Ann Med Res
Current issue list available at AnnMedRes
Annals of Medical Research
journal page: www.annalsmedres.org

The use of machine learning method in COVID-19 patient management

Zubeyir Cebeci[a,*], Omer Kaan Baykan[b], Ilker Coskun[a], Ebru Canakci[a]

[a]*Ordu University, Faculty of Medicine, Department of Anesthesiology and Reanimation, Ordu, Türkiye*
[b]*Konya Technical University, Faculty of Engineering and Natural Science, Department of Computer Engineering, Konya, Türkiye*

ARTICLE INFO

Keywords:
Machine learning
COVID-19
Outbreak

Received: Mar 18, 2024
Accepted: Oct 02, 2024
Available Online: 29.11.2024

DOI:
10.5455/annalsmedres.2024.03.054

Abstract

Aim: The COVID-19 pandemic, first originating in Wuhan, China in December 2019, has affected over 180 countries worldwide. The clinical spectrum of COVID-19 ranges from mild to severe pneumonia with acute respiratory distress syndrome. The sudden increase in COVID cases requiring hospitalization has made inpatient health institutions difficult to predict and manage. Machine learning models have been used to diagnose the disease, predict clinical course, and hospital stay.

Materials and Methods: Data from 322 PCR-positive patients were analyzed, including demographics, comorbidities, laboratory values, and radiological results. Machine learning algorithms such as Logistic Regression, Support Vector Machine, Ensemble Methods, and K-Nearest Neighbor were used for classification.

Results: Results showed that SVM provided the best classification performance. The model considered factors like age, gender, medical history, and test results to personalize treatment decisions. The study suggests that machine learning can improve patient care during the COVID-19 pandemic. Limitations include the need for validation with larger datasets from multiple centers.

Conclusion: This study aimed to show whether machine learning techniques can be used to make decisions about the hospitalization of COVID-19 patients.

Introduction

The world is grappling with the COVID-19 outbreak, which first appeared in Wuhan, China in December 2019 and was declared a pandemic by the World Health Organization in March 2020 [1]. The COVID-19 pandemic (COVID-19) associated with SARS-CoV-2 virus infection has affected more than 180 countries worldwide. The clinical spectrum of COVID-19 ranges from asymptomatic or mild upper airway symptoms to potentially fatal, severe pneumonia with acute respiratory distress syndrome (ARDS) [2]. The COVID-19 pandemic continued to affect the whole world for a long time, increasing the number of cases from time to time due to the mutation of the virus and the high contagiousness of its variants [3]. In particular, the sudden increase in COVID cases requiring hospitalization and long-term hospitalizations has turned into a chaos that inpatient health institutions cannot predict and have difficulty managing. In some studies, machine learning models using hemato-chemical from routine blood exams and radiological values were used to diagnose the disease, predict the clinical course and hospital stay [4-6]. The aim of this study was to show whether modelling using machine learning (ML) techniques can be used to make decisions about the hospitalisation of COVID-19 patients.

Materials and Methods

Study populations

After ethical approval for the study was given by the Clinical Research Ethics Committee of Ordu University (date:21.01.2021 number: 2021/19), PCR-positive patients who applied to the Ordu University Training and Research Hospital adult emergency service COVID area or COVID polyclinics between June 2020 and January 2021 were retrospectively screened from the hospital information management system. Age, gender, morbidity, laboratory data and places of stay (outpatient, intensive care unit or COVID service) of the people who applied to the hospital were recorded (337 patients). Of 337 patients, 322 with complete study data; were included in the study. Outpatients were classified as Group I and hospitalized patients as Group II.

*Corresponding author:
Email address: zubeyircebeci@gmail.com (Zubeyir Cebeci)

Study parameters

Age, gender, co-morbidity, hemogram values (neutrophil (NEU), lymphocytes (LYM) and neutrophil/lymphocytes ratio (NEU/LYM) biochemical markers (alanine aminotransferase (ALT), aspartate aminotransferase (AST), C-reactive protein (CRP), D-dimer, ferritin, lactate dehydrogenase (LDH)) and radiological imaging results (as consistent with COVID, incompatible with COVID) were recorded.

Preprocessing of data for machine learning classification

Logistic Regression, Support Vector Machine (SVM), Ensemble Methods, K-Nearest Neighbor (K-NN) methods were used with the MATLAB program to decide whether the patients who applied to the hospital should be treated on an outpatient basis or in the hospital. Age, gender, and laboratory values (NEU, LYM, NEU/LYM, AST, ALT, CRP, D-dimer, ferritin, LDH), obtained from medical records were evaluated as predictive variables by the machine learning algorithms. A 5-fold cross-validation method was used in the analysis.

Statistical analysis

Data normality was tested using the Kolmogorov–Smirnov test. Categorical variables were expressed as frequencies and percentages. For continuous variables, we used the mean and standard deviation or the median and range to summarize normally and non-normally distributed data respectively. Continuous variables were compared using the Student t-test or Mann–Whitney U test for normally and non-normally distributed data, respectively. For comparison of categorical variables, the Chi-square test and Fisher's exact test were used when appropriate. Feature correlations were measured using the Spearman correlation coefficient. The level of statistical significance was set at $p<0.05$.

Results

During the study period, a total of 337 confirmed COVID-19 patients were evaluated but only 322 patients with full study data were included. Of those, 132 confirmed COVID-19 patients were hospitalized in 89 COVID-19 designated departments and 43 in ICU departments. Outpatient treatments were planned for the other 190 patients. Of the patients surveyed, 169 were men and 153 were women. Although 177 patients had comorbidities, 145 had no comorbidities. The number of patients with infiltrates consistent with viral pneumonia on radiological imaging was 183 (Table 1).

Table 1. Results of patients in terms of gender distribution, presence of comorbidity and consistent radiological findings.

		Treatment		
		Group I (n)	Group II (n)	p-value*
Gender	Male	93	76	0.128
	Female	97	56	
Comorbidities	Yes	70	107	<0.0001*
	No	120	25	
Radiological appearance	Inconsistent	139	0	<0.0001*
	Consistent	51	132	

Table 2. Laboratory results and treatment modalities of COVID-19 patients.

	Treatment		
	Group I (n=190) Median(IQR)	Group II (n=132) Median(IQR)	p-value*
Age	50.5(36.75-60)	69.5(58.25-79)	<0.0001
ALT	20(14-31.25)	23(14-34)	0.391
AST	21(18-27)	28(19-41)	<0.0001
CRP	6.27(2.68-13.75)	80.05(30.85-144)	<0.0001
D_dimer	0.205(0.15-0.36)	0.535(0.28-1.16)	<0.0001
Ferritin	111.5(51.15-190.75)	393.5(206.25-871.50)	<0.0001
LDH	195.5(164.75-223)	294.5(231-423.25)	<0.0001
Lymphocyte	1.46(10.05-1.96)	0.89(0.60-1.34)	<0.0001
Neutrophil	5.22(4.15-6.69)	7.91(5.53-10.65)	<0.0001
Neutrophil/ Lymphocyte	3.52(2.58-4.84)	8.07(4.94-15.82)	<0.0001

LDH: Lactate dehydrogenase ALT: Alanine amino transferase AST: Aspartate amino transferase CRP: C-Reactive Protein.

Comparison between patients with outpatients vs. hospitalized COVID-19

The mean age of the patients included in the study was 57.5± 18.5. In terms of demographics and comorbidities, patients who were hospitalized had an older age ($p<0.001$) and had more comorbidities ($p<0.001$), but there was no difference in gender distribution ($p=0.128$) (Table 1).

When the hemogram values, biochemical parameters, and radiological evaluation at the time of admission to the hospital were compared, age ($p<0.001$, 95% CI,55.23-59.27), NEU ($p<0.001$,95% CI,6.51-7.62), LYM ($p<0.001$,95% CI,1.25-1.621), NEU/LYM ($p<0.001$, 95% CI,6.57-8.75), AST ($p<0.001$,95% CI,26.39-32.13), LDH ($p<0.001$,95% CI,251.76-290.57), CRP ($p<0.001$,95% CI,40.37-57.06), D-Dimer ($p<0.001$,95% CI,0.15-3.32), Ferritin ($p<0.001$,95% CI,289.96-421.71), Radiological evaluation ($p<0.001$) values were statistically significant in hospitalized patients (Table 2).There was a $p<0.001$ correlation between treatment and age, comorbidities, WBC, LYM/NEU, AST, LDH, CRP, ferritin and Radiology.

Machine learning

Analyses were performed with different kernel types (linear, polynomial, Gaussian / RBF kernel) using the one vs one method for the SVM algorithm. The best results were obtained when the Gaussian/ Radial Basis Function (RBF) was used, and the kernel scale was 14 (kernel scale). Different distance metrics (Euclidean, Chebyshev, Minkowski, Mahalanobis, Hamming) and different 'neighbors' inputs (5-10) were used for the K-NN Algorithm. The best results were obtained, when the Euclidean distance criterion was used, and the 'neighbors' inputs were 10. Bagging and boosting methods were used to classify with Ensemble algorithms. The analyses were repeated by

Table 3. Model accuracy.

Classifiers	Accuracy (%)	PPV	NPV	Sensitivity (%)	Specificity (%)	AUC
Logistic Regr.	90.1	92.47	86.76	90.52	89.39	0.97
SVM	91	92.15	89.31	92.63	88.63	0.97
KNN	89.4	88.23	91.52	94.73	81.81	0.96
Ensemble(bagged)	90.7	90.40	91.13	94.21	85.61	0.97

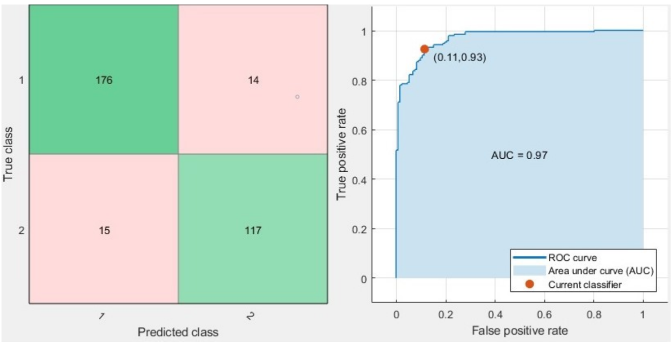

Figure 1. A: The confusion matrix of the SVM classifier, B: ROC curve of prediction.

changing the number of learner classifiers (10-30) and data split ratios. The best results were obtained in the Decision Tree algorithm when the number of classifiers was 30 and the bagging method was used.

The sensitivity, specificity, PPV, NPV, accuracy and AUC values reached by the models using Logistic Regression, Support Vector Machine (SVM), Ensemble Methods, and K-Nearest Neighbor (K-NN) classifier were presented in Table 3.

The best classification results were obtained in the model using the SVM classifier. The highest sensitivity value was found with KNN, and the highest specificity value was found with the Logistic regression classifier. The confusion matrix of the SVM classifier and the ROC curve were shown in Figure 1.

Discussion

The COVID-19 pandemic has necessitated the development of innovative machine learning (ML) models to enhance triage processes in healthcare settings. These models aim to optimize patient management during a time of overwhelming demand on medical resources. The integration of ML into triage systems can significantly improve the accuracy of patient assessments, thereby facilitating timely interventions and resource allocation.

Machine learning applications in COVID-19 triage, for predicting respiratory failure within 48 hours of admission, and predicting respiratory decompensation, have demonstrate ML's effectiveness in analyzing clinical data to stratify patients based on risk profiles. Thereby enabling individualized patient-level can decision-making and optimizing triage decisions through accurate prediction of disease severity and respiratory failure risk [7-9].

Many investigations on COVID-19 have shown that age, comorbidity, radiological alterations, neutrophilia, lymphopenia, AST, ALT, ferritin, D-dimer, and CRP elevation have an influence on the severity of the disease, mortality, hospital stay, and so on [2,10–18]. This situation leads us to believe that the results of the machine learning model that we will create using these parameters can be clinically meaningful.

When examining COVID-19 studies with machine learning models, it is discovered that different models produce different results. This demonstrates the importance of selecting a model that is appropriate for the expected result and the data set or parameters chosen [4,11-13,15–22].

Our model takes into account factors such as age, gender, medical history, chest radiographs, the results of inexpensive and rapid hematological and biochemical tests to determine whether an inpatient or outpatient treatment decision is appropriate for each individual patient. We have found that this approach allows us to provide more personalized and targeted care, resulting in better outcomes. Furthermore, our model is designed to be flexible and adaptive, so it can be quickly updated with new information or data.

Given the effectiveness of our modeling (accuracy between 89%-91%, sensitivity between 90%-94%), it is demonstrated that COVID-19 patients can be treated either as inpatients or outpatients using the data and calculation methods used. By leveraging the power of machine learning, we think it can provide the best possible decision for each patient.

Clinical evaluations performed in the past using scoring systems can now be performed using machine learning models, either now or in the near future. However, only time will tell if it will supplant the art of medicine in clinical decision-making or diagnosis.

The point we want to emphasize in our research is whether a system can be developed using machine learning that can compensate for the negative effects of insufficient manpower in events such as pandemics that put health systems in jeopardy.

There are some limitations in this study. Since it was a single-center study so the machine learning model should be tested by expanding the sample group with data obtained from other sources. After incorporating data from other centers, the prediction of the model might be slightly influenced by the variables.

Conclusion

This study has shown that machine learning algorithms that are appropriate for the data can be used to make decisions about the hospitalization of COVID-19 patients.

Ethical approval

Ethical approval for the study was given by the Clinical Research Ethics Committee of Ordu University (date:21.01.2021 number: 2021/19).

References

1. Huang C, Wang Y, Li X, Ren L, Zhao J et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet . 2020 Feb 15 ;395(10223):497–506.
2. Grasselli G, Zangrillo A, Zanella A, Antonelli M, Cabrini L, Castelli A, et al. Baseline Characteristics and Outcomes of 1591 Patients Infected with SARS-CoV-2 Admitted to ICUs of the Lombardy Region, Italy. JAMA - J Am Med Assoc. 2020 Apr 28;323(16):1574–81.
3. Kirby T. New variant of SARS-CoV-2 in UK causes surge of COVID-19. Vol. 9, The Lancet. Respiratory medicine. 2021. p. e20–1.
4. Li Wt, Ma J, Shende N, et al. Using machine learning of clinical data to diagnose COVID-19: a systematic review and meta-analysis. BMC Med Inform Decis Mak. 2020;20(1):247.
5. Vekaria B, Overton C, Wiśniowski A, et al. Hospital length of stay for COVID-19 patients: Data-driven methods for forward planning. BMC Infect Dis. 2021;21(1):700.
6. Yan L, Zhang HT, Goncalves J, et al. An interpretable mortality prediction model for COVID-19 patients. Nature Machine Intelligence. 2020;2(5):283-288.
7. Bolourani S, Brenner M, Wang P, McGinn T, Hirsch J, Barnaby D, Zanos T, Northwell COVID-19 Research Consortium A Machine Learning Prediction Model of Respiratory Failure Within 48 Hours of Patient Admission for COVID-19: Model Development and Validation J Med Internet Res 2021;23(2):e24246.
8. Burdick H, Lam C, Mataraso S, et al. Prediction of respiratory decompensation in Covid-19 patients using machine learning: The READY trial. Computers in Biology and Medicine. 2020;124:103949.
9. Rees EM, Nightingale ES, Jafari Y, Waterlow NR, Clifford S, Carl CA, et al. COVID-19 length of hospital stay: a systematic review and data synthesis. BMC Med. 2020 Sep 3;18(1):270.
10. Bird JJ, Barnes CM, Premebida C, Ekárt A, Faria DR. Country-level pandemic risk and preparedness classification based on COVID-19 data: A machine learning approach. PLoS One. 2020 Oct 1;15.
11. Alimadadi A, Aryal S, Manandhar I, Munroe PB, Joe B, Cheng X. Artificial intelligence and machine learning to fight COVID-19. Physiol Genomics. 2020;52(4):200-202.
12. Wu G, Yang P, Xie Y, Woodruff HC, Rao X, Guiot J, 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).
13. Yuan X, Huang W, Ye B, Chen C, Huang R, Wu F, et al. Detection of COVID-19 Infection from Routine Blood Examswith Machine Learning A Feasibility Study.pdf. N Engl J Med. 2020 Oct 1;382(1):1708–20.
14. Zahorec R, Hulin I, Zahorec P, Pei J, Lin X, Chen Q, et al. A novel haemocytometric COVID-19 prognostic score developed and validated in an observational multicentre European hospital-based study. IOP Conf Ser Earth Environ Sci. 2020 Oct 1;7(3):1–37.
15. Fu J, Kong J, Wang W, Wu M, Yao L, Wang Z, et al. The clinical implication of dynamic neutrophil to lymphocyte ratio and D-dimer in COVID-19: A retrospective study in Suzhou China. Thromb Res. 2020 Aug 1;192:3–8.
16. Gao Y, Cai GY, Fang W, Li HY, Wang SY, Chen L, et al. Machine learning based early warning system enables accurate mortality risk prediction for COVID-19. Nat Commun. 2020 Dec 1;11(1).
17. Zargari Khuzani A, Heidari M, Shariati SA. COVID-Classifier: an automated machine learning model to assist in the diagnosis of COVID-19 infection in chest X-ray images. Sci Rep. 2021;11(1).
18. Grasselli G, Zangrillo A, Zanella A, Antonelli M, Cabrini L, Castelli A, et al. Baseline Characteristics and Outcomes of 1591 Patients Infected With SARS-CoV-2 Admitted to ICUs of the Lombardy Region, Italy. JAMA. 2020 Apr 28;323(16):1574–81.
19. Yang HS, Hou Y, Vasovic L V., Steel PAD, Chadburn A, Racine-Brzostek SE, et al. Routine Laboratory Blood Tests Predict SARS-CoV-2 Infection Using Machine Learning. Clin Chem. 2020 Nov 1;66(11):1396–404.
20. Pan P, Li Y, Xiao Y, et al. Prognostic Assessment of COVID-19 in the Intensive Care Unit by Machine Learning Methods: Model Development and Validation. J Med Internet Res. 2020;22(11):e23128.
21. Famiglini L, Bini G, Carobene A, Campagner A, Cabitza F. Prediction of ICU admission for COVID-19 patients: A machine learning approach based on complete blood count data. Proc - IEEE Symp Comput Med Syst. 2021 Jun 1;160–5.
22. Vaid A, Somani S, Russak AJ, de Freitas JK, Chaudhry FF, Paranjpe I, et al. Machine learning to predict mortality and critical events in a cohort of patients with COVID-19 in New York City: Model development and validation. J Med Internet Res. 2020 Nov 1;22(11).