

Research Article

An Efficient CNN Model for COVID-19 Disease Detection Based on X-Ray Image Classification

Aijaz Ahmad Reshi ¹, Furqan Rustam ², Arif Mehmood ³, Abdulaziz Alhossan,^{4,5}
Ziyad Alrabiah,⁴ Ajaz Ahmad ⁴, Hessa Alsuwailem,⁴ and Gyu Sang Choi ⁶

¹Department of Computer Science, College of Computer Science and Engineering, Taibah University, Al Madinah Al Munawarah, Saudi Arabia

²Department of Computer Science, Khwaja Fareed University of Engineering and Information Technology, Rahim Yar Khan 64200, Pakistan

³Department of Computer Science & IT, The Islamia University of Bahawalpur, Bahawalpur, Punjab 63100, Pakistan

⁴Department of Clinical Pharmacy, College of Pharmacy, King Saud University, Riyadh 11451, Saudi Arabia

⁵Corporate of Pharmacy Services, King Saud University Medical City, Riyadh, Saudi Arabia

⁶Department of Information & Communication Engineering, Yeungnam University, Gyeongbuk 38541, Republic of Korea

Correspondence should be addressed to Aijaz Ahmad Reshi; aijazonnet@gmail.com

Received 2 December 2020; Revised 8 February 2021; Accepted 25 April 2021; Published 17 May 2021

Academic Editor: Átila Madureira Bueno

Copyright © 2021 Aijaz Ahmad Reshi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Artificial intelligence (AI) techniques in general and convolutional neural networks (CNNs) in particular have attained successful results in medical image analysis and classification. A deep CNN architecture has been proposed in this paper for the diagnosis of COVID-19 based on the chest X-ray image classification. Due to the nonavailability of sufficient-size and good-quality chest X-ray image dataset, an effective and accurate CNN classification was a challenge. To deal with these complexities such as the availability of a very-small-sized and imbalanced dataset with image-quality issues, the dataset has been preprocessed in different phases using different techniques to achieve an effective training dataset for the proposed CNN model to attain its best performance. The preprocessing stages of the datasets performed in this study include dataset balancing, medical experts' image analysis, and data augmentation. The experimental results have shown the overall accuracy as high as 99.5% which demonstrates the good capability of the proposed CNN model in the current application domain. The CNN model has been tested in two scenarios. In the first scenario, the model has been tested using the 100 X-ray images of the original processed dataset which achieved an accuracy of 100%. In the second scenario, the model has been tested using an independent dataset of COVID-19 X-ray images. The performance in this test scenario was as high as 99.5%. To further prove that the proposed model outperforms other models, a comparative analysis has been done with some of the machine learning algorithms. The proposed model has outperformed all the models generally and specifically when the model testing was done using an independent testing set.

1. Introduction

The virus called the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) had been discovered in late 2019. The virus which originated in China became a cause of a disease known as Corona Virus Disease 2019 or COVID-19. The World Health Organization (WHO) declared the disease as a pandemic in March 2020 [1, 2]. According to the reports issued and updated by global healthcare authorities

and state governments, the pandemic affected millions of people globally. The most serious illness caused by COVID-19 is related to the lungs such as pneumonia. The symptoms of the disease can vary and include dyspnea, high fever, runny nose, and cough. These cases can most commonly be diagnosed using chest X-ray imaging analysis for the abnormalities [3].

X-radiation or X-ray is an electromagnetic form of penetrating radiation. These radiations are passed through

the desired human body parts to create images of internal details of the body part. The X-ray image is a representation of the internal body parts in black and white shades. X-ray is one of the oldest and commonly used medical diagnosis tests. Chest X-ray is used to diagnose the chest-related diseases like pneumonia and other lung diseases [4], as it provides the image of the thoracic cavity, consisting of the chest and spine bones along with the soft organs including the lungs, blood vessels, and airways. The X-ray imaging technique provides numerous advantages as an alternative diagnosis procedure for COVID-19 over other testing procedures. These benefits include its low cost, the vast availability of X-ray facilities, noninvasiveness, less time consumption, and device affordability. Thus, X-ray imaging may be considered a better candidate for the mass, easy, and quick diagnosis procedure for a pandemic like COVID-19 considering the current global healthcare crisis.

Deep learning and ANNs have endorsed an exponential research focus over the last decade. The deep ANNs have outperformed other conventional models on many essential benchmarks. Thus, ANNs have generally proved to be the state-of-the-art technology across a wide range of application areas, including NLP, speech recognition, image processing, biological sciences, and other commercial as well as academic areas. The advancement of ANNs has massive potential in healthcare applications, specifically in medical data analysis, diagnosis through medical image processing, and analysis. As seen in recent times, various parts of the world face the healthcare crisis both in terms of the needed number of healthcare professionals and testing equipment. Considering the present pandemic situation, there is an appurtenant relationship between the detection of COVID-19 cases and chest X-ray image analysis and classification. In this work, an automatic diagnostic system has been developed using CNN which uses chest X-ray analysis results to diagnose whether a person is COVID-19-affected or normal. Preliminary analysis of this study has shown promising results in terms of its accuracy and other performance parameters to diagnose the disease in a cost-effective and time-efficient manner. This study used CNN with extra layers to improve the COVID-19 X-ray image classification accuracy. In neural networks, the CNN structure is specially designed to process the two-dimensional image tasks although it can also be used in one- and three-dimensional data. CNN is a type of DNN, inspired by the visual system of the human brain, and is most commonly used in the analysis of visual imagery. To train the CNN model, first, the dataset has been obtained from GitHub [5]. Since the dataset obtained for training the model was very small in size and imbalanced, to solve the problem of having very-limited-sized X-ray image dataset, it has been extended using data augmentation techniques to increase its size and also to make the model training feature rich. Image flipping and rotation at different angles have been used to generate more data. For dataset balancing in terms of proportion of images with different class labels, the dataset has been further extended with some more image instances of the minority class. After data augmentation and dataset balancing, the CNN model has been trained using a total of 800 images (400

COVID-19 and 400 normal) and then the model has been tested by using a test set. The CNN model performance evaluation has then been done using different performance metrics. These metrics include accuracy, precision, sensitivity, specificity, ROC AUC, and F_1 score. Later, the proposed CNN model has also been tested using an independent dataset obtained from the IEEE data port [6] for independent validation of the proposed CNN model. Various machine learning models have also been used for the comparative performance analysis in comparison with the proposed CNN model to show its significance over these models.

The following are some of the key findings of this study:

- (i) CNN with extra convolutional layers (e.g., six layers have been used in the CNN proposed in this study) performs best in COVID-19 diagnosis
- (ii) CNN models require a sufficient amount of images for efficient and more accurate image classification
- (iii) Data augmentation techniques are very effective to improve the CNN model performance remarkably by generating more data from an existing limited-size dataset
- (iv) Data augmentation is also effective in image classification as it gives the ability of invariance to CNNs
- (v) The proposed CNN model performance has been proved statistically significant in the performance of other ML models
- (vi) CNN-based diagnosis using X-ray imaging can be very effective for medical sector to handle the mass testing situations in pandemics like COVID-19

The rest of the paper is divided into various sections. Section 2 constitutes the related work. Section 3 presents the workflow. Section 4 contains the materials and methods used. Section 5 describes the results of the study and discussion, and in the end, Section 6 presents the conclusion.

2. Related Work

Deep learning has shown a dramatic increase in the medical applications in general and specifically in medical image-based diagnosis. Deep learning models performed prominently in computer vision problems related to medical image analysis. The ANNs outperformed other conventional models and methods of image analysis [7, 8]. Due to the very promising results provided by CNNs in medical image analysis and classification, they are considered as de facto standard in this domain [9, 10]. CNN has been used for a variety of classification tasks related to medical diagnosis such as lung disease [10], detection of malarial parasite in images of thin blood smear [11], breast cancer detection [12], wireless endoscopy images [13], interstitial lung disease [14], CAD-based diagnosis in chest radiography [15], diagnosis of skin cancer by classification [16], and automatic diagnosis of various chest diseases using chest X-ray image classification [17]. Since the emergence of COVID-19 in December 2019,

numerous researchers are engaged with the experimentation and research activities related to diagnosis, treatment, and management of COVID-19.

Researchers in [18] have reported the significance of the applicability of AI methods in image analysis for the detection and management of COVID-19 cases. COVID-19 detection can be done accurately using deep learning models' analysis of pulmonary CT [18]. Researchers in [19] have designed an open-source COVID-19 diagnosis system based on a deep CNN. In this study, tailored deep CNN design has been reported for the detection of COVID-19 patients using X-ray images. Another significant study has reported on the X-ray dataset comprising X-ray images belonging to common pneumonia patients, COVID-19 patients, and people with no disease [20]. The study uses the state-of-the-art CNN architectures for the automatic detection of patients with COVID-19. Transfer learning has achieved a promising accuracy of 97.82% in COVID-19 detection in this study. Another recent and relevant study has been conducted on validation and adaptability of Decompose-, Transfer-, and Compose-type deep CNN for COVID-19 detection using chest X-ray image classification [21]. The authors have reported the results of the study with an accuracy of 95.12%, sensitivity of 97.91%, and specificity of 91.87%.

Having reviewed the relevant and recent research work on the design, development, and possible applicability of CNNs in COVID-19 detection using medical images, particularly X-ray images, due to the availability of a very less amount of X-ray images of COVID-19 patients and the poor quality of some images in the dataset, the accuracy of the models was affected. This study is particularly focused on dataset preprocessing to fine-tune it, data augmentation, and design of a CNN with extra layers to increase further the performance of the COVID-19 diagnosis using CNNs as described in subsequent sections.

3. Workflow

As illustrated in Figure 1, the workflow of this study begins with collection of primary dataset containing two image classes: one class belonged to chest X-rays of COVID-19-confirmed cases and the other class of images belonged to the normal people without the disease. In the next phase of the study, the concerned medical professionals analysed the dataset and removed some of the X-ray images which were not clear in terms of quality and diagnostic parameters. Hence, the resulted dataset was very clean, as each X-ray image was of good quality as well as clear in terms of significant diagnostic parameters according to their expertise. In the third phase, the dataset was augmented using standard augmentation techniques to increase its size. The resulted dataset was used to train the model in the next phase. After training, the model was tested for its performance in the disease detection. The testing of proposed CNN model has been done using test dataset held from the primary dataset as well as using the independent validation dataset. Table 1 contains the details of datasets including the total number of

X-ray images in training set, testing set, validation set, and the proportion of X-ray images in the two prediction classes.

4. Materials and Methods

4.1. Dataset. In the experiments of this study, a primary dataset containing 178 X-ray images has been used as a base dataset. Of 178 images, 136 X-ray images belonged to confirmed COVID-19 patients and other 42 images belonged to normal or people with other diseases like pneumonia. The dataset used is available on GitHub [5]. The basic dataset consists of two classes of COVID-19 with 136 samples and others with 42 samples. Thus, the dataset was imbalanced and needed preprocessing to achieve promising results. As a first attempt, CNN was trained on the given original dataset and around 54% accuracy was achieved, which was not worthy of the current application domain. The main dataset sources used in this study are enlisted as follows:

- (i) Primary chest X-ray image dataset of COVID-19 patients collected from GitHub. The dataset has been collected by the University of Montreal's Ethics Committee no. CERSES-20-058-D from different hospitals and clinics [5].
- (ii) For dataset balancing, a collection of chest X-ray images were collected from Kaggle [22].
- (iii) Independent validation dataset containing a collection of 100 COVID-19 X-ray images for the real-world testing of the proposed CNN was collected from IEEE DataPort [6].

The experiments have been conducted using Core i7 7th-generation machine with 8 GB RAM, Microsoft Windows 10 platform using Python language with Anaconda 3 software and Jupyter Notebook.

4.2. Dataset Preprocessing

4.2.1. Balancing Dataset Classes. To balance the given dataset, in order to improve the performance of the proposed CNN models in the detection of COVID-19 cases, 136 normal chest X-ray images have been used. These concatenated extra X-ray images were downloaded from Kaggle [22]. After balancing the dataset when the models have been trained again on the resulted dataset, the accuracy of the given CNN models was improved to 69%. Still, the performance given by the models in terms of accuracy and other measures was not justified as an effective system for COVID-19 detection.

4.2.2. Analysis of X-Ray Images by Medical Experts. A deep analysis was done on the X-ray images by medical specialists. Out of 135 X-ray images of confirmed COVID-19 patients, only a set of 90 X-ray images was selected as a perfect candidate to train the models. The resulted dataset now was reduced to 90 COVID-19-confirmed cases and 90 normal X-ray images. The resulted dataset was again used in training the proposed CNN model; there was again an improvement

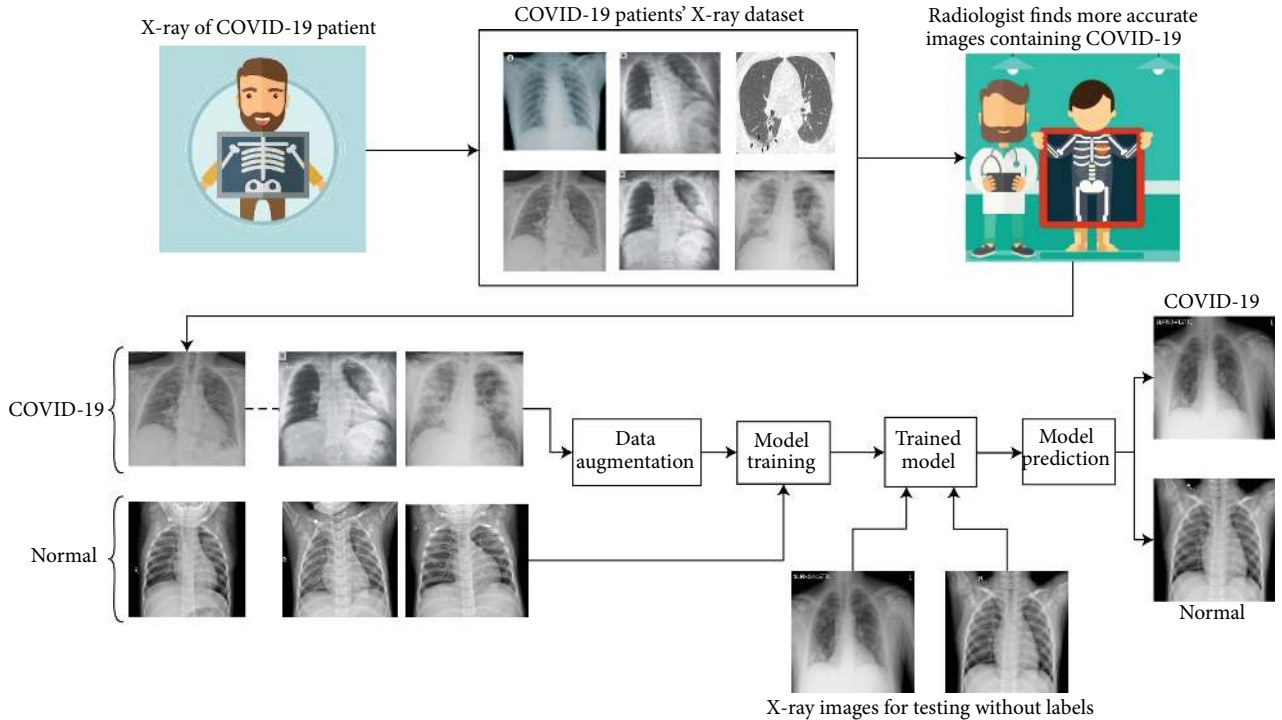


FIGURE 1: System development workflow.

TABLE 1: Dataset image count for training and testing.

Dataset	COVID-19 images	Normal images	Total images
Total data	450	450	900
Training data	400	400	800
Testing data	50	50	100
Independent validation data	100	100	200

in the performance of the model. Specifically, the accuracy was increased to 72% in the given scenario. Still, because the dataset was not containing a sufficient number of images for an effective training, there was not a significant increase in the accuracy and other performance metrics.

4.2.3. Data Augmentation. Data augmentation is a technique that can significantly increase the data instances of a dataset to train a model [23]. In the case of image datasets, the technique uses the basic image processing operations, such as flipping, rotating, cropping, or padding for augmentation. The dataset is then extended by these transformed images resulted from the existing image set, which increases the size of dataset to train the neural networks [24]. To solve the problem of the availability of a small size dataset that was affecting the performance of the proposed CNN, the data augmentation method has been used in this study. This technique increased the size of the dataset; in addition, it provides more learning features to the learning model. Two image processing operations, flipping and rotation, have been used in this study for data augmentation. In the first phase of data augmentation, the 90 X-ray images have been flipped to get extra 90 images. The resulted dataset was increased to contain 180 images after applying this

operation. In the second phase, the original 90 images have further been rotated by 90° angle to get 90 more images and then rotated by 180° angle to get 90 more images, and finally, the original 90 images were further rotated by 270° angle to get more 90 images. These operations resulted in a dataset containing 450 COVID-19 X-ray images. Table 2 shows the image processing operations performed on the image types and corresponding count of images resulted from the operation. Figure 2 shows the effect of augmentation techniques applied to the original sample image of the dataset used in this study.

4.3. Convolutional Neural Networks (CNNs). The CNNs are inspired by visual system of human brain. The idea behind the CNNs thus is to make the computers capable of viewing the world as humans view it. This way CNNs can be used in the fields of image recognition and analysis, image classification, and natural language processing [25]. CNN is a type of deep neural networks which contain the convolutional, max pooling, and nonlinear activation layers. The convolutional layer, considered as a main layer of a CNN, performs the operation called "convolution" that gives CNN its name. Kernels in the convolutional layer are applied to the layer inputs. All the outputs of the convolutional layers

TABLE 2: COVID-19 image count after data augmentation.

Image type	Count
Original	90
Original flipped	90
Original with a 90-degree rotation	90
Original with 180-degree rotation	90
Original with 270-degree rotation	90
Total	450

are convolved as a feature map. In this study, the Rectified Linear Unit (ReLU) has been used in the activation function with a convolutional layer which is helpful to increase the nonlinearity in input image, as the images are fundamentally nonlinear in nature. Thus, CNN with ReLU in the current scenario is easier and faster. Since the ReLU is zero for all negative inputs, it can be defined as

$$z = \max(0, i). \quad (1)$$

Here, the function implies that the output z is zero for all negative value and positive value remains the constant as shown in Figure 3.

The pooling layer or subsampling layer is also an important building block of CNN. On each feature map extracted through the convolution layer, the pooling layer operates independently. To minimize overfitting and the number of extracted features, it decreases the spatial size of the feature map and returns the important features. Pooling can be the max, average, and sum in the CNN model. In this study, max pooling has been used because others may not identify the sharp features easily as compared to max pooling. In addition, the batch normalization layer has been used in this study as it involved the training of a very deep neural network. So the technique adjusts the scaling and activation to normalize the input layer and speed up the learning procedure between hidden units. The dropout layer with a 20% dropout rate has also been used, which drops the neurons during the training chosen at random to reduce the overfitting problem. Towards the last stage of the CNN used in the study, there is a flattening layer to convert the output of convolutional layers into a single-dimensional feature vector. In other words, the flattening layer arranges all the pixel data output produced by convolutional layers in one vector. After flattening, the vector data is given as an input to the next layers of the CNN called fully connected layers or dense layers. In a fully connected layer, each neuron of the previous layer is directly connected to each of the neurons in its next layer. The main functionality of dense layers is to take flattened output results from the convolution and pooling layers and as input and classify the image to a specific class label. Each value of the flattened feature set represents the probability of a feature belonging to a specific class. Thus, on the basis of these probabilities, the fully connected network with dense layers finally drives the classification decision.

4.3.1. The Proposed CNN Architecture. The proposed CNN model consists of 38 layers in which 6 are convolutional (Conv2D), 6 max pooling layers, 6 dropout layers, 8

activation function layers, 8 batch normalization layers, 1 flatten layer, and 3 fully connected layers; CNN model input image shape is (150, 150, 3), i.e., 150-by-150 RGB image. In all Con2D layers, a 3×3 size kernel has been used but the filter size after every two Con2D layers increases. At the 1st and 2nd layers of Con2D, 64 filters have been used to learn from input and the 3rd and 4th layers of Con2D use 128 filters, and at the 5th and 6th layers, 256 filters have been used. After each Con2D layer, the max pooling layer with 2×2 pooling size has been used, the batch normalization layer has been used with the axis = -1 argument, the activation layer has been used with the ReLU function, and the dropout layer has been used with 20% dropout rate. The output of 256 output neurons of the final Con2D layer is followed by max pooling, batch normalization, activation, and dropout layer. Since the final pooling and convolutional layer gives a three-dimensional matrix as output, to flatten the matrix, a flattening layer has been used which converts them into a vector that will be input for 3 dense layers.

This study uses CNN for binary classification; that is the reason for using the binary crossentropy (BCE) loss function. In binary classification since only one output node is needed to classify the data to one of the two given classes, so in the case of BCE loss function, the output value is being given to a sigmoid activation function. The output given by the sigmoid activation function lies between 0 and 1. It finds the error between the predicted class and the actual class. The ‘‘Adam’’ optimizer has been used which changes the attribute weight and learning rate to reduce the loss of the learning model. The model parameter values are given in Table 3, and the model architecture is given in Figure 4. During the initial experiments, the CNN has been used with different configurations in terms of the usage of number of convolution layers in the model. The decision of how many convolution layers used in the model was made by using an incremental approach. First, the CNN was tested using only one convolutional layer and the results were analysed. Then, the CNN was built with two layers and results were analysed and so on. The approach had been continued till the results provided by the model were accurate and effective. The final model which was very feasible according to its results consisted of six convolution layers. The results of each increment of the model have been reported in the Results section.

5. Results and Discussion

After preprocessing of the dataset, the final dataset consisted of a total of 900 X-ray images. For training and testing the proposed CNN, the dataset was partitioned into two subsets. The training dataset contained 400 COVID-19 X-ray images and 400 normal X-ray images, making a total of 800 X-ray images. The testing dataset similarly contained 100 X-ray images, in which 50 X-ray images were from each class COVID-19 positive and normal. Then, the training subset containing 800 X-ray images has been passed to the model with 25% validation size. So, out of 800 X-ray images, with each epoch, 600 X-ray images train the model, and 200 X-ray Images validate the model. As mentioned in the proposed

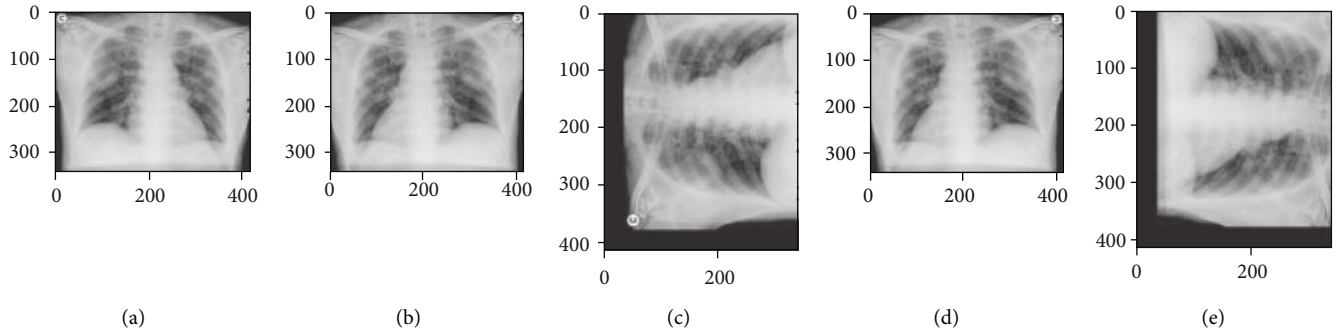


FIGURE 2: Effect of augmentation techniques on an X-ray image. (a) Original. (b) Flipped. (c) Rotated 90°. (d) Rotated 180°. (e) Rotated 270°.

architecture of the CNN model, it consisted of 38 layers in which 6 are convolutional, 6 max pooling layers, 6 dropout layers, 8 activation function layers, 8 batch normalization layers, 1 flattening layer, and 3 fully connected layers. The CNN model thus achieved an extraordinary performance with an accuracy of 100% with the test data subset used from the processed dataset of this study with a precision of 1.0, with the model parameter values given in Table 3. To evaluate the overall performance, in addition to accuracy, other important metrics have been adopted in this study including F_1 score, precision, sensitivity, specificity, and ROC AUC. The scores of these parameters are reported in Table 4.

The confusion matrix of the model is shown in Figure 5. Figures 6 and 7 show the curve of accuracy and loss between training and testing, respectively. According to the confusion matrix, the CNN model test uses the 100 X-ray images from the GitHub dataset, where 50 images belong to the COVID-19 class and 50 to the normal images. The CNN model shows significant performance on testing and predicts all 100 images correctly with 0% error rate as reported in the confusion matrix of Figure 5. Figure 6 shows the model accuracy during the training and validation as a graph where the curve drawn in blue color shows the training accuracy of CNN, while the curve with orange color shows the validation accuracy. Training accuracy of the CNN according to Figure 6 remains consistent after the 5 epochs and the CNN also shows a consistent validation accuracy after the 25 epochs. The plot in Figure 7 shows the loss during the training and validation of CNN. The training loss of CNN is minimum and consistent from the 1st epoch while validation loss becomes minimum after 5 epochs and remains consistent till the last epoch. The above results show the efficiency of the CNN model proposed in this study.

Figure 6 shows the plots drawn from the training and testing accuracy achieved by the proposed CNN model. Figure 7 shows the training and testing loss for the proposed CNN model. As can be observed in Figure 7, the proposed CNN is not taking a lot of time to converge, as in the first epoch the training loss is 31 and right after 5 epochs, it drops to 0.9; then after 23 epochs, it drops again to 0.0011, and at the last epoch, the total loss is 0.000058.

In addition to the above performance measurements, K fold cross-validation has been applied to further test the

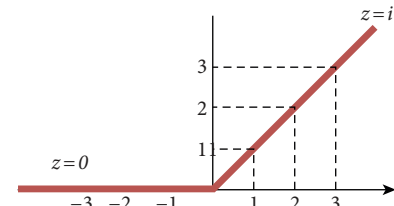


FIGURE 3: ReLU function demonstration.

TABLE 3: Model parameter values.

Parameter	Value
Input dimension	(150, 150, 3)
Filter to learn	64, 128, 256
Max pooling	2×2
Batch normalization	Axis = -1
Activation functions	ReLU, sigmoid
Dropout rate	20%
Kernel size	3×3
Epochs	50
Optimizer	Adam
Loss function	binary_crossentropy

proposed model for its skill. In this study, a 10-fold cross-validation has been used. The results provided are very effective as the average score of 10 iterations is 99.67% ($\pm 0.15\%$).

5.1. Testing of the CNN Using COVID-19 Independent Validation Data. As proof of the significance of the proposed CNN model in the classification for detection of COVID-19 from X-ray images, the trained model has been tested using an independent dataset obtained from the IEEE DataPort [6]. The independent test dataset contained 100 COVID-19 X-rays. This dataset was then extended by adding 100 normal images for testing. The model using the same settings performs very well with an accuracy of 0.995 and precision 1.000 along with other performance parameters reported in Table 5.

Figure 8 shows the confusion matrix of CNN model when tested on the independent test dataset which have been obtained from IEEE DataPort. The CNN model also performed very efficiently on independent test data giving 198

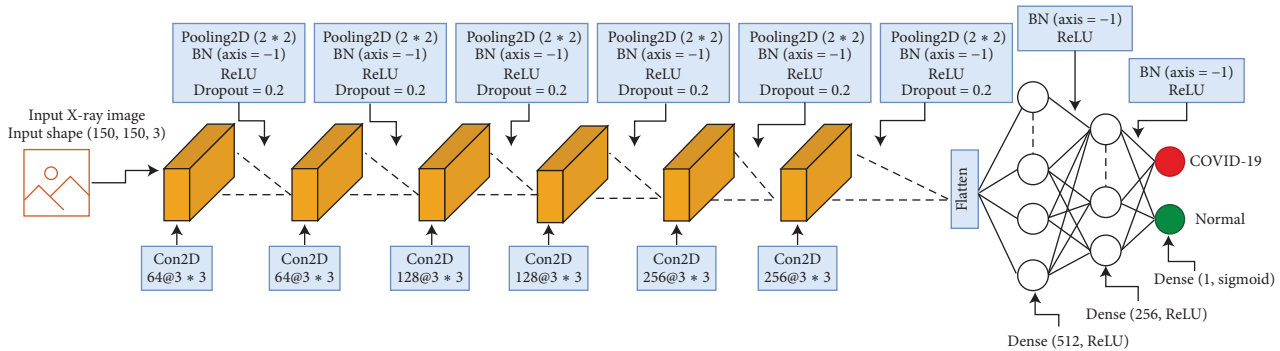


FIGURE 4: CNN model architecture.

correct classification results out of 200 input X-ray images. As can be seen in the results given by the confusion matrix, an equal number of images is obtained from both of the target classes (100 COVID-19 and 100 normal). The model classifies only one image falsely as normal from COVID-19 class.

Falsely classified image was examined by the experts. It was noted that the X-ray image belonged to a person on an early stage of COVID-19. As a result, the image does not contain the prominent patterns with which the image could have been differentiated from normal X-ray image class. Figure 9 shows the actual X-ray image of a COVID-19 positive case which was falsely classified by the model. The comparison of CNN model on test data and independent validation data in terms of different performance metrics is shown in Figure 10.

As mentioned in the proposed CNN model architecture section, the proposed model was constructed in an incremental approach. Starting with single convolutional layer model, in each following increment, a convolutional layer had been added and results were analysed. Table 6 illustrates the results of the model in terms of accuracy after every increment from single convolutional layer to a stable model consisting of six convolutional layers. The incremental approach comparison is shown in Figure 11.

5.2. Performance Comparison of Machine Learning Models with the CNN Model. In this study, experiments have also been conducted on some of the relevant machine learning models such as Random Forest (RF) [26, 27], gradient boosting machine (GBM) [28, 29], support vector classifier (SVC) [30], logistic regression (LR) [31], and k -nearest neighbor (KNN) [32] for comparative analysis of CNN with these models. These models have been used with their best parameter settings as shown in Table 7.

RF has been used with two hyperparameters as shown in the table. The $n_estimators$ define how many decision trees are generated under RF to make a prediction. The max_depth defines what should be the maximum depth of each decision tree in RF, so in this setting, the max_depth parameter restricts the decision tree to a maximum 300 level depth.

GBM has been used with three parameters, two are the same as in RF and one is $learning_rate$ which is a tuning

TABLE 4: Model performance on test data.

Evaluation parameters	Score
Accuracy	1.000
Precision	1.000
Sensitivity	1.000
Specificity	1.000
F_1 score	1.000
ROC AUC	1.000

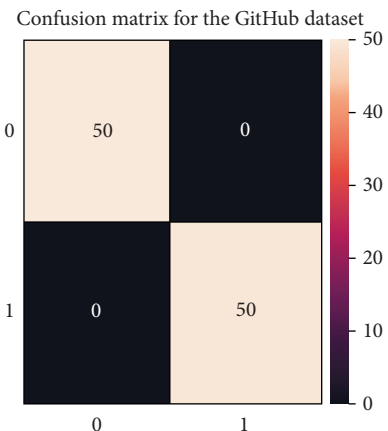


FIGURE 5: Confusion matrix of CNN model for primary/GitHub dataset. Here 0 represents the normal class and 1 represents the COVID-19.

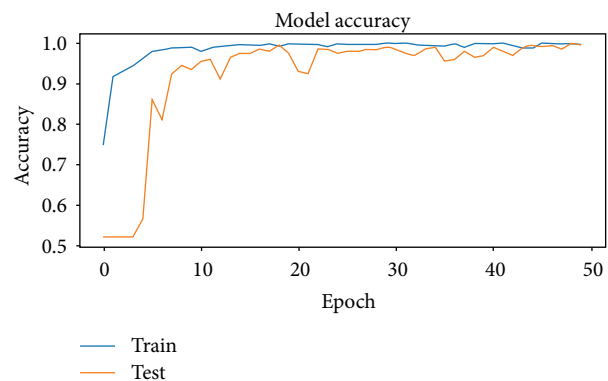


FIGURE 6: Training and testing accuracy plot achieved by the CNN model.

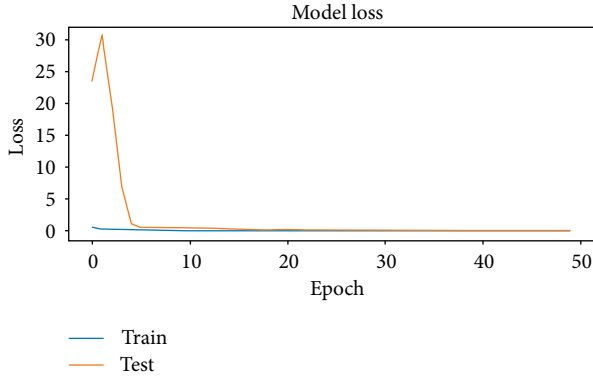


FIGURE 7: Training and testing loss plot by the CNN model.

TABLE 5: Model performance on independent validation data.

Evaluation parameters	Score
Accuracy	0.995
Precision	1.000
Sensitivity	0.990
Specificity	1.000
F_1 score	0.994
ROC AUC	0.990

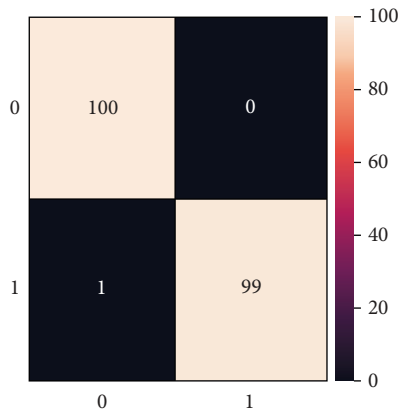


FIGURE 8: Confusion matrix of CNN model for independent validation data (0: normal class, 1: COVID-19).

parameter for optimization of algorithms to reduce the error to a minimum level [28]. SVC has been used with the linear kernel which is best for binary class classification and the second parameter C conveys SVC optimization the limit to prevent misclassification of the individual training examples. In contrary, even if the hyperplane misclassifies certain points, a very small value of C would cause the optimizer to search for a wider-margin separating hyperplane. LR has been used with the “liblinear” solver because it is preferred when there is a small dataset, and the second parameter is C as used in SVC. KNN has been used with all the default parameters setting.

Thus, the CNN and machine learning models have been trained using the original dataset of this study. Then, both CNN and each of the machine learning models have been tested using the training subset of the original dataset. In this

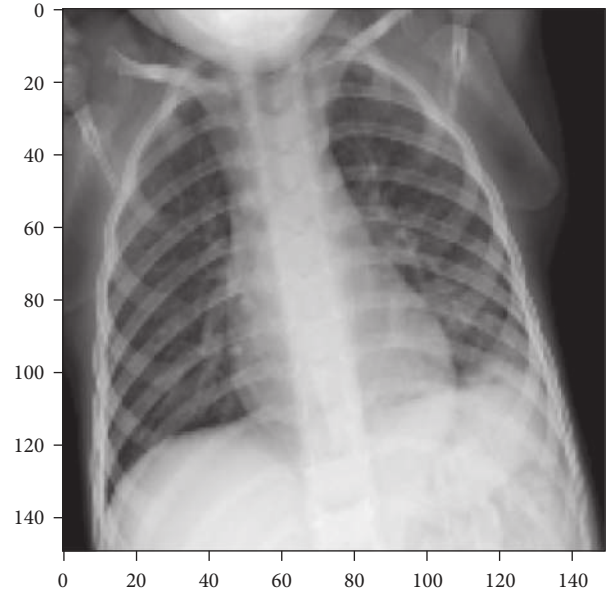


FIGURE 9: The falsely classified X-ray image.

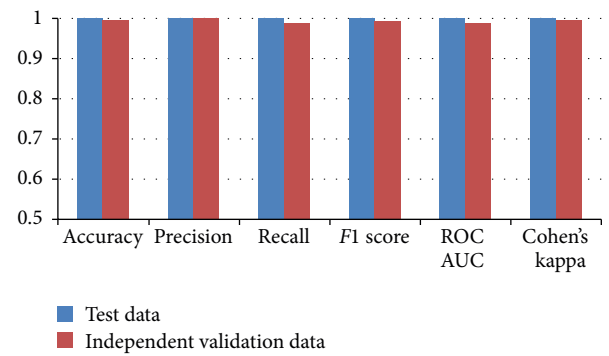


FIGURE 10: CNN model performance comparison on both test data and independent validation data.

TABLE 6: Accuracy score with different number of CNN layers.

Convolutional layer	Test data	Independent validation data
One Conv2D	0.715	0.455
Two Conv2D	0.940	0.895
Three Conv2D	0.995	0.957
Four Conv2D	0.995	0.980
Five Conv2D	0.995	0.995
Six Conv2D	1.000	0.995

scenario, three machine learning models SVC, LR, and KNN performed well almost the same as CNN as reported in Table 8. On the other hand, when the same machine learning models with the same settings were tested on the independent test set as mentioned earlier, their performance was degraded while the proposed CNN maintains the performance metrics in this scenario also. The results of the second scenario in which models were tested using the independent test set are reported in Table 9.

This comparative analysis and also comparing the overall performance results achieved by the proposed CNN model

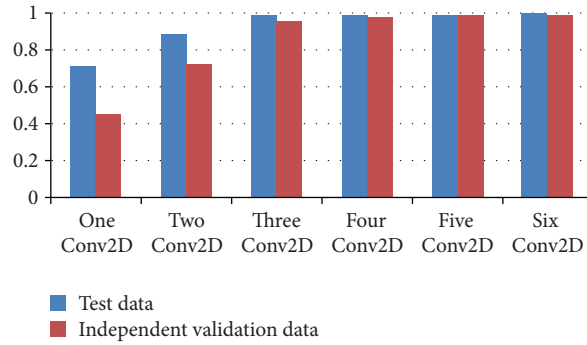


FIGURE 11: Accuracy score comparison with different number of CNN layers.

TABLE 7: Machine learning models hyperparameter setting.

Model	Hyperparameter settings
RF	$N_estimators = 300$, $max_depth = 300$
GBM	$N_estimators = 300$, $max_depth = 300$, $learning_rate = 0.2$
SVC	Kernal = linear, $C = 3.0$
LR	Solver = liblinear, $C = 3.0$
KNN	$n_neighbors = 3$

TABLE 8: Models' performance on test data.

Model	Accuracy	Precision	Sensitivity	Specificity	F_1 score	ROC AUC
RF	0.980	0.961	1.000	0.961	0.980	1.000
GBM	0.940	0.960	0.923	0.958	0.941	0.940
SVC	1.000	1.000	1.000	1.000	1.000	1.000
LR	1.000	1.000	1.000	1.000	1.000	1.000
KNN	1.000	1.000	1.000	1.000	1.000	1.000
CNN	1.000	1.000	1.000	1.000	1.000	1.000

TABLE 9: Models' performance on independent validation data.

Model	Accuracy	Precision	Sensitivity	Specificity	F_1 score	ROC AUC
RF	0.950	0.960	0.941	0.960	0.950	0.950
GBM	0.920	0.930	0.911	0.928	0.920	0.922
SVC	0.940	0.950	0.931	0.948	0.940	0.943
LR	0.940	0.960	0.923	0.958	0.941	0.944
KNN	0.930	0.990	0.880	0.990	0.930	0.930
CNN	0.995	1.000	0.990	1.000	0.995	0.995

provide strong evidence of possible applicability of the model in the COVID-19 diagnosis using X-ray image classification.

The two bar graphs in Figure 12 shows the comparisons of the performances given by different machine learning models and CNN. The bar graph labeled as test data is the result of the first test scenario, where each of the models was tested on the test subset extracted from the original dataset. The second bar graph labeled as independent validation data is the result of the second test scenario, where each of the models was tested on the independent test set.

To compare the performance results of the proposed CNN-based methodology for the current application domain, the results of other recent studies done by different

researchers have been collected and compared. The comparison of these study results along with the mythology used has been shown in Table 10.

5.3. Statistical Significance of the Proposed CNN Model. In order to test whether the proposed CNN model has a statistical significance over the other models, t -test [36] has been performed. To determine the significance, the alternate hypothesis H_a and null hypothesis H_o have been established as follows:

- (i) *Alternate Hypothesis (H_a).* There is a statistical significance in the performance given by the

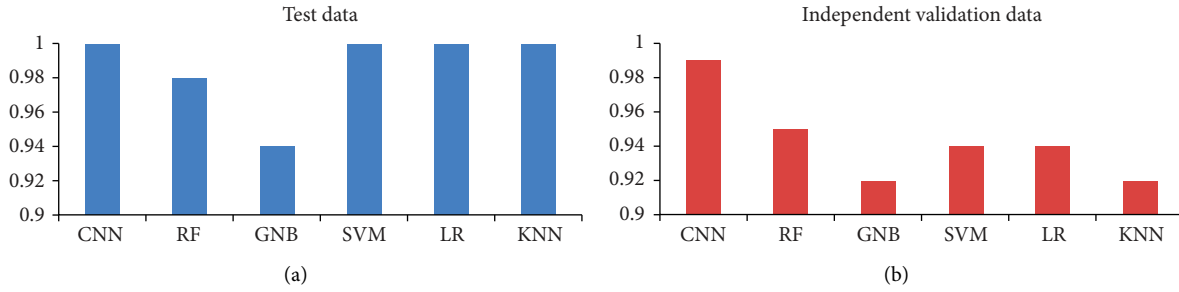


FIGURE 12: Accuracy of different machine learning models along with proposed CNN model. (a) Test data. (b) Independent validation data.

TABLE 10: Comparison with other studies performed on the same dataset.

Study	Model	Accuracy (%)
Sethy et al. [33]	SVM	98.66
Minaee et al. [34]	SqueezeNet	92.2
Das et al. [35]	CNN	97.4
Our study (test data)	CNN	100
Our study (validation data)	CNN	99.2

proposed CNN model over the performance of other models

- (ii) *Null Hypothesis (H_0)*. There is no statistical significance in the performance given by the proposed CNN model over the performance of other models

First, the significance of the proposed CNN model has been shown on two models, RF and GBM, when these models have been tested on the test data set derived from the primary dataset. The H_0 has been rejected in the case of RF and GBM. That means the proposed CNN has a statistical significance over RF and GBM only in this testing scenario. While H_0 was accepted in the case of SVC, LR, and KNN, that means the proposed CNN has no statistical significance over these models.

Second, the t -test was performed on the performance results of all the models when these models were tested on the independent validation data set. In this scenario, H_0 was rejected in comparison to all the other models. This implies that the CNN is statistically significant on all the given models in this model testing scenario.

6. Conclusion

This study has been conducted to demonstrate the effective and accurate diagnosis of COVID-19 using CNN which was trained on chest X-ray image datasets. The model training was performed incrementally with different datasets to attain the maximum accuracy and performance. The primary dataset was very limited in size and also imbalanced in terms of class distribution. These two issues with the primary dataset affected the performance of the models very badly. To overcome these issues, the dataset was preprocessed using different techniques, including dataset balancing technique, manual analysis of X-ray images by concerned medical

experts, and data augmentation techniques. To balance the dataset for model training and also to test its performance parameters, an ample number of chest X-rays were collected from different available sources. After training and testing the CNN model on the fully processed dataset, the performance results have been reported. In addition, to test further the model performance, particularly the accuracy, the proposed CNN model has been tested using an independent dataset as an independent validation and real-world test obtained from IEEE DataPort [6]. As reported in the results in both the testing scenarios, the proposed CNN model has shown highly promising results. Since this study uses an incremental approach in training the model using different sizes and types of datasets, the approach confirmed the fact that CNN models require an ample amount of image data for the efficient and more-accurate classification. The data augmentation techniques are very effective to significantly improve the CNN model performance by generating more data from an existing limited-size dataset and also by giving the ability of invariance to the CNN. The proposed CNN model's number of convolutional layers was also decided in an incremental approach; that is, in the first increment, only one convolutional layer was used and, then, on the basis of model performance metrics, one layer in each increment was increased till it reaches a stable and efficient stage in terms of its performance. The final version of the CNN consisted of six convolutional layers. A comparative analysis has also been done to further test the scope of the proposed CNN model by performance comparisons with some of the prominent machine learning models such as RF, GBM, SVC, LR, and KNN. The results prove that the proposed CNN has outperformed all the models particularly when each model was tested on the independent validation dataset. Considering the significant effect of data augmentation techniques on model performances, the authors are currently working on the application of other state-of-the-art data augmentation algorithms and techniques. In the future, the results obtained from the study concerned with the applicability of these modern data augmentation techniques in different application domains will be published.

Data Availability

The data supporting this study are from previously reported studies and datasets, which have been cited.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication of this paper.

Authors' Contributions

Aijaz Ahmad Reshi and Furqan Rustam contributed equally to this work.

Acknowledgments

The authors extend their appreciation to the Deanship of Scientific Research at King Saud University for supporting this research work through research group no. RG-1441-455.

References

- [1] D. Cucinotta and M. Vanelli, "WHO declares COVID-19 a pandemic," *Acta Biomedica: Atenei Parmensis*, vol. 91, pp. 157–160, 2020.
- [2] F. Rustam, A. A. Reshi, A. Mehmood et al., "COVID-19 future forecasting using supervised machine learning models," *IEEE Access*, 2020.
- [3] D. J. Cennimo, "Coronavirus disease 2019 (COVID-19) clinical presentation," vol. 8, pp. 101489–101499, 2020, <https://emedicine.medscape.com/article/2500114-clinical#b2>, 2020. Online.
- [4] X-ray (Radiography)-Chest, 2020, <https://www.radiologyinfo.org/en/info.cfm?pg=chestradi#overview>.
- [5] J. P. Cohen, "Github Covid19 X-ray dataset," 2020, <https://github.com/ieee8023/covid-chestxray-dataset>, 2020. Online.
- [6] Z. H. Chen, "Mask-RCNN detection of COVID-19 pneumonia symptoms by employing stacked autoencoders in deep unsupervised learning on low-dose high resolution CT," *IEEE Dataport*, 2020.
- [7] A. S. Lundervold and A. Lundervold, "An overview of deep learning in medical imaging focusing on MRI," *Zeitschrift für Medizinische Physik*, vol. 29, no. 2, pp. 102–127, 2019.
- [8] M. Ahmad, "Ground truth labeling and samples selection for hyperspectral image classification," *Optik*, vol. 230, Article ID 166267, 2021.
- [9] B. Kayalibay, G. Jensen, and P. van der Smagt, "CNN-based segmentation of medical imaging data," 2017, <http://arxiv.org/abs/1701.03056>.
- [10] Q. Li, W. Cai, X. Wang, Y. Zhou, D. D. Feng, and M. Chen, "Medical image classification with convolutional neural network," in *Proceedings of the 2014 13th International Conference on Control Automation Robotics & Vision (ICARCV)*, pp. 844–848, Singapore, December 2014.
- [11] M. Umer, S. Sadiq, M. Ahmad, S. Ullah, G. S. Choi, and A. Mehmood, "A novel stacked CNN for malarial parasite detection in Thin blood smear images," *IEEE Access*, vol. 8, pp. 93782–93792, 2020.
- [12] R. Rouhi, M. Jafari, S. Kasaei, and P. Keshavarzian, "Benign and malignant breast tumors classification based on region growing and CNN segmentation," *Expert Systems with Applications*, vol. 42, no. 3, pp. 990–1002, 2015.
- [13] M. Sharif, M. Attique Khan, M. Rashid, M. Yasmin, F. Afza, and U. J. Tanik, "Deep CNN and geometric features-based gastrointestinal tract diseases detection and classification from wireless capsule endoscopy images," *Journal of Experimental & Theoretical Artificial Intelligence*, pp. 1–23, 2019.
- [14] N. Asada, K. Doi, H. MacMahon et al., "Potential usefulness of an artificial neural network for differential diagnosis of interstitial lung diseases: pilot study," *Radiology*, vol. 177, no. 3, pp. 857–860, 1990.
- [15] S. Katsuragawa and K. Doi, "Computer-aided diagnosis in chest radiography," *Computerized Medical Imaging and Graphics*, vol. 31, no. 4-5, pp. 212–223, 2007.
- [16] A. Esteva, B. Kuprel, R. A. Novoa et al., "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.
- [17] Y. Dong, Y. Pan, J. Zhang, and W. Xu, "Learning to read chest X-ray images from 16000+ examples using CNN," in *Proceedings of the 2017 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*, pp. 51–57, Philadelphia, PA, USA, July 2017.
- [18] D. Dong, Z. Tang, S. Wang et al., "The role of imaging in the detection and management of COVID-19: a review," *IEEE Reviews in Biomedical Engineering*, vol. 14, pp. 16–19, 2020.
- [19] L. Wang and A. Wong, "COVID-Net: A tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images," 2020, <http://arxiv.org/abs/2003.09871>.
- [20] I. D. Apostolopoulos and T. A. Mpesiana, "Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks," *Physical and Engineering Sciences in Medicine*, vol. 43, pp. 635–640, 2020.
- [21] A. Abbas, M. M. Abdelsamea, and M. M. Gaber, "Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network," 2020, <http://arxiv.org/abs/2003.13815>.
- [22] P. Mooney, "Kaggle X rays dataset," 2020, <https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia> Online.
- [23] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, vol. 6, p. 60, 2019.
- [24] D. Ho, E. Liang, and R. Liaw, "1000x Faster Data Augmentation, Berkeley Artificial Intelligence Research," Berkeley, CA, USA, 2019, https://bair.berkeley.edu/blog/2019/06/07/data_aug/.
- [25] D. T. Mane and U. V. Kulkarni, "A survey on supervised convolutional neural network and its major applications," *International Journal of Rough Sets and Data Analysis*, vol. 4, no. 3, pp. 71–82, 2017.
- [26] A. Liaw and M. Wiener, "Classification and regression by random forest," *R News*, vol. 2, pp. 18–22, 2002.
- [27] F. Rustam, A. A. Reshi, I. Ashraf et al., "Sensor-based human activity recognition using deep stacked multilayered perceptron model," *IEEE Access*, vol. 8, pp. 218898–218910, 2020.
- [28] V. K. Ayyadevara, "Gradient boosting machine," in *Pro Machine Learning Algorithms* Springer, Berlin, Germany, 2018.
- [29] F. Rustam, A. Mehmood, S. Ullah et al., "Predicting pulsar stars using a random tree boosting voting classifier (RTB-VC)," *Astronomy and Computing*, vol. 32, Article ID 100404, 2020.
- [30] K. W. Lau and Q. H. Wu, "Online training of support vector classifier," *Pattern Recognition*, vol. 36, no. 8, pp. 1913–1920, 2003.
- [31] S. Dreiseitl and L. Ohno-Machado, "Logistic regression and artificial neural network classification models: a methodology review," *Journal of Biomedical Informatics*, vol. 35, no. 5-6, pp. 352–359, 2002.

- [32] R. Ramteke and K. Y. Monali, "Automatic medical image classification and abnormality detection using k-nearest neighbour," *International Journal of Advanced Computer Research*, vol. 2, p. 190, 2012.
- [33] P. K. Sethy, S. K. Behera, P. K. Ratha, and P. Biswas, "Detection of coronavirus disease (COVID-19) based on deep features and support vector machine," 2020.
- [34] S. Minaee, R. Kafieh, M. Sonka, S. Yazdani, and G. Jamalipour Soufi, "Deep-covid: predicting covid-19 from chest x-ray images using deep transfer learning," *Medical Image Analysis*, vol. 65, Article ID 101794, 2020.
- [35] N. N. Das, N. Kumar, M. Kaur, V. Kumar, and D. Singh, "Automated deep transfer learning-based approach for detection of COVID-19 infection in chest X-rays," *IRBM*, 2020.
- [36] E. B. Fatima, B. Omar, E. M. Abdelmajid, F. Rustam, A. Mehmood, and G. S. Choi, "Minimizing the overlapping degree to improve class-imbalanced learning under sparse feature selection: application to fraud detection," *IEEE Access*, vol. 9, pp. 28101–28110, 2021.