieved the best results with an accuracy and F1 score of 98.15% and 98.12%, respectively. The ensemble approach increased the accuracy and F1 score of DenseNet169 to 99.25% and 99.24%, respectively. Maheen et al. (2010) used different pre-trained CNN models to detect COVID-19 using chest X-ray images. The models are: AlexNet, VGG-16, MobileNet-V2, SqueezeNet, ResNet-34, ResNet-50 and COVIDX-Net. The dataset contains 406 images distributed evenly to COVID-19 and healthy classes. ResNet-34 achieved the best prediction accuracy of 98.33%. Shenoy et al. (2010) proposed a new CNN model to detect COVID-19. A dataset contains 4316 chest X-ray images (2158 COVID-19 negative scans and 2,158 COVID-19 positive scans) was used and data augmentation technique was used. The model achieved an accuracy of 95.5%. Hasoon et al. (2021) proposed many methods that combines between image processing and classifiers (i.e. K-Nearest Neighbor (KNN) and Support Vector Machine (SVM)) for classification and early detection of COVID-19. The dataset includes normal and pneumonia COVID-19 X-ray images. The method that combines Local binary pattern (LBP) and KNN achieved the best accuracy of 99%. Mohammed et al. (2022) proposed an integrated method for selecting the optimal deep learning model based on a novel crow swarm optimization algorithm for COVID-19 diagnosis. ResNet50 model achieved the best accuracy of 91.46%.

Detection of pneumonia together with COVID-19 is also considered by many researchers. In that case, it becomes a three-classes problem. One of the methods was proposed by Montalbo (2021) where DenseNet121 was modified to classify normal, COVID-19 and Pneumonia (Bacterial and Viral) classes. The resulting model, which has lower parameters and depth than the original one, achieved an accuracy of 97.99%. It did not achieve better accuracy compared to the base model but showed to be able to outperform against some state-of-the-art deep convolutional neural network models. Same author in another study (Montalbo 2022) applied a truncation method on various of famous deep convolutional neural networks to reduce the number of parameters of the models and make it applicable with low computing resources. Chest X-ray images were used and the results showed that the InceptionResNetV2 model achieved the best accuracy of 97.41% in three-classes classification

(Normal, COVID-19 and Pneumonia) after truncating it and reducing its parameters to 441 K. In addition, Shome et al. (2021) proposed a vision transformer-based deep learning pipeline for detecting COVID-19 using chest X-ray images. Dataset with three-classes (Normal, COVID-19 and Pneumonia) contains 30 K of chest X-ray images (10 K for each class) was used and the proposed model achieved an accuracy of 98% for binary classification (Normal and COVID-19) and 92% for multi-class classification. Nagi et al. (2022) used a relatively large dataset to check the performance of deep learning. Xception model was the best in terms of accuracy. The model achieved an accuracy of 94.21% while the Custom-Model (proposed model in the study) achieved an accuracy of 92.38%.

Transfer learning is a widely utilized practical tool in this three-class problem as well. Makris et al. (2020) used several well-known CNN model with a dataset containing 336 chest X-ray images. According to the results, VGG16 and VGG19 achieved the best accuracy score of 95%. El Asnaoui et al. (2020) used well-known CNN architectures, namely DenseNet201, InceptionV3, InceptionResNetV2, Resnet50, MobileNetV2, VGG16 and VGG19 to classify COVID-19. The database used in this work contains 6087 X-ray and CT images (231 COVID-19, 1493 Viral Pneumonia, 2780 Bacterial Pneumonia and 1583 Normal images). COVID-19 and Viral Pneumonia classes were considered as one class in the classification process. InceptionResNetV2 and Densnet201 achieved the best results with accuracy of 92.18% and 88.09%, respectively. Alqudah et al. (2020) used pre-trained and proposed models like ShuffleNet, MobileNet and AOCTNet to extract the automated features from the images, then they passed these features to Soft-max, Support Vector Machine (SVM), Random Forest (RF) and K-Nearest Neighbor (KNN) classifiers. It was shown that features extracted by MobileNet performed the best accuracy.

In addition to modification of available transfer learning models, there are some other studies in which specific CNN architectures are proposed. For instance, Antonchuk et al. (2021) proposed a new CNN model for detecting COVID-19 and influenza cases. The model achieved an accuracy score of 93% on a dataset consisting of 4,152 X-ray images for each class. The CNN architecture proposed by Atitalah et al. (2023) was tested on two different datasets. First dataset (COVIDx) contains 15,475 chest X-ray images (8851 Normal, 6,053 Pneumonia and 571 COVID-19) while the other (Enhanced COVID-19) includes 1092 chest X-ray images (364 images for each class). Data augmentation was applied on datasets and class weight method was applied on COVIDx dataset to re-balance it. The results showed that the proposed model achieved an accuracy of 94% and 99% for COVIDx and enhanced COVID-19 datasets, respectively. Liu et al. (2022) proposed an approach comprises of many stages. EfficientNetV2 was considered as backbone network

then ResNet101 (feature fusion), Convolutional Block Attention Module and SVM classifiers, respectively, were used. The dataset contains three-classes (COVID-19, Normal and Pneumonia) and data augmentation was applied. The results showed that the system achieved an accuracy of 99.89%.

Different from the studies considering the Viral and Bacterial Pneumonia as one single class, it is possible to take them as separate classes and eventually have a four-classes problem. One example of such works is proposed by Zeiser et al. (2021). In their work, pretrained DenseNet121, InceptionResNetV2, InceptionV3, MobileNetV2, ResNet50 and VGG16 models were used for classification of the X-ray images together with CLAHE as a preprocessing method. Their dataset contains 5,181 images categorized to four-classes as COVID-19, Normal, Viral Pneumonia and Bacterial Pneumonia. The results showed that VGG16 achieved the best classification performance with an accuracy of 85.11%, sensitivity of 85.25%, specificity of 85.16% and F1-score of 85.03%. Bolhassani (2021) used an unbalanced chest X-ray dataset together with ResNet50 and DenseNet121 models. To eliminate the effect of the class imbalance, they applied data augmentation and achieved an accuracy score of 80.0%. Sait et al. (2021) proposed a model based on InceptionV3 model and multilayer perceptron. Dataset consists of four-classes (COVID-19, Normal, Bacterial and Viral Pneumonia) of chest X-ray images was used without data augmentation. The dataset was split into train and validation sets with a ratio of 80:20. It is noted that the authors did not use part of the dataset as test data which is important to check the robustness of the model's performance. The proposed model achieved a validation accuracy of 91.3% on the chest ray images. In a study focused on determining the seriousness of lung disease using chest X-ray images, Rajinikanth et al. (2022) implemented a pre-trained InceptionV3 scheme with chosen multi-class classifiers to detect the pneumonia and check its severity level. The dataset contains four-classes (Normal, Mild, Moderate, and Severe Pneumonia). The result achieved by K-Nearest Neighbor (KNN) classifier was the best in this work with an accuracy of 85.18%.

Based on the explanations above, the existing studies about pneumonia and COVID-19 detection using X-ray images are summarized in Table 1. As can be seen from the available studies in the literature, there are very different approaches for pneumonia and COVID-19 diagnosis; however, the majority of these studies consider it as a binary-class problem or merge viral and bacterial pneumonia together in one class. In other words, the analysis made on the three mentioned lung diseases is very limited. In addition, most of the works that proposed new models do not consider the computational load of the model. Typically, deeper models with more convolutional layers may achieve better feature extraction eventually yielding

Table 1 Summary of related works

Reference No	Number of classes	Dataset classes	Number of images in dataset	Dataset reference	Transfer learning	Proposed model	Best accuracy
Reynaldi (2021)	2	Positive Covid-19 & negative Covid-19	2562	Sait et al. (2020) (Version 1)	Yes	No	99.61%
Hemdan et al. (2003)	2	Positive Covid19 & negative Covid-19	50	Cohen (2023), Simonyan et al. (2023)	Yes	No	90%
Narin et al. (2021)	2	Dataset 1: Positive Covid-19 and normal Dataset 2: Positive Covid-19 and viral pneumonia Dataset 3: Positive Covid-19 and bacterial pneumonia	7396	Cohen (2023), Wang et al. (2017), Chest X-Ray Images (2023)	Yes	No	96.1% 99.5% 99.7%
Ohata et al. (2020)	2	Dataset 1: Positive Covid-19 and healthy Dataset 2: Positive Covid-19 images same dataset 1 but different images for healthy class	388 for each dataset	Cohen (2023), Wang et al. (2017), COVID-19 X rays (2023)	Yes	No	98.46% 95.64%
Breve (2111)	2	Positive Covid-19 & negative Covid-19 with non-Covid-19 pneumonia	16,352	COVIDx Dataset (2023)	Yes	No	99.25%
Maheen et al. (2110)	2	Positive Covid-19 & healthy	406	Unstated	Yes	No	98.33%
Shenoy (2021)	2	Positive Covid-19 & negative Covid19	4316	COVID-19 Radiography Database (2023), Wang et al. (2020)	No	Yes	95.5%
Hasoon et al. (2021)	2	Normal and Covid-19 images	5000	Cohen (2023)	No	Yes	99%
Mohammed (2022)	2	Dataset 1: Covid-19 and healthy CT images Dataset 2: 632 Covid-19 CT images	746 632	Cohen (2023)	Yes	Yes	91.46%
Shome et al. (2021)	2	Positive Covid-19 & normal	20,000	Sait et al. (2020) (Version 1), El-Shafai and Abd El-Samie (2020) and Qi et al. (2021)	No	Yes	98%
Montalbo (2021)	3	Positive Covid-19, normal and pneumonia (bacterial and viral)	9208	Sait et al. (2020) (Version 1)	No	Yes	97.99%
Montalbo (2022)	3	Normal, Covid-19 and Pneumonia	9208	Sait et al. (2020) (Version 1)	Yes	No	97.41%
Shome et al. (2021)	3	Positive Covid-19, normal and pneumonia (bacterial and viral)	30,000	Sait et al. (2020); El-Shafai and Abd El-Samie (2020b) (Version 1) and Qi et al. (2021)	No	Yes	92%

Table 1 (continued)

Reference No	Number of classes	Dataset classes	Number of images in dataset	Dataset reference	Transfer learning	Proposed model	Best accuracy
Nagi et al. (2022)	3	Covid-19, lung opacity and healthy	19,820	Chowdhury et al. (2020), COVID-19 Radiography Database (2023)	Yes	Yes	94.21%
Makris et al. (2020)	3	Positive Covid-19, normal and pneumonia	336	Cohen (2023), Chest X-Ray Images (2023)	Yes	No	95%
Asnaoui and Chawki (2020)	3	Bacterial pneumonia, viral pneumonia with positive Covid-19 and normal	6087	Cohen (2023), Kermany et al. (2018)	Yes	No	92.18%
Alqudah et al. (2020)	3	Positive Covid-19, normal and pneumonia	930	Cohen (2023), Chest X-Ray Images (2023)	Yes	Yes	99.46%
Antonchuk, et al. (2021)	3	Covid-19, influenza virus and healthy	12,456	Unstated	No	Yes	93%
Atitallah et al. (2023)	3	Dataset 1: Positive Covid-19, normal and pneumonia Dataset 2: Positive Covid-19, normal and pneumonia	15,475 1092	Wang et al. (2020) and Canayaz (2021)	No	Yes	94% 99%
Liu et al. (2022)	3	Positive Covid-19, Normal and pneumonia	9208	Sait, et al. (2020) (Version 1)	Yes	Yes	99.89%
Zeiser (2021)	4	Positive Covid-19, normal, viral pneumonia and bacterial pneumonia	5181	Sait, et al. (2020) (Version 3)	Yes	No	85.11%
Bolhassani (2105)	4	Positive Covid-19, normal, viral pneumonia and bacterial pneumonia	5929	CoronaHack -Chest X-Ray-Dataset. Available from (2023)	Yes	No	84.60%
Sait et al. (2021)	4	Positive Covid-19, normal, viral pneumonia and bacterial pneumonia	9208	Sait, et al. (2020) (Version 1)	Yes	Yes	91.3% (Validation accuracy)
Our work	4	Positive Covid-19, normal, viral pneumonia and bacterial pneumonia	9207	Sait, et al. (2020) (Version 3)	Yes	Yes	89.89%

more successful classification. However, such models have major drawbacks like the requirement of large amount of images and expensive hardware with heavy computational capability. This problem is present especially in those studies considering the four-class problem (healthy, COVID-19, viral pneumonia, and bacterial pneumonia). Therefore, this situation has been addressed to some extent in this study by proposing a model with reduced convolutional layers as well as number of weights. Hence, it becomes more suitable for the detection task to be executed on a larger scale of digital devices including those with relatively lower computational power.

3 Methods

3.1 Dataset

In this work, a publicly available dataset of chest X-ray images have been used (Sait, et al. 2020). The dataset contains 9207 chest X-ray images categorized as follows: 3269 normal, 1281 COVID-19, 3001 bacterial pneumonia and 1656 viral pneumonia chest X-ray images. Figure 1 shows some sample images of the dataset.

The dataset was divided into training, validation and test sets with ratio of 60%, 20%, and 20%, respectively. After

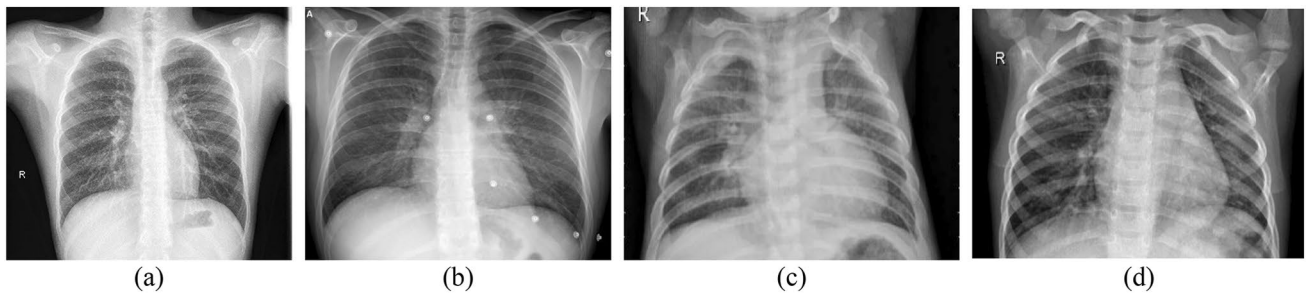


Fig. 1 Samples of normal **a**, COVID-19 **b**, Bacterial Pneumonia **c** and Viral Pneumonia **d** chest X-ray images

Table 2 The number of images in training set before and after applying data augmentation

	Training set	
	Before data augmentation	After data augmentation
Normal	1961	15,625
Covid-19	769	6145
Bacterial pneumonia	1801	14,345
Viral pneumonia	994	7938

dividing the dataset, data augmentation (Mikołajczyk 2018) technique was applied on the training set. Data augmentation is a technique which is used to increase the number of images in the dataset, this leads to increase its diversity and reduce the risk of overfitting. Horizontal flip and shifting

operations were applied. 10%, 30% and 50% were used for each of width shift and height shift. Table 2 shows the number of images in training set before and after applying data augmentation.

3.2 The proposed CNN architecture

For detecting COVID-19, viral and bacterial pneumonia samples, a lightweight sequential CNN architecture with small number of parameters was proposed. The successive convolutional and pooling layers in the model are followed by fully connected layers. Finally, the softmax function was used in last layer of the classification part for final prediction. Figure 2 shows a generic CNN architecture with convolutional, pooling and fully connected layers. In the feature extraction part of the proposed model, there are five convolutional and pooling layers. On the other hand, the

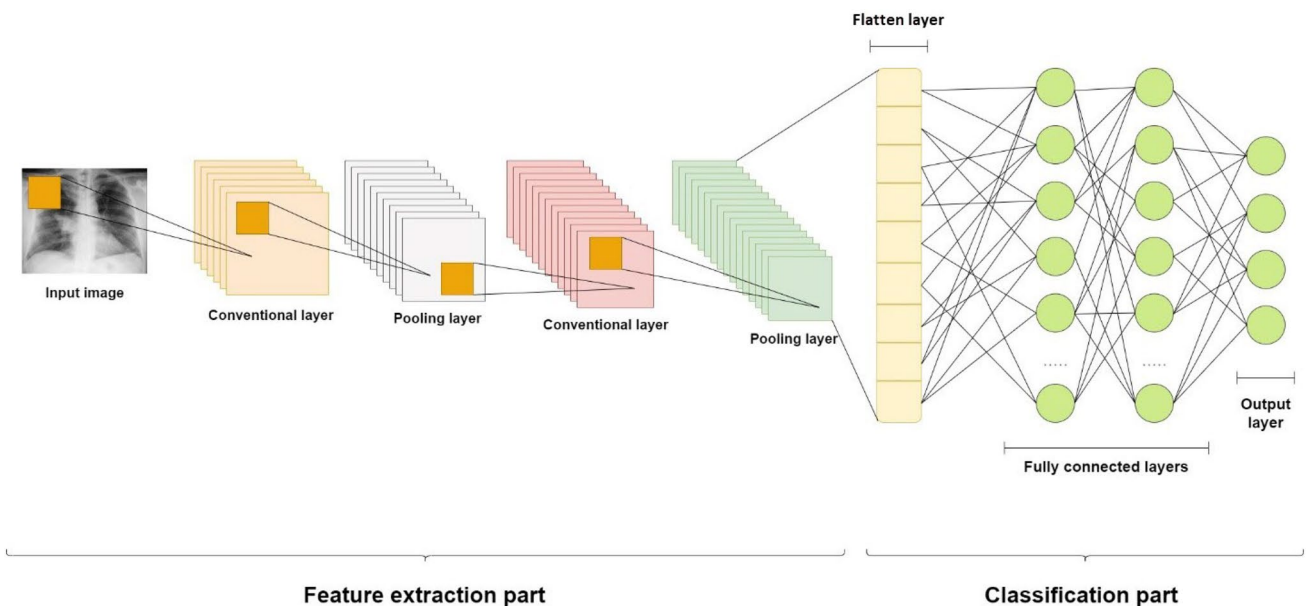


Fig. 2 A generic architecture of a sequential CNN

classification part involves three dense layers with dropout layers in between.

As Fig. 2 shows, the convolutional layers receive the input image and convolve it by a filter with specific dimensions. This process produces an output known as feature map. The feature map is being processed by a pooling layer and activation function. Rectified linear unit (ReLU) was used as non-linear activation function due to its ability to accelerate the training process and solve the vanishing gradient problem. ReLU returns all the negative inputs to zero while positive inputs pass without any change as Fig. 3 shows. The following equation can express the mathematical expression of ReLU:

$$f(x) = \begin{cases} 0, & x < 0 \\ x, & x \geq 0 \end{cases}$$

Pooling layers are responsible for reducing the size of the feature maps and max-pooling operation was used in the proposed model. This is accomplished by

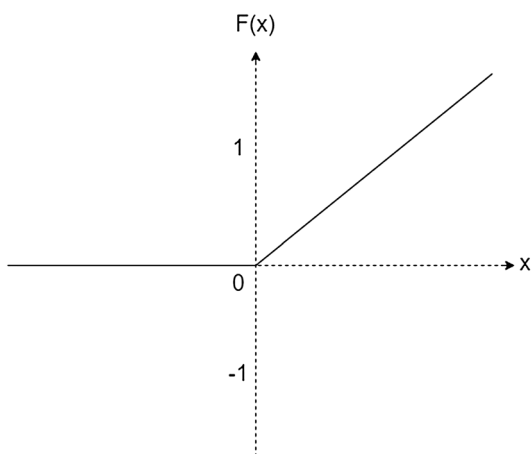


Fig. 3 ReLU function

two-dimensional filter that passes through the feature map. Max-pooling selects the maximum value covered by the filter. This process leads to a lower number of parameters in the model and can speed up the computational process. The max-pooling operation with a filter size of 2 × 2 and stride of 2 is illustrated in Fig. 4.

After passing the input through several convolutional and max-pooling layers, the flatten layer works on converting the resulted feature map from two dimensional and multichannel feature map to one dimensional vector. This operation is important as the fully connected layer expects a vector as input. The operation of flatten layer is given in Fig. 4.

Fully connected layers (FC) are responsible for final classification. It consists of input, hidden and output layers. In each layer there are many neurons. Softmax was chosen as activation function in the output layer because it converts the output to a probability distribution. The following equation shows the mathematical expression of softmax.

$$\sigma(z)_i = \frac{e^{z_i}}{\sum_j e^{z_j}}$$

Dropout layers were added to prevent overfitting and provide better generalization of the model. Dropout layers invalidate some neurons randomly in the fully connected layer during the training process. The number of such neurons are determined by the user defined dropout rate parameter. Tables 3 and 4 illustrate the hyperparameter values for each layer in the proposed model. These values affecting the model performance were determined empirically but taking care of the constraint that the model should possess smaller amount of convolutional layers as well as weights.

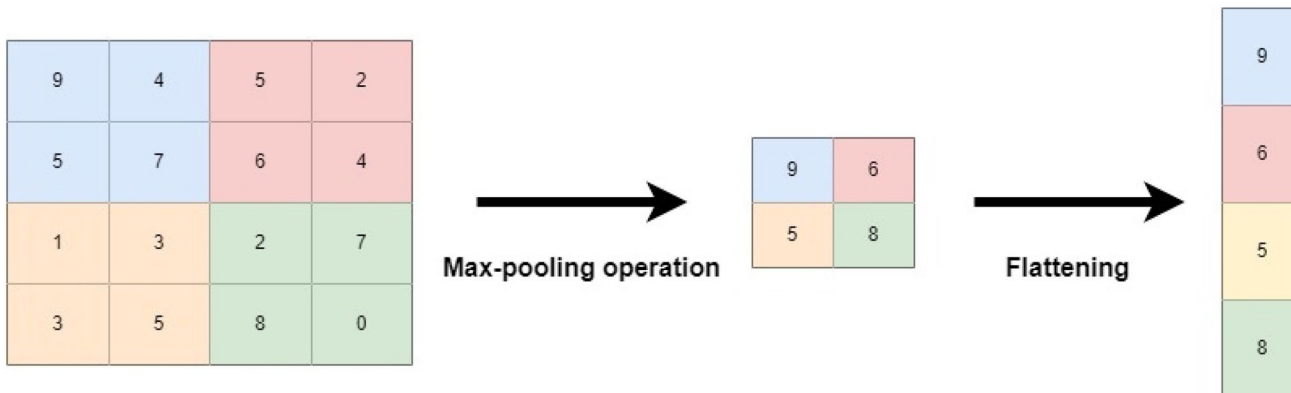


Fig. 4 Max Pooling operation using filter of 2 × 2 and flattening process

Table 3 The hyperparameter of the feature extraction part in the proposed model

	Number of filters	Kernel size	Max-pooling filter size	Activation function
Convolutional layer 1	35	5.5	2.2	ReLU
Convolutional layer 2	50	3.3	2.2	ReLU
Convolutional layer 3	65	3.3	2.2	ReLU
Convolutional layer 4	90	3.3	2.2	ReLU
Convolutional layer 5	125	3.3	2.2	ReLU

Table 4 The hyperparameter of the classification part in the proposed model

	Number of neurons	Dropout rate	Activation function
Dense layer 1	256	0.4	ReLU
Dense layer 2	128	0.4	ReLU
Dense layer 3	64	0.4	ReLU
Output layer	4	None	Softmax

3.3 Transfer learning models

For comparison with the proposed model, transfer learning models were used. Transfer learning is machine learning technique that depends on using the weights of pre-trained models as starting point for training the model on new task using new dataset. Subsequently, the images in our dataset were fed into different pre-trained CNN models which have different input image size, number of layers and number of parameters. The models are, namely, EfficientNet B0 (Tan and Le 2019), EfficientNet B2 (Tan and Le 2019), InceptionV3 (Szegedy et al.2016), InceptionResNetV2 (Szegedy et al. 2017), MobileNetV2 (Sandler et al. 2018), NASNet-Mobile (Zoph et al. 2018), ResNetV2_152 (He et al. 2016), VGG16 (Simonyan and Zisserman 2023) and VGG19 (Simonyan and Zisserman 2023). These models were trained on ImageNet (Russakovsky et al. 2015) dataset. The weights of layers in these models were frozen except of the output layer which was set to have four units. In addition, softmax was used as an activation function in the output layer. Table 5 shows the dimensions of input image, number of total parameters and number of trainable parameters in the transfer learning models and the proposed model.

3.4 Evaluation metrics

Standard metrics like accuracy, precision, recall and F1-Score were considered for evaluation of the pre-trained models and the proposed model. The components of

Table 5 Number of parameters and trainable parameters in transfer learning models and the proposed model

Model Name	Input Image Dimensions	Total number of parameters	Number of trainable parameters
VGG16	224×224 ×3	134,276,932	16,388
VGG19	224×224 ×3	139,586,628	16,388
MobileNetV2	224×224 ×3	2,263,108	5124
EfficientNet B0	224×224 ×3	4,054,688	5124
EfficientNet B2	260×260 ×3	7,774,198	5636
InceptionV3	299×299 ×3	21,810,980	8196
ResNetV2_152	224×224 ×3	58,339,844	8196
InceptionResNetV2	299×299 ×3	54,342,884	6148
NASNetMobile	224×224 ×3	4,273,944	4228
The proposed model	224×224 ×3	1,043,558	1,043,558

confusion matrix shown in Table 6 were used for calculating these metrics.

P_{CC} variable refers to the number of predictions where the images labeled as COVID-19 were correctly classified as a COVID-19. P_{CC} represents the True Positive ($TP_{Covid19}$) of COVID-19 class. On the other hand, P_{NC} , P_{BC} , and P_{VC} represent the COVID-19 images incorrectly labeled as Normal, Bacterial Pneumonia, and Viral Pneumonia, respectively.

True Negative (TN) for each class can be calculated by taking the sum of all the values of the confusion matrix except the values in row and column of the class being studied. The following equation shows the True Negative of COVID-19 class:

$$TN_{Covid19} = P_{NN} + P_{BN} + P_{VN} + P_{NB} + P_{BB} + P_{VB} + P_{NV} + P_{BV} + P_{VV}$$

False Positive (FP) is the sum of all the values in the column of the being studied class except the true positive value. The equation that presents false positive of COVID-19 class is:

$$FP_{Covid19} = P_{CN} + P_{CB} + P_{CV}$$

False Negative (FN) is the sum of all the values in the row of the being studied class except the true positive value. The equation that presents false negative of COVID-19 class is:

Table 6 The confusion matrix

		Predicted			
		Covid-19 (C)	Normal (N)	Pneumonia bacterial (B)	Pneumonia viral (V)
Actual	Covid-19 (C)	P_{CC}	P_{NC}	P_{BC}	P_{VC}
	Normal (N)	P_{CN}	P_{NN}	P_{BN}	P_{VN}
	Pneumonia bacterial (B)	P_{CB}	P_{NB}	P_{BB}	P_{VB}
	Pneumonia viral (V)	P_{CV}	P_{NV}	P_{BV}	P_{VV}

Table 7 Hyperparameter configuration

Hyperparameter	Value
β_1	0.9
β_2	0.999
Learning rate	0.001
Batch size	32 and 64
Number of training epochs	50 for the proposed model 1, 3 and 5 for the pre-trained models

$$FN_{Covid19} = P_{NC} + P_{BC} + P_{VC}$$

The rest components of the confusion matrix can be explained and calculated based on the above. Using these values, the metrics can be calculated as given below:

$$Precision = \frac{TP}{FP + TP}$$

$$Recall = \frac{TP}{TP + FN}$$

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

The accuracy has been calculated on the basis of the class-specific values where only the total true positives are divided by the total number of samples in the test set. As a result, the same accuracy value has been calculated for each class using the formula below.

$$accuracy = \frac{P_{CC} + P_{NN} + P_{BB} + P_{VV}}{\#samples\ in\ the\ test\ set}$$

3.5 Hyperparameters tuning

Adam optimizer (Kingma and Ba 2023) was used for training all the models mentioned in this work. The exponential

decay rate (beta 1 and beta 2) for the first and second moment estimates were set to default values of 0.9 and 0.999, respectively. Learning rate was chosen as 0.001. Batch size of 32 and 64 were chosen for all the experiments. Number of epochs was chosen as 50 for the proposed model while it was 1, 3 and 5 for the pre-trained models. The number of epochs was not increased because no improvement in the performance was observed. Table 7 shows these hyperparameter values.

The codes were written in Python (version 3.6.13) language using Tensorflow and Keras libraires. The execution of the codes was performed on Radeon RX 580 GPU.

4 Experimental results

4.1 Transfer learning results

The models mentioned in Sect. 3.3 were trained and confusion matrix was generated for each experiment. Table 8 shows the best result achieved for each model with corresponding confusion matrix, prediction accuracy, precision, recall and F1-score. These transfer learning experiments performed on the widely used benchmark models allows for understanding the best performing model together with the appropriate epochs and batch size.

As shown in Table 8, Efficient-Net B2 model achieved the best overall prediction accuracy. The model was the best to predict the images labeled as COVID-19 with precision of 0.99 and recall of 0.98. It is possible to consider this model as the most appropriate one for identifying COVID-19 images. The model also showed a good performance in predicting the images labeled as Normal (healthy). The performance of the model declined when predicting Bacterial and Viral Pneumonia where the number of misclassifications were high. ResNetV2_152 also achieved second best prediction accuracy; however, the model did not achieve satisfying results in COVID-19 prediction when compared to Efficient-Net B2 model. On the other hand, ResNetV2_152 was the best among the other models in detecting Viral Pneumonia. With regard to Bacterial Pneumonia, InceptionResNetV2 achieved the best accuracy in detecting this class; however,

Table 8 Prediction results of the benchmark models

Model	Epoch	Batch size	Label	Precision	Recall	F1-Score	Accuracy	Confusion matrix			
EfficientNet B0	3	32	COVID-19	0.98	0.97	0.98	84.8%	249	5	2	0
			Normal	0.92	0.94	0.94	1	632	13	8	
			Pneumonia Bacterial	0.76	0.88	0.82	2	26	530	42	
			Pneumonia Viral	0.75	0.46	0.57	1	24	155	151	
EfficientNet B2	5	32	COVID-19	0.99	0.98	0.99	85.7%	251	3	1	1
			Normal	0.94	0.96	0.95	0	630	8	16	
			Pneumonia Bacterial	0.78	0.86	0.82	1	17	516	66	
			Pneumonia Viral	0.69	0.55	0.61	0	17	133	181	
InceptionV3	5	64	COVID-19	0.98	0.96	0.97	85.0%	245	6	5	0
			Normal	0.92	0.97	0.94	1	633	5	15	
			Pneumonia Bacterial	0.79	0.84	0.81	3	26	503	68	
			Pneumonia Viral	0.69	0.55	0.61	0	24	124	183	
InceptionResNetV2	5	32	COVID-19	0.98	0.96	0.97	83.5%	245	1	8	2
			Normal	0.94	0.95	0.95	2	622	8	22	
			Pneumonia Bacterial	0.74	0.89	0.81	3	18	532	47	
			Pneumonia Viral	0.66	0.42	0.51	0	20	173	138	
MobileNetV2	5	64	COVID-19	0.99	0.96	0.97	84.6%	245	6	1	4
			Normal	0.93	0.98	0.95	0	642	0	12	
			Pneumonia Bacterial	0.80	0.81	0.80	1	20	484	95	
			Pneumonia Viral	0.63	0.56	0.59	1	25	118	187	
NASNetMobile	3	32	COVID-19	0.96	0.96	0.96	84.0%	245	5	6	0
			Normal	0.91	0.95	0.93	6	620	15	13	
			Pneumonia Bacterial	0.76	0.87	0.81	1	26	522	51	
			Pneumonia Viral	0.71	0.48	0.57	2	28	142	159	
ResNetV2_152	5	64	COVID-19	0.98	0.94	0.96	85.5%	241	7	5	3
			Normal	0.93	0.98	0.95	2	638	2	12	
			Pneumonia Bacterial	0.81	0.81	0.81	3	22	487	88	
			Pneumonia Viral	0.67	0.63	0.65	1	16	106	208	
VGG16	3	32	COVID-19	0.93	0.83	0.88	79.4%	212	9	34	1
			Normal	0.88	0.92	0.90	9	604	23	18	
			Pneumonia Bacterial	0.71	0.84	0.77	3	40	503	54	
			Pneumonia Viral	0.66	0.43	0.52	4	36	149	142	
VGG19	5	64	COVID-19	0.95	0.91	0.93	79.5%	233	14	9	0
			Normal	0.82	0.97	0.89	3	637	10	4	
			Pneumonia Bacterial	0.72	0.84	0.77	3	69	505	23	
			Pneumonia Viral	0.77	0.27	0.40	5	55	182	89	

The best values are shown in bold

the accuracy of classifying Viral Pneumonia drops sharply in this model.

On the other hand, VGG models achieved the lowest overall accuracy. These two models (especially VGG16) were not able to predict COVID-19 images properly and give very low sensitivity in predicting Viral Pneumonia class. The large number of parameters and very deep structure of these two models may be the reason for these models to perform poorly. This means that such models with large number of parameters may not be always feasible for problems with a relatively small number of classes.

It is notable from Table 8 that most of the misclassifications belong to Bacterial and Viral Pneumonia classes and this led to a decrease in the overall accuracy of the models. It is possible that the combining these two classes together and making the classification triple (i.e. Pneumonia, COVID-19 and Normal) instead of quadruple will give a higher accuracy as many studies showed. However, separating these two classes can give a better overview about the ability of these models to identify the lung diseases and surely will give more specific diagnosis.

4.2 The proposed model results

The number of parameters in the proposed model is significantly lower than the transfer learning models (see Table 5). In addition, the model has lower number of layers (lower depth). This can highlight the effect of the depth and the number of parameters on the prediction results.

The proposed model was trained from scratch (unlike transfer learning) with batch sizes of 32 and 64, respectively. The number of epochs was chosen as 50 because no improvement in the training accuracy was observed after this number. The images were resized to $224 \times 224 \times 2$. The weights belonging to epochs that gave highest validation accuracy were used for testing the model. The corresponding detailed results are given in Tables 9 and 10.

Table 9 shows that the proposed model with batch size of 64 achieved the best prediction accuracy of 89.89%. This result indicates that the proposed model was better in detecting the diseases than the transfer learning models given in Table 5. In particular, the proposed model was better in detecting Viral and Bacterial Pneumonia classes where the precision and recall values are higher than those in the transfer learning models.

On the other hand, when the performance values in Tables 8 and 10 are compared, it is possible to observe that Efficient-Net B2 and B0 models were a slightly better in predicting COVID-19 class than the proposed model. The precision and recall values related to detecting COVID-19 in Efficient-Net B2 model were 0.99 and 0.98, respectively, while it was 0.98 and 0.96, respectively, in the proposed model.

As mentioned before, the proposed model has relatively lower number of parameters and layers than the pre-trained models. Apparently, this relatively low amount was sufficient for the model to extract good features and perform the classification task; therefore, it gave a better overall accuracy. In contrast, a high number of layers, as the case in benchmark models, may be causing a negative effect on the classification task that has lower number of classes.

The average prediction and training times of the benchmark and the proposed models for one single image were also calculated and compared (Table 11). Among the benchmark models, EfficientNet B0 and InceptionResNetV2 are the most and least time-consuming models, respectively in the training process. With regard to the prediction time, VGG16 and InceptionResNetV2 are the most and least time-consuming models, respectively. When the time consumption of the proposed model is compared with the benchmark models, it is notable that the proposed model is significantly better in training and prediction phases.

4.3 Ablation study

Ablation study was conducted on the proposed model for better understanding of the network's behavior and to justify the robustness. Ablation in machine learning means to delete part of the network and train the model again to check the function or effect of the deleted layer on the overall performance. For this purpose, one convolutional layer was deleted from the network sequentially and the results were recorded. Table 12 shows the details of the best prediction accuracy after deleting each layer separately.

By comparing the results in Table 12 with the results of the proposed model in Table 9, it is notable that the prediction accuracy decreases when the models in the ablation study are used. As for the confusion matrix, the entries corresponding to false predictions are higher in general. In addition, the gap between the training and validation accuracies is increased during the ablation study. This means that the model is prone to overfitting when some layers are removed. However, the difference appears clearly in the number of parameters, training and prediction times as Table 13 shows.

As expected, the comparing between Tables 5, 11 and 13 shows that sequentially ablating of the convolutional layers caused a significant increase in the number of parameters, thus an increase in the time required for training and prediction. This increase in the number of parameters did not cause an increase in the prediction accuracy as Table 12 shows. This proves that low number of parameters as in the raw model is enough to achieve the task of prediction.

4.4 Optimizer effect

Optimizers are algorithms used to update the weights of the neural networks to reduce the overall loss and increase the performance. The effect of using different optimizers on the detection performance of the proposed model was evaluated as well. For this purpose, different optimizers like Adaptive Gradients (AdaGrad) (Duchi et al. 2011) and Stochastic gradient descent (SGD) (Bottou 2012) were used. Table 14 shows the best prediction accuracy obtained by the proposed model after using SGD and AdaGrad optimizers.

The corresponding results of Table 14 show that the proposed model with batch size of 64 and Adagrad optimizer achieved a prediction accuracy of 89.13% which is slightly lower than the result obtained by Adam optimizer. On the other hand, SGD failed to achieve competitive prediction accuracy. The training accuracy with SGD optimizer did not exceed 87.32% after 50 epochs of training, which means slow convergence.

In general, Adagrad optimizer frequently updates the learning rate for each iteration depending on the change in the parameters during the training process. This feature maybe led to better accuracy than SGD optimizer.

Table 9 Prediction performance of the proposed model

Batch size	Epoch No	Training accuracy	Validation accuracy	Confusion matrix				Prediction accuracy	
32	36	96.26%	96.88%	248	2	3	3	89.19%	
					2	636	0		16
					2	9	543		64
					0	6	110		215
64	38	97.26%	93.75%	247	1	7	1	89.89%	
					1	640	3		10
					2	8	555		35
					1	10	107		213

Table 10 Performance metrics of the proposed model

Batch size	Classes	Precision	Recall	F1 Score
32	Covid-19	0.98	0.97	0.98
	Normal	0.97	0.97	0.97
	Pneumonia Bacterial	0.83	0.91	0.86
	Pneumonia Viral	0.77	0.65	0.70
64	Covid-19	0.98	0.96	0.97
	Normal	0.97	0.98	0.97
	Pneumonia Bacterial	0.83	0.93	0.87
	Pneumonia Viral	0.82	0.64	0.72

In comparison between the confusion matrix of Adam optimizer in Table 9 and Adagrad optimizer in Table 14, it is notable that the performance of Adagrad optimizer is better in detecting Viral Pneumonia disease. However, using Adam optimizer led to better results in classifying the rest of the classes and gave better overall accuracy.

5 Discussion

The results show that the proposed model is able to outperform the benchmark models in detecting the lung diseases. The proposed model achieved an overall accuracy of 89.89% while Efficient-net B2, the best among the benchmark models, had an overall accuracy of 85.7%.

Figure 5 shows that the benchmark models Efficient-Net B0 and Efficient-Net B2 were slightly better in detecting COVID-19 than the proposed model. In addition, MobileNetV2 model was a bit more accurate in detecting Normal class. On the other hand, the proposed model was much better in detecting Bacterial Pneumonia class than the benchmark models and a little better in detecting Viral Pneumonia class.

By checking the confusion matrix for each of the proposed model and the benchmark models in Tables 8 and 9, it is notable that all these models have a significant decrease in the accuracy of detecting pneumonia classes (i.e. Viral and Bacterial Pneumonia) compared with detecting COVID-19 and Normal classes. The tables show that the accuracy of

Table 11 The training and prediction time for all the models

Model name	Average Training time for one epoch in seconds (batch size = 32)	Average Training time for one epoch in seconds (batch size = 64)	Prediction time for single image (in milliseconds)
VGG16	273 s	266 s	14 ms
VGG19	336 s	338 s	16 ms
MobileNetV2	263 s	257 s	21 ms
EfficientNet B0	209 s	224.4 s	28 ms
EfficientNet B2	414 s	414 s	43 ms
InceptionV3	301.2 s	297.5 s	30 ms
ResNetV2_152	368 s	355 s	47 ms
InceptionResNetV2	599 s	641 s	80 ms
NASNetMobile	286.3 s	249 s	68 ms
The proposed model	116 s	133 s	8 ms

The best values are shown in bold

Table 12 Best prediction accuracy after applying the ablation

Deleted layer	Epoch number	Training accuracy	Validation accuracy	Best prediction accuracy	Confusion matrix			
1st layer	38	98.22%	92.19%	89.46%	251	4	1	0
					2	634	6	12
					1	10	552	37
					3	9	117	202
2nd layer	49	98.49%	93.75%	88.21%	246	3	5	2
					1	634	4	15
					3	9	540	48
					4	8	120	199
3rd layer	42	98.14%	92.19%	89.46%	246	4	4	2
					1	632	8	13
					0	11	546	43
					3	10	102	216
4th layer	45	98.34%	92.19%	88.43%	252	3	1	0
					4	634	5	11
					3	13	533	51
					7	14	108	202
5th layer	37	97.39%	93.75%	89.51%	252	3	1	0
					2	633	5	14
					3	9	543	45
					3	14	99	215

predicting Viral Pneumonia is much less than the accuracy of predicting Bacterial Pneumonia. In addition, most of the misclassifications in these two classes are due to the confusion of the models in classifying the Viral Pneumonia as Bacterial Pneumonia, and vice versa. The reason for that may be associated with the small number of samples in Viral Pneumonia data compared with the number of samples in Bacterial Pneumonia data as shown in Table 2.

The imbalance in the data for these two types of classes (i.e. Bacterial and Viral Pneumonia) possibly led to a negative effect on training process and caused an inability in the models to differentiate between these two diseases. This limitation can be addressed if new chest X-ray images are obtained and added to the dataset as a future work. Obviously, obtaining and accessing the data is one of the difficulties that hamper the researchers, especially the medical data due to the privacy concerns.

Table 13 The consumed time after the ablation process

Deleted layer	Number of parameters	Training time (seconds)	Prediction time for single image (milliseconds)
1st layer	4,834,498	174	51
2nd layer	4,826,983	193	28
3rd layer	4,810,093	146	35
4th layer	4,770,693	144	37
5th layer	3,459,943	143	32

The low number of parameters is also an extra advantage that characterize the proposed model in this study. The number of parameters in the proposed model is around 1 million parameters while it is 7.7 million parameters in Efficient-Net B2 model that achieved the best prediction accuracy among the pre-trained models as Table 5 shows. The low number of parameters and layers in the proposed model leads to lower prediction and training time compared with the pre-trained models (Table 11). This means that the proposed model is faster, needs lower resources and more qualified to operate in places that do not have high computing power.

Table 5 also shows that the proposed model has lower parameters than MobileNetV2 model which designed to work with fewer operations. The low depth of the proposed model may cause a positive effect on the classification accuracy in chest X-ray images whereas, deeper models can cause a decay in the extracted features and thus lower accuracy in tasks with relatively low number of classes.

In the closest study to this work, Sait et al. (2021) (mentioned in the Table 1) used same dataset with same number and type of classes to check the ability of CNN to classify lung diseases. However, in their work, the authors did not use data augmentation techniques to increase the diversity of the dataset and did not verify the efficiency of the proposed model using test set. The results associated to both studies are summarized in Table 15.

Table 14 The results of using different types of optimizers

	Batch size	Epoch No	Training accuracy	Validation accuracy	Confusion matrix				Prediction accuracy
SGD	64	47	86.82%	89.06%	235	6	15	0	86.00%
					0	633	5	16	
					1	6	540	53	
					1	6	154	170	
Adagrad	64	28	96.68%	89.06%	244	6	4	2	89.13%
					1	634	3	16	
					2	12	519	67	
					1	10	82	238	

Table 15 shows that the proposed new model in this study has much lower number of parameters than those model (based on InceptionV3) proposed in Sait et al. (2021). This means that our model requires lighter computer resources and thus runs faster in terms of training. Another noticeable property of the other study is that the dataset was split into training and validation sets only; in other words, only validation set was utilized for final evaluation of the proposed method without test set. It is a well-known convention in such machine learning problems that the model performance

is assessed by evaluating it on a separated set of samples that are not used during training process. Using validation data, which is often used to optimize hyperparameters, to check the performance of the model mostly do not provide reliable results all the time.

Given that both works could not be compared using the test accuracy, the validation accuracy of the proposed model in this work has been included in Table 15. The proposed model in this study outperforms the other in terms of the validation accuracy.

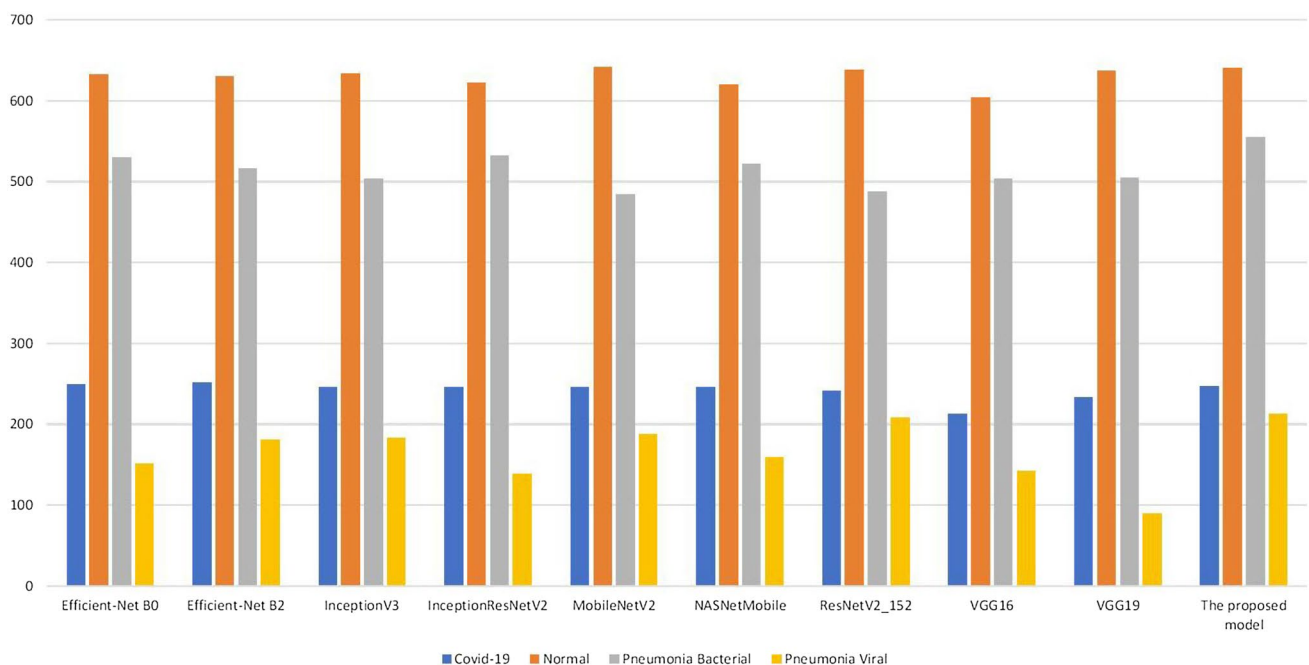


Fig. 5 Comparison of the number of true positives between the different models

Table 15 Comparison between this work and (Sait et al. 2021)

	Proposed a new model	Data augmentation	Dataset split	Validation accuracy	Test accuracy	Number of parameters
Sait et al. (2021)	Yes	No	Train and validation sets	91.3%	Not available	More than 21 million
This Study	Yes	Yes	Train, validation and test sets	93.75%	89.89%	1,043,558

6 Conclusion

In this work, a lightweight diagnosis model based on convolutional neural network was proposed to diagnose lung diseases like COVID-19, Bacterial and Viral Pneumonia. All experiments regarding the development and testing of the proposed model were carried out on a publicly available chest X-ray dataset. To validate and highlight the effectiveness of the proposed model, state of the art pre-trained CNN models were used for this prediction task and their corresponding performances were compared. Among these models, the pre-trained Efficient-Net B2 achieved the highest classification accuracy of 85.7%. On the other hand, the proposed model outperformed the pre-trained benchmark models by achieving an overall prediction accuracy of 89.89% with batch size of 64. A notably high accuracy in detection of COVID-19 samples was obtained in both the proposed model and the benchmark models; however, the pre-trained model Efficient-Net B2 showed a slightly better result in predicting COVID-19 with precision and recall of 0.99 and 0.98, respectively. In general, all the models used in this work showed a relatively poor precision for Viral Pneumonia class and confusion in distinguishing it from Bacterial Pneumonia class, and vice versa. This led to a decrease in overall accuracy.

The low number of data samples in Viral Pneumonia class may have hampered the models to extract better representations from the images content, thus obtaining a relatively low prediction performance for this class. This can be considered as the main limitation of the study. It is expected that supporting Viral Pneumonia class with more samples will improve the performance of the models.

Besides the performance of the proposed model, this study contributes to the related literature by submitting a model with significantly a low number of parameters. This advantage makes this model applicable in medical facilities and areas that do not have devices with high computational resources. Furthermore, the system can easily be integrated with a user interface on a regular computer and can be used by medical staff with no technical skills of computer usage.

Since such a deep learning-assisted diagnosis model is more suitable for computers with limited computational power, the model may be executed on edge devices or single board computers. Hence, the proposed study has another practical application possibility to be used as a part of

Internet of Things (IoT) systems. Provided that the decisions are made using the proposed model on an edge device, the application will have advantages like saving bandwidth and rapid assessment of the input image to perform diagnosis. Therefore, such applications may help more convenient diagnosis practices while preventing the spreading of viruses.

As a future work, image processing techniques like Contrast Limited Adaptive Histogram Equalization (CLAHE) can be applied to enhance the quality of the chest X-ray images used in this work. Also, ensemble methods may be utilized by considering class-specific correct detection rates of different classifier models. In addition, providing new images to Viral Pneumonias class can be considered to achieve the balance in the data and increase the capability of the models to extract better representations.

Author contributions MH: methodology, coding, writing. EA: methodology, supervision, writing.

Data availability The data used in this study is publicly available at the following link: <https://data.mendeley.com/datasets/9xkhgts2s6/4>.

Declarations

Conflict of interest None.

References

- Agarwal A, Patni K (2021) Rajeswari D Lung cancer detection and classification based on alexnet CNN. In: 2021 6th International Conference on Communication and Electronics Systems (ICCES). IEEE.
- Alqudah AM, Qazan S, Alqudah A (2020) Automated systems for detection of COVID-19 using chest X-ray images and lightweight convolutional neural networks.
- Al-Karawi A (2022) Stacked cross validation with deep features: a hybrid method for skin cancer detection. *Tehnički Glasnik* 16(1):33–39
- Antonchuk J, et al. (2021) COVID-19 Pneumonia and Influenza Pneumonia Detection Using Convolutional Neural Networks. arXiv preprint [arXiv:2112.07102](https://arxiv.org/abs/2112.07102).
- Avşar E (2021) Effects of Image Preprocessing on the Performance of Convolutional Neural Networks for Pneumonia Detection. In: 2021 International Conference on INnovations in Intelligent Systems and Applications (INISTA). IEEE.
- Atitallah SB, et al. (2023) Randomly Initialized Convolutional Neural Network for the Recognition of COVID-19 using X-ray Images. arXiv preprint [arXiv:2105a.08199](https://arxiv.org/abs/2105a.08199), 2021.

- Bottou L (2012) Stochastic gradient descent tricks. *Neural networks: Tricks of the trade*. Springer, pp 421–436
- Bolhassani M (2021) Transfer learning approach to Classify the X-ray image that corresponds to corona disease Using ResNet50 pre-trained by ChexNet. arXiv preprint arXiv:2105b.08382.
- Breve F (2021) COVID-19 Detection on chest x-ray images: a comparison of cnn architectures and ensembles. arXiv preprint arXiv:2111.09972.
- Cao G, Song W, Zhao Z (2019) Gastric cancer diagnosis with mask R-CNN. in 2019 11th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC). IEEE.
- Canayaz M (2021) MH-COVIDNet: Diagnosis of COVID-19 using deep neural networks and meta-heuristic-based feature selection on X-ray images. *Biomed Signal Process Control* 64:102257
- Chowdhury ME et al (2020) Can AI help in screening viral and COVID-19 pneumonia? *IEEE Access* 8:132665–132676
- Chest X-Ray Images (Pneumonia) (2023) <https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia>.
- CoronaHack -Chest X-Ray-Dataset (2023) <https://www.kaggle.com/datasets/praveengovi/coronahack-chest-xraydataset>.
- Cohen JP et al. (2023) Covid-19 image data collection: Prospective predictions are the future. arXiv preprint arXiv:2006.11988, 2020.
- COVID-19 X rays (2023) <https://www.kaggle.com/datasets/andrewmvd/convid19-x-rays>.
- COVIDx Dataset (2023) <https://github.com/lindawang/COVID-Net/blob/master/docs/COVIDx.md>.
- COVID-19 (2023) Radiography Database.
- COVIDx CXR-2 (2023) <https://www.kaggle.com/datasets/andyczhaov/covidx-cxr2>.
- Duchi J, Hazan E, Singer Y, Adaptive subgradient methods for online learning and stochastic optimization. *J Mach Learn Res* 2011. 12(7).
- El Asnaoui K, Chawki Y (2020) Using X-ray images and deep learning for automated detection of coronavirus disease. *Journal of Biomolecular Structure and Dynamics*: p. 1–12.
- El-Shafai, W. and F. Abd El-Samie, *Extensive COVID-19 X-Ray and CT chest images dataset*. Mendeley Data, V3. 2020.
- Gangopadhyay T et al (2022) MTSE U-Net: an architecture for segmentation, and prediction of fetal brain and gestational age from MRI of brain. *Netw Modeling Anal Health Inform Bioinform* 11(1):1–14
- Gunjan VK et al (2022) Detection of lung cancer in CT scans using grey wolf optimization algorithm and recurrent neural network. *Heal Technol* 12(6):1197–1210
- Hasoon JN et al (2021) COVID-19 anomaly detection and classification method based on supervised machine learning of chest X-ray images. *Results Phys* 31:105045
- He K, et al. (2016) Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*.
- Hemdan EE-D, Shouman MA, Karar ME (2020) Covidx-net: A framework of deep learning classifiers to diagnose covid-19 in x-ray images. arXiv preprint arXiv:2003.11055.
- Huang C, Yu Y, Qi M (2020) Skin Lesion Segmentation Based on Deep Learning. In: 2020 IEEE 20th International Conference on Communication Technology (ICCT). IEEE.
- Jaiswal AK et al (2021) Deep learning-based smart IoT health system for blindness detection using retina images. *IEEE Access* 9:70606–70615
- Kabiraj A, et al. (2022) Detection and classification of lung disease using deep learning architecture from x-ray images. In: *International Symposium on Visual Computing*. Springer.
- Kermany, D., K. Zhang, and M. Goldbaum (2018) Labeled optical coherence tomography (oct) and chest X-ray images for classification. *Mendeley data*. 2(2).
- Kingma DP, Ba J (2023) Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980, 2014.
- Liu J et al (2022) Deep feature fusion classification network (DFFC-Net): towards accurate diagnosis of COVID-19 using chest X-rays images. *Biomed Signal Process Control* 76:103677
- Makris A, Kontopoulos I, Tserpes K (2020) COVID-19 detection from chest X-Ray images using Deep Learning and Convolutional Neural Networks. in 11th Hellenic Conference on Artificial Intelligence.
- Maheen U, Malik KI, Ali G (2021) Comparative Analysis of Deep Learning Algorithms for Classification of COVID-19 X-Ray Images. arXiv preprint arXiv:2110a.09294.
- Medical Computer Vision (2023) <https://pyimagesearch.com/category/medical/>.
- Mikołajczyk A, Grochowski M (2018) Data augmentation for improving deep learning in image classification problem. In: 2018a international interdisciplinary PhD workshop (IIPhDW). IEEE.
- Mohammed MA, et al. (2022) Novel crow swarm optimization algorithm and selection approach for optimal deep learning COVID-19 diagnostic model. *Computational intelligence and neuroscience*
- Montalbo FJ (2022) Truncating fined-tuned vision-based models to lightweight deployable diagnostic tools for SARS-CoV-2 infected chest X-rays and CT-scans. *Multimedia Tools Appl* 81(12):16411–16439
- Montalbo FJP (2021) Truncating a densely connected convolutional neural network with partial layer freezing and feature fusion for diagnosing COVID-19 from chest X-rays. *MethodsX*: 101408.
- Nagi AT et al (2022) Performance Analysis for COVID-19 diagnosis using custom and state-of-the-art deep learning models. *Appl Sci* 12(13):6364
- Narin A, Kaya C, Pamuk Z (2021) Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks. *Pattern Analysis and Applications*: p. 1–14.
- Ohata EF et al (2020) Automatic detection of COVID-19 infection using chest X-ray images through transfer learning. *IEEE/CAA J Automatica Sinica* 8(1):239–248
- Pal D, Reddy PB, Roy S (2022) Attention UW-Net: a fully connected model for automatic segmentation and annotation of chest X-ray. *Comput Biol Med* 150:106083
- Qi X et al (2021) Chest X-ray image phase features for improved diagnosis of COVID-19 using convolutional neural network. *Int J Comput Assist Radiol Surg* 16(2):197–206
- Rajinikanth V, et al. Pneumonia Detection in Chest X-ray using InceptionV3 and Multi-Class Classification. in 2022 Third International Conference on Intelligent Computing Instrumentation and Control Technologies (ICICICT). 2022. IEEE.
- Reynaldi D et al. COVID-19 Classification for Chest X-Ray Images using Deep Learning and Resnet-101. In: 2021c International Congress of Advanced Technology and Engineering (ICOTEN). 2021. IEEE.
- Rajinikanth, V., et al. UNet with Two-Fold Training for Effective Segmentation of Lung Section in Chest X-Ray. in 2022 Third International Conference on Intelligent Computing Instrumentation and Control Technologies (ICICICT). 2022. IEEE.
- Roy S, Meena T, Lim S-J (2022) Demystifying Supervised Learning in Healthcare 4.0: A New Reality of Transforming Diagnostic Medicine. *Diagnostics*, 12(10): p. 2549.
- Russakovsky O et al (2015) Imagenet large scale visual recognition challenge. *Int J Comput Vision* 115(3):211–252
- Salçin K (2019) Detection and classification of brain tumours from MRI images using faster R-CNN. *Tehnički Glasnik* 13(4):337–342
- Sandler M, et al. Mobilenetv2: Inverted residuals and linear bottlenecks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018.
- Sait U, et al. Curated Dataset for COVID-19 Posterior-Anterior Chest Radiography Images (X-Rays). *Mendeley Data*, 2020. 1.

- Sait U et al (2021) A deep-learning based multimodal system for Covid-19 diagnosis using breathing sounds and chest X-ray images. *Appl Soft Comput* 109:107522
- Shenoy V, Malik SB CovXR: Automated Detection of COVID-19 Pneumonia in Chest X-Rays through Machine Learning. arXiv preprint arXiv:2110b.06398, 2021.
- Shome D et al (2021) COVID-Transformer: interpretable COVID-19 detection using vision transformer for healthcare. *Int J Environ Res Public Health* 18(21):11086
- Simonyan K, Zisserman A (2023) Very deep convolutional networks for large-scale image recognition. arXiv preprint [arXiv:1409.1556](https://arxiv.org/abs/1409.1556), 2014.
- Szegedy C et al. Rethinking the inception architecture for computer vision. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
- Szegedy C et al. Inception-v4, inception-resnet and the impact of residual connections on learning. in *Thirty-first AAAI conference on artificial intelligence*. 2017.
- Tan M, Le Q Efficientnet: Rethinking model scaling for convolutional neural networks. in *International Conference on Machine Learning*. 2019. PMLR.
- Wang Z et al (2019) Breast cancer detection using extreme learning machine based on feature fusion with CNN deep features. *IEEE Access* 7:105146–105158
- Wang L, Lin ZQ, Wong A (2020) Covid-net: A tailored deep convolutional neural network design for detection of COVID-19 cases from chest x-ray images. *Sci Rep* 10(1):1–12
- Wang X, et al. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.
- Zayet S et al (2020) Clinical features of COVID-19 and influenza: a comparative study on Nord Franche-Comte cluster. *Microbes Infect* 22(9):481–488
- Zeiser FA, et al. Evaluation of Convolutional Neural Networks for COVID-19 Classification on Chest X-Rays. In: *Brazilian conference on intelligent systems*. 2021. Springer.
- Zoph B, et al. Learning transferable architectures for scalable image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.