Journal of Biostatistics and Epidemiology

J Biostat Epidemiol. 2022;8(1): 77-89.

Original Article

Machine Learning-Based Clinical Decision Support System for Automatic Diagnosis of COVID-19 based on Clinical Data

Mohammad Reza Afrash[1], Leila Erfannia[2, 3], Morteza Amraei[4], Nahid Mehrabi[5], Saeed Jelvay[6], Raoof Nopour[7], Mostafa Shanbehzadeh[8*]

[1]School of Allied Medical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran.
[2]Department of Health Information Technology, Faculty of Paramedical, Zaheda University of Medical Sciences, Zahedan, Iran.
[3]Clinical Education Research Center, Shiraz University of Medical Sciences, Shiraz, Iran.
[4]Department of Health Information Technology, School of Allied Medical Sciences, Lorestan University of Medical Sciences, Khorramabad, Iran.
[5]Department of Health Information Technology, Aja University of Medical Sciences, Tehran, Iran.
[6]Instructor of Health Information Technology, Abadan University of Medical Sciences, Abadan, Iran.
[7]Department of Health Information Technology and Management, School of Allied Medical Sciences, Tehran University of Medical Sciences, Tehran, Iran.
[8]Department of Health Information Technology, School of Paramedical, Ilam University of Medical Sciences, Ilam, Iran.

ARTICLE INFO

Received 24.11.2021
Revised 21.12.2021
Accepted 07.01.2022
Published 15.03.2022

Key words:
COVID-19;
Coronavirus;
Machine learning;
Artificial intelligence;
Decision Support Systems.

ABSTRACT

Introduction: Needless to say that correct and real-time detection and effective prognosis of the COVID-19 are necessary to deliver the best possible care for patients and, accordingly, diminish the pressure on the healthcare industries. Hence our paper aims to present an intelligent algorithm for selecting the best features from the dataset and developing Machine Learning(ML) based models to predict the COVID-19 and finally opted for the best-performing algorithm.

Methods: In this developmental study, the clinical data of 1703 COVID-19 and non-COVID-19 patients Using a single-center registry from February 9, 2020, to December 20, 2020, were used. The Minimum Redundancy Maximum Relevance (mRMR) feature selection algorithm identified the most relevant variables. Then, chosen features feed into the several data mining methods, including K-Nearest Neighbors, AdaBoost Classifier, Decision Tree, HistGradient Boosting Classifier, and Support Vector Machine. A 10-fold cross-validation method and six performance evaluation metrics were used to evaluate and compare these implemented algorithms, and finally, the best model was implemented.

Results: Out of the 34 included features, 11 variables were selected as the essential features. The results of using ML algorithms indicated that the best performance belongs to the AdaBoost classifier with mean accuracy = 92.9%, mean specificity = 89.3%, mean sensitivity = 94.2%, mean F-measure = 91.6 %, mean KAPA = 94.3% and mean ROC = 92.1 %.

Conclusion: The empirical results reveal that the Adaboost model yielded higher performance than other classification models and developed our Clinical Decision Support Systems (CDSS) interface to discriminate positive COVID-19 from negative cases.

*.Corresponding Author: mostafa.shanbezadeh@gmail.com

Introduction

Recently, COVID-19, a highly contagious viral infection disease, has become a serious global health problem. It turned out that the clinical outcomes of the virus ranged from asymptomatic or mild symptoms to severe complications and even death reports in some cases.[1, 2] Thus, healthcare industries are looking for operative and cost-effective solutions to prevent the virus spread and deal with the COVID-19 crisis. In this way, early detection, diagnosis, screening, and consequently rapid isolation of infected or pre-symptomatic cases may play a pivotal role.[3, 4]

Besides, the rapid spread of COVID-19 has resulted in the shortage of medical resources, followed by the exhaustion of healthcare providers. Hence, effective prognosis and active detection procedures are necessary to diminish the burdensome pressure on the healthcare system and provide the best possible care for the patients. It is also essential for healthcare authorities to evaluate patients' conditions, triage them effectively, and properly manage the limited medical resources.[5, 6] As a result, developing intelligent diagnostic tools can help identify positive vs. negative COVID-19 cases and give support to cut harmful complications and deaths as soon as possible.[6, 7]

Unluckily, unpredictable concluding outcomes have made the clinicians commonly use and depend upon forecasts developed from different computational models. For instance, applying intelligence methods such as Artificial Intelligence (AI) can minimize diagnosis errors and inter-observer disagreement at any prediction, prognosis, and treatment level. Therefore, diagnostic and prognostic models can significantly contribute to identifying high-risk patients and the adoption of the most effective support and treatment plans.[8-13] These may decrease ambiguity through offering quantitative, objective, and evidence-based models for risk stratification, prediction, and eventually episode of the care plan. It might offer a better strategy for clinicians to lessen the complications, and improved patient survival likelihoods may result. Evidence indicates that the computational techniques can meet the requirements of disease classification and develop accurate diagnostic modeling.[14-16]

Specifically, Machine Learning (ML) techniques are essential to developing proper decision-making, including case identification, recognizing at-risk cases, triaging patients, and resource allocation.[5, 6, 17] As a branch of Artificial Intelligence (AI), the ML extracts high-quality and applicable knowledge and patterns from mining massive raw datasets.[18] ML algorithms can reduce uncertainty and ambiguity by offering evidence-based medicine for risk analysis, screening, prediction, and care plans and supporting reliable clinical decision-making, and, consequently, enhancing and improving patient outcomes and quality of care.[19, 20]

The previous research shows that a large number of ML algorithms were trained for the classification and identification of COVID-19 cases. Therefore, this study aimed to develop different ML models for early detection of COVID-19 and finally choose the best one. It, thus, presents a practical solution as a diagnostic intelligence model to ease the COVID-19 screening according to clinical data.

Material and Methods

The study was conducted in the form of a

retrospective and single-center survey in 2021 to discriminate the COVID-19 patients from Non-COVID-19 based on selected data-driven ML techniques in clinical data. It was carried out through six stages, including 1-data set description and participants, 2-ethical consideration, 3-study roadmap, and experiment environment, 4.1- data preprocessing, 4.2-patient selection criteria, 5- classifiers, and 6-data splitting and evaluation.

Data set description and participants

This study retrospectively reviewed a COVID-19 hospital-based registry database from Imam Khomeini hospital, Ilam city, West of Iran, from February 9, 2020, to December 20, 2020. During this period, 12885 suspected cases with COVID-19 had been referred to Imam Khomeini hospital ambulatory and Emergency Departments (EDs), of whom,1853 cases were introduced as positive COVID-19 2472 as negative via RT-PCR test. Inclusion/ exclusion criteria finally revealed1703 records entered in the study (350 non-COVID-19 and 1353 COVID-19). Thirty-four clinical features are correlated in disease diagnosis, and one attribute serves as an output variable (see Table1).

Ethical consideration

The ethical committee board approved the study of Ilam University of Medical Sciences (Ethics code: IR.MEDILAM.REC.1399.294). To protect the privacy and confidentiality of patients, we concealed the unique identification information of all patients in the process of data collection and presentation.

Study Roadmap and Experiment Environment

It should be noted that all experiments on the ML models (described in this paper)were run using Python programming languages (version 3.7.7). They were in three phases: preprocessing, training, evaluating, and using C# programming language to design a CDSS user interface. The ML models developed with Python are applied to various real-world issues. The python experiment environment also provides an excellent framework for developers to run and evaluate their ML algorithms. The road map of the proposed system for the detection of

Table 1. Clinical features studied on COVID-19 dataset

Modalities	Data classes	Variables
Input variables	Laboratory	absolute lymphocyte/neutrophil count, blood urea nitrogen (BUN), total bilirubin, aspartate aminotransferase (ASP), alanine aminotransferase(ALT), lactate dehydrogenase, activated partial thromboplastin time, prothrombin time, alkaline phosphatase, c-reactive protein, ferritin, D-dimer, erythrocyte sedimentation rate, hypersensitive troponin
	Clinical manifestation	Dyspnea, SPO2 rate, fever, dry cough, muscular pain
	Radiological	Existence of lesion, lesion dissemination (involved lobe), lesion staging
	Basic data	Age, sex, BMI, blood group
	Disease history	Cardiac disease, pneumonia, hypertension, smoking, alcohol addiction, diabetes, and other underline diseases
Output variables	COVID-19 diagnosis	Negative or Positive

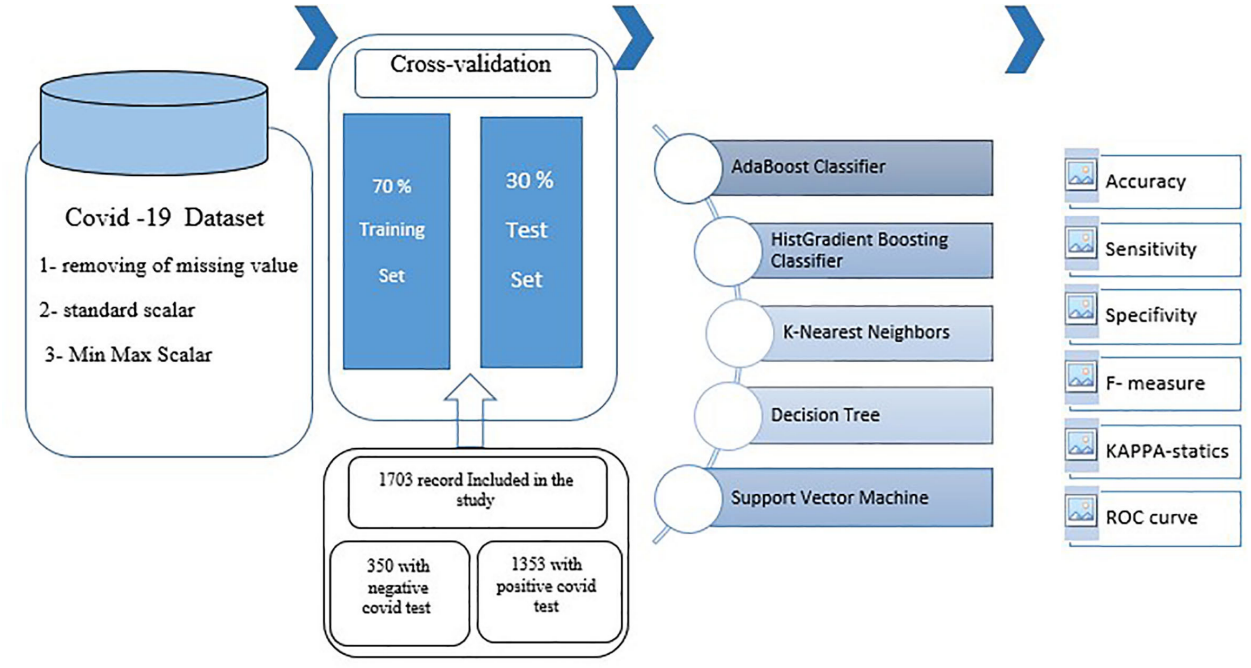

Figure 1. Road map of the proposed system for diagnosis of COVID-19

COVID-19 is displayed in Figure 1.

Data Pre-processing

In the preprocessing stage, the raw data were imputed using several preprocessing methods such as deleting missing values, minimum and maximum scalar, and standard scalar for effective use of data in the classification algorithms. The standard scalar guarantees that every character has the mean zero and variance one and brings all features to the same Coefficient. Likewise, Minimum and Maximum Scalar transfers the values such that all attributes are between 0 and 1, and rows with missing val¬ues (greater than 70%) were removed. Of course, two authors and two infectious disease and hematology specialists checked the noisy and abnormal values, errors, duplicates, and meaningless data (L: E and M: SH). For different interpretations about data preprocessing, we contacted the corresponding physicians accordingly.

Patient selection criteria

The data on 2521 eligible patients were extracted from the Imam Khomeini hospital registry database. Then, 228 incomplete case records with many missing data (more than 70%) were excluded from the analysis. Also, the missing values were imputed with the mean or mode of each variable. After applying exclusion criteria, the 1703 case records were ultimately chosen for the study (Figure 2).

Feature selection

Feature selection reduces the dimension of databases to the most important features predicting the output class. This process can lead to higher accuracy using the best features in the database. Some advantages of this process are 1- increasing the performance of algorithms. 2-getting the best predictors affecting the dependent variable. 3- increasing the speed of the algorithm's

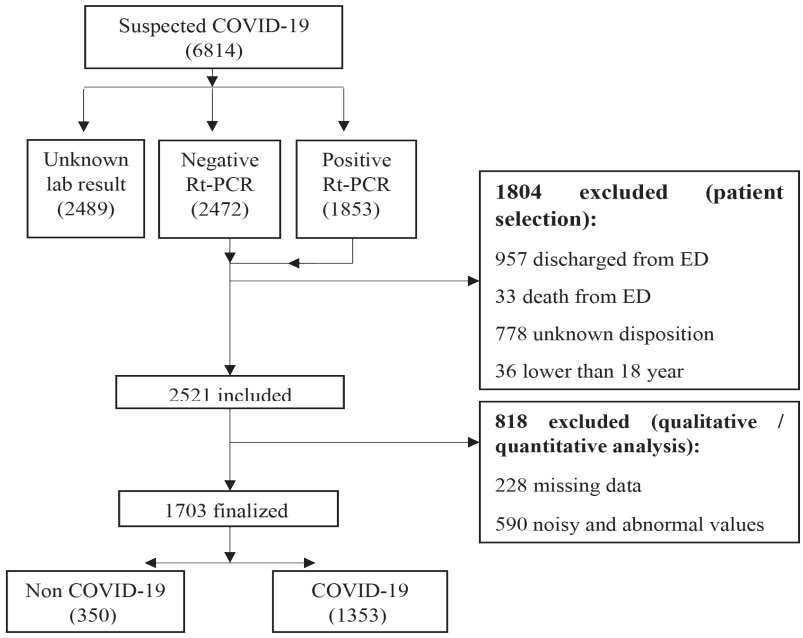

Figure 2. Flow chart describing patient selection

training. 4- preventing overfitting and 5-better understating the dataset by algorithms. The mRMR algorithm is applied to select the most relevant features in this work. Then, chosen features feed into the selected ML models.

Classifiers

To diagnose COVID-19 based on clinical data, we applied five classification techniques, including AdaBoost, HistGradient Boosting (HGB), K-Nearest Neighbors (K-NN), Decision Tree (DT), and Support Vector Machine (SVM), and also we used Grid environment. It should be noted that the grid search method is an effective technique to set the parameters in the learning phase of classification models and improve the generalization efficiency of an algorithm.

Data Splitting and Evaluation

In this section, the data related to the patient were randomly divided into training (70%) and testing (30%) sets. Then, this process was repeated ten times for ten independent runnings of ML models. In the present study, we applied 10-fold cross-validation and six performance evaluation metrics, including accuracy, specificity, sensitivity, F-measure, KAPA- Statics, and ROC curve (Equations 1 to 6) to compare the performance of implemented models. Also, we assess the effectiveness of five ML models in terms of time to build the model, correctly classified instances, and incorrectly classified samples to efficiently compare the performance of algorithms.

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

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

3) $$\text{c lassification specificity} = \frac{TN}{TN+FP} * 100$$

4) $$\text{classification error} = \frac{FP + FN}{TP + TN + FP + FN} * 100$$

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

Table 2. The descriptive statistics of selected variables

Variable name	Range	Mean (SD) / frequency
Age (year)	18-100	57.25 (17.8)
Height (centimeter)	126-195	163.53 (7.5)
Weight (kilogram)	42-123	85.20 (11.3)
Sex	Male, female	942, 761
Blood type	A-,A+	124, 256
	B-, B+	186, 326
	O-, O+	56, 96
	AB-, AB+	247, 412
Absolute lymphocyte count	2-95	23.74 (11.8)
Absolute neutrophil count	8-98	74.52 (12.3)
Blood urea nitrogen	0.5-251	42.52 (31.7)
Total bilirubin	0.01-10	0.72 (0.7)
Aspartate aminotransferase	3.8-924	44.45 (53.5)
Alanine aminotransferase	2-672	38.29 (41.6)
Lactate dehydrogenase	4.6-6973	555.68 (339.0)
Activated partial thromboplastin time	1-120	28.56 (11.4)
Prothrombin time	0.9-46.8	12.82 (1.9)
Alkaline phosphatase	9.6-2846	213.12 (139.2)
C-reactive protein	Positive	1263
	Negative	440
Ferritin	10.2-885	36.8 (48.9)
D-dimer	120-3800	818 (87.4)
Erythrocyte sedimentation rate	2-258	40.65 (28.8)
Hypersensitive troponin	Positive	236
	Negative	1467
SPO2	40-98	87(4)
Fever	Yes, no	714, 989
Dyspnea	Yes, no	697, 1006
Dry cough	Yes, no	885, 818
Muscular pain	Yes, no	1123, 580
Existence of lesion	Yes, no	432, 1271
If yes: lesion dissemination (involved lobe)	Unilateral, bilateral	189, 242
If yes: lesion staging	Early, progressive, severe	103, 253, 76
Cardiac disease	Yes, no	69, 1634
Pneumonia	Yes, no	92, 1611
Hypertension	Yes, no	207, 1496
Diabetes	Yes, no	108, 1595
Smoking	Yes, no	41, 1662
Alcohol consumption	Yes, no	18, 1685

Table 3 . Features selected by mRMR method

Order	Feature	Feature name	Score
1	1	Age	3.71
2	16	C-reactive protein	2.9
3	22	Fever	2.47
4	23	Dyspnea	2.21
5	24	Dry cough	1.80
6	18	D-dimer	1.64
7	7	Absolute lymphocyte count	1.33
8	21	SPO2	1.27
9	25	Muscular pain	1.24
10	12	Lactate dehydrogenase	1.17
11	8	Absolute neutrophil count	1.10

Table 4. Average evaluation metrics obtained from 10 runs of ML Models

Classifier		Mean Accuracy	Mean Specificity (%)	Mean Sensitivity	Mean F- measure	Kappa Statistic (KS)	ROC Rate
K-Nearest Neighbors	Mean	0.873	0.873	0.8813	0.883	0.88701	0.879
	95% CI	(0.841, 0.908)	(0.847, 0.8934)	(0.841, 0.90)	(0.83, 0.90)	(0.839, 0.91)	(0.847, 0.893)
	STD	0.0214	0.02743	0.01427	0.0238	0.02913	0.01846
AdaBoost Classifier	Mean	0.92907	0.8937	0.9427	0.916	0.9435	0.9214
	95% CI	(0.88, 0.96)	(0.85, 0.9292)	(0.923, 0.9675)	(0.871, 0.952)	(0.89, 0.97)	(0.88, 0.96)
	STD	0.2945	0.0204	0.0286	0.01834	0.19547	0.0183

XGB=XG Boost Classifier

Results

Characteristics of patients

For the retrospective study, 1703 eligible patients (1353 COVID-19 and 350 non-COVID-19) matched our inclusion/exclusion criteria, of whom 942(55.32%) were male, and 761 (44.68%) were female. The median age of the participant was 57.25 (interquartile 18-100) (Table 2).

Results of feature selection by mRMR Algorithm

mRMR feature selection method selects the most important features based on weight and

Table 5. Average evaluation metrics obtained from 10 runs of ML Models

Classifier		Mean Accuracy	Mean Specificity (%)	Mean Sensitivity	Mean F-measure	Kappa Statistic (KS)	ROC Rate
Decision Tree	Mean	0.923	0.917	0.919	0.9171	0.911	0.9104
	95% CI	(0.871, 0.953)	(0.883, 0.943)	(0.888, 0.941)	(0.883, 0.948)	(0.881,0.94)	(0.89, 0.933)
	STD	0.0477	0.0384	0.0342	0.0367	0.0368	0.0317
HGB	Mean	0.9113	0.9104	0.9002	0.9025	0.907	0.9034
	95% CI	(0.882, 0.9401)	(0.876, 0.957)	(0.882, 0.923)	(0.883, 0.9204)	(0.872, 0.931)	(0.885, 0.9217)
	STD	0.0288	0.0343	0.03547	0.0368	0.04231	0.03252
	Mean	0.879	0.8711	0.877	0.882	0.8833	0.8788
SVM	95% CI	(0.861, 0.884)	(0.851, 0.8992)	(0.859, 0.897)	(0.857, 0.911)	(0.869, 0.902)	(0.852, 0.897)
	STD	0.03272	0.03425	0.03125	0.0357	0.03234	0.0297

HGB, HistGradient Boosting Classifier; SVM, Support Vector Machine

Table 6. Performance of the classifiers

Evaluation criteria	Classifier				
	K-Nearest Neighbors	Decision Tree	AdaBoost Classifier	HGB	SVM
Best Time to build a model (s)	323	120	94	140	150
Correctly classified instances	452	456	474	465	447
Incorrectly classified instances	59	55	26	46	64

mutual information variables. The selected relevant and important 11 variables by mRMR Feature Selection method based on variables weight and mutual information are represented in Table 4. The feature selection in table 4 includes age, C-reactive protein, fever, dyspnea, dry cough, D-dimer, absolute lymphocyte count, SPO2, muscular pain, lactate dehydrogenase, and absolute neutrophil count are important features in the diagnosis of COVID-19 (see Table 3).

K-Fold Cross-Validation

In the present study, models were run ten independent times to measure the actual performance of classifiers better. Then we evaluated the actual performance of our classifier in terms of mean accuracy, mean F-measure, mean Specificity, mean Sensitivity, mean KAPA Statistic, and ROC rate. Tables 4 and 5 show the 10-fold cross-validation results of five classification models based on the test data set.

The experimental results on the test dataset in tables 3 and 4 show, the AdaBoost classifier obtained the best performance on the test dataset compared with the other three models based on the average of evaluation metrics. The AdaBoost classifier algorithm got the following standards: 0.92907 for accuracy, 0.8937 for specificity, 0.9427 for sensitivity, 0.916 for F-measure, 0.9135 for KAPA statistics, and, finally 0.9214 for ROC metrics. As shown in Table 6, the result obtained for classifiers regarding the time to run the model, correctly classified instances, and incorrectly classified cases.

It is apparent from Table 6 that the K-NN algorithm takes 323 (s) to build its model as the fastest model, whereas HGB Classifier takes about 140 (s) that was the slowest. Additionally, the AdaBoost Classifier algorithm takes 94 (s) to build its model. It should be said that the AdaBoost model was chosen as the best algorithm and used for developing our CDSS for the diagnosis of COVID-19. The confusion matrix and ROC curve of the Adaboost as the best ML models (in terms of the highest assessment metrics and AUC) in this study are displayed in Figure 3.

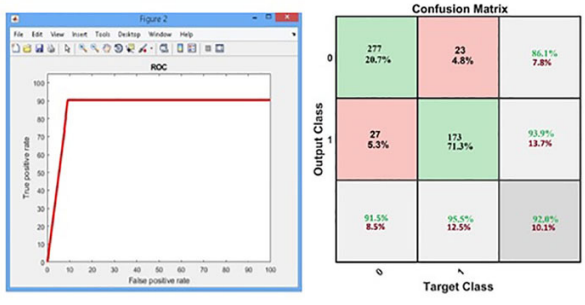

Figure 3. Confusion matrix and ROC chart of AdaBoost Classifier

System implementation

Using the developed AdaBoost classifier, we designed and imple¬mented a CDSS interface during June and July 2021. The CDSS consisted of two types of implementation codes: principles for the user interface implementation and regulations for the logic layer implementation (programming rule of CDSS according to AdaBoost classifier algorithm). The user interfaces in our study comprised of 2 pages: Welcome page (sign up and log in page) and CDSS modules. The COVID-19 diagnosis system user interface was developed with the C# programming language. Two screenshots of implemented system are shown in Figures 4 and 5.

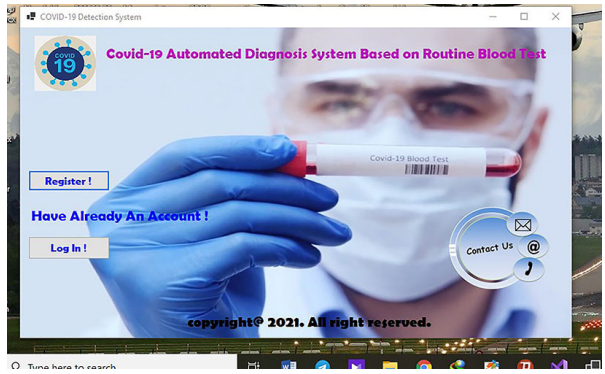

Figure 4. The welcome page of the clinical decision support system

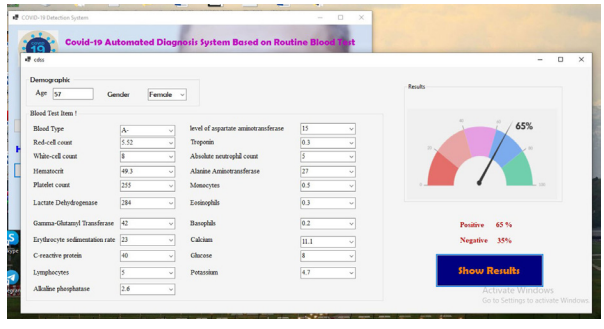

Figure 5. COVID-19 automated diagnosis system interface

Discussion

ML-based diagnostic models (intelligence systems) are proven to be useful for accurate case identification. They might minimize uncertainty and ambiguity by offering a systematic and evidence-based system for screening. This study intends to construct an intelligent CDSS via leveraging feature selection and ML-based predictive capabilities also, intends to compare the accuracy and efficiency of selected ML techniques for COVID-19 diagnosis. For this purpose, five ML methods, including KNN, AdaBoost, DT, HGB, and SVM, were trained in a dataset of 1703 de-identified case records of confirmed

and suspected COVID-19. For doing this, at first, we selected the 11 most relevant clinical predictors related to COVID-19 diagnosis by using the mRMR feature selection method. Then, chosen features feed into several ML methods.

So far, numerous studies have been conducted on the application of CDSS based ML techniques to the early detection and prognosis of COVID-19. For example, Kukar and their colleagues (2020) retrospectively studied the laboratory data of 5333 patients suffering from various bacterial and viral infections and 160 COVID-19 positive RT-PCR cases to construct an intelligent diagnostic model through selected ML algorithms. Finally, the XGBoost model with a sensitivity of 81.9% and a specificity of 97.9% revealed the best performance.[21] Also, Castiglioni et al. (2021) conducted a retrospective analysis on chest x-ray data of 250 COVID-19 and 250 non-COVID-19 subjects. Their results showed that the model developed using Convolutional Neural Networks (CNNs) with 0.80 sensitivity, 0.81 specificities, and 0.81 AUC enjoyed the best performance.[22] Besides, Tamal et al. (2020) compared three ML classification performances based on a total of 378 CXR images containing both normal lung and pneumonia for early diagnosis of COVID-19. Finally, Ensemble Bagging Model Trees (EBM) was introduced as the most suitable algorithm to distinguish between COVID-19 and other lung infections with an overall sensitivity of 87.8% and specificity of 97%.[23] Similarly, Baktash et al. (2021) showed an EBM model with an accuracy of 81.79%, sensitivity of 85.85, and specificity of 76.65%, which was reportedly the best rating.[24] Also, Alves et al. (2021) retrospectively analyzed 608 suspected COVID-19 patients (84 confirmed RT-PCR and 524 negatives). The experimental results showed that the Random Forest (RF) classifier achieved the best results with an accuracy of 0.88, F1–score of 0.76, the sensitivity of 0.66, specificity of 0.91, and AUROC of 0.86.[25] Mohammed and their colleagues (2021) proposed an ML predictive model based on an X-ray dataset of 400 healthy and 400 COVID cases. Finally, the most successful results were obtained concerning SVM with an accuracy of 95%.[26] Saha (2021) evaluated chest X-ray images and designed an automated detection ML-based system to identify COVID-19 patients. Finally XGBoost algorithm showed better performance with 98.91% accuracy, 100% precision, 97.82% recall, and 98.89% F1-score.[27] Similarly, Kim et al. (2020) assessed the performance of 55 ML algorithms to predict the prognosis of patients with COVID-19, and the best performance was reported from the XGBoost model (AUC= 0.897).[28] In this study, multiple ML-based diagnostic models, including KNN, AdaBoost, DT, HGB, and SVM, were trained and evaluated to determine the most optimal algorithm for COVID-19 detection. In our study, the results in 10 iteration execution in selected ML algorithms showed that the AdaBoost classifier enjoyed more discriminative power than other ML methods with mean accuracy = 92.9%, mean specificity = 89.3%, mean sensitivity = 94.2%, mean F-measure = 91.6 %, mean KAPA = 94.3% and mean ROC = 92.1 %.

The suggested ML-based CDSS can distinguish infected COVID-19 cases from healthy individuals with optimal performance. Hence, it provides a better plan for clinicians to improve patient outcomes and quality of care, especially in overwhelmed hospitals. It also may minimize the ambiguity by offering

a scientific and evidence-based model for resource utilization and the episode of care planning. ML models used for the diagnosis and prognosis of COVID-19 showed optimum discriminative performance. Given the power of the current study in the timely and accurate identification of positive cases, though, this study had some limitations. Thus, it is necessary to identify and tackle the causes of classification bias. First, we deal with a retrospective dataset that may lack unfilled and imbalanced data fields. Second, this study was conducted at a single center on a restricted number of 1703 data which undoubtedly confines the generalizability of the diagnostic model. Moreover, we used only five ML algorithms for prediction analyses based on some clinical features.

Conclusion

This study develops and evaluates some ML models for COVID-19 screening using the most important clinical parameters (11 variables). It was observed that the AdaBoost model performed much better on classification accuracy than the other algorithms. The developed CDSS can automatically and effectively discriminate between the admitted patients who are either positive or negative as early as possible. It demonstrates exemplary performance in COVID-19 classification and screening. The results suggest that frontline clinicians can use such systems to augment effective decision-making. To sum up, the ML algorithms coupled with qualitative and comprehensive clinical features in this study enable timely and accurate disease classification. However, it should be emphasized that the performance accuracy of our computational model will be improved if we test more ML techniques, using more extensive, multicenter, and perspective as well as various and reliable qualitative data.

Conflicts of interest

The authors declare that they have no conflict of interests.

Acknowledgments

This article is extracted from a research project supported by the Ilam University of Medical Sciences (IR.MEDILAM.REC.1399.294). We also thank the Ilam University of Medical Sciences Research Deputy for financially supporting of this project.

References

1. Dinicolantonio JJ, McCarty M, Barroso-Aranda J. Melatonin may decrease the risk for and aid treatment of COVID-19 and other RNA viral infections. Open Heart. 2021;8(1).

2. Tan JY, Sim XYJ, Wee LE, Chua YY, Cherng BPZ, Ng IM, et al. A comparative study on the clinical features of COVID-19 with non-SARS-CoV-2 respiratory viral infections. Journal of Medical Virology. 2021;93(3):1548-55.

3. Bai L, Yang D, Wang X, Tong L, Zhu X, Bai C, et al. Chinese experts' consensus on the Internet of Things-aided diagnosis and treatment of coronavirus disease 2019. Clinical eHealth. 2020.

4. Bayram M, Springer S, Garvey CK, Özdemir V. COVID-19 digital health

innovation policy: A portal to alternative futures in the making. OMICS: A Journal of Integrative Biology. 2020.

5. Yan L, Zhang H-T, Goncalves J, Xiao Y, Wang M, Guo Y, et al. An interpretable mortality prediction model for COVID-19 patients. Nature Machine Intelligence. 2020:1-6.

6. Hu H, Yao N, Qiu Y. Comparing Rapid Scoring Systems in Mortality Prediction of Critically Ill Patients With Novel Coronavirus Disease. Academic Emergency Medicine. 2020;27(6):461-8.

7. 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(7):e0236618.

8. Liu Y, Wang Z, Ren J, Tian Y, Zhou M, Zhou T, et al. A COVID-19 Risk Assessment Decision Support System for General Practitioners: Design and Development Study. Journal of medical Internet research. 2020;22(6):e19786.

9. Alom MZ, Rahman M, Nasrin MS, Taha TM, Asari VK. COVID_MTNet: COVID-19 Detection with Multi-Task Deep Learning Approaches. arXiv preprint arXiv:200403747. 2020.

10. Bansal A, Padappayil RP, Garg C, Singal A, Gupta M, Klein A. Utility of Artificial Intelligence Amidst the COVID 19 Pandemic: A Review. Journal of Medical Systems. 2020;44(9).

11. Lai C-C, Shih T-P, Ko W-C, Tang H-J, Hsueh P-R. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges. International Journal of Antimicrobial Agents. 2020;55(3):105924.

12. Hussain A, Bhowmik B, do Vale Moreira NC. COVID-19 and diabetes: Knowledge in progress. Diabetes Research and Clinical Practice. 2020;162.

13. Moujaess E, Kourie HR, Ghosn M. Cancer patients and research during COVID-19 pandemic: A systematic review of current evidence. Critical Reviews in Oncology/Hematology. 2020;150:102972.

14. Kanagarathinam K, Sekar K. Estimation of the reproduction number and early prediction of the COVID-19 outbreak in India using a statistical computing approach. Epidemiology and Health. 2020;42.

15. Briz-Redon A, Serrano-Aroca A. The effect of climate on the spread of the COVID-19 pandemic: A review of findings, and statistical and modeling techniques. Progress in Physical Geography-Earth and Environment.

16. Sarkodie SA, Owusu PA. Investigating the cases of novel coronavirus disease (COVID-19) in China using dynamic statistical techniques. Heliyon. 2020;6(4).

17. Malki Z, Atlam E-S, Hassanien AE, Dagnew G, Elhosseini MA, Gad I. Association between weather data and COVID-19 pandemic predicting mortality rate: Machine learning approaches. Chaos, Solitons & Fractals. 2020;138:110137.

18. Shanbehzadeh M, Nopour R, Kazemi-Arpanahi H. Comparison of Four Data Mining Algorithms for Predicting Colorectal Cancer Risk. Journal of Advances in Medical and Biomedical Research.29(133):100-8.

19. Shipe ME, Deppen SA, Farjah F, Grogan EL. Developing prediction models for clinical use using logistic regression: an overview. Journal of thoracic disease. 2019;11(Suppl 4): S574.

20. Gao Y, Cai G-Y, Fang W, Li H-Y, Wang S-Y, Chen L, et al. Machine learning-based early warning system enables accurate mortality risk prediction for COVID-19. Nature communications. 2020;11(1):1-10.

21. Kukar M, Gunčar G, Vovko T, Podnar S, Černelč P, Brvar M, et al. COVID-19 diagnosis by routine blood tests using machine learning. Scientific reports. 2021;11(1):1-9.

22. Castiglioni I, Ippolito D, Interlenghi M, Monti CB, Salvatore C, Schiaffino S, et al. Machine learning applied on chest x-ray can aid in the diagnosis of COVID-19: a first experience from Lombardy, Italy. European Radiology Experimental. 2021;5(1):1-10.

23. Tamal M, Alshammari M, Abdullah M, Hourani R, Alola HA, Hegazi TM. An integrated framework with machine learning and radiomics for accurate and rapid early diagnosis of COVID-19 from chest X-ray. Expert systems with applications. 2021;180:115152.

24. Baktash V, Hosack T, Rule R, Patel N, Kho J, Sekhar R, et al. Development, evaluation, and validation of machine learning algorithms to detect atypical and asymptomatic presentations of Covid-19 in hospital practice. QJM: An International Journal of Medicine. 2021.

25. Alves MA, Castro GZ, Oliveira BAS, Ferreira LA, Ramírez JA, Silva R, et al. Explaining machine learning-based diagnosis of COVID-19 from routine blood tests with decision trees and criteria graphs. Computers in Biology and Medicine. 2021;132:104335.

26. Mohammed MA, Abdulkareem KH, Garcia-Zapirain B, Mostafa SA, Maashi MS, Al-Waisy AS, et al. A comprehensive investigation of machine learning feature extraction and classification methods for automated diagnosis of covid-19 based on x-ray images. Computers, Materials, and Continua. 2020;66(3).

27. Saha P, Sadi MS, Islam MM. EMCNet: Automated COVID-19 diagnosis from X-ray images using convolutional neural network and ensemble of machine learning classifiers. Informatics in medicine unlocked. 2021;22:100505.

28. Kim H-J, Han D, Kim J-H, Kim D, Ha B, Seog W, et al. An Easy-to-Use Machine Learning Model to Predict the Prognosis of Patients With COVID-19: Retrospective Cohort Study. Journal of medical Internet research. 2020;22(11):e24225.