Scientific Journal of Impact Factor (SJIF): 5.71
E-ISSN (O): 2348-4470
P-ISSN (P): 2348-6406

International Journal of Advance Engineering and Research Development

Volume 10,Issue 4,April 2023

COVID-19 detection from Chest X-Ray Scans using Machine Learning

Ms Akanksha Mishra[1], Ms. Ruchi Dronawat[2], Dr. Ritu Shrivastava[3]

[1]MTech Scholar, Computer Science & Engineering, Sagar Institute of Research & Technology, Bhopal, MP,akanksha.itconsultant@gmail.com

[2]Assistant Professor, Computer Science & Engineering, Sagar Institute of Research & Technology, Bhopal, MP, dron.ruchi@gmail.com

[3]Professor, Computer Science & Engineering, Sagar Institute of Research & Technology, Bhopal, MP, ritushrivastava08@gmail.com

Abstract: *This paper presents an improved and effective Deep Learning (DL) based Covid-19 detection model wherein modifications has been done on feature extraction and classification phase. The main aim of our model is to enhance accuracy and also reduces computing complexity. Initially, medical chest X-Ray image samples were collected from Kaggle dataset, in which images are categorized into three classes of Normal, Covid-19 and Pneumonia. Since, these images have poor quality due to the presence of noisy, redundant and unnecessary data, therefore pre-processing technique is applied. During the pre-processing phase, the images are resized, noisy and redundant data is removed and colored images are converted into grayscale image for improving quality. Then, employing PCA and GLCM feature extraction algorithms, useful and crucial features were extracted from the processed x-ray pictures.Additionally, CNN is utilized as a classifier to divide images into the Covid-19, Normal, and Pneumonia classes. The usefulness and productivity of proposed CNN model is analyzed and verified bycomparing it with a few classical models including KNN, RF, and SVM in terms of several dependent aspects.*

Paper Received:20/04/2023 Paper Accepted:26/04/2023 Paper Published:10/05/2023

I. INTRODUCTION

In the year 2019, a novel disease named as, COVID-19 was discovered in the Wuhan city of China. It is a severe acute respiratory syndrome that mainly affects the respiratory organs and is caused by Coronavirus or SARS-COV-2 [1-2]. Since its emergence, COVID-19 spreaded like a wildfire and dominated more than 213 countries and territories thereby causing medical emergency. Owing to its varying forms and different levels of severity ranging from moderate to severe, with the possibility of organ failure and death, the covid-19 has proven to be a difficult condition. In case of the COVID19 epidemic, the government's efforts are extremely crucial to stop the spread of infectious disease. Some of the crucial steps taken by government include locking down the area to stop infection spread, getting ready the healthcare system to handle the epidemic, offering a crisis package to reduce the impact of the pandemic on the economy and population, and adopting policies that can be adjusted to the COVID19 situation and so on [3].The World Health Organization (WHO) advises the use of masks as part of a comprehensive package of preventative and control measures to lessen the propagation of COVID-19. However, even when utilized properly, a mask cannot provide adequate protection or source control.Therefore, other precautions like maintaining distance, covering your mouth while sneezing or coughing and isolating and seeking medical helpwhen symptoms like high fever, body ache, cough, cold gets adverse.

COVID-19 early diagnosis is critical not only for patient treatment, but also for public health, since it ensures patient isolation and pandemic containment. There were no recognized techniques to combat with the disease in the beginning, because of the rarity of the disease. But screening and fast diagnosis of infected individuals, as well as their seclusion from the general population, were regarded as critical measures by the researchers. Although being the standard method for diagnosing COVID-19, the RT-PCR diagnostic has drawbacks with certain traits that make the disease difficult to recognize. It is a time-consuming, complicated, expensive, and human dependent procedure [4-5].More importantly, numerous researchers have found that the RT-PCR test has a poor sensitivity. Various studies have found that the sensitivity of this diagnostic approach ranges from 30% to 60%. This implies that the accuracy of COVID-19 diagnosis has decreased in many situations. Its false-negative rate and inconsistent outcomes have also been mentioned in several researches.Scientists got prompted to work on cutting-edge research to help front-line medical workers mitigate, identify, and prevent the virus by the COVID-19 pandemic.Scientists using digital technology have devised unique approaches to combat the epidemic, apart from medical doctors and virologists. In the fight against COVID-19, two key scientific communities supported by digital technology may be identified. The primary digital endeavor in this area is automatic COVID-19 identification from CT scans and X-ray pictures, which comes from the Artificial Intelligence (AI) sector [6].ML approaches can help the battle against COVID-19 on numerous fronts using publicly available data sources. The most common use is ML-based COVID-19 diagnosis from CT scans and X-rays, which helps alleviate the strain on RT-PCR test kits, which are in short supply. However, the availability of open source data is the biggest barrier to using machine learning algorithms. The efficiency of scanning all high-dimensional datasets and the training of the model to find the most appropriate task are the two main challenges that ML algorithms face [7-8]. Therefore, the authors shifted their attention towards Deep Learning based systems. In the sphere of healthcare, the use of DL methods has led to new achievements. Numerous sources of medical data, such as magnetic resonance imaging (MRI), X-ray, Positron Emission Tomography (PET), and computerized tomography scan (CT scan), have given clinicians with massive amounts of data in the real world. Much research has been done on data processing linked to DL algorithms, particularly CNN, since the outbreak of COVID-19. These researches have focused on the detection and differential diagnosis of COVID-19 using various algorithms and DL structures.

The remaining section of paper is categorized as, Section 2 reviews some of the recently published models for detecting covid-19 using chest X-ray images. Section 3 gives an overview of the proposed model followed by its working mechanism. Section 4 represents the results obtained for given model and finally a conclusion of the paper is written in section 5.

II. LITERATURE SURVEY

In the past, a number of methods have been proposed by researchers for identifying COVID-19 through chest x-rays. Some of the recently published methods are reviewed and discussed briefly here; **D.A. Garzón, et al.[9],** utilized VGG19 and UNET techniques in their work and proposed an effective model that classify patients as positive or negative for covid-19. They also implemented a pre-processing method for segmenting lung regions and removing unnecessary data. Results simulated that proposed CNN model achieves an accuracy of 97% in detecting covid-19. **Rahaman MM, et al. [10],** an effective covid-19 detection model was presented that was based on deep transfer learning techniques. the authors of this paper examined 15 various previously trained CNN models in order to find the better covid-19 detection model. The simulation results revealed that VGG19 model is giving best results for detecting covid-19 with an accuracy of 89%. **Loey, M et al. [11],**used GAN along with deep transfer learning techniques for identifying and detecting covid-19 in patients through chest x-ray images at early stages. The authors used Alex net, Google net and Restnet18 classifiers for analyzing and classifying data. The efficacy of three classifiers were analyzed in three scenarios in which google net gives best accuracy result of 80% in first case, while as, in second and third case Alex net and google net performed best and achieved an accuracy of 85% and 99% respectively. **MRahimzadeh, et al. [12],**proposed an effective covid-19 detection model in which Xception and ResNet50V2 networks were concatenated along with NN for analyzing chest x-ray images effectively.Results simulated that proposed model was able to achieve an accuracy of 91 for detecting normal, covid-19 and pneumonia classes. **Abbas, A., et al. [13],** authors proposed a DeTraC (i.e. Decompose, Transfer, and Compose)model for detecting and classifying covid-19 chest x-ray images. Results simulated that proposed model was able to achieve an accuracy of 93.1% for detecting covid-19 in chest x-ray images. **S. Asif, et al. [14],**proposed a DCNN based inception V3 that was integrated along with transfer learning for the purpose of identifying covid-19 pneumonia by using chest x-ray images. Through extensive experimentation, the proposed model yielded an accuracy of more than 98%. **Qi, X., et al. [15],**a unique model that was based on multi feature CNN framework was developed for identifying the covid-19 patients via CXR images through multi feature classifications. Results showcased that multi featured based CNN model was giving more accurate and efficient results than single featured CNN model. **Agrawal, T. et al. [16],** presented a Deep CNN based network for identifying and detecting the Covid-19 in CXR images. The suggested model was validated on different datasets. **Degerli, A. et al. [17],** an effective model was proposed that can recognize joint localization, severity grading and covid-19 from chest x-ray images by producing infection maps. The model achieved an detection rate of 94% on large dataset.

After analyzing the literature survey, we observed that a number of approaches have been developed since the rise of COVID-19 pandemic for detecting and recognizing it in early stages. However, the current covid-19 detection model faced a lot of issues, degrading their detection rate and performance. Moreover, we also observed that majority of researchers utilized Machine learning (ML) classifiers for analyzing and detecting covid-19 in patients. Unfortunately, the problem with these ML based detection models is that they cannot handle large and huge datasets and often lead to overfitting issues. Furthermore, not much work was done on extracting texture features which play a crucial role in detection of covid-19. However, in some cases where feature extraction techniques were implemented also faced issues of low computational speed while detecting covid-19 in large scaled images. Keeping these facts in mind, a new and improved COVID-19 detection method will be presented in this paper that can overcome the limitations of current systems.

III. PROPOSED WORK

In this section, overview of the proposed covid-19 detection model along with its working mechanism is discussed. As discussed earlier that existing model undergo through various limitations that hinder their accuracy rate and make them computationally complex. Therefore, an improved and effective COVID-19 detection model is presented in this manuscript in which modifications have been done in feature extraction phase and classification phase. The main motive of the proposed approach is to decrease computational complexity while also increasing its detection rate. The proposed model works under different phases of data acquisition, pre-processing, feature extraction and classification, each phase is explained briefly in the coming up sub section. Initially, medical Chest X-ray image samples have been taken from Kaggle dataset, in which a total of 5863 images are present that are categorized into three classes of Normal, Covid-19 and pneumonia. Nonetheless, these images contain a lot of noisy and unnecessary data that degrade their quality and also make it difficult to extract features from it. Therefore, we have implemented pre-processing technique on these raw medical images and eliminate noisy and unwanted data. Moreover, the raw medical mages are resized and converted into gray scale so that features can be extracted from it efficiently. Furthermore, we have seen that exiting detection models were not paying any attention for extracting textual features from images, that play crucial role in x-ray images. To overcome this limitation, we used Principal Component Analysis (PCA) along with the GLCM feature extraction technique. The PCA is used to divide the data into new subcategories that have identical or lower dimensions. PCA initially determines which eigenvectors of a covariance matrix have the highest eigenvalues. A given set of x-ray pictures is utilized to determine second order statistical textural features using GLCM. The suggested technique is able to retrieve all the required textual features that are stored collectively to create a single feature vector by using the hybrid feature extraction technique.

In addition to this, the proposed model must also employ an efficient classifier that can deal with huge, complicated, and non-linear datasets. To overcome this, we have utilized DL classifier in our work that is able to handle large, complex and non-linear datasets. Deep learning techniques are further of two types one is CNN and other is RNN, but we have used CNN in the proposed model because it produces more accurate findings on images than RNN. In order to produce useful findings, each x-ray image is processed by multiple layers of CNN for analyzing them. Consequently, an efficient and successful model for the covid-19 disease is created by training the CNN classifier with characteristics produced by the PCA-GLCM approach. The following portion of this paper provides a thorough overview of the model.

IV. METHODOLOGY

The proposed CNN based Covid-19 detection model undergoes through various stages for attaining the desired objective. The step by step working mechanism of the approach is explained briefly here.

Data Acquisition: In the very beginning of our approach, all the necessary information needs to be collected from publicly available datasets. Here, we have collected medical chest x-ray sample images from Kaggle.com. The dataset comprises a total of 5863 X-ray images that are categorized into three classes of normal, covid-19 and pneumonia. The sample images taken from dataset for three classesis shown in figure 1.

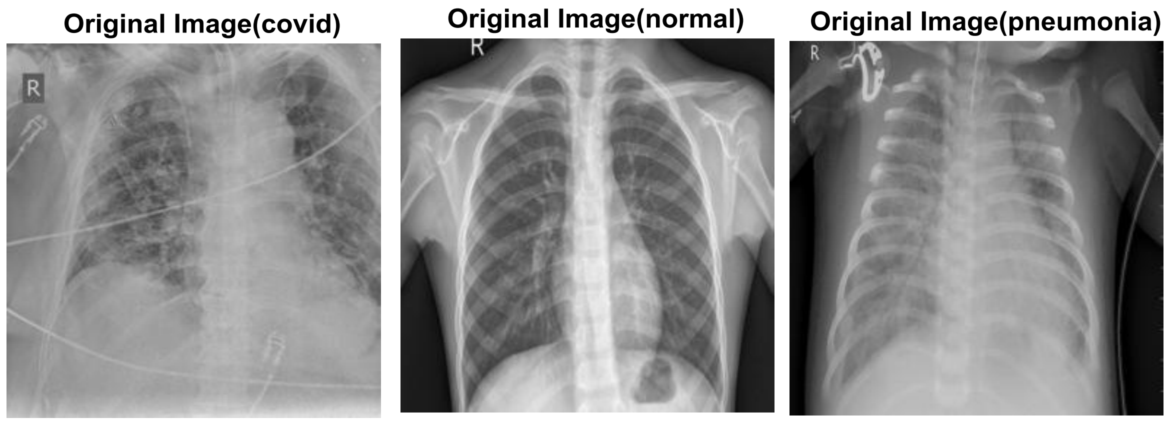

Figure 1. Original sample images taken from Dataset

Pre-processing: Since the images taken from the dataset are of poor quality and contain a lot of extraneous, irrelevant, and noisy data that reduces their quality, therefore applying pre-processing technique becomes crucial. In our approach, the raw medical images are processed by removing all the unnecessary and duplicate data. Also, the x-ray images are resized and converted into Grayscale for extracting features more efficiently and effectively.

Feature Extraction: After this, feature extraction technique is implemented on processed images in order to handle dimensionality and complexity issues. In our model, we have used PCA-GLCM feature extraction techniques for extracting only informative and crucial texture features from processed chest x-ray images for enhancing their quality. The enhanced image for three classes is depicted in figure 2.

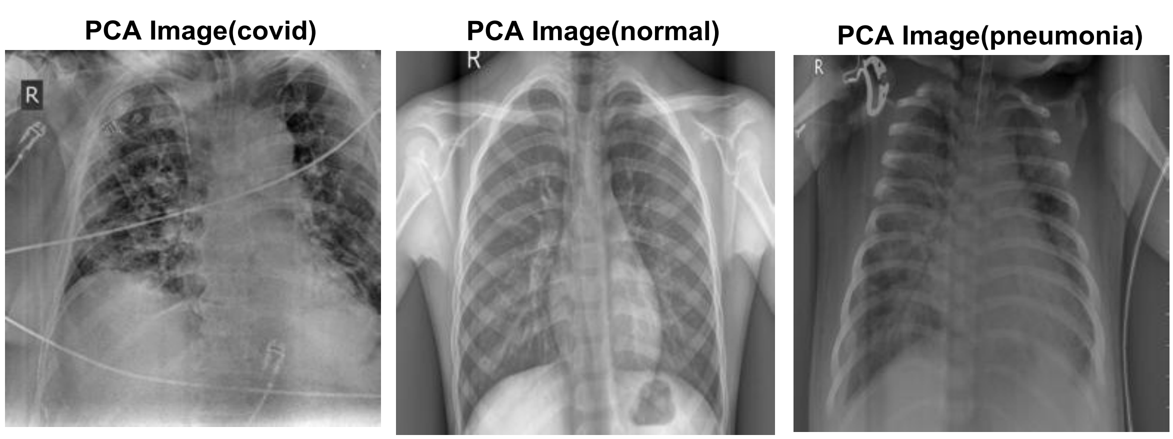

Figure 2.Enhanced images after applying PCA

A single feature vector is created by combining the characteristics that were gathered for the three categories. The data in this feature vector is then split into two groups—training data and testing data—in a proportion of 70:30. The classifier is trained using 70% of the data, and its performance is assessed using the remaining 30%.

CNN Initialization: In this stage,CNN model is initialized wherein, various parameters like epochs size, learning rate, total number of layers, batch size, input layer etc. are defined. Other than this, a number of other factors were also defined whose specific values are given in table 1.

Table 1. CNN initialization Parameters

CNN Parameter	Value
Max Epochs	50
Initial Learn Rate	0.001
Validation Frequency	20
Learn Rate Drop Period	5
No of Layer	15
Min. Batch Size	64
Input Layer	32 x 32
Conv. Filter Window size	3 x 3
No of stride	2 x 2
No of Pool Size	2 x 2
Training Set Size	70%
Testing Set	30%

Classification: Once the network is initialized**,** classification process starts wherein testing data is passed to the classifier. The CNN analyzes the given data and tries to match the features of the given image samples with the feature vector and eventually categorizes given image as normal, covid-19 and pneumonia.

Performance Analysis: In the last step of proposed approach, the efficacy and effectiveness of the suggested CNN model is validated and verified by comparing it with traditional models in terms of various performance dependency factors.

V. RESULT OBTAINED

The efficacy of proposed CNN model is analyzed and validatedin MATLAB software. Theresults were obtained by comparing the proposed CNN model with few traditional ML models interms of accuracy, precision, recall and F-score. The results obtained are thoroughly discussed in this section.

VI. PERFORMANCE EVALUATION

In order to prove the accuracy of proposed CNN approach over other similar models, we firstly compared its performance with ML classifiers like KNN, RF and SVM in terms of their overall accuracy. The comparison graph obtained for the same is shown in figure 3. According to the given graph, accuracy was observed at 99.57% in the suggested CNN model, compared to 88.89% in the standard KNN model, 96.2% in RF, and 98.14% in the SVM model. The specific value of accuracy obtained in each model is depicted in table 2.

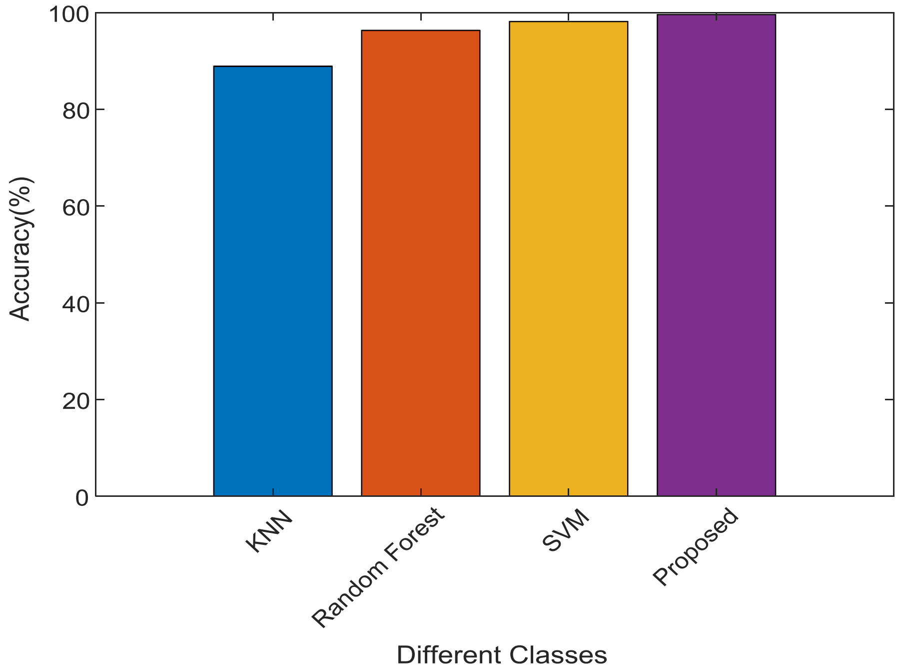

Figure 3. Comparison graph for accuracy

Table 2. Accuracy obtained in proposed CNN model

Algorithm	**Accuracy**
KNN	88.89
Random Forest	96.29
SVM	98.14
Proposed	99.573

Additionally, we have also examined and contrasted the precision performance of the suggested CNN approach with that of conventional KNN, RF, and SVM techniques for three classes. Figure 4, shows the comparison graph of precision obtained for three classes. The presented CNN model produced high precision results of1 for covid-19 and normal and 0.98 for identifying pneumonia in patients. On the other hand, the value of precision was mounted at merely 0.84, 0.91, and 1 in KNN; 0.93, 0.98, and 1 in RF; and 1, 0.98, and 0.95 in SVM for categorizing three classes of covid-19, normal and pneumonia. The specific value of precision is shown in table 3.

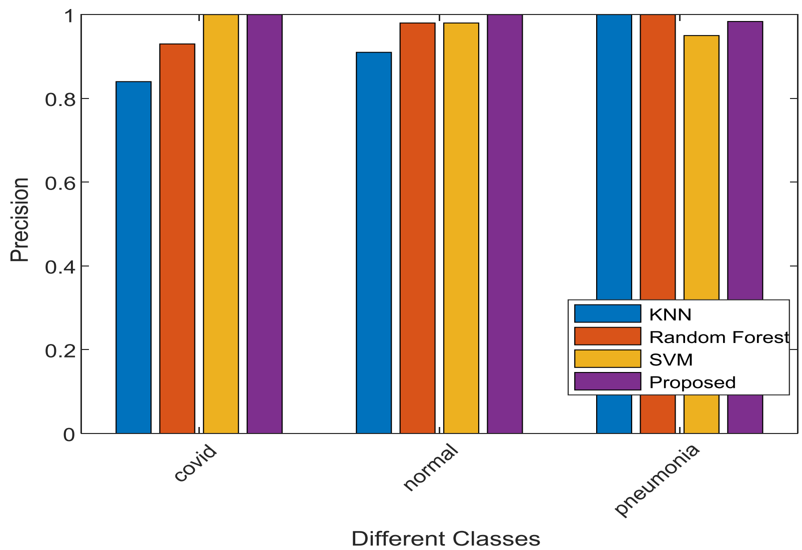

Figure 4. Comparison graph for precision for three classes

Table 3: Specific values of precision in various classifiers

Class	KNN-precision	RF-Precision	SVM-precision	CNN-Proposed
Covid-19	0.84	0.93	1	1
Normal	0.91	0.98	0.98	1
Pneumonia	1	0	0.95	0.98361

Moreover, the efficiency of the proposed CNN model is also analyzed and compared with traditional models (KNN, RF and SVM) models in terms of F-score for three classes of normal, covid-19 and pneumonia, as shown in figure 5. It is evident from the graph,that Fscore values for the KNN model, RF, and SVM models were 0.9, 0.93, & 0.74; 0.97, 0.95, &0.98; and 1, 0.95& 0.95 for covid-19, normal and pneumonia. On the other hand, in the suggested CNN model for identifying COVID-19, normal, and pneumonia in chest x-ray images, the value of Fscore was mounted at 0.98, 1 and 0.99.

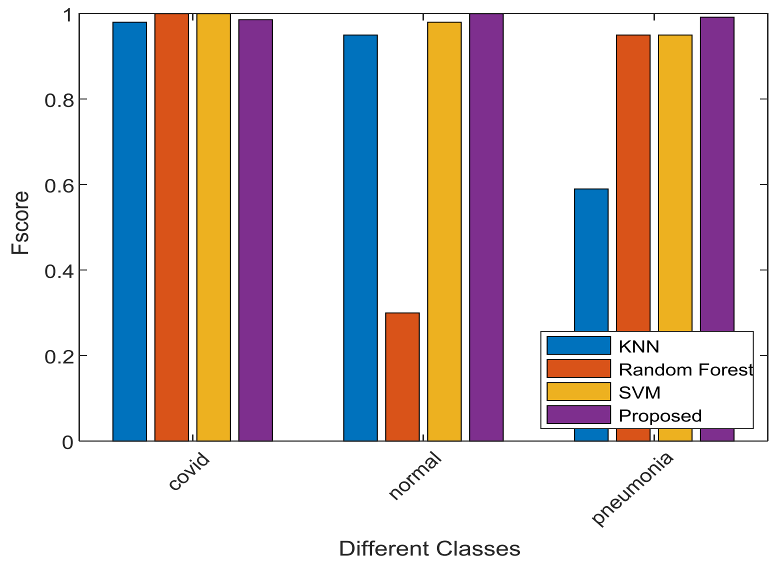

Figure 5. Comparison graph for Fscore

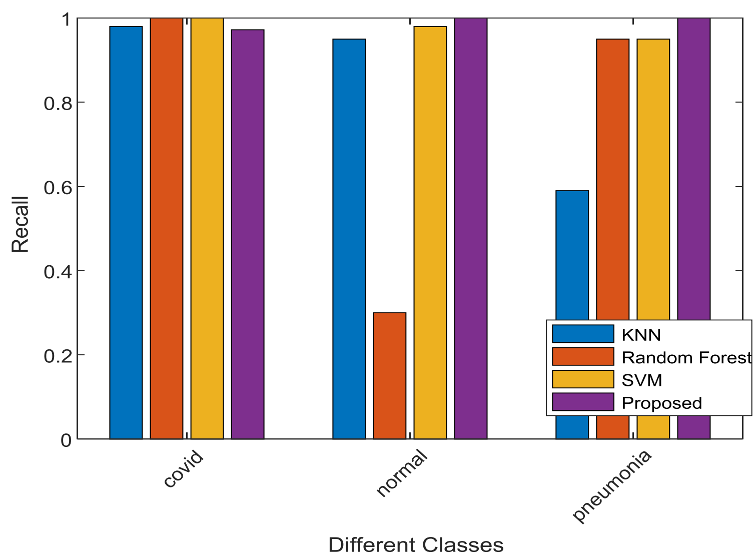

Figure 6.Comparison graph for Recall

Additionally, we have also validated the performance of proposed CNN model with respect to other models in terms of their Recall value for three classes (see figure 6). According to the graph, recall value for the COVID-19 sample was mounted at 0.97, while it was 1 for normal and pneumonia in proposed CNN model. Whereas, it was 1, 0.3, 095 in RF, 1, 0.98, 0.95 in SVM models, and 0.98, 0.95, 0.59 in KNN. These results demonstrate the proposed model's superior accuracy over other models. The specific value of precision and recall is recorded in tabular form and is shown in table 4.

Table 4: Specific values of F-Score and recall in various classifiers

Class	F-Score				Recall			
	KNN	RF	SVM	CNN	KNN	RF	SVM	CNN
Covid-19	0.9	0.97	1	0.98595	0.98	1	1	0.97222
Normal	0.93	0.95	0.95	1	0.95	0.3	0.98	1
Pneumonia	0.74	0.98	0.95	0.99174	0.59	0.95	0.95	1

From the above graphs and tables, it can be concluded that proposed CNN model is outperforming all traditional models in terms of accuracy, precision, recall and F-score as well to prove its dominance.

VII.CONCLUSION

In this paper, an effective and efficient covid-19 detection model is presented that is based on CNN. The performance and usefulness of the proposed model is analyzed and compared with few traditional models in MATLAB software. The simulated outcomes were determined in terms of accuracy, precision, recall and F-measure for three categories of Normal, covid-19 and pneumonia. The simulations showed that the proposed CNN model produces the maximum accuracy of 99.5%, compared to the classic KNN, RF, and DVM methods with 88%, 96%, and 98%. Additionally, the precision of the suggested CNN model was examined, and the results showed that its productivity was 1 for the COVID-19 and normal classes and 0.98 for the pneumonic class.On the contrary, for the covid-19, normal, and pneumonia categories, the measure of precision in classic KNN, RF, and SVM was set at 0.84, 0.91 & 1; 0.93, 0.98 & 1; and 1, 0.98, 0.95. Similarly, for the Covid-19, normal, and pneumonia classes, respectively, the recall scores were 0.98, 0.95, and 0.59 in KNN, 1, 0.3, 0.95 in RF, 1, 0.98, and 0.95 in SVM, while it was 0.97, 1 &1 in Proposed CNN. Additionally, we noted that the proposed model outperforms conventional classifiers in terms of Fscore, achieving values of 0.98, 1 and 0.99 for the covid-19, normal, and pneumonia classes, respectively.These statistics prove the supremacy of proposed CNN model over other similar covid-19 detection models.

REFERENCES

1. S. Shaikh, J. Gala, A. Jain, S. Advani, S. Jaidhara and M. Roja Edinburgh, "Analysis and Prediction of COVID-19 using Regression Models and Time Series Forecasting," 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), 2021, pp. 989-995, doi: 10.1109/Confluence51648.2021.9377137.
2. R. Punia, L. Kumar, M. Mujahid and R. Rohilla, "Computer Vision and Radiology for COVID-19 Detection," 2020 International Conference for Emerging Technology (INCET), 2020, pp. 1-5, doi: 10.1109/INCET49848.2020.9154088.
3. Pham, Q.-V., Nguyen, D. C., Huynh-The, T., Hwang, W.-J., &Pathirana, P. N. (2020). Artificial Intelligence (AI) and Big Data for Coronavirus (COVID-19) Pandemic: A Survey on the State-of-the-Arts. IEEE Access, 1–1.
4. G. Lippi, A.-M. Simundic, and M. Plebani, “Potential preanalytical and analytical vulnerabilities in the laboratory diagnosis of coronavirus disease 2019 (COVID-19),” Clinical Chemistry and Laboratory Medicine (CCLM), vol. 58, no. 7, p. 1, 2020.
5. C. Sheridan, “Coronavirus and the race to distribute reliable diagnostics,” Nature Biotechnology, vol. 38, no. 4, p. 382, 2020.
6. Shuja, J., Alanazi, E., Alasmary, W., &Alashaikh, A. (2020). COVID-19 open source data sets: a comprehensive survey. Applied Intelligence, 1–30. Advance online publication.

7. L. Huang, R. Han, T. Ai, P. Yu, H. Kang, Q. Tao, and L. Xia, ''Serial quantitative chest CT assessment of COVID-19: A deep learning approach,'' Radiol., Cardiothoracic Imag., vol. 2, no. 2, Apr. 2020, Art. no. e200075.
8. Bhattacharya, S., Reddy Maddikunta, P. K., Pham, Q. V., Gadekallu, T. R., Krishnan S, S. R., Chowdhary, C. L., Alazab, M., & Jalil Piran, M. (2021). Deep learning and medical image processing for coronavirus (COVID-19) pandemic: A survey. Sustainable cities and society, 65, 102589.
9. Daniel Arias-Garzón, et al., "COVID-19 detection in X-ray images using convolutional neural networks", Machine Learning with Applications,Volume 6,2021,100138,ISSN 2666-8270
10. Rahaman MM, Li C, Yao Y, Kulwa F, Rahman MA, Wang Q, Qi S, Kong F, Zhu X, Zhao X. Identification of COVID-19 samples from chest X-Ray images using deep learning: A comparison of transfer learning approaches. J Xray Sci Technol. 2020
11. Loey, M.; Smarandache, F.; M. Khalifa, N.E. Within the Lack of Chest COVID-19 X-ray Dataset: A Novel Detection Model Based on GAN and Deep Transfer Learning. *Symmetry* **2020**, *12*, 651
12. Mohammad Rahimzadeh, AbolfazlAttar,A modified deep convolutional neural network for detecting COVID-19 and pneumonia from chest X-ray images based on the concatenation of Xception and ResNet50V2,Informatics in Medicine Unlocked,Volume 19,2020,100360,ISSN 2352-9148
13. Abbas, A., Abdelsamea, M.M. & Gaber, M.M. Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network. *Appl Intell* **51,** 854–864 (2021).
14. S. Asif, Y. Wenhui, H. Jin and S. Jinhai, "Classification of COVID-19 from Chest X-ray images using Deep Convolutional Neural Network," *2020 IEEE 6th International Conference on Computer and Communications (ICCC)*, 2020, pp. 426-433
15. Qi, X., Brown, L.G., Foran, D.J. *et al.* Chest X-ray image phase features for improved diagnosis of COVID-19 using convolutional neural network. *Int J CARS* **16,** 197–206 (2021)
16. Agrawal, T., Choudhary, P. FocusCovid: automated COVID-19 detection using deep learning with chest X-ray images. *Evolving Systems* (2021).
17. Degerli, A., Ahishali, M., Yamac, M. *et al.* COVID-19 infection map generation and detection from chest X-ray images. *Health Inf Sci Syst* **9,** 15 (2021).