



COVID-19 DISEASE SEVERITY PREDICTION MODEL BASED ON BLOOD BIOMARKERS: A MACHINE LEARNING APPROACH

Andela Blagojević^{1,2}, Tijana Šušteršič^{1,2}, Ivan Lorencin³, Sandi Baressi Šegota³, Dragan Milovanović^{4,5}, Danijela Baskić⁴, Dejan Baskić^{4,5,6}, Zlatan Car³, Nenad Filipović^{1,2}, Nataša Zdravković^{4,5}, Sara Mijailović⁵, Nebojša Zdravković⁵

¹ Faculty of Engineering, University of Kragujevac, Sestre Janjić 6, 34000 Kragujevac, Serbia

² Bioengineering Research and Development Center (BioIRC), Prvoslava Stojanovića 6, 34000 Kragujevac, Serbia

³ Faculty of Engineering, University of Rijeka, Vukovarska 58, 51000 Rijeka, Croatia

⁴ Clinical Centre Kragujevac, Zmaj Jovina 30, 34000 Kragujevac, Serbia

⁵ Faculty of Medical Sciences, University of Kragujevac, Svetozara Markovića 69, 34000 Kragujevac, Serbia

⁶ Institute of Public Health Kragujevac, Nikole Pašića 1, 34000 Kragujevac, Serbia

e-mail: saramijailovic212@gmail.com

Abstract:

The use of artificial intelligence, especially machine learning methods in creating models that will be applied in clinical practice has reached its peak with the appearance of the COVID-19 pandemic. This study aims to determine the severity of the clinical condition of COVID-19 patients based on blood marker analysis. The study used data from 60 COVID-19 patients treated at the Clinical Center Kragujevac. The research methodology includes the selection of the most important laboratory parameters as well as the classification of patients depending on them using methods of supervised learning, regression and classification. With an accuracy of 90%, three parameters were selected that can mostly indicate the severity of the patient's condition, which are: lactate dehydrogenase (LDH), C-reactive protein (CRP), white blood cells (WBC). Laboratory biomarkers such as LDH, CRP and WBC may have an impact on predicting outcomes and help classify patients into an appropriate group based on symptoms.

Keywords: COVID-19, machine learning, biomarkers, modeling

1. Introduction

The use of artificial intelligence methods, but also the development of new methodologies, especially in the field of machine learning, has increased significantly since the beginning of the SARS-CoV-2 virus pandemic. The aim of such methods is mainly to predict the risk of worsening, as well as to predict the main outcomes of the disease [1]. A large number of studies have examined the predictor power of laboratory analyses measured in the admission of patients and during hospitalization. This knowledge can help in clinical practice and contribute to a better outcome of patient treatment. It has been shown that blood parameters such as D-dimer, creatinine, leukocyte count, lactate dehydrogenase, and C-reactive protein can predict the outcome, but the results of these studies are inconsistent with each other and require new approaches [2-4]. The aim of this study was to determine the severity of the clinical condition of COVID-19 patients based on blood marker analysis.

2. Methods

For the purpose of this study, a database was used that included COVID-19 patients treated at the Clinical Center in Kragujevac, Serbia. Admission to the hospital was considered as day one. Laboratory analyzes were collected at admission as well as on days 2, 5, 7, 9, 11 and 14 after admission. All patients were divided into one of 4 categories based on the severity of the clinical condition: mild, moderate, severe and critical. The following data were collected from the patients: demographic characteristics (sex, age, smoking status), symptoms at admission (presence of fever, cough, chest pain, headache, loss of smell, loss of taste). Laboratory analysis included measurement of complete blood count, measurement of gas analysis, inflammatory parameters, coagulation parameters, as well as renal and liver function parameters. After data collection, machine learning-based blood biomarker analysis was performed using methods of supervised learning. The supervised models use marked datasets that have to train the model so it can figure out which label to assign to new unlabelled data [1]. Also, methods of regression (Gradient boosting regressor) and classification (Extreme gradient boosting model (XGBoost[5])) were used. All data are divided into a test set and a training set. The analyzes collected on days 2, 5, 7, 9, 11 belong to the training set, while the last day belongs to the test set.

3. Results and discussion

A total of 60 patients with COVID-19 diagnoses were included in this study, of which 29 (48.3%) were female and 31 (51.7%) were male. The average age of the patients was 62.75 ± 10.61 . In this study, a total of 4 clusters were obtained, which is in accordance with the initial division of patients into groups: mild, moderate, severe and critical. Through a characteristics selection process we chose three of them with the greatest influence. The selected parameters and their average values are shown in Figure 1.

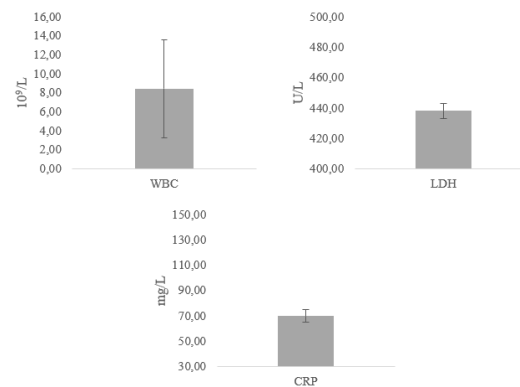


Fig. 1. Average values of selected most important laboratory biomarkers

The correlation of each of the two characteristics was used to assess the importance of the feature, after which the importance scores were calculated. The results indicate that lactate dehydrogenase (LDH) is a feature of the greatest importance (importance score is 685). It has been shown that the second most important is C-reactive protein (CRP) while the third most important parameter is white blood cells (WBC). Elevated lactate dehydrogenase level is one of the parameters that have been confirmed in a large number of studies to play a role in predicting patient outcomes which coincides with our results [6] Also, CRP has been shown to be an important blood parameter in previous research [7].

In order to predict the outcome after two weeks of hospitalization of COVID-19 patients, the classification model described in the Methods was used. This model achieved a prediction accuracy of 90%. Other measures of model precision were calculated and presented in Table 1.

State	Accuracy	Precision	Recall
Mild	0,90	0,89	1
Moderate	0,89	1	0,90
Severe	1	1	1
Critical	1	0,90	0,91

Table 1. Model precision parameters

4. Conclusions

Using the machine learning method, it has been shown that laboratory biomarkers such as LDH, CRP and WBC can have an impact in predicting outcomes, and special attention should be paid to the patient values of these parameters. A more detailed analysis would be needed to confirm the importance of these indicators or to single out new ones with greater prognostic power.

References

- [1] Negrini D, Danese E, Henry BM, Lippi G, Montagnana M. Artificial intelligence at the time of COVID-19: who does the lion's share?. *Clin Chem Lab Med*. 2022;10.1515/cclm-2022-0306. doi:10.1515/cclm-2022-0306
- [2] Xie J, Wu W, Li S, et al. Clinical characteristics and outcomes of critically ill patients with novel coronavirus infectious disease (COVID-19) in China: a retrospective multicenter study. *Intensive Care Med*. 2020;46(10):1863-1872. doi:10.1007/s00134-020-06211-2
- [3] Bayrak V, Şentürk Durukan N, Demirel Aydemir F, et al. Risk factors associated with mortality in intensive care COVID-19 patients: the importance of chest CT score and intubation timing as risk factors. *Turk J Med Sci*. 2021;51(4):1665-1674. Published 2021 Aug 30. doi:10.3906/sag-2101-89
- [4] Estenssoro E, Loudet CI, Ríos FG, et al. Clinical characteristics and outcomes of invasively ventilated patients with COVID-19 in Argentina (SATICOVID): a prospective, multicentre cohort study. *Lancet Respir Med*. 2021;9(9):989-998. doi:10.1016/S2213-2600(21)00229-0
- [5] Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining [Internet]*. New York, NY, USA: ACM; 2016. p. 785–94.
- [6] Zhang JJY, Lee KS, Ang LW, Leo YS, Young BE. Risk Factors for Severe Disease and Efficacy of Treatment in Patients Infected With COVID-19: A Systematic Review, Meta-Analysis, and Meta-Regression Analysis. *Clin Infect Dis*. 2020;71(16):2199-2206. doi:10.1093/cid/ciaa576
- [7] De la Cruz-Cano E, Jiménez-González CDC, Díaz-Gandarilla JA, et al. Comorbidities and laboratory parameters associated with SARS-CoV-2 infection severity in patients from the southeast of Mexico: a cross-sectional study. *F1000Res*. 2022;11:10. doi:10.12688/f1000research.74023.2