Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Computers in Biology and Medicine 138 (2021) 104869

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Artificial intelligence approach towards assessment of condition of COVID-19 patients - Identification of predictive biomarkers associated with severity of clinical condition and disease progression

Anđela Blagojević [a,b], Tijana Šušteršič [a,b], Ivan Lorencin [c], Sandi Baressi Šegota [c], Nikola Anđelić [c], Dragan Milovanović [d,e], Danijela Baskić [d], Dejan Baskić [e,f], Nataša Zdravković Petrović [d,e], Predrag Sazdanović [d,e], Zlatan Car [c], Nenad Filipović [a,b,*]

[a] University of Kragujevac, Faculty of Engineering, Sestre Janjić 6, 34000, Kragujevac, Serbia
[b] Bioengineering Research and Development Center (BioIRC), Prvoslava Stojanovića 6, 34000, Kragujevac, Serbia
[c] University of Rijeka, Faculty of Engineering, Vukovarska 58, 51000, Rijeka, Croatia
[d] Clinical Centre Kragujevac, Zmaj Jovina 30, 34000, Kragujevac, Serbia
[e] University of Kragujevac, Faculty of Medical Sciences, Svetozara Markovića 69, 34000, Kragujevac, Serbia
[f] Institute of Public Health Kragujevac, Nikole Pašića 1, 34000, Kragujevac, Serbia

ARTICLE INFO

Keywords:
COVID-19
Clinical condition assessment
Predictive blood biomarkers
Rule-based machine learning
Personalized model

ABSTRACT

Background and objectives: Although ML has been studied for different epidemiological and clinical issues as well as for survival prediction of COVID-19, there is a noticeable shortage of literature dealing with ML usage in prediction of disease severity changes through the course of the disease. In that way, predicting disease progression from mild towards moderate, severe and critical condition, would help not only to respond in a timely manner to prevent lethal results, but also to minimize the number of patients in hospitals where this is not necessary.

Methods: We present a methodology for the classification of patients into 4 distinct categories of the clinical condition of COVID-19 disease. Classification of patients is based on the values of blood biomarkers that were assessed by Gradient boosting regressor and which were selected as biomarkers that have the greatest influence in the classification of patients with COVID-19.

Results: The results show that among several tested algorithms, XGBoost classifier achieved best results with an average accuracy of 94% and an average F1-score of 94.3%. We have also extracted 10 best features from blood analysis that are strongly associated with patient condition and based on those features we can predict the severity of the clinical condition.

Conclusions: The main advantage of our system is that it is a decision tree-based algorithm which is easier to interpret, instead of the use of black box models, which are not appealing in medical practice.

1. Introduction

Since December 2019, global health issues have been caused by the outbreaks of the COVID-19 virus. It is still not known how SARS-CoV-2 triggers wide spectrum of heterogeneous clinical manifestation, from asymptomatic cases through acute respiratory distress syndromes to multiple organ failure and death [1–4]. Therefore, researchers all over the world have been looking to define and stratify predictors of the severity of COVID-19 disease in an attempt to properly guide medical management. Basic knowledge of pathogenesis of diseases and methods of discerning and assessing infection with COVID-19 have been established. Common blood hematology and clinical biochemistry tests are cheap, simple and widely accessible biomarkers. As such, they became the preferred method of tracking and predicting disease effects and forecasts it [5]. Recognizing the variation and phenotype of certain biomarkers as a result of multiple outcomes of COVID-19 will help

* Corresponding author. Faculty of Engineering, University of Kragujevac, Sestre Janjić 6, 34000 Kragujevac, Serbia.
E-mail addresses: andjela.blagojevic@kg.ac.rs (A. Blagojević), tijanas@kg.ac.rs (T. Šušteršič), ilorencin@riteh.hr (I. Lorencin), sbaressisegota@riteh.hr (S.B. Šegota), nandelic@riteh.hr (N. Anđelić), piki@medf.kg.ac.rs (D. Milovanović), d.baskic@gmail.com (D. Baskić), dejan.baskic@gmail.com (D. Baskić), silvester@sbb.com (N.Z. Petrović), predrag.sazdanovic@gmail.com (P. Sazdanović), car@riteh.hr (Z. Car), fica@kg.ac.rs (N. Filipović).

https://doi.org/10.1016/j.compbiomed.2021.104869
Received 22 June 2021; Received in revised form 10 September 2021; Accepted 12 September 2021
Available online 14 September 2021
0010-4825/

establish a risk-stratified strategy for the treatment of patients with this disease. Among other issues, it is very important to predict the unfavorable progression of the illness easily, reliably and timely. The possibility to detect cases that are at imminent risk of death has become an urgent, but necessary task [6].

Numerous researchers have studied predictors of disease severity in COVID-19 patients. Some studies indicated that serious or lethal cases of COVID-19 disease have been linked with elevated white blood cell count, creatinine, blood urea nitrogen, markers of liver and kidney function, interleukin-6 (IL-6), C-reactive protein (CRP), lower lymphocyte ($<1 \times 10^9$/L) and platelet counts ($<100 \times 10^9$/L) as well as albumin levels compared with milder cases of patients who survived the disease [7–9]. These research results provided an initial insight into of the effects of the infection by SARS-CoV-2, but due to geographical restrictions, specific experience of the clinical center and small cohorts, the result cannot be generalized [10]. Malik et al. found that particular biomarkers have been correlated with poor outcome results in COVID-19 hospitalized patients in meta-analysis of 32 trials reflecting 10,491 reported patients with COVID-19. The biomarkers included low lymphocytes, a lower count of platelets and elevated CRP, creatine kinase (CK), procalcitonin (PCT), D-dimer, lactate dehydrogenase (LDH), alanine aminotransferase (ALT), aspartate aminotransferase (AST) and creatinine [10]. Assandri et al. found that certain laboratory tests were characteristically altered in patients with of COVID-19 and proposed sensitive alternatives to recognize possible COVID-19 cases [1]. Specific biomarkers were found to be associated with the clinical outcome [11]. Some of the best prognostic markers were found to be neutrophil/lymphocyte ratio (in patients with a more serious illness) [12], CRP and inflammatory cytokines such as IL-6, IL-1 or tumor necrosis factor alpha (TNFα) [13]. Other studies also found that CRPO7 mg/dL will identify subjects that will develop critical condition [1,11]. After extensive research, it can be concluded that achieved results so far show inconsistency, biomarkers shown to be predictors of COVID-19 disease progression differ from research to research and sometimes are even contradictory [14]. Therefore, these is a strong need to investigate other strategies such as machine learning method to give a further insight into prognostic biomarkers.

1.1. Related work

The machine learning (ML) algorithms have been investigated in the field of COVID-19 for many purposes including epidemiological and clinical issues such as timely detection of disease outbreaks, fast diagnosis, classification and stratification of radiological images, risk factors analysis, as well as prediction of final clinical outcomes [15–20]. Most of the papers related to implementation of ML in COVID-19 investigation deal with analysis of medical images [21,22]. Another aspect analyzed is investigation of blood biomarkers as predictive features of clinical outcome. For example, Yan et al. [23] conducted research that used a blood sample database of 404 infected patients in the Wuhan region of China to classify important biomarkers of the disease seriousness in support of decision-making and logistical planners of health systems. To achieve this, three biomarkers have been selected using ML which had accuracy of more than 90%: LDH, lymphocyte and high sensitivity CRP (hs-CRP). This result is supported by current medical knowledge that elevated levels of LDH are related to the deterioration of tissues, which appear in pneumonia and other infectious and inflammatory diseases [24]. The main advantage of this paper is that it introduces a clear and operational formula to easily forecast and assess clinical outcome – death/survival, allowing the critical patients to be prioritized and potentially reduce the mortality rate. The main drawback of this study is that their classification is binary – survival/death, which may not be the best type of classification in situations where the healthcare systems are overloaded. Same group of researchers lead by Yan studied the samples of blood of 485 patients from the area of Wuhan, China retrospectively to recognize robust and relevant markers of risk mortality [6]. The goal was to identify the most discriminatory biomarkers of patient mortality by using state-of-the-art interpretable machine learning algorithms. Inputs contained the specific details, symptoms, the blood sample and laboratory results on the liver, renal functions, coagulation activity, electrolytes and inflammatory factors, collected from patients at different stages of disease, and related results for death or survival. The research employed a supervised XGBoost model classification. A tree-based machine learning model was used to predict the outcomes of individual patients (death/survival) using a sample blood test database. Same biomarkers (features) as in previous study were experimentally chosen with strong predictive degradation values or fatality of disease, also matching those in other literature [2,25–27]. As in the previous paper, the classification was binary and except predicting survival, does not help in reducing the burden put on healthcare system. Huang et al. discuss that although the studies by Yan et al. reveal that the mortality outcomes in COVID-19 patients are associated with three biomarkers using a single-tree XGBoost model, the forecast results are over-optimistic. The results of the forecast remain ambiguous (they indicate high variability). Hence, Yan et al. assertion on the successful prediction dates is insufficiently strong and not fully validated by the data provided in the papers. The prediction results may be unreliable.

Further, Gao et al. present a COVID-19 Mortality Risk Prediction (MRPMC) model that uses hospital admission data to streamline patients via risk of mortality, allowing prediction of physiological decline and death up to 20 days in advance. They implemented four methods of machine learning, namely Logistic Regression, Support Vector Machine, Gradient Boosted Decision Tree, and Neural Network. The MRPMC was tested on internal and external evaluation cohorts and achieved the average area under the curve (AUC) of 95% on internal and 97% and 92% on two different external cohorts, showing that it can have possible high utility in future clinical practice [15]. Research by Yao et al. [28] implemented the machine learning algorithms to build the COVID-19 disease detection model. Among several used algorithms (Support Vector Machine, Random Forest, K-nearest neighbors, AdaBoost etc.), SVM demonstrated most promising detection accuracy of 81.48% based on 32 features associated with the COVID-19 disease. Pulgar-Sánchez et al. [29] propose two methods - Multilayer Perceptron to predict severe-non severe cases of COVID-19 achieving 96.5% accuracy and C4.5 software, achieving 89.4% accuracy. Mahdavi et al. [30] proposed SVM for invasive and non-invasive groups achieving 0.92% accuracy for joint model. Cobre et al. [31] use biochemical tests and machine learning in order to predict positivity (positive negative) and disease severity (severe-non severe) with ANN as the best method achieving accuracy of both models more than 84%.

1.2. Motivation for the study

Although there are many papers that deal with implementing ML in proposing valid prognostic biomarkers and predictors of survival several days in advance, there is limited research in the field focused on classification of patient in more subtle severity-of-illness categories (e.g. mild, moderate, severe, critical) during the hospital stay. Such knowledge could help physician and hospital managers in decision-making process aiming to avoid not only patient's final unfavorable outcome, but also to improve other important secondary treatment endpoints and institutional performances which are deteriorated by inappropriate measures such as unnecessary prescription of adjunctive drugs (e.g. wide-spectrum antibiotics, immune-modulatory biologics), over-utilization of sophisticated and invasive diagnostics and inappropriate allocation of intensive care beds. Therefore, in this paper we propose a methodology based on ML to classify patients into several categories and predict the outcome in advance (change of severity of clinical condition). The main objectives and contributions of this paper are:

- examine easy interpretable algorithms which are suitable for implementation in clinical practices and compare with state-of-the art models
- classify patients into 4 distinct categories (mild, moderate, severe and critical) of COVID-19 disease using several machine learning methods (K-nearest neighbors (KNN), Support vector machine (SVM), Artificial neural network (ANN), Recurrent neural network (RNN) and Extreme gradient boost (XGBoost)) and compare the performances in order to select the final classification model.
- predict disease progression (mild to moderate, moderate to severe, severe to critical clinical condition) several machine learning methods (K-nearest neighbors (KNN), Support vector machine (SVM), Decision tree, Extra tree and Gradient Boost regressor (GBR)) and compare the performances in order to select the final regression model.
- work with a limited dataset, but implement different methods such as data imputation to overcome the drawbacks of small number of data

2. Materials and methods

This part of the paper explains the dataset first, and after that the proposed machine learning methodology. Since the methods can be divided into unsupervised (clustering) and supervised methods (regression and classification), and independent feature extraction, separate sections are created for each implemented sub-method.

2.1. Dataset

Blood biomarkers of the patients from two hospitals were used – Clinical Center of Kragujevac, Serbia (45 patients) and Clinical Center of Rijeka, Croatia (60 patients). In total, the results of blood analyses of 105 COVID-19 positive patients were collected (42% women, 58% men) and the age distribution of the patients in the form mean ± standard deviation was 52.77 ± 16.63. For all patients, fever was the most common symptom (83%), followed by cough (74.6%) and fatigue (45.7%). Described dataset is shown in Fig. 1.

We divide the clinical data into three subgroups:

- demographic data (gender and age)
- symptoms (fever, cough, fatigue, chest pain, muscle pain, headache, dyspnea, loss of taste or smell)
- blood analysis
  - Complete blood count (CBC)*: erythrocytes* (red blood cells (RBC)) – red cell indices: hemoglobin (HGB), mean corpuscular volume (MCV), mean corpuscular hemoglobin (MCH), mean corpuscular hemoglobin concentration (MCHC), red cell distribution width (RDW); *leucocytes (white blood cells (WBC))* – white cell differentials: neutrophils, lymphocytes, monocytes, eosinophils (EOS) and basophils (BASO), *platelet indices*: platelets (PLT) platelet distribution width (PDW), mean platelet volume (MPV)
  - Coagulation: prothrombin time (PT), international normalized ratio (INR), D-dimer
  - Kidney function: urea, creatinine (CREA)
  - Hepatic function: bilirubin – direct and total, alanine transaminase (ALT), aspartate transaminase (AST), gamma-glutamyltransferase (also γ-glutamyltransferase, GGT), albumin
  - Enzymes: creatine kinase (CK), also known as creatine phosphokinase (CPK) or phosphocreatine kinase; lactate dehydrogenase (LDH)

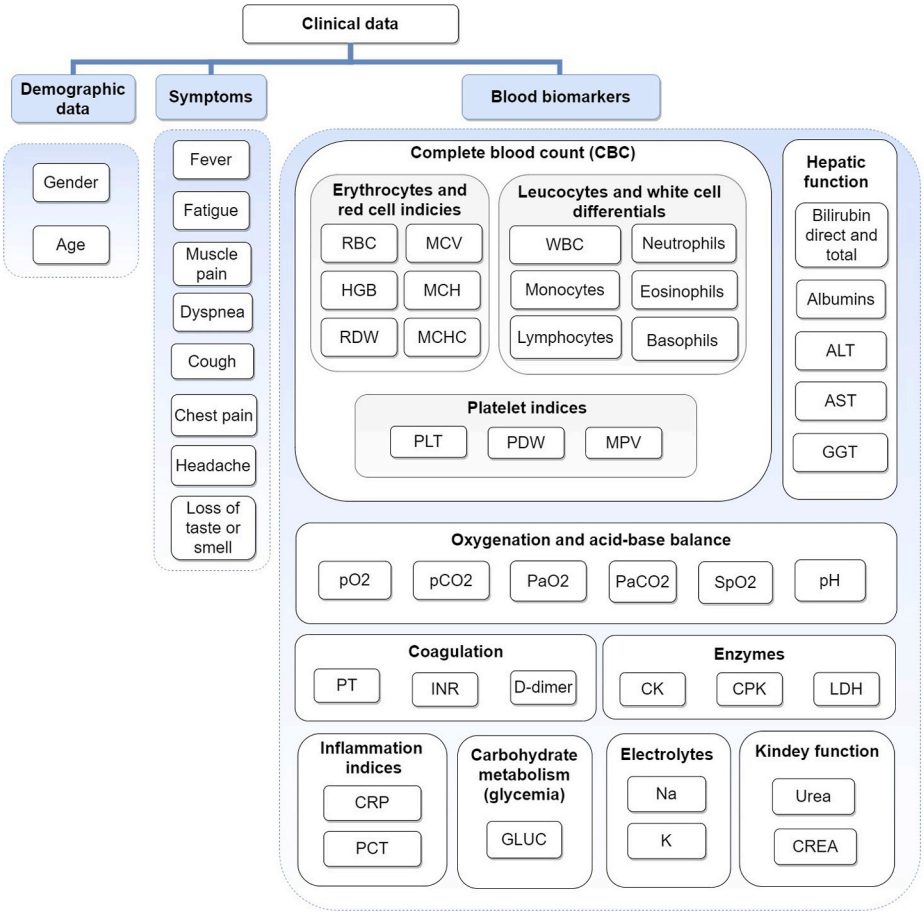

Fig. 1. Graphical representation of clinical data which consist of demographic data, symptoms and blood analysis.

- Electrolytes: Sodium (Na), potassium (K)
- Oxygenation and acid-base balance: arterial blood gas (ABG) analyses/tests: partial pressure of oxygen (pO2), arterial partial pressure of oxygen (PaO2) partial pressure of carbon dioxide (pCO2), arterial partial pressure of carbon dioxide (PaCO2), SpO2 (peripheral oxygen saturation s. oxygen saturation as measured by pulse oximetry), pH:
- Inflammation indices: C-reactive protein (CRP), procalcitonin (PCT)
- Carbohydrate metabolism (glycemia): GLUC – glucose

Most patients had multiple blood samples taken during their hospital stay. We have considered the day of admission to the hospital and days 2, 5, 7, 9, 11 and 14 after hospital admission. In order to deal with missing data, we used data imputation method. To fill in the missing values, we used the mean value of four different values (values from 2 days before and 2 days after). The aim of this type of data imputation is to include the values of analyzes of adjacent days that are not directly included in the prediction.

2.2. ML model development

Proposed machine learning methodology consists of two main tasks:

1. Prediction of the blood analysis in advance (in days)
2. Classification of the severity of clinical condition (mild, moderate, severe and critical).

Proposed methodology is shown in Fig. 2. Proposed methodology has been implemented in Python 3.7.4, Spyder environment 3.3.6, using CPU Intel Core i5. Given the collected dataset that consists of patient data (demographic data, physical symptoms and blood biomarkers) and patient condition, we undertake several preprocessing and main analysis steps. Since there were a lot of instances without labeled output - severity of clinical condition and omitting these instances were significantly reduced the size of the dataset, we have performed clustering (using K-means method). One group clustered around already known output label was assumed to associated with a same output label meaning the whole cluster was labeled as the same clinical condition. Further, feature selection (using ANOVA F-test) in preprocessing stage was performed to extract the most important blood biomarkers as predictors of the severity of clinical condition. In the main section of proposed method, classification (using several algorithms such as KNN, SVM, ANN, RNN and XGBoost) was used to determine the output class (mild, moderate, severe and critical), which corresponds to severity of clinical condition. In addition, regression (using SVM, Decision tree, KNN, Extra tree and GBR) was used to predict changes of biomarkers in time (up to 14th day), which would in return affect the change of the output class, as patient condition was assessed based on blood biomarkers.

2.2.1. Feature selection

To evaluate which biomarkers are crucial for determination of the severity of clinical condition, we have used Scikit learn function for feature selection that selects features according to the K highest scores and we have chosen the ten features that had the greatest influence in making the effective model. Proposed function for feature selection is based on principles of Analysis of variance (ANOVA) which can determine whether the means of three or more groups are different [32]. The ANOVA method assesses the relative size between group variance compared to the average variance within group variance. ANOVA uses F-test which determines the degree of how relatively greater the difference is between group variance compared to within group variance. Therefore, to get a statistical conclusion we may compare the F-test value calculated from the observed data with the P-values associated with the F-test. The features with the F-values below 0.05 (P-value) were usually considered to be statistically significantly associated with the outcome.

2.2.2. Unsupervised methods

In this section, we proposed the use of clustering methods to estimate the missing output labels. This section is particularly important, as without enough output labels, no ML method will be able to perform classification into different categories. In our case, should the instances without labels were omitted, the dataset size would significantly reduce. Therefore, omitting the instances was not possible and clustering was used as a method to estimate the data labels. We have applied several different clustering methods (k-means, agglomerative, mean shift) [33], all which gave similar results; therefore, we have adopted k-means algorithm as the final choice. K-means algorithm requires predefined number of clusters and for determining optimal number of clusters Elbow method was used. The method will be examined starting from two clusters and increasing it in each step by one until the optimal number of clusters is determined. For each given number of clusters, the sum of intra-cluster distances is computed. The variance of the intra-cluster distances between two consecutive numbers of clusters is computed. Then, the Elbow method looks into the percentage of variance explained as a function of the number of clusters. One should choose a number of clusters so that adding another cluster in the analysis does not give much better explanation of the variance.

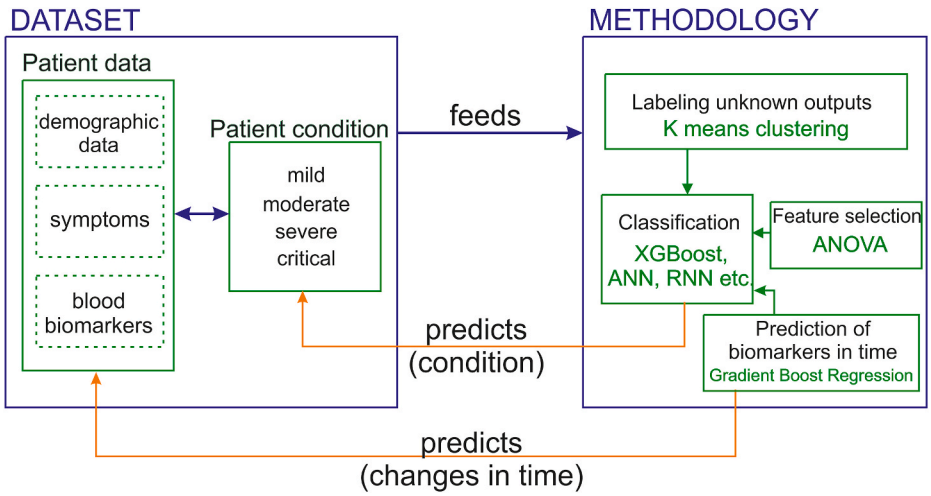

Fig. 2. Schematic representation of ML methodology for determination of the severity of clinical condition.

In order to better describe the relationship between available features and clusters, dependencies between clusters and several most descriptive features are presented in Fig. 3. The clusters are represented by color (purple, blue, green, yellow), while the manually annotated class is located next to each instance in the numeric form (class 0 refers to mild clinical condition, class 1 refers to moderate, class 2 refers to severe and class 3 refers to critical clinical condition).

If we take into consideration Fig. 3(a), clusters have been presented from the aspect of LDH and CRP. We can observe that green cluster is associated with higher LDH values, while yellow cluster is associated with lower LDH values and drastically higher CRP values. In Fig. 3(b), clusters have been presented from the aspect of Urea and Creatinine. In this case, it can be concluded that only yellow cluster has high values of Urea and Creatinine, which means that high values of these biomarkers occur only in blood analyses of patients with critical clinical condition. In Fig. 3(c), clusters have been presented from the aspect of Albumins and CRP. It can be concluded that only purple and blue clusters have higher values of Albumins, but smaller values of CRP which is in accordance of mild and moderate clinical conditions, manually annotated. Enormously large values of CRP have been related to the critical clinical condition i.e. cluster yellow. In Fig. 4(d), clusters have been presented from the aspect of WBC and Albumins. In this Figure, it can be concluded that only yellow cluster has high values of WBC and smaller values of Albumins simultaneously. Regarding the previous cluster-feature analysis, each class instance manually marked by the doctor has been labeled next to its cluster leading to conclusion that we can now translate the cluster to the appropriate class. As a result, the dataset was divided in four clusters now assigned to severity of clinical condition and the distribution is the following:

- 31.80% of total data belongs to class of mild clinical condition,
- 50.90% to moderate,
- 13.88% to severe,
- 3.42% belongs to class of critical clinical condition.

2.2.3. Supervised methods

In this section, we implement regression analysis to predict the values of important features (biomarkers) in the following days, in order to be able to further perform the classification of the patients. Such coupled methodology will enable us to estimate the patient condition development in time.

2.2.3.1. Regression.

Proposed methodology for prediction condition of the patient in time is to track and predict the change in biomarkers values. Based on changes in blood biomarkers during two weeks' time, we aim to predict the patient's clinical condition. The main limitation was the lack of data for biomarkers in time (full blood analysis on the admission day was available for all 105 patients, blood analysis from 104 patients were available for the second day, on 5th day data for 67 patients was available, on 7th day data for 68 patients was available, on 9th day data for 59 patients was available, on 11th day data for 51 patients was available and on 14th day only data for 44 patients was available). It should be emphasized that not always were the data for one patient available at each and every time point, but rather spread across some days (i.e. for patient 1, blood analysis was available on days 0, 5, 11, while for patient 2 for days 0, 7,9,14 etc.). Due to the small number of patients' data available in time, in order to predict blood biomarkers values, we decided to select 34 patients with a full blood analysis for all days.

We wanted to achieve time dependencies by creating additional feature which represents the subtraction between blood analysis for the day *t* and blood analysis for the previous observed day *t-1*. Among several approaches to reorganize the dataset in order to create time dependencies between values of biomarkers, auto-regressive model proved to be the most efficient way for creating dataset that would be appropriate for the regression task. Auto-regressive model by its definition is used for forecasting when there is some correlation between values in a time series and the adjacent ones (the values that precede and succeed them). In that manner, we have created time-dependencies between data because the biomarker values for day *t* certainly depends on the values of the biomarkers for day *t-1*. In the regression model itself we have used this additional feature alongside with blood analysis for the day *t-1*, in order to predict the values of blood biomarkers for the day *t*. This means that for day 5, day 2 as *t-1* and day 5 as *t* were taken as part of analysis. In this way, we have created time dependencies between a patient's blood analysis throughout time and expand the dataset for the regression problem. Fig. 4 shows a schematic representation of the described methodology.

After establishing the database, we divided data into training and test set - training set consists of data from days 2,5,7,9 and 11, meaning that the day 14 belongs to the test set.

For the prediction of each blood biomarker, several machine learning methods (KNN, SVM, Decision tree, Extra tree and Gradient Boost

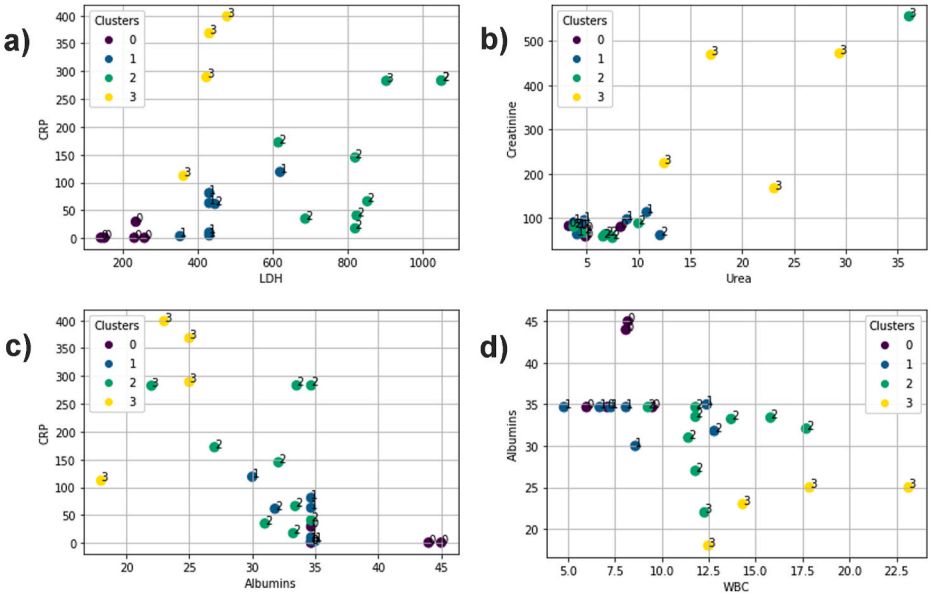

Fig. 3. Dependence between clusters and values of different blood biomarkers.

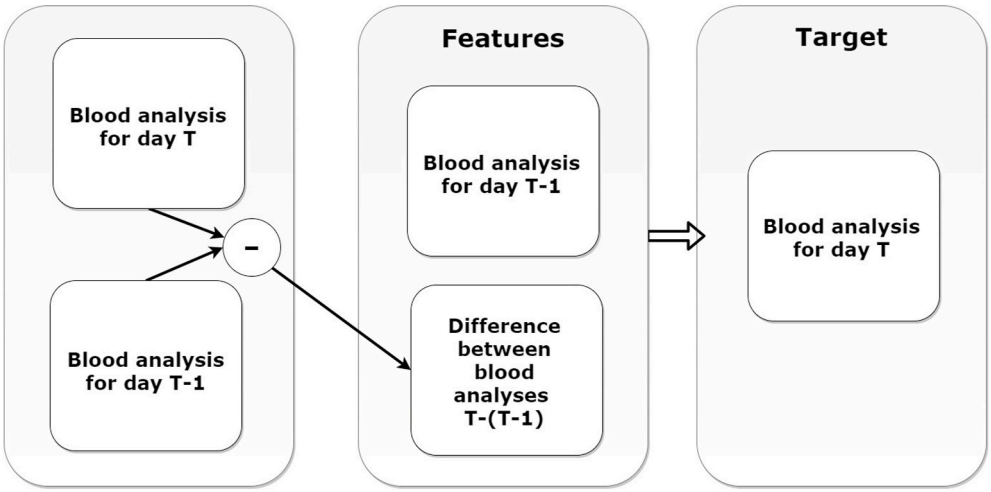

Fig. 4. Schematic representation of principles which was used for database organization.

regressor) were used and their performances have been compared in order to select the final regression model.

2.2.3.2. Classification. The main task of this paper is to assess how the COVID-19 develops over time in patients, meaning to determine the patient's condition 14 days after hospital admission. This was achieved by firstly assessing the values of biomarkers using the methods explained in the previous section, after which the patients were classified into one of 4 classes (mild, moderate, severe and critical). For classification purposes, we have included all patients for whom we had blood analysis on a particular day. This means that within this dataset, the same patients are repeated multiple times without any dependences between different days. This modification is justifiable as in this case, it is not important to observe the same patient in time, but to create as many different instances as possible, in order to expand the dataset and prepare it for the classification task. As a result, the dataset was expanded to 497 instances (158 instances belonged to the cluster of mild condition, 253 to moderate, 69 to severe and 17 instances belonged to the cluster of critical condition). For the described classification task, the aim was to create two types of models; the first considered type is based on black-box model such as ANN and RNN, which proved to be efficient in complex classification tasks. On the other hand, another type of model which tend to be more interpretable and simplified was considered. Therefore, from the aspect of simplified methods, we adopted SVM, KNN and XGBoost. Performances of all mentioned methods have been compared in order to select final classification method. For example, XGBoost is a supervised machine learning algorithm based on the recursive construction of a decision tree, and those trees that most influence the decision of the predictive model can be identified. At each decision step in the XGBoost trees, the significance of each feature is determined by its accumulated use [6,34]. This is an advantage of rule-based algorithms because internal model strategies are easy to interpret. Also, the importance of these algorithms in clinical prediction is reflected in the fact that the feature values for decision-making are known, unlike black box models whose rules and strategies are difficult to interpret.

The model performance was evaluated by assessing different classification metrics: accuracy, precision, sensitivity, specificity and F1 score. Equations of these metrics are defined below:

$$Accuracy = \frac{True\ positive + True\ negative}{True\ positive + True\ negative + False\ positive + False\ negative} \quad (1)$$

$$Precision = \frac{True\ positive}{True\ positive + False\ positive} \quad (2)$$

$$Sensitivity = \frac{True\ positive}{True\ positive + False\ negative} \quad (3)$$

$$Specificity = \frac{True\ negative}{True\ negative + False\ positive} \quad (4)$$

$$F1\ score = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \quad (5)$$

3. Results

Ten blood biomarkers were selected as the best features that had the greatest contribution to the classification task and most reliably described the development of COVID-19 disease in patients, according to the previously described methodology. These blood biomarkers and their values are presented in Table 1. In addition to values from our database presented in the form of mean ± standard deviation, biomarkers values from literature are given, in the form of median and mean ± standard deviation.

For the evaluation of the importance of features, Scikit learn function K highest scores was used. The importance of all ten features is shown in Fig. 5. Lactate dehydrogenase (LDH) has the highest importance score, which can be due to the fact that this enzyme is widely distributed in

Table 1
Ten blood analysis that had the greatest influence in COVID-19 patient's clinical condition assessment.

Blood biomarker	Median	Mean ± standard deviation	
	Mahdavi M. et al. [30]	Yao H. et al. [28]	Our
WBC count	7.8	0.1133 ± 0.3637	8.25 ± 4.72 10^9/L
Lymphocytes	1.22	9.0184 ± 8.5835	22.15 ± 11.06%
MCHC	34	340.1333 ± 14.6198	335.7 ± 8.16 g/L
RDW	14	13.2747 ± 1.9499	13.69 ± 1.07%
Hgb	/	126.1133 ± 21.8770	130.08 ± 21.05 g/L
Urea	/	/	7.35 ± 5.76 mmol/L
Creatinine	/	/	92.46 ± 69.62 μmol/L
Albumins	/	65.7480 ± 112.5770	33.57 ± 6.23 g/L
LDH	565	/	424.86 ± 239.26 U/L
CRP	/	79.5920 ± 93.2775	69.11 ± 93.55 mg/L

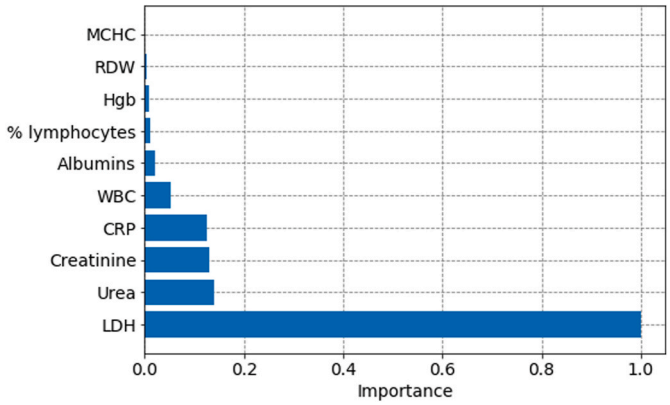

Fig. 5. Importance scores of ten best features.

Table 2
Root mean squared error between predicted and actual values of blood biomarkers.

Metric	RMSE				
Blood biomarker	SVM	Decision tree	KNN	Extra tree	Gradient Boost
White blood cell (WBC)	6.79	3.06	4.61	3.12	3.03
% Lymphocytes	8.97	2.83	4.51	2.25	1.57
MCHC	9.65	3.57	5.52	3.55	2.77
RDW	1.07	0.45	1.12	0.36	0.35
Hgb	16.87	12.35	8.9	8.50	10.12
Urea	9.79	2.61	5.8	2.87	2.54
Creatinine	140.36	97.84	103.65	47.73	55.23
Albumins	7.84	1.96	5.48	1.56	1.63
LDH	165.6	49.36	40.91	31.71	35.51
CRP	116.13	46.61	34.09	42.47	41.68

tissues and its elevated serum levels could be caused by systemic hypoxemia [35]. It was found that high serum activity of LDH in earlier stages of the disease is a good predictor of a lung injury and poor risk outcomes [6]. Beside LDH, which was shown to be by far the most important biomarker in predicting severity of clinical condition, other parameters which mostly influence the model are urea (2nd most important), creatinine (3rd most important), C-reactive protein (CRP) (4th most important), white blood cell (WBC) (5th most important), etc. (Fig. 5). The explanation for importance of biomarkers urea and creatinine could be found in the fact that both biomarkers are related to kidney function - one research that included the analysis of 701 COVID-19 patients' conditions, the conclusions were that incidence of acute kidney injury and death was significantly higher in patients which had elevated baseline serum creatinine levels in comparison to the patients with normal baseline values [36]. Further, CRP is used as a biomarker for different inflammatory and infectious conditions for clinical purposes. Elevated CRP is directly correlated with the level of inflammation and therefore could be correlated with severity of clinical condition. Generally, several studies have found that serious or lethal effects of COVID-19 disease are linked to liver and kidney function markers, C reactive protein (CRP), interleukin-6 (IL-6), lower lymphocytes and albumin blood levels relative to milder form of the disease in the survivors [10].

We can determine the importantance of certain biomarkers, if we exclude some of them with the greatest importance from the classification. For example, if we exclude biomarker of the greatest importance – LDH, classification metrics would be significantly affected. The values of statistical measures were sensitivity = 0.34, specificity = 0.57, accuracy = 0.5, precision = 0.3, F1-score = 0.31. This feature selection method reduces the amount of blood tests that need to be assessed. Values of all ten selected blood biomarkers are assessed for 34 patients on day 14 after the admission day. Due to the small number of patients' data available in time, we decided to select these 34 patients with a full blood analysis for all days, according to the described methodology in the previous section. Table 2 shows the root mean square error between predicted and actual values of blood analysis, using several algorithms – SVM, Decision Tree, KNN, Extra Tree and Gradient Boost.

As can be seen in the table, the best results (the lowest RMSE) can be observed with the Gradient boost and Extra tree methods. Although Gradient boost gives smaller RMSE for multiple biomarkers, we have examined both methods in order to see their influence on classification. It should be emphasized that the classification is performed on the estimated values of biomarkers. For this reason, it is very important to notice how the regression model affects classification process. If we take into account the classification metrics, after predicting the value by the Extra tree method, the classification algorithm reaches an accuracy of 0.91 and an F1-score of 0.91. It can be concluded that the change of the regression method downgrades the performances of the classification method, so we will adopt Gradient Boost regressor as the final method for the regression task (accuracy of 0.94 and F1-score of 0.94). This model was trained to obtain optimal hyperparameters settings using grid search method and the final setting included number of boosting stages set to 500, max depth equals to 4, min samples split equals to 5 and learning rate to 0.05.

It can be seen that values of biomarkers such as white blood cell (WBC), % lymphocytes, MCHC, RDW, urea and albumins fall under the narrower range (also seen from Table 1), so we can expect smaller RMSE. On the other hand, Hgb, creatinine, LDH and CRP have more deviation between real and predicted values. Results for each of these four biomarkers will be discussed individually.

In Fig. 6, comparison between actual and predicted values of Hgb is shown. RMSE is 10.12, which may be due to the fact that this biomarker had a wider range in patients in our study (from around 80 to around 150). Although RMSE for Hgb is larger in comparison to other biomarkers' RMSE, greater deviation from the actual value is expected, as the range of actual values (80–150) is wider compared to the ranges of other biomarkers (i.e. range of actual values of RDW is 11–18). For this reason, the value of RMSE does not affect the model's decision in assessment of clinical condition of COVID-19 patients.

In Fig. 7 comparison between actual and predicted values of CRP is given. RMSE for this analysis is 41.68, explanation of this RMSE values is similar to the RMSE value of Hgb, where CRP has even wider range of values (0–400) than Hgb.

Fig. 8 shows comparison between actual and predicted values of creatinine. This blood biomarker has the highest value of RMSE, and the cause may be the existence of some values that are drastically higher compared to the rest of the analyzes, meaning some outliers could be

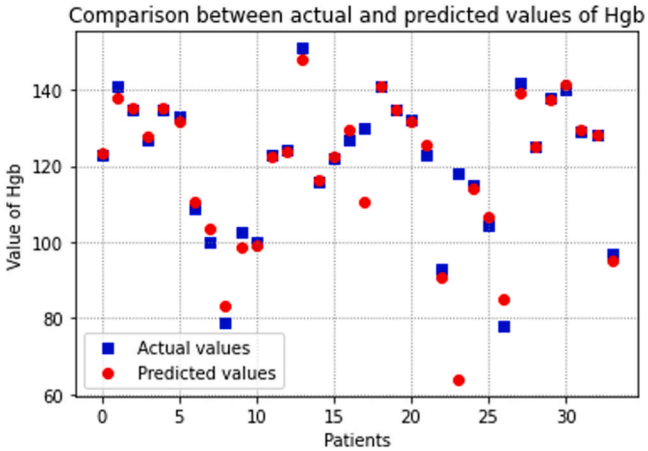

Fig. 6. Comparison between actual and predicted values of Hgb.

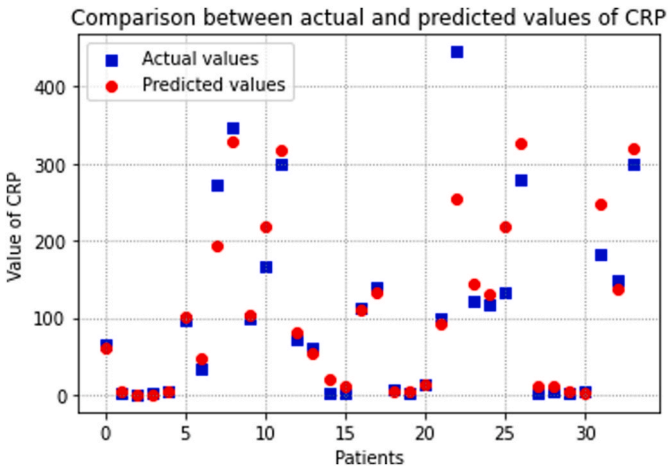

Fig. 7. Comparison between actual and predicted values of CRP.

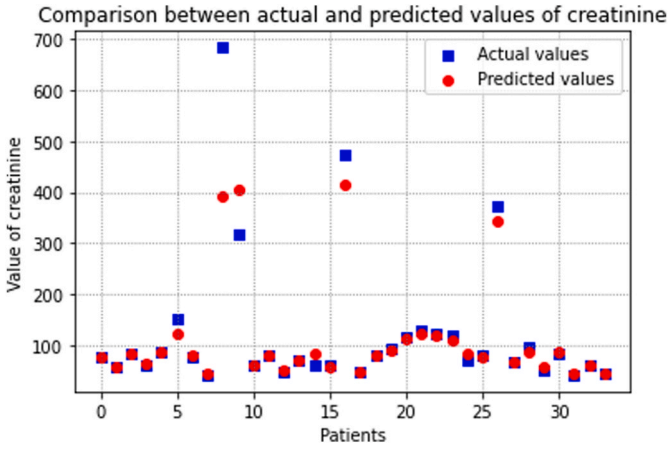

Fig. 8. Comparison between actual and predicted values of creatinine.

present in the dataset.

The same explanation could be given for greater RMSE for LDH analysis where comparison between predicted and actual values is shown in Fig. 9.

After evaluation of the results of patient's hematology and clinical biochemistry analyses, it is possible to predict the patient's clinical condition in advance. For the purposes of the classification task, several

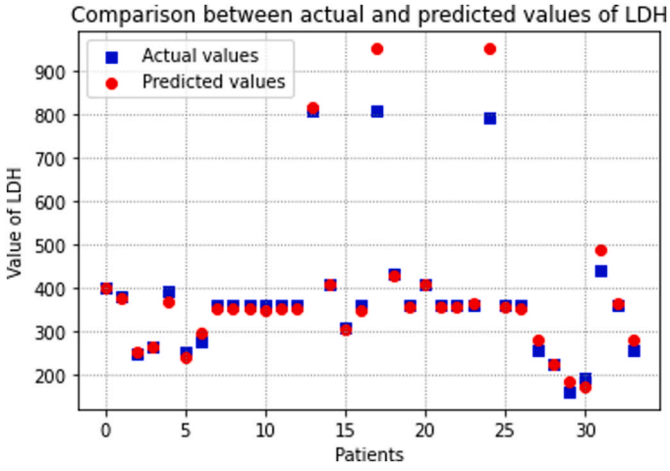

Fig. 9. Comparison between actual and predicted values of LDH.

classification methods have been considered. In order to obtain the most accurate and precise classification model, five algorithms have been trained and their performances were evaluated on test set. Table 3 shows the metrics for each considered method.

In order to get better insight in validation of performances of the proposed models, receiver operating characteristic (ROC) curve with area under the curve (AUC) and precision-recall (PR) with area under precision-recall (AUPR) were generated. Fig. 10(a) suggested that the classification performance of the XGBoost model is comparable with the SVM model. Although the AUC for XGBoost and SVM models is the same, AUPR (Fig. 10(b)) revealed differences between these two models (AUPR for XGBoost is 0.98 and AUPR for the SVM is 92). It can be concluded that XGBoost model has better performance in case of class imbalance.

In order to validate the effectiveness of XGBoost model, we have performed 10-fold cross-validation on the training set and the results are presented in Table 4.

XGBoost was trained to obtain optimal hyperparameters using grid search method and the final setting included number of tree estimators set to 100, max depth equals to 5, learning rate equal to 0.3, gamma equals to 1, 'subsample' and 'colsample bytree' both set to 1.Model was tested on 34 patients and achieved an accuracy of 94% in predicting the patient's condition on day 14 after the hospital admission. We computed confusion matrix for this test set which is shown in Fig. 11 both for normalized and regular number of patients.

Since the dataset is unbalanced, we considered other metrics such as precision, sensitivity, specificity and F1-score in addition to accuracy. In Table 5 all of these metrics for each class individually are shown. Despite the fact that dataset was rather are unbalanced, in terms of both tasks - regression and classification, we have achieved excellent results. In case that our model will not be able to recognize a class that is in the minority, we would consider one of the methods such as SMOTE and random undersampling.

Our validated XGboost model is a decision tree-based model, consisting of hundreds of trees whose rules are known and stand as a reason behind the final decision. Data scientists utilize XGBoost to tackle a variety of machine learning difficulties. Most deep learning algorithms are difficult to explain to medical professionals, both in terms of the process and the results they produce. In our study, interpretability is considered as a fundamental requirement for selecting machine learning model. Fig. 12 shows one such tree that is part of the XGBoost classifier, whose rules are understandable and based on if-then-else. This type of model is suitable for application in clinical practice due to its comprehensibility. In this example of a tree, it can be seen that decisions about determination of a class of patient, indicating severity of clinical decision, are based on a single tree. Although one tree is not enough for final decision, when several trees (i.e. hundreds) are investigated and analyzed, high accuracy will be achieved for the problem of classification. It should be emphasized that this analysis is automated and from a clinical point of view and final decision is met within seconds.

4. Discussion

Although there may seem that there are many studies on COVID-2019 disease prediction based on machine learning, most of these studies perform binary classification (positive-negative for COVID-19 or

Table 3
Overview for the classification metrics for each algorithm.

Algorithm	Accuracy	Specificity	Sensitivity	Precision	F1-score
XGBoost	0.94	0.98	0.97	0.95	0.96
KNN	0.88	0.94	0.89	0.92	0.90
SVM	0.82	0.93	0.88	0.80	0.83
ANN	0.76	0.91	0.85	0.73	0.77
RNN	0.56	0.77	0.37	0.51	0.37

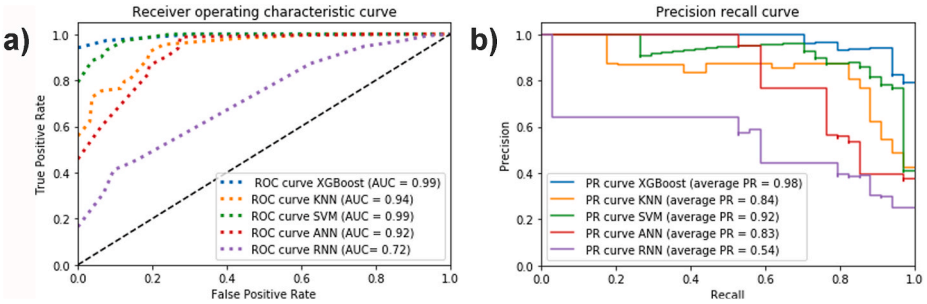

Fig. 10. a) Receiver operating characteristic curve (ROC) and b) Area under the precision-recall curve (AUPR).

Table 4
Classification metrics for each fold of training set.

10-fold cross-validation	Accuracy	Specificity	Sensitivity	Precision	F1-score
1st fold	0.98	0.99	0.99	0.99	0.99
2nd fold	1.00	1.00	1.00	1.00	1.00
3rd fold	0.94	0.97	0.91	0.96	0.93
4th fold	0.98	0.99	0.88	0.99	0.91
5th fold	0.72	0.74	0.73	0.70	0.71
6th fold	0.96	0.98	0.87	0.98	0.90
7th fold	0.94	0.97	0.96	0.97	0.96
8th fold	0.70	0.72	0.72	0.68	0.70
9th fold	0.75	0.75	0.75	0.75	0.75
10th fold	0.94	0.98	0.95	0.89	0.91
Overall	0.89	0.91	0.88	0.89	0.88

severe-not severe disease). Yao et al. (2020) performed the study that mentions the detection of disease severeness, however, in reality, the study performs binary classification of severely ill (positive) samples against the patients with mild symptoms [28]. Secondly, the number of patients with COVID-19 used in this paper is relatively small, which may limit the accuracy of detection model. Banerjee (2020) used a public database with 598 patients, only 39 of whom were positive for SARS-CoV-2, to create an ML system to predict COVID-19 diagnosis. The authors produced a model with high specificity (91%) but low sensitivity (43%), which may preclude the model from being used in practice for early illness detection [37]. Joshi et al. (2020) applied a logistic regression model previously trained with 390 samples, out of which only 33 were positive for COVID-19, achieving sensitivity and specificity values of 93% and 43%, respectively [38]. Furthermore, precise biomarkers linked with illness severity and patients' hospitalization in critical care units are yet unclear, which may impede the development of future tailored therapies [39].

All the previous studies, even those that mention severity determination, use binary classification (positive-negative samples, sever-not severe). The main advancement in our proposed methodology is that it is the only that uses multiclass classification – mild, moderate, severe, critical, as well as disease prerecession tracking in terms of change of the class (worsening or bettering of the condition). However, in order to compare the results of our study with other binary classification studies, Table 6 shows achieved metrics as reported in literature with our achieved results. We have used only the papers with high sensitivity/specificity and accuracy to compare with our proposed method.

Cobre et al. [31] in its prediction use besides the complete blood count test, data from biochemical, urinary, bacteriological, and virological tests aiming at identifying additional predictive biomarkers. However, in our study, we use only standard blood biomarkers as input, which adds to the daily praxis, especially in the times of pandemic, where there is no time for additional test and only standard tests can be used for high accuracy predication.

Added value of our study is also connected to the interpretability, as the best results in our study are achieved by XGBoost, which is an algorithm based on decision trees. Other prediction models such as ANNs do not provide clinically useful interpretable rules that could explain the reasoning process behind their predictions. They just produce the accuracy score, precision score, and recall score, which represent the likelihood that a patient would get ill. This is especially true for data sets that are unbalanced or biased [40]. Beside achieving the best results, a good prediction model should clearly show the decision processes to medical workers. Therefore, we adopt XGBoost as a final model to make a trade-off between predictive power and interpretability.

Table 5
Overview of classification metrics for each class on test data.

Class	Sensitivity	Specificity	Accuracy	Precision	F1-score
Mild	1.00	0.92	1.00	0.80	0.89
Moderate	0.89	1.00	0.89	1.00	0.94
Severe	1.00	1.00	1.00	1.00	1.00
Critical	1.00	1.00	1.00	1.00	1.00
Overall	0.97	0.98	0.94	0.95	0.96

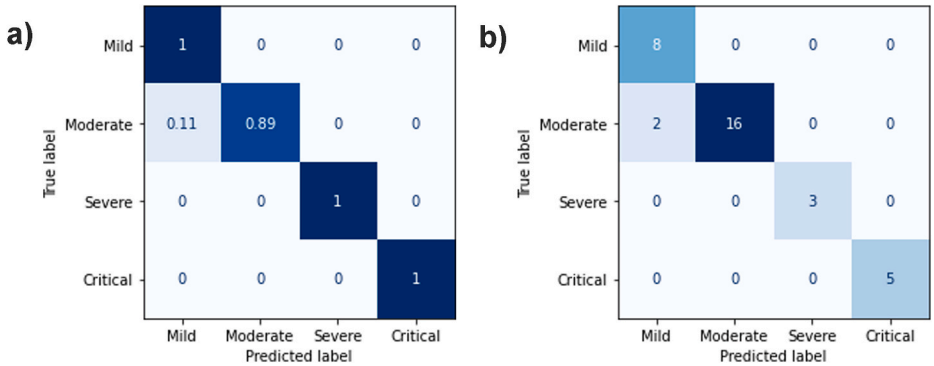

Fig. 11. Confusion matrix with normalized (left) and regular (right) values of patients.

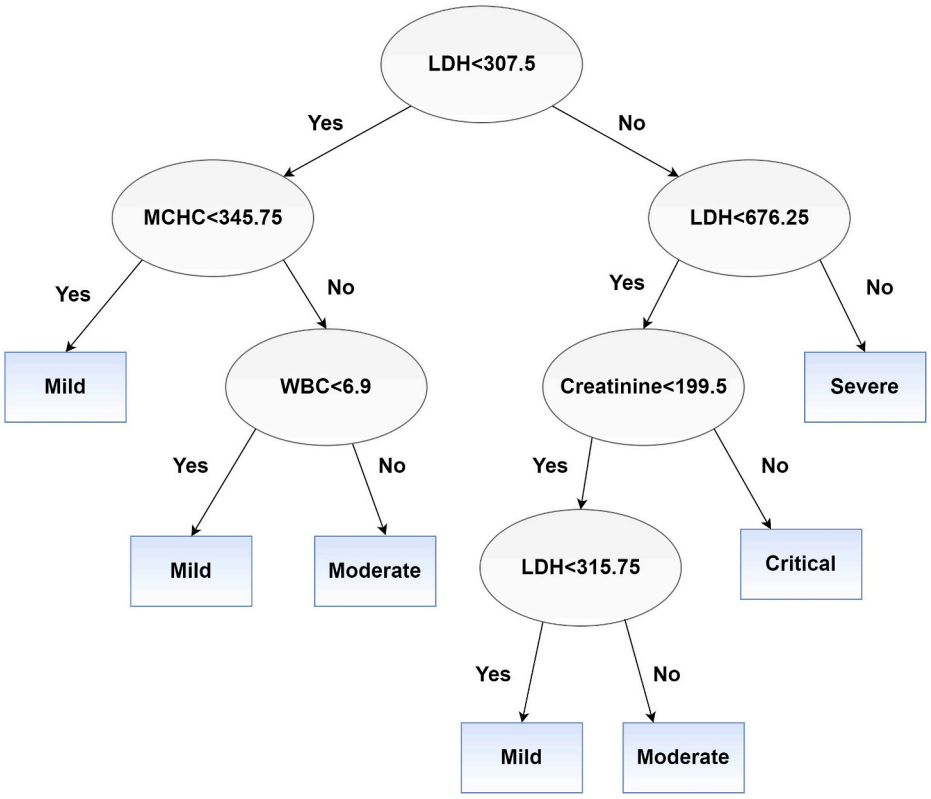

Fig. 12. An example of tree which is a part of XGBoost classification model.

Table 6
Comparison with existing literature.

Metrics	Yao et al. [28]	Pulgar-Sánchez et al. [29]	Mahdavi et al. [30]	Cobre et al. [31]		Proposed method
Best algorithm	SVM	MLP	SVM	Diagnostic ANN	Severity ANN	XGBoost
Accuracy	0.8148	96.5	0.92	0.94	0.98	0.94
Specificity	1.0	0.964	0.91	0.94	0.97	0.98
Sensitivity	0.833	0.965	0.81	0.93	0.99	0.97

5. Conclusion

This paper deals with automatic ML methods for classification of patients with COVID-19 into several categories, namely severity of clinical condition. Our research further analyses the possibility for prediction of the change of category in advance, which means that the model predicts the patient's clinical course during hospital stay until discharge or death. This research represents a proof of concept that a ML model is an efficient and informative method to gain insight into the COVID-19 disease process. From extensively implemented algorithms and hyper-parameter optimization, the following conclusions can be drawn:

- Ten most important variables (features), strongly associated with patient conditions, are extracted from blood biomarkers using ANOVA F-test feature selection method.
- Gradient boost regressor proved to be the best in predicting blood biomarkers, achieving root mean square error for WBC, lymphocytes, MCHC, RDW, urea, albumins less than 3, and some larger RMSE for Hgb, creatinine, LDH and CRP, which can be explained with wider ranges of those biomarkers.
- proposed methodology using XGBoost classifier is the most adequate for classification of patients into 4 distinct categories of clinical conditions (mild, moderate, severe, critical) with 94% of accuracy.
- since the proposed methodology is rule-based (XGBoost can be described using IF-then rules) rather than black box, it is more suitable for implementation in real clinical practice.
- coupled unsupervised and supervised algorithms are able to predict disease progression (mild to moderate, moderate to severe, severe to critical clinical condition) in advance.

The main limitation of our study the fact that the dataset included 105 patients. However, we have implemented different methods to overcome the drawbacks of small datasets and draw solid conclusions. Further research would be focused on collecting larger database of patients as well as investigation other ML models that can be coupled with existing in order to create hybrid model which would achieve even larger accuracy. We would also like to include patients from other countries as then, geographical position of the country could be taken into account as additional feature. In such a way, we could investigate how geographical area influences the disease development in patients and possible reaction to COVID-19 as a function of geographical location.

Ethics

The work has been approved by the appropriate ethical committees related to the institution(s) in which it was performed and that subjects

gave informed consent to the work.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This research is funded by Serbian Ministry of Education, Science, and Technological Development [451-03-68/2020-14/200107 (Faculty of Engineering, University of Kragujevac)]. This research was also supported by the CEI project "Use of Regressive Artificial Intelligence (AI) and Machine Learning (ML) Methods in Modelling of COVID-19 spread – COVIDAi" [Grant Nr. 305.6019-20].

References

[1] R. Assandri, E. Buscarini, C. Canetta, A. Scartabellati, G. Viganò, A. Montanelli, Laboratory Biomarkers predicting COVID-19 severity in the Emergency room, Arch. Med. Res. 51 (6) (2020) 598–599.
[2] C. Huang, et al., Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China, Lancet 395 (10223) (2020) 497–506.
[3] D. Wang, et al., Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus–infected pneumonia in Wuhan, China, Jama 323 (11) (2020) 1061–1069.
[4] W.-j. Guan, et al., Clinical characteristics of coronavirus disease 2019 in China, N. Engl. J. Med. 382 (18) (2020) 1708–1720.
[5] J.K. Aronson, R.E. Ferner, Biomarkers - a general review, Curr. Protoc. Pharmacol. 76 (1) (2017) 9–23.
[6] L. Yan, et al., An interpretable mortality prediction model for COVID-19 patients, Nat. Mach. Intell. (2020) 1–6.
[7] Q. Ruan, et al., Clinical predictors of mortality due to COVID-19 based on an analysis of data of 150 patients from Wuhan, China, Intensive Care Med. 46 (5) (2020) 846–848.
[8] W.-j. Guan, et al., Comorbidity and its impact on 1590 patients with covid-19 in China: a nationwide analysis, Eur. Respir. J. 55 (5) (2020).
[9] P. Goyal, et al., Clinical characteristics of Covid-19 in New York city, N. Engl. J. Med. 382 (2020).
[10] P. Malik, et al., Biomarkers and outcomes of COVID-19 hospitalisations: systematic review and meta-analysis, BMJ Evid. Based Med. 26 (3) (2020) 107–108.
[11] G. Benelli, et al., SARS-COV-2 Comorbidity Network and Outcome in Hospitalized Patients in Crema, Italy, 2020 medRxiv.
[12] Y. Liu, et al., Neutrophil-to-lymphocyte ratio as an independent risk factor for mortality in hospitalized patients with COVID-19, J. Infect. 81 (1) (2020) e6–e12.
[13] P. Sarzi-Puttini, et al., COVID-19, cytokines and immunosuppression: what can we learn from severe acute respiratory syndrome? Clin. Exp. Rheumatol. 38 (2) (2020) 337–342.
[14] A. Cheng, et al., Diagnostic performance of initial blood urea nitrogen combined with D-dimer levels for predicting in-hospital mortality in COVID-19 patients, Int. J. Antimicrob. Agents 56 (3) (2020) 106110.
[15] Y. Gao, et al., Machine learning based early warning system enables accurate mortality risk prediction for COVID-19, Nat. Commun. 11 (1) (2020) 1–10.
[16] A. Cho, AI systems aim to sniff out coronavirus outbreaks, Science 368 (2020) 810–811.
[17] B.H. X, et al., AI augmentation of radiologist performance in distinguishing COVID-19 from pneumonia of other etiology on chest CT, Radiology (2020), 201491.
[18] X. Mei, et al., Artificial Intelligence-Enabled Rapid Diagnosis of Patients with COVID-19, 2020 medRxiv.
[19] G. Wu, et al., Development of a clinical decision support system for severity risk prediction and triage of COVID-19 patients at hospital admission: an international multicenter study, Eur. Respir. J. 56 (2) (2020).
[20] L. Wynants, al. et, Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal, BMJ 369 (2020).
[21] W.M. Shaban, A.H. Rabie, A.I. Saleh, M.A. Abo-Elsoud, A new COVID-19 Patients Detection Strategy (CPDS) based on hybrid feature selection and enhanced KNN classifier, Knowl. Base Syst. 205 (2020) 106270.
[22] R.M. Pereira, D. Bertolini, L.O. Teixeira, C.N. Silla Jr., Y.M. Costa, COVID-19 identification in chest X-ray images on flat and hierarchical classification scenarios, Comput. Methods Progr. Biomed. (2020), 105532.
[23] L. Yan, et al., A Machine Learning-Based Model for Survival Prediction in Patients with Severe COVID-19 Infection, 2020. MedRxiv.
[24] V. Jurisic, S. Radenkovic, G. Konjevic, The actual role of LDH as tumor marker, biochemical and clinical aspects, Adv. Canc. Biomarkers (2015) 115–124.
[25] X. Li, et al., Risk factors for severity and mortality in adult COVID-19 inpatients in Wuhan, J. Allergy Clin. Immunol. 146 (1) (2020) 110–118.
[26] X. Li, et al., Clinical characteristics of 25 death cases with COVID-19: a retrospective review of medical records in a single medical center, Wuhan, China, Int. J. Infect. Dis. 94 (2020) 128–132.
[27] L. Tan, et al., Lymphopenia predicts disease severity of COVID-19: a descriptive and predictive study, Signal Trans. Targeted Ther. 5 (1) (2020) 1–3.
[28] H. Yao, N. Zhang, R. Zhang, M. Duan, T. Xie, J. Pan, G. Wang, Severity detection for the coronavirus disease 2019 (COVID-19) patients using a machine learning model based on the blood and urine tests, Front. Cell Develop. Biol. 8 (2020) 683.
[29] M. Pulgar, K. Chamorro, M. Fors, F.X. Mora, H. Ramírez, E. Fernandez-Moreira, S. J. Ballaz, Biomarkers of severe COVID-19 pneumonia on admission using data-mining powered by common laboratory blood tests-datasets, Comput. Biol. Med. (2021), 104738.
[30] M. Mahdavi, H. Choubdar, E. Zabeh, M. Rieder, S. Safavi-Naeini, Z. Jobbagy, E. Kamrani, A machine learning based exploration of COVID-19 mortality risk, PloS One 16 (7) (2021), e0252384.
[31] A. de Fátima Cobre, D.P. Stremel, G.R. Noleto, M.M. Fachi, M. Surek, A. Wiens, R. Pontarolo, Diagnosis and prediction of COVID-19 severity: can biochemical tests and machine learning be used as prognostic indicators? Comput. Biol. Med. (2021), 104531.
[32] H.-Y. Kim, Analysis of variance (ANOVA) comparing means of more than two groups, Restorative eentistr. Endodontics 39 (1) (2014) 74–77.
[33] N. Altman, M. Krzywinski, Points of significance: clustering, Nat. Methods 14 (2017) 545–546.
[34] T. Chen, C. Guestrin, Xgboost: a scalable tree boosting system, in: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, 2016, pp. 785–794.
[35] M. Panteghini, Lactate dehydrogenase: an old enzyme reborn as a COVID-19 marker (and not only), Clin. Chem. Lab. Med. 1 (2020) (ahead-of-print).
[36] Y. Cheng, et al., Kidney disease is associated with in-hospital death of patients with COVID-19, Kidney Int. 97 (5) (2020) 829–838.
[37] A. Banerjee, S. Ray, B. Vorselaars, J. Kitson, M. Mamalakis, S. Weeks, M. Baker, L. Mackenzie, Use of machine learning and artificial intelligence to predict SARS-CoV-2 infection from full blood counts in a population, Int. Immunopharm. 86 (2020), 106705.
[38] R. Joshi, V. Pejaver, N. Hammarlund, H. Sung, S. Lee, A. Furmanchuk, H. Lee, G. Scott, S. Gombar, N. Shah, S. Shen, A. Nassiri, D. Schneider, F. Ahmad, D. Liebovitz, A. Kho, S. Mooney, B. Pinsky, N. Banaei, A predictive tool for identification of SARS-CoV-2 PCR-negative emergency department patients using routine test results, J. Clin. Virol. 129 (2020), 104502.
[39] A.J. Heffernan, Host diagnostic biomarkers of infection in the ICU : where are we and where are we going ? Curr. Infect. Dis. Rep. 23 (2021) 4.
[40] L. Liu, Y. Yu, Z. Fei, M. Li, F.X. Wu, H.D. Li, J. Wang, An interpretable boosting model to predict side effects of analgesics for osteoarthritis, BMC Syst. Biol. 12 (6) (2018) 29–38.