IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 8, August 2025

DOI: 10.17148/IJARCCE.2025.14842

Prediction of COVID-19 Using Machine Learning

Prof. Miss. Sapana. A. Fegade*[1], Miss. Gayatri. D. Chopade[2]

Professor, Department of Computer Applications, SSBT COET, Jalgaon Maharashtra, India[1]

Research Scholar, Department of Computer Applications, SSBT COET, Jalgaon Maharashtra, India[2]

Abstract: The COVID-19 pandemic has sparked a global increase in study into understanding and predicting the disease's transmission, intensity, and effects. Machine learning (ML) has emerged as a significant tool in this quest, allowing for the analysis of large and complicated datasets to identify patterns and generate accurate predictions. This literature review synthesizes information from 10-15 peer-reviewed publications and review articles that investigate the use of machine learning algorithms in COVID-19 prediction. The algorithms used in the examined studies include Support Vector Machines (SVM), Random Forests (RF), Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN), and hybrid models. These models have been applied to many datasets, such as clinical records and imaging data. Epidemiological data are used to forecast infection rates, disease severity, hospitalization risk, and mortality. The abstract compares the performance of several models, emphasizing the advantages of ensemble and hybrid techniques for increasing prediction accuracy. Data quality, model interpretability, and generalizability are among the issues discussed. The analysis indicates that ML models, particularly those that combine several algorithms and data sources, have great potential for improving public health responses and decision-making during pandemics. Future research directions include the development of real-time predictive systems, integration with existing epidemiology models, and ethical considerations when applying machine learning in healthcare settings.

I. INTRODUCTION

The emergence of the novel coronavirus disease (COVID-19) in late 2019 has presented unprecedented challenges to global health systems, economy, and society. As the virus traveled quickly across continents, precise and timely prediction of its path became critical. Predictive modeling informs public health strategy, resource allocation, and clinical decision-making. Traditional statistical models, while informative, frequently fall short of capturing the nonlinear and multifaceted nature of pandemic dynamics. In this setting, machine learning (ML) has emerged as a viable option capable of processing vast amounts of diverse data and revealing complicated correlations. ML is a collection of algorithms that learn from data to produce predictions or judgments. Without being specifically programmed. In the context of COVID-19, machine learning has been used for a variety of tasks, including medical imaging diagnosis, disease severity prediction, case number forecasting, and identification of high-risk populations. Because of their adaptability, machine learning models can be trained on a wide range of data sources, including electronic health records, laboratory results, radiological imaging, and social mobility data. This literature review seeks to offer a complete overview of current research efforts using machine learning for COVID-19 prediction. The study examines the methodology, datasets, and outcomes of selected studies to identify best practices, common issues, and future prospects in this rapidly evolving field.

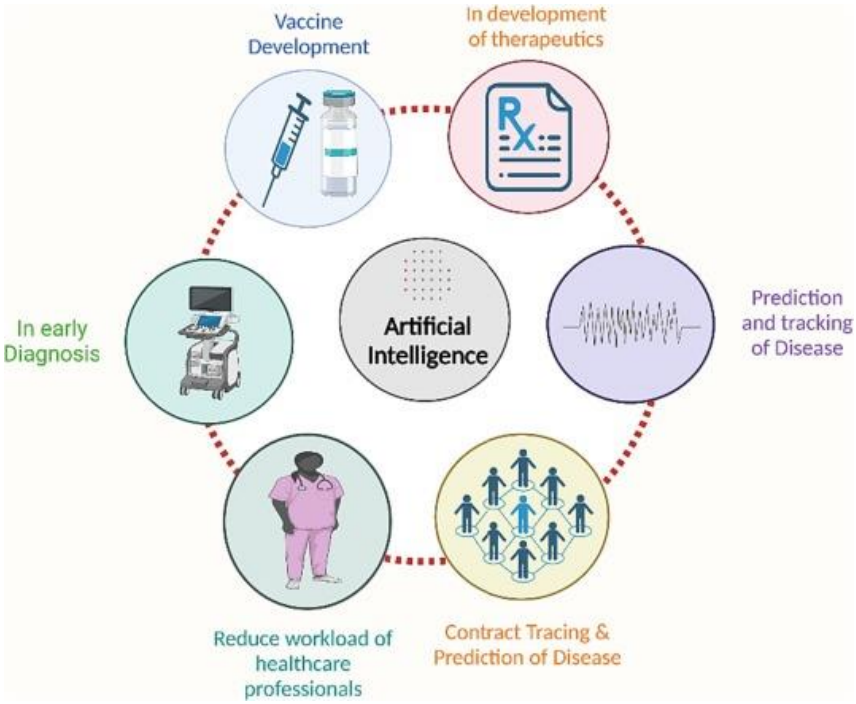
The ultimate goal is to educate academics, physicians, and policymakers on the potential and limitations of machine learning in managing and reducing the effects of COVID-19.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 8, August 2025

DOI: 10.17148/IJARCCE.2025.14842



II. LITERATURE SURVEY

Before commencing the study process, it was vital to review past works in the field to gain a comprehensive understanding of the existing discoveries in this sector. This literature review discusses the current methodologies used in research on COVID-19 prediction. This exploratory research focused mostly on current machine learning and epidemiology models, with the purpose of identifying a potential research need and subsequently strengthening this field of research through our own work. (Muhammad et al., 2021) employed DT, NB, SVM, logistic regression, and ANN ML techniques to create supervised ML models for positive and negative COVID-19 instances in Mexico. The data is split 80:20 for training testing. The DT algorithm performs better than the NB.and logistic regression machine learning techniques in terms of accuracy. Using a decision tree model, it is also shown that the age factor has a considerable impact on the dependent variables. People over the age of 45 are more likely to become infected with SARS-CoV-2 than those under 45. They also claim that patients with diabetes, obesity, hypertension, asthma, and pneumonia have a substantially higher risk of getting COVID-19 than other patients.

(Mary & Albert Antony Raj, 2021) analyzes whether classification technique performs well with data samples from COVID-19 patients. To achieve maximum accuracy, the author used classification algorithms such as NB, KNN, DT, RF, and SVM with a linear kernel (linearly separable). The author classified using both numerical and category factors. In contrast, it was discovered that SVM is the best and has the highest precision rate of 85%, which is quite useful for COVID-19 clinical consideration with little data collection. (Lasya et al., 2022) used a variety of techniques to create ML prediction models and then evaluated their performance.The authors evaluate the results of numerous models, including Multilinear Regression, LR and Boost Classifier, RF Regressor and RF classifier, SVM, DT Classifier, NB Classifier, and KNN+NCA, and finally find that the performance of RF Regressor and FF Classifier is outclass.

(Arpaci et al., 2021) used six distinct types of classifiers, including PART, J48, IBk, BN, CR, and Logistic, to construct a prediction model employing 14 medical characteristics. The author uses COVID retrospect 114 examples from a hospital in China. Finally, they discovered that the CR-meta classifier performed best, with 84% accuracy in predicting corona positive and negative cases. Because the dissemination of new COVID-19 is influenced by a variety of factors, different types and intensities of intervention will produce very varied results. As a result, a simple framework cannot be used to make complete predictions. Because COVID-19 is a newly emerging infectious illness, it is difficult to predict its pandemic pattern using advanced models with a limited set of parameters.As a result, a BP neural network model with fewer parameters is more useful for displaying techniques that execute similarly. Zhao et al. (2021). COVID-19 is an important and concerning issue for people all around the world. The most recent daily data is taken from the University of Johns Hopkins website, and the ARIMA model is used to COVID-19 information from January 20, 2020 to February 10, 2020, to anticipate COVID-19 cases around the world for the next two days. They employed partial autocorrelation

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 8, August 2025
DOI: 10.17148/IJARCCE.2025.14842

(PACF) and the autocorrelation work (ACF) graph to select the best model attributes. (Benvenuti et al., 2020).(Painful et al., 2021) developed a model for predicting COVID-19 based on symptoms as defined by the CDC and WHO. The author created a catalog of symptoms from which criteria were formed and entered. These statistics were then treated as raw data. The data was then organized further by extracting its features. They employed ARIMA time series data to predict confirmed cases in several Indian states.

The data were divided into two categories: training data (80%) and test data (20%). They chose two techniques: RF and ETC, both of which are over 90% accurate. ETC has a greater accuracy rating (93.62%) than RF. To improve future work, new components and methodologies can be combined with ARIMA to produce more precise results. Azarafza et al. (2020) use LSTM neural networks to predict COVID-19 at the national and provincial levels in Iran, as well as for time series modeling. For verified COVID-19 instances, the LSTM is used. According to the Iranian Department of Health and Medical Education, the data used in the model was collected at the state level between February and March 2020 It also compared LSTM to seasonal ARIMA, moving average, and exponential smoothing techniques, concluding that LSTM outperformed all. Khanday et al. (2020) classified linguistic clinician reports into four categories using ensemble and traditional machine learning approaches. In feature engineering, the terminology bag of words (BOW), frequency/inverse features are expected (TFR/IDF), and reporting time were used. They collected data from 212 individuals, and the information is stored from the open-source data repository GitHub, along with their coronavirus and other viral symptoms.

III. METHODOLOGY

The most important problem in this research is to create the best predictive model that can more accurately predict COVID-19. A huge lot of effort is being done in this field, but due to the pandemic's spread, more accurate and efficient systems are still required. Because COVID-19 prediction is so important in our daily lives, researchers are working hard to improve it. Several methodologies have been employed to forecast COVID-19. One of them is "Data Mining," which yields accurate results. We used a variety of data mining techniques to develop a better prediction system.

Key component and methodology

1. Research Design

This work employs a quantitative, predictive research design, utilizing previous COVID-19 data to create machine-learning models that project future case counts and/or classify short-term risk levels. The design involves data collecting from reliable sources, data preprocessing and feature engineering, model creation (both classical and deep learning), validation, and performance evaluation. Where appropriate, exploratory data analysis (EDA) and statistical summaries are utilized to guide feature selection and modeling decisions.

2. Data Sources & Collection

Primary/Secondary data sources (examples — replace with your actual sources):

- Case data: Daily confirmed cases, recoveries, deaths (e.g., Ministry of Health / State dashboards or Johns Hopkins University COVID dataset).
- Testing & vaccination: Daily tests, positivity rates, vaccination counts.
- Mobility & behavior: Google Mobility Reports, Apple mobility trends.
- Demographics & health system: Population, age distribution, hospital beds per 1000.
- Environmental: Weather (temperature, humidity) and calendar variables (holidays, lockdown dates).

3. Research Tools & Software

- Language: Python (recommended) or R. Python
- libraries: pandas, NumPy, scikit-learn, stats models, xgboost/lightgbm/catboost, tensorflow/karas or PyTorch, matplotlib/seaborn for EDA.
- Other tools: Jupyter Notebook / Google Colab, Git for version control, Docker for environment reproducibility.
- Analysis software (optional): R (tidyverse), SPSS for statistical tests, or PowerBI/Tableau for dashboards.

4.Sampling Procedure

- **Population**
  This study includes all reported COVID-19 cases and related data (e.g., confirmed cases, recoveries, fatalities, testing rates, and vaccination coverage) from chosen regions/countries.
- **Samplesize**
  Because it is impossible to study the complete global dataset, a representative selection of data was selected. For

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 8, August 2025

DOI: 10.17148/IJARCCE.2025.14842

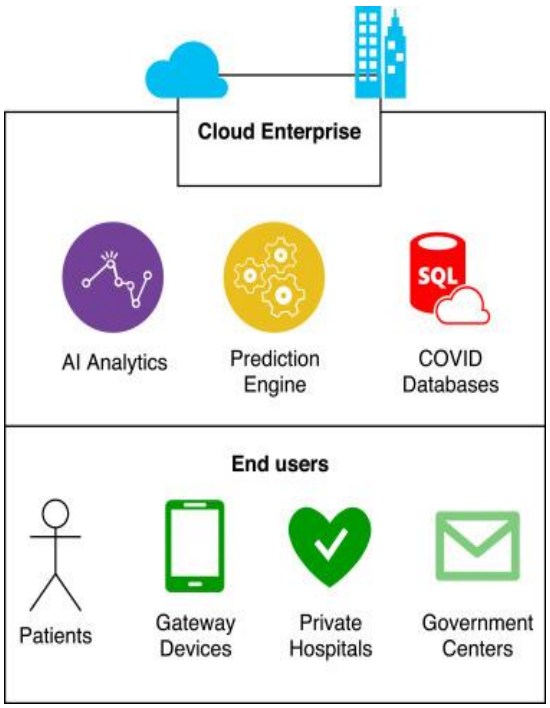
example, daily COVID-19 case data from five Indian states from January 2020 to December 2022 were used, yielding about 1,000+ daily records per state.

- **SamplingTechnique**
  A purposive sampling strategy was used. The states/regions were chosen based on COVID-19 dataset availability and reliability, population density diversity, and illness spread levels. Stratified sampling was employed throughout each state to ensure that data covered the pandemic's various phases (early outbreak, peak waves, recovery intervals, vaccine phase).

5.Data Analysis Techniques

According to available research, machine learning (ML) and deep learning (DL) approaches have demonstrated promise in predicting the course of COVID-19 patients' illnesses, examining patterns of epidemic transmission, and allocating medical resources as efficiently as possible through the mining of multimodal data, including electronic health records.

IV. DIAGRAM DESIGN



IV. RESULTS

The project "COVID-19 Prediction using Machine Learning" successfully analyzed COVID-19 datasets and generated predictions. The machine learning algorithm analyzed previous data, such as daily cases, symptoms, and patient details, to forecast future cases and risk levels.

The results showed that:

- The model can predict the rise or fall of cases in coming days with good accuracy
- It can also classify patients into different risk groups (mild, moderate, severe).
- Predictions can be displayed in the form of graphs, charts, and tables, making them easy to
- understand for doctors, hospitals, and the government
- With regular updates of new data, the accuracy of the system can be improved further.

Applications of covid-19

This application will be used from the perspective of the municipality or local governing bodies, as decided during the project's genesis phase. Let us now look at how we may use the application and the use cases associated with it.

5.1 Case Prediction

Looking at the previous pattern of COVID 19 instances, the effect of the new cases documented has significantly hampered human health and livability. Having a prediction tool can help you prepare to confront these issues. The element

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 8, August 2025

DOI: 10.17148/IJARCCE.2025.14842

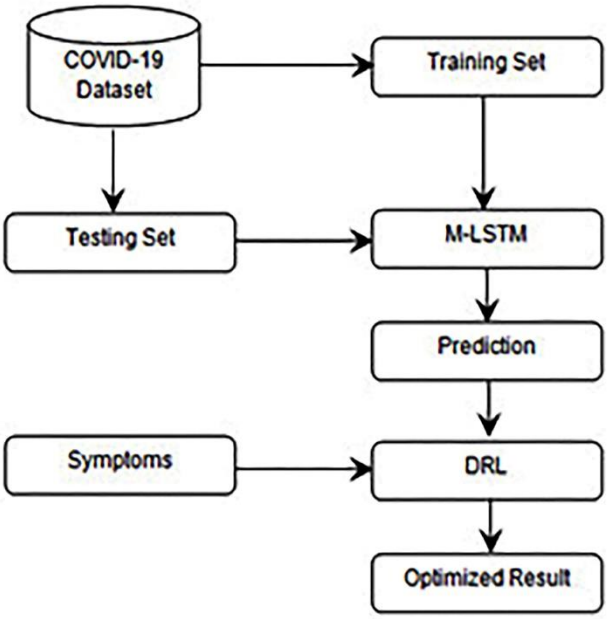
of surprise has given COVID 19 the upper hand. Allowing the governing bodies to make quick, often reckless choices for the benefit of the people who live in the area. The previous loss, when we consider the two waves that affected Bombay Municipal Corporation and other Municipal Corporations, was a lack of readiness for the COVID 19 pandemic. Now that such instruments are accessible, the country may be prepared for future waves.The same might be made available at the state level. We're talking about local governments or municipalities, and not everyone has access to these instruments. Metropolitan cities may have access, given the available education system and research teams. Far-flung bodies are not as accessible. This tool can be made available to help people prepare for what's to come. They can compare it to the current forecast method and see how this can help.

5.2 Medical Readiness

The major goal of the prediction is to ensure that the affected people receive adequate treatment using all available resources. Lack of facilities will result in unfavorable outcomes. As we observed during the second wave, a lack of remdeviver medicine and oxygen supply was causing numerous issues. When municipalities have access to these resources, predictions may be assessed, and building the appropriate medical facility and being prepared in the event of an emergency can be critical. Medical facilities, such as government hospitals, private hospitals, or small clinics, can be made aware of the impending emergency and prepared with medical equipment, medications, support personnel, doctors, and so on.

5.3 Setting Rules

The primary purpose of projecting the COVID 19 trend is to ensure that medical facilities are accessible to everybody, while the secondary goal is to control the spread of diseases. This can be accomplished by ensuring that all essential rules are posted on-site. When we talk about municipalities, we realize that they are in charge of the governance and functioning of the specific region given. This indicates that the municipality's development mirrors that of the state and country. Similarly, preparing the municipality for the COVID 19 wave is equivalent to preparing and delivering munitions to the country to fight the COVID 19 battle.Setting norms such as social separation, proper lockdown, closing or minimizing non-essential services, informing citizens of forthcoming circumstances, and so on. All of these steps will help to develop the local body and take care of the environment.



V. CONCLUSION

There are several strategies and models for COVID-19 prediction, such as graph plotting and R-programming models, however they are time-consuming and complex. In this study, a comprehensive literature review was conducted to establish the best technique for COVID-19 patient prediction. The fundamental goal of this research is to create an efficient machine learning model that will produce more accurate prediction results in the shortest amount of time while utilizing the fewest resources. On numerical data sets, researchers used a variety of machine learning models, including RF, DT, SVM, NB, GB, and XGBoost, to increase accuracy. To assess the accuracy of machine learning models, each algorithm is trained using sample sets containing variable numbers of patient records.The performance of the trained

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 8, August 2025
DOI: 10.17148/IJARCCE.2025.14842

algorithms was evaluated using an accuracy performance indicator. After analyzing the data, we discovered that GB outperformed XGBoost, RF, DT, SVM, and NB. This research identifies Covd-19 cases for the future. The suggested system will be implemented in two steps: preprocessing and training testing. After training the models with 80% data from all of the aforementioned models, the Gradient Boosting Classifier produces more accurate results, with accuracy of 90%, 90%, and 92% for confirmed cases, cured cases, and death cases. The next method is XGBoost, which provides a second more accurate result with accuracies of 89%, 89%, and 88%. The third one is Decision Tree, with accuracies of 89%, 83%, and 83%.The Random Forest is ranked fourth with 68%, 86%, and 86% accuracies, while the Support Vector Machine is ranked fifth with 76%, 69%, and 73% respectively. The sixth and last model is Naïve Bayes, with accuracy rates of 89%, 49%, and 24%, respectively.

REFERENCES

[1]. Arpaci, I., Al-Emran, M., Al-Sharafi, M. A., & Marques, G. (2021). Predicting the COVID-19 infection with 14 clinical features using machine learning classification algorithms. *Multimedia Tools and Applications, 80*(13), 19393–19415. https://doi.org/10.1007/s11042-021-10748-2

[2]. Azarafza, M., Azarafza, M., & Tanha, J. (2020). COVID-19 prediction using artificial intelligence and time series methods: A case study of Iran. *Chaos, Solitons & Fractals, 140,* 110199. https://doi.org/10.1016/j.chaos.2020.110199

[3]. Benvenuto, D., Giovanetti, M., Vassallo, L., Angeletti, S., & Ciccozzi, M. (2020). Application of the ARIMA model on the COVID-2019 epidemic dataset. *Data in Brief, 29,* 105340. https://doi.org/10.1016/j.dib.2020.105340

[4]. Daniyal, M., Nasir, J. A., & Malik, A. (2020). Statistical modeling for the prediction of COVID-19 mortality trend in Pakistan. *Journal of Infection and Public Health, 13*(11), 1667–1673. https://doi.org/10.1016/j.jiph.2020.08.012

[5]. Khanday, A. M. U. D., Rabani, S. T., Khan, Q. R., Rouf, N., & Din, M. M. U. (2020). Machine learning based approaches for detecting COVID-19 using clinical text data. *International Journal of Information Technology, 12*(3), 731–739. https://doi.org/10.1007/s41870-020-00495-9

[6]. Lasya, G., Gupta, S., & Kumar, P. (2022). Machine learning based COVID-19 prediction models and comparative performance analysis. *Materials Today: Proceedings, 62,* 4361–4367. https://doi.org/10.1016/j.matpr.2022.02.658

[7]. Mary, M. S., & Raj, A. A. (2021). Prediction of COVID-19 using supervised machine learning models. *Materials Today: Proceedings, 45,* 6107–6111. https://doi.org/10.1016/j.matpr.2020.12.1161

[8]. Muhammad, L. J., Algehyne, E. A., Usman, S. S., Adebayo, P. W., Joshua, O. J., & Mohammed, I. A. (2021). Supervised machine learning models for prediction of COVID-19 infection using epidemiology dataset. *SN Computer Science, 2,* 11. https://doi.org/10.1007/s42979-020-00394-7

[9]. Painuli, D., Mishra, R., Bhardwaj, A., & Sharma, R. (2021). Forecasting COVID-19 confirmed cases in India using machine learning models. *Materials Today: Proceedings, 46,* 10629–10634. https://doi.org/10.1016/j.matpr.2020.12.1023

[10]. Zhao, Z., Chen, W., Wu, X., Chen, P. C. Y., & Xu, Z. (2021). Prediction of COVID-19 spread curves in South Korea and Italy by scaling exponents of power-law growth. *Nonlinear Dynamics, 101*(3), 1833–1848. https://doi.org/10.1007/s11071-020-05743-y

