Journal of Environmental Health and Sustainable Development

Proposing an Intelligent Monitoring System for Early Prediction of Need for Intubation among COVID-19 Hospitalized Patients

Mohammad Reza Afrash [1], Hadi Kazemi-Arpanahi [2,3], Raoof Nopour [4], Elmira Sadat Tabatabaei [5], Mostafa Shanbehzadeh [6]*

[1] *Department of Medical Informatics, Student Research Committee, School of Allied Medical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran.*
[2] *Department of Health Information Technology, Abadan University of Medical Sciences, Abadan, Iran.*
[3] *Department of Student Research Committee, Abadan University of Medical Sciences, Iran.*
[4] *Department of Health Information Management, Student Research Committee, School of Health Management and Information Sciences Branch, Iran University of Medical Sciences, Tehran, Iran.*
[5] *Department of Genetics, Islamic Azad University, Tehran Medical Branch, Tehran, Iran.*
[6] *Department of Health Information Technology, School of Paramedical, Ilam University of Medical Sciences, Ilam, Iran.*

ARTICLE INFO

ORIGINAL ARTICLE

Article History:
Received: 25 May 2022
Accepted: 10 July 2022

***Corresponding Author:**
Mostafa Shanbehzadeh

Email:
mostafa.shanbezadeh@gmail.com

Tel:
+989300833691

Keywords:
COVID-19,
Coronavirus,
Artificial Intelligence,
Machine Learning,
Intubation,
Prognosis.

ABSTRACT

Introduction: Predicting acute respiratory insufficiency due to coronavirus disease 2019 (COVID-19) can diminish the severe complications and mortality associated with the disease. This study aimed to develop an intelligent system based on machine learning (ML) models for frontline clinicians to effectively triage high-risk patients and prioritize who needs mechanical intubation (MI).

Materials and Methods: In this retrospective-design study, the data regarding 482 COVID-19 hospitalized patients from February 9, 2020, to July 20, 2021, was analyzed by six ML classifiers. The most critical clinical variables were identified by a minimal-redundancy-maximal-relevance (mRMR) feature selection technique. In the next step, the models' performance was assessed using confusion matrix criteria and, finally, the best model was adopted.

Results: Proposed models were implemented using 23 confirmed variables. Results of comparing six selected ML algorithms indicated the extreme gradient boosting (XGBoost) classifier with 84.7% accuracy, 76.5 % specificity, 90.7% sensitivity, 85.1% f-measure, 87.4% Kappa statistic, and 85.3% for receiver operating characteristic (ROC) had the best performance in the intubation prediction.

Conclusion: It is found that ML enables a satisfactory accuracy level in calculating intubation risk in COVID-19 patients. Therefore, using the ML-based intelligent models, notably the XGBoost algorithm, actually enables recognizing high-risk cases and advising correct therapeutic and supportive care by the clinicians.

Citation: Afrash MR, Kazemi-Arpanahi H, Nopour R, et al. ***Designing an Intelligent Decision Support System for Early Prediction of Intubation Need among COVID-19 Hospitalized Patients.*** J Environ Health Sustain Dev. 2022; 7(3): 1698-707.

Introduction

Coronavirus disease 2019 (COVID-19) has been puzzling for millions of people globally [1] due to its extra contagious power, high mortality rate, lack of effective medicine options, slow pace of vaccination, privation of sufficient capacity in hospitals, intensive care unit (ICU) beds allocation and medical staff exhaustion [2]. Furthermore, many

[DOR: 20.1001.1.24766267.2022.7.3.1.4]
[DOI: 10.18502/jehsd.v7i3.10719]

[DOR: 20.1001.1.24766267.2022.7.3.1.4]

[DOI: 10.18502/jehsd.v7i3.10719]

healthcare systems face severe challenges in care management and resource utilization[3]. Most of the COVID-19 patients have mild symptoms, but approximately 15% to 20% of symptomatic people onrush to intense pneumonia disease requiring hospitalization. The critical phase of COVID-19 is specified by intense complications such as acute respiratory insufficiency, multi-organ failure, loss of consciousness, coma, and even death[4].

The COVID-19 patients with ARDS require medical intubation and supplemental oxygen [5]. Therefore, there is a critical demand for detecting patients who need mechanical intubation (MI) facilities. More clearly, to manage the respiratory ventilator scarceness, a precise clinical decision is necessary for the triage of cases in order to assure proper and effective intubation services[6]. During the COVID-19 pandemic, the need for reliable and evidence-based decision-making is essential, especially where the health care system is facing a growing number of hospitalization and a shortage of critical hospital resources [7].

Recently, researchers have shown great interest in introducing novel and intelligent digital technologies such as artificial intelligence (AI) that can effectively detect the patient at the risk of clinical deterioration [8]. It is proven that these methods can minimize diagnostic errors and disagreements between observers at any level of prediction, prognosis and treatment. Moreover, it may make it easier for at-risk cases to detect and implement the most operative supportive and treatment plans [9]. Machine learning (ML) can be utilized to foretell future intubation risks based on data gathered and available in clinical environments routinely. ML-equipped CDSS can help healthcare providers by rendering effective recommendations [10, 11]. ML techniques in various areas are used for COVID-19 management, such as early detection and screening, disease diagnosis as well as identifying and predicting patient deterioration [12]. Thus, the present study aimed to train and compare selected ML classifiers for the prediction of MI risk among COVID-19 hospitalized patients.

Materials and Methods

This developmental study was performed in the form of a retrospective and mono-center analysis between February 2020 and July 2021. The study aimed to accurately predict the intubation risk among COVID-19 hospitalized patients by using six popular ML algorithms and selecting the best performing.

Study mind map and environment

The study mind map of the preferred system was classified into five stages: 1- dataset description, 2- data preprocessing, 3-feature selection, 4- prediction models, and 5- models performance assessment metrics. All classifiers were implemented using Python programming language (version 3.7.7) in three phases of preprocessing, training, and evaluation. Programming language was used for designing the CDSS user interface and developing the modules. The study mind map is shown in Figure1.

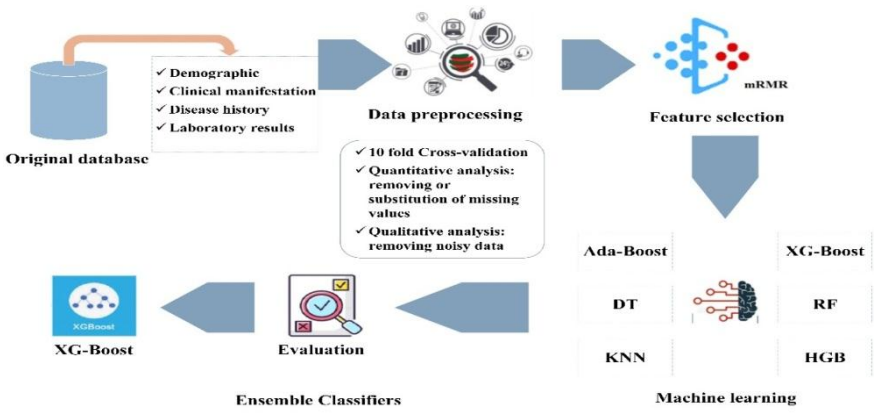

Figure 1: The study mind map

Dataset description

This study retrospectively reviewed a COVID-19 registry database at Ayatollah Taleghani Hospital in Abadan, Iran, from 2020-02-09, to 2021-07-20. In this time span, 36,854 suspected COVID-19 people who referred to this center (14,800 positive COVID-19 cases, 19,525 healthy individuals, and 2,529 unspecified cases). After applying the exclusion criteria, 9,530 hospitalized case records remained: 9,000 belonged to non-intubated and 530 were associated with the intubated cases. The exclusion criteria for the patient selection were: 1-non-COVID-19 individuals, 2-non-hospitalized COVID-19 patients, 3-patients who were under 18 years of age, 4-case records with missing more than 70%) and 5-admission time either before or after the defined time span. Based on Table 1, the number of 60 clinical features in six classes including 1-patients' demographic data (six features), manifestations (14 features), patient history (eight features), laboratory tests (28 features), remedies (one feature), imaging indicators (two features) and output (0: non-intubation and 1: intubation) were derived from the registry database.

Data preprocessing

The data preprocessing phase is vital for effectively enhancing the data quality and data mining performance. Preprocessing methods, such as deleting missing values, minimal scalar and standard scalar, were applied to the dataset. In the dataset used in this paper, incomplete case records with missing more than 70% were removed from the analysis. Noisy, duplicate and meaningless data were investigated by two authors in cooperation with one infectious diseases specialist.

Feature selection

For feature selection in this paper, we used the minimal-redundancy-maximal-relevance (mRMR) technique. This method selects the most relevant variable for prediction intubation in the patients with COVID-19.

Classifiers

Some ML-based models, including decision tree (DT), Hist gradient boosting (HGB), extreme gradient boosting (XGBoost), k-nearest neighbors (K-NN), and random forest (RF) were used in this study for creating the intubation models. The grid search method was employed to optimize the hyperparameters of the applied ML techniques. The grid search method is an efficient method for setting the parameters in the training stage of prediction models and enhancing the generalization efficiency of an algorithm [13]. The developed algorithms are described below:

Adaboost: This algorithm is an iterative ensemble method which constructs a robust learning algorithm by uniting multiple week learners to achieve high accuracy. The idea behind Adaboost is to set the multiple weights of ML algorithms and train the data in each iteration, hence it warrants the accurate prediction of unusual observations. Adaboost is of significance in the situation where developing a robust learning algorithm directly is a too hard task [14].

XGBoost: It is a library of gradient boosting algorithms developed to adjust the errors created by the existing classifiers. Some of the main returns of XGBoost are its high scalability, so that it runs more quickly than other ML methods and uses less memory [15].

HBG: The HGB algorithm is more highly effective when there is a great size of dataset. Also, this method decreases training time without reducing the accuracy level. Indeed, HGB is an algorithm for quickly training DTs used in the gradient boosting ensemble. As the HGB algorithm is more efficient in both memory consumption and training speed, we used it in our study.

DT: The DT algorithm is a tree-structured model that starts from the root node, continues through the internal nodes and, then, reaches the external nodes or leaves. Some benefits of this algorithm are the ability to classify numerical and qualitative variables, better understand knowledge discovery through simple structure, extract rules with the if-then structure and achieve higher performance with better classification ability of the study samples. But, we may face the overfitting process during training time due to the increased

[DOR: 20.1001.1.24766267.2022.7.3.1.4]
[DOI: 10.18502/jehsd.v7i3.10719]

understanding of these algorithms [16-18].

RF: This algorithm is suitable for classifying a tremendous amount of data cases. The splitting process occurs randomly in the classification process. This algorithm is considered a more advanced technique for using multiple sub-algorithms to optimize the classification capability. The RF uses the council mode of sub-algorithms to obtain the performance; in other words, the most frequent algorithms with better classification capability are considered for the classification process. One significant flaw of this algorithm is the "black box" phenomenon, referring to the complexity of the algorithm's structure and challenging interpretation [19-21].

KNN: The KNN is the classifier and predictor algorithm and can be regarded as the semi-supervised algorithm. It predicts the output classes through similar values in near cases with specified Euclidean distance quantification. Some advantages of this algorithm are the potential of working on a high-dimensional dataset due to high speed of performance, high capability of prediction with the straightforward interpretation of relationships between the elements of the datasets and simple implementation. Due to the instance-based way of prediction associated with this algorithm, poor performance capability is one of the disadvantages, especially in large datasets [22-24].

Validation method of classifiers

In the present study, to evaluate and compare the capability of selected classifiers in predicting intubation, we applied the k-fold cross-validation and confusion matrix performance assessment metrics. In the k-fold cross-validation method, the dataset is split into k equal sizes of sections. In this regard, K-1 part (tenfold-cross-validation: 10-1) is used to learn the predictors and the remaining section is applied to test the classifier's performance in each phase.

Performance evaluation metrics

We assessed the performance and effectiveness of six ML classifiers in terms of accuracy (overall number of cases classified correctly), sensitivity (total amount of positive cases classified correctly), specificity (proportion of true negatives classified correctly), F-measure(probability that a positive prediction is correct), Kappa indicator (comparison of the observed versus expected accuracy), receiver operating characteristic (ROC) rate, time to build the model and number of correctly and incorrectly classified cases (Equation 1-5).

$$1)\ \text{classification accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \times 100$$

$$2)\ \text{classification sensitivity} = \frac{TP}{TP+FN} \times 100$$

$$3)\ \text{classification specificity} = \frac{TN}{TN+FP} \times 100$$

$$4)\ f-\text{measure} = 2\ \frac{\text{precision} * \text{sensitivity}}{\text{precision} + \text{sensitivity}}$$

$$5)\ \text{kappa statistic} = \frac{P0-PE}{1-PE}$$

In these formulas, True Positive (TP) and True Negative (TN) point to the COVID-19 and Non-COVID-19 cases correctly categorized by the model. Also, False Negative (FN) and False Positive (FP) refer to wrongly classified samples.

Ethical Issue

This study is extracted from the research project with ethical code of IR.ABADANUMS.REC.1400.071 that was accepted by the Ethics Committee of Abadan University of Medical Sciences.

Results

Feature selection by mRMR method

mRMR algorithm selects the most important feature according to feature weight. We experimented with different numbers of variables (60 features) and the result represented the performance of predictors is high based on 24 variables. In this study, we only reported the results of prediction algorithms on 24 variables in our simulation results. Figure 2 shows the scores obtained for features by the mRMR algorithm. The selected 24 most important variables by MRMR are given in Figure 2.

[DOR: 20.1001.1.24766267.2022.7.3.1.4]
[DOI: 10.18502/jehsd.v7i3.10719]

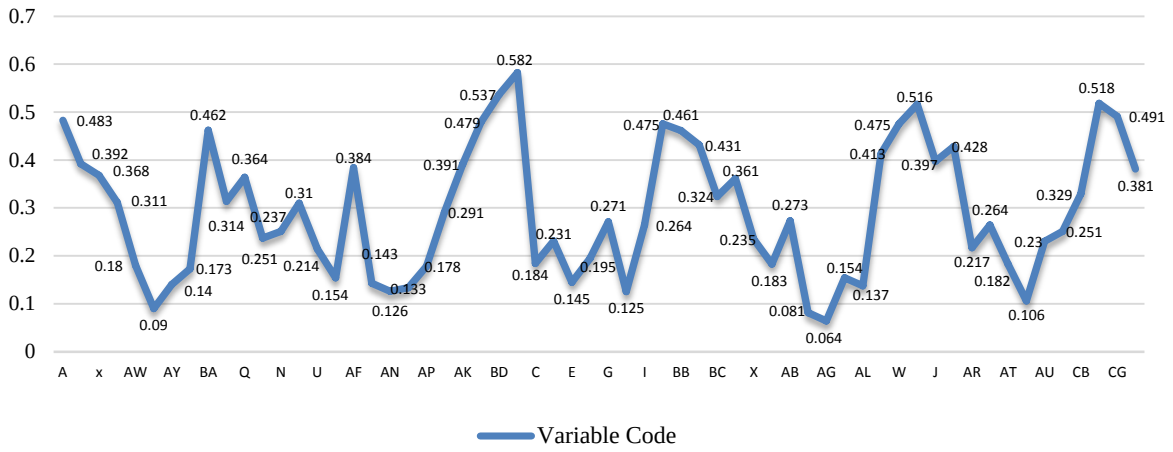

Figure 2: Scores for all dataset features based on the mRMR algorithm

K-fold cross-validation on dataset with the selected feature

In this research, the features selected by mRMR were tested on six ML techniques using a tenfold-cross-validation method. In the ten-fold cross-validation, 90% of the dataset was utilized for learning algorithms and 10% for testing classifiers. For better performance, the actual performance of the classifiers and mean for included metrics of tenfold cross-validation were measured. Table 2 shows the ten-fold cross-validation of six models' performance based on the best features for predicting intubation. For better presentation of the results, the confusion matrix and ROC curve of the best ML-based predictive model on the selected dataset are displayed in Figure 3.

According to Table 2, the XGBoost classifier with 84.7% accuracy, 76.5 % specificity, 90.7% sensitivity, 85.1% f-measure, 87.4% Kappa statistic and 85.3% for ROC metrics attained the best performance in predicting intubation risk among hospitalized patient with COVID-19. (See Figure 3).

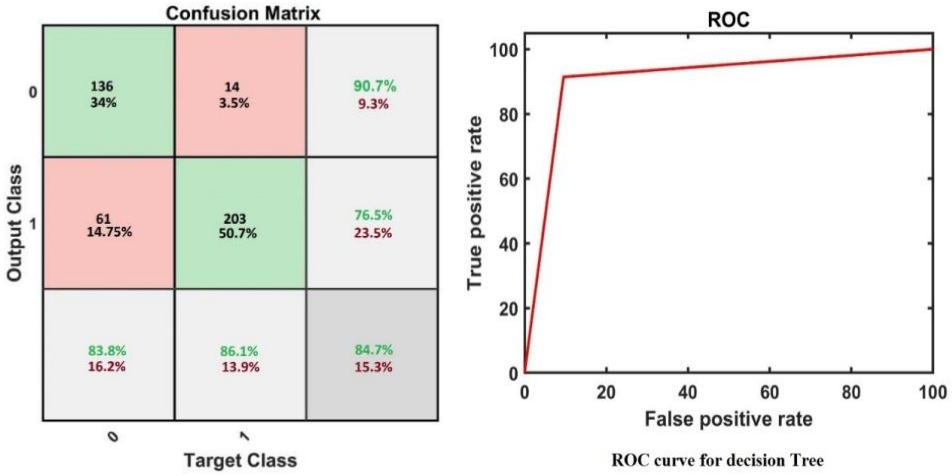

Figure 3: Confusion matrix and ROC obtained for XGBoost classifier

Based on the confusion matrix depicted in Figure 2 and the information obtained from Table 2, the 83.8% precision rate was obtained and the value for accuracy was 84.7%.

System implementation

The CDSS for the intubation prediction was implemented during July - August 2021. All the prediction classifiers were implemented using Python programming language by scikit-learn library version 0.23.2. The system programming included three kinds of implementation codes: designing the user interface, logic layer

[DOR: 20.1001.1.24766267.2022.7.3.1.4]
[DOI: 10.18502/jehsd.v7i3.10719]

implementation and codes associated with the database. Our study's user interfaces comprised two pages: the welcome page and the CDSS module (two pages). The developed CDSS menu-driven interface was developed using C# programming language in the visual studio environment (Figure 4).

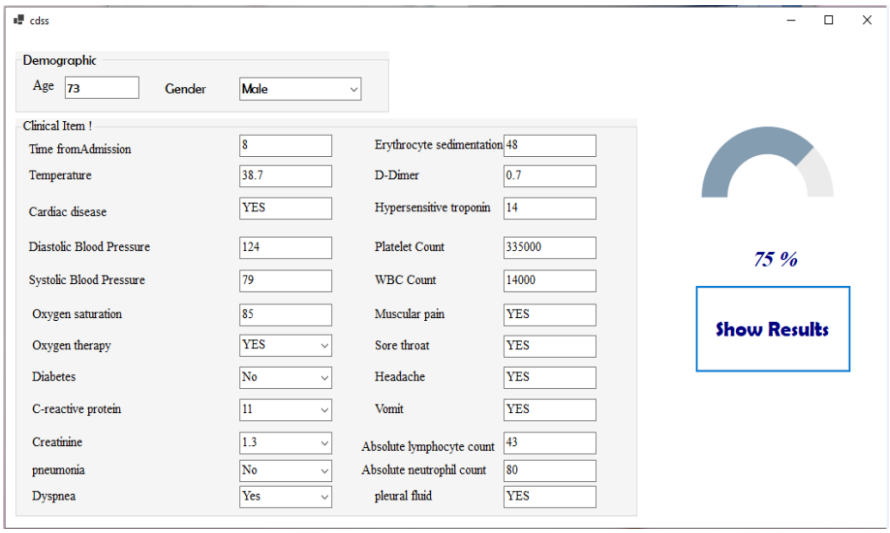

Figure 4: Module page of designed CDSS

Discussion

In this study, six popular ML algorithms such as DT, HGB, XGBoost, K-NN, RF and AdaBoost were trained using the most important features affecting the risk of COVID-19 intubation selected by performing mRMR feature selection. Several studies have also obtained the most important factors using various feature selection techniques to predict COVID-19 outcomes such as entering to ICU, need for MI, length of stay (LOS), readmission, and mortality. The important variables in the reviewed studies [25-32] included age, high BMI, dyspnea, low consciousness, fever, decreased SPO_2, low lymphocyte count, increased CRP, D dimer, ALT and AST, cardiovascular disease, cancer, pneumonia and chronic renal disease. In the present study, we used a combination of mRMR to score the important predictors. In this regard, the time of hospitalization (0.483), body temperature (0.392), cardiac disease (0.368), diastolic blood pressure (0.311), systolic blood pressure (0.329), oxygen saturation (0.518), oxygen therapy (0.491), diabetes (0.381), C-reactive protein (0.462), creatinine (0.314), pneumonia (0.364), dyspnea (0.475), erythrocyte sediment rate (0.461), D-Dimer (0.431), hypersensitive troponin (0.324), platelet count (0.361), white blood cell count (0.384), muscular pain (0.413), sore throat (0.475), headache (0.516), vomit (0.397), absolutes lymphocyte rate (0.428), absolutes neutrophil rate (0.391) and pleural fluid (0.479) were selected as the best set predictive features, respectively. In general, the high performance was obtained using the best variables in current studies, similar to other works. Our study proved that ML algorithms, especially XGBoost classifiers, increase analytical accuracy and diagnostic efficiency. Therefore, six classification algorithms were trained using selected features. Finally, the XGBoost with 84.7% accuracy, 76.5% specificity, 90.7% sensitivity, 85.1% f-measure, 87.4% Kappa statistic and 85.3% ROC metrics outperformed from others.

Presently, several researches have investigated the importance of data mining algorithms in detecting patient deterioration and severe complications of COVID-19 infection. Yadaw et al. (2020) studied 3,841 patient data to propose a prediction model through four ML algorithms for death anticipation. Finally, the XGBoost model

[DOR: 20.1001.1.24766267.2022.7.3.1.4]
[DOI: 10.18502/jehsd.v7i3.10719]

with the ROC of 0.91% gained the best performance [30]. Similarly, in a retrospective study, Ryan et al. developed the XGBoost model for COVID-19 mechanical ventilation (MV), ICU and mortality prediction with AUC-ROC of 91%, 82% and 87%, respectively [33]. Arvind's study resulted that the XGBoost with AUC = 0.83 and PRC = 0.32 was significantly better than the ROX index in predicting the intubation risk. However, their discoveries also found that ML can rapidly evaluate the patients for intubation requirements with the best performance and reduce mortality [34]. Domínguez-Olmedo et al. developed the CDSS system using the data mining approach to predicting COVID-19 intubation. The best result was belonged to the XGBoost with AUC-ROC = 97%, accuracy = 94%, F-score = 77%, sensitivity = 93% and specificity = 95% [35]. Aljouie et al. in their analysis, evaluated the functionality of four selected ML classifiers including the RF, linear support linear regression (LR), vector machine (SVM), and XGBoost to predict the COVID-19 outcome. Their study resulted that the XGBoost with an AUC of 0.81 was selected as the best model for predicting the need for intubation among COVID-19 hospitalized patients[36]. Burdicka et al. estimated the need for ventilation among COVID-19 patients using the XGBoost. The ML algorithm showed better performance using MEWS for predicting ventilation within 24 h. Their sensitivity was 90%, specificity 58% and AUC 86% [37]. In the study by Cobre et al., data from 5643 cases with positive and negative COVID-19 diagnostic tests were analyzed to predict the intensity of COVID-19 symptoms through the best-known data mining models. They concluded the XGBoost with 86% accuracy yielded better performance than other techniques [38]. The experimental results of the present work, similar to the reviewed studies, showed that the XGBoost classifier with the ROC of 85.3% had the best prediction ability for intubation risk in COVID-19 hospitalized patients. The XGBoost classifier is an extended and more advanced version of the gradient boosting algorithm. Compared to others, some advantages of the XGBoost algorithm can be noted as the ability to perform parallel processing, handle missing numerical values and splitting different levels to make decisions on the maximum depth. In this algorithm, optimization of hyper-parameters is essential and, in the conditions that these hyper-parameters are not adjusted, the problem of the overfitting process will occur [39, 40].

The results of our study can inform clinicians and hospital managers in a correctly, accurately, and rapid manner to assess the need for intubation among COVID-19 patients, reducing severe complications of the disease and consequently decreasing the mortality rates. Despite using a size dataset for developing ML algorithms, our ML models performed well, specifically the XGBoost model. Also, the CDSS designed in this study can help the frontline physicians for predicting the need for intubations in healthcare settings with its design straightforwardness, user-friend, and flexible manner. However, the developed models in our study showed satisfactory results in predicting the intubation risk among COVID-19 patients, but there are some limitations that must be considered. First, due to the retrospective dataset analyzed in this study, naturally, some records may have missing values, duplicate fields, noisy and meaningless values. Second, the used dataset extracted from a single-center hospital registry limits the generalizability of developed models.

In addition, we used just six ML techniques to predict the intubation risk in COVID-19 patients using clinical features available at the admission time; however, according to the study aim these features are sufficient, but using other para-clinical and imaging features obtained during hospitalization can improve the models' results. Finally, dynamic changes in some critical features need to be followed to early identify those patients who are exposed to deteriorating outcomes. In the future, the model's capability and its generalizability will be improved if we use more ML algorithms in a more significant, multi-center and prospective dataset, having more qualitative and validated data. Compared with traditional statistical risk analysis methods, the proposed intelligent ML-based prediction model in the

[DOR: 20.1001.1.24766267.2022.7.3.1.4]
[DOI: 10.18502/jehsd.v7i3.10719]

current study can precisely predict the deterioration risk of COVID-19 hospitalized patients. Combining this model with conventional risk analysis systems (such as critical care systems) can create more added values and identify patients' cases quickly and actively.

Conclusion

Given the significant challenges to ICU hospital resources during the COVID-19 epidemic, accurate estimates of the patients needing MI can provide vital guidance for patient prioritization and limited resource utilization. This paper analyzed hospital records and test models that could predict the need for MV in COVID-19 hospitalized patients based on 24 clinical features. These developed predictive models may have an advantage in providing better care, diminishing physician workload and reducing mortalities. In addition, early detection of such individuals may take planned measures for intubation and reduce some of the known risks associated with immediate intubation. Finally, the results disclosed satisfactory performance of the selected models, particularly the XGBoost model, indicating that adopting this model for CDSS designing is acceptable.

Acknowledgement

This study is extracted from the research project with ethical code of IR.ABADANUMS.REC.1400.071 that was approved by the research committee of Abadan University of Medical Sciences.

Funding

Abadan University of Medical Sciences

Conflict of interest

The authors declare that they have no conflict of interest.

References

1.Braam DH, Srinivasan S, Church L, et al. Lockdowns, lives and livelihoods: the impact of COVID-19 and public health responses to conflict affected populations-a remote qualitative study in Baidoa and Mogadishu, Somalia. Confl Health. 2021;15(1):47.
2.Supady A, Staudacher D, Bode C, et al. Hospital networks and patient transport capacity during the COVID-19 pandemic when intensive care resources become scarce. Critical Care. 2021;25(1):28.
3.Leclerc T, Donat N, Donat A, et al. Prioritisation of ICU treatments for critically ill patients in a COVID-19 pandemic with scarce resources. Anaesth Crit Care Pain Med. 2020;39(3):333-9.
4.Chow N, Fleming-Dutra K, Gierke R, et al. CDC COVID-19 Response Team. Preliminary estimates of the prevalence of selected underlying health conditions among patients with coronavirus disease 2019-United States, February 12–March 28, 2020. MMWR Morb Mortal Wkly Rep. 2020;69(13):382-6.
5.Hawkins A, Stapleton S, Rodriguez G, et al. Emergency tracheal intubation in patients with COVID-19: A single-center, retrospective cohort study. West J Emerg Med. 2021;22(3):678.
6.Zhang K, Jiang X, Madadi M, et al. DBNet: a novel deep learning framework for mechanical ventilation prediction using electronic health records. Proceedings of the 12th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics; 2021.
7.Yu S, Qing Q, Zhang C, et al. Data-Driven Decision-Making in COVID-19 Response: A Survey. IEEE Trans Comput Soc Syst. 2021;8(4):989-1002.
8.Khan M, Mehran MT, Haq ZU, et al. Applications of artificial intelligence in COVID-19 pandemic: A comprehensive review. Expert Syst Appl. 2021;185:115695.
9.Rodríguez-Rodríguez I, Rodríguez JV, Shirvanizadeh N, et al. Applications of artificial intelligence, machine learning, big data and the internet of things to the COVID-19 pandemic: A scientometric review using text mining. Int J Environ Res Public Health. 2021;18(16):8578.
10. Karthikeyan A, Garg A, Vinod PK, et al.

Machine learning based clinical decision support system for early COVID-19 mortality prediction. Front public health. 2021;12(9):626697.
11. Wu G, Yang P, Xie Y, et al. Development of a clinical decision support system for severity risk prediction and triage of COVID-19 patients at hospital admission: An international multicentre study. Eur Respir J. 2020;56(2):2001104.
12. Agieb R. Machine learning models for the prediction the necessity of resorting to icu of covid-19 patients. International Journal of Advanced Trends in Computer Science and Engineering. 2020;9(5):6980-4.
13. Eskandari A, Milimonfared J, Aghaei M. Optimization of SVM classifier using Grid Search Method for Line-Line Fault Detection of Photovoltaic Systems. Conf Rec IEEE Photovolt Spec Conf; 2020.
14. Ying C, Qi-Guang M, Jia-Chen L, et al. Advance and prospects of AdaBoost algorithm. Zidonghua Xuebao. 2013;39(6):745-58.
15. Sharma A, Verbeke WJ. Improving diagnosis of depression with XGBOOST machine learning model and a large biomarkers Dutch dataset. Front big Data. 2020;3:15.
16. Charbuty B, Abdulazeez A. Classification based on decision tree algorithm for machine learning. Journal of Applied Science and Technology Trends. 2021;2(01):20-8.
17. Yu Y, Zhong-liang F, Xiang-hui Z, et al. Combining classifier based on decision tree. 2009 WASE International Conference on Information Engineering; 2009: IEEE.
18. Mienye ID, Sun Y, Wang Z. Prediction performance of improved decision tree-based algorithms: a review. Procedia Manuf. 2019;35:698-703.
19. Wiesmeier M, Barthold F, Blank B, et al. Digital mapping of soil organic matter stocks using Random Forest modeling in a semi-arid steppe ecosystem. Plant Soil. 2011;340(1):7-24.
20. Chen J, Li K, Tang Z, et al. A parallel random forest algorithm for big data in a spark cloud computing environment. IEEE Trans Parallel Distrib Syst. 2017;28(4):919-33.
21. Yeşilkanat CM. Spatio-temporal estimation of the daily cases of COVID-19 in worldwide using random forest machine learning algorithm. Chaos Solitons Fractals. 2020;140:1-8.
22. Maillo J, Ramírez S, Triguero I, et al. kNN-IS: An Iterative Spark-based design of the k-nearest neighbors classifier for big data. Knowl Based Syst. 2017;117:3-15.
23. Adeniyi DA, Wei Z, Yongquan Y. Automated web usage data mining and recommendation system using K-Nearest Neighbor (KNN) classification method. Applied Computing and Informatics. 2016;12(1):90-108.
24. Gallego AJ, Calvo-Zaragoza J, Valero-Mas JJ, et al. Clustering-based k-nearest neighbor classification for large-scale data with neural codes representation. Pattern Recognition. 2018;74:531-43.
25. Allenbach Y, Saadoun D, Maalouf G, et al. Development of a multivariate prediction model of intensive care unit transfer or death: A French prospective cohort study of hospitalized COVID-19 patients. PloS One. 2020;15(10):e0240711.
26. Assaf D, Gutman Ya, Neuman Y, et al. Utilization of machine-learning models to accurately predict the risk for critical COVID-19. Intern Emerg Med . 2020;15(8):1435-43.
27. Das AK, Mishra S, Gopalan SS. Predicting CoVID-19 community mortality risk using machine learning and development of an online prognostic tool. PeerJ. 2020;8:e10083.
28. Hu H, Yao N, Qiu Y. Comparing rapid scoring systems in mortality prediction of critically Ill patients with novel coronavirus disease. AEM Performance Electronics. 2020;27(6):461-8.
29. Wu G, Yang P, Xie Y, et al. Development of a clinical decision support system for severity risk prediction and triage of COVID-19 patients at hospital admission: an international multicentre study. Eur Respir J. 2020;56(2):2001104.
30. Yadaw AS, Li Yc, Bose S, et al. Clinical features of COVID-19 mortality: development and validation of a clinical prediction model. The Lancet Digital Health. 2020;2(10):e516-e25.
31. Zhang Y, Xin Y, Li Q, et al. Empirical study of seven data mining algorithms on different characteristics of datasets for biomedical

classification applications. BioMed Eng. 2017;16(1):125.
32. Zhou Y, He Y, Yang H, et al. Exploiting an early warning Nomogram for predicting the risk of ICU admission in patients with COVID-19: a multi-center study in China. Scand J Trauma Resusc Emerg Med. 2020;28(1):1-13.
33. Ryan L, Lam C, Mataraso S, et al. Mortality prediction model for the triage of COVID-19, pneumonia, and mechanically ventilated ICU patients: a retrospective study. Ann Med Surg. 2020;59:207-16.
34. Arvind V, Kim JS, Cho BH, et al. Development of a machine learning algorithm to predict intubation among hospitalized patients with COVID-19. J Crit Care. 2021;62:25-30.
35. Domínguez-Olmedo JL, Gragera-Martínez Á, Mata J, et al. Machine learning applied to clinical laboratory data in spain for COVID-19 outcome prediction: model development and validation. J Multidiscip Healthc. 2021;23(4): e26211.
36. Aljouie AF, Almazroa A, Bokhari Y, et al. Early prediction of COVID-19 ventilation requirement and mortality from routinely collected baseline chest radiographs, laboratory, and clinical data with machine learning. J Multidiscip Healthc. 2021;14:2017-33.
37. Burdick H, Lam C, Mataraso S, et al. Prediction of respiratory decompensation in Covid-19 patients using machine learning: The READY trial. Comput Biol Med. 2020;124: 103949.
38. de Fátima Cobre A, Stremel DP, Noleto GR, et al. Diagnosis and prediction of COVID-19 severity: can biochemical tests and machine learning be used as prognostic indicators?. Comput Biol Med. 2021;134:104531.
39. Aydin ZE, Ozturk ZK. Performance analysis of XGBoost classifier with missing data. Manchester Journal of Artificial Intelligence and Applied Sciences. 2021;2(02):2021.
40. Wang C, Deng C, Wang S. Imbalance-XGBoost: Leveraging weighted and focal losses for binary label-imbalanced classification with XGBoost. Pattern Recognit Lette. 2020;136: 190-7.