Original Article

Performance analysis of data mining algorithms for diagnosing COVID-19

Raoof Nopour[1], Hadi Kazemi-Arpanahi[2,3], Mostafa Shanbehzadeh[4*], Akbar Azizifar[5]

Abstract:

BACKGROUND: An outbreak of atypical pneumonia termed COVID-19 has widely spread all over the world since the beginning of 2020. In this regard, designing a prediction system for the early detection of COVID-19 is a critical issue in mitigating virus spread. In this study, we have applied selected machine learning techniques to select the best predictive models based on their performance.

MATERIALS AND METHODS: The data of 435 suspicious cases with COVID-19 which were recorded from the Imam Khomeini Hospital database between May 9, 2020 and December 20, 2020, have been taken into consideration. The Chi-square method was used to determine the most important features in diagnosing the COVID-19; eight selected data mining algorithms including multilayer perceptron (MLP), J-48, Bayesian Net (Bayes Net), logistic regression, K-star, random forest, Ada-boost, and sequential minimal optimization (SMO) were applied in data mining. Finally, the most appropriate diagnostic model for COVID-19 was obtained based on comparing the performance of the selected algorithms.

RESULTS: As the result of using the Chi-square method, 21 variables were identified as the most important diagnostic criteria in COVID-19. The results of evaluating the eight selected data mining algorithms showed that the J-48 with true-positive rate = 0.85, false-positive rate = 0.173, precision = 0.85, recall = 0.85, F-score = 0.85, Matthews Correlation Coefficient = 0.68, and area under the receiver operator characteristics = 0.68, respectively, had the higher performance than the other algorithms.

CONCLUSION: The results of evaluating the performance criteria showed that the J-48 can be considered as a suitable computational prediction model for diagnosing COVID-19 disease.

Keywords:

Artificial intelligence, coronavirus, COVID-19, data mining, diagnosis, machine learning

[1]Department of Health Information Management, School of Health Management and Information Sciences, Iran University of Medical Sciences, Tehran, Iran, [2]Department of Health Information Technology, Abadan University of Medical Sciences, Abadan, Iran, [3]Student Research Committee Department, Abadan University of Medical Sciences, Abadan, Iran, [4]Department of Health Information Technology, School of Paramedical, Ilam University of Medical Sciences, Ilam, Iran, [5]Department of English Language, School of Medicine, Ilam University of Medical Science, Ilam, Iran

Address for correspondence:
Dr. Mostafa Shanbehzadeh, Department of Health Information Technology, School of Paramedical, Ilam University of Medical Sciences, Ilam, Iran.
E-mail: mostafa. shanbezadeh@gmail.com

Received: 30-01-2021
Accepted: 22-02-2021
Published: 30-11-2021

Introduction

COVID-19 is a zoonotic, contagious, and infectious disease caused by a virus called acute respiratory syndrome coronavirus 2,[1] which causes serious damage to the respiratory system, creating pneumonia, and in some cases leads to human death.[2,3] This disease was first reported in December 2019 in Wuhan, Hubei Province in China, and gradually, spread throughout the country and then the world.[4] The most important symptoms of this disease can be mentioned as fever, dry cough, shortness of breath, headache, abdominal pain, gastrointestinal symptoms, vomiting, chest tightness, lung lesions, and respiratory failure.[5,6] The viral disease can be transmitted through oral and nasal secretions of the infected people when coughing or sneezing or talking to healthy people.[7] The incubation period of this disease in most people is up to 14 days, but in some cases, up to 24 days has been reported.[8] The number of cases and deaths is increasing worldwide, while Iran is among the top 10 countries in terms of the prevalence of the disease.[9] According to data published by the World Health Organization (WHO), by August 16, 2020, the number of cases of corona disease has reached 22,492,312 people, the number of

deaths due to this disease has been announced to 788,503 people, in Iran, in this period, there are 352,558 cases and 20,264 deaths due to this disease.[10] Due to the rapid spread of this disease, subsequently, the increasing number of this disease, and deaths worldwide, as well as adverse effects on the economic and social conditions of societies, the use of new and advanced technological methods to quickly diagnose the disease in the early stages, isolation and quarantine cases after that, and effective monitoring are very important in preventing the disease.[11,12] Furthermore, due to the ambiguous and sensitive nature of this disease and the lack of proven therapies and the lack of methods for rapid detection of the disease, the use of innovative and noninvasive methods to identify the disease in earlier stages can be very useful and determinative.[13] The most important new and advanced technologies today that have a great role in controlling the epidemic of COVID-19 can be pointed to artificial intelligence (machine learning and deep learning), cloud computing, the internet of things (IoT), big data mining, robotic technologies, and intelligent systems.[14,15] The data mining process involves discovering important patterns from the vast amount of data that combines machine learning, statistics, and database systems.[9] Given the huge amount of data available in the field of medicine, this process has so far played a very effective role in the management of various diseases such as prognosis, diagnosis, and treatment.[16,17] So far, several studies have been conducted on the application of data mining methods in the management and control of COVID-19 disease with the aim of monitoring and predicting disease trends and epidemiological assessments, providing diagnostic models, determining the best treatment and care model, and predicting the probability of death.[18,19] In a review study by Albahri *et al*., on the role of data mining and artificial intelligence in the detection and diagnosis of COVID-19, the study's quantitative and qualitative criteria were discussed. Finally, the results showed that the use of this technology to provide diagnostic models and identify the most optimal and effective data mining algorithm can be significantly helpful in a timely, effective, and economical diagnosis of the disease.[20] Hence, due to the high prevalence of the disease in our country and the existence of some limitations and scarcity of resources, the purpose of this study was to create an effective and efficient diagnostic model based on comparing the performance of data mining algorithms for COVID-19 disease based on different diagnostic criteria to be able to be a suitable solution for physicians to increase their diagnostic accuracy and reduce the prevalence and death rate in our community.

Materials and Methods

This research methodology consisted of three steps as follows.

Study design and setting

This retrospective and the descriptive-applied study was performed in 2020 in the central hospital for diagnosis and treatment of COVID-19 disease affiliated with Ilam University of Medical Sciences aimed to develop a Clinical Decision Support System based on various data mining techniques and to present the best diagnostic model.

Study participants and sampling

The population was people who had been referred to Mostafa Khomeini Hospital in Ilam city for COVID-19 diagnosis and treatment. The data of 435 COVID-19 and non-COVID-19 cases (positive or negative reverse transcription-polymerase chain reaction) were stored in the electronic medical records (EMRs) of the patient database along with the 40 diagnostic criteria. Out of this number, the information of the 35 cases which had a lot of missing data (more than 70%) was excluded from the analysis. Furthermore, the missing value was imputed with the mean or mode of each variable. The final sample size used in this analysis was 400 COVID-19 and non-COVID-19 cases.

Data collection tool and technique

The data of the 400 cases of patients with the COVID-19 disease and non-COVID-19 were stored on the center's EMRs to be used for data mining. The diagnostic criteria used in the study included demographic findings, clinical and care, medical and personal history, and epidemiological data. The output variable of the study was the same as the diagnostic result of COVID-19 disease, in which the numbers of 0 and 1 were assigned to people who had a negative or positive COVID-19 disease test result.

Ethical consideration

This study was approved by the ethical committee board of Ilam University of Medical Sciences (Ethic code: IR.MEDILAM.REC.1399.294).

Statistical analysis

In this step, before performing the data mining process, due to the number of the diagnostic fields in the research and also to increase the efficiency of the results obtained from different data mining algorithms, some diagnostic criteria that were less important in the research were removed to have a maximum speed to improve the data mining performance. Therefore, because of the existence of two and multistate qualitative variables, the Pearson Chi-square method was used. Using this statistical method, all the diagnostic criteria were weighted according to their degree of importance. The basis of this statistical method is the comparison between the obtained actual value and the expected value, which indicates that the difference between the actual and

expected values is statistically significant and the results are completely random.

The formula for calculating the Chi-square along with how to calculate the expected frequency has been shown in Equations 1 and 2, respectively.

$$X^2 = \sum_{i=1}^{n} (Fe_i - Fo_i)^2 / Fe_i \qquad (1)$$

$$Fe_i = \frac{n_i * n_j}{n} \qquad (2)$$

In Equation 1, the Fo is the observed frequency and Fe is the expected frequency obtained from Equation 2. Furthermore, in this study, the method of selecting the most important diagnostic criteria is based on the Chi-square criterion, at $P < 0.05$.

Data mining and performance evaluation

After determining the most important diagnostic criteria based on the Pearson method and Chi-square test, data mining was performed using WEKA 3.9 software. The eight well-known data mining algorithms including multilayer perceptron (MLP), J-48 decision tree, Bays Net, logistic regression (LR), K-star, random forest (RF), Ada-boost, and sequential minimal optimization (SMO) were used to evaluate and generate the best COVID-19 diagnostic model. The 10% fold cross-validation had been considered for measuring the performance criteria. The performance of the sample classification rate of these data mining algorithms was evaluated based on the true positive (TP) (The COVID-19 cases are truly classified as positive by the data mining model), false positive (FP) (The non-COVID-19 cases falsely classified as positive by the data mining model), false negative (FN) (The COVID-19 cases are falsely classified as negative by the data mining model), and true negative (TN) (The non-COVID-19 cases are truly classified as negative by the data mining model), and their performance was evaluated based on various evaluation criteria such as TP-Rate, FP-Rate, Precision, Recall, F-Score, Matthews Correlation Coefficient (MCC), and the area under the receiver operator characteristics (AUC) diagram of each algorithm in WEKA software were examined and compared.

Results

After weighing the research variables, the most important diagnostic criteria were obtained, 21 COVID-19 diagnosing variables had a meaningful correlation with the output class (positive and negative COVID-19 diagnosis) at $P < 0.05$. Table 1 shows the most important diagnostic criteria, which were weighted using the Chi-square relationship.

Based on the findings have been represented in Table 1, the variable of lung lesions with the highest Chi-square

value ($\chi^2 = 179.21$) was considered as the most important diagnostic criterion for COVID-19 disease. The results of classifying data mining algorithms from people with disease and people without the disease are shown in Table 2.

Based on information obtained from Table 2, the Ada-boost data mining algorithm with the correct classification of the 231 cases (out of 250) of people with COVID-19 disease had the highest classification ability for positive COVID-19 cases, But the Bayesian Net (Bayes Net) with 149 correct samples of people without the disease out of the 150 had the best classification compared to the other data mining algorithms. The results of measuring the TP-Rate, FP-Rate, Precision, Recall, and MCC of all selected data mining algorithms with 10% fold cross-validation are shown in Figure 1.

The results of comparing the performance of algorithms based on the TP-Rate, FP-Rate, Precision, Recall, and MCC showed that the Bayes Net with the FP-Rate of 0.147 better than the other algorithms had the higher performance in classifying the samples. The J-48 and Bayes Net with the Precision of 0.85 had the best performance in categorizing the sample, but generally, the J-48 data mining algorithm with the TP-Rate, FP-Rate, Precision, Recall, and MCC of 0.85, 0.173, 0.85, 0.85, and 0.68 had better performance than the other algorithms, respectively. Figure 2 shows the F-Score obtained from different data mining algorithms.

Figure 2 shows that the F-score obtained from the J-48 data mining algorithm was higher than the other algorithms (0.85); therefore, it had a better performance than other algorithms. The Bayes Net (0.761) had the worst performance in this respect. The results of the AUC of each of the data mining algorithms are shown in Figure 3. Horizontal and vertical vertices show a true-positive rate (TPR) and a false-positive rate, respectively.

All algorithms used in this study had a ROC value above 0.9 (except the SMO algorithm with AUC = 0.80) and the
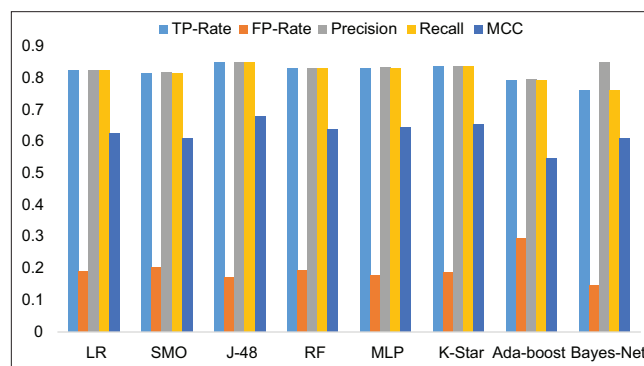

Figure 1: Comparing the performance of different data mining algorithms

Table 1: The most important COVID-19 diagnostic criteria

Variable name	Variable type	Features	χ^2
Lung lesion existence	Binominal	Haven't Have	179.21
Fever	Binominal	Haven't Have	113.26
Contact with suspected people	Binominal	Haven't Have	111.26
O_2 saturation in the blood	Poly nominal	>95% 85%-95% <85%	102.4
Rhinorrhea	Binominal	Haven't Have	96.4
Dyspnea	Binominal	Haven't Have	90.1
Digestive sign (diarrhea)	Binominal	Haven't Have	81.7
Nausea and vomiting	Binominal	Haven't Have	75.5
Traveling to high-risk area history	Binominal	Haven't Have	63.3
History of use of immunosuppressive drugs	Binominal	Haven't Have	58.2
History of respiratory failure	Binominal	Haven't Have	46.6
History of respiratory tract infection	Binominal	Haven't Have	40.2
Cough	Binominal	Haven't Have	34.2
History of taking Vitamin D	Binominal	Haven't Have	33.6
Disability sensation	Binominal	Haven't Have	32.9
Chest pain	Binominal	Haven't Have	31.1
Tremor	Binominal	Haven't Have	30.8
Age	Poly nominal	Young Middle-aged Adult	29.7
Headache	Binominal	Haven't Have	27.6
Consciousness	Poly nominal	Complete Relative Without	27.3
Throat pain	Binominal	Have Haven't	25.5

classification of the research samples was close to TPR, which indicated the high classification strength of the algorithms, but the J-48 classification algorithm with a ROC value of 0.917 with a slight difference compared to the other algorithms had a better performance, relatively.

Comparing the performance of the data mining algorithms based on the different evaluation criteria such as TP-rate, FP-rate, precision, recall, MCC, F-score, and ROC obtained from different algorithms,

generally, the J-48 data mining algorithm had a higher performance than other algorithms. Figure 4 shows a pruned J-48 decision tree with the features described below along with a maximum algorithm efficiency based on the most important diagnostic criteria were considered by the tree.

Some important characteristics of the J-48 decision tree algorithm method had been used for sample classification with the highest performance.

The J-48 important characteristics for making tree

Batch size = 100, confidence factor = 0.2, minimum number of objects = 2, number of folds = 3, and number of seeds = 1.

Based on the decision tree drawn, the root node is the presence of lung lesions in the individual, which was considered as the most important diagnostic criterion based on the Chi-square. Based on this decision tree, all COVID-19 disease diagnostic rules were extracted with 85% diagnostic accuracy. The simplest rule derived from this tree was that if a person had a lung lesion, he or she was considered as a person with COVID-19 disease, which was confirmed by the 148 samples of the research. The interpretation of the other rule derived from this decision tree was that if a person had no lung lesions, fever, and blood oxygen saturation >95%, the decision tree algorithm would classify this person with 89% probability in the group with non-COVID-19 disease. In this leaf node, 85 samples were classified, from which the 9 samples were classified incorrectly, and therefore, 76 samples confirmed the diagnosis of non-COVID-19 disease (89%).

Figure 2: The comparison of the F-Score of the different data mining algorithms

Discussion

Today, COVID-19 disease has become a universal problem that has endangered the health, social, economic, and developmental conditions of many countries; thus, WHO in its report has identified the disease as a serious threat to community health.[21,22] In such circumstances, the need to use new and advanced technologies to monitor disease trends, track contacts, and identify the disease's patterns as well as to prevent, screen, treat, and monitor the patients has become more realized.[23,24] To counter this threat to public health, many countries have emphasized the need to use technological infrastructure.[25] During the COVID-19 epidemic, the promotion of public health potential in the fight against coronavirus requires attention to the fact that the cause of problems and risks threatening public health, as well as solutions to these problems must follow a complex technological adaptation system.[26] The digital health technology revolution in the early 2020s was marked by the epidemic of coronavirus due to the widespread use of
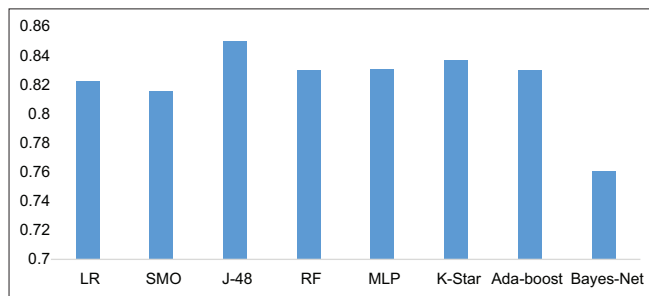
Table 2: The results of sample classification with COVID-19 disease and non-COVID-19

Algorithm	TP	FP	FN	TN
MLP	211	39	29	121
J-48	221	29	31	119
Bayes net	155	95	1	149
LR	211	39	32	118
Ada-boost	231	19	64	86
RF	216	34	34	116
K-star	218	32	33	117
SMO	210	40	32	116

MLP=Multilayer perceptron, LR=Logistic regression, RF=Random forest, SMO=Sequential minimal optimization, TP=True positive, FP=False positive, FN=False negative, TN=True negative
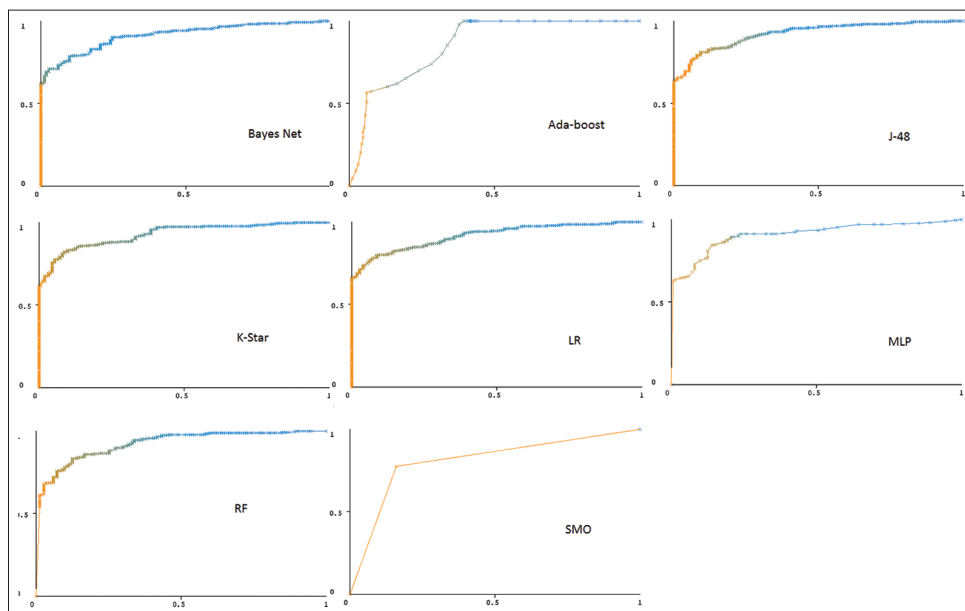

Figure 3: The receiver operator characteristics of different data mining algorithms

```
J48 pruned tree
------------------

Lung-lesion = No
|   Fever = No
|   |   SPO2 = >95: 0 (85.0/9.0)
|   |   SPO2 = 85-95
|   |   |   Cough = No: 1 (5.0/1.0)
|   |   |   Cough = Yes: 0 (10.0/2.0)
|   |   SPO2 = <85: 0 (0.0)
|   Fever = Yes
|   |   History-of-contact = No
|   |   |   Cough = No: 0 (32.0/6.0)
|   |   |   Cough = Yes
|   |   |   |   Diarrhea = No
|   |   |   |   |   Dyspnea = No: 0 (16.0/5.0)
|   |   |   |   |   Dyspnea = Yes
|   |   |   |   |   |   Nausea-vomiting = No: 0 (3.0/1.0)
|   |   |   |   |   |   Nausea-vomiting = Yes: 1 (3.0)
|   |   |   |   Diarrhea = Yes: 1 (25.0/8.0)
|   |   History-of-contact = Yes
|   |   |   Dyspnea = No: 1 (61.0/13.0)
|   |   |   Dyspnea = Yes
|   |   |   |   Rhinorhea = No: 1 (6.0)
|   |   |   |   Rhinorhea = Yes: 0 (6.0/1.0)
Lung-lesion = Yes: 1 (148.0)
```
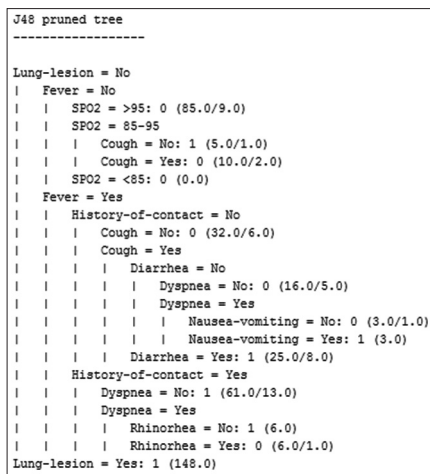
Figure 4: The pruned J-48 algorithm in diagnosing COVID-19

artificial intelligence, cloud computing, IoT, and big data mining technologies.[27] Today, the use of innovative and advanced computing technologies in the form of machine learning systems, deep learning, and big data mining, provides deep, effective, and noninvasive analytical capabilities to support the decision of physicians and health policymakers over traditional statistical methods, clinical trials, and evaluations.[28] With the spread of COVID-19 disease and the widespread involvement of many communities, the health-care systems of many countries will not be able to meet the growing needs of patients to diagnose, treat, and care for. In such circumstances, the design and implementation of clinical decision-making systems equipped with custom and effective predictive algorithms will be critical to identify patients and differentiate them from healthy individuals as well as individuals with other similar respiratory diseases.[29] Due to the sensitive, ambiguous, and multidimensional nature of the disease and also the lack of specific treatment, long incubation period, relatively difficult paraclinical evaluations, and differential diagnosis, clinical decision technologies play an important role in improving the quality of diagnosis and provide customized care. Furthermore, increasing the number of infected cases, rapid and effective identification of patients, proper prioritization of health resources, screening, and implementation of quarantine programs help reduce the burden of the disease.[30] Therefore, the main purpose of this study was to evaluate the performance of some data mining models such as MLP, J-48 Decision Tree, Bayes Net, LR, Ada-boost, SMO, RF, and K-star for diagnosing COVID-19 disease and differentiating between infected and healthy individuals to improve the quality and effectiveness of clinical decisions. In the field of data mining applications (machine learning and deep learning), several types of research have been done, among which the comparison of different algorithms to suggest the most efficient ones present an optimal diagnostic model

is on the agenda. The results of Alakus study (2020) to design a diagnostic model of COVID-19 based on deep learning showed the precision, F1 score, accuracy, recall, and AUC of 88.66%, 91.89%, 86.75%, 99.42%, and 62.50% of the system to diagnose the disease, respectively.[31] In the study of Narin *et al.*, the performance of three types of artificial neural networks (ANNs) of ResNet50, InceptionV3, and Inception-ResNetV2 was compared to diagnose the COVID-19 disease. The results showed the ANN of ResNet50 with 98% accuracy was more efficient than the other data mining algorithms.[32] In the study of Elaziz *et al.*, they used the machine learning method to analyze the chest image in individuals and the algorithm was able to identify people with COVID-19 disease and healthy individuals based on chest image analysis, the results of the algorithm evaluation showed that the accuracy of 96% and 98% in the two different data sets of images.[33] In a study conducted by Brunese *et al.*, an interpretive deep learning method was developed for the rapid detection of lung and COVID-19, by performing the statistical analysis on 6523 chest photographs belonged to the different institutions, the results of the study showed a 97% accuracy in 2.5 s in diagnosing the disease.[34] In Rodriguez's study, by comparing the two mathematical methods (Gompertz and logistic), and a computational method (ANN), the results showed better performance of the computational models in diagnosing COVID-19.[35] In the present study, based on comparing the performance of the data mining algorithms considering the different evaluation criteria such as TP-Rate, FP-Rate, Precision, Recall, MCC, F-Score obtained, the J-48 data mining algorithm with 0.85, 0.173, 0.85, 0.85, 0.68, and 0.85 of these evaluation criteria, respectively, had a higher diagnostic performance than the other algorithms.

In this research, by introducing a noninvasive, effective, and evidence-based method, we will be able to identify and propose the best diagnostic model for COVID-19 based on the most effective diagnostic criteria. The results of evaluating the performance of the different data mining techniques showed that the proposed diagnostic models based on machine learning have a higher performance than the traditional paraclinical approaches, and also can be used by physicians to improve their diagnostic performance.

Limitation and suggestion
This study had several limitations that need to be addressed. First, this is a backward-looking study that there was some documented data that was erroneous, irregular, uneven, or imbalanced. Thus, we balanced them by manual checking of noisy and abnormal values, errors, duplicates, and meaningless data. Furthermore, by using the Synthetic Minority Oversampling Technique, the bias can be minimized via class balancing. Furthermore, the missing values

were imputed with the mean or mode of each variable. To solve the imbalanced data set problem, in which the number of records related to the positive class is significantly lower than the negative, different criteria were measured to assess the performance of each ML algorithm. Second, this was a retrospective study based on a single-center registry system with a limited number of samples. This may limit the generalizability of the proposed model. Nevertheless, we will use a multicenter dataset to perform an external validation of the proposed model to augment the generalized prediction. In the future, the performance of our computational model will be improved if we test more classification techniques at multiorganization datasets. Furthermore, larger cohorts, prospective settings, and clinical trials are needed before elucidating its contribution to improving the outcome of COVID-19.

Conclusion

Given the limited capacity of the health-care industries to deal with the current pandemic, effective prognosis, diagnosis, and triage of patients can help in properly managing, scheduling, and maximize the use of restricted hospital resources. Designing a true and valid diagnostic model may be improving the quality of care and increasing the survival rate of the patients. Therefore, the diagnostic model can greatly contribute to identifying high-risk patients and the adoption of the most effective support and treatment plans. This study may assist clinicians in enabling early detection, effective intervention, and possibly a decrease in the death of COVID-19 patients. This led to decreasing ambiguity by offering quantitative, objective, and evidence-based models for risk stratification, prediction, and eventually episode of the care plan. It offers a better strategy for clinicians to lessening the complications and improved patient survival likelihoods.

Conflicts of interest
There are no conflicts of interest.

References

1. Bikdeli B, Talasaz AH, Rashidi F, Sharif-Kashani B, Farrokhpour M, Bakhshandeh H, *et al*. Intermediate versus standard-dose prophylactic anticoagulation and statin therapy versus placebo in critically-ill patients with COVID-19: Rationale and design of the INSPIRATION/INSPIRATION-S studies. Thromb Res 2020;196:382-94.

2. Tang D, Comish P, Kang R. The hallmarks of COVID-19 disease. PLoS Pathog 2020;16:e1008536.

3. Mehta P, McAuley DF, Brown M, Sanchez E, Tattersall RS, Manson JJ, *et al*. COVID-19: Consider cytokine storm syndromes and immunosuppression. Lancet 2020;395:1033-4.

4. Shereen MA, Khan S, Kazmi A, Bashir N, Siddique R. COVID-19 infection: Origin, transmission, and characteristics of human coronaviruses. J Adv Res 2020;24:91-8.

5. Zhao Z, Chen A, Hou W, Graham JM, Li H, Richman PS, *et al*. Prediction model and risk scores of ICU admission and mortality in COVID-19. PLoS One 2020;15:e0236618.

6. Hu H, Yao N, Qiu Y. Comparing rapid scoring systems in mortality prediction of critically ill patients with novel coronavirus disease. Acad Emerg Med 2020;27:461-8.

7. Thomson G. COVID-19: Social distancing, ACE 2 receptors, protease inhibitors and beyond? Int J Clin Pract 2020;74:e13503.

8. Xu G, Yang Y, Du Y, Peng F, Hu P, Wang R, *et al*. Clinical pathway for early diagnosis of COVID-19: Updates from experience to evidence-based practice. Clin Rev Allergy Immunol 2020;59:89-100.

9. Ayyoubzadeh SM, Ayyoubzadeh SM, Zahedi H, Ahmadi M, R Niakan Kalhori S. Predicting COVID-19 incidence through analysis of google trends data in Iran: Data mining and deep learning pilot study. JMIR Public Health Surveill 2020;6:e18828.

10. James P, Das R, Jalosinska A, Smith L. Smart cities and a data-driven response to COVID-19. Dialogues Hum Geogr 2020;10:255-9.

11. Peck KR. Early diagnosis and rapid isolation: Response to COVID-19 outbreak in Korea. Clin Microbiol Infect 2020;26:805-7.

12. Prabu S, Velan B, Nelson SC, Jayasudha FV, Visu P, Janarthanan K. Mobile technologies for contact tracing and prevention of COVID-19 positive cases: A cross- sectional study. Int J Pervasive Comput Commun 2020 Aug 18; 8:185-93 [doi: 10.1108/IJPCC-07-2020-0086].

13. Shaban WM, Rabie AH, Saleh AI, Abo-Elsoud MA. A new COVID-19 Patients Detection Strategy (CPDS) based on hybrid feature selection and enhanced KNN classifier. Knowl Based Syst 2020;205:106270.

14. Syed-Abdul S, Hsu MH, Iqbal U, Scholl J, Huang CW, Nguyen PA, *et al*. Utilizing health information technology to support universal healthcare delivery: Experience of a National Healthcare System. Telemed J E Health 2015;21:742-7.

15. Urbaczewski A, Lee YJ. Information technology and the pandemic: A preliminary multinational analysis of the impact of mobile tracking technology on the COVID-19 contagion control. Eur J Inf Syst 2020;29:405-14.

16. Lavrač N, Zupan B. Data Mining in Medicine. Data Mining and Knowledge Discovery Handbook. Boston: Springer; 2005. p. 1107-37. [doi: 10.1007/0-387-25465-X_52].

17. Zhang Y, Xin Y, Li Q, Ma J, Li S, Lv X, *et al*. Empirical study of seven data mining algorithms on different characteristics of datasets for biomedical classification applications. Biomed Eng Online 2017;16:125.

18. Pan P, Li Y, Xiao Y, Han B, Su L, Su M, *et al*. Prognostic assessment of COVID-19 in the intensive care unit by machine learning methods: Model development and validation. J Med Internet Res 2020;22:e23128.

19. Shanbehzadeh M, Nopour R, Kazemi-Arpanahi H. Comparison of four data mining algorithms for predicting colorectal cancer risk. J Adv Med Biomed Res 2021;29:100-8. [doi: 10.30699/jambs.29.133.100].

20. Albahri AS, Hamid RA, Alwan JK, Al-Qays ZT, Zaidan AA, Zaidan BB, *et al*. Role of biological data mining and machine learning techniques in detecting and diagnosing the novel

coronavirus (COVID-19): A systematic review. J Med Syst 2020;44:122.

21. Foddai A, Lubroth J, Ellis-Iversen J. Base protocol for real time active random surveillance of coronavirus disease (COVID-19)-Adapting veterinary methodology to public health. One Health 2020;9:100129.

22. Xu T, Chen C, Zhu Z, Cui M, Chen C, Dai H, *et al*. Clinical features and dynamics of viral load in imported and non-imported patients with COVID-19. Int J Infect Dis 2020;94:68-71.

23. Shipe ME, Deppen SA, Farjah F, Grogan EL. Developing prediction models for clinical use using logistic regression: An overview. J Thorac Dis 2019;11:S574-84.

24. Gao Y, Cai GY, Fang W, Li HY, Wang SY, Chen L, *et al.* Machine learning based early warning system enables accurate mortality risk prediction for COVID-19. Nat Commun 2020;11:1-10.

25. Xiao C, Zheng L, Chen F, Xiao Y. Design and research of a smart monitoring system for 2019-nCoV infection-contact isolated people based on blockchain and internet of things technology. Reasearch Square. 2020;6(3):e19399. [doi: 10.21203/rs. 3.rs-18678/v1].

26. Allam Z, Jones DS. On the coronavirus (COVID-19) outbreak and the smart city network: Universal data sharing standards coupled with artificial intelligence (AI) to benefit urban health monitoring and management. Healthcare (Basel) 2020;8:46.

27. Bayram M, Springer S, Garvey CK, Özdemir V. COVID-19 digital health innovation policy: A portal to alternative futures in the making. OMICS 2020;24:460-9.

28. Agieb R. Machine learning models for the prediction the necessity of resorting to icu of covid-19 patients. Int J Adv Trends Comput Sci Eng 2020;9(5): 6980-4. [doi: 10.1161/STROKEAHA.118.024293].

29. Govindan K, Mina H, Alavi B. A decision support system for demand management in healthcare supply chains considering the epidemic outbreaks: A case study of coronavirus disease 2019 (COVID-19). Transp Res E Logist Transp Rev 2020;138:101967.

30. Bredmose PP, Diczbalis M, Butterfield E, Habig K, Pearce A, Osbakk SA, *et al*. Decision support tool and suggestions for the development of guidelines for the helicopter transport of patients with COVID-19. Scand J Trauma Resusc Emerg Med 2020;28:43.

31. Alakus TB, Turkoglu I. Comparison of deep learning approaches to predict COVID-19 infection. Chaos Solitons Fractals 2020;140:110120.

32. Narin A, Kaya C, Pamuk Z. Automatic detection of coronavirus disease (COVID-19) using x-ray images and deep convolutional neural networks. arXiv 2020;3(5):10849. [doi: 10.1080/07391102.2021.1875049].

33. Elaziz MA, Hosny KM, Salah A, Darwish MM, Lu S, Sahlol AT. New machine learning method for image-based diagnosis of COVID-19. PLoS One 2020;15:e0235187.

34. Brunese L, Mercaldo F, Reginelli A, Santone A. Explainable deep learning for pulmonary disease and coronavirus COVID-19 detection from x-rays. Comput Methods Programs Biomed 2020;196:105608.

35. Torrealba-Rodriguez O, Conde-Gutiérrez RA, Hernández-Javier AL. Modeling and prediction of COVID-19 in Mexico applying mathematical and computational models. Chaos Solitons Fractals 2020;138:109946.