

Comparing Supervised Machine Learning Models for Covid-19 patient detection using a Combination of Clinical and Laboratory Dataset

Narges Mohebbi^a, Mehdi Tutunchian^a, Meysam Alavi^a, Mehrdad Kargari^{a*}, Amir Behnam Kharazmi^b

^a Department of Information Technology, Tarbiat Modares University, Tehran, Iran;

narges.mohebbi@modares.ac.ir, mehditutunchian@modares.ac.ir, meysamalavi@modares.ac.ir, M_kargari@modares.ac.ir*

^b Department of Internal Medicine, School of Medicine, Shahid Beheshti University of Medical Sciences, Imam Hossein Hospital, Tehran, Iran; Drkharazmy@sbmu.ac.ir

ABSTRACT

COVID-19 is a new variant of SARS-COV-2 which can lead to mild to severe infection in humans. Despite the remarkable efforts to contain the epidemic, the virus spread rapidly around the world and its prevalence continued with different degrees of clinical symptoms in many countries. Although common strategies including prevention, diagnosis, and care are necessary to curb this epidemic, early and accurate diagnosis can play an important role in reducing the speed of the epidemic. In this regard, the use of technologies based on artificial intelligence can be of great help. For this reason, since the outbreak of COVID-19, many researchers have tried to use machine learning techniques as a subset of artificial intelligence for the early diagnosis of COVID-19. Considering the importance and role of using clinical and laboratory data in the diagnosis of people with covid-19, in this paper K-NN, SVM, decision tree, random forest, Naive Bayes, neural network and XGBoost models are the most common machine learning models, and a dataset containing 1354 records consisting of clinical and laboratory data of patients in Imam Hossein Hospital in Tehran has been used to diagnose patients with covid-19. The results of this research indicate that based on the evaluation criteria, XGBoost and K-NN models have the most accuracy among the mentioned models and can be considered suitable predictive models for the diagnosis of COVID-19.

Keywords: COVID-19; Coronavirus; Early Detection; Machine Learning Techniques; Supervised Model.

1. Introduction

The first records of the coronavirus date back to 1930 [1]. But in the last days of 2019, a new type of this virus, which is also highly contagious, was reported from the city of Wuhan, Hubei Province, China, and was named sars-cov-2 [2], [3]. The names of the virus include Covid-19, 2019nCoV, and 2019 Novel Coronavirus. It reports the sequence of the bat gene, but the intermediate hosts between the bat and the human are still being studied. The disease is transmitted through the respiratory system of a patient that is inhaled, coughing, or sneezing in the air and is transmitted to a healthy person's respiratory system [4]. Covid-19 has spread rapidly worldwide due to its high transmission rate and contagiousness [5], making it the fifth deadliest virus in history [6], with the World Health Organization declaring the virus to be the sixth leading cause of global health emergency [7]. The virus has a similar impact on the global economy [4], [5], and according to the World Economic Cooperation and Development Organization, the emergence of Covid-19 has put economic conditions at their worst in history [8]. Some of the other impacts of the virus can be attributed to the impact on the education system and the impact on mental and social health [6]. Strategies for managing the disease worldwide include force lockdown, travel restrictions, social distancing, and quarantine [5]. Due to the damage that the virus has done

to the world, early detection, improvement of screening, prevention and control of this disease are urgent and are highly encouraged [4-6], [9], [10]. All over the world diagnosis of the disease is based on the patient's history and physical examination, making the possibility of a mistake or late diagnosis of the disease inevitable. Once the disease is diagnosed quickly, the chances of successful treatment will increase, and the patient's rehabilitation costs will decrease. Most medical errors are due to a lack of medical tools when making decisions. For example, the RT-PCR test, the golden standard method for diagnosing the disease, is not efficient enough; it's expensive, needs too much time and is not very sensitive at all [4], [10]. In recent years, after the start of Covid-19, artificial intelligence (AI) methods have become very popular due to features such as high accuracy, sensitivity and evaluation criteria, as well as the use of low resources and high speed in medicine and biology [2], [11]. Nowadays, very useful and popular tools are being made to help the decision-making process more accurately and with the least possible mistakes in the medical field, which are based on artificial intelligence techniques and algorithms, and machine learning has an undeniable role in health care and disease diagnosis [12-16].

The power of machine learning creates intelligent systems whose output can be used as a useful and efficient tool in finding the signs of disease in potential patients which will



<http://dx.doi.org/10.22133/ijwr.2022.348799.1127>.

Citation N. Mohebbi, M. Tutunchian, M. Alavi, M. Kargari, and A. B. Kharazmy, "Comparing Supervised Machine Learning Models for Covid-19 patient detection using a Combination of Clinical and Laboratory Dataset," *International Journal of Web Research*, vol.5, no.1, pp.8 -18, 2022.

*Corresponding Author

Article History: Received 26 June 2022; Revised 11 July 2022; Accepted 29 July 2022

Copyright © 2022 University of Science and Culture. Published by University of Science and Culture. This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International license (<https://creativecommons.org/licenses/by-nc/4.0/>). Noncommercial uses of the work are permitted, provided the original work is properly cited.

lead to reducing the huge burden on the health care system in all countries of the world [15], [17]. Studies show that This case has been the intention and goal of many researchers who are active in this field [17], [18]. Most of these studies have focused on the diagnosis of patients with covid-19 based on CT scan and X-ray images of the lungs, and especially, deep learning algorithms are considered very successful in this field because they are unique in discovering patterns and high-level features from images. [19]. In addition, artificial intelligence techniques are widely to predict whether people are sick or not based on their medical information. A person's medical information can include things such as symptoms and history of underlying diseases (clinical data), results of various tests such as blood or urine tests (laboratory data) and general information such as gender and age (demographic data) [18]. The diagnosis of covid-19 using clinical and laboratory findings is not an exception to this rule. Accordingly, in this study, a brand new and up-to-date dataset of the mentioned data of patients with covid-19 who were admitted to Imam Hossein Hospital in Tehran was collected and seven models of the most used supervised machine learning models were selected and trained by this dataset. Finally, the performance of these models has been examined and compared in terms of accurate disease diagnosis.

The structure of the continuation of the paper is organized as follows: In the second part, the literature review is briefly reviewed. In the third part, the research method is introduced. In the fourth part, the results are evaluated. The fifth part is related to conclusions and suggestions for future work.

2. Literature Review

As stated in the introduction of the study, several studies have been conducted in this field that has tried to use intelligent systems for the early diagnosis of diseases. In the following, we will categorize and review a selection of these studies. There are basically three types of disease diagnostic models; The first group is models whose basis is working on CT scan images or X-ray images of the patient's lungs. For example, the authors in [20] used a dataset consisting of 2481 CT scan images to train their models. Finally, the CNN-based model was able to register a better performance with 95% precision. Another research [21] has trained a CNN model to diagnose the disease with a dataset consisting of CT scans of 131 people with Covid-19 infection and 150 people without Covid-19 infection (a total of about 10,000 images), and its final precision had reached up to 90%. A study was performed on medical images of the lungs, which in the first stage analyzes the images and extracts important features using machine vision techniques, and in the second stage uses five machine learning models, including k Nearest Neighbor (k-NN), Decision Tree (DT), Multinomial Logistic Regression (MLR), Naive Bayes (NB) and Support Vector Machine (SVM) to diagnose Covid-19. They have reached the accuracy of 0.85, 0.79, 0.91, 0.61, 0.89, respectively [10]. X-ray images are also used to train disease diagnosis models, for example [22] used a set of 392 of these images to train their CNN model, which resulted in 98.4% precision. Also, among the research, one can find examples that have used both CT scan images and X-ray images to train their models at the same time [23-25].

In addition to the studies that used patient lung images to diagnose COVID-19, other studies used patients' clinical

information. For example, the study [26] used a database of 1,624 patients (with a 52% positive COVID-19 rate) that included white blood cell count, monocytes (%), lymphocyte (%), and age features to build a stacking machine learning-based system for covid-19 diagnosis. The above system was measured by the parameters of accuracy, precision, recall, and F-Score, which achieved 91.45%, 91.44%, 91.44%, and 91.45%, respectively. In the final section, this system leads to another nomogram-based scoring system which scored 93.34%, 92.02%, 90.54%, and 93.73% for accuracy, precision, recall, and F-Score parameters, respectively. In [9], a dataset was formed based on routine blood values (RBV), which includes 28 features for two groups of patients named severely (192 patients) and mildly (4010 patients), and the writer claims that the effect of all 28 features on the diagnosis of Covid-19 in the patient is medically approved. In the next step, this dataset was analyzed using supervised machine learning algorithms. The algorithms used and their performance based on accuracy is (LWL)-97.86%, K-star (K*)-96.31%, Naive-Bayes (NB)-95.36%, and k-nearest-neighbor (KNN)-94.05%. Moreover, they have reached the AUC of LWL-0.95%, K*-0.91%, NB-0.85%, and KNN-0.75%. Another study [27] used four machine learning models, including RF, XGBoost, SVM, and logistic regression, to detect Covid-19. The data used in this study consists of the results of six clinical tests (CC), 14 blood tests (CBC), and PCR test results as the golden standard that were collected for 1391 patients. Using the mentioned algorithms, eight models have been made using CBC and combining CBC and CC. The RF model on the combined data performed best with 85.3% accuracy, 79.6% sensitivity, and 91.2% specificity, while the accuracy of this model was 82.8% on using only CBC. Also, the SVM model using combined data has the best performance on external data with 91.18% accuracy, 100% sensitivity, and 84.21% specificity. In [28], we also see that machine learning algorithms (NB, NLP, KNN, RF, LR & DT) were used to diagnose Covid-19 using clinical patient datasets. Nevertheless, the different point in this research is the attention paid to optimization of a dataset and recognizing the importance of different features (Feature selection) for detecting COVID-19, with different algorithms (Boruta (shadow features), Ridge Regression (RR) & Random Fourier features (RFF)) and the results of all models Are reported for all cases. Although, in general, in all cases, we see a curacy of over 90%, feature selection has improved the performance of different models in terms of classification and processing time. For example, the accuracy of the NB model with the Boruta algorithm is 83%, with the RR algorithm 84%, and the RFF algorithm is 93%. Also, the best performance among all cases is the RF model with the Boruta algorithm with 95% accuracy, 94% precision, 95% recall, and 94 F-Score. In [29], the authors used the blood test results of 279 patients to form their dataset and train their models. Different machine learning models were trained using these data and sensitivity in the range of (92 to 95) % and accuracy in the range of (82 to 86) % were obtained. The authors in [30] have also tried to provide models for disease diagnosis using ML algorithms and the data of 235 patients. Finally, the SVM model showed the best performance among the five models built in this research. Another study [31] has focused on predicting the recovery of covid-19 patients. SVM, KNN, DT and RF models were used for this purpose, and the DT model performed better than the others with 99.85%

precision. The different performance of conventional machine learning models in different researches is also significant and in different conditions, different models have shown better performance. For example, in another study [32], which used the results of blood tests of patients to train the models, the NB model achieved the highest accuracy value and the lowest computational costs among other models. In the literature on the subject, one can find many other examples similar to what was stated [33-36].

To be comprehensive in the present report, we should also mention the studies that used a combination of image and clinical data to train COVID-19 diagnostic models. For example, [37] used a CXR image dataset and a clinical data dataset of Covid-19 patients to build a model that performed better than the image-based and the clinical data-based models. The first two independent models were trained and evaluated, then fusion techniques were used to combine the two models, and an accuracy of 0.970, a recall of 0.986, precision of 0.978, and an F-Score of 0.982 were obtained.

3. Methodology

The methodology of this research consists of four stages, which are data collection, data pre-processing, classifier model design and evaluation Figure 1. In the rest of this section, each stage will be briefly introduced.

3.1 Data collection

The dataset used in this research was collected from Imam Hossein Hospital in Tehran, which includes 1354 patients referred to this hospital for testing for covid-19. It took about 40 working days to collect this dataset and it was obtained from reviewing the files of 1354 patients who were referred to Imam Hossein Hospital between October 2019 and October 2014. In this data set, 754 samples of records with the label of confirmed infection with covid-19 and 600 samples with the label of non-confirmation of infection with covid-19 have been recorded.

3.2 Data Preprocessing

The dataset collected from patient files has 30 features, which are classified into three categories of clinical, laboratory and population data. Feature selection is considered as selecting the appropriate variable from a dataset, in other words, feature selection is a process in which essential information is extracted from a large dataset to reduce processing time with better performance [17]. According to the multiplicity of features and also the existence of a correlation between some features, with the opinion of a specialist doctor, essential features were selected and reduced from 30 basic features to 12 basic features, which include 4 clinical features and 8 laboratory findings. The details of the selected features can be seen in Table 1 and Table 2. In Table 1, row 13 is the record class label, where the numbers 0 and 1 are taken to mean being negative or positive of the Covid-19 disease.

3.3 Building model

The dataset used in this research has specific classes, so 7 of the most used supervised machine learning algorithms in medicine [38], [39] have been used for the early diagnosis of

patients with covid-19. The models used are as follows: XGBoost, Neural Networks (NN), Naïve Bayes (NB),

Table 1. Information About The Dataset

No.	Feature	Min	Max	Mean	Std
1	Saturation o2	46	100	92.22	8
2	RR	8	80	16.42	4.01
3	PR	50	160	79.48	14.45
4	Clinical symptoms	0	8	2.10	1.58
5	CRP	0	357	39.9	49.9
6	CPK	14	5280	207.9	344.2
7	LDH	218	3094	564.5	319.4
8	D-Dimer	1	5000	675.9	996.1
9	WBC	0.8	123.8	7.55	4.78
10	Lymphocytes	11.62	10275.4	185.4	307.8
11	ESR	2	1545	357.5	456.3
12	Ferritin	9	1863	372.8	392.2
13	Result	0	1	0.55	0.49

Table 2. Description of Data Set Features

Type	Feature	Description
Clinical Variable	Saturation o2	Percent of hemoglobin that is bound to molecular oxygen
	RR(Respiratory Rate)	Number of person breaths in one minute
	PR(Pulse Rate)	The average of heart beats in one minute
	Clinical symptoms	Patient symptoms and signs, disease manifestation in a patient
Laboratory Variable	CRP(C-Reactive protein)	It is a substance that produces in the liver and rises in response to inflammation
	CPK(Creatine phosphokinase)	It is an enzyme that is found mainly in the heart, brain and skeletal muscle and can rise in response to viral infections.
	LDH(lactate dehydrogenase)	It is an enzyme that is found in all live cells and can rise in response to viral infections.
	D-Dimer	It is a protein fragment that's made when a blood clot dissolves in the body and rises in response to inflammation.
	WBC(White Blood Cell)	It is the cells of the immune system and involve in protecting the body against infectious disease.
	Lymphocytes	It is a type of white blood cell in the immune system
	ESR(Erythrocyte Sedimentation Rate)	It is a blood test that measures the rate of settling red blood cells at the bottom of a test tube and rises in response to inflammation.
	Ferritin	It is a blood protein that stores iron and rises in response to inflammation

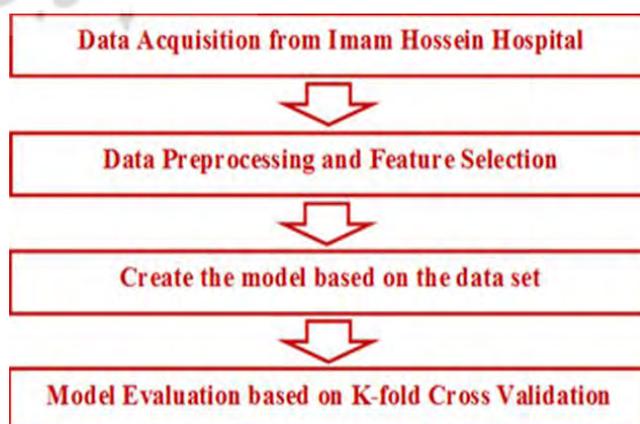


Figure 1. Research methodology stages

Random Forest (RF), Decision Tree (DT), Support Vector Machine (SVM) and K-Nearest Neighbor (KNN).

3.4 Model evaluation

In this paper, the K-Fold Cross Validation method has been used to evaluate the results of the mentioned models. In this research, the results are evaluated based on k=5, k=10

4. Comparison And Evaluation

A system with specifications of Intel(R) Core(TM) i5-3210M CPU @ 2.50 GHz, 8 GB memory and Windows 10 (64-bit) operating system was used in this research to implement and evaluate classification models.

4.1 Correlation Coefficient analysis

In order to clarify the nature of the collected dataset as much as possible, the analysis of the correlation coefficient between the 12 features in the dataset has been used. Figure 2 and Figure. 3 show the results of these analyzes that show the relationship of each feature with all other features (and itself) in the numerical format in the range of 0 and 1 and the color spectrum from light to dark (both respectively meaning the lowest and highest value) are displayed. Independent features and dependent features can be identified as a result of this analysis.

4.2 Evaluation Criteria

The evaluation criteria of the models used in this research include Accuracy, Precision, Recall, F1-Score, and the method of calculating these criteria can be seen in Equ. (1) to (4).

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

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (4)$$

4.3 Results

The evaluation results of different models based on the evaluation criteria for the values of 5 and 10 k can be seen in Figure 2 and Figure 3.

As can be seen in Figure 2 and Figure 3, for both values of k, K-NN and XGBoost models have the best performance compared to other models. Another noteworthy point is that the neural network model has shown better performance for K equal to 5 compared to K equal to 10. Also, the confusion matrix of the seven models are presented in the Figure 4 and Figure 5.

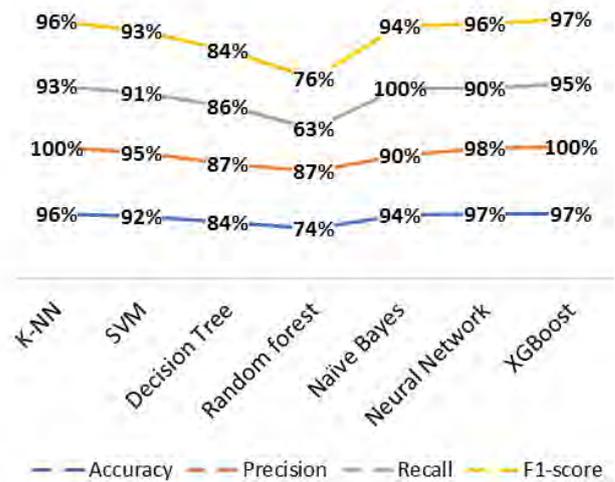


Figure. 2. Performance evaluation Result (5-fold)

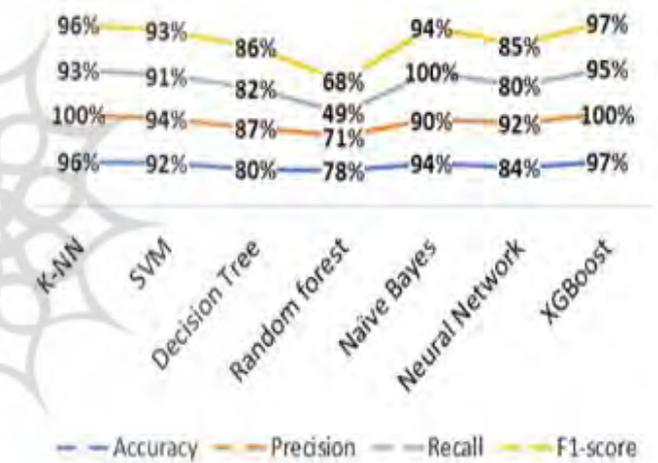


Figure. 3. Performance evaluation Result (10-fold)

In order to deepen the results of the ROC diagram, the AUC value of each model was calculated and as can be seen in Figure 6 and Figure 7, the ROC diagram for each model is presented separately. Accordingly, Naive Bayes and K-NN show higher AUC values.

Based on the results of the XGBoost model, we found that in our prediction model, WBC, Lymphocytes, LDH Serum, D-Dimer, Ferritin, ESR, CPK, and saturation are the most important features Figure 8.

According to the order of importance of features obtained in XGBoost modeling, in the continuation of our research, we re-evaluated all seven models with 9 important features of the XGBoost model. As can be seen in Figure 9 and Figure 10, the comparison of models using 9 features in two evaluation modes with different values of k are shown. The results indicate that in this case, K-NN and XGBoost models have shown the best performance compared to other models. After these two, the Naive Bayes model is in third place.



پرویشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

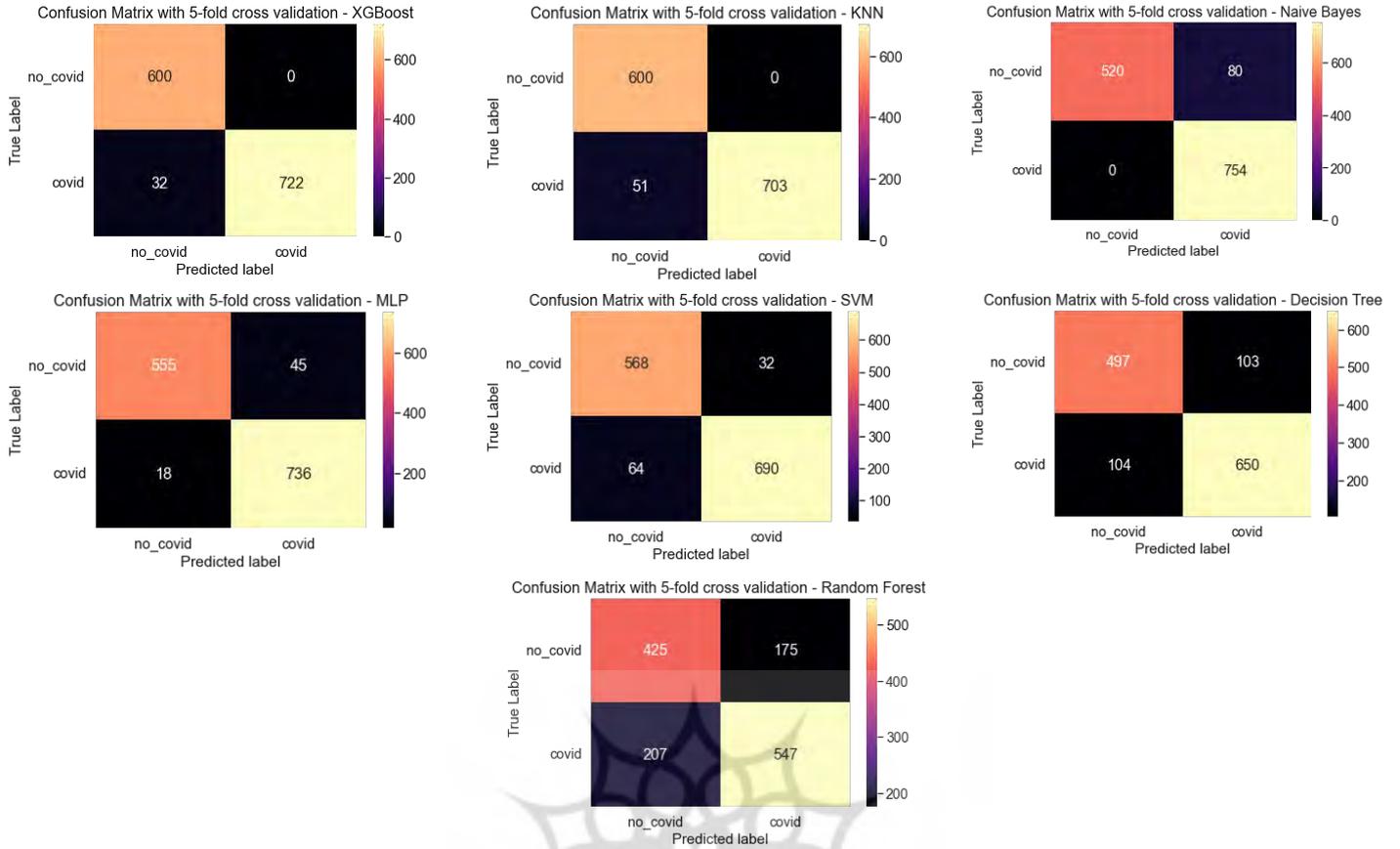


Figure 4. Confusion matrix of prediction(5-Fold Cross Validation)

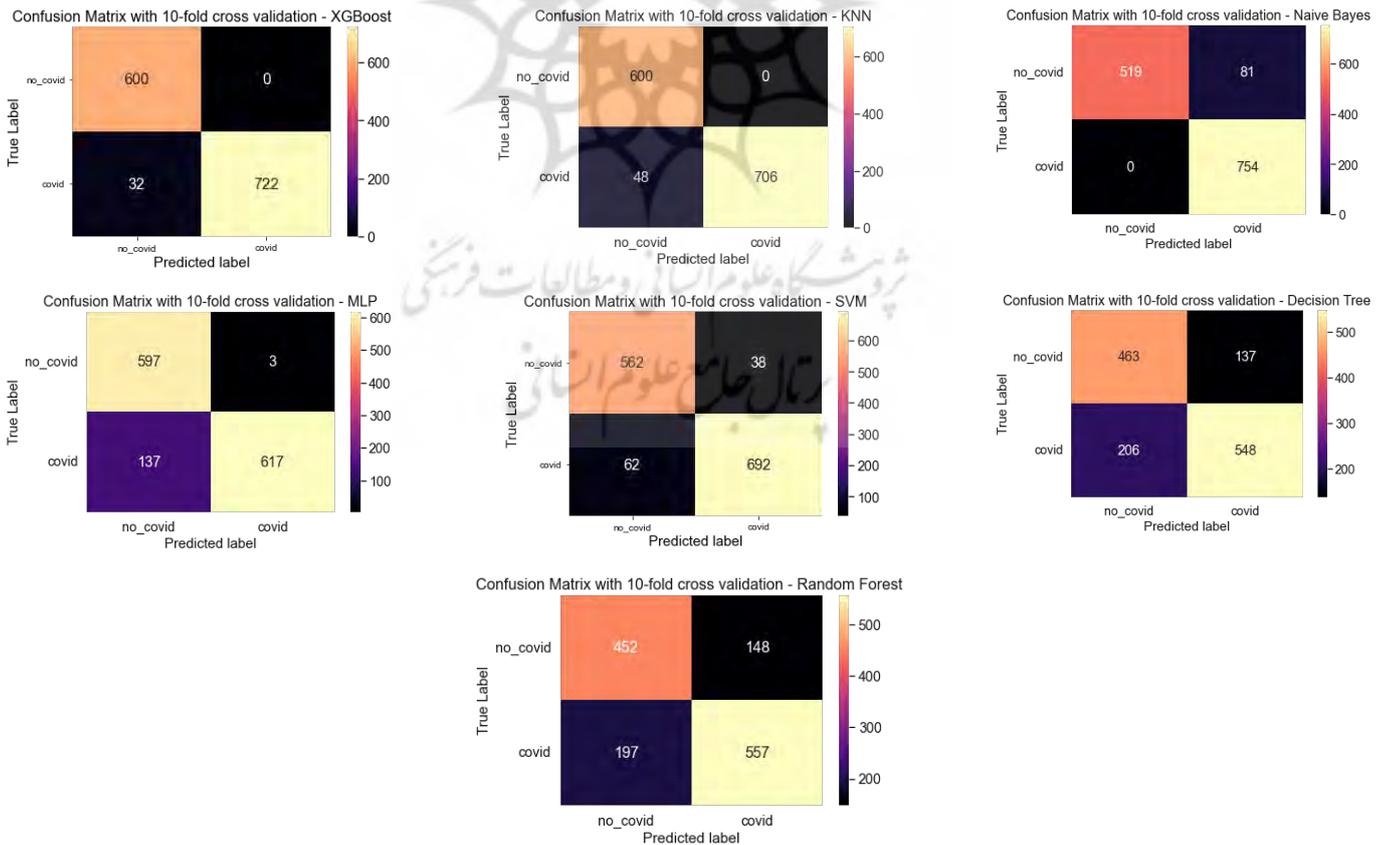


Figure 5. Confusion matrix of prediction(10-Fold Cross Validation)

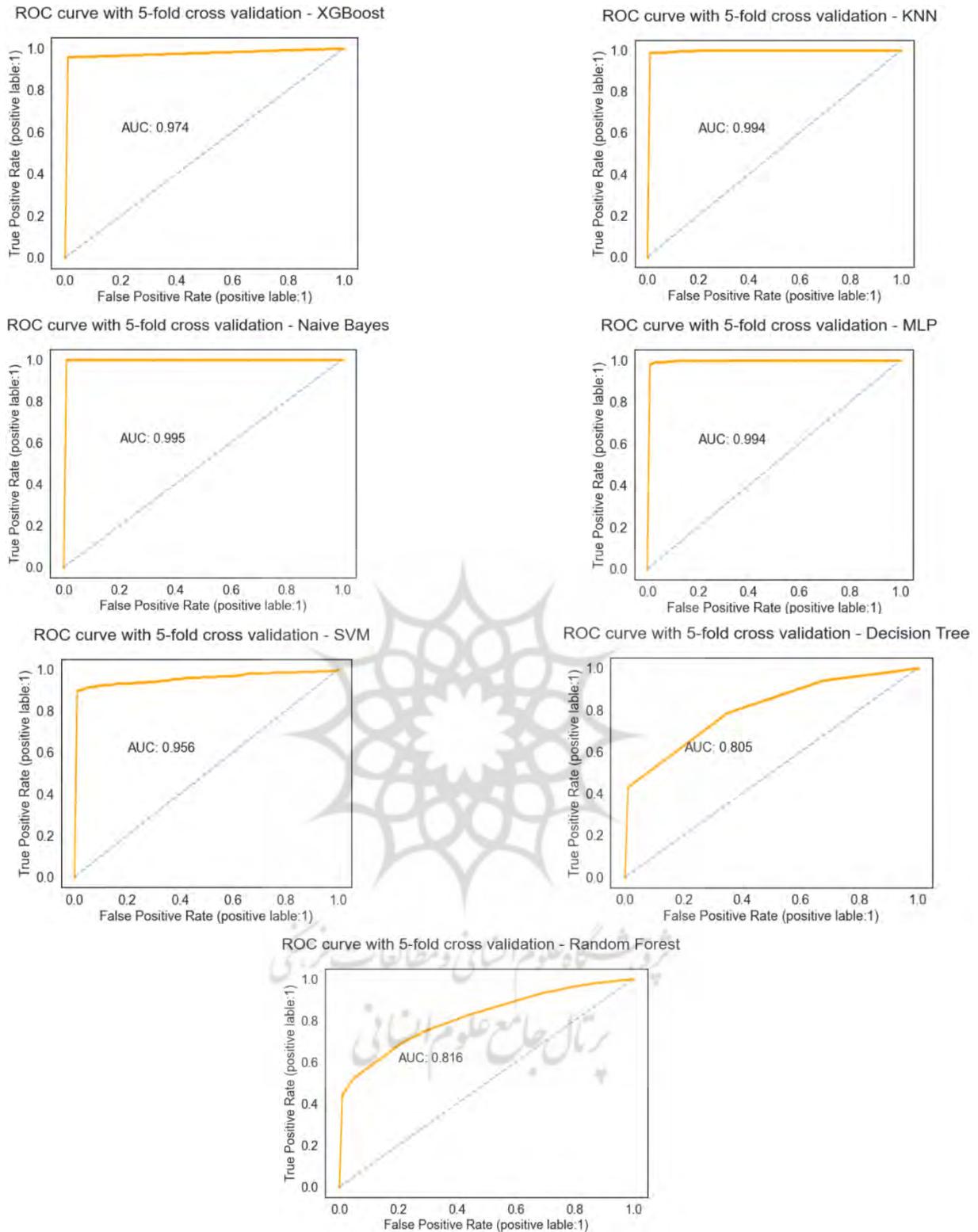


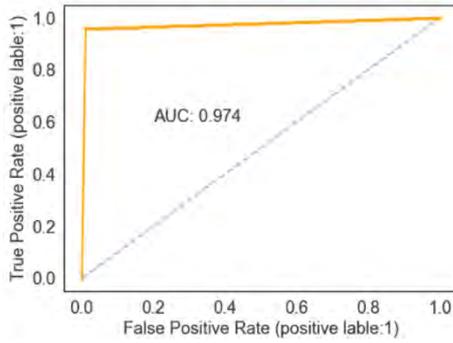
Figure. 6. ROC curve of the Different models(5-Fold Cross Validation)

Regarding the importance of features in the XGBoost model, we evaluated again the models using 9 of the most important features.

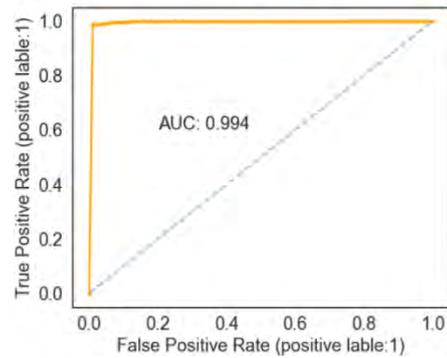
Figure 9 and Figure 10, show a comparison between the results of the seven models used in this paper in terms of evaluation criteria for 5 and 10-fold Cross-Validation using nine features.

Also, as seen in Figure 11 and Figure 12, the Accuracy criterion, as one of the important criteria in comparing models, has the same values in both K-NN and XGBoost models in two evaluation modes with 9 and 12 features. In other words, reducing the features from 12 to 9 in the two mentioned models did not decrease the accuracy.

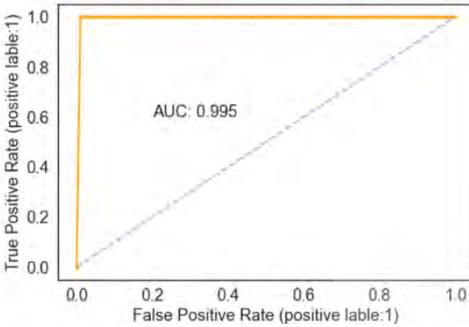
ROC curve with 10-fold cross validation - XGBoost



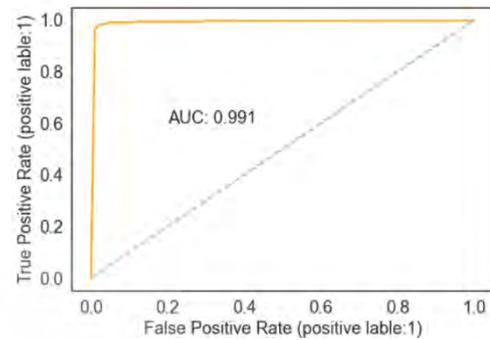
ROC curve with 10-fold cross validation - KNN



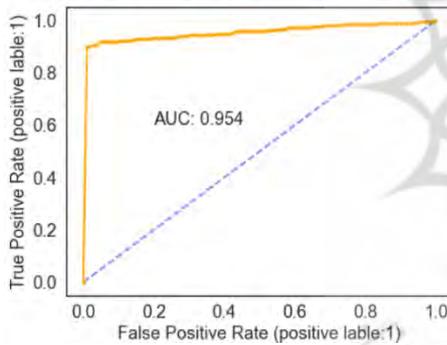
ROC curve with 10-fold cross validation - Naive Bayes



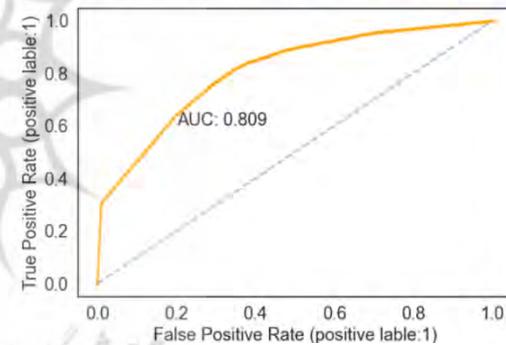
ROC curve with 10-fold cross validation - MLP



ROC curve with 10-fold cross validation - SVM



ROC curve with 10-fold cross validation - Decision Tree



ROC curve with 10-fold cross validation - Random Forest

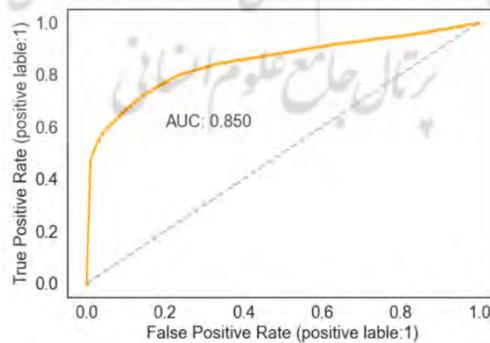


Figure 7. ROC curve of the Different models(10-Fold Cross Validation)

As shown in Figure 8 and Figure 9, the XGBoost and K-NN in both experiments perform better than the other models based on 5 and 10-fold Cross-Validation. Also, Naïve Bayes and neural networks are in the next position. In addition, the results showed that the decision tree and the random forest in the dataset of this study do not perform well compared to other models. Figure 13 shows the Accuracy of all models in

terms of evaluation criteria for 5 and 10-fold Cross-Validation. XGBoost and K-NN have the highest Accuracy.

As can be seen, based on the comparison of the results in 9 and 12 features, the XGBoost and KNN methods have performed better than other methods. In other words, by reducing the features from 12 to 9 while reducing the overhead, almost similar results can be obtained.

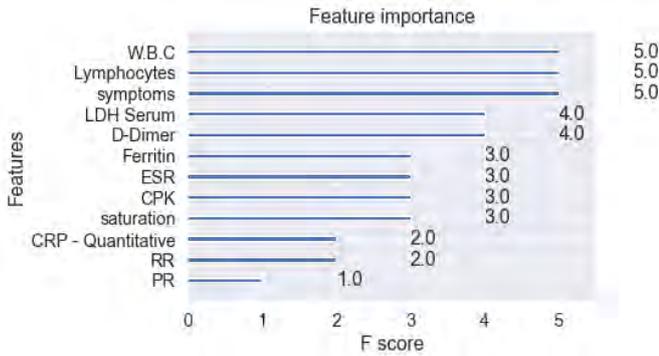


Figure 8. Variables most important for classification, listed in decreasing order of importance

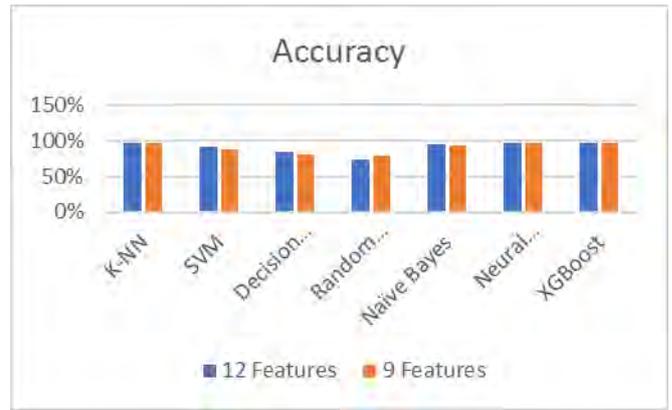


Figure 11. The Resulted Accuracy of the Different models(5-Fold Cross Validation)

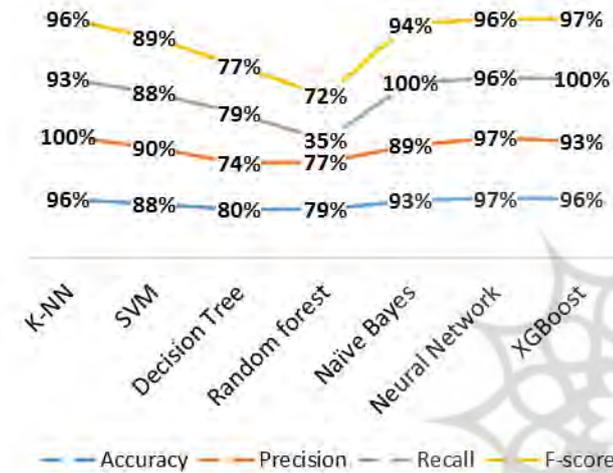


Figure 9. Performance evaluation result(5-Fold Cross Validation)

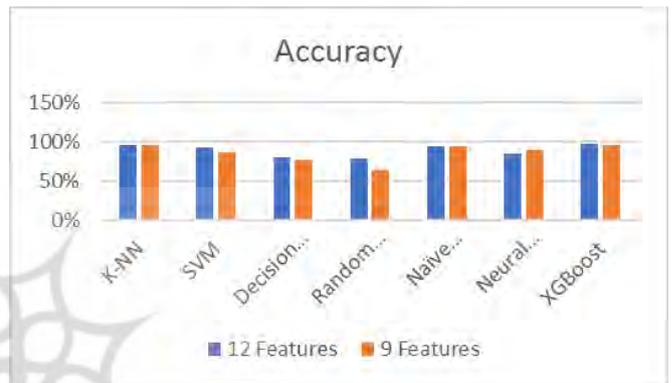


Figure 12. The Resulted Accuracy of the Different Models(10-Fold Cross Validation)

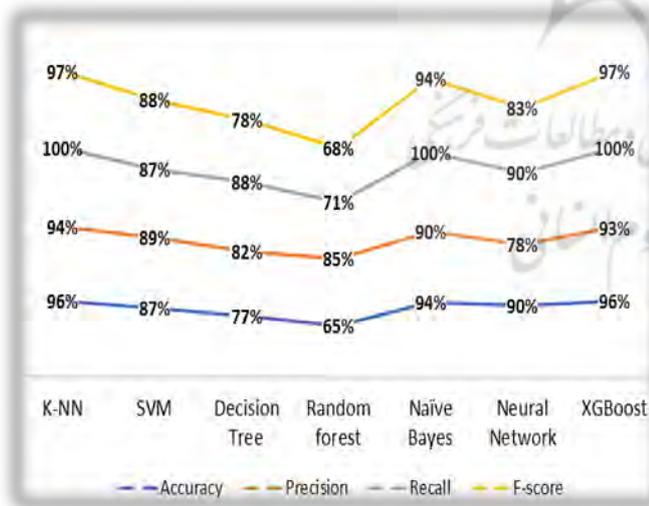


Figure 10. Performance evaluation result(10-Fold Cross Validation)

5. Conclusion

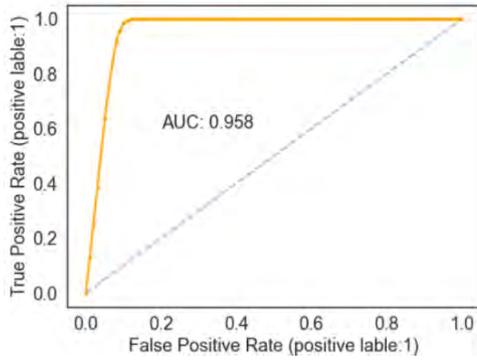
The COVID-19 pandemic has had a significant impact on the life of people and public health around the world, putting the world's health care systems at risk and causing irreparable damage. In order to prevent the spread of this disease, the first step is early diagnosis of it and, as a result, the reduction of morbidity and mortality. In this regard, the use of new technologies based on artificial intelligence, especially in the field of machine learning, can play a significant role in this matter.

In this research, seven supervised machine learning techniques, namely K-NN, SVM, Decision Tree, Random Forest, Naive Bayes, Neural Network and XGBoost, were tested and evaluated for the diagnosis of Covid-19 based on a brand new Iranian dataset. The results of these evaluations were presented in detail in the form of evaluation criteria, ROC curves, confusion matrices, and other comparative charts in the text of the paper. Finally, the results of the research show that the K-NN and XGBoost models have the best performance compared to the rest of the models.

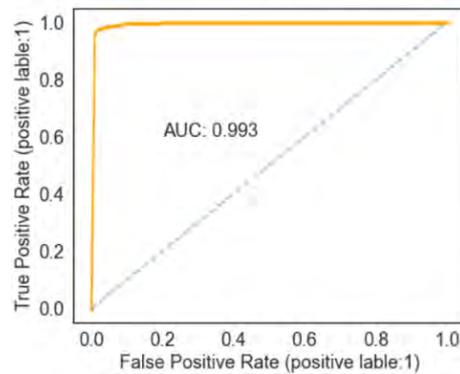
In order to check the results of the ROC diagram and the AUC value based on 9 features in more detail, and as can be seen in Figure 7 and Figure 13, Naive Bayes and K-NN show higher AUC values. As shown in Figure 13, using 9 of the most important features, the AUC is reduced in the XGBoost and SVM models. In contrast, this value for the decision tree and the random forest has increased.

The dataset used in this research included 12 features for 1354 patients that one by one were reported from rereading the files and medical records of the patients in the hospital, it is obvious that the cost of collecting such a dataset is very high. Therefore, in the following, we tried to distinguish more important features from less important features based on the results of the XGBoost algorithm in order to reduce the

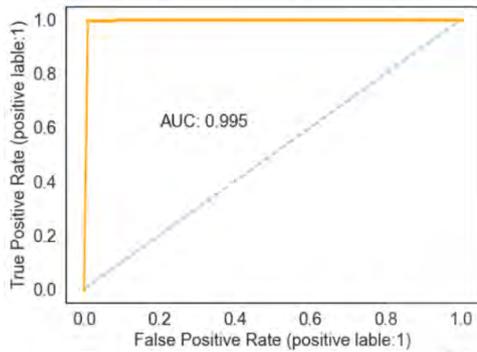
ROC curve with 10-fold cross validation - XGBoost



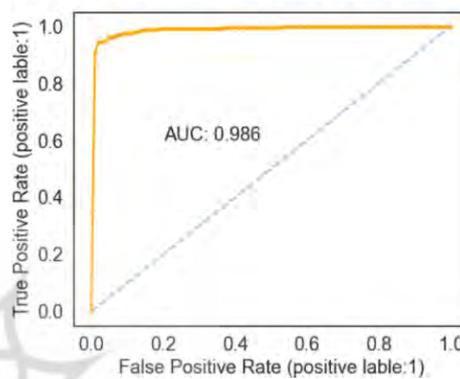
ROC curve with 10-fold cross validation - KNN



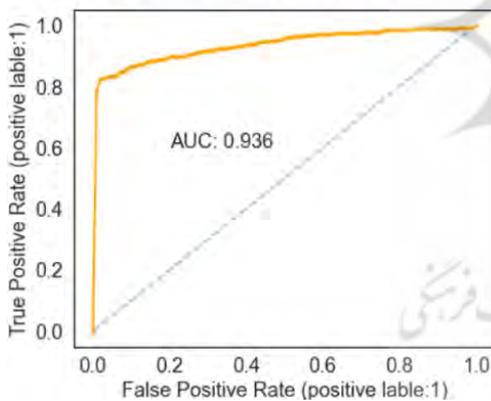
ROC curve with 10-fold cross validation - Naive Bayes



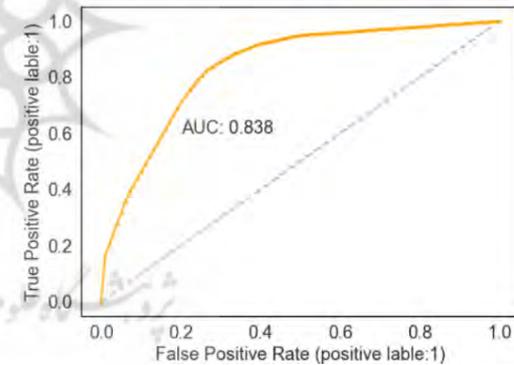
ROC curve with 10-fold cross validation - MLP



ROC curve with 10-fold cross validation - SVM



ROC curve with 10-fold cross validation - Decision Tree



ROC curve with 10-fold cross validation - Random forest

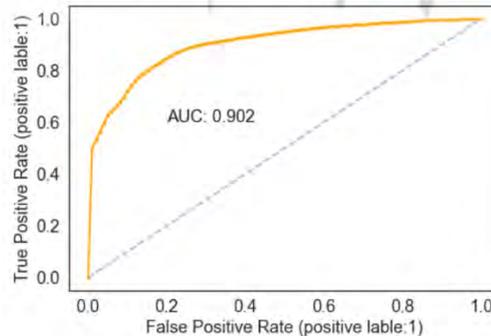


Figure. 13. ROC Curve Of The Different Models For Nine Features(10-Fold Cross Validation)

dataset collection costs as well as the system computing costs as well. Based on the results, 3 features, namely PR, RR and CRP-Quantitative were recognized as less important features

and to prove the correctness of such a claim, all models were trained and evaluated again by removing these three features from the dataset (only 9 features remained) and the

performance results of this experiment were also reported in detail. The final results showed that there was no significant change in the performance of the models, and if necessary, the three mentioned features can be ignored to reduce costs. Also, as stated, the data used in this research were the clinical and laboratory data of the patients. The use of other types of data such as the images of the lungs of the patients (CT-scans or X-rays) in a combined manner along with these data and comparing the results of the models can be a good solution for further studies on this issue.

Declarations

Funding

This research did not receive any grant from funding agencies in the public, commercial, or non-profit sectors.

Authors' contributions

NM: Study design, acquisition of data; MT: Study design, acquisition of data; MA: Statistical analysis, drafting the manuscript, revision of the manuscript; MK: Supervision, revision of the manuscript; AK: Study design, Interpretation of the results.

Conflict of interest

The authors declare that there is no conflict of interest.

References

- [1] J. Nayak, B. Naik, P. Dinesh, K. Vakula, B. K. Rao, W. Ding, and D. Pelusi. "Intelligent system for COVID-19 prognosis: A state-of-the-art survey", *Applied Intelligence*, vol. 51, no. 5, pp. 2908-2938, 2021.
- [2] M. Greco, G. Angelotti, P. F. Caruso, A. Zanella, N. Stomeo, E. Costantini, A. Protti, A. Pesenti, G. Grasselli, and M. Cecconi. "Outcome prediction during an ICU surge using a purely data-driven approach: a supervised machine learning case-study in critically ill patients from COVID-19 Lombardy outbreak", *International Journal of Medical Informatics*, vol. 164, p. 104807, 2022.
- [3] D. Tiwari, B. S. Bhati, F. Al-Turjman, and B. Nagpal. "Pandemic coronavirus disease (Covid-19): World effects analysis and prediction using machine-learning techniques", *Expert Systems*, vol. 39, no. 3, p. e12714, 2022.
- [4] O. R. Shahin, H. H. Alshammari, A. I. Taloba, and R. M. Abd El-Aziz, "Machine Learning Approach for Autonomous Detection and Classification of COVID-19 Virus". *Computers and Electrical Engineering*, vol. 101, p. 108055, 2022.
- [5] K. M. Kuo, P. C. Talley, and C. S. Chang, "The accuracy of machine learning approaches using non-image data for the prediction of COVID-19: A meta-analysis", *International Journal of Medical Informatics*, vol. 164, p. 104791, 2022.
- [6] M. Moosazadeh, P. Ifaei, A. S. Tayerani Charmchi, S. Asadi, and C. Yoo, "A machine learning-driven spatio-temporal vulnerability appraisal based on socio-economic data for COVID-19 impact prevention in the US counties", *Sustainable Cities and Society*, vol. 83, p. 103990, 2022.
- [7] N. Mohebbi, M. Alavi, and M. Kargari, "Determining COVID-19 Severity with Fuzzy Inference System". In *2022 27th International Computer Conference, Computer Society of Iran (CSICC)*, IEEE, 2022, pp. 1-5.
- [8] A. Anuja, and A. Bansal, "Novel coronavirus (COVID-19) diagnosis using computer vision and artificial intelligence techniques: a review", *Multimedia tools and applications*, vol. 80, no. 13, pp. 19931-19946, 2021.
- [9] M. T. Huyut, "Automatic Detection of Severely and Mildly Infected COVID-19 Patients with Supervised Machine Learning Models", *IRBM*, In Press, 2022.
- [10] E. Avuçlu, "A Novel Method Using Covid-19 Dataset and Machine Learning Algorithms FOR THE MOST ACCURATE DIAGNOSIS That can be Obtained in Medical Diagnosis", *Biomedical Signal Processing and Control*, vol. 77, p. 103836, 2022.
- [11] K. Goel, R. Sindhgatta, S. Kalra, R. Goel, and P. Mutreja, "The effect of machine learning explanations on user trust for automated diagnosis of COVID-19", *Computers in Biology and Medicine*, vol. 146, p. 105587, 2022.
- [12] A. de Fátima Cobre, M. Surek, D. P. Stremel, M. M. Fachi, H. H. L. Borba, F. S. Tonin, and R. Pontarolo, "Diagnosis and prognosis of COVID-19 employing analysis of patients' plasma and serum via LC-MS and machine learning", *Computers in biology and medicine*, vol. 146, p. 105659, 2022.
- [13] M. Shanbehzadeh, H. Kazemi-Arpanahi, and R. Nopour, "Performance evaluation of selected decision tree algorithms for COVID-19 diagnosis using routine clinical data", *Medical Journal of the Islamic Republic of Iran*, vol. 35, p. 29, 2021.
- [14] R. Nopour, H. Kazemi-Arpanahi, M. Shanbehzadeh, and A. Azizifar, "Performance analysis of data mining algorithms for diagnosing COVID-19", *Journal of Education and Health Promotion*, vol. 10, p. 405, 2021.
- [15] G. B. Huang, Q. Y. Zhu, and C. K. Siew, "Extreme learning machine: theory and applications", *Neurocomputing*, vol. 70, no. 1-3, pp. 489-501, 2006.
- [16] A. Zimmerman, and D. Kalra, "Usefulness of machine learning in COVID-19 for the detection and prognosis of cardiovascular complications", *Reviews in Cardiovascular Medicine*, vol. 21, no. 3, pp. 345-352, 2020.
- [17] D. Tiwari, B. S. Bhati, F. Al-Turjman, and B. Nagpal, "Pandemic coronavirus disease (Covid-19): World effects analysis and prediction using machine-learning techniques", *Expert Systems*, vol. 39, no. 3, p. e12714, 2022.
- [18] E. S. Adamidi, K. Mitsis, and K. S. Nikita, "Artificial intelligence in clinical care amidst COVID-19 pandemic: A systematic review", *Computational and Structural Biotechnology Journal*, vol. 19, pp. 2833-2850, 2021.
- [19] L. Li, L. Qin, Z. Xu, Y. Yin, X. Wang, B. Kong, J. Bai et al., "Using artificial intelligence to detect COVID-19 and community-acquired pneumonia based on pulmonary CT: evaluation of the diagnostic accuracy", *Radiology*, vol. 296, no. 2, E65-E71, 2020.
- [20] A. Bansal, G. Thakur, and D. Verma, "Detection of COVID-19 Using the CT Scan Image of Lungs", In *ISIC*, vol. 21, pp. 25-27, 2021.
- [21] A. Karimiyan Abdar, S. M. Sadjadi, H. Soltanian-Zadeh, A. Bashirgonbadi, and M. Naghibi, "Automatic detection of coronavirus (COVID-19) from chest CT images using VGG16-based deep-learning", In *2020 27th National and 5th International Iranian Conference on Biomedical Engineering (ICBME)*, IEEE, 2020, pp. 212-216.
- [22] V. Madaan, A. Roy, C. Gupta, P. Agrawal, A. Sharma, C. Bologa, and R. Prodan, "XCOVNet: chest X-ray image classification for COVID-19 early detection using convolutional neural networks", *New Generation Computing*, vol. 39, no. 3, pp. 583-597, 2021.
- [23] S. Thakur, and A. Kumar, "X-ray and CT-scan-based automated detection and classification of covid-19 using convolutional neural networks (CNN)", *Biomedical Signal Processing and Control*, vol. 69, p. 102920, 2021.
- [24] E. Benmalek, J. Elmhamdi, and A. Jilbab, "Comparing CT scan and chest X-ray imaging for COVID-19 diagnosis", *Biomedical Engineering Advances*, vol. 1, p.100003, 2021.
- [25] S. Liang, H. Liu, Y. Gu, X. Guo, H. Li, L. Li, Z. Wu, M. Liu, and L. Tao, "Fast automated detection of COVID-19 from medical images using convolutional neural networks", *Communications Biology*, vol. 4, no. 1, pp. 1-13, 2021.
- [26] T. Rahman, A. Khandakar, F. F. Abir, M. A. A. Faisal, M. S. Hossain, K. K. Podder, T. O. Abbas et al., "QCovSML: A reliable COVID-19 detection system using CBC biomarkers by a stacking machine learning model", *Computers in Biology and Medicine*, vol. 143, p. 105284, 2022.
- [27] H. C. Çubukçu, D. İ. Topcu, N. Bayraktar, M. Gülşen, N. Sarı, and A. H. Arslan, "Detection of COVID-19 by machine learning using routine laboratory tests". *American journal of clinical pathology*, vol. 157, no. 5, pp. 758-766, 2022.
- [28] S. Ali, Y. Zhou, and M. Patterson, "Efficient analysis of covid-19 clinical data using machine learning models", *Medical & Biological Engineering & Computing*, vol. 60, pp. 1815-1825.

- [29] A. F. de Moraes Batista, J. L. Miraglia, T. H. R. Donato, and A. D. P. Chiavegatto Filho, "COVID-19 diagnosis prediction in emergency care patients: a machine learning approach". *MedRxiv*, 2020.
- [30] Brinati, Davide, Andrea Campagner, Davide Ferrari, Massimo Locatelli, Giuseppe Banfi, and Federico Cabitza. "Detection of COVID-19 infection from routine blood exams with machine learning: a feasibility study." *Journal of medical systems* 44, no. 8 (2020): 1-12.
- [31] L. J. Muhammad, M. Islam, S. S. Usman, and S. I. Ayon, "Predictive data mining models for novel coronavirus (COVID-19) infected patients' recovery", *SN Computer Science*, vol. 1, no. 4, pp. 1-7, 2020.
- [32] V. A. de Freitas Barbosa, J. C. Gomes, M. A. de Santana, J. E. D. A. Albuquerque, R. G. de Souza, R. E. de Souza, and W. P. dos Santos, "Heg. IA: an intelligent system to support diagnosis of Covid-19 based on blood tests", *Research on Biomedical Engineering*, vol. 38, no. 1, pp. 99-116, 2022.
- [33] M. A. Alves, G. Z. Castro, B. A. S. Oliveira, L. A. Ferreira, J. A. Ramirez, R. Silva, and F. G. Guimarães, "Explaining machine learning based diagnosis of COVID-19 from routine blood tests with decision trees and criteria graphs", *Computers in Biology and Medicine*, vol. 132, p. 104335, 2021.
- [34] H. S. Yang, Y. Hou, L. V. Vasovic, P. A. Steel, A. Chadburn, S. E. Racine-Brzostek, P. Velu et al., "Routine laboratory blood tests predict SARS-CoV-2 infection using machine learning", *Clinical chemistry*, vol. 66, no. 11, pp. 1396-1404, 2020.
- [35] M. AlJame, I. Ahmad, A. Imtiaz, and A. Mohammed, "Ensemble learning model for diagnosing COVID-19 from routine blood tests", *Informatics in Medicine Unlocked*, vol. 21, pp. 100449, 2020.
- [36] A. Banerjee, S. Ray, B. Vorselaars, J. Kitson, M. Mamalakis, S. Weeks, M. Baker, and L. S. Mackenzie, "Use of machine learning and artificial intelligence to predict SARS-CoV-2 infection from full blood counts in a population". *International immunopharmacology*, vol. 86, p. 106705, 2020.
- [37] Y. Cao, Z. Xu, J. Feng, C. Jin, X. Han, H. Wu, and H. Shi, "Longitudinal assessment of COVID-19 using a deep learning-based quantitative CT pipeline: illustration of two cases", *Radiology: Cardiothoracic Imaging*, vol. 2, no. 2, p. e200082, 2020.
- [38] A. M. Rahmani, E. Yousefpoor, M. S. Yousefpoor, Z. Mehmood, A. Haider, M. Hosseinzadeh, and R. A. Naqvi, "Machine learning (ML) in medicine: review, applications, and challenges", *Mathematics*, vol. 9, no. 22, p. 2970, 2021.
- [39] S. Uddin, A. Khan, M. E. Hossain, and M. A. Moni, "Comparing different supervised machine learning algorithms for disease prediction", *BMC medical informatics and decision making*, vol. 19, no. 1, pp. 1-16, 2019.



Meysam Alavi was born in Hamadan, Iran. He received his B.Sc. and an M.Sc. degree in Computer Engineering in 2005 and 2010, respectively. Also, he received his Second M.Sc. degree in Information Technology at Tarbiat Modares University in 2020. His research interests include medical image processing, machine learning, data mining, and social network analysis.



Mehrdad Kargari received his PhD degree in industrial engineering from Tarbiat Modares University of Iran. He is currently an assistant professor at the department of Information Engineering, in Tarbiat Modares University. His research interests are in the fields of machine learning, artificial intelligence, IoT and their applications in health or Banking.



Amir behnam kharazmi was born in 1985. He is an internal medicine specialist, pulmonary subspecialist, and assistant professor at Shahid Beheshti medical university. Also, he is a member of the European Respiratory Society(ERS) and American thoracic society(ATS). His research interest is in the field of the application of artificial intelligence in medicine.



Narges Mohebbi was born in 1996 in Tehran. She received her B.Sc. in Computer Engineering from Arak University in 2018. Currently, she is an M.Sc. student in Information Technology at Tarbiat Modares University. Her research interests include data mining and machine learning.



Mehdi Tutunchian was born in 1996 in Tabriz. He received his B.Sc. in Information Technology Engineering from Tabriz University in 2018. Currently, he is an M.Sc. student in Information Technology at Tarbiat Modares University. His research interests include medical image processing and machine learning.