A Deep Learning Approach for Automated COVID-19 Detection

Original Scientific Paper

Amanpreet Singh
Department of Electronics and Communication Engineering
Punjabi University, Patiala, Punjab, India.
asodhi.pta@gmail.com

Charanjit Singh
Department of Electronics and Communication Engineering
Punjabi University, Patiala, Punjab, India.
channisingh@yahoo.com

Abstract *– Nowadays, COVID-19 is a life-threatening virus for human beings, and the reason behind it is its attack on the respiratory system. A large number of cases of infection were reported with minor to no symptoms. So, detection of the disease at an earlier stage can decrease the death rate in the patients. Chest X-Rays scans can be used primarily for analyzing the infection. X-ray technology is chosen over CT scans because its equipment is readily available, results can be obtained quickly, and the process is quite affordable in terms of cost. This paper proposed a solution using a deep learning approach to detect COVID-19 infection in human lungs using Chest X-Ray scans. Here, we have used CLAHE (Contrast Limited Adaptive Histogram Equalization) to enhance the contrast of X-ray images and then Convolutional Neural Network on CLAHE processed images to improve the accuracy of the overall model.*
Further, these scans are classified using machine learning classifiers among COVID-19 infected and normal. The proposed model is trained and validated on a publicly available COVID-19 X-ray dataset containing 15917 X-ray Images. Confusion matrices and ROC curves have been generated to analyze the model's efficiency. Training and validation graphs are developed to calculate the other parameters like validation accuracy and training Accuracy. The model's accuracy is 99.8%, which is better than its existing state-of-the-art approaches. These results show that this model is promising for physicians to classify the chest X-Rays scans of infected patients with COVID-19.

Keywords*: Deep Learning, Medical Images, COVID-19 Detection, X-Ray Images*

1. INTRODUCTION

COVID-19 is a lethal disease caused by a newly confirmed Coronavirus infection. In December 2019, it was first transmitted to humans, and it would be transmitted from person to person through droplets that form when a person speaks, coughs, or sneezes [1-6]. This virus primarily attacks the lungs and can harm the muscles of an infected human being.

Because of global climate change, there are so many diseases from which people are already suffering, and at the same time, the impact of coronavirus is immense. Healthcare professionals and researchers in different regions across the globe are working to find a stable solution and improve testing capacity by employing multifunctional tests to control the spread of contamination and protect themselves from the same.

Recently, RT-PCR (reverse transcriptase-polymerase chain reaction) diagnostics have proved effective in detecting infection. Though, this method has the disadvantages of a longer detection time and a slower virus detection speed. [7, 8]. Many scholars are working globally to overcome the limits of RT-PCR tests and expand the diagnosis process of COVID-19. Deep learning algorithms with CNN are also used to diagnose viruses through image classification techniques. According to the WHO recommendations, chest X-rays effectively diagnose the clinical symptoms of infected people who have been improved [9].

Recent studies show that CNNs are very useful and have a perfect effect in identifying COVID-19 through image processing. CNN can be a multi-layer neural network that recognizes image patterns with the help of different image pre-processing tools. Some CNN Models like Resnet50 [10], AlexNet [11], VGG16 [12], and VGG19 are also available and perform well in classifying COVID-19 chest X-Rays scans.

Here we have proposed a fusion of CLAHE (Contrast Limited Adaptive Histogram Equalization) along with the CNN (Convolutional Neural Network) to increase the model's accuracy. CLAHE is the developed version of the adaptive histogram equation used to improve the contrast of the images by performing a stretching-out mechanism on frequent intensity values of the image. In the proposed method, we have used CLAHE to pre-process the medical images, and then processed images are provided to the CNN network for classification. A detailed explanation of CLAHE and the CNN network is provided in the upcoming sections. The main contributions of the work are:

- The work presented here provides an improved Deep learning model trained to spot COVID-19 contamination using chest X-Ray images and classify them into infected and normal subjects. This method proposes a new fusion of the CLAHE (Contrast Limited Adaptive Histogram Equalization) CNN classification. In addition, the coloring of indexed images is used in pre-processing to enhance the accuracy of this model.
- We have used data augmentation strategies for COVID-19 discovery to avoid overfitting issues.
- A publicly available large X-ray image dataset is used in this work, showing better accuracy and other metrics than existing methods.

2. RELATED LITERATURE

Over the past few months, many researchers have examined and analyzed chest X-rays using machine learning algorithms to detect the infection. Several AI learning-based approaches are available for COVID-19 detection using X-ray scans. In the healthcare area, deep learning [DL], a sub-branch of artificial intelligence, is a current and increasingly evolving CAD (Computer-Aided Design) tool to help clinicians/radiologists better predict disease. DL methods can guide practitioners in advancing the quality of COVID-19 detection.

Chowdhury et al. [13] worked on breast X-rays and created a framework, PDCOVIDNet, based on dilated parallel conventional neural network. The proposed method used small convolution in a similar stack to capture and stretch the required properties to obtain an accuracy of 96.58%.

Khan et al. [14] presented a new X-ray analytical architecture such as COVID-19 with pre-charged machine learning models such as VGG16, ResNet50, DensNet121, and VGG19, in which VGG16 and 19 have shown the best accuracy. This proposed model consists of 2 phases, such as pre-processing and data dissemination and learning transfer, and lastly indicates an accuracy of around 99.3%.

Minae et al. [15] reported wide-ranging research showing COVID-19 infection in chest X-ray imaging using four integrated models: SqueezeNet, ResNet18, ResNet50, and DensNet-121. This plan used the data expansion to create an altered variety of the COVID-19 image to raise the number of testers and ultimately achieve 90% specificity and 98% sensitivity.

Sekeroglu et al. [16] designed a prototype using different machine-learning techniques that performed 38 experiments to identify infection using high-precision X-ray imaging. Of these, he served ten experiments, five various deep learning algorithms, and 14 trials with hi-tech pre-trained systems for educational exchange. These procedures found an accuracy of 98.50%, an accuracy of 99.18%, and a sensitivity of 93.84%. They conclude that the process developed by CNN can complement the detection of COVID-19 in low-resolution images with minimal processing and no pre-processing.

Khalifa et al. [17] introduced a method of classifying coronavirus cure goals in the human brain grounded on handling type and therapeutic level by using deep learning (DL) and machine learning (ML). The processing distribution accuracy obtained by the model reaches 98.05% in comparison with other ML models, such as SVM and DT. The DCNN model lacked an accuracy rate (98.2%) compared with the DT (98.5%) for estimating the clinical trials. Broadcast models (e.g., Alexnet) were used in the study.

School et al. [18] reported a possible development of hybrid delivery methods using CNN and marine hunters for COVID-19 imaging obtained by international chest radiologists. They used the CNN design model to extract features and the competitive marine predator algorithm to select the most important images. However, scientific research has yet to determine the fusion pathway to improve the distribution and presentation of the COVID-19 image.

Mohammad Marufur Rahman et al. [19] proposed a HOG (histogram of oriented gradient) and CNN-based model for the classification of COVID-19 and Pneumonia from X-ray images and achieved an accuracy of 96.74% in image classification.

Most reports have used X-ray scanning to spot the contamination, highlighting the significance of chest X-ray scans as an accurate means for physicians and electrotherapists. Though, in some cases, the distributions do not provide the desired results due to inconsistencies in the control data and the inability to qualify from the image. To eliminate these limitations, in this research, we have proposed a combination of CLAHE and CNN to improve the system's overall accuracy.

3. PROPOSED METHODOLOGY

Over the past few years, multiple classifier systems have gained everybody's consideration in the domain of artificial intelligence. These systems are proved very effective in resolving many existing complications, like health care and computer vision difficulties. These systems combine different features from different models to boost the system's overall efficiency. In this proposed system, we combine CLAHE with CNN to increase the accuracy of the entire system.

3.1. SYSTEM ARCHITECTURE

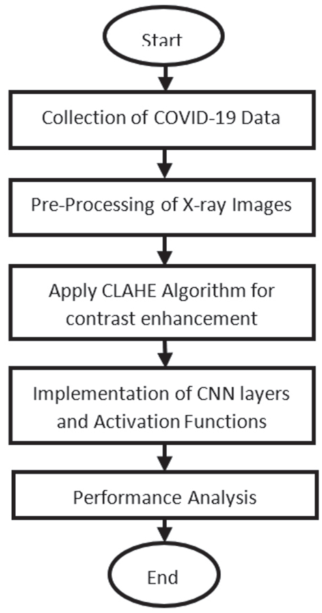

Fig.1. Flow chart of the Proposed Model

Fig. 1 above shows the flow chart of the proposed model. This proposed system takes X-ray scans as input. The first step is to resize the images and convert indexed images to RGB, as CLAHE works much better on Colored images than any other format. Then this colored converted image is sent to the CLAHE for contrast enhancement; then, we applied the CNN model to the same image. These two features were merged and used as strategies to form a distribution model. A detailed explanation of the whole procedure is given in the following sections. Fig 2. below shows the architecture of the proposed system.

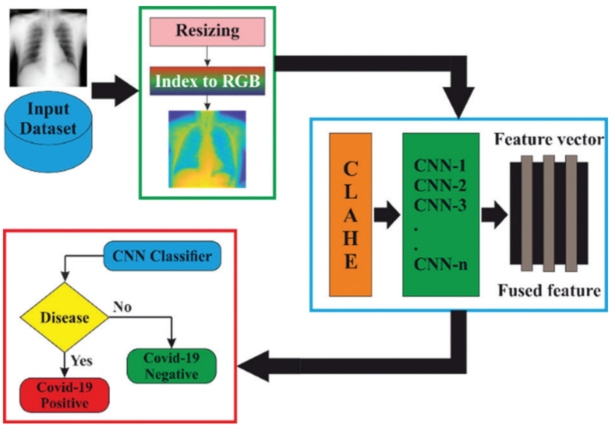

Fig. 2. Proposed System Architecture

As we can see in Fig. 2 above, firstly input image goes to the resizing block and then to the RGB block for conversion; after conversion, the same image goes to the CLAHE block for contrast enhancement so the powerful CNN layers can perform their necessary actions on it. Then data is fed to the CNN network, where images are classified as positive or negative for COVID-19.

3.2. DATASET USED

The dataset used in the process is publicly available on Kaggle, named COVID-19 X-Ray dataset. This dataset contains 15917 X-Ray scans of infected and non-infected patients, 2186 and 13731, respectively. The files contain images of various sizes that range from 512×512 to 657×657 pixels.

3.3. DATA PRE-PROCESSING

Firstly the dataset is normalized, resizing the images to 70 × 70 is done, and then the images are mixed and divided into training and test data. The training data contains 11142 images, which are divided into two folders named COVID and NONCOVID. The COVID folder under training data contains 1530 images, and the NONCOVID folder contains 9612 images. Similarly, test data contains 4775 images and two subfolders, COVID and NONCOVID, which include 656 and 4119 images, respectively. It can be precisely seen in Table 1 below.

Table 1. Number of images in different categories used in the training process

Dataset	COVID	NONCOVID	Total
Train	1530	9612	11142
Test	656	4119	1800

Training and test dataset images were selected by excruciating the complete dataset of images (i.e., training and test dataset combined). If multiple images are present for the same patient, we have ensured that images are marked as either training or test data so results can be manageable because of patient overlap. Then we converted the indexed images into color images using the MATLAB function ind2rgb(X, map). After all, this dataset is ready for the training of the model.

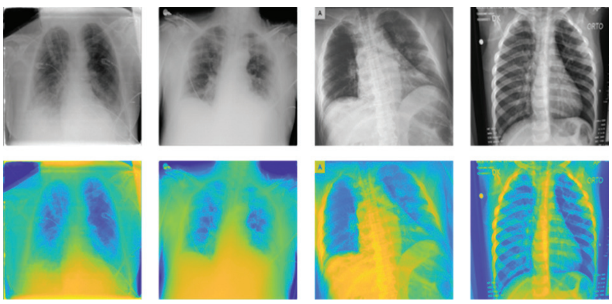

Fig. 3. Some Sample Indexed and Converted Color Images

3.4. CLAHE (CONTRAST LIMITED ADAPTIVE HISTOGRAM EQUALIZATION)

CLAHE is a variant of Adaptive Histogram Equalization (AHE) that deals with contrast over-enhancement. CLAHE works with small zones of the image, termed mosaics. Adjacent tiles are then joined using bilinear interpolation to eliminate artificial boundaries. This algorithm can be used to enhance the contrast of images.

We can also use CLAHE on color images, which generally works on the luminance channel. The results are much better after fitting to just the luminance channel of an HSV image. The standard architecture of CLAHE is shown in Fig. 4 below.

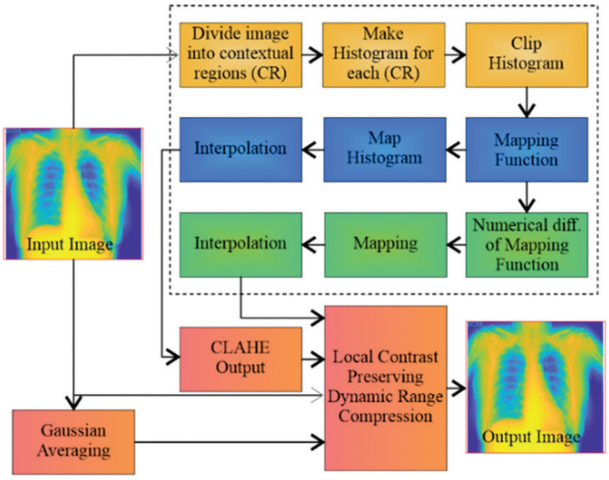

Fig. 4. Architecture of CLAHE

3.5. CNN (CONVOLUTIONAL NEURAL NETWORK)

A Convolutional neural network, generally called ConvNet or CNN, tends to be a deep learning algorithm that can take an input image and classify it among other images. One main thing differentiating CNN from other algorithms is that it requires much lower pre-processing than different algorithms; additionally, ConvNet has self-learning capabilities.

The general architecture of ConvNet is shown in Fig. 5 below. Here we have an RGB image as input. The element that performs the convolution operation at the very first is called the kernel; in the image, the kernel is shown with red color. The kernel will have the same depth as the input image if the image has multiple channels. The kernel will move all over the image to extract the high-level features. Then we have pooling layers, the same as the convolution layers pooling layers are responsible for reducing the spatial size of the convolved features. It reduces the computational power requirements to process extensive data. At last, fully connected layers are used to learn the non-linear combinations of features represented by the output of the convolutional layer.

3.6. EXPERIMENTAL ENVIRONMENT

All experimental simulations are carried out on a system with Intel i7 Processor, 8 GB RAM, and NVIDIA G-force 2 GB Graphic card. The simulation software used for all this is MATLAB 2020a. While the training process, we resized all images to 70 × 70, so all the illustrations should be constant according to size.

3.7. PERFORMANCE METRICS

The performance of the planned system is demonstrated in the form of a confusion matrix. Classification and validation accuracy is also used to calculate the projected model's performance.

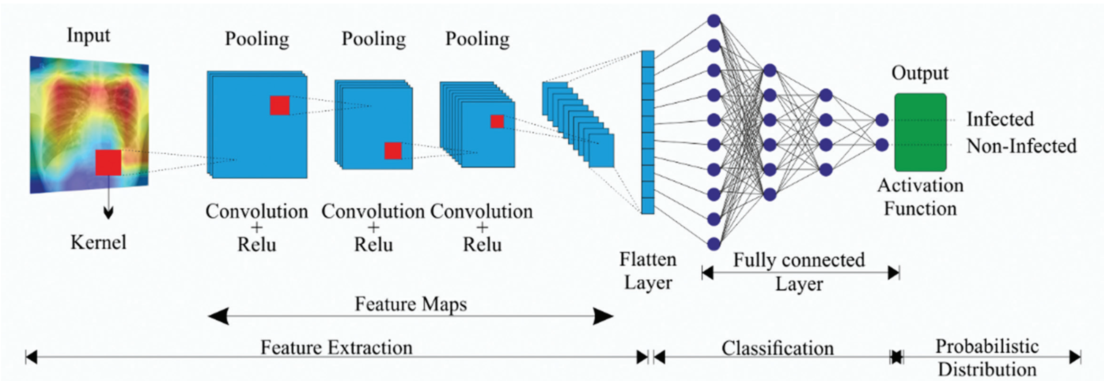

Fig. 5. Architecture CNN

Classification Accuracy is the most vital recital evaluation metric. One can calculate it as a (true positive + true negative) ratio with the total length.

$$Classification\ accuracy = \frac{TP + TN}{N}$$

Specificity is the ratio of true negatives with the sum of true negatives and false positives. Mathematically it can be represented as

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

Precision is the ratio of true positives with the sum of true positives and false positives. Mathematically it can be represented as

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

The Recall is the ratio of true positives with the sum of true positives and false negatives. Mathematically it can be represented as

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

F1 Score is the function of Precision and Recall. Mathematically it can be represented as

$$F1 = 2 \times \frac{Precision * Recall}{Precision + Recall}$$

Where **TP** = True Positive (How many times did the model correctly classify a Positive sample as Positive?)

FP = False Positive (How many times did the model incorrectly classify a Negative sample as Positive?)

TN = True Negative (How many times did the model correctly classify a Negative sample as Negative?)

FN = False Negative (How many times did the model incorrectly classify a Positive sample as Negative?)

N = Total Length = (*TP*+*TN*+*FP*+*FN*)

In the confusion matrix, correctly classified COVID-19 positive cases by the model are represented as TP, and those incorrectly classified as COVID-19 negative are represented as FP. Similarly, adequately categorized COVID-19 negative cases are categorized as TN, and wrongly classified as COVID-19 positive are termed FN.

4. EXPERIMENTAL RESULTS

This section provides a performance analysis of the planned system to categorize the COVID-19 infected X-ray scans. The model is trained for 50 epochs. The proposed method achieved an average accuracy of 99.8%, Precision of 92.6%, Recall of 99.81%, and F1 Score of 99.32% while determining the COVID-19 contamination among X-ray scans. To analyze the given model's effectiveness, some plots have been generated. The figures below show the confusion matrix ROC curve and the training and validation graph obtained. In the training and validation graph, spikes in blue show the obtained classification accuracy of the proposed system, and the markers in black are the validation lines.

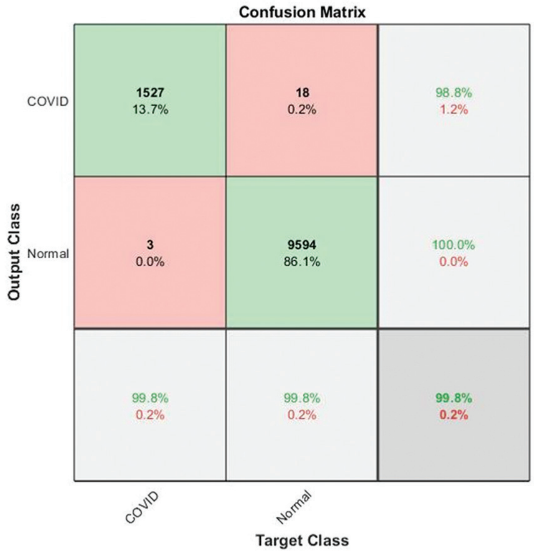

Fig. 6. Confusion Matrix

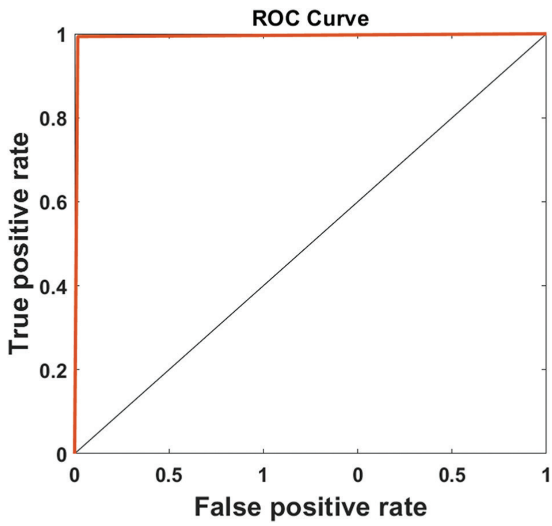

Fig. 7. ROC Curve

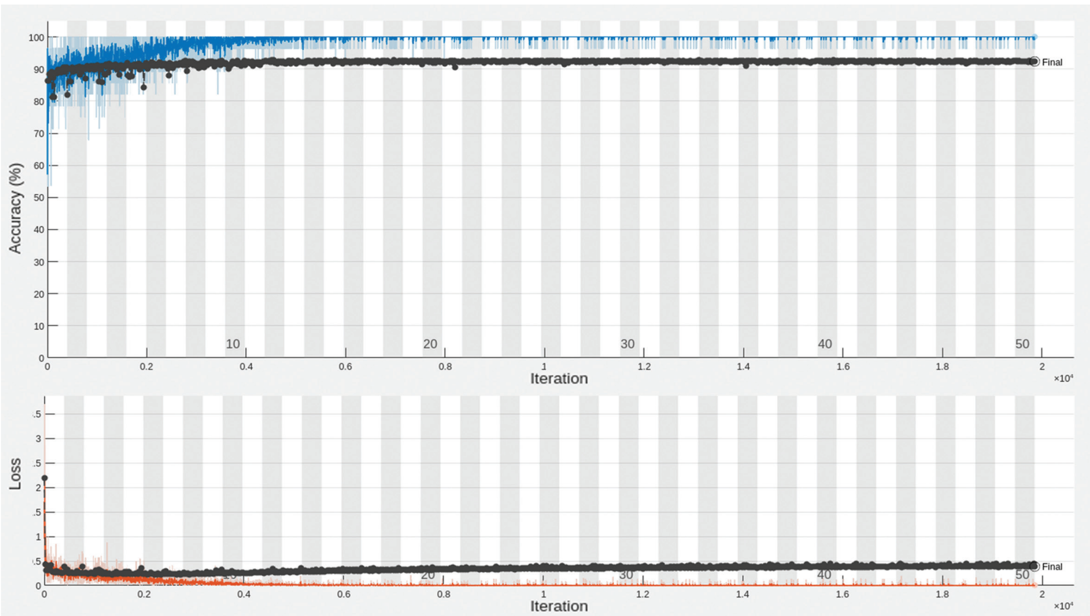

Fig. 8. Training and validation graph

4.1. K-FOLD CROSS VALIDATION

Additionally, the given model is tested with K-fold cross-validation. It is a statistical practice to analyze the performance of machine learning algorithms. This entire dataset is divided into k no of folds, and the performance is analyzed when new data is given. Here we have chosen three as the value of K to perform the analysis. We also tried with the values 4 and 5, but the results are almost similar, so we have decided to go with the minimum value of 3. The results obtained from k-fold validation are as follows. Where the three folds have given the accuracy of 99.7%, 99.8%, and 99.9%, respectively, if we take the mean of these values, we will get the same result as our model, which is 99.80% accuracy.

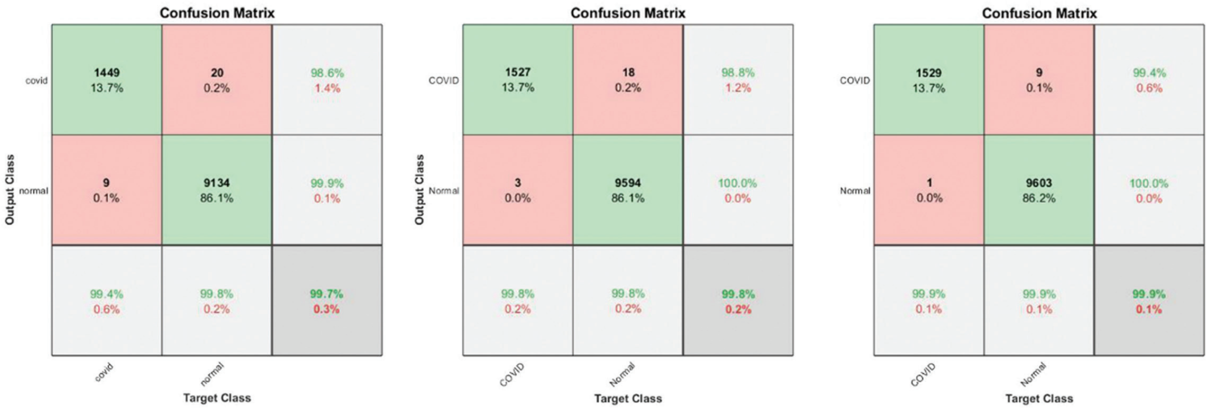

Fig. 9. Confusion matrix (K-Fold cross-validation)

As we can see from fig. 9 above, If we take the mean of these values, the three folds have given the accuracy of 99.7%, 99.8%, and 99.9%, respectively.

$$Average\ Accuracy\ = \frac{1}{K}\sum_{k=0}^{3}(99.7+99.8+99.9)$$

$$Average\ Accuracy\ = 99.80\%$$

As we can see, the final result is the same as the obtained result from a proposed model in terms of accuracy; now, we can say the proposed model is validated through k fold validation process.

4.2. COMPARISON WITH OTHER MODELS

In this section, we compare the proposed model with other existing models. The table below shows the comparison of different algorithms, such as PDCOVIDNET [10], VGG16 [18], ResNet50 [21], and HOG+CNN, with the proposed algorithm in terms of accuracy.

Table 2. Comparison with other existing algorithms

Method	Accuracy	Specificity	Precision	Recall	F1
PDCCOVIDNet	96.58	NE	96.58	96.59	96.58
VGG 16	93.8	NE	93.84	93.86	93.83
ResNet50	94.74	NE	92.67	92.15	92.12
HOG+CNN[20]	99.49	95.7	NE	NE	NE
Proposed CLAHE+CNN	99.81	99.81	98.83	99.81	99.32

*Where NE is not evaluated

PDCOVIDNET [13] in Dec 2020 used small convolution in a similar stack to capture and stretch the required properties to obtain a recognition accuracy of 96.58%, a precision of 96.58%, and a recall of 96.59%, and F1 Score of 96.58%. VGG16 [12] model as consisting of two stages, one is pre-processing, and the second is data dissemination and learning transfer, and lastly shows an accuracy of around 99.8%, Precision of 93.84%, Recall of 93.86%, and the F1 Score of 93.83%. ResNet50 [10] has the lowest parameters compared to all other models, with an accuracy of 94.74%, Precision of 92.6%, Recall of 92.15%, and the F1 Score of 92.12%. HOG+CNN [20] by Noor-A-Alam achieved an accuracy of 99.49% and a Specificity of 95.7%. The proposed system CLAHE+CNN has achieved the highest Accuracy of 99.81%, Precision of 92.6%, Recall of 99.81%, and F1 Score of 99.32%.

4.3. LIMITATIONS

Although the proposed algorithm has achieved a very high accuracy of 99.81%, every technique has some limitations. The proposed system works only on indexed images that can be converted into RGB color format. This was challenging for authors to find a large image dataset with all the images with the same graphical properties. The proposed technique is a promising tool in healthcare to classify COVID-19 infection.

5. CONCLUSION

We have presented a deep learning model by combining CLAHE and CNN to detect COVID-19 infection using chest X-ray images. CLAHE is used before CNN to enhance the contrast of X-ray images so the CNN model can classify the X-ray images more precisely. This research developed an intelligent system to identify COVID-19 infection using chest X-rays Images with great accuracy of 99.8% and low complexity. This is very encouraging that X-ray images are used to detect COVID-19 contamination at this level. This proposed system was more accurate than the results obtained from personal isolation techniques like HOG and CNN. Additionally, the given approach is validated with the same accuracy by using the k-fold authentication procedures. For future work, we will develop contactless image-capturing methods for front-end healthcare workers.

6. REFERENCES

[1] F. Wu, et al. "A new coronavirus associated with human respiratory disease in China", Nature, Vol. 579, 2020, pp. 265-269.

[2] W. Guan, Z. Y. Ni, Y. Hu, W. H. Liang, C. Q. Ou, J. X. He, L. Liu, C. L. Lei, "Clinical characteristics of coronavirus disease 2019 in China", New England Journal of Medicine, Vol. 382, No. 18, 2020, pp. 1708-1720.

[3] N. Chen, M. Zhou, X. Dong, J. Qu, F. Gong, Y. Han, Y. Qiu, J. Wang, Y. Liu, Y. Wei, L. Zhang, "Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study", The Lancet, Vol. 395, No. 10223, 2020, pp. 507-513.

[4] C. Wang, P. W. Horby, F. G. Hayden, G. F. Gao, "A novel coronavirus outbreak of global health concern", The Lancet, Vol. 395, No. 10223, 2020, pp. 470-473.

[5] N. Zhu, D. Zhang, W. Wang, X. Li, B. Yang, J. Song, et al., "A novel coronavirus from patients with pneumonia in China", New England Journal of Medicine, Vol. 382, 2020, pp. 727-733.

[6] Q. Li, X. Guan, P. Wu, X. Wang, L. Zhou, Y. Tong, R. Ren, K. K. S. Leung, E. H. Y. Lau, J. Y. Wong, et al., "Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia", New England Journal of Medicine, Vol. 382, 2020, pp. 1199-1207.

[7] L. Wang, A. Wong, "COVID-Net: A Tailored Deep Convolutional Neural Network Design for Detection of COVID-19 Cases from Chest X-Ray Images", arXiv:2003.09871, 2020.

[8] A. Afzal, "Molecular diagnostic technologies for COVID-19: Limitations and challenges", Journal of Advanced Research, Vol. 26, 2020, pp. 149-159.

[9] World Health Organization, "Use of Chest Imaging in COVID-19", www.who.int/publications/i/ item/use-of-chest-imaging-in-COVID-19 (accessed: 2021)

[10] K. He, X. Zhang, S. Ren, J. Sun, "Deep Residual Learning for Image Recognition", arXiv:1512.03385, 2015.

[11] A. Krizhevsky, S. Ilya, G. E. Hinton, "Imagenet classification with deep convolutional neural networks", Communications of the ACM, Vol. 60, No. 6, 2017, pp. 84-90.

[12] K. Simonyan, A. Zisserman, "Very deep convolutional networks for large-scale image recognition", arXiv:1409.1556, 2014.

[13] N. K. Chowdhury, M. M. Rahman, M. A. Kabir, "PDCOVIDNet: A parallel-dilated convolutional neural network architecture for detecting COVID-19 from chest X-ray images", Health Information Science and Systems, Vol. 8, 2020, pp. 1-14.

[14] I. U. Khan, N. Aslam, "A Deep-Learning-Based Framework for Automated Diagnosis of COVID-19 Using X-ray Images", Information, Vol. 11, No. 9, 2020.

[15] S. Minaee, R. Kafieh, M. Sonka, S. Yazdani, G. J. Soufi, "Deep-COVID: Predicting COVID-19 from chest X-ray images using deep transfer learning", Medical Image Analysis, Vol. 65, No. 101794, 2020.

[16] B. Sekeroglu, I. Ozsahin, "Detection of COVID-19 from Chest X-Ray Images Using Convolutional Neural Networks", SLAS TECHNOLOGY: Translating Life Sciences Innovation, Vol. 25, No. 6, 2020, pp. 553-565.

[17] N. E. M. Khalifa, M. H. N. Taha, G. Manogaran, M. Loey, "Retracted article: a deep learning model and machine learning methods for the classification of potential coronavirus treatments on a single human cell", Journal of Nanoparticle Research, Vol. 22, No. 11, 2020, pp. 1-13.

[18] A. T. Sahlol, D. Yousri, A. A. Ewees, M. A. A. Al-Qaness, R. Damasevicius, M. A. Elaziz, "COVID-19 image classification using deep features and fractional-order marine predator's algorithm", Scientific Reports, Vol. 10, No. 1, 2020, pp. 1-15.

[19] M. M. Rahman, S. Nooruddin, K. M. A. Hasan, N. K. Dey, "HOG+CNN Net: Diagnosing COVID-19 and Pneumonia by Deep Neural Network from Chest X-Ray Images", SN Computer Science, Vol. 2, No. 5, 2021.

[20] N. A. Alam, M. Ahsan, Md. A. Based, J. Haider, M. Kowalski, "COVID-19 detection from chest X-ray images using Feature fusion and deep learning", Sensors, Vol. 21, No. 4, 2021.