Journal of Computer and Communications, 2023, 11, 29-41
https://www.scirp.org/journal/jcc
ISSN Online: 2327-5227
ISSN Print: 2327-5219

COVID-19 Detection from Chest X-Ray Images Using Convolutional Neural Network Approach

Md. Harun Or Rashid, Muzakkir Hossain Minhaz, Ananya Sarker, Must. Asma Yasmin, Md. Golam An Nihal

Department of Computer Science and Engineering, Bangladesh Army University of Engineering & Technology (BAUET), Qadirabad Cantonment, Natore, Bangladesh
Email: harun.ruetbd@gmail.com, minhaz.amj@gmail.com, ananya.ruet@gmail.com, asma.yasmin.cu@gmail.com, nihal.bauet@gmail.com

How to cite this paper: Rashid, M.H.O., Minhaz, M.H., Sarker, A., Yasmin, M.A. and Nihal, M.G.A. (2023) COVID-19 Detection from Chest X-Ray Images Using Convolutional Neural Network Approach. *Journal of Computer and Communications*, **11**, 29-41.
https://doi.org/10.4236/jcc.2023.115003

Received: March 26, 2023
Accepted: May 23, 2023
Published: May 26, 2023

Abstract

COVID-19 is a respiratory illness caused by the SARS-CoV-2 virus, first identified in 2019. The primary mode of transmission is through respiratory droplets when an infected person coughs or sneezes. Symptoms can range from mild to severe, and timely diagnosis is crucial for effective treatment. Chest X-Ray imaging is one diagnostic tool used for COVID-19, and a Convolutional Neural Network (CNN) is a popular technique for image classification. In this study, we proposed a CNN-based approach for detecting COVID-19 in chest X-Ray images. The model was trained on a dataset containing both COVID-19 positive and negative cases and evaluated on a separate test dataset to measure its accuracy. Our results indicated that the CNN approach could accurately detect COVID-19 in chest X-Ray images, with an overall accuracy of 97%. This approach could potentially serve as an early diagnostic tool to reduce the spread of the virus.

Keywords

COVID-19, Chest X-Ray Images, CNN, Virus, Accuracy

1. Introduction

COVID-19, caused by the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), rapidly turned into a global pandemic after first emerging in December 2019 in Wuhan, China. As of December 16, 2022, it has led to over 647.97 million confirmed cases and 6.62 million deaths globally [1]. The pandemic has caused catastrophic impacts on public health and the global economy. Common COVID-19 symptoms include fever, cough, sore throat, headache, fatigue, muscle pain, and shortness of breath [2]. The timely identification of COVID-19-positive cases is crucial in slowing the spread of the pandemic. The following figure (**Figure 1**) is the representation of the most common symptoms of COVID-19.

Reverse transcriptase-polymerase chain reaction (RT-PCR) testing is considered the standard for diagnosing COVID-19 patients. This test detects the presence of SARS-CoV-2 in respiratory specimens collected through nasopharyngeal or oropharyngeal swabs [4]. However, RT-PCR testing is time-consuming, requires specialized laboratory expertise, can yield false-negative results, and is costly [5]. Alternatively, chest radiography imaging, such as computed tomography (CT) or chest X-ray, can be examined by a radiologist to identify visual indicators associated with SARS-CoV-2 [6]. While CT scans provide greater image detail, chest X-Ray (CXR) images are more accessible, portable, and offer rapid triaging. CXR imaging is more accessible in most healthcare systems than CT scanners, which require expensive equipment and maintenance. The portability of the CXR system reduces the risk of COVID-19 transmission by performing exams within isolation rooms, which is impossible with fixed CT scanners. Importantly, CXR allows for the rapid triaging of suspected COVID-19 cases in severely affected countries such as the USA, Spain, and Italy, where there is a shortage of both testing capacities and supplies for RT-PCR testing [7]. Combining laboratory results with radiological image features can speed up the COVID-19 detection process.

Artificial Intelligence (AI) applications, particularly deep learning (DL), can speed up the COVID-19 diagnosis process using chest radiological imaging. DL enables AI-based models to achieve accurate results without manual feature extraction [8]. For COVID-19 detection from chest X-Ray images, Convolutional Neural Network (CNN) has gained popularity among the research community of AI in medicine. CNN is a supervised DL approach that has recently shown the best classification accuracy compared to other classification techniques such as Artificial Neural Network (ANN), Support Vector Machine (SVM), and K-Nearest Neighbor (KNN) [9].

Convolutional Neural Network (CNN) models are typically constructed using

Figure 1. The most common symptoms of COVID-19 [3].

a combination of a convolution layer, a pooling layer, and a fully connected layer. These layers work together to extract features from input, minimize computational performance, and classify images. The CNN model adjusts its internal parameters to accomplish specific tasks, such as classifying chest X-rays. To improve CNN model performance, data augmentation, and CNN hyper-parameters can be optimized [10] [11].

According to KFF [12] data, COVID-19 deaths among people aged 65 and older have decreased since the peak of Omicron in early 2022. However, deaths more than doubled between April and July 2022 (125%) and exceeded 11,000 in both July and August 2022. For people under 65, deaths also increased during this period, but at a slower rate compared to older adults (52%), and reached approximately 1900 in both July and August 2022. Between the following figures, **Figure 2** is for normal chest X-Ray and **Figure 3** is for and COVID-19 affected chest X-Ray.

In this paper, we propose a deep learning-based approach to identify and locate COVID-19 in chest X-Ray images. Our approach involves training a Convolutional Neural Network (CNN) model on a dataset of chest X-Ray images to classify them as either COVID-19 positive or negative. To improve the performance of our model, we experiment with different threshold values and hyper-parameters.

The key contributions of our work include the development of a deep learning-based model for COVID-19 detection from chest X-Ray images, which can aid in the rapid triaging of suspected COVID-19 cases. Our approach can potentially overcome the challenges associated with standard RT-PCR testing, such as time consumption, false-negative results, and cost. Our study is unique in its focus on chest X-Ray images and its exploration of threshold values to improve

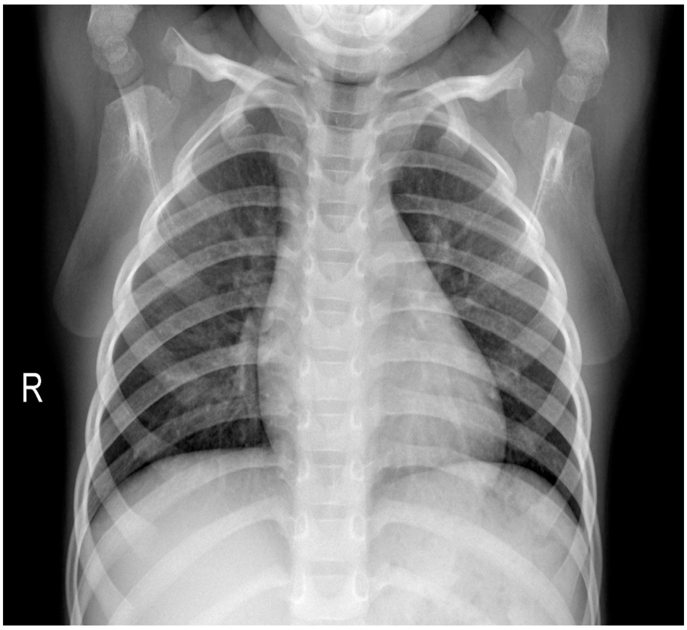

Figure 2. Normal Chest X-Ray.

Figure 3. COVID-19 Affected Chest X-Ray.

model performance. We believe that our work has the potential to significantly impact the field of AI-based medical imaging and COVID-19 diagnosis.

2. Materials and Methods

To improve the speed and accuracy of COVID-19 detection using chest X-Ray images, our study proposes a CNN-based approach. We utilized transfer learning techniques to leverage pre-trained models, reducing the amount of data needed for training. The dataset we used contained both COVID-19 positive and negative cases, ensuring that the model was trained on representative and diverse data. Our approach focused on optimizing the CNN architecture, classification, and prediction, resulting in an overall accuracy of 97% on the test dataset.

Data Preprocessing

Data preprocessing refers to the process of transforming raw data into a format that is suitable for machine learning models. It is a crucial step in building such models since real-world data is often noisy, contains missing values, and may not be directly usable for analysis. Therefore, data needs to be cleaned and formatted before using it for machine learning tasks. By performing data preprocessing, we can increase the accuracy and efficiency of the models since the data is now properly cleaned and transformed to meet the requirements of the models.

Normalization

Normalization is a crucial step in preparing data for machine learning. It involves transforming the data to a common scale or range, such as [0, 1], to ensure that differences in value ranges are not distorted. Some machine learning algorithms, particularly those that use Euclidean distance, can benefit greatly from normalization or standardization. One common approach is to transform the data into a unit sphere, which is also known as unit normalization.

Noisy Data Handling

Noisy data refers to data with a significant amount of irrelevant and meaningless information, which can include data corruption and any data that the system cannot interpret accurately. It poses a challenge in data analysis, but there are various strategies available to reduce its impact on the results. Effective handling of noisy data requires a thorough understanding of the data characteristics and the specific analysis being performed. By utilizing these strategies and experimenting with different approaches, the accuracy and reliability of the results can be improved.

3. Results and Discussion

The study presented training strategies and test results for a CNN-based model designed to classify chest X-Ray images as either abnormal (labeled as COVID) or normal. Before training, all images were resized to match the target network model. The resulting adapted model combined CNN-based feature extraction with a supervised classification algorithm, leading to an optimal solution for the classification task.

3.1. Dataset Description

The dataset used for COVID-19 detection was obtained from Kaggle, a publicly available online dataset. The dataset is structured into a main folder called "dataset" which contains two sub-folders labeled "COVID" and "normal". The dataset consists of a total of 98 X-Ray images that are in either JPEG or PNG format. The two categories of the dataset are "COVID" and "normal", with a total of 70 X-Ray images labeled as COVID-19 positive and 28 X-Ray images labeled as normal. The following figures (**Figure 4** and **Figure 5**) are taken from dataset which represent Normal Patients X-Ray Images and COVID-19 Patients X-Ray Images respectively. Amongst the following array of images, one (**Figure 6**) is selected from Normal Patients X-Ray Images and another (**Figure 7**) is selected from COVID-19 Patients X-Ray Images to show clearly.

A normal chest radiograph (left panel) shows clear lungs without any areas of abnormal specification on the image.

Dataset Distribution

This is a common understanding of the purpose of a training dataset in machine learning. The training dataset serves as the set of examples used to train a model, allowing it to learn and generalize the underlying patterns and features of the data. In each iteration or epoch, the same training data is repeatedly presented to the model for it to update its weights and adjust its parameters, improving its accuracy and ability to make predictions on new data.

Training Data

Training data is a set of data that is used to train a machine learning model. It

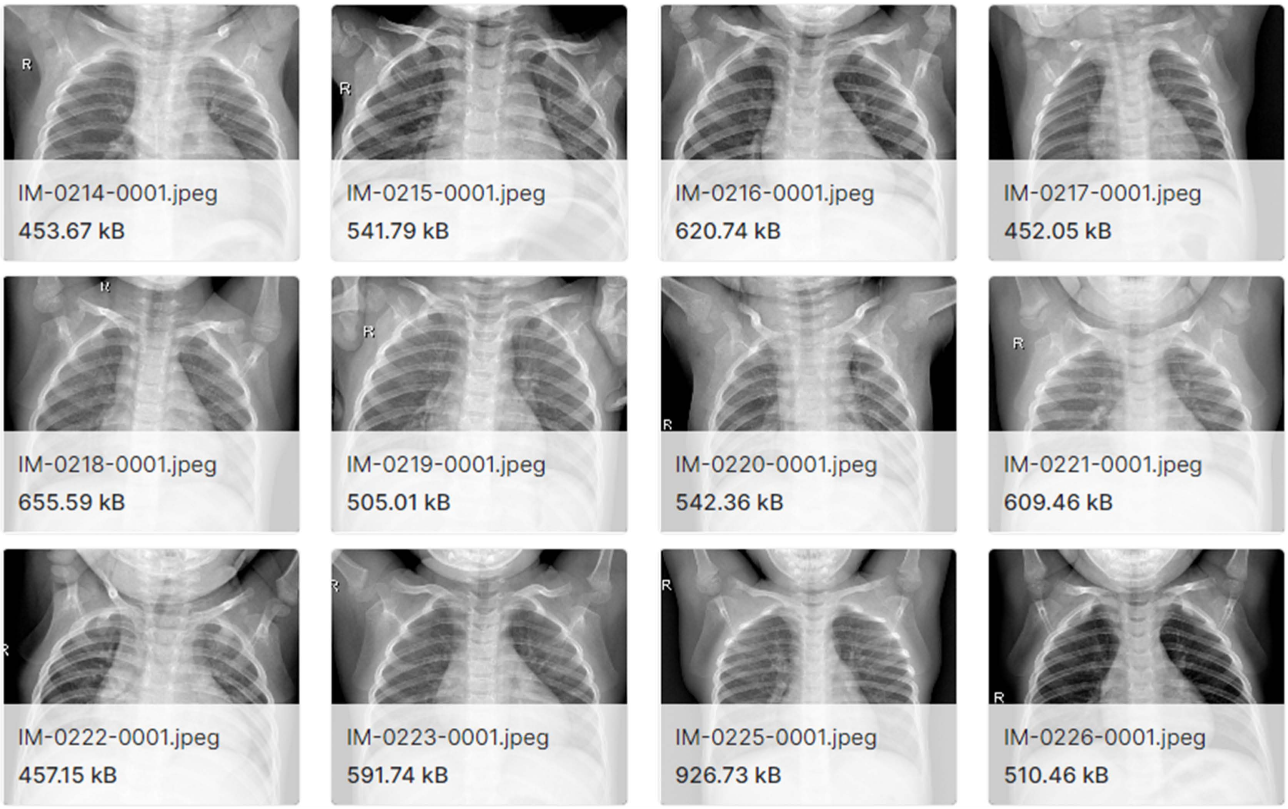

Figure 4. Normal Patients X-Ray Images.

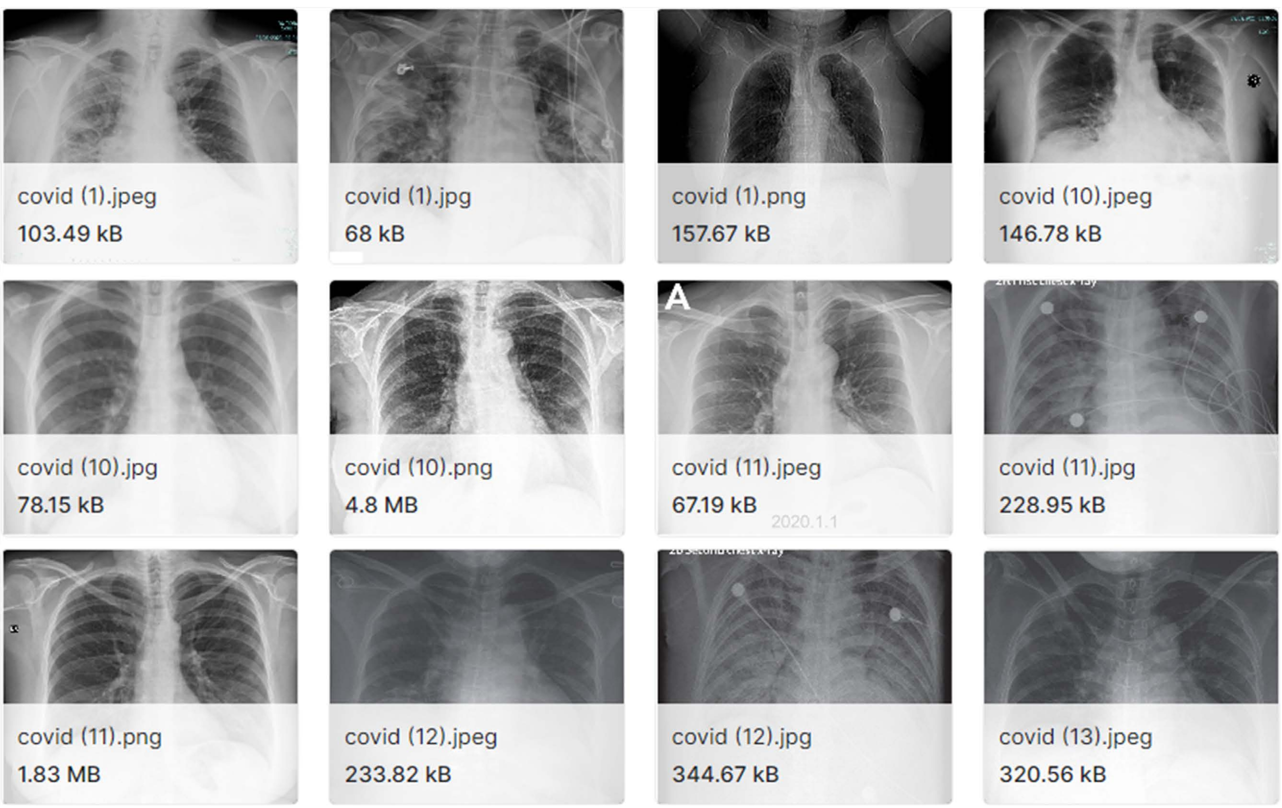

Figure 5. COVID-19 Patients X-Ray Images.

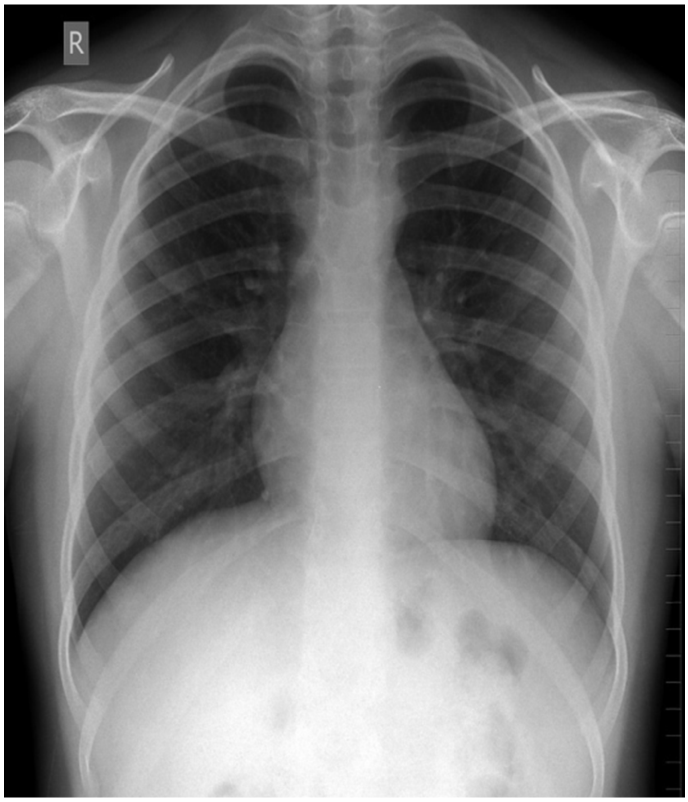

Figure 6. Normal Chest X-Ray.

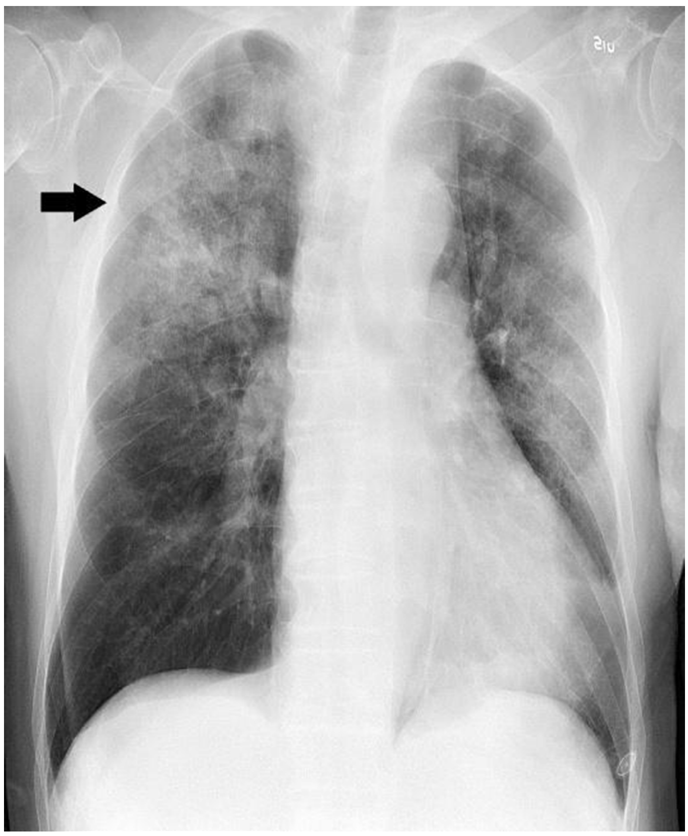

Figure 7. COVID-19 Affected Chest X-Ray.

is a subset of the overall dataset and is used to teach the model to recognize patterns and make predictions based on the input data. The training data is labeled with the correct output values or target variables, which the model tries to learn to predict from the input data. The quality and quantity of the training data are critical factors in determining the performance and accuracy of the trained model. The training data should be diverse, balanced, and representative of the problem at hand to ensure that the model can learn to generalize to new, unseen data. In this research, 70% of the available dataset was used as the training set for the proposed model.

Testing Data

The testing data is a set of data used to evaluate the performance of a machine learning model. A portion of the original dataset is typically held out for testing purposes. In this case, the dataset was split into a 70/30 ratio for training and testing, respectively. The images in the dataset were resized to 150 × 150 for consistency. The test set is used to assess the accuracy and precision of the trained model, providing an unbiased evaluation of its performance. The remaining 30% of the data was used as the test set for this proposed model.

Validation Data

The validation data is a subset of data that is used to tune the hyperparameters of a machine learning model. It is usually extracted from the training data and used to evaluate the performance of the model during the training process. The main goal of using validation data is to prevent overfitting, which occurs when a model performs well on the training data but poorly on unseen data. By using validation data, the model developer can adjust the hyperparameters to find the optimal balance between model complexity and performance, ensuring that the model generalizes well to unseen data.

System Setup

The data preprocessing, experimentation, and model evaluation are performed using the Python programming language. The proposed architecture is implemented using the TensorFlow and Keras libraries, while NumPy is used for mathematical operations on the architecture.

3.2. Evaluation

Before we discuss our result, we want to show our parameter list: (**Table 1**).

This study aims to evaluate the effectiveness of a convolutional neural network in detecting COVID-19 from chest X-Ray images. The experimental research was conducted on a Kaggle dataset, and the impact of chest radiographs was investigated.

The figures (**Figure 8** and **Figure 9**) demonstrate an increase in the accuracy of our proposed model, with a significant increase in the validation accuracy. It is worth noting that the model's validation accuracy is higher than the training accuracy.

Figure 10 and **Figure 11** show the loss values of our model during the training

Table 1. Parameter list.

No.	Parameter Name	Values
01	Total Parameter	23,041,294
02	Trainable Parameter	23,041,288
03	Non-Trainable Parameter	6
04	Kernel Size	3
05	Activation Function	Relu, Sigmoid
06	Optimizer	Adam
07	No. Of Epoch	20
08	No. Of Batch Size	4

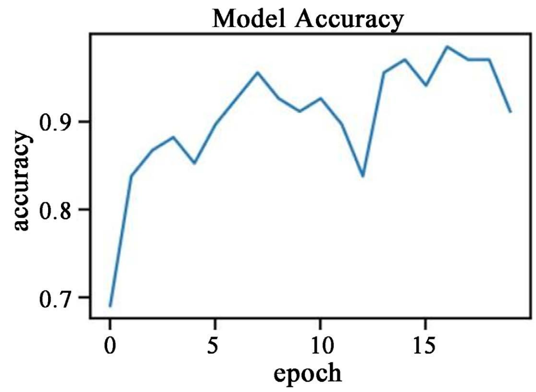

Figure 8. Analysis of model accuracy.

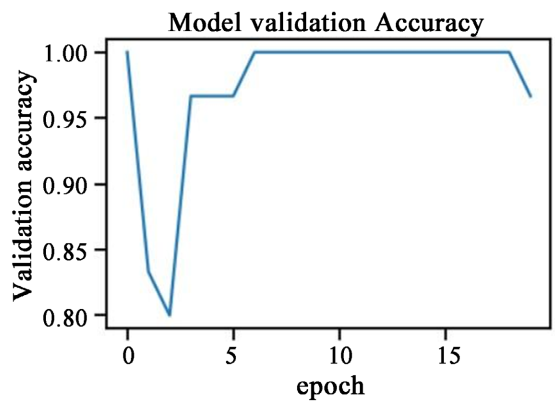

Figure 9. Analysis of model validation accuracy.

process. The model loss decreases steadily, indicating that the model is learning from the training data. Meanwhile, the model validation loss also decreases, suggesting that the model is generalizing well to unseen data. It is important to note that the model loss is usually lower than the model validation loss, which is expected since the model is optimized to perform well on the training data.

Precision

To obtain an accuracy value, we divide the total number of correctly classified

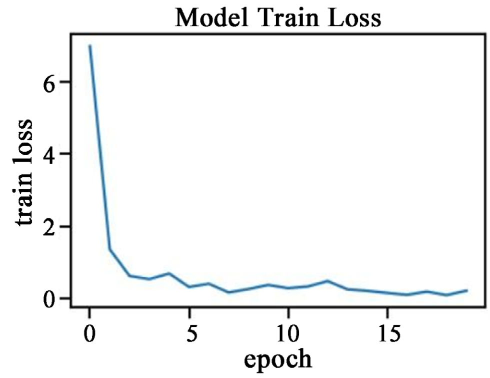

Figure 10. Analysis of model train loss.

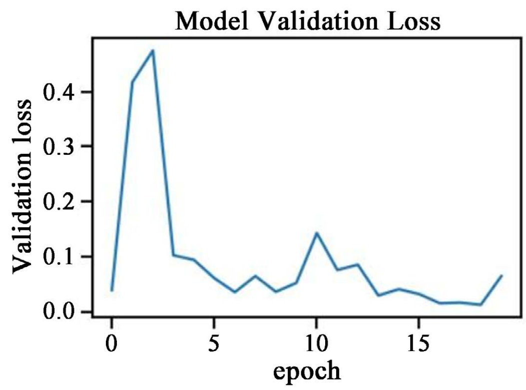

Figure 11. Analysis of model validation loss.

positive examples by the total number of predicted positive examples. High accuracy means that an example marked as positive is truly positive (small number of FPs). Precision is given by the relation:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{1}$$

Recall

Recall can be defined as the proportion of the total number of correctly classified positive examples divided by the total number of positive examples. High Recall means that the class is correctly recognized (small number of FNs). The relation gives the recall:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{2}$$

F-Measure

Since we have two measures (Precision and Recall), it is useful to have a measure that represents both. We calculate an F-measure that uses the harmonic mean instead of the arithmetic mean because it penalizes extreme values more.

$$\text{F-measure} = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \tag{3}$$

Confusion Matrix

The confusion matrix is a summary of the prediction results of a classification problem. The number of correct and incorrect predictions is summarized with count values and broken down by class. This is the key to the confusion matrix. The confusion matrix shows the ways in which your classification model is confused when it makes a prediction.

It gives us an insight not only into the mistakes the classifier makes but above all into the types of mistakes it makes.

Classification Rate/Accuracy

Proportion of correct predictions made by the model, calculated as

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{4}$$

3.3. The Result of Our Work

Consider the following Figure 12, Table 2 & Table 3 for the result of our proposed method.

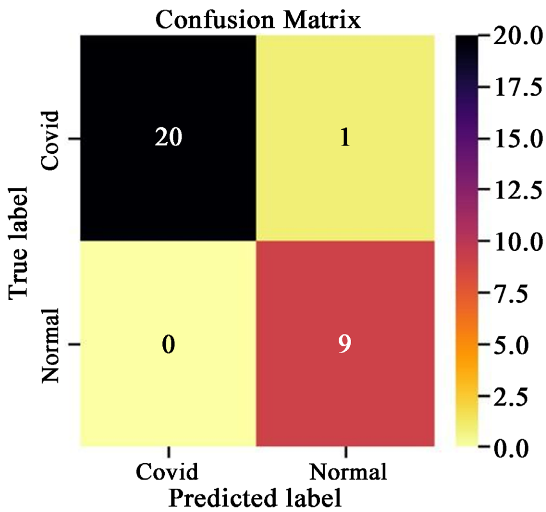

Figure 12. Confusion matrix.

Table 2. Result against performance metrics.

Performance Metrics	Result
Accuracy on the testing set	$\frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} = 97\%$
Precision on the testing set	$\frac{\text{TP}}{\text{TP} + \text{FP}} = 100\%$
Recall on the testing set	$\frac{\text{TP}}{\text{TP} + \text{FN}} = 69\%$

Table 3. Model accuracy and model loss.

Parameters	Value (%)
Model Accuracy	97
Model Loss	03

The accuracy metric is commonly used to evaluate the performance of a machine learning model. It represents the ratio of correct predictions to the total number of predictions. However, it is important to note that a high accuracy score does not necessarily mean that the model is the best, as other factors such as precision, recall, and F1 score also play an important role in evaluating model performance.

Precision is a measure of how many of the predicted positive cases are actually positive, while recall is a measure of how many actual positive cases are predicted as positive. F1-score is a harmonic mean of precision and recall.

In our study, we evaluated the performance of our proposed model for detecting COVID-19 from chest X-Ray images using a convolutional neural network. We used accuracy as one of the metrics to evaluate our model's performance.

4. Conclusion

The goal of this research paper is to explore and implement a CNN-based approach for the detection of COVID-19 from chest X-Ray images with the objective of quickly identifying whether a person is positive or negative for COVID-19. The proposed method achieved a high accuracy of 97% when evaluated on a publicly available dataset. This demonstrates the potential of CNNs for early detection of COVID-19. Future research could investigate the use of other deep learning models and datasets further to enhance the performance of automated COVID-19 detection systems.

Conflicts of Interest

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

References

[1] WHO Coronavirus (COVID-19) Dashboard. https://COVID19.who.int/

[2] Singh, A. (2021). A Review of Coronavirus Disease-2019. *International Journal of Pharmaceutical Science & Clinical Research*, **15**, 17-24. https://doi.org/10.22159/ajpcr.2022.v15i7.44547

[3] The Most Common Symptoms of COVID-19. https://deputyprimeminister.gov.mt/en/hearth-promotion/COVID-19/Pages/symptoms.aspx

[4] Wang, W., Xu, Y., Gao, R., *et al.* (2020) Detection of SARS-CoV-2 in Different Types of Clinical Specimens. *JAMA,* **323**, 1843-1844. https://doi.org/10.1001/jama.2020.3786

[5] West, C.P., Montori, V.M. and Sampathkumar, P. (2020) COVID-19 Testing: The Threat of False-Negative Results. *Mayo Clinic Proceedings*, **95**, 1127-1129. https://doi.org/10.1016/j.mayocp.2020.04.004

[6] Ng, Y., Lee, Y.P., *et al.* (2020) Imaging Profile of the COVID-19 Infection Radiologic Findings and Literature Review. *Radiology*: *Cardiothoracic Imaging*, **2**. https://doi.org/10.1148/ryct.2020200034

[7] Rubin, G.D., Ryerson, C.J., *et al*. (2020) The Role of Chest Imaging in Patient Management during the COVID-19 Pandemic: A Multinational Consensus Statement from the Fleischner Society. *Radiology*, **296**, 172-180. https://doi.org/10.1148/radiol.2020201365

[8] LeCun, Y., Bengio, Y. and Hinton, G. (2015) Deep Learning. *Nature*, **512**, 436-444. https://doi.org/10.1038/nature14539

[9] Ahsan, M., Based, M.A., Haider, J. and Kowalski, M. (2021) COVID-19 Detection from Chest X-Ray Images Using Feature Fusion and Deep Learning. *Sensors*, **21**, 1480. https://doi.org/10.3390/s21041480

[10] Wang, X., Peng, Y., Lu, L.Z. and Summers, R.M. (2018) Tienet: Text-Image Embedding Network for Common Thorax Disease Classification and Reporting in Chest X-Rays. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 9049-9058. https://doi.org/10.1109/CVPR.2018.00943

[11] Monshi, M.M.A., Poon, J. and Chung, V. (2019) The Convolutional Neural Network to Detect Thorax Diseases from Multi-View Chest X-Rays. In *International Conference on Neural Information Processing*, **1142**, 148-158. https://doi.org/10.1007/978-3-030-36808-1_17

[12] Kaiser Family Foundation (KFF). https://www.kff.org/