International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 09 Issue: 04 | Apr 2022 www.irjet.net p-ISSN: 2395-0072

# Role of Machine Learning Techniques in COVID-19 Prediction and Detection

**Chithra S. Prasad[1], Reshma Suku[2], Smitha C. Thomas[3]**

*[1]MTech Student, Computer Science and Engineering, APJ Abdul Kalam Technological University, Kerala, India*
*[2, 3]Asst. Professor, Computer Science and Engineering, Mount Zion College of Engineering, Kadammanitta, Kerala, India*

---

**Abstract** – *Currently, the detection of coronavirus disease (COVID-19) is one of the main challenges in the world, given the rapid spread of the disease. Recent statistics indicate that the number of people diagnosed with COVID-19 pandemic is increasing exponentially, with more than 1.6 million confirmed cases. The disease is spreading to many countries across the world. In this study, we analyses the incidence of COVID-19 distribution across the world. Machine learning is an innovative approach that has extensive applications in prediction. This technique needs to be applied for the COVID-19 pandemic to identify patients at high risk, their death rate, and other abnormalities. It can be used to understand the nature of this virus and further predict the upcoming issues.*

***Key Words:*** **COVID-19, Pandemic, Machine Learning, Convolutional Neural Network, Kaggle database**

## 1. INTRODUCTION

COVID-19 is a global health crisis, with more than 51 crore people infected and more than 63 lakhs deaths reported worldwide. The resulting impact on health care systems is that many countries have overstretched their resources to mitigate the spread of the pandemic. In addition, a high degree of variance in COVID-19 symptoms has been reported, with symptoms ranging from a mild flu to acute respiratory distress syndrome (ARDS) or fulminant pneumonia. There is an urgent need for effective drugs and vaccines for COVID-19 treatment and prevention. Owing to the lack of validated therapeutics, most containment measures to curtail the spread of the disease rely on social distancing, quarantine measures, and lockdown policies. The transmission of COVID-19 has been slowed as a result of these measures, but not eliminated.

Moreover, with the ease of restrictions, a fear of the second wave of infection is prevalent. To prevent the next potential outbreak of COVID-19, there is a need for advanced containment measures such as contact tracing and identification of hotspots.

Machine learning techniques have been employed in the health care domain on different scales ranging from the prediction of disease spread trajectory to the development of diagnostic and prognostic models.

A study by Ye et al identified and evaluated various health technologies, such as big data, cloud computing, mobile health, and AI, to fight the pandemic. These technologies and a wide range of data types, including data from social media, radiological images, omics, drug databases, and public health agencies, have been used for disease prediction. Several studies have focused on reviewing publications that discuss AI applications to support the COVID-19 response. One of the early studies by Vaishya et al identified 7 critical areas where AI can be applied to monitor and control the COVID-19 pandemic. However, given that this was an early work, this review lacked publications in all the 7 areas. In a later study, Lalmuanawma et al built upon these 7 areas by identifying and performing a rapid review of the then available studies; however, considering this was a rapid review, only limited studies were included, and the qualification criteria were not clear. Furthermore, a study by Shi et al focused-on AI applications to radiological images, and a study by Wynants et al focused on critical appraisal of models that aimed to predict the risk of developing the disease, hospital admission, and disease progression. Nevertheless, the majority of epidemiological studies that aimed to model disease transmission or fatality rate, among other factors, were excluded in this study.

Machine learning is an innovative approach that has extensive applications in prediction. This technique needs to be applied for the COVID-19 pandemic to identify patients at high risk, their death rate, and other abnormalities. It can be used to understand the nature of this virus and further predict the upcoming issues. This literature-based review is done by searching the relevant papers on machine learning for COVID-19 from the databases of SCOPUS, Academia, Google Scholar, PubMed, and ResearchGate. This research attempts to discuss the significance of machine learning in resolving the COVID-19 pandemic crisis. This paper studied how machine learning algorithms and methods can be employed to fight the COVID-19 virus and the pandemic. It further discusses the primary machine learning methods that are helpful during the COVID-19 pandemic.

We further identified and discussed algorithms used in machine learning and their significant applications.

Machine learning is a useful technique, and this can be witnessed in various areas to identify the existing drugs, which also seems advantageous for the treatment of COVID-

19 patients. This learning algorithm creates interferences out of unlabeled input datasets, which can be applied to analyze the unlabeled data as an input resource for COVID-19. It provides accurate and useful features rather than a traditional explicitly calculation-based method. Further, this technique is beneficial to predict the risk in healthcare during this COVID-19 crisis. Machine learning also analyses the risk factors as per age, social habits, location, and climates.

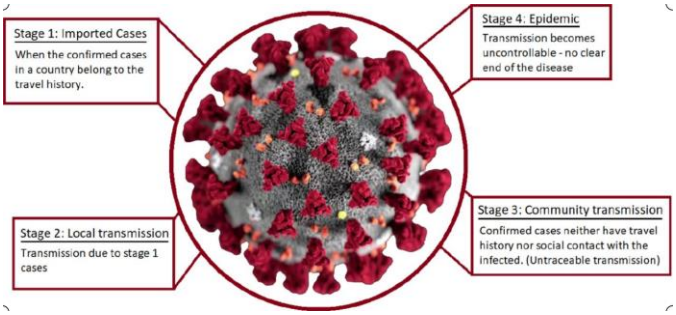


**Fig**: COVID-19 Stages

The primary aim of this study was to conduct a comprehensive systematic literature review on the role of AI as a technology to combat the COVID-19 crisis and to assess its application in the epidemiological, clinical, and molecular advancements. Specifically, we summarized the areas of AI application, data types used, types of AI methods employed and their performance, scientific findings, and challenges experienced in adopting this technology.

## 2. RELATED WORKS

Since the last decade, digital technologies are playing critical roles in major health sector problems including disease prevention, the present worldwide health emergency also seeking technological support to tackle COVID-2019.The analysis and detection of COVID-19 have been extensively investigated in the last few months. The first part of this section addresses issues related to COVID-19 detection based on deep-learning approaches using CT scans and chest X-ray images. The second part reviews the related literatures to assess future estimates of the number of COVID-19 confirmations, recoveries, and deaths. COVID-19 has now become a global pandemic owing to its rapid spread. It is very challenging to detect exposed persons because they do not show disease symptoms immediately.

Thus, it is necessary to find a method of estimating the number of potentially infected persons on a regular basis to adopt the appropriate measures. AI can be used to examine a person for COVID-19 as an alternative to traditional time-consuming and expensive methods. Although there are several studies on COVID-19, this study focused on the use of AI in forecasting COVID-19 cases and diagnosing patients for COVID-19 infection through chest X-ray images.

One of the main advantages of AI is that it can be implemented in a trained model to classify unseen images. In this study, AI was implemented to detect whether a patient is positive for COVID-19 using their chest X-ray image.

### 2.1 COVID-19 Diagnosis Using Machine Learning

The use of machine learning (ML) has been rapidly increasing in various fields including malware detection, mobile malware detection, medicine and information retrieval. A modern ML system called deep learning was introduced, which is based on a convolutional neural network (CNN). It won the ImageNet classification competition, the world's best-known computer-vision competition. Deep-learning algorithms enable computational models composed of multiple processing layers to learn data representation through several abstraction layers. They train a computer model to perform classification tasks directly from pictures, texts, or sounds. According to LeCun et al., deep-learning models feature high accuracies and can improve human output in certain instances. Artificial intelligence approaches have repeatedly given accurate and dependable outcomes in applications that use image-based data. Using deep learning techniques, researchers have been investigating and analyzing chest X-ray images to identify COVID-19 in recent years.

### 2.2 X -Ray Diagnosis Using Deep Learning

The images were normalized to extract enhanced features, which were then fed into image classification algorithms utilizing deep learning techniques. Five cutting-edge CNN systems, VGG19, MobileNetV2, Inception, Exception, and InceptionResNetV2, on a transfer-learning scenario, were tested to detect COVID-19 from control and pneumonia images. Experiments were conducted in two parts: one with 224 COVID-19 pictures, 700 bacterial pneumonia images, and 504 control images, and another with the prior normal and COVID-19 data but 714 instances of bacterial and viral pneumonia. In the two- and three-class classifications, the MobileNetV2 net had the greatest results, with 96.78% and 94.72% accuracy, respectively. Both VGG16 CNN and Resnet50, which were trained on color camera images from ImageNet, were utilized to perform transfer learning.

To assess the feasibility of utilizing chest X-rays to diagnose COVID-19, 10-fold cross-validation was performed to obtain an overall accuracy of 89.2%.

Three CNN architectures (ResNet50, InceptionV3, and InceptionRes-NetV2) were evaluated in relation to COVID-19 identification in utilizing a database of just 50 controls and 50 COVID-19 cases. ResNet50 achieved the highest accuracy of 98%.


**Fig:** Chest Images

In a successful performance in diagnosis accuracy found in this research demonstrates that deep CNNs could correctly and efficiently distinguish 21,152 normal and abnormal chest radiographs. The CNN model pre-trained on datasets of adult patients and fine-tuned on pediatric patients obtained an accuracy of 94.64%, a sensitivity of 96.5% and a specificity of 92.86% for normal versus pneumonia categorization. The proposed model in was based on an X-ray image dataset and shows that COVID-CAPS outperforms prior CNN-based models. COVID-CAPS attained an accuracy of 95.7%, a sensitivity of 90%, and a specificity of 95.8% despite having a much lower number of trainable parameters than other models. From 400 chest X-ray images, individuals with COVID-19 symptoms were identified using eight different deep learning techniques: VGG16, InceptionResNetV2, ResNet50, DenseNet201, VGG19, MobilenetV2, NasNet Mobile, and ResNet15V2. NasNet Mobile beat all other models in chest X-ray datasets, attaining an accuracy of 93.94%.

The authors utilized a database of 127 COVID-19, 500 controls, and 500 pneumonia patients collected from various sources for the binary classification of COVID-19 and controls, as well as the multiclass classification of COVID-19, controls, and pneumonia. The Darknet model was modified for transfer-learning and five-fold cross-validation, yielding 98% accuracy in binary classification and 87% accuracy in multiclass classification.

### 2.3 COVID-19 Prediction Using Machine Learning Techniques

Machine Learning is the science of training machines using mathematical models to learn and analyze data. Once ML is implemented in a system, the data are analyzed, and interesting patterns are detected.

The validation data are then categorized according to the patterns learned during the learning process. As COVID-19 infection has rapidly spread worldwide and international action is required, it is important to develop a strategy to estimate the number of potentially infected people on a regular basis to adopt the appropriate measures. Currently, decision-makers rely on certain decision-making statistics such as imposing lockdowns on infected cities or countries. Therefore, ML can be used to predict the behaviors of new cases to stop the disease from spreading. Li et al. developed a prediction model using ML algorithms to combat COVID-19 in mainland China and in other infected countries in the world. The authors developed a model to estimate the number of reported cases and deaths in mainland China and in the world. The data used to build the models were collected between 20 January 2020 and 1 March 2021. The authors also stated that COVID-19 would be controlled at the beginning of April 2020 in mainland China and in mid-June 2020 across the world. They concluded that the estimated number of COVID-19 cases would be approximately 89,000 in China and 403,000 worldwide during the outbreak. As of 17 April 2021, the estimated number of deaths was 4000 in mainland China and 18,300 worldwide. It is clear that their forecast was similar to the actual situation in China as the total numbers of infected cases and deaths had exceeded 82,367 and 3342, respectively.

## 3. SYSTEM DESIGN

Our proposed deep learning-based COVID-19 detection comprises several phases, as illustrated in the below Figure.

The phases are summarized in the following five steps:
Step 1: Collect the chest X-ray images for the dataset from COVID-19 patients and healthy persons.
Step 2: Generate 1000 chest X-ray images using data augmentation.
Step 3: Represent the images in a feature space and apply deep learning.
Step 4: Split the dataset into two sets: a training set and a validation set.
Step 5: Evaluate the performance of the detector on the validation dataset

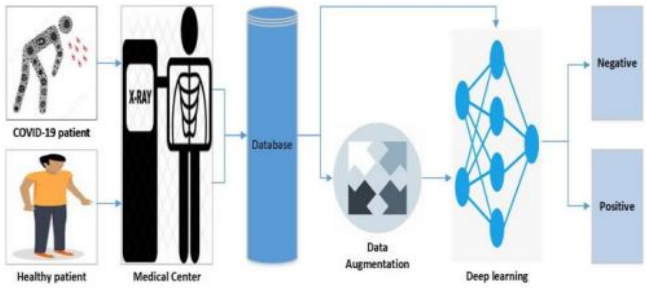


**Fig:** Proposed system Architecture

### 3.1 Dataset Preparation

Two types of datasets were used in the evaluation, the original dataset (without augmentation) and the augmented dataset, which are summarized in Tables. The dataset contained the following: a) a healthy dataset containing chest X-ray images of healthy persons and b) a COVID-19

dataset containing chest X-ray images of COVID-19 patients. The original dataset was obtained from the Kaggle database. we generated our dataset using data augmentation.

*Table 1 Original dataset (without augmentation)*

| *X-ray images* | *Number* |
|---|---|
| Healthy | 28 |
| COVID-19 | 70 |
| **Total** | **128** |

*Table 2 Augmented dataset*

| *X-ray images* | *Number* |
|---|---|
| Healthy | 500 |
| COVID-19 | 500 |
| Total | **1000** |

**Fig:** Augmented data set

Data augmentation is an AI method for increasing the size and the diversity of labelled training sets by generating different iterations of the samples in a dataset. Data augmentation methods are commonly used in ML to address class imbalance problems, reduce overfitting in deep learning, and improve convergence, which ultimately contributes to better results. The total number of images in the dataset became 1000 after applying augmentation, as presented in Table.

### 3.2 Environment

A computer with Microsoft Windows 10 was used for the experiment. It has the following specifications: Intel Core i7-8565U 1.80-GHz processor, 16 GB of DDR4 RAM, and 1 TB of hard disk. We installed the virtual machine tool VMware Workstation Pro version 14.1.8 build-14921873 on it.
Then, we installed Ubuntu 18.04.4 (64 bit) on the virtual machine and the following libraries and software:

- Python
- NumPy
- ARMIA
- Image Data Generator
- Fbprophet
- Matplotlib
- KERAS
- LSTM
- Pandas

All the results and predictions made in this study have been uploaded to the Kaggle database. We believe that by making the system and solution publicly available, we draw attention to the most affected areas, thereby preventing the spread of the COVID-19 outbreak and fostering the use of deep-learning techniques in COVID-19 research.

## 4. RESULT

During the times of pandemics, faster diagnosis plays a key role in the response efforts to contain the disease as well as reducing its spread. Computer-aided detection would save time and increase the quality of diagnosis in comparison with manual human diagnosis.

This study developed a CNN-based COVID-19 detection model that was tested with both the original and the augmented datasets. All the chest X-ray images used were resized to 224×224 pixels while ignoring the aspect ratio. Figures 14a and b present the chest X-ray images of healthy and COVID-19-infected patients, respectively. The collected dataset was randomly split into a training data subset and a testing data subset. The COVID-19 detector was trained and tested on the collected dataset, 80% of which was used for training and the remaining 20% was used for testing. The weights of the CNN were randomly initialized, and the batch size was varied up to 25 and empirically set to 25 to avoid overfitting and to achieve the highest training accuracy.
Based on the study results, the following results were drawn:

- PA delivered the best performance for COVID-19 prediction over 7 days, compared to LSTM and ARIMA.
- The predictions will enable people in both countries to predict their medical needs for tackling the spread of COVID-19.
- ARIMA cannot make predictions over the next 1, 2, and 3 days.
- After investigating the number of COVID-19 confirmations, recoveries, and deaths in various countries, we found that coastal areas are significantly impacted by the disease.
- The use of chest X-ray images is recommended for diagnosing COVID-19 because X-rays are easily obtained at nearby hospitals or clinics fairly quickly and at low costs.
- Our CNN-based COVID-19 detector delivered superior performance in terms of precision, recall, and F-measure.
- The application of ML techniques for COVID-19 diagnosis using our CNN-based COVID-19 detector is recommended.

Our COVID-19 detector obtained better results when using augmentation. A better training process was achieved as the gap between the training and validation became smaller.

## 5. CONCLUSIONS

The world is under the grasp of COVID-19 virus. Early prediction of the transmission can help to take necessary actions. This article proposed to utilize the machine learning and deep learning models for epidemic.

Future prediction of potential infections will enable authorities to tackle the consequences effectively. Furthermore, it is necessary to keep up with the number of infected people by performing regular check-ups, and it is often vital to quarantine infected people and adopt medical measures

Prediction models such as the PA, ARIMA, and LSTM algorithms were used to predict the number of COVID-19 confirmations, recoveries, and deaths over the next 7 days. PA delivered the best performance and a diagnosis model using VGG16 was proposed to detect COVID-19 using chest X-ray images. The model allows the rapid and reliable detection of COVID-19, enabling it to achieve an F-measure of 99% using an augmented dataset. In a future study, we will consider diagnosing COVID-19 in chest CT scan images using the VGG-XX versions and compare their performances using larger datasets.

A further contribution of this study is the analysis of the COVID-19 spread and its related statistical data based on its global regional distributions. Thus, two main conclusions were drawn using our AI-based analysis: (i) the most highly infected areas have similar characteristics, and (ii) the spread of the disease in coastal areas is significantly higher than that in other non-coastal areas. Therefore, extra care and attention should be given to coastal cities. In our future work, we will investigate the effects of temperature, humidity, and terrain on the COVID-19 spread in cities and countries.

## REFERENCES

[1] Wang L, Wong A (2020) COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest radiography images.

[2] Beck BR, Shin B, Choi Y, Park S, Kang K. Predicting commercially available antiviral drugs that may act on the novel coronavirus (SARS-CoV-2) through a drug-target interaction deep learning model

[3] Cohen et al (2020) COVID-19 image data collection

[4] Ting, Daniel Shu Wei, Lawrence Carin, Victor Dzau, and Tien Y. Wong. “Digital technology and COVID-19.”

[5] Abed MH, Al-Rammahi AHI, Radif MJ (2019) Real-time color image classification based on deep learning network

[6] Deb, Soudeep, and Manidipa Majumdar. “A time series method to analyze incidence pattern and estimate reproduction number of COVID-19.” arXiv preprint arXiv:2003.10655 (2020).

[7] Jia, L., Li, K., Jiang, Y., Guo, X. (2020). Prediction and analysis of Coronavirus Disease 2019. arXiv preprint arXiv:2003.05447.

[8] Sujath R, Chatterjee JM, Hassanien AE. A machine learning forecasting model for COVID-19 pandemic in India. Stoch Environ Res Risk Assess. 2020 May 30;34(7):959–972. doi: 10.1007/s00477-020-01827-8

[9] Cohen JP, Bertin P, Chester VF, A web delivered locally computed chest x-ray disease prediction system.

[10] Ivanov, D. Predicting the impacts of epidemic outbreaks on global supply chains: A simulation-based analysis on the coronavirus outbreak (COVID-19/SARS-CoV-2) case. Transp. Res. Part E Logist. Transp.Rev. 2020, 136, doi:10.1016/j.tre.2020.101922.

[11] Fu L, Li Y, Cheng A, Pang P, Shu Z. a novel machine learning-derived radiomic signature of the whole lung differentiates stable from progressive COVID-19 infection: a retrospective cohort study. J Thorac Imaging. 2020 Jun 16; doi: 10.1097/RTI.0000000000000544

[12] Hick JL, Hanfling D, Wynia MK, Pavia AT. Duty to plan: health care, crisis standards of care, and novel coronavirus SARS-CoV-2. *NAM Perspectives.* 2020 Mar 5; doi: 10.31478/202003b.

[13] Liu C, Wang X, Liu C, Sun Q, Peng W. Differentiating novel coronavirus pneumonia from general pneumonia based on machine learning. *Biomed Eng Online.* 2020 Aug 19;19(1):66. doi: 10.1186/s12938-020-00809-9.

[14] Clark K, Vendt B, Smith K, Freymann J, Kirby J, Koppel P, Moore S, Phillips S, Maffitt D, Pringle M, Tarbox L, Prior F. The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. *J Digit Imaging.* 2013 Dec;26(6):1045–57. doi: 10.1007/s10278-013-9622-7

[15] Sedik A, Iliyasu AM, Abd El-Rahiem B, Abdel Samea ME, Abdel-Raheem A, Hammad M, et al. Deploying Machine and Deep Learning Models for Efficient Data-Augmented Detection of COVID-19 Infections. 2020;12(7):769. PMID: doi:10.3390/v12070769

[16] Cyranoski D. 'We need to be alert': Scientists fear second coronavirus wave as China's lockdowns ease. Nature. 2020 Mar 30. PMID: 32231319. doi: 10.1038/d41586-020-00938-0.

[17] Sear RF, Velasquez N, Leahy R, Restrepo NJ, El Oud S, Gabriel N, et al. Quantifying COVID-19 Content in the Online Health Opinion War Using Machine Learning. Ieee Access. 2020;8:91886-93.

**BIOGRAPHIES**

Chithra S Prasad, currently pursuing MTech degree in Computer Science and Engineering from APJ Abdul Kalam Technological University, Kerala, India at Mount ZionCollege of Engineering, Kadammanitta, Kerala, India.

Smitha C Thomas, received the MTech degree in Computer science and Engineering. She is currently working as Assistant Professor in the Department of Computer science and Engineering at Mount Zion College of Engineering, Kadammanitta, Kerala, India.

Reshma Suku, received the MTech degree in Computer science and Engineering. She is currently working as Assistant Professor in the Department of Computer science and Engineering at Mount Zion College of Engineering, Kadammanitta, Kerala, India.