Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 08, No. 02, June 2022.

# DEVELOPMENT OF ENSEMBLE MACHINE LEARNING MODEL TO IMPROVE COVID-19 OUTBREAK FORECASTING

Meaad Alrehaili and Fatmah Assiri

*(Received: 22-Dec.-2021, Revised: 27-Feb.-2022, Accepted: 25-Mar.-2022)*


## *ABSTRACT*

*The world is currently facing the coronavirus disease 2019 (COVID-19 pandemic). Forecasting the progression of that pandemic is integral to planning the necessary next steps by governments and organizations. Recent studies have examined the factors that may impact COVID-19 forecasting and others have built models for predicting the numbers of active cases, recovered cases and deaths. The aim of this study was to improve the forecasting predictions by developing an ensemble machine-learning model that can be utilized in addition to the Naïve Bayes classifier, which is one of the simplest and fastest probabilistic classifiers. The first ensemble model combined gradient boosting and random forest classifiers and the second combined support vector machine and random-forest classifiers. The numbers of confirmed, recovered and death cases will be predicted for a period of 10 days. The results will be compared to the findings of previous studies. The results showed that the ensemble algorithm that combined gradient boosting and random-forest classifiers achieved the best performance, with 99% accuracy in all cases.*



## *KEYWORDS*

*COVID-19, Coronavirus disease, Coronavirus, Pandemic, Epidemic prediction, Future forecasting, Machine learning, Ensemble machine learning algorithms, Naive Bayes, Support vector machine, Random forest, Gradient boosting.*


## 1. INTRODUCTION

The world is currently in the midst of a critical pandemic, the coronavirus disease 2019 (COVID-19 pandemic), which has spread throughout the world and is expected to continue doing so. By the end of 2019, there had been 7,000 deaths due to COVID-19 in 150 countries, prompting the World Health Organization (WHO) to declare the COVID-19 outbreak a global pandemic in March 2020 [1]. Previous research about pandemics allows for prediction of future pandemic cases or expansions, but the historical data must be reliable to ensure prediction accuracy [2].

Forecasting a pandemic's progression is extremely important for governmental and organizational actions, such as those within the fields of transportation, health-care and supplies. Prediction of the next phases of a pandemic will give the decision-making parties early notice of actions that they should undertake in order to minimize or even avoid catastrophes [3]. Additionally, successful forecasting will help control the situation by assisting the authorities in taking the right action at the right time to contain the crisis, thus preventing major losses. Many studies have been conducted on COVID-19 forecasting [3] and the existence of recently collected datasets and commitments of support from governments, health organizations and social parties enrich the opportunity to use machine-learning models to predict the progression of the pandemic.

A study used supervised machine-learning models to forecast pandemic development; classifying COVID-19 dataset using four classifiers or machine-learning models: linear regression (LR), support vector machine (SVM), exponential smoothing (ES), least absolute shrinkage and selection operator (LASSO) [4]. Then, the models were trained and evaluated using the $R^2$ score, adjusted $R^2$, mean absolute error (MAE), mean square error (MSE) and root mean square error (RMSE). Each classifier was evaluated separately; the results revealed that ES achieved the highest accuracy in forecasting. LR and LASSO also performed well in forecasting the numbers of deaths and confirmed cases. Previous studies trained machine-learning models for COVID-19 forecasting; however, a small dataset was utilized [4]. Thus, there is a pressing need for more accurate models that use larger dataset.

M. Alrehaili and F. Assiri are with College of Computer Science and Engineering, University of Jeddah, Jeddah, Saudi Arabia. Emails: malrehaili0006.stu@uj.edu.sa and fyassiri@uj.edu.sa

The contribution of this paper is to develop a more accurate early-forecast model for COVID-19 using ensemble machine-learning algorithms. Furthermore, we also conducted a comparative analysis based on four measurements ($R^2$ score, adjusted $R^2$, MAE, MSE and RMSE) that compared the performance of the proposed ensemble algorithms with the results of a similar study [4]. We used the same dataset that was used by [4], but we applied the Naïve Bayes classifier, the simplest and fastest probabilistic classifier [5], which can be used for COVID-19 forecasting by requiring several linear parameters for the number of features or predictors as a variable in a learning problem.

In addition, we propose two ensemble models for improving COVID-19 forecasting: the gradient boosting and random-forest (GBRF) ensemble model and the support vector machine and random-forest (SVM+RF) ensemble model. Following the approach used by [4], we made predictions regarding the new confirmed cases, recovered cases and deaths that would occur in the next 10 days. We also used the same evaluation methods to compare the results of the proposed ensemble methods to those of the previous work by computing the $R^2$ -score, adjusted $R^2$, MAE, MSE and RMSE [4]; this allowed to show which model has the highest degree of accuracy in COVID-19 forecasting.

The rest of the paper is organized as follows: Section 2 describes the literature review. The dataset description is given in Section 3 and the methodology is presented in Section 4. The results and discussion are presented in Section 5. The conclusion is presented in Section 6.

## 2. LITERATURE REVIEW

Accurate forecasting serves various crucial clinical purposes, particularly for health-based systems. Computer-aided clinical predictive models have been used in various areas, including for predicting the progression of different diseases. In this study, we applied different prediction models to build a predictive model for COVID-19. In a recent study, the researchers proposed a system for detecting COVID-19 movements and progression by forecasting cases based upon real-time data [6]. SVM was compared with seven other classifiers and achieved the highest accuracy (92.95%). The results of the SVM and k-nearest neighbors' classifiers were the same.

Machine learning has also been used for COVID-19 survival analysis and discharge time likelihood prediction [7]. Several machine-learning algorithms were used for these purposes, including gradient boosting, component-wise gradient boosting and SVM. The results indicated that the gradient boosting survival model is the best model for prediction of patient survival. However, the details of the dataset were not mentioned.

Random forest was also used for COVID-19 patient health prediction [8]. Different algorithms (i.e., decision tree, support vector, Gaussian Naive Bayes and boosted random forest), were used for this purpose and their performances were compared. The boosted random-forest algorithm was the best-performing model (94%) for patient health prediction. COVID-19 management and progression predictions were performed through the use of mathematical modeling and artificial intelligence [5]. The results of the Naive Bayes and other classifiers were compared by computing their respective prediction accuracy levels. The Naive Bayes classifier showed 99.4% accuracy, which was the highest. However, the different classifiers were tested on different datasets, which might have affected their accuracy levels.

Another use of the Naive Bayes classifier was proposed in [9]. A model for obtaining computed tomography images for predicting the progression of COVID-19 was implemented and multiple classifiers were used. The Naive Bayes classifier showed 92.15% accuracy in feature selection and its accuracy was enhanced to 96.07% when another dataset was used. The average Naive Bayes classifier accuracy was 94.11%, similar to that of the convolution neural network (CNN). No advantage of using the Naive Bayes classifier over the CNN classifier or *vice versa* was reported. Several studies have tried to solve the problems related to predicting the movement or progression of the COVID-19 pandemic, but the current models' robustness needs to be improved. In this study, we applied different machine-learning algorithms to address the prediction accuracy limitations of the previous studies.

Convolution Neural Networks (CNNs) have also been used for the diagnosis of COVID-19 based on the classification of chest X-ray images [10]. A total of 178 X-ray images were used, of which 136 images were for COVID-19 patients and the rest of non-infected people. The proposed CNN model was integrated incrementally, starting with a single layer, and adding a convolutional layer at each increment.

The results exhibited 99.5% accuracy. In another study, the Facebook Prophet model predicted the number of future infections over 90 days, taking into account the peak dates of confirmed cases for 6 of the most affected countries in the world [11]. A comparative analysis of the use of machine-learning and soft-computation models in the prediction of COVID-19 was also conducted [12]. The results showed that multi-layered perceptron and adaptive network-based fuzzy inference systems are among the best ones.

To improve prediction accuracy, a hybrid machine-learning model that consists of an adaptive network-based fuzzy inference system (ANFIS) and a multi-layered perceptron-imperialist competitive algorithm was proposed to perform COVID-19 forecasts in Hungary [13]. The model predicted the number of infected cases and mortality rates within 9 days based on time-series data. To further improve the performance of the prediction models, in [14], a grey wolf optimizer and an artificial neural network were applied to the same data used [13] and the results were promising.

Another study applied supervised machine learning to perform sentiment analysis as a decision support tool to better manage the pandemic [15]. The contribution of this work lies in the features set identified by the authors. The results showed that extra tree classifiers performed best compared to other algorithms, with an accuracy of 93%. Most recent work utilized mobile sensors to collect users' vital signs such as temperature and coughing patterns, which was subsequently combined with information entered by the users through mobile applications. Next, all data were used by applying an Artificial Neural Network (ANN) as a symptom-prediction algorithm to predict the likelihood of a user having COVID-19 [16].

## 3. DATASET

For a fair comparison between the proposed ensemble algorithms and those developed by Rustam [4], we used the novel COVID-19 dataset obtained from Johns Hopkins University (JHU), which contains data beginning on January 22, 2020 and is updated daily. The data is sourced from governments, national agencies across the world and the WHO [17], [24].

The number of global confirmed cases when the study was conducted was 10,853, 589 and the number of global deaths was 2,393,707. Figure 1 shows the average number of daily deaths and daily confirmed cases. The aforementioned dataset was accessed from the COVID-19 Data Repository of the Centre for Systems Science and Engineering at JHU. The data features include the state, region, date, number of confirmed cases, death cases and number of recovered cases. To meet the needs of this study, we further pre-processed the dataset.

Below are the descriptions of the features (attributes) that were used in this study [17]:

- Confirmed cases: The counts included the reported confirmed and probable cases.
- Deaths: The counts included the reported confirmed and probable cases.
- Recovered cases: Estimates based on local media reports and on state and local-government reports were considered where available; thus, this may be substantially lower than the true number.

## 4. METHODOLOGY

In this study, we used the Naive Bayes classifier, because it is one of the simplest and fastest classifiers, especially in the training stage [5]. We also utilized two ensemble models, GBRF and SVM+RM. SVM and random-forest machine-learning models complement each other; random forest computes the probability of belongings to a class, while SVM computes the distance to the boundary. Random forest can also complement gradient boosting, which is sensitive to noise and can cause overfitting [20].

Voting is a technique used to combine the results of many classifiers. There are three types of voting: unanimous, majority and plurality voting. In *unanimous voting*, all classifiers agree on a final decision. Majority voting makes the decision based on the number of voters; if one half or more of the votes go for one option, then it gets selected. In *plurality voting*, if most votes go to one option, it is selected as the final decision. In this study, we combine classifiers using majority voting, as it has been the most used one in prior studies.

The *Naive Bayes classifier* assumes that the classes' features are not related to each other, and it is not

affected by the classification assumption. It only requires a small training dataset to estimate the means and variances needed for classification [21]. Gradient boosting is a machine-learning model that generates a forecasting model in the form of an ensemble of weak-prediction models to increase the prediction performance. Gradient boosting is used for regression and classification problems [7].

*Random forest* is a common machine-learning method for developing prediction models in many research settings. To minimize the burden of data collection and to improve its efficiency, the random-forest model can be used as a prediction model to decrease the number of variables required to achieve a prediction [23]. Equation (1) presents RF regression model:

$$h(x) = \frac{1}{P}\sum_{n}^{p=1} h(y, \lambda P) \quad (1)$$

*SVM* is a statistical classifier that is used for linear and non-linear pattern classification [22]. The data is converted into high-dimension representations *via* non-linear mapping and SVM searches the new representations for the most appropriate data classification. SVM classifies data by increasing the margins of the classes; at the same time, it decreases the classification errors [21].

To predict the total number of people that might be affected in terms of new confirmed COVID-19 cases, deaths and expected recoveries for the upcoming 10 days, the Naive Bayes classifier and the ensemble models were trained using a dataset spanning from January to March 2020 [24]. The size of the training dataset was 66 days and that of the testing dataset was 10 days, following the approaches used in [4]. We evaluated the performances of the learning models in terms of $R^2$ score, adjusted $R^2$, MAE, MSE and RMSE, which are commonly used in the evaluation of predictive problems.

The $R^2$ score is only useful for simple linear regression. When using multiple linear regression, the value of the $R^2$ score grows as the number of independent variables increases, even if the independent variable is small. Adjusted $R^2$, on the other hand, increases only when the independent variable is significant and impacts the dependent variable. Equations (2) and (3) present the evaluation of $R^2$ and adjusted $R^2$, respectively:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \quad (2)$$

$$Adjusted\ R^2 = 1 - \frac{(1-R^2)(N-1)}{N-p-1} \quad (3)$$

Mean Absolute Error (MAE) measures the differences between target values and predicted values, as shown in Equation (4):

$$MAE = \frac{\sum_{i=1}^{n}|X_i - \hat{X}_i|}{n} \quad (4)$$

Mean Square Error (MSE) takes the square of the differences between the actual and predicted values. This removes any negative values, as shown in Equation (5):

$$MSE = \frac{\sum_{i=1}^{n}(X_i - \hat{X}_i)^2}{n} \quad (5)$$

Root Mean Square Error (RMSE): detects the error rate from the regression model and compares the error size against the size of the target value. Equation (6) presents the evaluation of RMSE:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \hat{X}_i)^2}{n}} \quad (6)$$

Figure 2 shows the study's overall methodology.

## 5. Results and Discussion

This evaluation study was designed to answer the following research questions:

- RQ1 (accuracy of the Naive Bayes classifier): What is the degree of accuracy of the Naive Bayes classifier?
- RQ2 (accuracy of the ensemble classifiers): What are the degrees of accuracy of the GBRF and SVM+RF ensemble classifiers?

This paper aims to build a predictive model using machine-learning algorithms for potential prediction of COVID-19 cases. The analysis provides details on regular estimates for the total number of confirmed, recovered and death cases around the world. The total of death and confirmed cases increased

daily; this is obviously worrying. The following sub-sections discuss the results of the proposed models in terms of new infected, recovered and death cases.

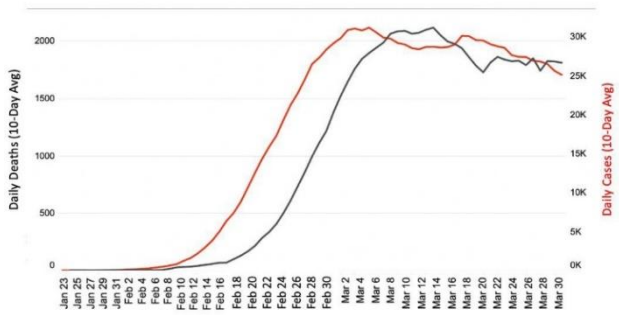


Figure 1. Daily deaths and daily confirmed cases (10-day average).

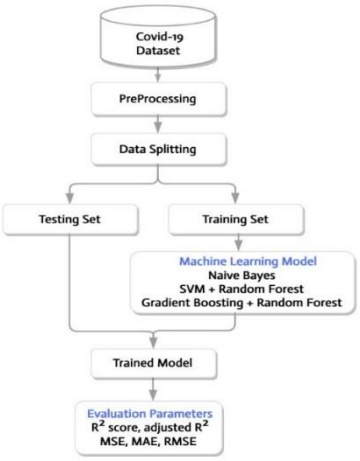


Figure 2. Proposed methodology workflow.

## 5.1 Future Forecasting of New Infections

The results of the Naive Bayes classifier for the number of cases of COVID-19 showed that the predicted number of cases was lower than the actual number of cases. As the attempted prediction period grew, the gap between the predicted and actual values increased. Figure 3 shows the Naive Bayes classifier predictions. The results of the SVM+RF ensemble model that was used to predict the number of new confirmed cases of COVID-19 showed that the predicted number of cases did not match the number of actual cases. The gap between the predicted values and the actual values increased with the number of upcoming days. Figure 4 shows the SVM+RF ensemble model predictions.

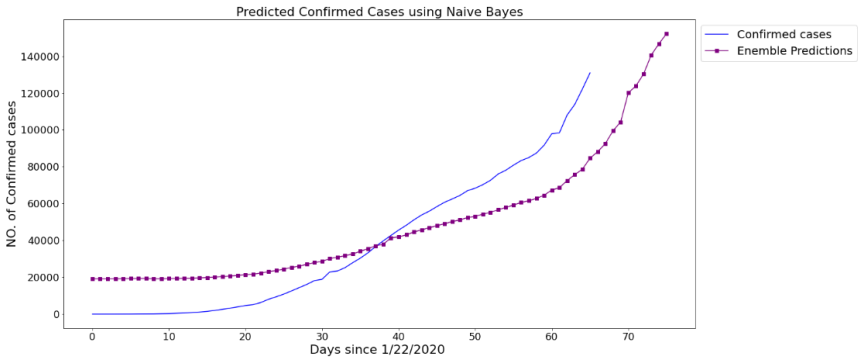


Figure 3. New infected cases for the upcoming 10 days using Naive Bayes.

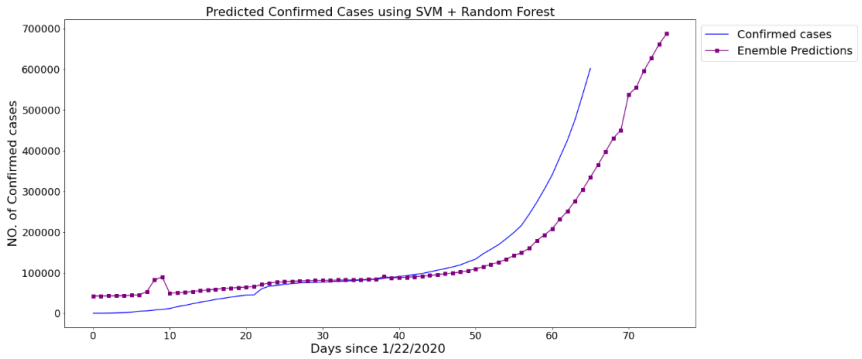


Figure 4. New infected cases for the upcoming 10 days using SVM and random forest.

The results of the GBRF ensemble model, which was used to predict the number of confirmed cases, showed that the predicted values matched the number of actual cases. Figure 5 shows the GBRF ensemble model predictions. Table 1 shows the results of the Naive Bayes classifier and the two

proposed ensemble models. GBRF gives the best results when predicting the newly infected cases for the upcoming 10 days. In contrast, Naïve Bayes and SVM+RF performed poorly.

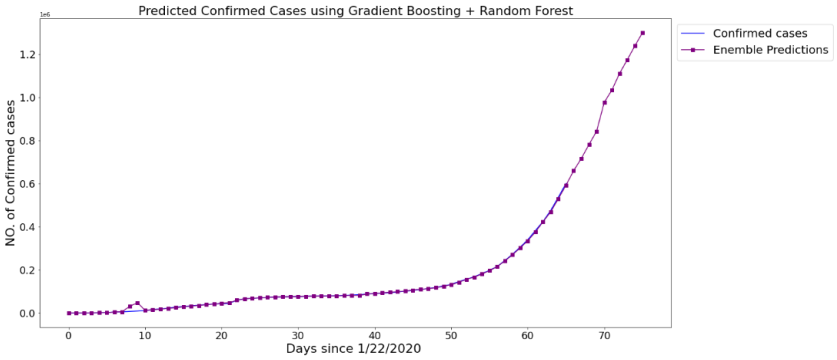


Figure 5. New infected cases for the upcoming 10 days using gradient boosting and random forest.

Table 1. Model performance of future forecasting for newly infected cases.

| **Models** | **$R^2$ score** | **Adjusted $R^2$** | **MSE** | **MAE** | **RMSE** |
|---|---|---|---|---|---|
| Naive Bayes | 0.71 | 0.70 | 4628031693.91 | 39781.21 | 68029.64 |
| SVM+RF | 0.68 | 0.67 | 34836665919.42 | 98018.14 | 186645.83 |
| GBRF | 0.99 | 0.99 | 69138158.83 | 4092.88 | 8314.94 |

## 5.2 Future Forecasting of Recovered Cases

The results of the Naive Bayes classifier for recovered cases showed that the predicted number of recovered cases was lower than the actual number. With an increase in the number of upcoming days, the gap between the predicted and the actual values increased. Figure 6 shows the Naive Bayes classifier predictions.

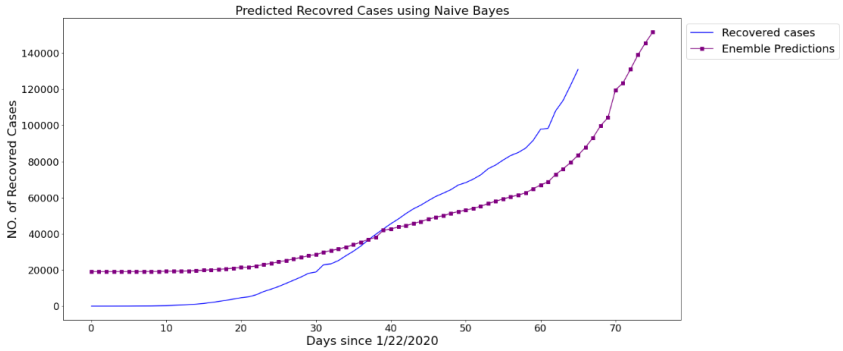


Figure 6. Recovery rate prediction for the upcoming 10 days using Naive Bayes.

The results of the SVM+RF ensemble model prediction of the number of recovered cases did not match the actual number of recovered cases. The predicted values were less than the actual cases and after five more days, the predicted number of cases became more than the number of actual cases. Figure 7 shows the SVM+RF ensemble model predictions.

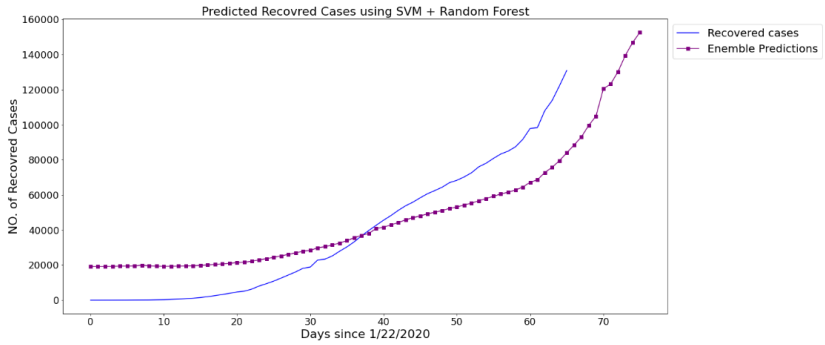


Figure 7. Recovery rate prediction for the upcoming 10 days using support vector machine and random forest.

The results of the GBRF ensemble model prediction for the number of recovered cases did not match the actual number of cases. Figure 8 shows the GBRF ensemble model predictions. The performance results of the models when predicting the number of recovered cases is shown in Table 2. GBRF gave the best results when predicting the recovered cases for the upcoming 10 days, while Naive Bayes and SVM+RF gave approximately the same results.

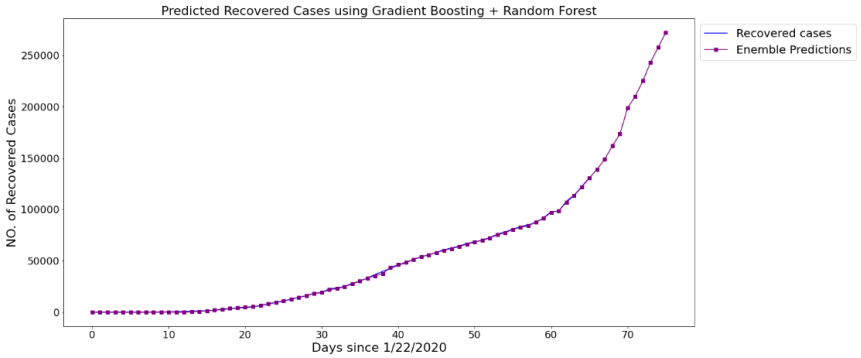


Figure 8. Recovery rate prediction for the upcoming 10 days using gradient boosting and random forest.

Table 2. Model performance for future forecasting for recovered cases.

| **Models** | **$R^2$ score** | **Adjusted $R^2$** | **MSE** | **MAE** | **RMSE** |
|---|---|---|---|---|---|
| Naive Bayes | 0.72 | 0.73 | 4628031693.91 | 41781.21 | 68029.64 |
| SVM+RF | 0.71 | 0.70 | 1335741927.85 | 25459.42 | 36547.8 |
| GBRF | 0.99 | 0.99 | 1682739.6 | 711.0 | 1297.2 |

## 5.3 Future Forecasting of the COVID-19 Death Rate

The results of the Naive Bayes classifier prediction of the number of COVID-19 deaths showed that the number of predicted cases was lower than the actual number. With an increase in the number of upcoming days, the gap between the predicted and actual values increased. Figure 9 shows the Naive Bayes classifier predictions. The SVM+RF ensemble model predictions for the number of deaths did not match the actual number of deaths. Figure 10 shows the SVM+RF ensemble model predictions.

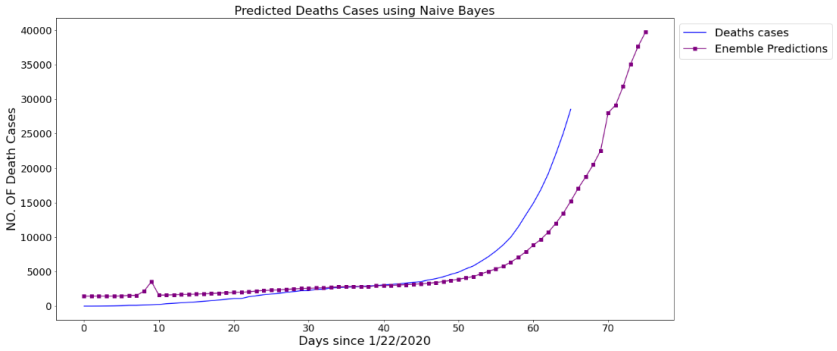


Figure 9. Death prediction for the upcoming 10 days using Naive Bayes.

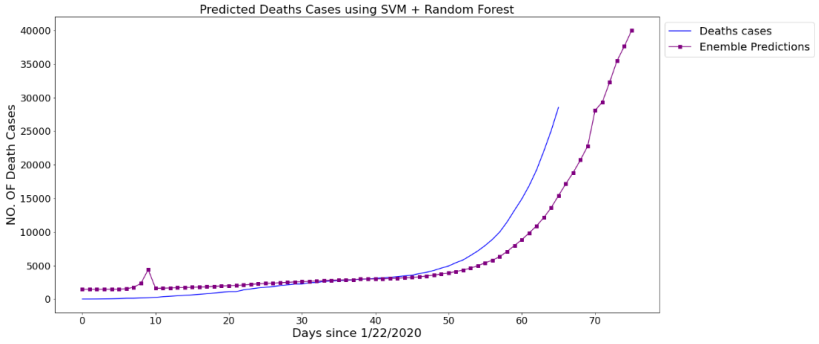


Figure 10. Death prediction for the upcoming 10 days using support vector machine and random forest.

The results of the GBRF predictions of the number of deaths showed that the predicted value was almost the same as the number of actual deaths. A gap appeared only in the eighth and ninth days, as shown in Figure 11 The death rate prediction performance results are shown in Table 3. GBRF gave the best results when predicting death cases for the upcoming 10 days. In contrast, Naive Bayes and SVM+RF performed poorly.

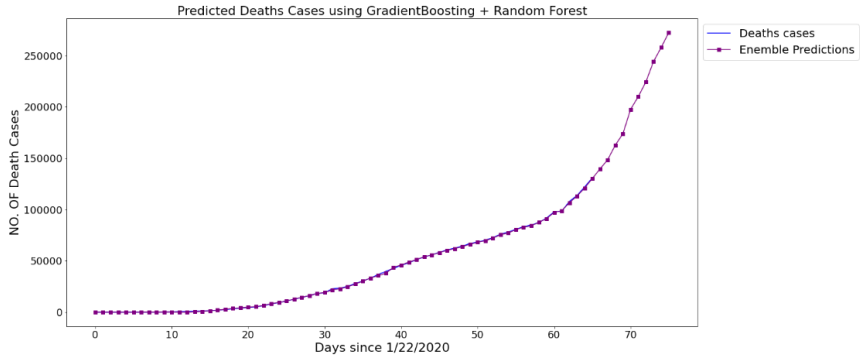


Figure 11. Death prediction for the upcoming 10 days using gradient boosting and random forest.

Table 3. Model performance of future forecasting for death cases.

| **Models** | **$R^2$ score** | **Adjusted $R^2$** | **MSE** | **MAE** | **RMSE** |
|---|---|---|---|---|---|
| Naive Bayes | 0.72 | 0.69 | 10833945.26 | 41181.5 | 68029.62 |
| SVM+RF | 0.69 | 0.68 | 106804953.79 | 5032.5 | 10334.65 |
| GBRF | 0.99 | 0.99 | 200297.36 | 215.2 | 447.55 |

To summarize our findings, the GBRF ensemble model reached the highest accuracy (99%) in all the cases, performing better than the Naive Bayes classifier and the SVM+RF ensemble model.

## 5.4 Model Performance within 10-Day Prediction Intervals

We compared the results of the applied models (i.e., Naive Bayes classifier, SVM+RF ensemble model and GBRF ensemble model) with the results of other models (i.e., linear regression (LR), support vector machine (SVM), exponential smoothing (ES), least absolute shrinkage and selection operator (LASSO)) tested in [4]. Table 4 shows the comparison results between the applied ensemble machine-learning models and other models tested in [4]. We then compared the two best models, GBRF and ES in three aspects:

- Confirmed-cases Prediction: GBRF gave higher $R^2$ score and higher adjusted $R^2$ values, which means that GBPR is better than SE at predicting the number of newly confirmed cases. Also, the MSE, RMSE and MAE results for GBRF were less than those of ES.
- Recovered-cases Prediction: The $R^2$ for GBRF was better than that for ES. The adjusted $R^2$ value was the same for GBRF and ES. MSE, RMSE and MAE values for GBRF were less than those for ES.
- Number-of-deaths Prediction: The results were similar to the results of the confirmed-cases prediction. The results of GBRF were better than those of ES in all evaluation measurements used in this work.

Table 4. Comparison between the proposed ensemble machine-learning models and the models for future forecasting in [4].

| **Model** | **Evaluation** | **Confirmed** | **Recovered** | **Death** |
|---|---|---|---|---|
| Naïve Bayes | $R^2$ score | 0.71 | 0.72 | 0.72 |
| | Adjusted $R^2$ | 0.70 | 0.73 | 0.696 |
| | MSE | 4628031.7 k | 4628031.7 k | 10834 k |
| | MAE | 39781.21 | 41781.21 | 1900.88 |
| | RMSE | 68029.64 | 68029.64 | 3291.5 |

| | | | | |
|---|---|---|---|---|
| SVM+RF | $R^2$ score | 0.68 | 0.71 | 0.69 |
| | Adjusted $R^2$ | 0.67 | 0.70 | 0.68 |
| | MSE | 34836666 k | 1335742 k | 106805 k |
| | MAE | 98018.14 | 25459.42 | 5032.5 |
| | RMSE | 186645.83 | 36547.8 | 10334.65 |
| GBRF | $R^2$ score | 0.999 | 0.999 | 0.999 |
| | Adjusted $R^2$ | 0.99 | 0.99 | 0.99 |
| | MSE | 69138 k | 1683 k | 200 k |
| | MAE | 4092.88 | 711.0 | 215.2 |
| | RMSE | 8314.94 | 1297.2 | 447.55 |
| ES | $R^2$ score | 0.98 | 0.99 | 0.98 |
| | Adjusted $R^2$ | 0.97 | 0.99963 | 0.97 |
| | MSE | 0.67 | 34836665919.42 | 98018.14 |
| | MAE | 8867.43 | 1827.85 | 406.08 |
| | RMSE | 16828.58 | 2243.48 | 813.77 |
| LR | $R^2$ score | 0.83 | 0.39 | 0.96 |
| | Adjusted $R^2$ | 0.79 | 0.21 | 0.95 |
| | MSE | 1472986 k | 480922 K | 840240.11 |
| | MAE | 30279.55 | 30705.27 | 723.11 |
| | RMSE | 38390.51 | 21929.95 | 916.64 |
| LASSO | $R^2$ score | 0.98 | 0.29 | 0.85 |
| | Adjusted $R^2$ | 0.97 | 0.08 | 0.81 |
| | MSE | 234489 k | 1462144 k | 3244066.79 k |
| | MAE | 11693.97 | 30705.27 | 1430.29 |
| | RMSE | 15322.11 | 38237.99 | 1801.12 |
| SVM | $R^2$ score | 0.59 | 0.24 | 0.53 |
| | Adjusted $R^2$ | 0.02 | 0.99963 | 0.39 |
| | MSE | 5760890 k | 13121148 k | 160162 k |
| | MAE | 60177.9 | 106739.82 | 3129.74 |
| | RMSE | 75911.28 | 114547.58 | 4002.02 |

To summarize our findings, GBRF performed best in the current forecasting domain given the nature and size of the dataset, followed by ES. LR and LASSO performed fairly well in forecasting death rates and newly confirmed cases. The Naive Bayes classifier and the SVM+RF ensemble model showed approximately the same degree of accuracy. SVM produced poor results in all the scenarios.

## 6. CONCLUSION

Forecasting the movement and progression of a pandemic facilitates governmental or organizational actions needed to contain that pandemic. During the current COVID-19 pandemic, forecasting is essential to prevent high numbers of active cases and deaths. Machine learning-based prediction models have been proposed for predicting the risk of COVID-19 outbreak. In this study, we used the Naive Bayes classifier and two ensemble algorithms: GBRF and SVM+RF. We compared these models with the prediction models tested in [4].

Our results showed that the best performance when forecasting new infections, recovered cases and deaths was achieved by the GBRF ensemble model (99%). This is an improved performance over ES, which reached 98% in [4]. The Naive Bayes classifier and SVM+RF ensemble model showed approximately the same performance, reaching 71% and 68%, respectively. SVM performed poorly in all scenarios. These results could be due to the nature of the significance improvement of ensemble models, in which two or more algorithms complement each other to provide better results [25]. Deep-learning algorithms have been known for their advantages over machine-learning models and we will consider them in our future work.

## ACKNOWLEDGMENT

The authors would like to thank Dr. Kaouther Laabidi, Professor at the Department of Computer and Network Engineering, College of Computer Science and Engineering, University of Jeddah, Saudi Arabia and University of Tunis El Manar, Tunis, for her valuable discussions.

## REFERENCES

[1] Centers for Disease Control and Prevention, "Severe Outcomes among Patients with Coronavirus Disease 2019 (COVID-19) : United States," MMWR Morb. Mortal. Wkly. Rep., vol. 69, no. 12, pp. 343-346, February 12–March 16, 2020.

[2] I. M. Hall et al., "Real-time Epidemic Forecasting for Pandemic Influenza," Epidemiology and Infection, vol. 135, no. 3, pp. 372-385, 2007.

[3] K. Sarkar, S. Khajanchi and J. Nieto, "Modeling and Forecasting the COVID-19 Pandemic in India," Chaos, Solitons & Fractals, vol. 139, DOI : 10.1016/j.chaos.2020.110049, 2020.

[4] F. Rustam, A. Reshi, A. Mehmood, S. Ullah, B. Won On et al., "COVID-19 Future Forecasting Using Supervized Machine Learning Models," IEEE Access, vol. 8, pp. 101489–101499, 2020.

[5] Y. Mohamadou, A. Halidou and P. T. Kapen, "A Review of Mathematical Modeling, Artificial Intelligence and Datasets Used in the Study, Prediction and Management of COVID-19," Applied Intelligence, vol. 50, pp. 3913–3925, 2020.

[6] M. Otooma, N. Otoumb, M. A. Alzubaidi, Y. Etoom and R. Banihani, "An IoT-based Framework for Early Identification and Monitoring of COVID-19 Cases," Biomedical Signal Processing and Control, vol. 62, DOI : 10.1016/j.bspc.2020.102149, 2020.

[7] M. Nemati, J. Ansary and N. Nemati, "Machine-learning Approaches in COVID-19 Survival Analysis and Discharge-time Likelihood Prediction Using Clinical Data," Patterns, vol. 1, no. 5, 2020.

[8] C. Iwendi, A. K. Bashir, A. Peshkar et al., "COVID-19 Patient Health Prediction Using Boosted Random Forest Algorithm," Frontiers in Public Health, vol. 357, DOI : 10.3389/fpubh.2020.00357, 2020.

[9] A. Farid, G. Selim and H. Khater, "Novel Approach of CT Images Feature Analysis and Prediction to Screen for Corona Virus Disease (COVID-19)," International Journal of Scientific and Engineering Research, vol. 11, no. 3, DOI:10.14299/ijser.2020.03.02, 2020.

[10] A. A. Reshi et al., "An Efficient CNN Model for COVID-19 Disease Detection Based on X-ray Image Classification," Complexity, vol. 2021, DOI: 10.1155/2021/6621607, 2021.

[11] S. Dash et al., "Intelligent Computing on Time-series Data Analysis and Prediction of COVID-19 Pandemic," Pattern Recognition Letters, vol. 151, pp. 69-75, 2021.

[12] S. F. Ardabili et al., "COVID-19 Outbreak Prediction with Machine Learning," Algorithms, vol. 13, no. 10, DOI: 10.3390/a13100249, 2020.

[13] G. Pinter et al., "COVID-19 Pandemic Prediction for Hungary; A Hybrid Machine Learning Approach," Mathematics, vol.8, no.6, DOI: 10.3390/math8060890, 2020.

[14] S. Ardabili et al., "Coronavirus Disease (COVID-19) Global Prediction Using Hybrid Artificial Intelligence Method of ANN Trained with Grey Wolf Optimizer," Proc. of the 3rd IEEE International Conference and Workshop in Óbuda on Electrical and Power Engineering (CANDO-EPE), DOI: 10.1109/CANDO-EPE51100.2020.9337757, Budapest, Hungary, 2020.

[15] F. Rustam et al., "A Performance Comparison of Supervized Machine Learning Models for COVID-19 Tweets Sentiment Analysis," Plos One, vol. 16, no. 2, DOI: 10.1371/journal.pone.0245909, 2021.

[16] H. Khaloufi et al., "Deep Learning Based Early Detection Framework for Preliminary Diagnosis of COVID-19 *via* Onboard Smartphone Sensors," Sensors, vol. 21, no. 20, DOI: 10.3390/s21206853, 2021.

[17] J. H. C. R. Center, Johns Hopkins Coronavirus Resource Center, [Online], Available: https://coronavirus.jhu.edu/.

[18] The COVID Tracking Project, [Online], Available: https://covidtracking.com/data.

[19] U. N. Dulhare, "Prediction System for Heart Disease Using Naive Bayes and Particle Swarm Optimization," Biomedical Research, vol. 29, pp. 2646–2649, DOI:10.4066/biomedicalresearch.29-18-620, 2018.

[20] W. Zhang, C. Wu, H. Zhong, Y. Li and L. Wang, "Prediction of Undrained Shear Strength Using Extreme Gradient Boosting and Random Forest Based on Bayesian Optimization," Geoscience Frontiers, vol. 12, no. 1, pp. 469–477, 2021.

[21] V. Mohan, "Liver Disease Prediction Using SVM and Naive Bayes Algorithms," International Journal of Science, Engineering and Technology Research (IJSETR), vol. 4, no. 4, pp. 816–820, 2015.

[22] M. Loey, G. Manogaran, M. Taha and N. Khalifa, "A Hybrid Deep Transfer Learning Model with Machine Learning Methods for Face Mask Detection in the Era of the COVID-19 Pandemic," Measurement, vol. 167, DOI: 10.1016/j.measurement.2020.108288, 2020.

[23] J. L. Speiser, M. E. Miller, J. Tooze and E. Ip, "A Comparison of Random Forest Variable Selection Methods for Classification Prediction Modeling," Expert Systems with Applications, vol. 134, pp. 93–101, DOI: 10.1016/j.eswa.2019.05.028, 2019.
[24] J. Hopkins, Johns Hopkins University Data Repository, [Online], Available : https://github.com/CSSEGISandData.
[25] N. Singhal et al., "Comparative Study of Machine Learning and Deep Learning Algorithm for Face Recognition," Jordanian Journal of Computers and Information Technology (JJCIT), vol. 7, no. 3, pp. 313-325, Sep. 2021.

**ملخص البحث:**

تهدف هذه الدراسة الى تحسين توقُّع انتشار جائحة فيروس كورونا عن طريق تطوير نموذج موحَّد من نماذج تعلُّم الآلة من الممكن الاستفادة منه بإضافته الى مصنِّف من نوع (Naïve Bayes) الذي يُعدّ واحداً من أبسط المصنِّفات الاحتمالية وأسرعها. النّموذج الأول يجمع بين تعزيز الميْل و مصنِّفات الغابة العشوائية (RF) بينما يجمع الثاني بين آلة متّجهات الدّعم (SVM) ومصنِّفات الغابة العشوائية. وسيتم توقُّع حالات الإصابة المؤكّدة، وحالات الشّفاء، وحالات الوفاة بسبب الجائحة لمدّة عشرة أيام. كذلك ستجري مقارنة نتائج هذه الدّراسة مع نتائج دراسات سابقة.

وبيّنت النتائج أنّ النّموذج الذي يجمع بين تعزيز الميْل ومصنِّفات الغابة العشوائية حقّق الأداء الأفضل بدقّة بلغت 99% في جميع الحالات.