

## Research Article

# Decision Tree Ensembles to Predict Coronavirus Disease 2019 Infection: A Comparative Study

**Amir Ahmad** ,<sup>1</sup> **Ourooj Safi** ,<sup>2</sup> **Sharaf Malebary** ,<sup>3</sup> **Sami Alesawi** ,<sup>4</sup> and **Entisar Alkayal** <sup>3</sup>

<sup>1</sup>*College of Information Technology, United Arab Emirates University, Al-Ain, Abu Dhabi, UAE*

<sup>2</sup>*Independent Researcher, Al-Ain, UAE*

<sup>3</sup>*Department of Information Technology, King Abdulaziz University, Rabigh 21911, Saudi Arabia*

<sup>4</sup>*Department of Computer Science, King Abdulaziz University, P. O. Box 344, Rabigh 21911, Saudi Arabia*

Correspondence should be addressed to Amir Ahmad; amirahmad@uaeu.ac.ae

Received 22 January 2021; Revised 25 March 2021; Accepted 24 April 2021; Published 8 May 2021

Academic Editor: Atila Bueno

Copyright © 2021 Amir Ahmad et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The coronavirus disease 2019 (Covid-19) pandemic has affected most countries of the world. The detection of Covid-19 positive cases is an important step to fight the pandemic and save human lives. The polymerase chain reaction test is the most used method to detect Covid-19 positive cases. Various molecular methods and serological methods have also been explored to detect Covid-19 positive cases. Machine learning algorithms have been applied to various kinds of datasets to predict Covid-19 positive cases. The machine learning algorithms were applied on a Covid-19 dataset based on commonly taken laboratory tests to predict Covid-19 positive cases. These types of datasets are easy to collect. The paper investigates the application of decision tree ensembles which are accurate and robust to the selection of parameters. As there is an imbalance between the number of positive cases and the number of negative cases, decision tree ensembles developed for imbalanced datasets are applied. F-measure, precision, recall, area under the precision-recall curve, and area under the receiver operating characteristic curve are used to compare different decision tree ensembles. Different performance measures suggest that decision tree ensembles developed for imbalanced datasets perform better. Results also suggest that including age as a variable can improve the performance of various ensembles of decision trees.

## 1. Introduction

SAR-CoV-2 is a new strain of coronavirus causing Covid-19, which was first identified in Wuhan city, Hubei province of China (WHO) [1, 2]. It was declared a pandemic by the World Health Organization (WHO) [1, 2]. On 24 November 2020, there have been more than 59.2 confirmed cases of Covid-19 and more than 1.3 million deaths [3], reported to WHO.

Covid-19 is a respiratory infection and in severe cases may have Covid-19 pneumonia [4]. The reverse transcript polymerase chain reaction (RT-PCR) is considered as the gold standard for Covid-19 diagnosis [5]. It requires at least 24 hours to produce a result. Clinicians use chest imaging tests to diagnose Covid-19 disease when awaiting RT-PCR test results.

Machine learning algorithms are very effective for prediction [6]. Chest X-ray images and computed

tomography (CT) scans have been used to train machine learning models [7–9]. Trained models are then used to predict the Covid-19 positive cases. Using Point-of-care ultrasound, machine learning methods can accurately predict which Covid-19 patients are at a greater risk of death on initial lung scans [10]. Machine learning methods can also help distinguish Covid-19 infection from the community-acquired pulmonary infection [11].

Various studies have been carried out to investigate whether Covid-19 positive cases can be predicted using commonly taken laboratory tests such as hematocrit, hemoglobin, platelets, red blood cells, lymphocytes, and leukocytes [12–14].

Some of these datasets are imbalanced datasets [12, 14]; in other words, the number of positive cases is very small as compared to the number of negative cases. There are various

machine learning methods which have been specifically developed for these datasets [15, 16]. The performance measures for these datasets should be carefully selected as, in some datasets such as medical datasets, the penalty for false negative is more than the penalty for false positive [17].

An ensemble is a combination of many classifiers [18]. An ensemble performs better than an individual classifier if member classifiers are accurate and diverse. Classifier ensembles have been used in many applications [19, 20]. A decision tree is a popular machine learning algorithm [6]. Random forest [21], eXtreme Gradient Boosting (XGBoost) [22], Bagged decision trees [23], and so forth are examples of decision tree ensembles. Ensembles of decision trees are accurate and quite robust to the selection of parameters. Decision tree ensembles have also been developed for imbalanced datasets [24, 25]. Various Covid-19 datasets based on laboratory tests [12–14] are imbalanced datasets. However, they have not been studied using machine learning algorithms developed for imbalanced datasets.

The paper will investigate the application of decision tree ensembles for Covid-19 dataset that is based on commonly taken laboratory tests. The dataset is imbalanced as the ratio of positive class to negative class data points is 1 to 6.5. Therefore, decision tree ensembles that are specifically developed for imbalanced datasets are also applied. Undersampling and oversampling techniques [15, 16] are used to address the data imbalance problem. As the data is imbalanced, classification accuracy is not an appropriate performance measure to compare different classifiers. F-measure, precision, recall, area under the precision-recall curve, and area under the receiver operating characteristic curve [17] performance measures are employed to compare the performances of different decision tree ensembles.

The paper is organized in the following way. Section 2 discusses the research works that apply machine learning algorithms on various kinds of datasets to predict the Covid-19 positive cases. Information on the Covid-19 dataset, decision tree ensembles, and performance measures used in the experiments is presented in Section 3. Section 4 has experiments and discussion. The paper ends with a conclusion and future work section.

## 2. Related Work

Machine learning algorithms have been applied on various types of datasets such as X-ray images, CT scans, and point-of-care ultrasound to predict the Covid-19 positive cases [7, 10]. However, in this section, we concentrate on the research works related to the application of machine learning algorithms to predict Covid-19 positive cases using only routinely collected laboratory findings of the patients [12–14].

Batista et al. [13] collected data from 235 adult patients from the Hospital Israelita Albert Einstein in São Paulo, Brazil, from 17 to 30 March 2020. 15 variables which include age and gender other than 13 laboratory tests such as hemoglobin, platelets, and red blood cells are used to create a prediction model. Five machine learning algorithms (neural networks, random forests, XGBoost, logistic regression, and

support vector machines (SVM)) were tested on this dataset. The best predictive performance was obtained by the SVM algorithm. Schwab et al. [14] used anonymised data from a cohort of 5644 patients seen at the Hospital Israelita Albert Einstein in São Paulo, Brazil, in the early months of 2020. In the dataset, the rate of positive patients was around 10%. They used 97 routine clinical, laboratory, and demographic measurements as the features. They applied the same five different machine learning algorithms as applied by Batista et al. In their experiments. XGBoost performed the best. They also did the experiments for predicting hospitalisation and ICU admission. Random forests performed best for predicting hospital admissions for Covid-19 positive patients, whereas SVM performed best for predicting ICU admission for Covid-19 positive patients. Alakus and Turkoglu [12] modified the data used by Alakus and Turkoglu [12]. The new dataset has 18 laboratory findings of 600 patients. In the dataset, the rate of positive patients was around 13%. They carried out a comparative study of deep learning approaches. Six different model types—Artificial Neural Network (ANN), Convolutional Neural Networks (CNN), Long-Short Term Memory (LSTM), Recurrent Neural Networks (RNN), CNNLSTM, and CNNRNN—were applied to this dataset. LSTM performed best for the 10-fold cross-validation approach. The datasets used by Schwab et al. [14] and Alakus and Turkoglu [12] are imbalanced. Therefore, it will be interesting to apply algorithms that have been developed for imbalanced datasets.

## 3. Data and Classifiers Used in the Experiments

The section will discuss the Covid-19 dataset, decision tree ensembles, and performance measures which are used in our experiments.

**3.1. Data.** 111 laboratory findings of 5644 patients seen at the Hospital Israelita Albert Einstein in São Paulo, Brazil, were collected to detect Covid-19 in the early months of 2020 [14]. Alakus and Turkoglu [12] selected the 18 lab findings that play the most important role in Covid-19. They removed those data points that had missing values for these 18 lab findings. The final data has 600 data points: 520 data points are negative and 80 data points are Covid-19 patients. The ratio of negative data (majority class) to positive data (minority class) is 6.5 to 1 which makes the data imbalanced data. This dataset is used in the experiments.

**3.2. Decision Tree Ensembles.** A decision tree has been a very successful classifier that has been applied in many domains [6]. Decision trees are built using a recursive partition process in which data points are split at each node by using the selected split criterion. A path from the root node to a leaf is a rule which is used for the prediction. An ensemble of classifiers consists of many classifiers [18]. The final decision is the combination of all the member classifiers. An ensemble generally performs better than individual members if individual members are accurate and diverse. Decision tree ensembles are quite robust to the selection of parameters and

perform well. In the experiments, many decision tree-based ensembles are used. As the data is imbalanced data, decision tree-based ensembles that are developed for imbalanced datasets are also used. We will discuss these ensembles and their implementation.

**3.2.1. C4.5 Decision Tree.** Various split criteria have been proposed for decision trees. C4.5 uses the information gain ratio split criterion which reduces a bias towards multi-valued attributes [26]. For all the ensembles, we use C4.5 or its version as the base classifier.

**3.2.2. Decision Tree Ensembles.** Different kinds of ensembles methods have been proposed. Some of them are general methods such as bagging [23] and AdaBoost [27] which can be used with any classifier, whereas some of them are specific to a decision tree such as random forests [21]. XGBoost [22] is a scalable gradient tree boosting ensemble method, which has produced excellent results in many domains.

**3.2.3. Decision Tree Ensembles for Imbalanced Datasets.** Many approaches have been proposed to handle imbalanced datasets. Undersampling of the majority class and oversampling of the minority class are two important approaches to reduce the imbalance of the datasets [15, 16]. Random undersampling (RUS) [15, 16] selects some data points from the majority class and combines them with the minority class to reduce the imbalance. Synthetic Minority Over-sampling Technique (SMOTE) [28] generates synthetic minority data points which are combined with the original dataset to reduce the imbalance. SMOTEBagging [29] combines SMOTE algorithm and the boosting procedure, whereas SMOTEBagging [30] is a combination of SMOTE and Bagging algorithm. RUSBoost [31] combines data undersampling and boosting. RUSBagging [32] is the combination of random undersampling and bagging. Balanced random forests [25] use undersampling of the majority class to create balanced data for each tree of random forests.

Different packages such as Weka [33], imblearn [32], ebmc [34, 35], and XGBoost [36] are used in the experiments. Table 1 shows the classifiers and the related packages used in the experiments.

The default values of parameters were used for all the classifiers. However, the sizes of ensembles were fixed to 50. A 10-fold cross-validation procedure was used in the experiments. The average results are presented in different tables (Tables 2–8).

**3.3. Performance Measures.** There are different measures to compute the performance of the classifiers. Accuracy is one of the most commonly used performance measures. However, for imbalanced datasets, accuracy measure is not very useful [17]. Accuracy, precision, recall, and F1-measure are used in the experiments [17]. We discuss these measures in detail in the appendix. The area under the receiver operating characteristic curve (AUROC) and the area under the

precision-recall curve (AUPRC) are also used to compare the performance of decision tree ensembles [17].

- (i) AUROC: The ROC is the curve between TPR and FPR at different decision threshold settings. The area under the curve is used as a performance measure. AUROC values range from 0 to 1. The baseline (random classifier) is 0.5 [37].
- (ii) AUPRC: The precision-recall curve is the curve between precision and recall at different decision threshold settings. The area under the curve is used as a performance measure. AUPRC takes the values from 0 to 1. The baseline is not constant for AUPRC. It depends on the ratio of positive and negative samples. It is equal to positive/(positive + negative) [37]. For the dataset, which was used in the experiment, it is equal to  $80/600 = 0.13$ . It has been demonstrated [37] that, for imbalanced datasets, AUPRC is a better performance measure as compared to AUROC.

## 4. Results and Discussion

Different experiments were carried out to study the performance of different decision tree ensembles. This section will present the results of those experiments.

**4.1. The Comparative Study of Various Decision Tree Ensembles.** Different types of decision tree ensembles are compared using various performance measures. Results are presented in Table 2. For two (accuracy and precision) out of six performance measures, the standard classifiers performed best, whereas for the other four performance measures (F1-measure, recall, AU-ROC, and AUPRC) decision tree ensembles for imbalanced datasets perform best. Random forests performed best for two performance measures (accuracy and precision). Balanced random forest (RUS) performed best for three performance measures: recall, F1-measures, and AUPRC. RUSBagging performed best for AUROC. AUROC and AUPRC are widely used performance measures for imbalanced datasets. RUSBagging performed best for AUROC with the value 0.881, whereas balanced random forest (RUS) gave the best AUPRC with the value 0.561. The study demonstrates that decision tree ensembles for imbalanced datasets perform better for this Covid-19 dataset.

**4.2. Standard Decision Tree Ensembles with Different Sampling Techniques.** Sampling is an approach to overcome the imbalance of datasets. We use standard decision tree ensembles with two sampling approaches: SMOTE and RUS. As AUROC and AUPRC are mostly used performance measures for imbalanced datasets, only these performance measures were used. Results with SMOTE oversampling sampling method are presented in Table 3. We carried out experiments with different ratios of minority class and majority class data points. The results on the original data points were also presented. Results suggest that the best

TABLE 1: Decision tree ensembles and the related packages.

Classifier	Package
Single decision tree	Weka
Random forest	Weka
Bagging	Weka
XGBoost	XGBoost (Python)
AdaBoost	Weka
Balanced random forest (RUS)	Imblearn (Python)
SmoteBagging	Ebmc (R)
RUSBagging	Imblearn (Python)
SmoteBoost	Ebmc (R)
RUSBoost	Imblearn (Python)
SMOTE oversampling (single decision tree)	SMOTE (Weka), Weka
SMOTE oversampling random forest	SMOTE (Weka), Weka
SMOTE oversampling Bagging	SMOTE (Weka), Weka
SMOTE oversampling AdaBoost	SMOTE (Weka), Weka
SMOTE oversampling XGBoost	SMOTE (Weka), Weka
Random undersampling (single decision tree)	Imblearn (Python), XGBoost (Python)
RUS random forest	SpreadsubSample (Weka), Weka
RUSBagging	SpreadsubSample (Weka), Weka
RUS XGBoost	SpreadsubSample (Weka), Weka
RUS AdaBoost	Imblearn (Python), XGBoost (Python)
	SpreadsubSample (Weka), Weka

TABLE 2: The comparative study of various ensemble methods. Different performance measures are used to study their performance.

Ensemble	Accuracy	Precision	Recall	F1-measure	AUROC	AUPRC
Single decision tree	0.845	0.386	0.275	0.321	0.609	0.247
Random forest	0.878	0.585	0.300	0.397	0.872	0.499
Bagging	0.876	0.568	0.313	0.403	0.858	0.506
XGBoost	0.865	0.533	0.575	0.536	0.863	0.505
AdaBoost	0.860	0.462	0.300	0.364	0.844	0.410
Ensembles for imbalanced datasets						
Balanced random forest (RUS)	0.816	0.410	0.812	0.540	0.879	0.561
SmoteBagging	0.858	0.518	0.512	0.493	0.845	0.486
RUSBagging	0.835	0.432	0.675	0.517	0.881	0.516
SmoteBoost	0.868	0.544	0.562	0.530	0.849	0.514
RUSBoost	0.781	0.316	0.562	0.397	0.716	0.340

Bold numbers show the best performances.

TABLE 3: The performance of the Standard decision tree ensembles with the SMOTE oversampling method for different ratios of minority and majority class points. AUROC and AUPRC performance measures are used in the experiments. Bold numbers show the best results.

Ensemble	Original		Ratio = 1		Ratio = 0.75		Ratio = 0.50	
	AUROC	AUPRC	AUROC	AUPRC	AUROC	AUPRC	AUROC	AUPRC
Single decision tree	0.609	0.247	0.679	0.245	0.673	0.273	0.664	0.275
Random forest	0.872	0.499	0.863	0.480	0.869	0.502	0.86	0.492
Bagging	0.858	0.506	0.838	0.481	0.855	0.497	0.846	0.471
XGBoost	0.863	0.505	0.859	0.499	0.859	0.504	0.854	0.500
AdaBoost	0.844	0.410	0.823	0.431	0.847	0.482	0.834	0.434

TABLE 4: The performance of Standard decision tree ensembles with RUS undersampling method for different ratios of minority and majority class points. AUROC and AUPRC performance measures are used in the experiments.

Ensemble	Original		Ratio = 1		Ratio = 0.75		Ratio = 0.50	
	AUROC	AUPRC	AUROC	AUPRC	AUROC	AUPRC	AUROC	AUPRC
Single decision tree	0.659	0.307	0.731	0.231	0.733	0.26	0.745	0.291
Random forest	0.872	0.499	0.852	0.454	0.859	0.476	0.855	0.454
Bagging	0.858	0.506	0.839	0.422	0.848	0.445	0.835	0.442
XGBoost	0.863	0.505	0.845	0.475	0.846	0.510	<b>0.873</b>	<b>0.554</b>
AdaBoost	0.844	0.410	0.826	0.389	0.830	0.401	0.849	0.425

Bold numbers show the best results.

TABLE 5: The performance (AUROC) of various ensembles of decision trees for different sizes of ensembles.

Ensemble	Size			
	20	50	100	200
Random forest	0.852	0.872	0.882	0.876
Bagging	0.859	0.858	0.856	0.857
XGBoost	0.863	0.864	0.867	0.868
AdaBoost	0.827	0.844	0.850	0.855
Ensembles for imbalanced datasets				
Balanced random Forest (RUS)	0.871	0.879	0.889	<b>0.893</b>
SmoteBagging	0.857	0.846	0.854	0.859
RUSBagging	0.874	0.881	0.880	0.883
SmoteBoost	0.853	0.849	0.864	0.864
RUSBoost	0.7516	0.716	0.717	0.715

Bold numbers show the best performance.

TABLE 6: The performance (AUPRC) of various ensembles of decision trees for different sizes of ensembles.

Ensemble	Size			
	20	50	100	200
Random forest	0.484	0.499	0.523	0.517
Bagging	0.524	0.506	0.490	0.488
XGBoost	0.494	0.505	0.507	0.511
AdaBoost	0.415	0.410	0.441	0.453
Ensembles for imbalanced datasets				
Balanced random forest (RUS)	0.557	0.561	0.586	<b>0.596</b>
SmoteBagging	0.478	0.486	0.473	0.489
RUSBagging	0.508	0.516	0.532	0.552
SmoteBoost	0.537	0.514	0.518	0.516
RUSBoost	0.364	0.340	0.361	0.334

Bold numbers show the best performance.

TABLE 7: The effect of age variable on the performance (AUROC) of various ensembles. “+” sign shows the better performance of with age or without age.

Ensemble	Size		Size = 100	
	With age	Without age	With age	Without age
Random forest	0.867	0.872 (+)	0.878	0.882 (+)
Bagging	0.861 (+)	0.858	0.861 (+)	0.856
XGBoost	0.860	0.863 (+)	0.866	0.867 (+)
AdaBoost	0.850 (+)	0.844	0.865 (+)	0.850
Ensembles for imbalanced datasets				
Balanced random Forest (RUS)	<b>0.892</b> (+)	0.879	0.895 (+)	0.889
SmoteBagging	0.862 (+)	0.845	0.859 (+)	0.854
RUSBagging	0.883 (+)	0.881	0.887 (+)	0.880
SmoteBoost	0.863 (+)	0.849	0.857	0.863 (+)
RUSBoost	0.755 (+)	0.716	0.727 (+)	0.7171

TABLE 8: The effect of age variable on the performance (AUPRC) of various ensembles. “+” sign shows the better performance of with age or without age.

Ensemble	Size = 50		Size = 100	
	With age	Without age	With age	Without age
Random forest	0.541 (+)	0.499	0.575 (+)	0.523
Bagging	0.539 (+)	0.506	0.548 (+)	0.490
XGBoost	0.517 (+)	0.505	0.537 (+)	0.507
AdaBoost	0.485 (+)	0.410	0.527 (+)	0.441
Ensembles for imbalanced datasets				
Balanced Random Forest (RUS)	<b>0.606</b> (+)	0.561	0.602 (+)	0.585
SmoteBagging	0.552 (+)	0.486	0.546 (+)	0.473
RUSBagging	0.582 (+)	0.516	0.598 (+)	0.532
SmoteBoost	0.554 (+)	0.514	0.547 (+)	0.518
RUSBoost	0.333	0.340 (+)	0.329	0.361 (+)

AUROC was 0.872 by random forest with the original dataset, whereas the best AUPRC was 0.648 by bagging with the original dataset. For bagging and random forest, the results improved with oversampling. Except for a single decision tree, SMOTE oversampling method had a negative effect on the performance of other ensembles. The presence of noisy minority points may be the reason for the poor performance of SMOTE oversampling method [38].

RUS was done on the majority class to change the ratio of minority class and majority class. The best results (AUROC and AUPRC) were obtained with XGBoost. The best AUROC was 0.873 with the data with a ratio of 0.5 whereas the best AUPRC was 0.554. XGBoost and Ada-Boost performed best for the dataset with a ratio of 0.5 whereas random forest performed best for the original data. The results suggest that the sampling methods did not have similar effects on all the ensembles. However, the best results among all the classifiers were obtained with the XGBoost for the data with a ratio of 0.5 created with RUS.

**4.3. Effects of the Ensemble Size.** An ensemble is a combination of classifiers. The number of classifiers in an ensemble is the size of the ensemble. An experiment is carried out to study the effect of ensemble sizes on different classifier ensembles. For the study, those ensembles were selected which performed better with default values. AUROC and AUPRC performance measures are used for the study. Results for AUROC and AUPRC are presented in Tables 5 and 6, respectively. Results suggest that generally the performance slightly improves or remains constant with size. The results are consistent with the theory of classifier ensembles that most of the performance improvement comes with the first few classifiers in ensembles, adding more classifiers may not be very useful [39].

**4.4. Effects of the Age Variable.** The data has 18 lab findings. The data also has the age of each patient. The previous experiments [12] used 18 lab tests for the Covid-19 prediction. They did not use age in the experiments. The age of a patient plays a very important part in the severity of Covid-19. Therefore, it is important to understand the effect of age on the Covid-19 prediction. An experiment is carried out with a dataset with 19 attributes (18 lab findings + age variable) and the results are compared with that of 18 attributes. Experiment settings were the same as discussed for the dataset with 18 attributes. Two sets of experiments were done, one with the ensemble size equal to 50 and the other with ensemble size equal to. Results are presented in Tables 7 and 8. For the AUROC performance measure with an ensemble size of 50, seven out of nine ensembles performed better for the dataset with 19 attributes. Similarly, with an ensemble size of 100, six out of nine ensembles performed better with the dataset with 19 attributes. For the AUPRC performance measure, eight out of nine ensembles performed better for the dataset with 19 attributes, with ensemble sizes of 50 and 100. The results suggest that including

age variable in the dataset can improve the prediction performance.

Following are the findings of the experiments:

- (I) Covid-19 dataset based on laboratory test [12] is an imbalanced dataset. However, it has not been studied using machine learning algorithms developed for imbalanced datasets. We studied them using decision tree ensembles for imbalanced datasets. The study demonstrates that decision tree ensembles for imbalanced datasets perform better for this Covid-19 dataset.
- (II) Experiments with different sampling techniques suggest that that the sampling methods did not have similar effects on all the general decision tree ensembles. However, the best results among all the classifiers were obtained with the XGBoost for the data with a ratio of 0.5 created with RUS.
- III. The previous experiments [12] did not use the AUPRC performance measure, which is a better performance measure as compared to AUROC [37]. The results could be misleading with inappropriate performance measures. As in Table 2, the performance of balanced random forest (RUS) is best for AUPRC, whereas RUSBagging is best for AUROC. The classifier based on the AUPRC performance measure is used.
- (IV) The dataset has 19 attributes (18 lab tests + age). The previous experiments [12] used 18 lab tests for the Covid-19 prediction. We also studied the effect of the age attribute on the Covid-19 prediction. It is found that including age attribute with 18 lab tests can improve the Covid-19 prediction accuracy.

## 5. Conclusion and Future Work

The prediction of Covid-19 positive cases is an important step for managing Covid-19 positive patients. Machine learning algorithms can be useful for this classification task. Various kinds of decision tree ensembles are applied on a Covid-19 dataset. The dataset has commonly taken laboratory tests. Therefore, these kinds of datasets are easy to collect. The dataset has a class imbalance problem. The results demonstrate that decision tree ensembles developed for imbalanced datasets perform better than standard decision tree ensembles. This suggests that the selection of classification methods should be based on the properties of data. If the data is imbalanced, classifiers developed for imbalanced datasets should be used. Similarly, the appropriate performance measures should be used for a given classification problem. Otherwise, the results could be misleading. The results also suggest that combining the age variable with the other laboratory tests can improve the prediction performance. In the future, we will compare the performance of decision tree ensembles with the other type of classifiers such as SVM and deep learning classifiers. We will further study the combination

of laboratory tests with X-ray data for the prediction of Covid-19 positive cases. The prediction of the severity of Covid-19 by using datasets with laboratory tests will also be investigated.

## Appendix

We discuss various performance measures in detail. We define terms which are used to define these measures for a binary dataset. The dataset has two classes: positive and negative. In our experiments, the minority class is taken as a positive class.

True positive (TP): correctly predicted positive class data points.

False positive (FP): negative class data points predicted as a positive class. True negative (TN): correctly predicted negative class data points.

False negative (FN): negative class data points predicted as a positive class.  $P$ : total positive class data points.

$N$ : total negative class data points.

True positive rate (TPR) =  $TP/P$ .

False positive rate (FPR) =  $FP/N$ .

Performance measures are defined using the following equations:

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + FP + TN + FN)}, \quad (\text{A.1})$$

$$\text{Precision} = \frac{TP}{(TP + FP)}, \quad (\text{A.2})$$

$$\text{Recall} = \frac{TP}{(TP + FN)}, \quad (\text{A.3})$$

$$F\text{-measure} = \frac{2 * \text{precision} * \text{recall}}{(\text{precision} + \text{recall})}. \quad (\text{A.4})$$

## Data Availability

The data are publicly available data.

## Conflicts of Interest

The authors declare no conflicts of interest.

## References

- [1] World Health Organization, *Naming the Coronavirus Disease (Covid-19) and the Virus that Causes it*, World Health Organization, Geneva, Switzerland [https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-\(covid-2019\)-and-the-virus-that-causes-it](https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-(covid-2019)-and-the-virus-that-causes-it) Accessed 26th May 2020.
- [2] World Health Organization, *WHO Director-General's Opening Remarks at the Media Briefing on Covid-19*, World Health Organization, 11 march 2020 <https://www.who.int/dg-speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19—11-march-2020> Accessed, Geneva, Switzerland.
- [3] Johns Hopkins University, *Covid-19 Dashboard by the Center for Systems Science and Engineering (Csse)* at, Johns Hopkins University, Baltimore, MD, USA <https://gisanddata.maps.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6> Accessed 24th Nov. 2020.
- [4] World Health Organization, *Coronavirus Disease 2019 (Covid-19)*, World Health Organization, Geneva, Switzerland <https://www.mayoclinic.org/diseases-conditions/coronavirus/symptoms-causes/syc-20479963> Accessed 26th Nov 2020.
- [5] World Health Organization, *Coronavirus Disease 2019 Testing Basics*, World Health Organization, Geneva, Switzerland <https://www.fda.gov/consumers/consumer-updates/coronavirus-disease-2019-testing-basics> Accessed 26th Nov 2020.
- [6] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer-Verlag New York Inc, New York, NY, USA, 2008.
- [7] Md. Milon Islam, F. Karray, R. Alhajj, and Z. Jia, "A review on deep learning techniques for the diagnosis of novel coronavirus (Covid-19)," 2020, <https://arxiv.org/abs/2008.04815>.
- [8] M.-H. Tayarani, "Applications of artificial intelligence in battling against covid-19: a literature review," *Chaos, Solitons & Fractals*, Article ID 110338, 2020.
- [9] S. Lalmanawma, J. Hussain, and L. Chhakchhuak, "Applications of machine learning and artificial intelligence for covid-19 (sars-cov-2) pandemic: a review," *Chaos, Solitons & Fractals*, vol. 139, Article ID 110059, 2020.
- [10] N. Bonadia, A. Carnicelli, A. Piano et al., "Lung ultrasound findings are associated with mortality and need for intensive care admission in covid-19 patients evaluated in the emergency department," *Ultrasound in Medicine & Biology*, vol. 46, no. 11, pp. 2927–2937, 2020.
- [11] L. Li, L. Qin, Z. Xu et al., "Using artificial intelligence to detect covid-19 and community-acquired pneumonia based on pulmonary ct: evaluation of the diagnostic accuracy," *Radiology*, vol. 296, no. 2, pp. E65–E71, 2020.
- [12] T. B. Alakus and I. Turkoglu, "Comparison of deep learning approaches to predict covid-19 infection," *Chaos, Solitons & Fractals*, vol. 140, Article ID 110120, 2020.
- [13] A. F. D. M. Batista, J. L. Miraglia, T. H. R. Donato, and A. D. P. C. Filho, "Covid-19 diagnosis prediction in emergency care patients: a machine learning approach," 2020, <https://www.medrxiv.org/content/10.1101/2020.04.04.20052092v2>.
- [14] P. Schwab, A. DuMont Schutt, B. Dietz, and S. Bauer, *Predcovid-19: A Systematic Study of Clinical Predictive Models for Coronavirus Disease 2019*, <https://www.scienceopen.com/document?vid=abfb14b2-e267-4292-ad9a-f62e82f0a705>, 2020.
- [15] Y. Sun, A. K. C. Wong, M. S. Kamel, and M. s. Kamel, "Classification of imbalanced data: a review," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 23, no. 4, pp. 687–719, 2009.
- [16] H. Guo, Y. Li, J. Shang, M. Gu, Y. Huang, and B. Gong, "Learning from class-imbalanced data: review of methods and applications," *Expert Systems with Applications*, vol. 73, pp. 220–239, 2017.
- [17] Q. Gu, L. Zhu, and Z. Cai, "Evaluation measures of the classification performance of imbalanced data sets," in *Communications in Computer and Information Science*, Z. Cai, Z. Li, Z. Kang, and Y. Liu, Eds., Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 461–471, 2009.

- [18] L. I. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms*, Wiley Publishing, Hoboken, NJ, USA, 2nd edition, 2014.
- [19] K. Wang, S. Li, Q. Wang, and C. Hou, "Identification of hormone-binding proteins using a novel ensemble classifier," *Computing*, vol. 101, no. 6, pp. 693–703, 2019.
- [20] A. Amal and M. E. B. Menai, "SVM ensembles for named entity dis- ambiguation," *Computing*, vol. 102, no. 4, pp. 1051–1076, 2020.
- [21] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [22] T. Chen and C. Guestrin, "Xgboost: a scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16*, pp. 785–794, Association for Computing Machinery, San Francisco, CA, USA, August 2016.
- [23] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, no. 2, pp. 123–140, 1996.
- [24] B. Krawczyk, M. Woźniak, and G. Schaefer, "Cost-sensitive decision tree ensembles for effective imbalanced classification," *Applied Soft Computing*, vol. 14, pp. 554–562, 2014.
- [25] A. L. C. Chen and L. Breiman, "Using random forest to learn imbalanced data," Technical Report 666, Department of Statistics, UC Berkeley, 2004.
- [26] J. Ross, *Quinlan. C4.5: Programs for Machine Learning*, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1993.
- [27] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of Computer and System Sciences*, vol. 55, no. 1, pp. 119–139, 1997.
- [28] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, no. 1, pp. 321–357, 2002.
- [29] N. V. Chawla, A. Lazarevic, L. O. Hall, and K. W. Bowyer, "Smoteboost: improving prediction of the minority class in boosting," in *Knowledge Discovery in Databases: PKDD 2003*, N. Lavrač, D. Gamberger, L. Todorovski, and H. Blockeel, Eds., pp. 107–119, Springer Berlin Heidelberg, Berlin, Heidelberg, 2003.
- [30] S. Wang and X. Yao, "Diversity analysis on imbalanced data sets by using ensemble models," in *Proceedings of the 2009 IEEE Symposium on Computational Intelligence and Data Mining*, pp. 324–331, Nashville, Tennessee, April 2009.
- [31] C. Seiffert, T. M. Khoshgoftaar, J. Van Hulse, and A. Napolitano, "RUSBoost: a hybrid approach to alleviating class imbalance," *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 40, no. 1, pp. 185–197, 2010.
- [32] PYPI, *Imbalanced-Learn 0.7.0*, <https://pypi.org/project/imbalanced-learn/> Accessed 30th November, 2020.
- [33] E. Frank, M. A. Hall, G. Holmes, R. Kirkby, B. Pfahringer, and I. H. W.. Weka, *A Machine Learning Workbench for Data Mining*, pp. 1305–1314, Springer, Berlin, Germany, 2005.
- [34] R Foundation for Statistical Computing, *R Core Team. R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, AT, USA, 2013.
- [35] H. H. Chen, *R Package version 1.0.0*, Ebmc, 2017, <https://cran.r-project.org/bin/windows/base.old/>.
- [36] "Xgboost python package," 2020, <https://xgboost.readthedocs.io/en/latest/python/pythonintro.html>.
- [37] T. Saito and M. Rehmsmeier, "The precision-recall plot is more informative than the roc plot when evaluating binary classifiers on imbalanced datasets," *PLoS One*, vol. 10, no. 3, pp. 1–21, 2015.
- [38] X. W. Liang, A. P. Jiang, T. Li, Y. Y. Xue, and G. T. Wang, "Lr-smote — an improved un- balanced data set oversampling based on k-means and svm," *Knowledge-Based Systems*, vol. 196, p. 105845, 2020.
- [39] G. Fumera, F. Roli, and A. Serrau, "A theoretical analysis of bagging as a linear combination of classifiers," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 7, pp. 1293–1299, 2008.