**Evaluating Median Accuracy of ResNet50 and VGG16 Models in COVID-19 Detection**

**Mallisetty Siva Mahesh1,Kattamuri B N Ayyappa2, Maddela Murali3 ,and Mididoddi Surendra Babu4**

1 Designation, Department, College/University/Organization, City, Country

2 Second affiliation, Address, City, Country

3 Third affiliation, Address, City, Country

**ORCID ID** of First Author-

**ORCID ID** of Second Author-

**ORCID ID** of Third Author-

Correspondence should be addressed to First Author Name; abc@gmail.com

Copyright © 2023 Made First Author Name 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.

**ABSTRACT-** The increasing number of Covid-19 cases and the lack of reliable, quick-to-use testing tools herald a new era in X-ray analysis employing deep learning methods .The Covid-19 virus's emergence poses a threat to human existence. Therefore, it will take time to develop a quick and accurate method of identifying the Covid-19 virus in patients. The reference method is the conventional RT-PCR technique .The goal of this research is to create an automated system that uses CNN models such as Resnet50, VGG16, and Grad-CAM to analyze X-ray images in order to provide a reliable and effective method of diagnosing Covid-19 infection .The created models use image processing techniques to preprocess the X-ray picture. Afterwards, deep learning is used to classify the images after they have been segmented and transformed. The CNN model that is being used provides strong classification accuracy and shows the location in the lung where the disease is attacked, Even for a normal person, we can anticipate the likelihood of where the COVID may affect them. Our model utilizes a convolution neural network that is trained on the standard COVID-19 Radiography Dataset.

KEYWORDS-CNN,Resnet50,VGG16,Grad-CAM,Covid

# **INTRODUCTION**

In early spring of 2020, the virus had already made its presence felt in most countries. By the end of March 2020, the World Health Organization (WHO) officially labeled the new virus as a pandemic [8]. As per the data provided by the WHO, by April 30, 2021, the virus had affected more than 157 million individuals, with the death toll surpassing 3 million [31].

While coronaviruses are not a new phenomenon, SARS-CoV-2 is anything but ordinary [28]. It is highly likely that this virus originated from an animal reservoir [32]. The treatment protocols for COVID-19 differ from those used for other coronavirus infections [33]. Our understanding of the disease remains limited and is continuously evolving along with the progress of the pandemic [24]. Commonly reported symptoms of COVID-19 include fever, coughing, fatigue, a sore throat, and body aches. Loss of taste or smell has been widely reported worldwide. In rarer but more severe cases, patients may experience difficulty breathing, high fever, chills, fatigue, muscle aches, or even succumb to the illness [34].

Artificial Intelligence (AI) models represent a promising solution [15]. The deep learning approach, known for its exceptional accuracy, has gained widespread acceptance and success in medical image classification applications. Recent advancements in deep learning technology have contributed to the development of intelligent diagnostic systems, aiding healthcare professionals in making informed decisions about patients' health. For instance, Lopez et al. addressed the issue of skin lesion classification, particularly the early identification of melanoma, in their research [23]. They proposed a deep learning method to address the challenge of classifying and identifying skin lesions as either malignant or benign.

# **LITERATURE SURVEY**

There are several works that use deep learning on chest X-ray images to recognize patients suffered from COVID-19. Here we limit our attention to those closely related to our proposal.

Ioannis et al. [3] assess the performance of the latest CNN architectures used in recent years for medical image classification, namely VGG19, MobileNet v2, Xception, Inception-ResNetv2 and Grad-CAM. The author uses deep learning because it performs well in detecting various abnormalities in small medical image data sets [9]. They used 1,442 X-ray data sets from patients, including 714 cases of bacterial pneumonia and viral pneumonia, and 224 cases of confirmed COVID-19 disease and 504 healthy cases. The results show that among the remaining CNNs, MobileNet-v2 and VGG19 offer the best classification in terms of accuracy. While VGG19’s outperforms the other techniques in terms of accuracy (reaching 98.75 %), MobileNet-v2 shows better performance regarding the sensitivity and the specificity (reaching 99.10 and 97.09 %, respectively).

Narin and others solved the problem of the limited supply of COVID-19 test kits accessible in public hospitals. [20] proposes to apply an automatic detection system as another rapid diagnosis recourse to avert the spread of COVID-19 and its pressure on medical institutions. In order to detect patients infected with coronavirus pneumonia, the author proposed three CNN based models (Inception-ResNetV2, Grad-CAM and ResNetV2), using totally 100 chest Xray images (50 health images and 50 COVID-19 images). In view of the high-performance results obtained, we can conclude that compared with the other two proposed models, the pretrained ResNet50 model achieves 98 % accuracy (Grad-CAM reach an accuracy of 97 % and Inception-ResNetV2 reach only an accuracy of 87 %).

For the sake of accurately detect COVID-19 and help overcome the lack of specialist doctors in remote villages, Ozturk et al. [21] propose a new model that uses raw radiographic images to automatically detect COVID-19. Among the proposed models, the DarkNet model is utilized as the classifier of the real-time object detection system (YOLO), conceived of 17 convolutional layers. The authors proceed by different filtering at the level of each layer. The purpose of this algorithm is to offer an accurate diagnosis for two-class classification (COVID/no-findings) and a multi-class classification (COVID/no-findings/pneumonia). The classification accuracy for the binary classification is 98.08 %, and for the multi-classification is 87.02 %.

# **METHODOLOGY**

In this study, we train, evaluate and test three well-known pre‐trained deep learning architectures to classify chest radiography images for a two-class classification: COVID-19 and Normal chest X-ray. VGG16 , ResNet50 and Grad-CAM are mainly used as deep learning models.

Figure 1 provides a general overview of the methodology of this study that describes the deep learning proposed method based on a simple standard pipeline, that is, chest image preprocessing, and then the classification model obtained through deep learning. After preprocessing the data, a deep model is trained. For instance, we will do a fine-tuning which includes an unfreezing of some of the top layers of the frozen model that is utilized for feature extraction, and then training both the freshly added part of the model (in our experiment, the fully-connected classifier) and the unfreezing top layers.

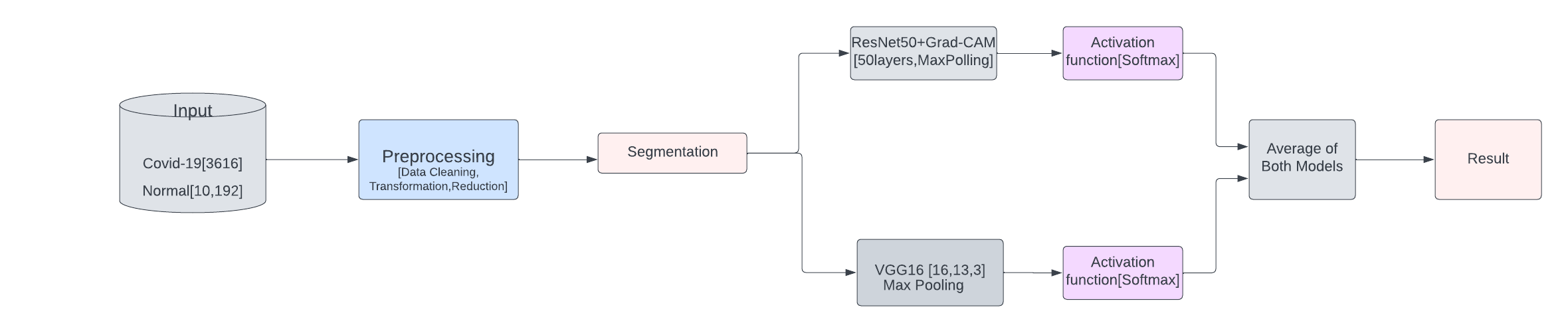


Fig. 1 Overview for COVID-19 and non-COVID-19 Chest X-ray images classification

# **SYSTEM IMPLEMENTATION**

**4.1 Dataset**

## **4.1.1 Creation**

We start by preparing the dataset, as it is the first step to apply deep learning. As COVID-19 addresses the epithelial cells lining our airways, we use a chest X-ray image in order to analyze the health of a patient’s lung.

In this article, we adopt chest X-ray images instead of computer tomography scans to finetune the three proposed classification models. Compared with higher radiation exposure, time consuming CT scans, and expensive, X-rays are a lot cheaper, faster, lower doses for the patient and more available. In addition, portable X-ray machines can be tested in isolation wards, thereby reducing the risk of hospital infections and reducing the number of personal protective equipment used.

Furthermore, chest X-ray image analysis is a practical alternative to the PCR method. They can provide a variety of assistance from the discovery of the disease to the selection of high risk patients for isolation and prioritization, as well as selective testing to identify false negative PCR cases, they can provide a variety of help. However, because most cases of viral pneumonia are similar and overlapping, it is hard for radiologists and doctors to distinguish adequate details visually, and it is very time-consuming. Using the deep learning models can be an accurate solution.

In our experiment, we focus on reducing false positives and false negatives by using the deep learning process with 3 Convolutional Neural Network (CNN) on our collected chest X-ray images. The constructed dataset for this work contains a total of 5000 images:

– 10000 images for normal chest X-ray were selected from image databases: Kaggle repositories “Covid-19 Radiography Dataset” [19].

– 3653 chest X-ray COVID-19 images were collected from Covid-19 Radiography Data Set [19]. Table 1 below presents the content of the prepared dataset, that was divided into two folders namely Normal and COVID-19.

Figure 2 illustrates two examples of chest X-ray images taken from our prepared dataset.

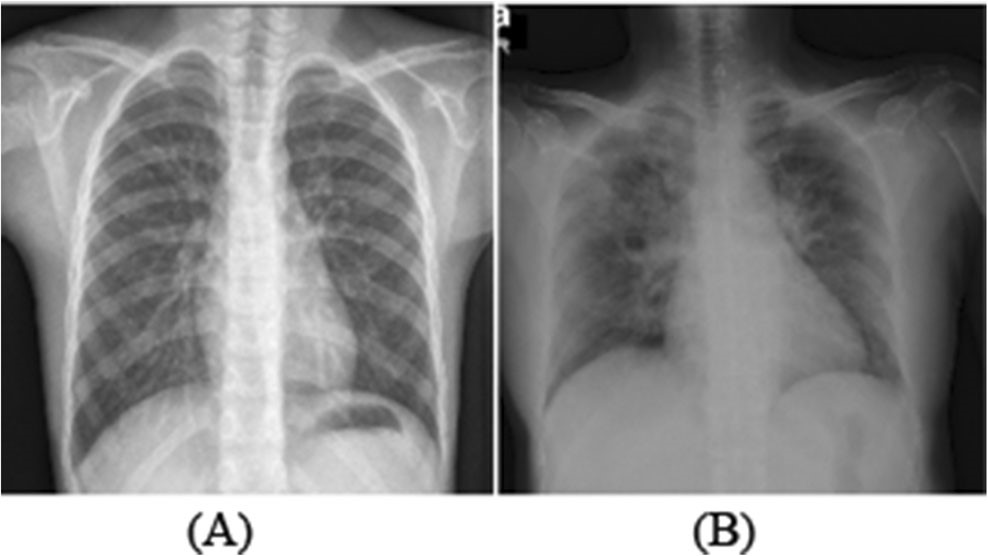


Fig. 2  (A): Chest X-ray image of a Normal person (B): COVID-19 chest X-ray image

## **4.1.2 Data preprocessing**

During data preprocessing, it is possible to resize the X-ray images. It’s due to the fact that the various algorithms require different image inputs. The images should be normalized according to the given model standards.

The input images were in different original size, then they were all processed and they were made uniform by changing the dimensions to 224 × 224 pixels.

**4.2 The proposed framework**

Due to the insufficient number of free COVID-19 radiography images, it could not be possible to develop a CNN model from scratch to automatically identify COVID-19 from X-ray images. In order to control this problem, we adopt  some deep learning models and fine-tune three well-known pre-trained models on the prepared data set.

***4.2.1 COVID-19 Detection using VGG16:***

The VGGnet architecture was proposed in 2014 by Simonyan et al. and referred to as “Very Deep Convolutional Networks for Large-scale Image Recognition” [27]. The characteristics of VGG series networks are the 3 × 3 convolutional layers that are one on top of the other, and the depth is getting larger and larger. Reducing the volume size is treated by a maximum pooling.

The VGG16 architecture is composed as following:

* Two Convolutional layers with 64 filters followed by Max pooling layer
* Two Convolutional layers with 128 filters followed by Max pooling layer
* Three Convolutional layers with 256 filters followed by Max pooling layer
* Two stack each with 3 convolutional layers with 512 filters and separated by a max pooling layer. A final Max pooling layer Two fully connected layers with 4096 channels
* Softmax output layer with 1000 classes

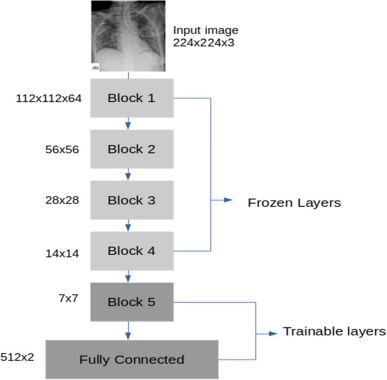


Fig. 3 Proposed VGG16 architecture

## **4.2.2 COVID-19 detection using ResNet50**

ResNet-50 is a CNN contains 50 layers; it is deeper than VGG16. Since a global average pool is used instead of a fully connected layer, the size of the model is actually much smaller, which reduces the model size of ResNet50 to 102 MB [29]. The special part of ResNet is residual block learning. This means that each layer should feed into the next layer as well as directly into the layers about 2–3 hops away. Its architecture is composed as follow:

* A convolutional layer with 64 filters and kernel size of 7 × 7. This is followed by a max pooling layer with a stride size of 2.
* Then, a convolutional layer with 64 filters and a kernel size of 1 \* 1, followed by a second convolutional layer with 64 filters and a kernel size of 3 \* 3. Then, we have another convolutional layer with 256 filters and a kernel size of 1 \* 1. These 3 layers are replicated in total 3 time and 9 layers are obtained at this stage.
* Next, 3 convolutional layers, the first one is with 128 filters and a kernel size of 1 \* 1, the second one is with 128 filters and kernel size of 3 \* 3, and the third one is with 512 filters and a kernel size of 1 \* 1. These layers are replicated 4 time to give us 12 layers at this stage.
* Afterwards, we have convolutional layer with 256 filters and a kernel size of 1 \* 1, and two others with 256, 1024 filters and a kernel size of 3 \* 3, 1 \* 1. This is replicated 6 time to give us totally 18 layers.
* Then, we have a convolutional layer with 512 filters and a kernel size of 1 \* 1, with two others with 512, 2048 and a kernel size of 1 \* 1, 3 \* 3. This is replicated 3 times to give us totally 9 layers.

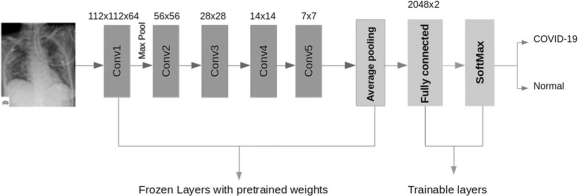


Fig. 4 Proposed Resnet50 architecture

Finally, we apply an average pooling and finish it with a fully connected layer (with 1000 nodes) and then a softmax function to give us 1 layer as a final stage.

1. **EXPERIMENTAL RESULTS**

The hidden layers are all activated by the Rectified Linear Unit activation function. The input shape is [224, 224, 3]. We fine-tune all the models for 25 epochs. We set the batch size to 32 and the learning rate to 0.0001, and we use ADAM as a loss function. We chose to rain all the models with a cross-entropy loss function.

We divide the dataset, described in Sec. 3.1, into two groups, for training (80 %) and for validation (20 %), where the first is used for the training process and the second is used for testing the final evaluation. Six performance criteria are utilized to measure the performance: Sensitivity, Accuracy, Specificity, Recall, Precision, and finally F1 score. Here are the obtained results.

***5.1 Training results***

The Training results of the proposed models are recorded and presented in the form of plots in Figs. 6, 7. The orange curve is for validation and the blue curve is according to training.

The results for the three models show that the training accuracy rate is as high as 97 %, and the training loss is reduced to 0.1, which is highlighted in each figure. This can be seen as a good sign for good classification results, especially in the field of medical diagnosis.

***5.2 Performance criteria***

There are a multiple of performance criteria that we can use to assess the performance of a classification model, by using the aforementioned criteria: sensitivity, accuracy, specificity, recall, precision and finally F1 score.

The two criteria of specificity and sensitivity could be used to evaluate a model. They are indeed widely used in the domain of health [18].

## **5.2.1 Definition of the Terms**

To assess the performance of this classifier, we should distinguish four types of elements that are classified for the desired class: TP (True Positive), TN (True Negative), FP (False Positive), and FN (False Negative).

* TP: It’s when the model correctly predicts the positive class. Here, positive class refers to a patient suffering from COVID 19.
* TN: It’s when the model correctly predicts the negative class. Here, negative class refers to a patient NOT suffering from COVID 19.
* FP (Type 1 Error): It’s when the model incorrectly predicts the positive class. Predicted that a patient suffering from COVID-19 but it’s wrong.
* FN (Type 2 Error): It’s when the model incorrectly predicts the negative class. Predicted that a patient NOT suffering from COVID-19 but it’s wrong.

Let us first define the performance criteria used to evaluate the performance of the pretrained models we used.

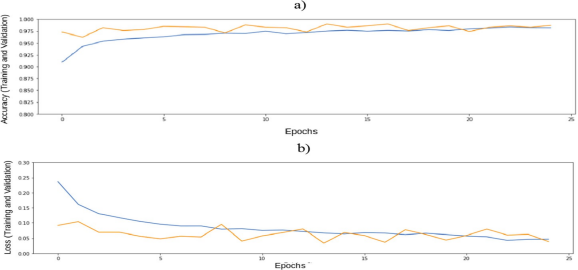


Fig. 7 Plots of (a) Training and validation accuracy and (b) Training and validation loss by using training epochs-VGG16

* Classification accuracy = TP+ TN / (TP + TN + FP + FN): The accuracy is defined as the rate of correctly classified images.
* Sensitivity = TP / (FN + TP): Measures how the model detects events in the positive category. Therefore, given that COVID-19 is a positive category, sensitivity can quantify how much X-ray images are correctly predicted as COVID-19.

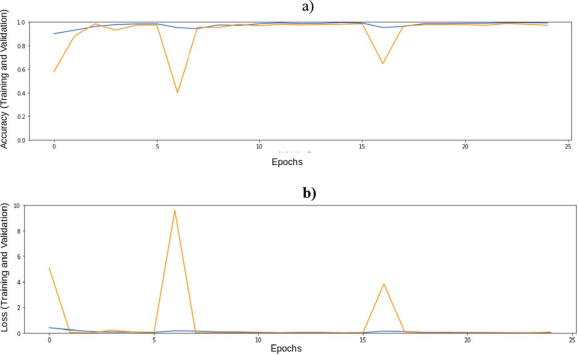


Fig. 7 Plots of (a) Training and validation accuracy and (b) Training and validation loss by using training epochs-Resnet50

* Specificity = TN/ (FP + TN): Specificity determines the proportion of actual negatives which are correctly detected.
* Precision= TP/(TP + FP): It is the proportion of the number of correctly classified positive categories to the number of predicted positive categories. In other words, precision is the response of the question: among all patients predicted as positive how many are really infected by COVID-19. The Precision should be high.
* Recall = TP / (FN + TP): The recall rate is the proportion of the number of correctly classified positive subjects to the number of positive subjects. The aim is to have it as high as possible.
* F1 score = 2 \* (precision \* recall)/ (precision + recall): To compare two models with high recall but low precision, or with low recall but high precision, is not an easy task. F1-Score is generally used, to make this comparison feasible. It enables the measuring of precision and recall at the same time. In practice, we replace the Arithmetic Mean by the Harmonic Mean. The result is that we further penalize the extreme values.

## **4.2.2 Results**

Table 2 shows the sensitivity, accuracy, specificity, recall, precision as well as F1 score of Resnet50, VGG16, and Grad-CAM. It points out that the two proposed fine-tuned versions of VGG16 and Grad-CAM outperform the proposed fine-tuned version of Resnet50 with respect to all the six performance criteria. These two fine-tuned versions have exactly the same performance with respect to three performance criteria, namely recall, precision, and F1 score, as depicted in the last three columns of the last two rows of Table 2.

The modified version of the VGG16 model shows slightly better results than the fine-tuned Grad-CAM with respect to accuracy and specificity. However, the fine-tuned Grad-CAM shows better results than VGG16 in terms of sensitivity. We can then conclude that overall, the fine-tuned Grad-CAM is the choice that we recommend.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Modified version of | Accuracy | Sensitivity | Specificity | Precision | Recall | F1 Score |
| Resnet50 | 97.20 % | 98.25 % | 97.00 % | 97.00 % | 96.00 % | 97.00 % |
| VGG16 | 98.30 % | 98.25 % | 98.33 % | 98.00 % | 98.00 % | 98.00% |
|  |  |  |  |  |  |  |

We provide the confusion matrix for the three performed models. Figures 9, 10 show the confusion matrix of the fine-tuned ResNet50, Grad-CAM, and VGG16 models on 1000 test image sets.

For VGG16 model, the confusion matrix shows only 10 out of 600 COVID-19 images are involved in the normal class (false negative), and only 7 out of 400 normal images (non-COVID-19) are classified as COVID-19 class (false positive). As for Resnet50, we find 21 false negative and 7 false positive cases. The confusion matrix of the final model, InveptionV3, shows also 16 false negative and only 3 false positive cases.

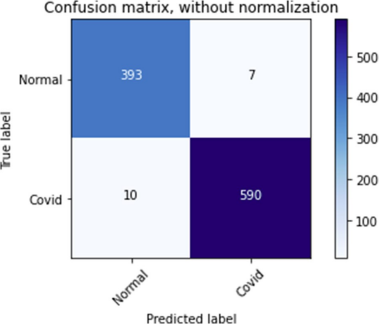


Fig. 9 The confusion matrix of the proposed VGG16 model

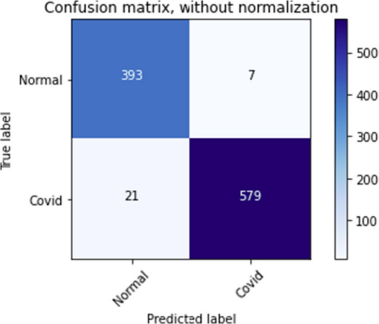
Our results show that by using deep learning models like VGG16, Resnet50, Grad-CAM, an accurate CNN model can be constructed.

Fig. 10 The confusion matrix of the proposed Resnet50 model

The encouraging results of the deep learning model for detecting COVID-19 in radiographic image detection indicate that in the near future, deep learning will play a greater clinical support role for fighting against this epidemic. Some of the limitations of this study can be overcome by performing a more analysis when we get more available data (symptomatic and asymptomatic patients).

# **CONCLUSION**

It is critical to diagnose the new coronavirus as soon as possible in order to stop the virus from spreading to other people. In tandem with this work, we develop a deep transfer learning-based approach that combines chest X-ray pictures of patients infected with COVID-19 and individuals uninfected with the virus to automatically identify the illness. With an accuracy of above 98%, the proposed classification model for COVID-19 detection aids in identifying the lung location where the virus attacks. Our research results indicate that because of its great overall performance, we think it is natural for it to assist medical professionals in making therapeutic judgments. This work provides a thorough understanding of how to employ deep transfer learning methodologies in order to find COVID-19 as soon as feasible.

COVID-19 kills millions of individuals and poses a threat to the global healthcare community. Doctors are often pressed for time due to the high volume of patients seen in emergency rooms or outdoors. Computer-aided analysis has the potential to save lives by providing early detection and tailored care.

Through effective training on a very limited image collection, our refined models demonstrate superior performance in COVID-19 pneumonia classification. We firmly believe that the suggested computer-aided diagnosis method might significantly enhance COVID-19 case diagnosis.

This is particularly useful during a pandemic when medical resources are few and preventive actions are required to keep the disease from spreading.

Deep learning research is continually trying to improve reality representations and develop models that can learn these representations from enormous amounts of unlabeled data. A portion of these depictions are grounded in the most recent advancements across many domains. For instance, Ahmad Ali et al. [2] study temporal and geographical linkages in their research using deep learning techniques. They recommend using a dynamic deep hybrid spatiotemporal neural network to accurately forecast traffic flow in every part of the metropolis.

In order to improve the outcome, we are considering combining the three models that were suggested in this work and training each layer separately as a novel technique.

# **REFERENCES**

1. Ai T et al (2020) Correlation of chest CT and RT-PCR testing in coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases. Radiology: 200642
2. Ali A, Zhu Y, Zakarya M (2021) A data aggregation-based approach to exploit dynamic spatio-temporal correlations for citywide crowd flows prediction in fog computing.Multimed Tools Appl. [https://doi.org/10. 1007/s11042-020-10486-4](https://doi.org/10.1007/s11042-020-10486-4)
3. Apostolopoulos ID, Mpesiana TA (2020) Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks. Phys Eng Sci Med 43(2):635–640
4. Axell-House DB, Lavingia R, Rafferty M, Clark E, Amirian ES, Chiao EY (2020) The estimation of diagnostic accuracy of tests for COVID-19: A scoping review. J Infect 81(5):681–697
5. Bloice MD, Roth PM, Holzinger A (2019) Biomedical image augmentation using Augmentor.

Bioinformatics 35(21):4522–4524.

1. Cohen J, Paul et al (2020) Covid-19 image data collection: Prospective predictions are the future. ArXiv preprint arXiv:2006.11988
2. covid-chest xray-dataset. <https://github.com/ieee8023/COVID-chestxray-dataset>. Accessed 25 Mar 2020
3. Eurosurveillance Editorial Team (2020) Note from the editors: World Health Organization declares novel corona virus (2019-nCoV) sixth public health emergency of international concern. Eurosurveillance 25(5): 200131e
4. Gazzah S, Bencharef O (2020) A Survey on how computer vision can response to urgent need to contribute in COVID-19 pandemics. 2020 International Conference on Intelligent Systems and Vision C (ISCV). IEEE, New York
5. Globalpulse. Need for greater cooperation between practitioners and the AI community. [https://www.](https://www.unglobalpulse.org/2020/05/need-for-greater-cooperation-between-practitioners-and-the-ai-community/)

[unglobalpulse.org/2020/05/need-for-greater-cooperation-between-practitioners-and-the-ai-community/.](https://www.unglobalpulse.org/2020/05/need-for-greater-cooperation-between-practitioners-and-the-ai-community/) Accessed 27 May 2020

1. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 770–778
2. Holzinger A et al (2018) Current advances, trends and challenges of machine learning and knowledge extraction: from machine learning to explainable AI. International Cross-Domain Conference for Machine Learning and Knowledge Extraction. Springer, Cham, 2018
3. Holzinger A et al (2019) Causability and explainability of artificial intelligence in medicine. Wiley InterdiscRev Data Min Knowl Discov 9(4):e1312
4. Isa A. Computational intelligence methods in medical image-based diagnosis of COVID-19 infections.Computational Intelligence Methods in COVID-19: Surveillance, Prevention, Prediction and Diagnosis. Springer, Singapore, pp 251–270
5. Kallianos K et al (2019) How far have we come? Artificial intelligence for chest radiograph interpretation. Clin Radiol 74(5):338–345
6. Khan A, Sohail A, Zahoora U, Qureshi AS (2020) A survey of the recent architectures of deep convolutional neural networks. Artif Intell Rev 53(8):5455–5516
7. Krizhevsky A, Sutskever I, Hinton GE (2017) Imagenet classification with deep convolutional neural networks. Commun ACM 60(6):84–90
8. Lalkhen AG, McCluskey A (2008) Clinical tests: sensitivity and specificity. Contin Educ Anaesth Crit Care Pain 8(6):221–223