ORIGINAL RESEARCH

COVINet: a convolutional neural network approach for predicting COVID-19 from chest X-ray images

Muhammad Umer[1] · Imran Ashraf[2] · Saleem Ullah[1] · Arif Mehmood[3] · Gyu Sang Choi[2]

Abstract

COVID-19 pandemic is widely spreading over the entire world and has established significant community spread. Fostering a prediction system can help prepare the officials to respond properly and quickly. Medical imaging like X-ray and computed tomography (CT) can play an important role in the early prediction of COVID-19 patients that will help the timely treatment of the patients. The x-ray images from COVID-19 patients reveal the pneumonia infections that can be used to identify the patients of COVID-19. This study presents the use of Convolutional Neural Network (CNN) that extracts the features from chest x-ray images for the prediction. Three filters are applied to get the edges from the images that help to get the desired segmented target with the infected area of the x-ray. To cope with the smaller size of the training dataset, Keras' ImageDataGenerator class is used to generate ten thousand augmented images. Classification is performed with two, three, and four classes where the four-class problem has X-ray images from COVID-19, normal people, virus pneumonia, and bacterial pneumonia. Results demonstrate that the proposed CNN model can predict COVID-19 patients with high accuracy. It can help automate screening of the patients for COVID-19 with minimal contact, especially areas where the influx of patients can not be treated by the available medical staff. The performance comparison of the proposed approach with VGG16 and AlexNet shows that classification results for two and four classes are competitive and identical for three-class classification.

1 Introduction

The coronavirus disease broke out in Wuhan China at the end of 2019 and infected thousands within the early few weeks. Initially, it was named the Wuhan virus but later the disease was title COrona VIrus Disease 2019 (COVID-19)

✉ Imran Ashraf
ashrafimran@live.com

Muhammad Umer
umersabir1996@gmail.com

Arif Mehmood
arif.mehmood@iub.edu.pk

[1] Department of Computer Science, Khwaja Fareed University of Engineering and Information Technology, Rahim Yar Khan, Pakistan

[2] Department of Information and Communication Engineering, Yeungnam University, Gyeongsan 38541, Korea

[3] Department of Computer Science & Information Technology, The Islamia University of Bahawalpur, Bahawalpur 63100, Pakistan

by the World Health Organization (WHO) Singhal (2020), Lai et al. (2020). COVID-19 is an ongoing pandemic that is spreading rapidly with every passing day and the cases have been reported from 213 countries and territories by the end of April 2020. Figure 1 shows total number of confirmed cases (2,810,325) and deaths (193,825) in all the countries affected from it Organization (2020). The suspected cases are those with signs are fever, sore throat, and cough which later turns into severe pneumonia. Study Chen et al. (2020a) reports 50% fever and 38% cough signs among the initial patients from Wuhan. Other reported symptoms are dyspnea, headache, and rhinorrhea Di Gennaro et al. (2020). However, the patients can be asymptomatic showing no signs in the initial days and still carry the disease.

The period of COVID-19 symptoms to death ranges from 6 to 41 days while the median is 14 days. The period depends on the age and immunity of the patient. For example, the study Wu et al. (2020) shows a short period for patients with age> 70 than those under the age of 70 years. The suspected patients are initially diagnosed using real-time fluorescence (RT-PCR), throat swabs, and secretions of
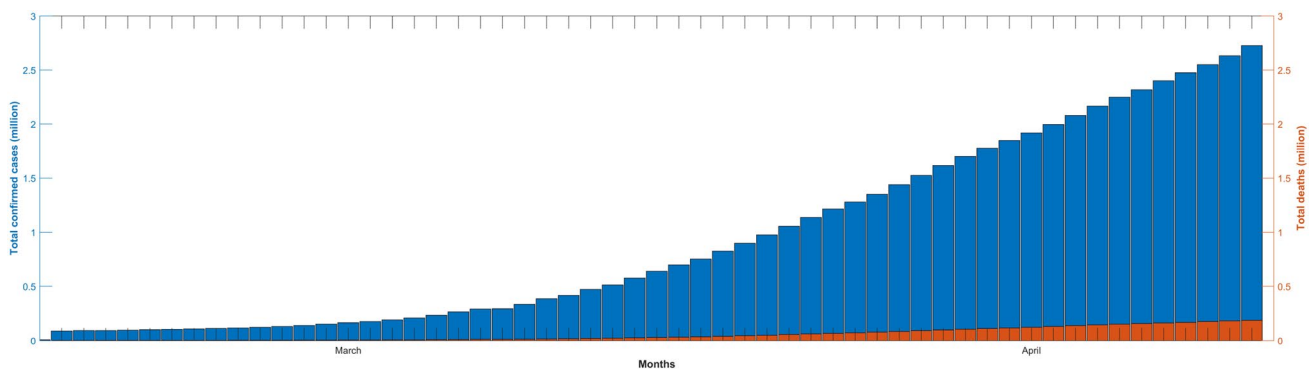
Fig. 1 COVID-19 cases for all the countries, left y-axis shows total confirmed cases in million, and right y-axis shows total number of deaths in million Organization (2020)

the lower tract samples Lippi et al. (2020). Chest Computed Tomography (CT) and chest X-ray are among the imaging findings that can help detect the abnormalities in the lungs of COVID-19 patients. The features extracted from CT and X-ray images are called radiological features.

Chest CT proves to be efficient for the detection of COVID-19 cases at the early stage and various imaging findings are reported in Kanne (2020), Kim (2020). Previous studies show that the imaging findings of radiological features can be very helpful for COVID-19 patients detection as patients have shown GGO lesions, bilateral patchy shadowing, and local patchy shadowing when CT is performed Huang et al. (2020a), Wang et al. (2020), Pan et al. (2020). Chest X-ray shows patchy shadowing, interstitial abnormalities, septal thickening, and crazy-paving pattern in COVID-19 patients Tu et al. (2020). Although the imaging findings from COVID-19 patients are non-conclusive, 17.9% of non-severe and 2.9% of severe cases did not have any CT or X-ray abnormality Chen et al. (2020c), X-ray can be used for early detection of the patients.

With the exponential growth of the malady, the available medical staff is unable to keep pace with the efficient diagnosis of the patients. Automate diagnosis of patients from radiological features is possible to lessen the burden of medical staff and increase their efficacy. Deep learning techniques have proven fruitful for disease detection from chest CT Li et al. (2020), Shi et al. (2020). Chest X-rays, however, have not been used so often for COVID-19 detection. This study aims at devising a Convolutional Neural Network (CNN) based model that can classify the patients into COVID-19 and normal pneumonia patients using chest X-rays. In summary, the proposed system makes the following contributions:

– A CNN base model is proposed that can accurately classify the patients into COVID-19 and pneumonia patients based on the chest X-rays of the patients

– The proposed approach is tested for three scenarios which involve two (COVID-19, normal people), three (COVID-19, normal people, and virus pneumonia) and four (COVID-19 and normal people, virus pneumonia, and bacterial pneumonia) class classification.
– Image preprocessing steps are defined that can help in accurate edge detection and segmentation of the infected area and thus increase the accuracy of the proposed model.
– An open dataset of COVID-19 patients containing chest X-rays is utilized to evaluate the performance of the proposed model and the results are compared with state-of-the-art deep learning-based approaches including VGG16 and AlexNet.
– Lack of training images is compensated with images generated augmented image data generator of Keras. Total images are 10,000 for training and testing the proposed approach.

The rest of the paper is organized in the following manner. Section 2 describes the research work related to the current study. The components and functionality of the proposed system are discussed in Sect. 3. Results are given in Sect. 4, while Sect. 5 contains the discussions and conclusion.

2 Related work

Deep learning is a well known research area in artificial intelligence. It provides promising results with end to end modeling with out manual feature engineering in medical image classification Umer et al. (2020), multi-label image classification He et al. (2020), text categorization Imtiaz et al. (2020), lung cancer detection Yamunadevi and Ranjani (2020), ECG classification Huang et al. (2020b), glaucoma diagnosis Ajesh et al. (2020) and athlete gesture tracking Long (2019). Since the start of COVID-19, researchers

start focusing on vaccine development, detection of SARS-CoV-2 using medical images, salivary specimen detection Bajaj et al. (2020), factors affecting mortality of physicians and nurses Jackson et al. (2020) and clinical feature analysis Zhao et al. (2020).

Research Li et al. (2020) presents an approach which utilizes CNN on the chest CT for detecting COVID-19 patients'. A deep learning model COVID-19 detection neural network (COVNet) is designed which extracts visual features from chest CT. The CT exams for community-acquired pneumonia and non-pneumonia CT exams are added in the dataset as well to evaluate the proposed model. Results show that the model sensitivity and specificity are 114 of 127 (90%) and 294 of 307 (96%) for detecting COVID-19 patients.

Segmentation is an important and pivotal step for machine learning-based approaches that aim to detect COVID-19 patients through imaging techniques. It delimits the infected areas called regions of interest (ROIs) that can be used for further processing and analysis. So many research works proposed deep learning-based approaches for CT segmentation for the quantification and prediction of COVID-19. The U-Net designed by Ronneberger et al. (2015) is a famous technique used in general-purpose segmentation. It has been adopted by many authors to segment COVID-19 patient's CT images. For example, authors in Zheng et al. (2020) used a pre-trained UNet to segment lung regions of CT images of the patients. A total of 499 and 131 CT images are used for training and testing with the proposed DeCoVNet which is a weakly-supervised deep learning model. The receiver operating characteristic curve (PR AUC) value is 0.975 for the tested CT images and the sensitivity and specificity values are larger than 0.9. Similarly, authors in Gozes et al. (2020) utilize deep learning approaches to classify COVID-19 and non-COVID-19 patients from CT images. Segmentation of ROIs is done using UNet while the classification of patients is achieved through the Reset-50 2D deep convolutional neural network He et al. (2016). Results are 0.996 AUC, 98.2% sensitivity and 92.2% specificity.

Another similar work that uses CT images to distinguish between COVID-19 and non-COVID-19 patients is Chen et al. (2020b). The proposed deep learning approach makes use of CT images of 51 confirmed COVID-19 patients, and 55 control patients from other diseases to train the model. Image segmentation is done using UNet++ and later a CNN is trained for classification Zhou et al. (2018). The proposed approach shows accuracy comparable to that of radiologists' and can considerably reduce the reading time of the radiologists. Authors in Jin et al. (2020) design a system that automatically analyzes the features from CT images to detect COVID-19 pneumonia features and help physicians in the classification of the patients. A training dataset comprising of 1,136 CT images (723 positives for COVID-19) is used for this purpose. The 3D U-Net++ is leveraged for image

segmentation while the classification is performed using ResNet He et al. (2016). The proposed approach achieves a sensitivity of 0.974 and a specificity of 0.922 for the used dataset.

The above-cited research works employ deep learning models on the CT images for COVID-19 detection. CT images are high-quality 3D images achieved from tomography. CT images are 3D images and contain hundreds of slices. It requires a substantial amount of time and computational resources to preprocess these images before we can put them to the training models. On the other hand, X-ray images are more common and easy to process than those of CT images. Hence various researchers proposed machine learning models that can work with X-ray images.

Authors in Narin et al. (2020) introduces three different models, i.e., ResNet50, InceptionV3, and Inception-ResNetV2 to classify COVID-19 patients from X-ray images. The models are trained on chest X-ray images of 50 COVID-19 patients and 50 normal people. The achieved accuracy is 98.0%, 97.0% and 87% for ResNet50, InceptionV3 and Inception-ResNetV2, respectively. A similar ResNet approach is presented in Zhang et al. (2020) whereby the trained model is used to classify the patients and detect the anomaly. Anomaly detection is used to improve COVID-19 classification. Classification is performed to separate COVID-19 patients from pneumonia patients. Results show the sensitivity of 96.0%, the specificity of 70.07%, and AUC of 0.952. Another deep learning model is worked out by authors in Wang and Wong (2020) for COVID-19 patient classification. The model i.e., COVID-Net is based on a deep CNN and uses X-ray images of 1203 healthy people, 931 bacterial pneumonia patients, 660 patients with viral pneumonia, and 45 patients confirmed for COVID-19. The testing accuracy of COVID-19 is 83.50%.

The studies that utilize X-ray images to classify COVID-19 patients and healthy subjects train on a small dataset of 45 to 70 images Shi et al. (2020). With the limited number of X-ray images, the robustness and accuracy of the proposed approach can not be determined conclusively. Also, the results can not be generalized with a smaller dataset. We, therefore, use Keras ImageDataGenerator class to augment images for increasing the number of X-ray images. Later we work out image pre-processing technique and a customized CNN model to increase the prediction accuracy for COVID-19 patients.

3 Materials and methods

This section provides the details of the proposed COVID-19 prediction approach, preprocessing phases, and the structure of the CNN used for prediction.

3.1 Description of dataset used for experiments

This study uses the X-ray datasets from two sources. Dataset-1 is available at Dataset (2020) and contains 79 images each for virus and bacterial pneumonia. Dataset 2 is available at Kaggle (2020) and it contains 78 X-ray images of COVID-19 patients and 28 images for normal people. Figure 2 shows a few sample images from the dataset. Further images are generated using Dataset-2 using the ImageDataGenerator class form Keras to a total of 10,000. The size of the training dataset can potentially influence the performance of deep learning models. Deep learning is a data intensive approach and requires a large amount of data for training. The smaller dataset of X-ray images is not appropriate to produce generalized results. Among the widely used methods are Keras ImageDataGenerator class and Generative Adversarial Neural Network (GAN). Authors in Shorten and Khoshgoftaar (2019) perform analysis of augmentation techniques for image data and point out that ImageDataGenerator is preferred over GAN. It is superior to GAN and avoids overfitting as well.

In the light of results given in Shorten and Khoshgoftaar (2019), this study uses the ImageDataGenerator class for generating more images Documentation (2018). Keras provided an image generator class that defines the configuration for image augmentation. Its capabilities include ' random rotation, shift, shear and flips', 'whitening' and 'dimension reordering', etc. Table 1 provides the
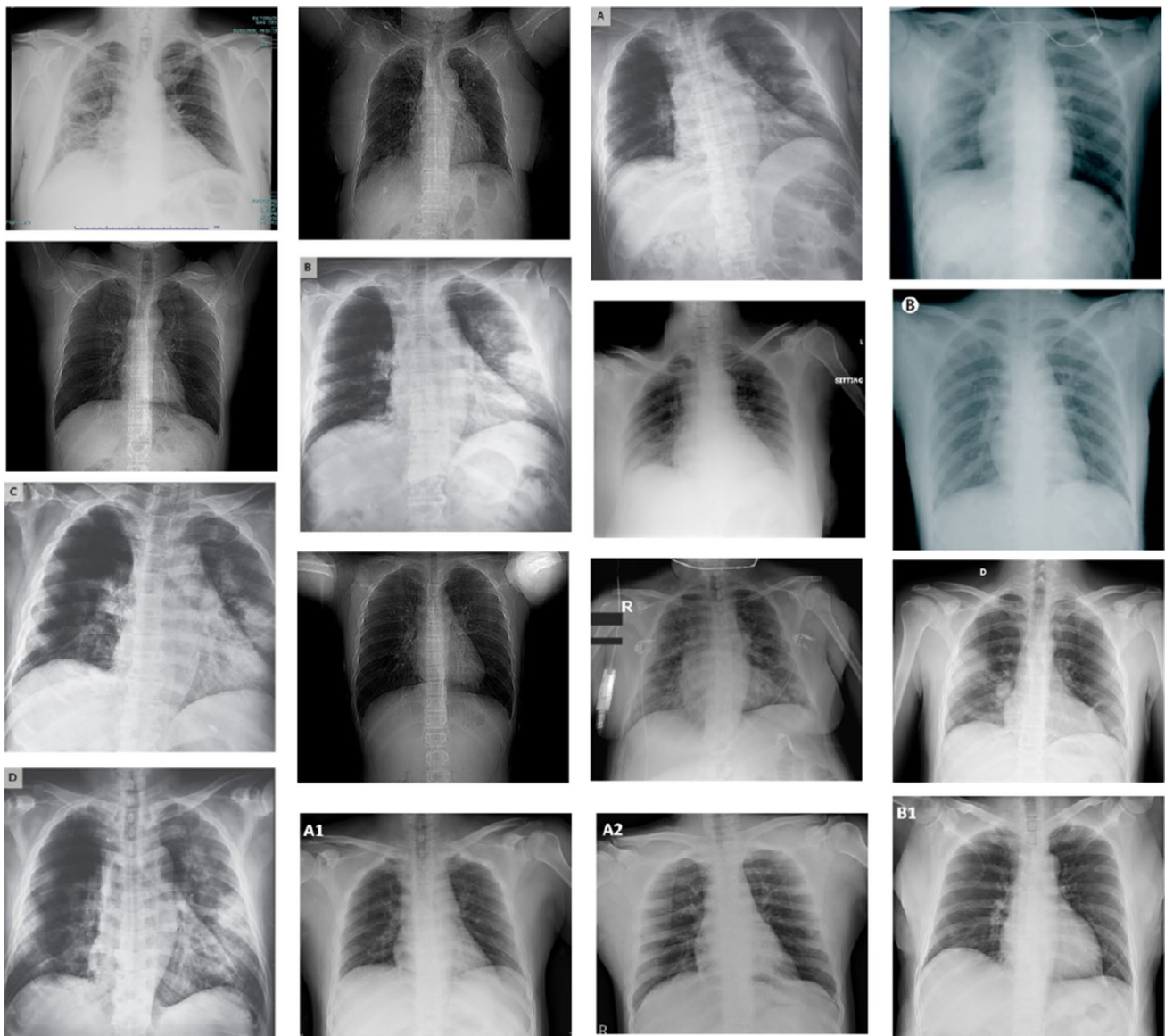

Fig. 2 Sample images from X-ray dataset. X-ray images come in different size

names and values of the parameters used in the current study.

3.2 Proposed approach

The proposed system utilizes X-ray images from the dataset. The architecture of the proposed approach is shown in Fig. 3. The proposed approach comprises of two modules: image preprocessing and CNN. These modules are described here in detail.
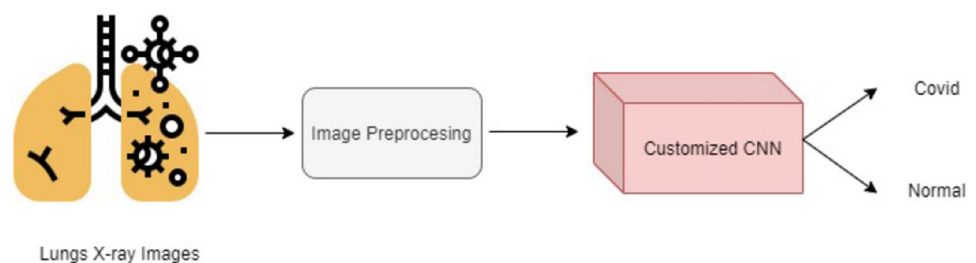
3.2.1 Image preprocessing

The preprocessing aims at removing the noise in X-ray images to improve the training process of CNN. Predominantly, input images are large which increases the training time. The first step is to reduce the size of the X-rays images. The size of X-ray images in the dataset is different for X-ray images as shown in Fig. 2. In the first step, we reduce this size to $120 \times 120 \times 3$ as shown in Fig. 4b. For edge detection, a value-based filter ([0,-1,0],[-1,6,-1],[0,-1,0]) is applied on the images which results in images with edges as shown in Fig. 4c. As the third step, Blue Green Red (BGR) image is converted to the luma component, blue projection, and red projection (YUV). It reduces the resolution of the U and V channels but keeps Y at full resolution. Because luminance is more important than color. And reducing U and V channels, the size of CNN can be reduced substantially. Figure 4d shows the results of BGR to YUV conversion. As the last step, we transform the YUV images back to BGR. It performs histogram normalization and smooths the edges.

3.2.2 The architecture of the proposed cnn

Deep learning-based approaches have shown superior performance than those of traditional machine learning approaches. Owing to their significant accuracy, deep learning-based models has attracted considerable attention during recent days. They have been applied in a large variety of domains like object detection, scene recognition, scene analysis, etc. Convolutional neural networks (CNN) have been specifically utilized for computer vision tasks. CNN comprises of a large number of convolutional, as well as, pooling and fully connected layers, each layer performing a different task. For example, the convolutional layer uses a fixed size filter called kernel to extract local features from the input image.
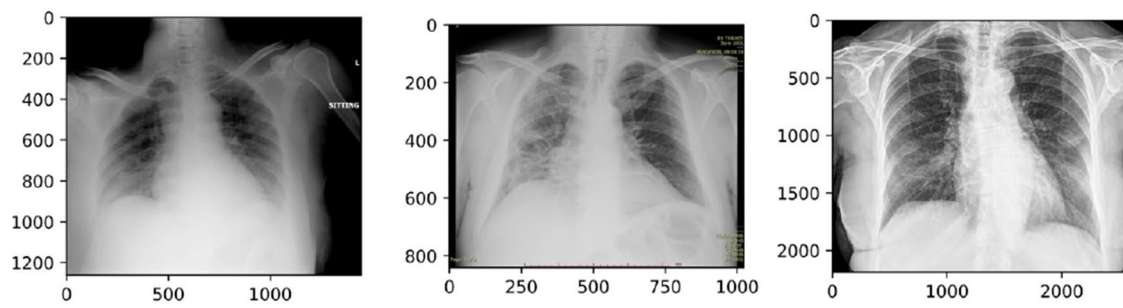
A new convolved image is obtained each time a convolution is applied. Each convolved image contains features that have been extracted from the image of the previous step. Let I(x, y) be a 2D input image and let f(x, y) be the 2D kernel applied for convolution, then the convolution is Nielsen (2015)

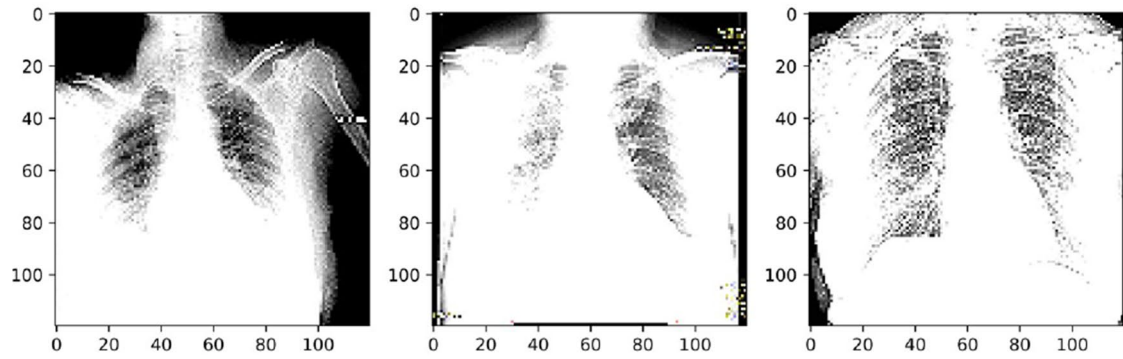$$y(i,j) = (I,f)(x,y) = \sum_{-\infty}^{\infty} \sum_{-\infty}^{\infty} I(x-u, y-v)f(u,v) \qquad (1)$$

When the convolution is applied, the pixel values at the edges can be ignored or padding can be applied. The output of the convolution can be transform using a nonlinear activation Patterson and Gibson (2017):

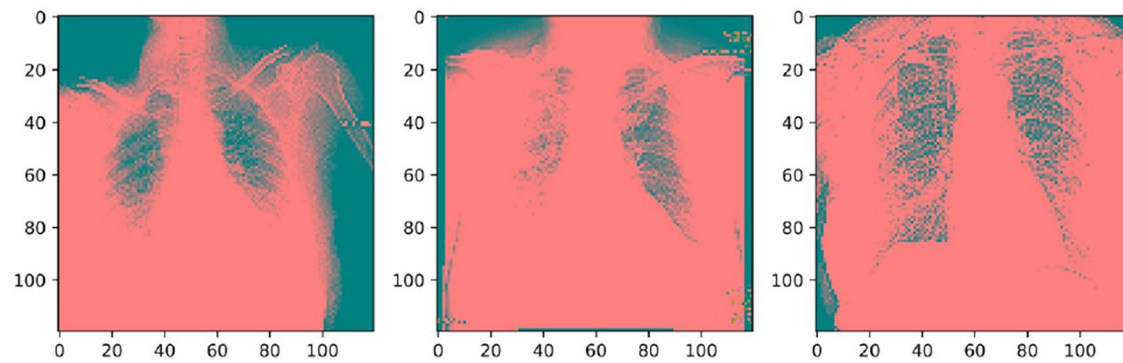$$sigmoid(x) = \frac{1}{1 + e^{-x}} \qquad (2)$$

Other than the convolutional layers, CNN contains pooling and fully connected layers. The pooling layer is used to summarize the local patches of convolutional layers. It subsamples the convolutional layer to reduce the size of the feature map. The pooling layer calculates the *maximu* and *average* function over the convolutional layer and are called *max pooling* and *average pooling* with respect to the function they perform. Spacing in the pixels of the image is used with pooling and is called *stride*. There is no activation function in pooling layers; they use a rectified linear unit (ReLU)
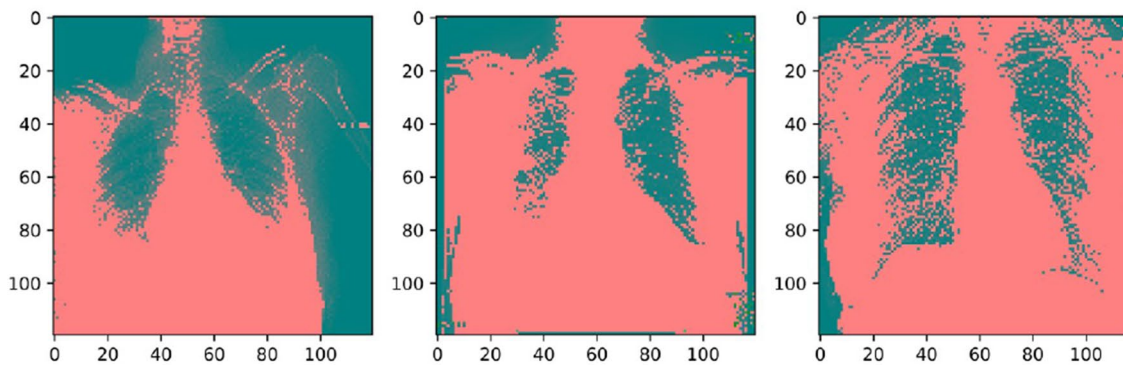
Table 1 Parameters used for 'ImageDataGenerator' to augment images

Paramter	Value
rotation_ range	30
zoom_ range	0.15
width_ shift_ range	0.2
height_ shift_ range	0.2
shear_ range	0.15
horizontal_ flip	True
fill_ mode	nearest

Fig. 3 The architecture of the proposed approach. The proposed approach uses the image preprocessing for X-ray images to remove noise. Later CNN is trained on the processed images for COVID-19 prediction

Lungs X-ray Images

(a)

(b)

(c)

(d)

instead. The pooling average for each convolutional layer can be calculated by Zhu et al. (2017):

$$X_{ij}^{[l]} = \frac{1}{MN} \sum_{m}^{M} \sum_{n}^{N} X_{iM+m,jN+n}^{[l-1]} \tag{3}$$

where i and j show the positions of the output map, while M and N are the pooling sample sizes.

Besides convolutional and pooling layers, fully connected layers are added in CNN to perform classification. The features from convolutional layers are given to the fully connected layer for classification. Fully connected layers have a different weight associated with each link and require substantial computing resources. Figure 5 shows the architecture of the proposed CNN used for the classification.

ReLU is used as activation with convolutional layers. Average pooling is used with a stride of 3 after the third convolutional layer. Dropout layers are used to prevent complex co-adaptations on the training data and avoid overfitting of the model. Originally, dropout layers were used with fully connected layers by Hinton et al. (2012), however, they have been used with convolutional layers as well Park and Kwak (2016). The last fully-connected layer uses the sigmoid function to output the prediction to 0 or 1. The standard sigmoid function Bishop (2006) is given as

$$S(t) = \frac{1}{1 + e^{-t}} \tag{4}$$

The details of the parameters used in the proposed CNN are given in Table 2.

3.3 Performance evaluation metrics

This study uses accuracy, precision, recall, F-score, AUC, sensitivity, and specificity as the performance evaluation metrics. These metrics are based on four terms, i.e., True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN). TP refers to patients that have a disease and the test is positive while FP are those patients who do not have the disease but the test is positive. Similarly, TN refers to those patients who do not have the disease and the test is negative and FN are the patients with disease but the test is negative. Based on these terms, sensitivity is calculated as

$$Sensitivity = \frac{TP}{TP + FN} \tag{5}$$

A 100% sensitivity indicates the classifier has correctly classified all the patients with the disease Lalkhen and McCluskey (2008). High sensitivity is important for detecting a serious disease. Specificity is calculated as

$$Specificity = \frac{TN}{TN + FP} \tag{6}$$

Both sensitivity and specificity do not consider a cut-off point for the test. A cut-off point affects the number of false negatives and false positives. For example, higher cut-off values result in higher false negatives and lower cut-off value raises false positives. The former indicates that the test highly specific but low sensitive while the latter shows a test that is highly sensitive but less specific. We have used AUC, i.e., the area under the receiver operating characteristic (ROC) to measure the discriminatory capability of the proposed model Jiménez-Valverde (2012). The model is considered to have a better discriminatory capacity if the AUC curve is higher than 0.5 Krzanowski and Hand (2009). Accuracy is one of the widely used metric to evaluate a classifiers performance and is calculate using

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{7}$$

Precision and recall are among the commonly used metrics for classifier performance evaluation. Precision measures the predicted positive cases which are positive and is calculated using

$$Precisioin = \frac{TP}{TP + FP} \tag{8}$$

Besides the above-mentioned metrics, F-score is measured as well. F score is a statistical measure used in classification. It considers precision and recall of a model/classifiers to compute a value between 0 and 1 indicative of classifiers lowest to highest performance Mining (2006). F score is calculated as

$$F = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{9}$$

4 Results and discussions

The proposed model is tested with three scenarios which involve a different number of classes as follows

– Scenario 1—Training and testing is performed with two classes, i.e., COVID-19 and Normal.
– Scenario 2—Three classes are used for training and testing, i.e., COVID-19, Normal, and Virus Pneumonia.
– Scenario 3—Training and testing is performed with four classes, i.e., COVID-19, Normal, Virus Pneumonia, and Bacterial Pneumonia.
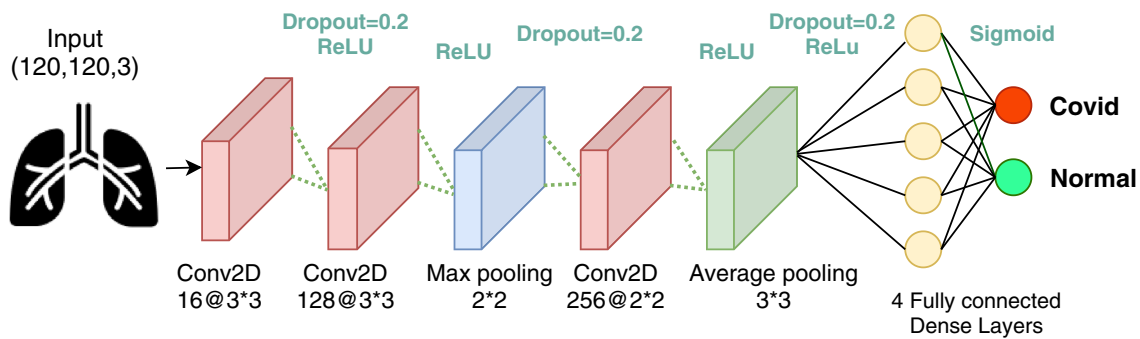
Fig. 5 The architecture of the proposed convolutinal neural network for COVID-19 patient classification

Table 2 Detail of the layers structure used in the proposed CNN model

Name	Description
Convolution	Filters=$(3 \times 3, @16)$, Strides=(1×1)
Convolution	Filters=$(3 \times 3, @128)$, Strides=(1×1)
Max pooling	Pool_ size=(2×2), Strides=(2×2)
Convolution	Filters=$(2 \times 2, @256)$, Strides=(1×1)
Average pooling	Pool_ size=(3×3), Strides=(1×1)
Layer	Flatten()
Fully connected	Dense (120 neurons)
Fully connected	Dense (60 neurons)
Fully connected	Dense (10 neurons)
Sigmoid	Sigmoid (2-class)

4.1 performance analysis of the proposed model

The proposed model is trained and tested with 10,000 X-ray images of COVID-19 and normal people combined for two classes. For three and four class problems, 79 X-ray images of virus and bacterial pneumonia each are added. The training and validation are done with 70% and 10% of the data while the testing data is 20% of the total data. Training is performed using a Tesla K80 Tensor Processing Unit (TPU) available at Google Colab. It provides 180 TFlops, 16 GB Random Access Memory (RAM), and 128 GB of disk space. Training took 1.5 h to run 12 epochs on the dataset for two classes. Figure 6a shows the accuracy of training and validation while Fig. 6 shows the curves for loss of training and validation.

Figure 7 shows the confusion matrix for the proposed approach. Besides two-class prediction, the proposed approach is tested with three and four classes as well to evaluate its performance. Table 3 shows the results for accuracy, precision, recall and F-score while Table 4 shows sensitivity, specificity and AUC values for two, three and four classes using the proposed approach. Results show that the accuracy with two classes is very good, i.e., 0.9721,
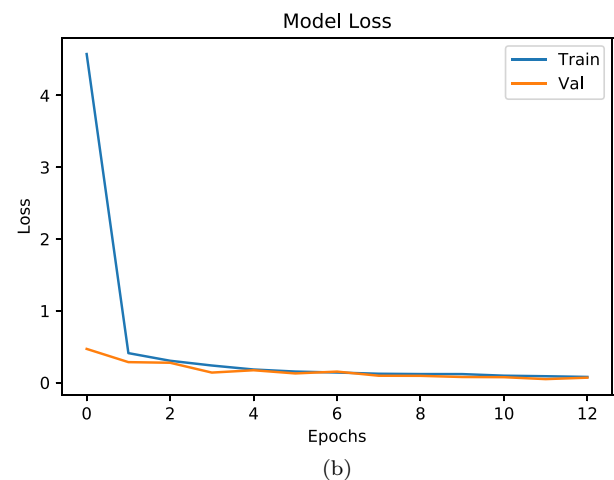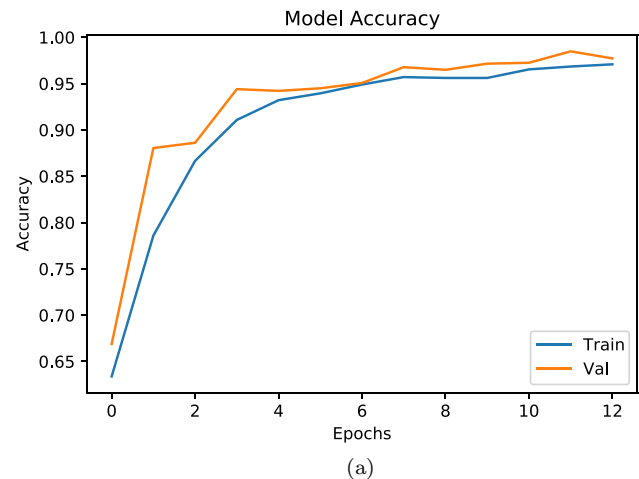

Fig. 6 Training and validation curves of the proposed approach, **(a)** Training and validation accuracy, **(b)** Loss curve for training and validation

however, it is reduced when we add X-ray images for more classes. Results for two-class classification indicated that the proposed approach can discriminate COVID-19 patients from normal people with high accuracy. The adopted image
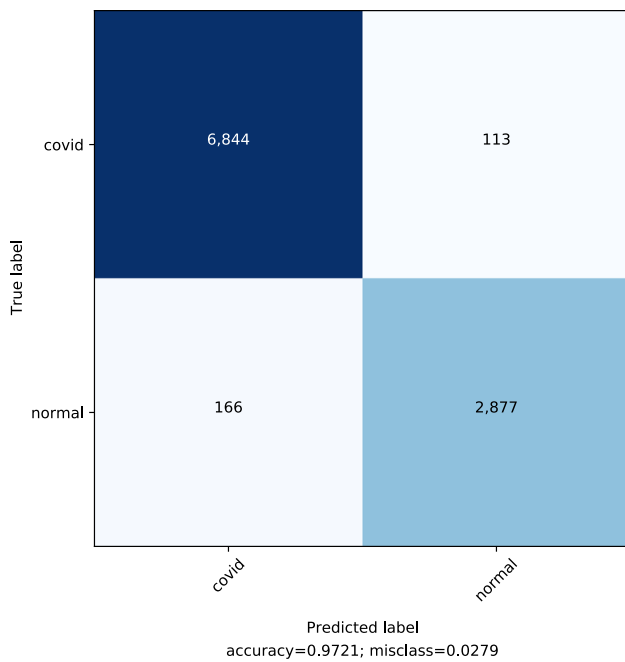
Fig. 7 Confusion matrix for the proposed approach

Table 3 Statistics for the performance of the proposed approach

No. of classes	Accuracy	Precision	Recall	F-score
2 classes	0.97210	0.98214	0.98761	0.97986
3 classes	0.89855	0.91410	0.97320	0.95512
4 classes	0.84762	0.89293	0.98994	0.93893

Table 4 Sensitivity, specificity and AUC of the proposed approach

No. of classes	Sensitivity	Specificity	AUC
2 classes	0.98761	0.96221	0.9817
3 classes	0.97320	0.94600	0.5541
4 classes	0.98994	0.92190	0.5948

preprocessing strategy helps training customized CNN to achieve that. Other indicators like precision, recall, and F-score is equally good. Accuracy for three-class classification has been reduced to 0.8986. However, it is still good considering the fact that the X-ray images from COVID-19 and virus pneumonia may have similarities. Additionally, the ribs overlying soft tissues, and low contrast can make the classification very challenging Zhang et al. (2020).

Precision, recall, and F-score values for three class scenarios are still higher than 0.91. Four class discrimination scenario has an accuracy of 0.8476 which is lower than that of the other two scenarios. The customized CNN extract biomarker features from X-ray images for training and

classification. X-ray images from COVID-19, virus pneumonia and bacterial pneumonia makes this process complicated which reduces the accuracy. Even then the accuracy is higher than other research works that used four classes with deep learning approaches. For example, the accuracy of CNN with four classes is 83.50% in Wang and Wong (2020).

4.2 comparison of models' performance with VGG16 and AlexNet

Two deep learning-based classification models are selected with whom the performance of the proposed model is compared, i.e. VGG16 and AlexNet. AlexNet outperformed previous models and won ImageNet Large Scale Visual Recognition Competition (ILSVRC) in 2012 Krizhevsky et al. (2012). Its structure is similar to LeNet but it is deeper having more filters with stacked convolutional layers Lecun et al. (1998). Approximately 60 million parameters and 650,000 neurons are trained in AlexNet to perform image classification when the input image belongs to one of 1000 different classes. Comprising of 11×11, 5×5, 3×3, convolutions, max pooling, dropout, data augmentation, it attaches ReLU to every convolutional and fully-connected layer. It shows that nonlinear ReLU can help in fast training of deep CNNs than that of using tanh or sigmoid. VGG16 is a CNN model that won ILSVRC in 2014 Simonyan and Zisserman (2015). Contrary to other models that focus on a large number of hyper-parameters, VGG16 adopts a different approach. It focuses on the use of convolutional layers of 3×3 filters with a stride of 1 and always use the same padding and max pool layer of 2×2 filters with a stride of 2. Two fully-connected layers are put at the end followed by a softmax for output. The 16 in VGG16 is indicative of 16 layers that have weights. Improvements are made through multiple 3×3 filters.

Figure 8 shows the confusion matrices for VGG16 and AlexNet for classification. It shows that the performance of the proposed approach is competitive with VGG16 for two and four class problems and identical when three classes are used for classification. AlexNet shows poor performance when used for COVID-19 and normal people X-ray images.

Table 5 shows the performance comparison of the proposed approach with VGG16 and AlexNet. VGG16 achieves the highest accuracy when training data from COVID-19 and normal people is used; the accuracy of the proposed model is marginally low. AlexNet performs poorly with two classes. The accuracy of Alexnet with 2 classes is much lesser than other classifiers because it starts overfitting when it reaches the accuracy of 85% during model training. Testing accuracy starts falling until it reaches to 66% at the end of 12 epochs. An important point worth mentioning here is that the proposed image preprocessing is used with VGG16 and AlexNet. If we change
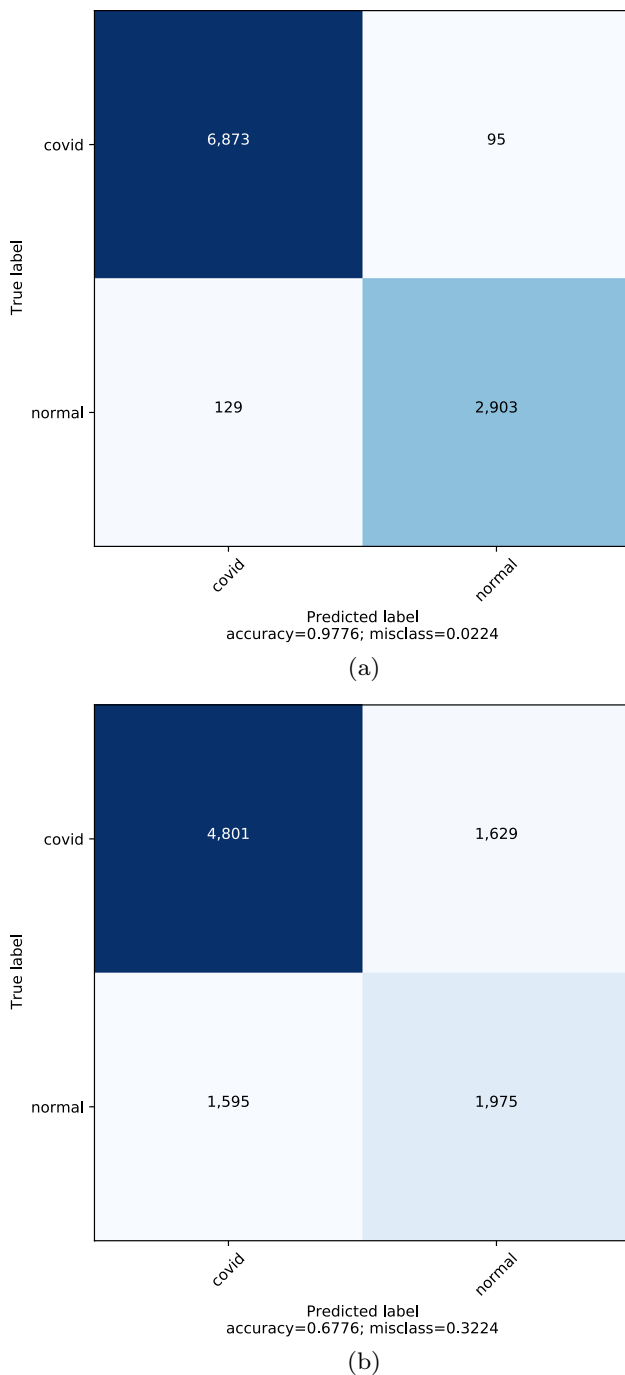
Fig. 8 Confusion matrices for techniques used for performance analysis, **(a)** Confusion matrix for VGG16, and **(b)** Confusion matrix for AlexNet

the preprocessing, the results for VGG16 and AlexNet degrade. Precision, recall, and F-score for the proposed approach are slightly lower than VGG16 but better than AlexNet. Accuracy of the proposed with three class problems is the same as that of VGG16 and AlexNet and so are the values for precision, recall, and F-score.

Table 5 Statistics for the performance metrics for the classifiers

Model	No. of classes	Accuracy	Precision	Recall	F-score
Proposed	2 classes	0.97210	0.98214	0.98761	0.97986
	3 classes	0.89855	0.91410	0.97320	0.95512
	4 classes	0.84762	0.89293	0.98994	0.93893
VGG16	2 classes	0.97768	0.99647	0.96478	0.98036
	3 classes	0.89855	0.91410	0.98876	0.95512
	4 classes	0.85714	0.89397	0.97427	0.94402
AlexNet	2 classes	0.67768	0.69254	0.90430	0.81834
	3 classes	0.89855	0.91410	0.94530	0.95512
	4 classes	0.85714	0.89397	0.96458	0.94402

When trained with data from COVID-19, normal people, virus pneumonia, and bacterial pneumonia X-ray images the accuracy of the proposed approach as well as, VGG16 and AlexNet is largely decreased. Accuracy of VGG16 and AlexNet is 0.85714 while the proposed approach is 0.00452 lower, i.e. 0.85262. The proposed approach can show similar performance for three and four class classification to that of VGG16 and AlexNet. Specificity, sensitivity, and AUC values are suggestive of good performance as well.

Table 6 demonstrates the results for sensitivity, specificity and AUC for the proposed as well as, VGG16 and AlexNet. AUC is an important factor to analyze the performance of a classifier to accurately discriminate between "diseased" and "non-diseased" patients Hajian-Tilaki (2013). AUC value for the proposed approach is marginally low for two classes than that of VGG16 and higher than that of AlexNet. For three-class classification, it is identical to VGG16 but lower than AlexNet. AUC value of the proposed approach for four-class problem is indicative of its good performance than those of both VGG16 and AlexNet.

Tables 5 and 6 indicate that accuracy of VGG16 is slight higher than the proposed approach, however, we need to consider the complex architecture of VGG16 and the training time it requires as well to compare its performance with

Table 6 Comparison of sensitivity, specificity and ACU for the classifiers

Model	No. of classes	Sensitivity	Specificity	AUC
Proposed	2 classes	0.98761	0.96221	0.98170
	3 classes	0.97320	0.94600	0.55410
	4 classes	0.98994	0.92190	0.59480
VGG16	2 classes	0.96478	0.96831	0.99900
	3 classes	0.98876	0.92420	0.55410
	4 classes	0.97427	0.92291	0.57280
AlexNet	2 classes	0.90430	0.54800	0.68470
	3 classes	0.94530	0.92420	0.57371
	4 classes	0.96458	0.92290	0.57180

Table 7 Statistics for required training time for classifiers

Model	Training time
Proposed approach	1.5 h
VGG16	3.25 h
AlexNet	2.5 h

the proposed approach. Table 7 shows the training time which is required for the proposed approach in comparison to VGG16 and AlexNet. Considering the difference of training time and accuracy, we can say that the proposed approach performs better than VGG16 and AlexNet to discriminate among healthy people, and patients from COVID-19, virus and bacterial pneumonia.

5 Conclusion

This study presents a convolutional neural network to discriminate COVID-19 patients from normal people using X-ray images. Deep learning is a data intensive approach but the dataset of COVID-19 patients is small which makes it very difficult to evaluate the robustness and generalizability of the deep learning-based models. To overcome this issue, Keras' ImageDataGenerator class is used to augment X-ray images. Image preprocessing is proposed which helps in the segmentation of the infected area in the X-ray images. Testing is performed with two, three, and four classes, i.e., COVID-19 patients, normal people, virus pneumonia, and bacterial pneumonia. Results indicate that the accuracy of the proposed approach is 0.97, 0.90, and 0.85 for two, three, and four classes, respectively. Precision, recall, and F-score values are also very good. The proposed approach has sensitivity and specificity of 0.98994, and 0.92190 respectively and AUC value is 0.5948 when four classes are used for training and testing. AUC values of VGG16 and AlexNet are 0.57280 and 0.57180 which are lower than that of the proposed approach for four-class classification.

The performance of the proposed approach is compared with VGG16 and AlexNet. The comparison indicates that the results of the proposed are marginally lower than that of VGG16 for two and four class problems while equal for three-class classification. AlexNet does not perform well for two-class, however, proves equally good for three and four class classification. Both VGG16 and AlexNet show the performance when used with the proposed image preprocessing strategy. The architecture of both VGG16 and AlextNet is complex and require higher training time than the proposed approach. The proposed approach is robust and produces good accuracy. If a higher number of X-ray images of COVID-19 patients are available, it is expected to improve the results further.

Compliance with ethical standards

Conflict of interest "The authors declare no conflict of interest. The funding agency had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results".

References

Ajesh F, Ravi R, G R, (2020) Early diagnosis of glaucoma using multi-feature analysis and dbn based classification. J Ambient Intell Human Comput. https://doi.org/10.1007/s12652-020-01771-z

Bajaj N, Granwehr BP, Hanna EY, Chambers MS (2020) Salivary detection of sars-cov-2 (covid-19) and implications for oral healthcare providers. Head & Neck

Bishop CM (2006) Pattern recognition and machine learning. springer, Berlin

Chen F, Liu Z, Zhang F, Xiong R, Chen Y, Cheng X, Wang W, Ren J (2020a) First case of severe childhood novel coronavirus pneumonia in china. Zhonghua er ke za zhi= Chinese Journal of Pediatrics 58:E005–E005

Chen J, Wu L, Zhang J, Zhang L, Gong D, Zhao Y, Hu S, Wang Y, Hu X, Zheng B, et al. (2020b) Deep learning-based model for detecting 2019 novel coronavirus pneumonia on high-resolution computed tomography: a prospective study. medRxiv

Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, Qiu Y, Wang J, Liu Y, Wei Y et al (2020c) Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. The Lancet 395(10223):507–513

Dataset (2020) Covid data. https://drive.google.com/uc?id=1coM7x3378f-Ou2l6Pg2wldaOI7Dntu1a, online: accessed 22 April 2020

Di Gennaro F, Pizzol D, Marotta C, Antunes M, Racalbuto V, Veronese N, Smith L (2020) Coronavirus diseases (covid-19) current status and future perspectives: a narrative review. Int J Environ Res Pub Health 17(8):2690

Documentation K (2018) Imagedatagenerator class. https://keras.io/preprocessing/image /#imagedatagenerator-methods, online: accessed 28 April 2020

Gozes O, Frid-Adar M, Greenspan H, Browning PD, Zhang H, Ji W, Bernheim A, Siegel E (2020) Rapid ai development cycle for the coronavirus (covid-19) pandemic: Initial results for automated detection & patient monitoring using deep learning ct image analysis. arXiv preprint arXiv:200305037

Hajian-Tilaki K (2013) Receiver operating characteristic (roc) curve analysis for medical diagnostic test evaluation. Caspian J Int Med 4(2):627

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

He T, Zhang L, Guo J, Yi Z (2020) Multilabel classification by exploiting data-driven pair-wise label dependence. Int J Intell Syst

Hinton GE, Srivastava N, Krizhevsky A, Sutskever I, Salakhutdinov RR (2012) Improving neural networks by preventing co-adaptation of feature detectors. arXiv preprint arXiv:12070580

Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Zhang L, Fan G, Xu J, Gu X et al (2020a) Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. The Lancet 395(10223):497–506

Huang JS, Chen BQ, Zeng NY, Cao XC, Li Y (2020b) Accurate classification of ecg arrhythmia using mowpt enhanced fast compression deep learning networks. J Ambient Intell Human Comput. https://doi.org/10.1007/s12652-020-02110-y

Imtiaz Z, Umer M, Ahmad M, Ullah S, Choi GS, Mehmood A (2020) Duplicate questions pair detection using siamese malstm. IEEE Access 8:21932–21942

Jackson D, Anders R, Padula WV, Daly J, Davidson PM (2020) Vulnerability of nurse and physicians with covid-19: Monitoring and surveillance needed. J Clin Nurs

Jiménez-Valverde A (2012) Insights into the area under the receiver operating characteristic curve (auc) as a discrimination measure in species distribution modelling. Global Ecolo Biogeogr 21(4):498–507

Jin S, Wang B, Xu H, Luo C, Wei L, Zhao W, Hou X, Ma W, Xu Z, Zheng Z, et al. (2020) Ai-assisted ct imaging analysis for covid-19 screening: Building and deploying a medical ai system in four weeks. medRxiv

Kaggle (2020) Covid-19 patients lungs x ray images 10000. https://www.kaggle.com/nabeelsajid917/covid-19-x-ray-10000-images, online: accessed 20 April 2020

Kanne JP (2020) Chest ct findings in 2019 novel coronavirus (2019-ncov) infections from wuhan, china: key points for the radiologist

Kim H (2020) Outbreak of novel coronavirus (covid-19): What is the role of radiologists?

Krizhevsky A, Sutskever I, Hinton G (2012) Imagenet classification with deep convolutional neural networks. Neural Inform Process Syst 25

Krzanowski WJ, Hand DJ (2009) ROC curves for continuous data. Crc Press

Lai CC, Shih TP, Ko WC, Tang HJ, Hsueh PR (2020) Severe acute respiratory syndrome coronavirus 2 (sars-cov-2) and corona virus disease-2019 (covid-19): the epidemic and the challenges. Int J Antimicrobial Agents p 105924

Lalkhen AG, McCluskey A (2008) Clinical tests: sensitivity and specificity. Continuing Education in Anaesthesia Critical Care & Pain 8(6):221–223

Lecun Y, Bottou L, Bengio Y, Haffner P (1998) Gradient-based learning applied to document recognition. Proc IEEE 86(11):2278–2324

Li L, Qin L, Xu Z, Yin Y, Wang X, Kong B, Bai J, Lu Y, Fang Z, Song Q, et al. (2020) Artificial intelligence distinguishes covid-19 from community acquired pneumonia on chest ct. Radiology p 200905

Lippi G, Simundic AM, Plebani M (2020) Potential preanalytical and analytical vulnerabilities in the laboratory diagnosis of coronavirus disease 2019 (covid-19). Clin Chem Lab Med (CCLM) 1(ahead-of-print)

Long T (2019) Research on application of athlete gesture tracking algorithms based on deep learning. Journal of Ambient Intelligence and Humanized Computing pp 1–9, https://doi.org/10.1007/s12652-019-01575-w

Mining WID (2006) Data mining: concepts and techniques. Morgan Kaufinann 10:559–569

Narin A, Kaya C, Pamuk Z (2020) Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks. arXiv preprint arXiv:200310849

Nielsen MA (2015) Neural Netw Deep Learn, vol 25. Determination press, San Francisco

Organization WH (2020) Coronavirus (covid-19). https://covid19.who.int/, online: accessed 27 April 2020

Pan F, Ye T, Sun P, Gui S, Liang B, Li L, Zheng D, Wang J, Hesketh RL, Yang L, et al. (2020) Time course of lung changes on chest ct during recovery from 2019 novel coronavirus (covid-19) pneumonia. Radiology p 200370

Park S, Kwak N (2016) Analysis on the dropout effect in convolutional neural networks. In: Asian Conference on Computer Vision, Springer, pp 189–204

Patterson J, Gibson A (2017) Deep learning: a practitioner's approach. "O'Reilly Media, Inc."

Ronneberger O, Fischer P, Brox T (2015) U-net: Convolutional networks for biomedical image segmentation. In: International conference on medical image computing and computer-assisted intervention, Springer, pp 234–241

Shi F, Wang J, Shi J, Wu Z, Wang Q, Tang Z, He K, Shi Y, Shen D (2020) Review of artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for covid-19. IEEE Rev Biomed Eng pp 1–1

Shi F, Wang J, Shi J, Wu Z, Wang Q, Tang Z, He K, Shi Y, Shen D (2020) Review of artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for covid-19. IEEE Rev Biomed Eng

Shorten C, Khoshgoftaar TM (2019) A survey on image data augmentation for deep learning. J Big Data 6:1–48

Simonyan K, Zisserman A (2015) Very deep convolutional networks for large-scale image recognition. In: International conference on learning representations

Singhal T (2020) A review of coronavirus disease-2019 (covid-19). Indian J Pediatrics pp 1–6

Tu YF, Chien CS, Yarmishyn AA, Lin YY, Luo YH, Lin YT, Lai WY, Yang DM, Chou SJ, Yang YP et al (2020) A review of sars-cov-2 and the ongoing clinical trials. Int J Mol Sci 21(7):2657

Umer M, Sadiq S, Ahmad M, Ullah S, Choi GS, Mehmood A (2020) A novel stacked cnn for malarial parasite detection in thin blood smear images. IEEE Access 8:93782–93792

Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, Wang B, Xiang H, Cheng Z, Xiong Y et al (2020) Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China. Jama 323(11):1061–1069

Wang L, Wong A (2020) Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest radiography images. arXiv preprint arXiv:200309871

Wu C, Chen X, Cai Y, Zhou X, Xu S, Huang H, Zhang L, Zhou X, Du C, Zhang Y, et al. (2020) Risk factors associated with acute respiratory distress syndrome and death in patients with coronavirus disease 2019 pneumonia in wuhan, china. JAMA Int Med

Yamunadevi M, Ranjani S (2020) Efficient segmentation of the lung carcinoma by adaptive fuzzyglcm (af-glcm) with deep learning based classification. J Ambient Intell Human Comput. https://doi.org/10.1007/s12652-020-01874-7

Zhang J, Xie Y, Li Y, Shen C, Xia Y (2020) Covid-19 screening on chest x-ray images using deep learning based anomaly detection. arXiv preprint arXiv:200312338

Zhao Y, Zhou J, Pan L, Zhang Y, Wang H, Wu W, He J, Chen J, Huang H (2020) Detection and analysis of clinical features of patients with different covid-19 types. J Med Virol

Zheng C, Deng X, Fu Q, Zhou Q, Feng J, Ma H, Liu W, Wang X (2020) Deep learning-based detection for covid-19 from chest ct using weak label. medRxiv

Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J (2018) Unet++: A nested u-net architecture for medical image segmentation. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, Springer, pp 3–11

Zhu Y, Ouyang Q, Mao Y (2017) A deep convolutional neural network approach to single-particle recognition in cryo-electron microscopy. BMC Bioinformatics 18(1):348

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