Dicle University
Journal of Engineering

https://dergipark.org.tr/tr/pub/**dumf**
duje.dicle.edu.tr

Research Article

Bagging and Boosting Methods for Predicting Mortality of Patients with COVID-19

Hilal ARSLAN[1*]

[1] Ankara Yıldırım Beyazıt University, Software Engineering Department, hilalarslanceng@gmail.com, Orcid No: 0000-0002-6449-6952

ARTICLE INFO

Article history:

Received 30 March 2022
Received in revised form 5 May 2022
Accepted 8 May 2022
Available online 28 June 2022

Keywords:

COVID-19, SARS-CoV-2, Ensemble learning, Bagging, Boosting

Doi: 10.24012/dumf.1095858

* Corresponding author

ABSTRACT

COVID-19 pandemic has been going on for more than two years and an increasing number of deaths has been occurring. The mortality prediction of the COVID-19 patient is crucial to reduce the risk of imminent death as well as to apply effective clinical treatment strategy. Ensemble learning techniques are effectively employed to predict the outcome of the patients with COVID-19. In this study, we perform bagging and boosting methods to predict mortality of the patients with COVID-19. The six different decision tree methods, C4.5, Random tree, REPTree, Logistic Model Tree, Decision Stump, and Hoeffding Tree are performed for base learners in bagging and boosting. The results are obtained using a real-world dataset including information obtained from 1085 patients. Experimental results present that bagging using REPTree as a base learner achieves an accuracy of 97.24%. Furthermore, when we compare our results with other classification algorithms, the proposed method has a higher performance with respect to the accuracy, and presents an admirable performance.

Introduction

Severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) belongs to a betacoronavirus family and has caused a disease known as COVID-19. COVID-19 has been announced as a worldwide pandemic and has caused many deaths worldwide. According to the World Health Organization (WHO) report published on March 15, 2022, the total number of patients who were positive in the diagnosis of COVID-19 is 458,479,635 and the number of patients who died from COVID-19 is 6,047,653. The common symptoms of the disease are tiredness, cough, fever, sore throat as well as problems in breathing and there is no proven treatment for COVID-19. Zoabi et al. [1] investigated importance of these symptoms and they reported that headache and sore throat were identified as the most important symptoms. Several studies have shown that people who have a chronic respiratory disease, older and male are more affected by COVID-19 disease. Furthermore, people with crucial medical diseases like cancer and cardiovascular illness are affected seriously from COVID-19.

After increasing transmissibility of COVID-19, several variants of SARS-CoV-2 have been emerged. The main variants of SARS-CoV-2 are the Alpha, Beta, Gamma, Delta, and Omicron, and mRNA vaccines are effective against SARS-CoV-2 variants. Machine learning (ML) techniques are actively used in COVID-19 detection from genome sequences [2,3] and common symptoms of COVID-19 [1]. Furthermore, ML techniques are used to estimate the severity of COVID-19 infected in the patient [4]. Besides, deep learning (DL) techniques are efficiently used to predict COVID-19 from medical imaging [5,6]. DL techniques are also efficiently used to interpret clinical findings from various types of cancers [7] and biomedical studies [8].

In this study, we applied bagging and boosting methods to predict the mortality of patients with COVID-19. We used six different decision tree methods as base learners of the bagging and boosting methods. The rest of the paper is organized as follows. In Section 2, related work is summarized. In Section 3, the methods are given. In Section 4, the results are presented and compared with methods published in the literature. The last section presents conclusion as well as the future directions.

Related Work

Various types of ML and DL algorithms recently published to detect COVID-19 that can be found in [2,3,6,9-16].

Schwab et al. [17] proposed a method for predicting COVID-19 mortality from electronic health records. Deniz et al. [18] introduced a technique for predicting severity of COVID-19. They applied a multithreaded genetic algorithm to choose the optimum set of features and combine with extreme learning machine. Mydukuri et al. [19] used Gaussian neuro-fuzzy classifier to predict COVID-19. Zoabi et al. [1] employed an ML algorithm to diagnose COVID-19. They designed their methods based on basic information and symptoms without using any medical equipment. Their method achieved an auROC of 0.86. Cabitza et al. [20] evaluated five machine learning techniques on the data including blood tests. They performed various types of ML methods. They achieved satisfactory results and concluded that ML techniques based on blood tests can detect COVID-19 cases fast compared to the RT-PCR tests. Unal and Dudak [21] implemented naive bayes, SVM, KNN, and decision tree methods on the dataset including 19 features which are sex, age, the state of pneumonia as well as the state of various types of diseases such as asthma, diabetes, kidney failure, and hypertension. They showed that the SVM achieved an accuracy of 100%.

Alakus and Turkoglu [5] compared DL approaches to diagnose COVID-19 using laboratory findings. Their methods achieved an accuracy of 68.6%. Jiang et al. [22] developed a ML tool to predict patients at risk for COVID-19. They used the data containing 11 features which are blood count, hemoglobin, temperature, Na+, creatinine, K+, a liver enzyme, myalgias, gender, lymphocyte count, and age. They performed logistic regression, KNN, decision tree, and SVM classifiers. They concluded that myalgias, hemoglobin, and the liver enzyme are important features and the most predictive. Batista et al. [23] used SVM, random forest, neural network, gradient boosted trees and logistic regression to diagnose COVID-19 using laboratory findings. Their method achieved the best AUC score with 0.84 when SVM and random forest methods were used. Schwab et al. [24] evaluated predictive models using logistic regression, neural network, random forest, different SVM methods and gradient boosting using demographic, clinical and blood analysis data containing 111 features. They obtained the best performance with 66% AUC score when gradient boosting method is used. Halasz et al. [25] investigated mortality in the patients with COVID-19 pneumonia. They developed a method based on the Piacenza score and their method achieved satisfactory results. Quanjel et al. [26] used decision trees to predict COVID-19 mortality from laboratory data.

Methods

In this section, we present bagging and boosting methods applied in this study. Furthermore, we explain various types of decision tree methods used as base learners in bagging and boosting methods.

Bagging Using Decision Tree Classifiers

Bagging, a bootstrap ensemble method, generates multiple training sets using a bootstrap method [27,28]. Each training set contains N samples where N is the size of the original training set and is created by random and repeatable distribution of the original training set. Each individual classifier is operated on different training set in parallel, and the final prediction is achieved by aggregating the predictions obtained from individual classifiers. We performed bagging method using various types of decision tree based classifiers, C4.5, Random tree, REPTree, LMT, Decision Stump, and Hoeffding Tree algorithms, respectively.

Boosting Using Decision Tree Classifiers

Boosting, an ensemble method, combines a set of weak learners to achieve a strong learner [29]. Equal weighted samples are retrieved from the training dataset, and next decision tree classifiers, C4.5, Random tree, REPTree, LMT, Decision Stump, and Hoeffding Tree are applied on this training dataset to form the first weak learner, respectively. Thus, each model copes with the weakness of its predecessor. The samples are weighted with respect to classification accuracies and the new training set is created. The process of creating weak learners using new training set is repeated until a strong classifier is achieved.

Adaptive Boosting (AdaBoost) [30] combines many weak learners into a strong one such that the weak classifiers try to improve classification of the data samples that were classified incorrectly by the previous classifier. In this study we used AdaBoost algorithm using decision tree methods. In the next sections, we briefly explain the decision tree methods used as base learners for bagging and boosting methods.

Decision Tree Methods

C4.5 [31] is among widely known decision tree methods and was introduced by Ross Quinlan [32]. We used J48, which is an implementation of C4.5 in Weka. C4.5 method uses the information gain to specify attributes for each node [31]. In Random Tree method, a decision tree is randomly selected among various number of decision trees constructed by randomly selected samples from the dataset. Logistic Model Tree (LMT) [33] combines decision trees with logistic model trees. Decision Stump [34] is a one-level decision tree. The method builds a decision tree with one internal node (the root) that is connected to its leaves. It is commonly used as a base learner in bagging and boosting. Hoeffding Tree [35,36] is an incremental decision tree technique for big data streams based on the assumption that the distribution of the data does not change with time. It expands decision tree incrementally with respect to Hoeffding bound.

Experimental Results

Data Description

The original dataset used in this study can be founded at Kaggle as "Novel Corona Virus 2019 Dataset" [37] and gathered from different sources including the WHO and John Hopkins University.

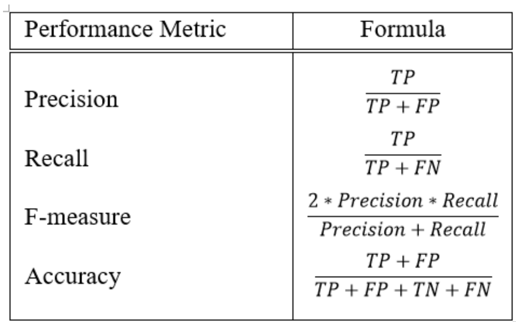
The original dataset includes some missing as well as redundant values. Therefore, the dataset was preprocessed by Deniz et al. [18] and we used preprocessed version of it. The dataset includes information belonging to 1085 patients, and 63 out of 1085 patients are recorded as deceased. The information related to the patients includes location, country, gender, age, date that the patient has symptoms, the date that the patient visits hospital, the knowledge whether the patient visited Wuhan, and the knowledge the patient from Wuhan. The dataset also includes knowledge about the number of the days the first symptom was occurred. Furthermore, the dataset includes 24 unique symptoms listed below and includes 34 features in total.

- Abdominal pain
- Breathing difficulty
- Chest pain
- Chills
- Cold
- Cough
- Diarrhea
- Fatigue
- Fever
- Flu
- Headache
- Joint pain
- Loss of appetite
- Muscle pain
- Nausea
- Physical discomfort
- Pneumonia
- Reflux
- Runny nose
- Sneeze
- Sore throat
- Sputum
- Thirst
- Vomiting

Performance Measurements

We used precision, recall, F-measure, and accuracy values that are presented in Table 1. Precision is the ratio of the samples correctly classified as positive to the total number of samples classified as positive. Recall is the ratio of the samples correctly classified as positive to total number of positive samples. F-measure is the harmonic mean of the precision and recall, and the accuracy is the most common measure and ratio of the samples correctly classified to the total number of samples.

Table 1. Performance measures used in this study



Performance Metric	Formula
Precision	$\frac{TP}{TP+FP}$
Recall	$\frac{TP}{TP+FN}$
F-measure	$\frac{2 * Precision * Recall}{Precision + Recall}$
Accuracy	$\frac{TP+FP}{TP+FP+TN+FN}$

Results and Discussion

In this section, we show the results of bagging and boosting methods using various types of decision tree classifiers. The number of iteration is set to 10. 5-fold cross validation technique is used.

Table 2 presents precision and recall results of bagging and boosting methods using six different decision tree methods. First, we evaluate the results of AdaBoost method using decision tree methods. AdaBoost method achieves the best precision with 97.50% when REPTree is used as a base learner. The precision of the AdaBoost method using C4.5, Random Tree, LMT, and Decision Stump as base learners are about 97%, and AdaBoost method has a precision of 94.19% when Hoeffding Tree method is used as a base learner. On the other hand, the recall values of the AdaBoost method using six different decision tree methods are close to 100%.

Second, we evaluate the results of bagging method using the decision tree methods. The precision values range from 94.19% to 97.33%, and the best precision is obtained when LMT is used as a base learner. When we look at recall results, bagging using C4.5, REPTree, LMT, and Hoeffding Tree methods achieve a recall about 100% and bagging using Random Tree and Decision Stump achieves a recall of about 99%.

Table 2. Precision and recall results of bagging and boosting methods using six different decision tree methods

	Precision		Recall	
Method	AdaBoost	Bagging	AdaBoost	Bagging
C4.5	97.40	97.15	99.02	99.90
RandomTreee	97.22	97.03	99.12	99.22
REPTree	97.50	97.24	99.22	99.90
LMT	97.22	97.33	99.22	99.80
Decision Stump	96.86	95.56	99.51	99.02
000Hoeffding	94.19	94.19	100	100

Table 3. F-measure and accuracy results of bagging and boosting methods using six different decision tree methods.

	F-measure		Accuracy	
Method	AdaBoost	Bagging	AdaBoost	Bagging
C4.5	98.20	98.50	96.59	97.14
RandomTreee	98.16	98.11	96.50	96.41
REPTree	98.35	98.55	96.87	97.24
LMT	98.21	98.55	96.59	97.24
Decision Stump	98.17	97.26	96.50	94.75
Hoeffding Tree	97.01	97.01	94.19	94.19

Now, we analyze F-measure and accuracy results of bagging and boosting methods shown in Table 3. First, we evaluate the F-measure and accuracy values for the Adaboost algorithm using the decision tree methods. The best F-measure value is 98.35% and achieved when the AdaBoost method using REPTree. Similarly, the AdaBoost method achieves the best accuracy that is 96.87% when the REPTree method is used. Finally, we analyze F-measure and accuracy results of bagging using the decision tree methods. While F-measure values change from 97.01% to 98.55%, the accuracy values range from 94.19% to 97.24%. The bagging using the REPTree or LMT method achieves the best F-measure and accuracy values.

Comparison with the other state-of-the-art methods

Deniz et al. [18] reported the accuracy values of the traditional machine learning methods shown in Figure 1 on the original COVID-19 dataset. With respect to their results, the best accuracy was obtained with 90.14% when Random forest method is used.

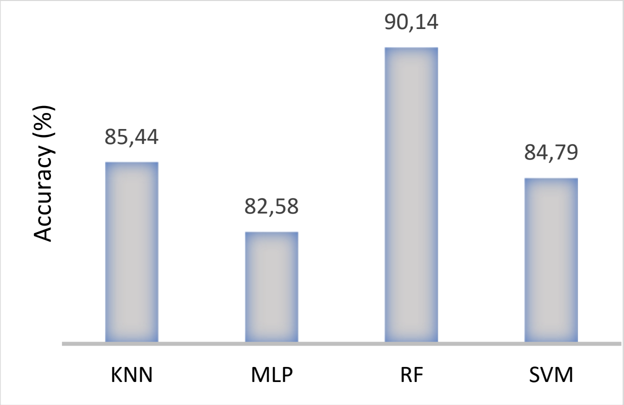


Figure 1. Results of ML methods

Figure 2 gives the accuracy results of the-state-of-the-art methods on the COVID-19 dataset. Iwendi et al. [38] used Random Forest enhanced with AdaBoost algorithm. The accuracy of their method is 94.0%. Deniz et al. [18] used a multi-threaded genetic algorithm, and their method achieves an accuracy of 96.22%. Mydukuri et al. [19] applied feature selection method and they used Gaussian neuro-fuzzy classifier. The accuracy of their method is 95.0%. Too and Mirjalili [39] used Dragonfly Algorithm to select the most relevant features and their method reached an accuracy of 92.21%. When we compare our method with these methods, the proposed method achieves the highest accuracy with 97.24%.

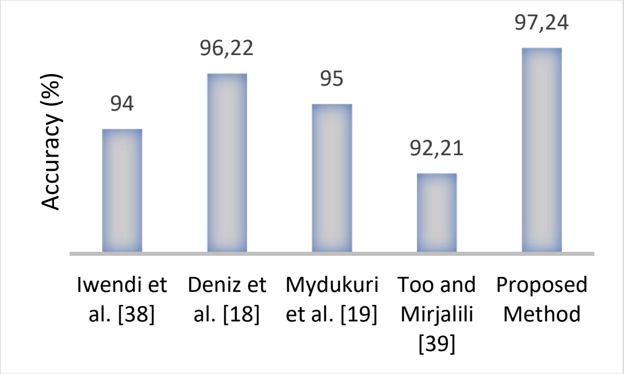


Figure 2. Accuracy results of the state-of-the-art methods

Conclusion

COVID-19 disease has caused severe and deadly complications. Determining the severity level of the patient is important to follow a better treatment strategy. In this paper, we apply bagging and boosting techniques using several types of decision tree methods on a real dataset including symptoms information as well as gender, age, and location information of the patients. Experimental results indicate that performance of the bagging method is higher than the boosting method, and bagging using REPTree as a base learner achieves an accuracy of 97.24%. This means that the information about the patients including symptoms knowledge as well as the basic information has a severe effect to predict mortality of the patients. In future studies, we plan to combine deep learning techniques with high performance techniques to predict severity level of the patients, and we will also consider the laboratory information of the patients.

Ethics committee approval and conflict of interest statement

There is no need to obtain permission from the ethics committee for the article prepared.

There is no conflict of interest with any person / institution in the article prepared.

Authors' Contributions

All parts of the paper are prepared and implemented by Hilal Arslan.

References

[1] Y. Zoabi, S. Deri-Rozov, and N. Shomron. Machine learning-based prediction of COVID-19 diagnosis based on symptoms. *npj Digital Medicine*, 4(1), 3. 2021. DOI: 10.1038/s41746-020-00372-6

[2] H. Arslan and H. Arslan. A new covid-19 detection method from human genome sequences using cpg island features and knn classifier. *Engineering Science and Technology, an International Journal*, 2021. DOI: 10.1016/j.jestch.2020.12.026

[3] H. Arslan, Machine learning methods for covid-19 prediction using human genomic data, *MDPI Proceedings,* vol.74 no.1. 2021. DOI: 10.3390/proceedings2021074020

[4] W. Shang, J. Dong, Y. Ren, M. Tian, W. Li, J. Hu and Y. Li, The value of clinical parameters in predicting the severity of COVID- 19. *Journal of Medical Virology*, 92(10), 2188–2192, 2020. DOI: 10.1002/jmv.26031

[5] T. B. Alakus and I. Turkoglu, “Comparison of deep learning approaches to predict covid-19 infection,” *Chaos, Solitons Fractals*, vol. 140, p. 110120, 2020.

[6] M. Alazab, A. Awajan, A. Mesleh, A. Abraham, V. Jatana, and S. Alhyari4, “Covid-19 prediction and detection using deep learning,” *International Journal of Computer Information Systems and Industrial Management Applications*, vol. 12, pp. 168–181, 2020.

[7] V. Andriasyan, A. Yakimovich, F. Georgi, A. Petkidis, R. Witte, D. Puntener, and U. F. Greber, “Deep learning of virus infections reveals mechanics of lytic cells,” Oct. 2019. DOI: 10.1101/798074

[8] A. W. Senior et al. (2020). “Improved protein structure prediction using potentials from deep learning,” *Nature*, vol. 577, no. 7792, pp. 706–710. DOI: 10.1038/s41586-019-1923-7

[9] Y. Zoabi, S. Deri-Rozov, and N. Shomron, “Machine learning-based prediction of COVID-19 diagnosis based on symptoms,” *npj Digital Medicine*, vol. 4, no. 1, p. 3, Dec. 2021.

[10] L. J. Muhammad, E. A. Algehyne, S. S. Usman, A. Ahmad, C. Chakraborty, and I. A. Mohammed, “Supervised Machine Learning Models for Prediction of COVID-19 Infection using Epidemiology Dataset,” *SN Computer Science*, vol. 2, no. 1, p. 11, Feb. 2021. [Online] http://link.springer.com/10.1007/s42979-020-00394-7

[11] S. F. Ardabili, A. Mosavi, P. Ghamisi, F. Ferdinand, A. R. arkonyi-Koczy, U. Reuter, T. Rabczuk, and P. M. Atkinson, “COVID-19 Outbreak Prediction with Machine Learning,” Algorithms, vol. 13, no. 10, p. 249, Oct. 2020. [Online]. Available: https://www.mdpi.com/1999-4893/13/10/249

[12] M.-H. Tayarani N., “Applications of artificial intelligence in battling against covid-19: A literature review,” *Chaos, Solitons & Fractals,* vol. 142, p. 110338, Jan. 2021. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0960077920307335

[13] S. Kushwaha, S. Bahl, A. Bagha, K. Parmar, M. Javaid, A. Haleem, and R. Singh, “Significant applications of machine learning for covid-19 pandemic,” *Journal of Industrial Integration and Management*, vol. 5, no. 4, Dec. 2020.

[14] F. De Felice and A. Polimeni, “Coronavirus Disease (COVID-19): A Machine Learning Bibliometric Analysis,” *In Vivo*, vol. 34, no. 3 suppl, pp. 1613–1617, 2020. [Online]. Available: http://iv.iiarjournals.org/lookup/doi/10.21873/invivo.11951

[15] H. Arslan, “COVID-19 prediction based on genome similarity of human SARS-CoV-2 and bat SARS-CoV-like coronavirus,” *Computers & Industrial Engineering*, vol. 161, p. 107666, Nov. 2021. [Online]. Available: https://doi.org/10.1016/j.cie.2021.107666

[16] H. Arslan and B. Aygün, “Performance analysis of machine learning algorithms in detection of covid-19 from common symptoms,” in *2021 29th Signal Processing and Communications Applications Conference (SIU)*, 2021, pp. 1–4.

[17] P. Schwab, A. Mehrjou, S. Parbhoo, L. A. Celi, J. Hetzel, M. Hofer, B. Scholkopf, and S. Bauer, “Real-time prediction of COVID-19 related mortality using electronic health records,” *Nature Communications*, vol. 12, no. 1, Feb. 2021. [Online]. Available: https://doi.org/10.1038/s41467-020-20816-7

[18] A. Deniz, H. E. Kiziloz, E. Sevinc, and T. Dokeroglu, “Predicting the severity of covid-19 patients using a multi-threaded evolutionary feature selection algorithm,” *Expert Systems*, Feb. 2022. [Online]. Available: https://doi.org/10.1111/exsy.12949

[19] R. V. Mydukuri, S. Kallam, R. Patan, F. Al-Turjman, and M. Ramachandran, “Deming least square regressed feature selection and gaussian neuro-fuzzy multi-layered data classifier for early COVID prediction,” *Expert Systems*, Mar. 2021. [Online]. Available: https://doi.org/10.1111/exsy.12694

[20] F. Cabitza, A. Campagner, D. Ferrari, C. D. Resta, D. Ceriotti, E. Sabetta, A. Colombini, E. D. Vecchi, G. Banfi, M. Locatelli, and A. Carobene, “Development, evaluation, and validation of machine learning models for covid-19 detection based on routine blood tests,” *Clinical Chemistry and Laboratory Medicine (CCLM)*, vol. 59, no. 2, pp. 421–431, 2021. [Online]. Available: https://doi.org/10.1515/cclm-2020-1294

[21] Y. Unal and M. N. Dudak, “Classification of covid-19 dataset with some machine learning methods,” *Journal of Amasya University the Institute of Sciences and Technology*, vol. 1, pp. 30 – 37, 2020.

[22] X. Jiang, M. Coffee, A. Bari, J. Wang, X. Jiang, J. Huang, J. Shi, J. Dai, J. Cai, T. Zhang, Z. Wu, G. He, and Y. Huang, “Towards an artificial intelligence framework for data-driven prediction of coronavirus clinical severity,” *Computers, Materials & Continua*, vol. 62, no. 3, pp. 537–551, 2020. [Online]. Available: https://doi.org/10.32604/cmc.2020.010691

[23] A. F. de Moraes 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,” Apr. 2020. [Online]. Available: https://doi.org/10.1101/2020.04.04.20052092

[24] P. Schwab, A. D. Sch¨utte, B. Dietz, and S. Bauer, “Clinical predictive models for COVID-19: Systematic study,” *Journal of Medical Internet*

Research, vol. 22, no. 10, p. e21439, Oct. 2020. [Online]. DOI: 10.2196/21439

[25] Halasz, G., Sperti, M., Villani, M., Michelucci, U., Agostoni, P., Biagi, A., Rossi, L., Botti, A., Mari, C., Maccarini, M., Pura, F., Roveda, L., Nardecchia, A., Mottola, E., Nolli, M., Salvioni, E., Mapelli, M., Deriu, M. A., Piga, D., & Piepoli, M., "A Machine Learning Approach for Mortality Prediction in COVID-19 Pneumonia: Development and Evaluation of the Piacenza Score", In *Journal of Medical Internet Research,* vol. 23, issue 5, 2021. JMIR Publications Inc. DOI:10.2196/29058

[26] Quanjel, M. J. R., van Holten, T. C., Gunst-van der Vliet, P. C., Wielaard, J., Karakaya, B., Söhne, M., Moeniralam, H. S., & Grutters, J. C. (2020). "Replication of a mortality prediction model in Dutch patients with COVID-19", In *Nature Machine Intelligence* (Vol. 3, Issue 1, pp. 23–24). Springer Science and Business Media LLC. https://doi.org/10.1038/s42256-020-00253-3

[27] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, no. 2, pp. 123–140, Aug. 1996. [Online]. Available: https://doi.org/10.1007/bf00058655

[28] B. Efron and R. J. Tibshirani, An Introduction to the Bootstrap. *Springer US*, 1993. [Online]. Available: https://doi.org/10.1007/978-1-4899-4541-9

[29] R. E. Schapire, "The strength of weak learnability," *Machine Learning*, vol. 5, no. 2, pp. 197–227, Jun. 1990. [Online]. Available: https://doi.org/10.1007/bf00116037

[30] R. E. Schapire and Y. Singer, *Machine Learning*, vol. 37, no. 3, pp. 297–336, 1999. [Online]. Available: https://doi.org/10.1023/a:1007614523901

[31] S. Ruggieri, "Efficient c4.5 [classification algorithm]," *IEEE Transactions on Knowledge and Data Engineering*, vol. 14, no. 2, pp. 438–444, 2002.

[32] J. R. Quinlan, C4.5: Programs for Machine Learning, 1993.

[33] N. Landwehr, M. Hall, and E. Frank, "Logistic model trees," *Machine Learning*, vol. 59, no. 1-2, pp. 161–205, May 2005. [Online]. Available: https://doi.org/10.1007/s10994-005-0466-3

[34] S. K. Jha, P. Paramasivam, Z. Pan, and J. Wang, "Decision stump and Stacking C-based hybrid algorithm for healthcare data classification," in *Cloud Computing and Security. Springer International Publishing*, 2018, pp. 205–216. [Online]. Available: https://doi.org/10.1007/978-3-030-00018-919

[35] P. Domingos and G. Hulten, "Mining high-speed data streams," in *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM Press, 2000. [Online]. Available: https://doi.org/10.1145/347090.347107

[36] G. Hulten, L. Spencer, and P. Domingos, "Mining time-changing data streams," in proc. *of the 2001 acm sigkdd intl. conf. on knowledge discovery and data mining*, 2001, pp. 97–106.

[37]"Covid-19 dataset," https://www.kaggle.com/sudalairajkumar/novel-coronavirus-2019-dataset, accessed: 2022-03-01.

[38] C. Iwendi, A. K. Bashir, A. Peshkar, R. Sujatha, J. M. Chatterjee, S. Pasupuleti, R. Mishra, S. Pillai, and O. Jo, "COVID-19 patient health prediction using boosted random forest algorithm," *Frontiers in Public Health*, vol. 8, Jul. 2020. [Online]. Available: https://doi.org/10.3389/fpubh.2020.00357

[39] J. Too and S. Mirjalili, "A hyper learning binary dragonfly algorithm for feature selection: A COVID-19 case study," *Knowledge-Based Systems*, vol. 212, p. 106553, Jan. 2021. [Online]. Available: https://doi.org/10.1016/j.knosys.2020.106553