Advances in Nonlinear Variational Inequalities
ISSN: 1092-910X
Vol 28 No. 1s (2025)

Machine Learning-Based Prediction of COVID-19 Patient Outcomes: Enhancing Clinical Decision-Making

Nitin N. Sakhare[1], Gajanan M. Walunjkar[2], Ganesh Narkhede[3], Dharmesh Dhabliya[1]

[1]Department of Computer Engineering, Vishwakarma Institute of information technology, Kondhwa, Pune, Maharashtra, India

[2]Department of Information Technology, Army Institute of Technology, Pune, Maharashtra, India

[3]Department of Mechanical Engineering, Vishwakarma Institute of Information Technology, Pune, Maharashtra, India

Article History:

***Received**: 30-07-2024*

***Revised:** 27-09-2024*

***Accepted**: 13-10-2024*

Abstract:

The COVID-19 epidemic has created unprecedented challenges to healthcare systems around the world, demanding the development of reliable techniques for predicting patient outcomes to guide clinical decisions. This paper comprehensively explores machine learning-based methodologies for predicting COVID-19 patient outcomes. Using a broad dataset that includes demographic information, symptoms, comorbidities, laboratory test results, and imaging data, we use cutting-edge machine-learning approaches to create prediction models. Data preprocessing techniques including feature engineering and selection are applied to enhance model performance and reliability. Various machine learning algorithms, support vector machine, logistic regression, support and Naïve Bayes, are evaluated for their efficacy in predicting disease severity, hospitalization, and mortality outcomes. Model performance is assessed and evaluated using standard evaluation criteria, assuring the robustness of all models. These models provide useful insights into COVID-19 patient prognosis and can assist healthcare staff in triaging patients, optimizing treatment regimens, and allocating resources more efficiently. Ethical concerns around patient data privacy and model interpretability are thoroughly followed throughout the study. In all, this work highlights the promise of machine learning techniques for improving COVID-19 patient care and public health response efforts.

Keywords: Prediction, IoT, Health Monitoring, COVID-19, Sensors, Machine Learning.

1. Introduction

The highly contagious respiratory disease COVID-19, commonly known as the coronavirus disease 2019, spread to become a worldwide pandemic. The global pandemic of COVID-19 has had a significant impact on daily lives, economy, and public health. To stop the virus from spreading, governments and health groups have put in place several precautions, including travel restrictions and lockdowns. Many symptoms are caused by COVID-19 and might manifest two to fourteen days after exposure. Fever, coughing, exhaustion, shortness of breathe, and loss of taste or smell are typical symptoms. On the other hand, some individuals might not exhibit any symptoms, or they might be asymptomatic. The main reasons behind the spread of the virus are respiratory droplets when an infected person coughs, sneezes, talks, or breathes. Additionally, it can spread through skin-to-face contact with contaminated surfaces.by the virus SARS-CoV-2. Vaccines against COVID-19 have been created to offer protection from the virus. Many vaccinations from various producers have been approved for use in emergency situations in several different nations. Global vaccination initiatives

are in place to stop the virus's spread and stop serious sickness. The course of the treatment is determined by the severity of the COVID-19 infection. People with severe symptoms are treated with supportive care, which may include oxygen therapy. To treat COVID-19, several drugs and treatments have been approved for use in an emergency. The pandemic of COVID-19 has posed numerous challenges for international healthcare systems, emphasizing the need for accurate predictive methods to forecast patient outcomes. In the face of this urgency, machine learning methods have surfaced as viable means of predicting the course of COVID-19 cases and directing clinical judgment. Through the utilization of varied patient data, such as demographic profiles, clinical symptoms, comorbidities, laboratory results, and imaging results, machine learning models can identify trends and forecast outcomes that span from the severity of a disease to its eventual death. An overview of the approaches used in applying machine learning to COVID-19 patient outcome prediction is provided in this introduction, with a strong focus on building a robust model system, and possible effects on resource allocation and healthcare delivery. This project intends to benefit patient care methods and support current efforts to prevent the COVID-19 pandemic by conducting a thorough investigation of machine learning-based approaches.

The literature review of the research project suggested in this study is presented in Section 2. A succinct analysis of the problem description and suggested remedy is given in Section 3. The results are summarized in Section 4 along with a comparison of the several machine learning techniques that were employed. The research work's conclusion is provided in Section 5.

2. LITERATURE SURVEY

Most of the researchers used various machine learning techniques to automatically detect coronaviruses. Authors in [1] used data on suspects of COVID-19 and rates at which these suspects are admitted to the hospitals for treatment. The data available from Israelita Albert Einstein Medical Center, Sao Paulo, Brazil has been used for machine learning model training. The clinical details like blood cell counts (CBC), details about the critical body organs like liver, kidney and diabetes of suspected and actual COVID-19 patients were recorded in the data. Using six machine learning methods that included different ensemble techniques, such as bagging, gradient boosting, and adaptive boosting, the authors trained models on the dataset. Based on the ROC curve, confidence score, test accuracy, and training accuracy of the machine learning models, they assessed them. 0.97 and 0.95, was the best accuracies recorded for training and validation datasets respectively, were attained by the random forest method. Using a variety of machine learning techniques, authors in [2] investigated the detection of the COVID-19 epidemic's beginning. It is anticipated that the authors' research on the disease's origins, forecasting, and time-series monitoring would help with the disease's future management. At 98% accuracy, KNN with artificial oversampling achieved the best results. Authors in [3] focused on employing the SVM machine learning model for sophisticated coronavirus identification. A modified cuckoo search strategy popular for hyperparameter optimization was used to tune the hyperparameters of SVM and then increase its accuracy. COVID-19 and healthy cases were sorted using an innovative feature selection framework called mRMR (Minimum Redundancy Maximum Relevance). Authors in [4] employed four different machine learning approaches to determine whether COVID-19 was present in each patient. The XGBoost model had shown a recall 99.26% with an accuracy level of 97.71%. The SVM method earned the highest degree of accuracy at 98.38%. Using traditional machine learning techniques, Authors in [5] investigated the COVID-19

infectious patterns, the rate of fatality and treatment measures. Additional efforts were also made to forecast the virus's future expansion. The naive Bayes methodology produced the most accurate COVID-19 disease projection, according to the authors' findings. Authors in [6] used majority rule-based ensemble approaches for the prediction of death rate because of COVID-19 infected individuals. In this work, approaches to feature selection, synthetic oversampling, and multivariate imputation were used which showed an accuracy of 86.9% and F1 coefficient of 71.6%. The XGBoost model had shown a better result with 94.11% accuracy. A customized attention-based CNN framework was designed by authors in [7] for the detection of COVID-19. A total of 936 CT scans of the chest were used. Overall accuracy reported was 96.32% with an F-1 score of 96.33%. To identify coronavirus, Authors in [8] used a specially designed CNN framework combined with a multi-way picture augmentation method. Their greatest accuracy was 96.36%, and their F1 index was 96.35%. Authors in [9] started the autonomous forecasting of COVID-19 by various machine and deep learning based with a wrapper feature selection technique. CNN predicted COVID-19 with 80% accuracy using the BGA method. Authors in [10] used a variety of machine learning and neural network models to predict coronavirus illness positivity and severity. The scientists classified positive and negative cases with greater than 84% accuracy by using artificial neural network, DT, and KNN models. Using the decision tree approach, patients were classified according to the severity of their disease with an accuracy of over 92%.

3. Methodology

About 2405 rows and 6 columns make up the dataset utilized for analysis. The dataset was gathered using an implementation basis that was in real time and appropriate analysis was conducted in accordance with the research's input criteria. Every requirement for an input was subject to several circumstances, and the final data output is the result of identifying COVID patients. It was discovered that conflicting values may be detected in the dataset that was gathered from the real-time implementation. Thus, consistency was maintained while manipulating the data. The gathered dataset needs to be transformed into a suitable format, with the dataset being transformed into Boolean values for both detection and non-detection. There are no variable quantities in the dataset because it was gathered via the real-time implementation of various testing sensors. Many attributes are included in the dataset, including temperature, pulse, IR, gender, age, and SPo2. Therefore, each measure helps identify patients with COVID-19. Several AI algorithms are used in data modeling, some of which may be the most accurate and best fit. Thus, we have coaching datasets with values for heart rate, breathing rate, oxygenation rate, and so forth. Thus, a variety of models were employed to forecast COVID-19 patients.

Table 1 Threshold Values

Component	**Range**
TEMPERATURE	normal range between 36 to 98 Fahrenheit
SPO2	normal range above 95
IR	0 and 1 to detect
PULSE RATE	Normal range between 60 to 100

A method of displaying records and processed information in a graphical representation is called data visualization. A variety of consulting tools, including maps, charts, graphs, and many more, can be

used in data visualization. The visualization system makes it easy to understand the styles and attributes present in the created data. Big Data technology advancements have led to an unexpected desire to interpret the enormous volumes of records. Displaying information and ensuring that the ML model is running correctly are crucial due to the advancements in machine learning and analytics techniques. Furthermore, when compared to numerical clusters of records, the graphical records are translated higher. It makes it easier to bind variables to each other in order to establish the relationship between the variables. This is a quite clean instance of records visualization in records analysis. Based on diverse parameters of covid-19 we had been plotting the graph.

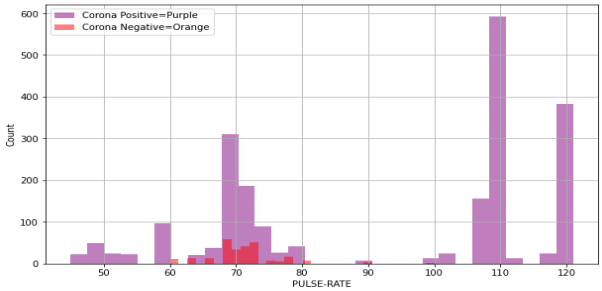

Fig. 1. Pulse Rate Detection

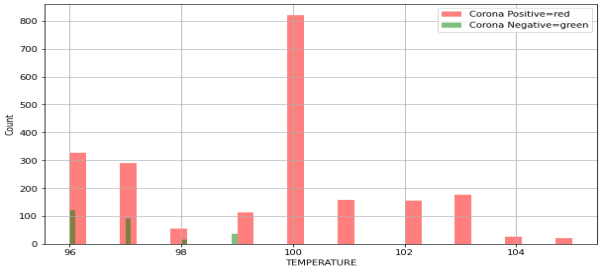

Fig. 2. Temperature Detection

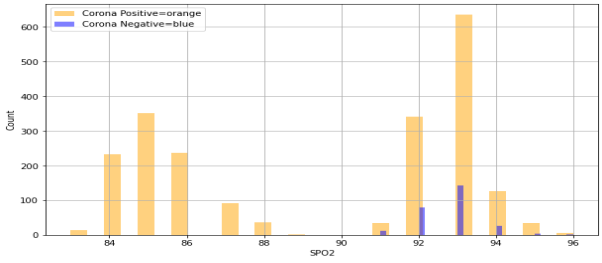

Fig. 3. SPO2 Detection

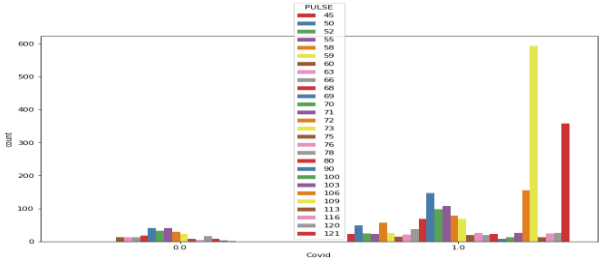

Fig. 4. Pulse Rate vs. Covid

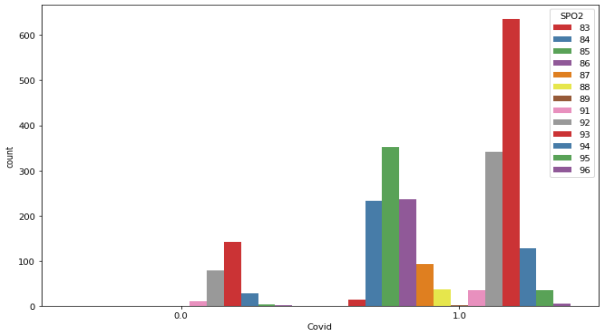

Fig. 5 SPO2 vs. Covid

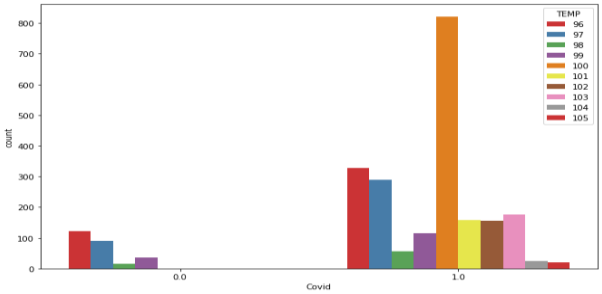

Fig. 6 Temperature vs Covid

Machine Learning Algorithms

A. Logistics Regression

Logistic regression is a classification algorithm, which predicts the outcomes of dependent variable in a binary format. The binary values indicate 0 as negative class and 1 as positive class. The Logistic Regression uses a cost function called "Sigmoid" function that makes it a complex algorithm in comparison to Logistic Regression algorithm. Hypothesis on Logistics regression limit its outcomes between 0 and 1. The result of hypothesis on Logistic Regression fails to represents the hypothesis ranging less than 0 and greater than 1. In Logistic Regression, the "Sigmoid" function maps the

predicted hypothesis results. Below is the formula to calculate sigmoid function:

$$f(X) = \frac{1}{1 + e^{-x}}$$

Where,
f(x) = probabilistic outcome of independent variable x = input given
e = base of natural log
This algorithm predicts COVID-19 patients based on multiple characteristics, including coaching, testing, and overall accuracy, which has been found to be 95% on multiple occasions.

B. Support Vector Machines

Algorithms for supervised machine learning include Support Vector Machine. The SVM method performs well in issues involving regression and classification. To quickly classify newly entered data into the most appropriate categories, SVM seeks to create the decision based boundary needed to sort N-dimensional spaces into classes. The Hyperplane is the name given to the decision border. When building a hyperplane, SVM often selects the vector points that are the most correct. Support Vectors is the term given to these extreme vectors, which identify the SVM algorithm. Below diagram show the hyperplane plotting for SVM: This algorithm predicts coaching, testing, and overall accuracy, which comes out to be 97% many times, by utilizing several characteristics to forecast COVID patients.

C. Naive Bayes

To handle classification problems, Naïve Bayes is categorized as a supervised learning algorithm. The algorithm is predicated on the Bayes theorem in mathematics. The technique performs optimally when used to high dimensionality training datasets. The algorithm is easy to use, incredibly efficient, and predicts outcomes quickly. To find the optimal solution for the given problem statement, the following stages are taken: first, the dataset to be trained is transformed into frequency tables; next, the feature's probability is discovered; and last, the posterior probability is computed using the Bayes theorem.

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)}$$

P(c|x) = Posterior Probability, P(x|c) = Likelihood, P(c)= Class Prior Proability, P(x)= Predictor Prior Probability
This model forecasts the coaching, testing, and overall accuracy of 98% for COVID-19 patients based on a number of different parameters.

4. Results

Confusion Matrix
A confusion matrix is a matrix-based structure which is used to analyze the performance of a learning model. It contrasts the desired values with what the machine learning model predicts. Using 2x2 matrixes, the Binary classification model provides four values, as an associate degree example.

Table 2: Confusion Matrix for Logistic Regression

	Covid Negative (Actual)	Covid Positive (Actual)
Covid Negative (Predicted)	641	0
Covid Positive (Predicted)	16	65

Table 3: Confusion Matrix for Support Vector Machine

	Covid Negative (Actual)	Covid Positive (Actual)
Covid Negative (Predicted)	646	8
Covid Positive (Predicted)	11	57

Table 4: Confusion Matrix for Naïve Bayes

	Covid Negative (Actual)	Covid Positive (Actual)
Covid Negative (Predicted)	652	2
Covid Positive (Predicted)	5	63

A. Precision

Precision makes certain that a classifier doesn't label a negative instance example as positive. for each category, its mentioned due to the fact the significance relation of proper positives to the be an actual positive and false positive.

Precision = $\frac{True\ Positives}{True\ Positives+False\ Positives}$

In the true positive case, we generally tend to have an affected person that Covid and is detected as positive. In false positive the affected person isn't Covid-detected but the version predicted the affected person as Covid affected person.

B. Recall

Recall is for a classifier, types all the true positive instances. Its published due to the quantitative relation among fact positives and moreover the upload of true positives with false positives.

Recall- Fraction of positives that had been nicely better-recognized.

Precision = $\frac{True\ Positives}{True\ Positives+False\ Negatives}$

C. F1-score

F1 Score is that the weighted common among exactness and recollect values. It gives stability among the exactness and recollect values. F-1 rating considers each of the parameters. fake positives and fake negatives. It measures the accuracy of a version on a given dataset. It offers binary outcomes, positive or negative. $F1 = 2 \times \frac{Precision*Recall}{Precision+Recall}$

D. Accuracy= (Total Number of Predictions/Number of Correct Predictions) ×100%

Table 5 Classification report: Logistic Regression

	TPR	FPR	Precision	Recall	F1-score
Covid negative	0.97	0	1.00	0.98	0.99
Covid positive	1.00	0.02	0.80	1.00	0.89

Table 6 Classification report: SVM

	TPR	FPR	Precision	Recall	F1-score
Covid negative	0.98	0.12	0.98	0.98	0.98
Covid positive	0.87	0.02	0.83	0.87	0.85

Table 7 Classification report: Naive Bayes

	TPR	FPR	Precision	Recall	F1-score
Covid negative	0.99	0.03	0.98	0.99	0.99
Covid positive	0.97	0.01	0.93	0.97	0.95

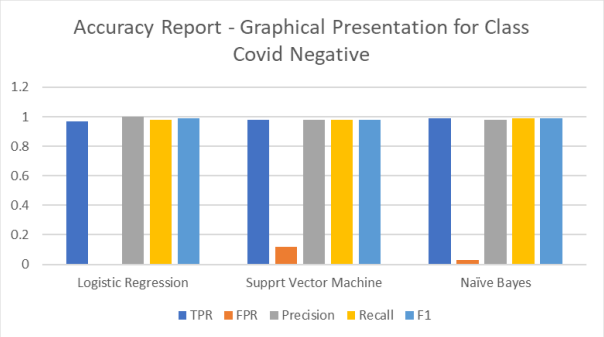

Fig. 7 Accuracy report – Graphical Presentation for class Covid negative

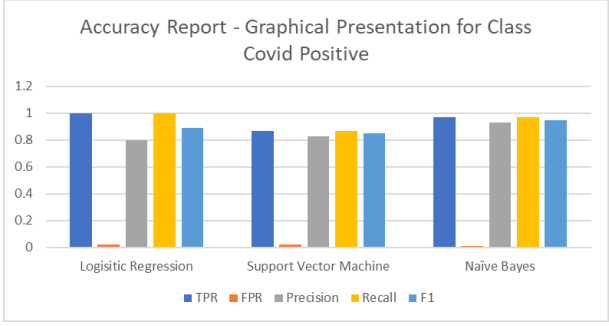

Fig. 8 Accuracy report – Graphical Presentation for class covid positive

Table 8: Training and Testing Accuracy Report

	Training Accuracy	Testing Accuracy
Logistic Regression	97.78%	92.63%
Support Vector Machine	97.37%	90.50%
Naïve Bayes	99.03%	95.33%

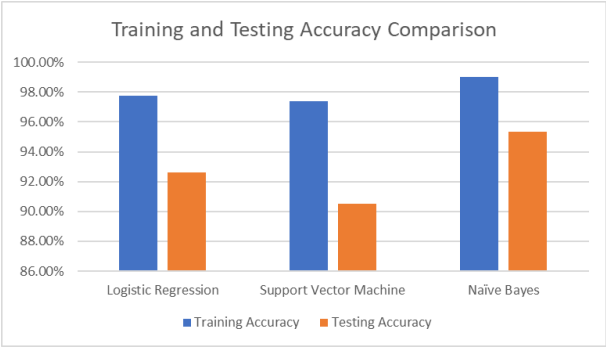

Fig 9. Training and Testing Accuracy Prediction

5. CONCLUSION

In conclusion, the utilization of machine learning methods to forecast COVID-19 patient outcomes is a significant development in healthcare decision-making during the current pandemic. Machine learning provides useful insights into disease prognosis through the study of various patient data and other supportive parameters This helps doctors and medical staff to efficiently prioritize patient care, optimize resource allocation, and customize treatment plans, ensuring minimum workload. Even with the encouraging results of this study, some issues persist, such as the requirement for bigger and more varied datasets, the interpretability of the model, and moral concerns about patient privacy. Interdisciplinary teams made up of data scientists, doctors, and healthcare experts must work together to find effective solutions to these hurdles. Predictive models must also be continuously validated and monitored to maintain their relevance and dependability in dynamic healthcare environments.

References

[1] Solayman S, Aumi SA, Mery CS, Mubassir M, Khan R. Automatic COVID-19 prediction using explainable machine learning techniques. International Journal of Cognitive Computing in Engineering. 2023 Jun; 4:36–46. doi: 10.1016/j.ijcce.2023.01.003. Epub 2023 Jan 25. PMCID: PMC9876019.

[2] Zoabi, Y., Deri-Rozov, S. & Shomron, N. Machine learning-based prediction of COVID-19 diagnosis based on symptoms. *npj Digit. Med.* **4**, 3 (2021). https://doi.org/10.1038/s41746-020-00372-6

[3] Alimadadi, A., Aryal, S., Manandhar, I., Munroe, P. B., Joe, B., and Cheng, X. (2020). Artificial Intelligence and Machine Learning to Fight Covid-19. *Physiol. Genomics* 52, 200–202. doi:10.1152/physiolgenomics.00029.2020

[4] Meraihi, Y., Gabis, A.B., Mirjalili, S. et al. Machine Learning-Based Research for COVID-19 Detection, Diagnosis, and Prediction: A Survey. SN COMPUT. SCI. 3, 286 (2022). https://doi.org/10.1007/s42979-022-01184-z

[5] Chen, Y., Ouyang, L., Bao, S., Li, Q., Han, L., Zhang, H., et al. An Interpretable Machine Learning Framework for Accurate Severe vs Non-severe Covid-19 Clinical Type Classification. medRxiv (2020). doi:10.1101/2020.05.18.20105841

[6] Junling Luo, Zhongliang Zhang, Yao Fu, Feng Rao, Time series prediction of COVID-19 transmission in America using LSTM and XGBoost algorithms, Results in Physics, Volume 27, 2021, ISSN 2211-3797, https://doi.org/10.1016/j.rinp.2021.104462

[7] Asif, S., Zhao, M., Tang, F. et al. A deep learning-based framework for detecting COVID-19 patients using chest X-rays. Multimedia Systems 28, 1495–1513 (2022). https://doi.org/10.1007/s00530-022-00917-7

[8] Kabid Hassan Shibly, Samrat Kumar Dey, Md Tahzib-Ul Islam, Md Mahbubur Rahman, COVID faster R–CNN: A novel framework to Diagnose Novel Coronavirus Disease (COVID-19) in X-Ray images, Informatics in Medicine Unlocked, Volume 20, 2020, ISSN 2352-9148, https://doi.org/10.1016/j.imu.2020.100405

[9] Afifi, A.; Hafsa, N.E.; Ali, M.A.S.; Alhumam, A.; Alsalman, S. An Ensemble of Global and Local-Attention Based Convolutional Neural Networks for COVID-19 Diagnosis on Chest X-ray Images. Symmetry 2021, 13, 113. https://doi.org/10.3390/sym13010113

[10] Norah Alballa, Isra Al-Turaiki,Machine learning approaches in COVID-19 diagnosis, mortality, and severity risk prediction: A review, Informatics in Medicine Unlocked, Volume 24, 2021, ISSN 2352-9148, https://doi.org/10.1016/j.imu.2021.100564.