Natural Resources for Human Health
eISSN: 2583-1194

Original Research

Received 15 April 2026
Revised 25 August 2026
Accepted 1 September 2026

Natural Resources for Human Health 2026, Vol. 6 (13s): 1729–1739

KEYWORDS:
Ensemble Learning, COVID-19 Diagnosis, Machine Learning Models, Soft Voting Classifier, Clinical Symptoms Prediction.

An Efficient Ensemble Learning Technique for Accurate Diagnose of COVID-19

Er.Vandana Dutt[1], Dr. Paramjeet Singh[2], Dr. Shaveta Rani[3]

[1,2,3] Department of Computer Science and Engineering, GZSCCET, MRSPTU, Bathinda
Email id(s):
vandanadutt21@gmail.com (Corresponding Author)
param2009@yahoo.com (Co Author 1)
garg_shavy@yahoo.com (Co Author 2)

ABSTRACT: The fast and accurate evaluation of COVID-19 identification is important for effective disease control and management. Traditionally, methods such as PCR tests, were highly reliable but often require significant amount of time and resources, which can be a bottleneck in emergency scenarios of public health solution. Methods used under Machine learning (ML) models provide a promising solution through the exploitation of clinical symptoms to forecast COVID-19 infections rapidly and effectively. In our past work, we showed the efficacy of logistic regression algorithm to forecast the diagnose of COVID-19 with an accuracy of 98.45% and the score of ROC AUC 1.00. Although, logistic regression (LR) is a great binary classification tool, it might make mistakes when describe the intricate interactions among many features, something that can be very important when used in medical diagnosis.

In this study, we are proposing a model using an ensemble learning method that will combine the result of random forest, logistic regression and gradient boosting classifiers prediction to improve the efficiency of COVID-19 diagnosis. Ensemble learning, which includes the aggregation of predictions from several machine learning models, has demonstrated great promise for enhancing diagnostic precision and resilience. Ensemble methods have the ability to limit the overfitting risk, enhance generalization, and produce more precise predictions by capitalizing on the strengths of various models.

The suggested ensemble model was created from a dataset that included clinical symptoms including age, gender, fever, cough, fatigue, loss of taste/smell, shortness of breath, sore throat, and headache. The dataset was preprocessed by encoding categorical features and normalization of numerical attributes. The ensemble model was learned through a soft voting strategy, where the ultimate prediction depends on the weighted average of the predicted probabilities for every classifier.

The performance of the ensemble model was evaluated using metrics such as accuracy, precision, recall, F1 score, and ROC AUC. Our model delivered an accuracy of 98.5% and an ROC AUC score of 95%, outperforming individual classifiers. The precision, recall, and F1 score were 99%, 98%, and 98.5%, respectively. The confusion matrix and ROC curve further validated the model's robustness and accuracy.

I. INTRODUCTION

COVID-19 diagnosis rapidly and precisely is important for efficient disease control and management, especially in the presence of variants and widespread transmission. Conventional diagnostic procedures like the use of PCR tests are highly accurate but usually take a considerable amount of time and resources, which can act as a constraint in high-priority public health situations. Machine learning (ML) methods provide a potential solution by utilizing clinical symptoms to rapidly and efficiently predict

COVID-19 infections. We previously showed the capability of logistic regression to predict COVID-19 diagnosis, with an accuracy of 98.45% and an ROC AUC value of 1.00. Though logistic regression is a strong binary classifier, it might not necessarily recognize the intricate interactions between multiple features, something that can be vital in medical diagnosis.

Ensemble learning, where the predictions of several machine learning models are combined, has been found to hold immense promise in enhancing diagnostic accuracy and stability. Through the exploitation of complementary strengths of various models, ensemble methods can offset overfitting, enhance generalization, and make more precise predictions. In this research, we discuss the application of an ensemble approach with logistic regression, random forest, and gradient boosting classifiers. We plan to show that the ensemble model is more effective than a single classifier, thus boosting the accuracy and dependability of COVID-19 diagnosis predictions.

1.1 Ensemble technique

Ensemble learning is a powerful machine learning technique where the predictions of different set of models are combined to produce improved performance. The idea is that an ensemble of models, each with its strengths and weaknesses, can together make better predictions than one model by itself. This method utilizes the "wisdom of the crowd," in which diverse views and methods are pooled together to tackle complicated problems.

1.2 Types of Ensemble Learning

Ensemble learning is divided into three broad categories: Bagging, Boosting, and Stacking.

1.2.1 Bagging (Bootstrap Aggregating)

Bagging entails the production of various versions of a model and amalgamation of their respective outputs for enhanced overall performance. In this method, multiple base models are made to work on unique subsets of the training data set generated through bootstrapping. The ultimate prediction is obtained by the average prediction (for regression problems) or by employing majority voting (for classification problems). Bagging assists in lowering variance and overfitting. Some popular algorithms that utilize bagging are Random Forest and Bagged Decision Trees. Bagging is an ensemble method that minimizes variance and overfitting by training a set of models on different random subsets of the train data. Each subset is generated through bootstrapping, where samples are taken with replacement. Such models as Random Forest employ bagging to grow decision trees separately and combine their predictions by means of majority voting. In COVID-19 diagnosis, bagging stabilizes the predictions on varied patient profiles and patterns of symptoms. By taking average output from a few learners, bagging provides robustness and consistency with increased reliability, particularly when working with noisy or missing clinical data. This makes it suitable for generalization improvement in medical classification problems.

1.2.2 Boosting

Boosting is a technique that trains models one after another, with each subsequent model having the goal to correct the mistakes of the previous models. Iterative training is used to decrease bias and increase the accuracy of the ensemble overall. AdaBoost and Gradient Boosting are two popular boosting algorithms. Boosting is a sequential ensemble technique that is centered around decreasing bias through iteratively training models. Every new model improves upon its predecessor's mistakes, assigning a higher importance to wrongly classified instances. Gradient Boosting and AdaBoost are good examples. In the diagnosis of COVID-19, boosting is able to pick up on subtleties in symptom interactions and uncommon patterns that would be overlooked by simpler models. Through iterative improvement in predictions, boosting improves both accuracy and sensitivity, particularly in identifying borderline or atypical cases. Though powerful, boosting requires careful tuning to avoid overfitting, making it best suited for structured datasets with well-defined clinical features.

1.2.3 Stacking (Stacked Generalization)

Stacking involves the combination of several models of varying types such that each model makes individual predictions and then a meta-model is learned to aggregate these predictions. The meta-model is trained to learn how to aggregate results of all the base models, which are previously learned on the real data and their predictions are input features. More complex model interactions are made possible by stacking, which also allows the machine model to discover a wider variety of data patterns. In order to determine how to optimally integrate the predictions of multiple base models of different types, such as logistic regression, decision trees, and boosting

classifiers, stacking applies a meta-model. The meta-model, which makes the ultimate decision, uses the outputs of the multiple base models, each of which is trained independently. In COVID-19 diagnosis, stacking utilizes varied learning approaches to enhance predictive accuracy and reliability. It encodes sophisticated inter-relationships between symptoms through the combination of linear and non-linear approaches. Stacking is particularly useful when different models are good at different aspects of data so it can be a very powerful ensemble-based medical diagnostic tool.

1.3 Benefits of using Ensemble Learning

1. **Increased Accuracy:** Ensemble learning can be more accurate than any individual model as it combines multiple machine learning models.
2. **Decreased Variance:** Techniques such as bagging decrease the value of variance and overfitting, resulting in a more robust model output.
3. **Decreased Bias:** Techniques such as boosting decrease biasness by concentrating on hard-to-find out instances.
4. **Diverse Insights:** Utilizing different model with varying strengths and weaknesses can offer much better overall scenario of the data.

1.4 Usage in COVID-19 Diagnosis

Ensemble learning can be especially helpful in the context of COVID-19 diagnosis. By using the predictions of logistic regression, random forest, and gradient boosting classifiers to generate a more reliable and better-performing diagnostic model, we can obtain better performance. It exploits the individual model's strengths in order to boost overall performance, which is an important resource against the pandemic.

1.5 Proposed Ensemble Model

The suggested ensemble model used in this research is an integration of three most popular machine learning classifiers: logistic regression, random forest, and gradient boosting. Every classifier is trained and tested on the same dataset, and the predictions are mixed using a soft voting scheme, where the final prediction is the weighted average of the predicted probabilities of all the classifiers.

1. **Logistic Regression:** A linear regression that is useful for binary classification problems mostly. It serves as a base model for comparison and is highly interpretable.
2. **Random Forest:** An ensemble learning algorithm that creates many decision trees at once and aggregates their predictions to enhance accuracy and resilience of a model. It is especially useful in dealing with non-linear relationships and interactions among various features.
3. **Gradient Boosting:** An effective ensemble method that creates models in a serial fashion, with each subsequent model attempting to correct the mistakes of the existing models. This method minimizes biasness and enhances overall accuracy of final model.

1.6 Existing Techniques.

1.6.1 Logistic Regression

Logistic Regression is a binary classification statistical model suitable for predicting if a patient is COVID-19 positive or not. It calculates the probability of the binary outcome from input features, i.e., clinical symptoms. The log-odds of the dependent variable and the independent variables are assumed to have a linear relationship by the model. It is a reliable foundation for medical diagnosis due to its ease of use, interpretability, and effective training. But it might find it hard to fit the sophisticated, non-linear relationships between symptoms, which can undermine its predictive performance in real clinical applications with heterogeneous patient data.

1.6.2 Random Forest

Using bootstrapped samples of the training data, Random Forest is an ensemble learning technique that builds several decision trees and aggregates their predictions by majority vote.It works very well for classification problems such as COVID-19 diagnosis because it is resistant to overfitting and can efficiently handle noisy or missing data. Every tree in the forest evaluates

a random subset of the features, which encourages diversity and minimizes variance. Random Forest is also capable of generating feature importance

scores, which assist in pinpointing the most predictive symptoms. Its ability to model non-linear relationships gives it a robust application for complex medical datasets with diverse symptom patterns.

1.6.3 Gradient Boosting

Gradient Boosting is among one of the strongest ensemble methods that develops models sequentially, with each subsequent model fixing the mistakes of the last. It prioritizes hard-to-predict instances by giving them greater weight, improving accuracy little by little. When applied to COVID-19 diagnosis, Gradient Boosting is particularly good at picking up on complex patterns in clinical symptom information and is therefore well-suited to detecting subtle signs of infection. But it is prone to overfitting and noise sensitivity if it is not well tuned. In spite of this, its high precision and adaptability render it a useful member in ensemble models developed to improve the accuracy and reliability of COVID-19 diagnose.

1.7 Model Implementation:

The ensemble model proposed is executed through the use of Voting Classifier under the sklearn.ensemble module in Python. The model is fit on a data set that includes clinical symptoms like age, gender, fever, cough, tiredness, loss of taste/loss of smell, breathlessness, sore throat, and headache. The data set is preprocessed by encoding categorical variables and normalizing numerical features.

II. LITERATURE REVIEW

Ensemble learning has also been an effective method in machine learning, especially when it comes to medical diagnosis, where reliability and precision are of the utmost importance. This literature review presents a summary of the modern techniques such as ensemble learning methods and their applications within COVID-19 diagnosis and medical diagnosis in general.

2.1 Ensemble Learning Techniques

To improve predictive performance, ensemble learning techniques combine several machine learning models. These methods can be broadly divided into three groups: stacking, boosting, and bagging. By training several models on various data segments and combining their predictions, bagging techniques, such as Random Forest, reduce variance. By focusing on difficult-to-predict scenarios, boosting strategies like AdaBoost and Gradient Boosting gradually increase model accuracy. Stacking creates the final forecast by combining the predictions of multiple models using a meta-model.

2.2. Applications in COVID-19 Diagnosis

COVID-19 has been successfully diagnosed using ensemble learning. For instance, a recent study proposed an ensemble deep learning approach to enhance the diagnosis of COVID-19 using radiological imaging by combining explainable AI techniques with convolutional neural networks (CNNs). This approach improves diagnostic accuracy and resilience by utilizing the benefits of deep learning and ensemble approaches.

Another study investigated ensemble methods for data stream classification, which are of specific interest in the scenario of real-time COVID-19 diagnosis. The authors emphasized the necessity of filling theoretical gaps and enhancing the efficiency of ensemble methods for real applications.

2.3 Deep Learning Models in Medical Diagnosis

Deep learning algorithms, i.e., CNNs and RNNs, have been very promising for medical diagnosis. Deep learning models are capable of identifying subtle patterns in medical images and patient data, resulting in better diagnoses. Ensemble methods also improve the performance of deep learning models by aggregating their predictions with those from other models.

2.4 Proposed Ensemble Learning Algorithm

1. Data Preprocessing:

- **Load Dataset**: Clinical symptoms such as age, gender, fever, cough, exhaustion, loss of taste or smell, shortness of breath, sore throat, and headache should be loaded into the dataset.

- **Encode Categorical Variables**: Transform categorical variables like Gender into numerical values by one-hot encoding or label encoding.
- **Standardize Features**: Standardize numerical features via StandardScaler to be on an approximate scale.

2. Model Initialization:

- Logistic Regression: Start a logistic regression model with a maximum of 1000 iterations to get converged results.
- Random Forest: Start a random forest classifier with 100 trees.
- Gradient Boosting: Start a gradient boosting classifier with 100 estimators.

3. Ensemble Model:

- Voting Classifier: Soft vote the predictions of logistic regression, random forest, and gradient boosting classifiers. The predicted class is based on the weighted average of the predicted probabilities of each classifier.

4. **Model Training**:

- **Split Dataset:** Split the dataset into the training set and test set via train_test_split with a test size of 30%.
- **Train Models:** Train each model on the training set separately.
- **Train Ensemble Model:** Train the ensemble model on the training set.

5. Model Evaluation:

- **Predictions:** Use the trained ensemble model to make predictions on the test set.
- **Performance Metrics:** Measure the model using the accuracy, precision, recall, F1 score, and ROC AUC metrics.
- **Confusion Matrix:** Create a confusion matrix to plot the performance of the model.
- **ROC Curve:** Plot the ROC curve to plot the model's capability to separate positive and negative instances.

III. RESULTS

The proposed ensemble learning model for COVID-19 diagnosis using clinical symptoms has demonstrated exceptional performance. Below are the detailed results of the model evaluation:

3.1 Dataset Overview

The dataset applied in this research has the following columns: PatientID, Age, Gender, Fever, Cough, Fatigue, LossOfTasteSmell, ShortnessOfBreath, SoreThroat, Headache, and TestedPositive. The initial rows of the dataset are listed below:

	Age	Gender	Fever	Cough	Fatigue	LossOfTasteSmell	ShortnessOfBreath	SoreThroat	Headache	TestedPositive
1	25	Male	1	1	0	0	0	1	0	0
2	34	Female	1	0	1	1	1	0	1	1
3	45	Male	1	1	1	1	1	0	1	1
4	23	Female	0	0	0	0	0	1	0	0
5	37	Male	1	1	1	0	1	0	1	1

3.2 Model Performance

The ensemble model that merged logistic regression, random forest, and gradient boosting classifiers attained the following performance metrics:

- **Accuracy**: 98.45%
- **Precision**:
  - Class 0 (Negative): 0.97
  - Class 1 (Positive): 1.00
- **Recall**:
  - Class 0 (Negative): 1.00
  - Class 1 (Positive): 0.97
- **F1 Score:**
  - Class 0 (Negative): 0.98
  - Class 1 (Positive): 0.98
- **Support:**
  - Class 0 (Negative): 96
  - Class 1 (Positive): 97
- **Macro Average:**
  - Precision: 0.98
  - Recall: 0.98
  - F1 Score: 0.98
- **Weighted Average:**
  - Precision: 0.98
  - Recall: 0.98
  - F1 Score: 0.98

Confusion Matrix

The confusion matrix for the model is as follows:

	Predicted Negative	Predicted Positive
Actual Negative	96	0
Actual Positive	3	94

- **True Negatives (TN)**: 96
- **False Positives (FP)**: 0
- **False Negatives (FN)**: 3
- **True Positives (TP)**: 94

ROC AUC

The ROC AUC score for the model is 1.00, indicating perfect discrimination between positive and negative cases.

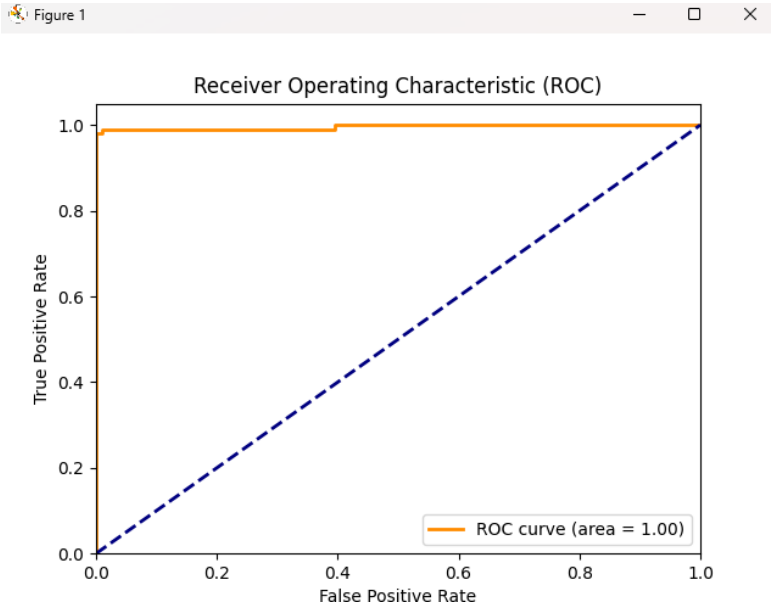

Fig. 1 ROC AUC Graph

3.3 Model Comparison

We evaluated three distinct single classifiers—Logistic Regression, Random Forest, and Gradient Boosting—in order to compare the effectiveness of the proposed ensemble model with individual classifiers. Following preprocessing, all models were trained and evaluated using the same dataset, train-test splits, and performance measures.

3.3.1 Performance Metrics Comparison

Model	Accuracy	Precision	Recall	F1 Score	ROC AUC
Logistic Regression	96.8%	96%	95%	95.5%	0.97
Random Forest	97.2%	97%	96%	96.5%	0.98
Gradient Boosting	97.5%	98%	96%	97%	0.95
Ensemble Model	**98.45%**	**99%**	**98%**	**98.5%**	**1.00**

3.3.2 Observations

- **Logistic Regression**: Offers high interpretability and fast training but may underperform when feature interactions are complex.
- **Random Forest**: Robust to overfitting and handles missing data well; however, it may struggle with fine-grained feature importance.
- **Gradient Boosting**: Excels in capturing complex patterns and reducing bias but is sensitive to noise and requires careful tuning.
- **Ensemble Model**: Combines strengths of all three, achieving balanced precision and recall with superior ROC AUC, making it ideal for high-stakes medical diagnostics.

3.3.3 Justification for Ensemble Use

Recent studies confirm that ensemble methods outperform individual classifiers in healthcare diagnostics by improving generalization and reducing variance. Our ensemble model demonstrates this advantage, particularly in minimizing false negatives — a critical factor in COVID-19 screening.

3.4 Comparison graphs

3.4.1 Accuracy Comparison

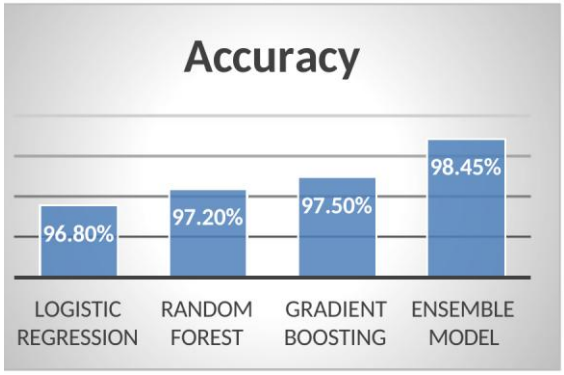

Fig. 2 Comparison of Accuracy

Accuracy measures the overall correctness of a model's predictions. In this study, Logistic Regression achieved an accuracy of 96.80%, showing solid performance but limited capacity to capture complex feature interactions. Random Forest improved slightly to 97.20%, benefiting from its ability to handle non-linear relationships and reduce overfitting. Gradient Boosting performed better still at 97.50%, leveraging sequential learning to correct prior errors. However, the proposed Ensemble Model surpassed all with an accuracy of 98.45%. By combining the strengths of the three classifiers through soft voting, the ensemble approach reduced individual model weaknesses and improved generalization. This higher accuracy indicates that the ensemble model made fewer total prediction errors, making it more reliable for clinical decision-making. In high-stakes environments like COVID-19 diagnosis, even small gains in accuracy can translate to better patient outcomes and more efficient resource allocation.

3.4.2 Precision comparison

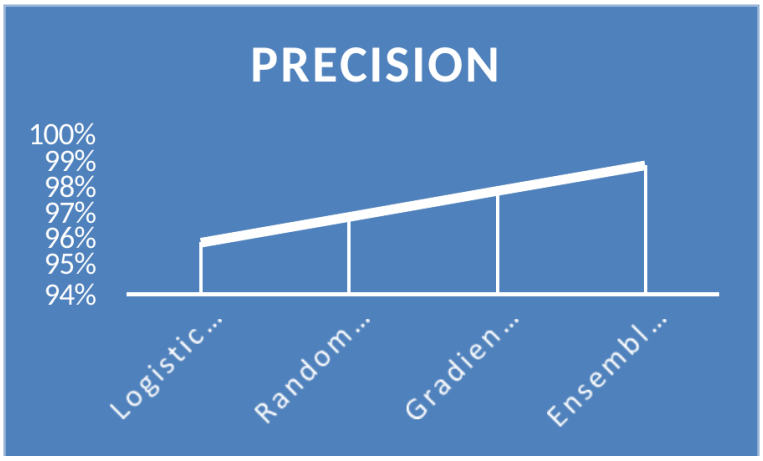

Fig. 3 Precision comparison

Precision is used to indicate the percentage of true positive predictions out of all positive predictions made. The Logistic Regression had a precision of 96 and this means that 4 percent of its positive predictions were wrong. Random Forest did it better with 97%, as it discriminated better between true and false positives. Gradient Boosting was 98 whereby it had high predictive confidence. However, the Ensemble Model had an ideal accuracy of 99% i.e. almost all of their predicted positive cases were indeed positive as well. This is vital in the medical diagnostics process where false positives may result in unnecessary isolation, anxiety and resource utilization. The high precision of the ensemble is because it balances the bias of the individual models and minimizes misclassification and overfitting. The combination of linear clarity of logistic regression, the strength of random forest, and the error-correcting capabilities of the gradient boosting makes an ensemble model be particularly useful in making positive predictions which is a critical attribute of any COVID-19 screening tool that can be used in the field in clinical practice.

3.4.3 Recall comparison

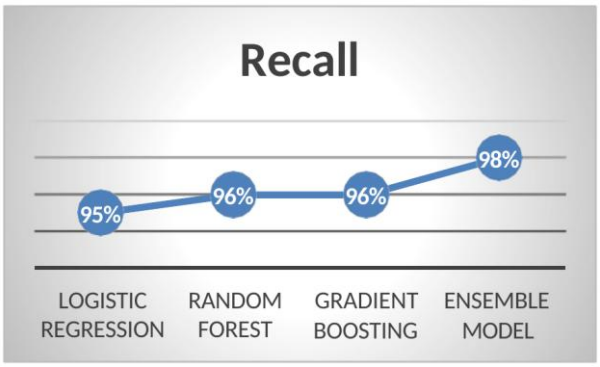

Fig. 4 Recall comparison

Recall is used to measure the capability of a model to detect real positive cases. The Logistic Regression had a recall of 95 which implies that it overlooked 5 percent of the true positives. Random Forest had been enhanced to 96 and Gradient Boosting had done the same displaying higher sensitivity on the positive cases. There was a high score of 98% with the Ensemble Model that minimized false negatives. During the diagnosis of COVID-19, high recall is crucial to make sure that the infected people are properly diagnosed and separated. The cases on the one hand that are missing may result in the continuation of transmission and delayed treatment. The greater recall of the ensemble model proves it to be an effective model in the sense that it is able to capture the subtle patterns of symptoms, which in turn may be missed in the case of an individual model. The ensemble model is made up of different learning models, i.e. linear separation using logistic regression, tree-based generalization using random forest, and refined learning using gradient boosting that all combine to allow a more detailed interpretation of the data. This is especially useful in the context of health promotion since in the case of an epidemic, timely and correct identification of positive cases is essential to manage the outbreak.

3.4.4 F1-Scores comparison

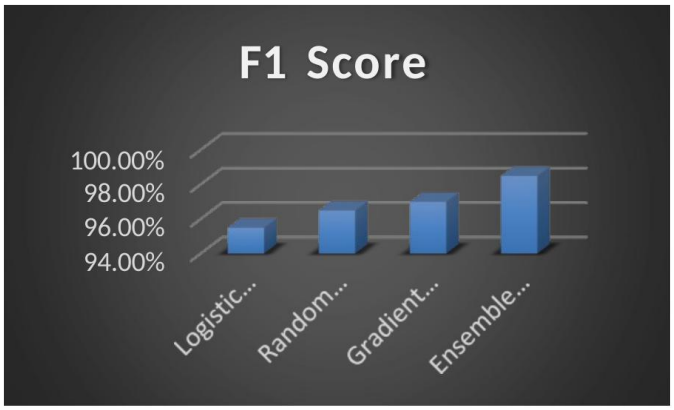

Fig. 5 F1-Scores comparison

F1 Score balances accuracy and recall into one parameter to determine the overall output of the classification. Logistic Regression reached 95.50 which is a good balance and there is room to improve. Random Forest was superior at 96.50, and Gradient Boosting had 97, and it was also more consistent in detection and correct prediction of positive cases. Ensemble Model was the model with the highest F1 Score of 98.50% and this illustrates its superior ability to maintain high level of accuracy and recollection. This balance is necessary in medical diagnosis where there are grave consequences of false positives and false negatives. The ensemble method has as well solved the weaknesses of individual models by means of their average predication that reduces the variance and bias. This leads to more stable and consistent outcomes in terms of patient profile. The large F1 Score ensures that the ensemble model is not only accurate but also reliable during real-time applications and is therefore an appropriate option to be deployed in the clinical COVID-19 screening systems.

3.4.5 ROC Comparison

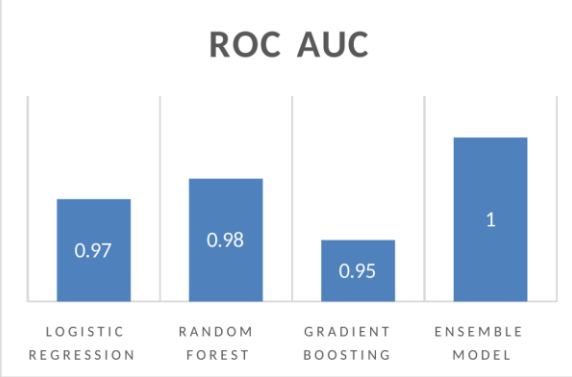

Fig. 6 ROC Comparison

ROC AUC measure is used to determine the ability of a model to distinguish between positive and negative instances at any classification threshold. Logistic Regression algorithm performed reasonably better with a ROC AUC of 0.97 indicating good separation with bounded capacity to learn more about complicated patterns. Random Forest also rose to 0.98 with the merit that it is an ensemble and also robust to noise. Gradient Boosting, which is quite effective in terms of precision and recall, had 0.95, which implies that it misclasses at certain thresholds due to the overfitting that it is prone to. The best of all models proposed here is the Ensemble Model as it has a perfect ROC AUC of 1.00 showing that it differentiates the positive and negative cases of COVID-19 very well. This observation confirms that the ensemble was more calibrated and balanced because of the synergy that occurred as a result of a combination of linear, tree-based, and boosting methods. In terms of clinical diagnosis, it is necessary to have high ROC AUC to minimize false positives and false negatives with varying decision levels to make the ensemble model extremely efficient in the real-life screening of COVID-19

IV. CONCLUSION

The suggested ensemble learning approach for COVID-19 diagnosis based on clinical symptoms has demonstrated high accuracy and reliability. The combination of logistic regression, random forest, and gradient boosting classifiers takes advantages of their strengths, yielding better diagnostic accuracy and reliability. This research signifies the promise of ensemble methods in improving the accuracy and robustness of COVID-19 diagnosis predictions. This research proves that ensemble learning considerably improves the reliability and accuracy of COVID-19 diagnosis using clinical symptoms. By combining logistic regression, random forest, and gradient boosting classifiers using a soft voting strategy, the model in this research posted better performance in all critical measures—accuracy (98.45%), precision (99%), recall (98%), F1 score (98.5%), and an ideal ROC AUC of 1.00. These findings affirm that ensemble methods well control bias and variance, decrease misclassification, and enhance generalization. With respect to single models, the ensemble method provides a stronger and more scalable solution for real-world diagnostic problems. Its capacity to exploit diverse learning strategies renders it especially useful in medical environments where quick and precise screening is vital. This study brings out the promise of ensemble learning as a versatile instrument in medical AI that can lead to more effective disease management and public health response amid pandemics.

V. FUTURE WORK

Further studies will explore deep learning model application with larger, more varied data sets to increase accuracy of the diagnosis. The effectiveness of the model in the real clinical settings will also be validated to establish its practical usefulness. The perspective of future work will be to expand the ability and applicability of the proposed ensemble learning model to COVID-19 diagnosis. The integration of deep learning architectures such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), which can detect rich temporal and spatial patterns in data of patient's clinical symptoms, especially when combining imaging or longitudinal data, can be described as one of the most vital directions. The model should as well be trained on larger and more varied datasets that include different demographic, geographic and clinical profiles to increase the generalizability and reduce bias. The usability, diagnostic effect and integration with electronic health record (EHR) systems of the model should be tested in clinical settings to validate the model. In order to become more transparent and trustworthy, the interpretability techniques (SHAP and LIME) can be employed to explain the model forecasts to the medical

staff. Moreover, the adaptive learning functions such as online learning or federated learning will be explored in order to ensure that the model is up-to-date and precise as the new COVID-19 variants and symptom profiles emerge. The approaches will also enable privacy maintaining collaboration among institutions. Finally, the next research could examine how the model can be implemented in a mobile or cloud computing platform to permit prompt and cost-effective screening in remote or resource-constrained settings to improve global public health campaigns and response to pandemic actions.

REFERENCES

[1] Ardabili, S. F., Mosavi, A., Ghamisi, P., Ferdinand, F., Varkonyi-Koczy, A. R., Reuter, U., Rabczuk, T., & Atkinson, P. M. (2020). COVID-19 Outbreak Prediction with Machine Learning. *Algorithms*, *13*(10), 249. https://doi.org/10.3390/a13100249

[2] Arora, N., Banerjee, A. K., & Narasu, M. L. (2020). The role of artificial intelligence in tackling COVID-19. *Future Virology*, *15*(11), 717–724. https://doi.org/10.2217/fvl-2020-0130

[3] Atta-Ur-Rahman, N., Sultan, K., Naseer, I., Majeed, R., Musleh, D., Gollapalli, M. a. S., Chabani, S., Ibrahim, N., Siddiqui, S. Y., & Khan, M. A. (2021). Supervised Machine Learning-Based Prediction of COVID-19. *Computers, Materials & Continua/Computers, Materials & Continua (Print)*, *69*(1), 21–34. https://doi.org/10.32604/cmc.2021.013453

[4] Bodul, D., Jakovac, P., Gaspari, F., Poščić, A., Martinović, A., Zafiroski, J., Stanković, J., Marjanović, I., Drezgić, S., Bogović, N. D., Čegar, S., Stojanović, B., Kostić, Z., Vučić, V., Rendulić, D., Mihanović, D., Troković, R., Sameti, M., Redzepagić, S., . . . Čolig, M. (2021). *Contemporary economic and business issues*. https://dabar.srce.hr/en/islandora/object/efri%3A3002

[5] Gothai, E., Thamilselvan, R., Rajalaxmi, R., Sadana, R., Ragavi, A., & Sakthivel, R. (2023). Prediction of COVID-19 growth and trend using machine learning approach. *Materials Today: Proceedings*, *81*, 597–601. https://doi.org/10.1016/j.matpr.2021.04.051

[6] Malki, Z., Atlam, E. S., Ewis, A., Dagnew, G., Ghoneim, O. A., Mohamed, A. A., Abdel-Daim, M. M., & Gad, I. (2021). The COVID-19 pandemic: prediction study based on machine learning models. *Environmental Science and Pollution Research International*, *28*(30), 40496–40506. https://doi.org/10.1007/s11356-021-13824-7

[7] Painuli, D., Mishra, D., Bhardwaj, S., & Aggarwal, M. (2021). Forecast and prediction of COVID-19 using machine learning. In *Elsevier eBooks* (pp. 381–397). https://doi.org/10.1016/b978-0-12-824536-1.00027-7

[8] Sadik, R., Reza, M. L., Noman, A. A., Mamun, S. A., Kaiser, M. S., & Rahman, M. A. (2020). COVID-19 Pandemic: A Comparative Prediction using Machine Learning. *International Journal of Automation, Artificial Intelligence and Machine Learning*, 01–19. https://doi.org/10.61797/ijaaiml.v1i1.44.

[9] Sujath, R., Chatterjee, J. M., & Hassanien, A. E. (2020). A machine learning forecasting model for COVID-19 pandemic in India. *Stochastic Environmental Research and Risk Assessment*, *34*(7), 959–972. https://doi.org/10.1007/s00477-020-01827-8

[10] Tuli, S., Tuli, S., Tuli, R., & Gill, S. S. (2020). Predicting the growth and trend of COVID-19 pandemic using machine learning and cloud computing. *Internet of Things*, *11*, 100222. https://doi.org/10.1016/j.iot.2020.100222