Available online at www.medicinescience.org

ORIGINAL RESEARCH

Medicine Science International Medical Journal

Medicine Science 2021;10(4):1516-23

Ensemble learning-based prediction of COVID-19 positive patient groups determined by IL-6 levels and control individuals based on the proteomics data

Seyma Yasar[1], Zeynep Kucukakcali[1], Adem Doganer[2]

[1]Inonu University, Faculty of Medicine, Biostatistics and Medical Informatics, Malatya, Turkey
[2]Kahramanmaras Sutcu Imam University, Faculty of Medicine, Department of Biostatistics and Medical Informatics, Kahramanmaras, Turkey

Received 17 September 2021; Accepted 04 November 2021
Available online 01.12.2021 with doi: 10.5455/medscience.2021.09.283

Abstract

Coronavirus disease (COVID-19) is a newly found coronavirus that causes an infectious disease. COVID-19, which has a detrimental impact on many people, has varied effects on different people. Therefore, proteomic analysis is an important approach used to develop early diagnosis and treatment strategies. This research to classify COVID-19 positive patient groups represented by interleukin 6 (IL-6) levels (low, medium, high) and control groups based on proteomic analysis using ensemble learning methods (Adaboost, Bagging, Stacking, and Voting). The public dataset from a website consists of 49 subjects (31 COVID-19 positives and 18 controls) and 493 proteins achieved from blood samples. The dataset was handled to estimate the relation between disease severity and proteins using ensemble learning approaches (Adaboost, Bagging, Stacking, and Voting) using ten-fold cross-validation. Predictions were evaluated with accuracy, sensitivity,etc. performance metrics. The accuracy of Adaboost (96.00%) was higher as compared to Voting (93.88%) and Bagging (91.84%). However, the Stacking ensemble learning method produced the highest accuracy (97.92%). IL6, SERPINA3, SERPING1, SERPINA1, and GSN were the five most important proteins associated with disease severity. In comparison to the other methods, the suggested ensemble learning model (Stacking) produced the best estimation of disease severity based on proteins. The results indicate that changes in blood protein levels correlated with the severity of COVID-19 may be benefited to follow early diagnosis/treatment of the COVID-19 disease.

Keywords: Adaboost, bagging, COVID-19 severity, ensemble learning, stacking, voting

Introduction

COVID-19 is a global danger caused by the coronavirus that causes severe acute respiratory syndrome 2 (SARS-CoV-2). Its symptoms include a prolonged dry cough, dyspnea, myalgia headache, loss of taste or smell, and gastrointestinal discomfort. Clinically, the course of COVID-19 disease is highly varying from individual to individual. In some patients, COVID-19 occurs with mild symptoms, while in quite a significant part, it occurs with symptoms that progress to acute respiratory distress. To date, no specific antiviral approach has proven successful in treating the disease. The majority of COVID-19 patients recover spontaneously without any antiviral treatment. However, early detection and treatment for severe and critical patients are very important issues that require urgent investigation. The research findings to assess the severity of COVID-19 show that proinflammatory cytokines play a very important role in lung damage pathophysiology in COVID-19 patients. [1]. Interleukin-6 (IL-6), a proinflammatory cytokine, is one of the main mediators of the inflammatory and immune response resulting from infection or injury, and more than half of COVID-19 patients have elevated levels of IL-6 [2]. Changes in IL-6 levels appear to be associated with inflammation, respiratory failure, need for mechanical ventilation/intubation, and mortality in COVID-19 patients [3, 4]. As considering all these results, it can be concluded that determining the effect of IL-6 levels on the proteome of COVID-19 patients plays an important role in predicting the severity of the disease. Machine learning methods have been widely employed to diagnose diseases and clinical decision support systems in recent years. Therefore, during these periods when the COVID-19 disease peaked, the need for studies in which algorithms capable of classifying with high accuracy by combining the dataset(s) obtained from COVID-19 studies and machine learning methods have increased considerably. While machine learning methods provide high accuracy performance in many complex data sets with powerful algorithms, the

***Corresponding Author:** Seyma Yasar, Inonu University, Faculty of Medicine, Biostatistics and Medical Informatics, Malatya, Turkey
E-mail: seyma.yasar@inonu.edu.tr

approaches perform classifications with high variance and low accuracy values in some data sets. Different methods have been proposed to prevent performance loss in the classification and estimation processes. One of these methods is ensemble learning methods. Generalization ability for the predictions of more than one algorithm (i.e., ensemble methods) is stronger than that of a base algorithm and can perform very precise estimations. As a result, the main idea of ensemble learning methods is based on the idea of combining many base classifiers to obtain a more accurate and reliable model (meta classifier) compared to the classification success that a base classifier (model) can achieve [5].

This study intends to classify COVID-19 positive patients represented by IL-6 levels (low, medium, high) and control group individuals using ensemble learning methods (i.e., Adaboost, Voting, Bagging, Stacking) based on the data obtained by proteomic analysis of blood sera of subjects.

Materials and Methods

Research design and dataset

The relevant data of experimental research were acquired from a public website address (http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PXD020601). In the open-sourced dataset used in this study, the thirty-three COVID-19-positive patients were included in Columbia University Irving Medical Center/New York-Presbyterian Hospital as determined by SARS-CoV-2 nucleic acid testing of nasopharyngeal swabs. In this sample, the magnitude of the disease was inferred from serum IL-6 levels as determined by CLIA-certified ELISA measurements. The control group comprised 16 participants, all of whom were nasopharyngeal swab-negative for SARS-CoV-2 when drawing blood samples. The proteomics analyses were conducted on sera from 49 subjects. Results are declared comprehensively, which identifies 493 proteins. COVID-19 patients were divided using path analysis to have low (<10 pg/mL), medium (10−65 pg/mL), and high (>90 pg/mL) IL-6 levels. Two subjects in the patient group were included in the control group according to the values obtained from the proteomic analysis. As a result, COVID-19 positive groups are determined as sixteen subjects in the low IL-6 group, five subjects in the medium IL-6 group, ten subjects in the high IL-6 group, and eighteen subjects in the control groups [6].

Data Preprocessing

Many real-world datasets may include missing values for various reasons. Therefore, serious problems arise in many statistical analyzes made with these datasets and in the performance of data mining algorithms. Removing the missing observations from the data set causes the sample size to decrease and the statistical power of the analysis to decrease [7]. For this reason, missing values in the data set are imputed using the "Impute Missing Value" operator in RapidMiner Studio. The Random Forest learner for estimating missing values was placed in the subprocess of this operator [8]. Instead of excluding observations with missing values from the data set, Random Forest allows them to stay in the data set with an algorithm that calculates the proximity measure [9]. Another problem of machine learning methods is the unbalanced classes in the dataset. Therefore, the classes in the dataset are balanced using the Synthetic minority over-sampling technique (SMOTE). Based on feature space similarities between existing minority observations, the SMOTE algorithm creates synthetic/artificial data. The working steps of the algorithm can be summarized as follows:

Step-1: The k nearest neighbors of each observation belonging to the minority class are searched,

Step-2: The difference between the observation belonging to the minority class and the observation with k close neighbors (kNN) is taken,

Step-3: A random number (α) is chosen between (0,1), this number is multiplied by the difference found in Step 2,

Step-4: A new synthetic observation is obtained using the following equation.

$$\chi_{new}=\chi_i+(\chi_j-\chi_i)*\alpha$$

Step-5: Repeat Steps 1-4 to generate the desired number of synthetic observations [10]. "Boruta" variable selection method was used to increase the performance of machine learning methods to be used in our study. The algorithm of the Boruta is intended as a wrapper around an algorithm for the classification of Random Forest and iteratively excludes the features found by a statistical test to be less significant than random samples [11].

Machine learning approaches implemented in ensemble learning

The machine learning algorithms described below were used to build ensemble learning approaches.

Deep Learning

Although deep learning is based on artificial neural networks as a basic approach, it is much more than that. Deep learning (DL) is an artificial neural network model with an increased number of layers. Many hidden layers have been added to the artificial neural network, and with the help of these layers, the output layer has been found. Since there are many hidden layers in deep learning, the number of parameters is quite high. Each layer consists of multiple trainable layers placed behind itself [12].

Decision Tree

The Decision Tree (DT), which can be described as a recursive partitioning of the instance space, is one of the most commonly used practical approaches compared to other algorithms, and the classification algorithm is a structure that divides the data set into sub-sections with appropriate procedures. Advantages of decision tree algorithms; establishing understandable rules, having the ability to work with very large data, the model is easy to interpret and understand, to be able to work with continuous and categorical data, be operable even if there are missing values in the data set. In addition to all these advantages, there are also negative aspects such as lowering the performance of the algorithm in multi-class problems when there are continuous variables and the number of training data is small or limited [13].

Random Forest

Random Forest (RF) is an method developed based on a decision tree. Decision trees that exist in more than one form are combined with each other and transformed into multiple decision trees, and all decision trees are independent of each other. The final classification results are selected according to simple and multiple selective voting methods. If a random forest algorithm will be used to solve a regression problem, the mean square error (MSE) formula is used. If a random forest algorithm will be used to solve a classification problem, the Gini index formula is used [14].

Gradient Boosted Trees

Gradient boosted trees (GBT's) is a learning algorithm based on the optimization of the loss function. This method, which Friedman introduced in 2001, generally uses the mean squared error, which is also known as multiple additive regression trees. It is aimed that the estimates made by the model have the lowest value of the loss function. For this purpose, the estimates are updated with the learning coefficient determined using the gradient descent algorithm, and the MSE value is tried to be minimized [15].

Ensemble learning approaches

Bagging (Bootstrap aggregating)

In the Bagging ensemble learning method based on the Bootstrap sampling method, single algorithms are trained with different datasets created by dividing the training data set into equal sample numbers. At the same time, all classifiers classify different sub-training sets. The bagging method uses the majority vote technique to combine the classifiers' estimates. In this method, the classifiers' majority estimation among all the estimators' classification predictions is accepted as the classification guess of the ensemble method [16]. In this study, the RF algorithm was used as a nested classifier in the Bagging ensemble learning method.

Adaboost

Adaptive Boosting is one of the powerful and widely applied ensemble learning methods in boosting ensemble learning methods. This method, proposed by Freund and Shapire, aims to achieve a strong result by combining the weak results obtained from the data. In the first step, it distributes the data evenly and then makes a classification. As a result of this classification, it finds the weakest classifier and re-weighs it. During the re-weighting process, it focuses on the lowest outcome [14]. It combines several weak classifiers to create a successful classifier. Its purpose is to increase its success in terms of classification. Thus, it is possible to reduce errors and increase correct classifications at every stage [16]. In this study, the RF algorithm, which was explained earlier, was used as a classifier in the Adaboost ensemble learning method.

Voting

The main idea of the Voting method, which is one of the ensemble learning methods, is to estimate the highest voted class label by collecting the estimates of each basic classifier it will combine. In the voting method, the weight of all basic classifiers is equal. The Voting method often achieves a higher accuracy rate than the classifier that provides the highest accuracy in the ensemble. When the base classifiers are different from each other (diverse) in ensemble learning methods, the approaches can make different types of errors and produce lower predictions. That is why the formed meta classifier gives more successful results [5]. In this study, RF, DL, and GBT's algorithms were used as classifiers in the Voting ensemble learning method.

Stacking

The stacking ensemble learning method developed by Wolpert is a simple ensemble learning technique that creates a meta classifier by combining base, multiple classification models. In other words, the Stacking model is another ensemble model that is trained by combining the estimates of two or more basic classifier models. Predictions made from models created by the base classifier are used as input for each ordered layer and are combined to create a new set of predictions. In the stacking method, base classification models are trained on the original training data set. The meta-classifier is then created based on the outputs (estimates) of the ensemble's basic classification models [17]. In this study, RF, DC, and DL algorithms were used as classifiers in the Stacking ensemble learning method. GBT's algorithm was determined as a meta classifier. The information gain ratio technique was implemented to determine the related predictors with COVID-19.

Resampling procedure and performance evaluation metrics

Cross-validation is a procedure of resampling generally used on a small dataset(s) to validate machine learning models. The ten-fold cross-validation was applied to test the validity of the models. These approach make it possible to use the whole data set during the modeling phase. In this approach, the dataset is randomly divided into ten equal parts. Nine of these equal parts were used as training data and one as test data. In this way, an accuracy calculation is made, then an accuracy calculation is made by replacing the test data set and training data sets, and the accuracy rate of the model is calculated by taking the average of the accuracy values [18]. Performance metrics for all models are given with accuracy, classification error, kappa, F1 measure, sensitivity, specificity, and G-mean.

Data Analysis

The whole data set consists of quantitative variables. Therefore, the conformity of all variables to the normal distribution was checked with the Shapiro Wilk test. Data are summarized with mean, standard deviation, median, and min-max. Kruskal Wallis test and one-way analysis of variance test were used for statistical analysis. After the Kruskal Wallis test, the Conover test was used for multiple comparisons, while the Tukey and Tamhane T2 tests were used where appropriate for the One-Way analysis of variance. The effect size was calculated to evaluate the effects of each protein on the COVID-19 positive and control groups [19]. According to the statistical analyzes used (Kruskal-Wallis, one-way analysis of variance), the interpretation values of the literature are generally accepted as small effect size between 0.01-0.06, medium effect size between 0.06-0.14, and large effect size greater than 0.14 [20]. $p < 0.05$ was considered statistically significant. Data analysis was performed with the programming languages "Statistical Analysis Software"[21], "DTROC: Diagnostic Tests and ROC Analysis Software" [22], RStudio Version 3.6.2 [20], and RapidMiner Studio Version 9.8 [23].

Results

Baseline characteristics of the chosen variables by feature selection

In the dataset, variables with missing observations were assigned missing values using the RF algorithm within the "Impute Missing Value" operator in RapidMiner Studio. Then, the unbalanced groups in the dataset were balanced using SMOTE to include 16 subjects in the low IL-6 group, 18 subjects in the medium IL-6 group, ten subjects in the high IL-6 group, and 18 subjects in the control group. When the variable selection method was applied to the Boruta, 28 proteins remained in the dataset. Descriptive statistics for the new 28 protein datasets obtained as a result are represented in Table 1.

When considering Table 1, the differences between the groups regarding all the proteins in the dataset are statistically significant ($p<0.001$).

Table 1. Descriptive statistics of the proteins in the preprocessing dataset by the groups

Proteins Names	Groups				Effect Size	p-value*
	Low IL-6	Medium IL-6	High IL-6	Control		
	Mean ± SD	Mean ± SD	Mean ± SD	Mean ± SD		
ITIH4	46587.5a±5422.9	90533.06b±19751.49	91720b±20925.78	100650b±32204.55	0.50 (Large)	**<0.001***
GSN	26437.5a±3544.17	12882.22b±1663.08	10039b±3197.72	12342.22b±4684.51	0.78 (Large)	**<0.001***
F13B	10964.38a±1967.09	7101.33b±1354.58	5893b±1642.37	6689.44b±1831.21	0.57 (Large)	**<0.001***
SERPINA4	11773.75a,c±2051.77	13892.83a±2162.65	6033b±1934.34	9095.56b,c±5101.5	0.43 (Large)	**<0.001***
CPN1	1581.25a±398.53	3457.33b±845.18	3470b±824.51	4288.33b±1923.89	0.44 (Large)	**<0.001***
C9	12870a±2479.79	22409.94b±4308.64	28990b±10031	28211.67b±11761.84	0.40 (Large)	**<0.001***
IGHA1	82593.75a±22234.49	125291.56b±17629.78	159700b±26008.76	150350b±48831.45	0.47 (Large)	**<0.001***
CFI	9680.63a±1147.89	15699.28b±3024.72	14255b±3281.47	15656.67b±5908.75	0.32 (Large)	**<0.001***
TTR	105318.75a±14087.36	105831.22a±15685.04	67210b±22095.62	105827.78a±49808.28	0.19 (Large)	**<0.001***
C1QB	18206.25a±2799.64	27569.33b±9384.94	21930b±2855.81	25466.67b±4836.5	0.29 (Large)	**<0.001***
	Median (Min-Max)	**Median (Min-Max)**	**Median (Min-Max)**	**Median (Min-Max)**		
IL6	5(5-5)	5.46b(3,1-8,7)	32.75a(11.4-61.3)	315b(96-315)	0.83 (Large)	**<0.001****
SERPINA1	409000a(332000-550000)	656177b(615000-927000)	801000b(621000-1000000)	748000b(482000-1400000)	0.61 (Large)	**<0.001****
VWF	7680a(6020-11400)	21054.5b(8560-28300)	18600b(6900-47500)	20600b(11200-56200)	0.49 (Large)	**<0.001****
CFB	42550a(30300-55600)	68159b(43200-112000)	79150b(52300-120000)	83150b(27700-169000)	0.51 (Large)	**<0.001****
SERPINA3	37550a(27400-43500)	145063b(125000-186000)	157500b(123000-209000)	156500b(65800-359000)	0.56 (Large)	**<0.001****
C7	11100a(8510-16400)	19298b(12200-27000)	16850b(12600-49300)	21550b(14400-33000)	0.54 (Large)	**<0.001****
SERPING1	29100a(20700-38500)	72521b(57900-73600)	59350b(48700-98100)	66700b(47100-132000)	0.62 (Large)	**<0.001****
C1S	14000a(11600-17100)	16847,5a(15100-21000)	21100a,b(15400-25500)	25600b(14100-70500)	0.62 (Large)	**<0.001****
APOE	23050a(14200-30600)	43874b(30900-64000)	35450b(23000-109000)	42600b(29000-93600)	0.51 (Large)	**<0.001****
LBP	1415a(879-2730)	3327b(2790-4600)	6080b,c(3990-17200)	7025c(1290-28400)	0.62 (Large)	**<0.001****
SERPINF1	2995a(2200-4540)	4219a(3970-5670)	6835b(3420-9270)	6135b(3230-14600)	0.65 (Large)	**<0.001****
ORM1	12250a(5050-21000)	27174.5b(24600-37000)	32850b(24900-42400)	25650b(14400-48400)	0.56 (Large)	**<0.001****
LRG1	6465a(4130-10700)	13505b(10600-21800)	27350c(16900-38900)	21000d(7790-39900)	0.67 (Large)	**<0.001****
SAA1	789.5a(373-1540)	24177a,b(4410-51300)	32550b,c(4910-102000)	24300b,c(300-138000)	0.51 (Large)	**<0.001****
C5	45800a(37700-63700)	80377b(72800-110000)	75450b(61500-103000)	69200c(23900-110000)	0.53 (Large)	**<0.001****
HRG	31150a(23000-38500)	12432b(4980-27900)	11750b(7260-29900)	22100a,b(8660-70800)	0.41 (Large)	**<0.001****
RBP4	10575a(4040-22200)	14783a(12900-25300)	19650a,b(5150-35700)	27950b(4620-53000)	0.34 (Large)	**<0.001****
S100A9	1140a(581-2890)	4620a(1530-9600)	5040a,b(1450-20000)	5325b(1060-32100)	0.49 (Large)	**<0.001****

a, b,c: Different characters in each row show a statistically significant difference ($p<0.05$); *: One-way analysis of variance; **: Kruskal-Wallis test

Findings of the constructed ensemble models

In this study, ensemble learning models (Adaboost, Bagging, Stacking, and Voting) are composed to classify three COVID-19 positive patient groups (low, medium, high) and a control group based on proteomic analysis from sera. Considering the model performance metric results in Table 2, the Stacking ensemble learning models gave the most successful result. The accuracy rate based on Adaboost ensemble learning models (96.00%) was more successful than the accuracy rate based on the other ensemble learning models (Bagging 92.00%, Voting 94.00%). In kappa statistics, which measure the reliability of the statistical fit, the Adaboost, Bagging, Stacking, and Voting ensemble learning models represent a perfect fit with the values of 0.947, 0.887, 0.971, and 0.917.

Figure 1 exhibits the pseudo-code of the Stacking ensemble learning model, which gives the best results in classifying COVID-19 positive represented by IL-6 levels and control individuals based on the proteomics data.

1: Input: Training data $D = \{x_i, y_i\}_{i=1}^{m}$
2: Output: ensemble classifier H
3: Step 1: learn base-level classifiers
4: for t=1 to T do
5: learn h_t *based on D*
6: end for
7: Step 2: construct new data set of predictions
8: for i= 1 to m do
9: $D_h = \{x_i', y_i\}$, *where* $x_i' = \{h_1(x_1), \ldots, h_T(x_i)\}$
10: end for
11: Step 3: learn a meta-classifier
12: learn H based on D_h
13: return H

Figure 1. The pseudo-code of the Stacking ensemble learning model

Table 3 and Figure 2 display the importance levels of top ten proteins in COVID-19 positive represented by IL-6 levels and control individuals on the severity of the disease in the Stacking ensemble learning modeling with the value of information gain ratio (IGR).

Table 2. Performance metrics of the ensemble learning models

Models	Groups	Sensitivity (%)	Specificity (%)	Precision (%)	F1 (%)	G-mean (%)	Accuracy (%)	Class. Error	Kappa
Stacking	Low IL6	100	100	100	100	100	97.92	2.08	0.971
	Medium IL-6	80.00	100	100	88.89	98.88			
	High IL-6	100	97.44	90.91	95.24	95.35			
	Control	100	100	100	100	100			
Adaboost	Low IL6	100	96.97	94.12	96.97	97.01	96.00	4.00	0.947
	Medium IL-6	80.00	100	100	88.89	98.86			
	High IL-6	100	97.44	90.91	95.24	95.35			
	Control	94.44	100	100	97.14	98.37			
Voting	Low IL6	100	100	100	100	100	93.88	6.00	0.917
	Medium IL-6	60.00	97.72	75.00	66.67	84.66			
	High IL-6	90.00	94.87	81.82	85.71	89.26			
	Control	100	100	100	100	100			
Bagging	Low IL6	100	96.97	94.12	96.97	97.01	91.84	8.00	0.887
	Medium IL-6	60.00	100	100	75.00	97.70			
	High IL-6	90.00	94.87	81.82	85.71	89.22			
	Control	94.44	96.77	94.44	94.44	95.49			

Table 3. Variable importance values for the Stacking ensemble learning model

Proteins	IGR	Proteins	IGR	Proteins	IGR	Proteins	IGR
IL6	1.00	SAA1	0.69	VWF	0.45	IGHA1	0.31
SERPINA3	0.99	CPN1	0.69	APOE	0.40	S100A9	0.26
SERPING1	0.99	C7	0.56	F13B	0.38	C1S	0.25
SERPINA1	0.80	CFB	0.52	CFI	0.37	HRG	0.17
GSN	0.80	LRG1	0.49	C1QB	0.35	SERPINA4	0.11
ITIH4	0.69	ORM1	0.49	C5	0.34	RBP4	0.03
LBP	0.69	SERPINF1	0.47	C9	0.33	TTR	0.00

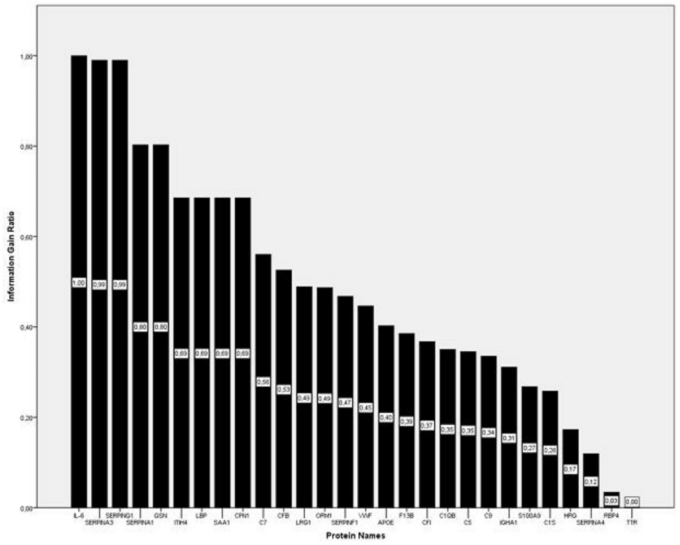

Figure 2: The graphical representation of variable importance values for the Stacking ensemble learning model

The IL6 (1.00), SERPINA3 (0.99), SERPING1 (0.99), SERPINA1 (0.80), and GSN (0.80) provided the highest IGR, while the lowest IGR were for IGHA1 (0.31), S100A9 (0.27), C1S (0.26), HRG (0.17), SERPINA4 (0.12), and RBP4 (0.03) from the Stacking ensemble learning model.

Discussion

Nowadays, COVID-19 has been a severe public health topic globally. In clinics for the SARS-CoV-2 infection diagnosis, nasopharyngeal swab samples and serological tests are regularly used. However, biomarkers are yet to be detected for disease prognosis until it could lead to lethal symptoms. The comprehension of the host response against the viral infection might provide an important sign about the advancement of disease severity. Nowadays, proteomics approaches are applied for a detailed understanding of the structure of disease. Omics models that allow understanding disease structure support clinicians in coping with the COVID-19 outbreak [24]. Although COVID-19 can be diagnosed at an early stage with methods based on proteomic analyses, detecting critical COVID-19 patients before the appearance of symptoms to decrease mortality is uniformly important. Hence, the development of methods based on artificial intelligence and machine learning approaches made with the data obtained from proteomic analysis plays an important role in the early diagnosis and diagnosis of the disease [25]. In this research, three positive (mild/extreme/critical) COVID-19 patient groups represented by IL-6 levels and a control group may be recognizable based on ensemble learning models (i.e., Adaboost, Bagging, Stacking, and Voting) related to proteomic analysis of serum from COVID-19 patients. Considering the empirical results from the current research, it can be terminated that the ensemble models based on the results of proteomic analyses generate promising prediction outcomes in classifying COVID-19 positive (mild/extreme/critical) represented by IL-6 levels and the control group. When the ensemble learning methods' prediction results are compared according to the performance metrics (i.e., accuracy, sensitivity,etc.), the Stacking ensemble learning method outperforms Adaboost, Bagging, and Voting on aforementioned the classification. A newly published research has reported that the accuracy (89.36%) of the stacking model is considerably superior to those of the single models (MLP, KNN, CART, and SVM) for determining transformer faults [26]. Another recent study has explored various techniques for ensemble learning, such as bagging, boosting, and stacking for landslide susceptibility mapping, and has demonstrated that Stacking can offer a promising method for stable and enhanced modeling of the landslide [27]. The results of the outlined surveys are in line with the outcomes of ensemble learning techniques in the present study.

The five top proteins, IL6, SERPINA3, SERPING1, SERPINA1, and GSN, calculated from the best performing Stacking ensemble learning method, can be used as biomarkers in the COVID-19 severity classification. In infection and tissue injury, IL-6 is promptly secreted as an immune response by strictly regulated transcriptional and post-transcriptional mechanisms [28]. Present studies in literature proposed that the rapid progression of the disease's severity in the outbreak of COVID-19 might be due to the cytokine storm or cytokine release syndrome [29, 30]. Hence, the violently elevated levels of IL-6 play a very important role in the deterioration of the health of COVID-19 patients [31]. The elevated levels of IL6 observed in COVID-19 positive patients can progress from acute respiratory distress syndrome to

severe pneumonia, leading to multisystem organ failure and high mortality [32, 33]. A study has shown that with an increase in IL-6 level, upcoming respiratory failure can be predicted with high accuracy and can help doctors accurately distinguish patients by disease severity at an early stage [6]. Many studies in the literature have indicated that IL-6 can be used as a biomarker to determine the severity of COVID-19 [34-36]. Thence, the proposed Stacking ensemble learning model's findings indicate that the IL-6 protein is significantly associated to COVID-19 severity, as described by the past works.

Similarly, the top five proteins identified by the proposed model are three types of SERPIN components (SERPINA3, SERPING1, and SERPINA1). Proteolytic events within various biological processes, including digestion, coagulation, inflammation, and immune responses, are regulated by serine protease inhibitors (serpins) [37]. A study has shown that the aforementioned SERPIN's proteins are at high expression levels in COVID-19 patient sera representing high IL-6 levels. In this context, the Covid-19 severity may be associated with an increase in the levels of SERPIN proteins [6]. On the other hand, when the effect sizes for selected proteins are examined, the four proteins with the highest values are IL-6 (0.83), GSN (0.78), LRG1 (0.67), and SERPINF1 (0.65), respectively. The findings of the proposed Stacking ensemble model demonstrate that the five proteins (i.e., SERPINA3, IL6, SERPING1, SERPINA1, and GSN) are significantly identify with COVID-19 severity, as expressed by the preceding work.

There are many studies on proteomics data to classify COVID-19 severity based on machine learning approaches. In one study, they performed plasma proteomics of a population of COVID-19 patients, including non-survivors and survivors suffering from moderate to severe symptoms, profiling host responses to COVID-19 and discovered various plasma protein changes consistent with COVID-19. Using the proteomics, they used machine learning-based models (penalized logistic regression) to distinguish patients of different severity (fatal, severe, mild, health). The area under the curve for the fatal, severe, mild and health groups of the created model is 0.952, 0.917, 0.974, and 0.983, respectively [38]. In another study, a machine learning approach (ExtraTrees classifier) was used to estimate the severity of COVID-19 as a result of multi-omic analysis of 128 blood samples from COVID-19-positive and COVID-19-negative patients with different disease severity and consequences. The importance degree of COVID-19 of the model created was 0.80 with an accuracy value [39]. Various ensemble learning methods have been used in the classification of Covid-19 disease with different radiological images [40, 41]. However, the number of studies using one or more of the omic technologies together remained quite limited. However, as far as we know, no study investigates serum proteomics data of COVID-19 patients regarding IL-6 levels based on ensemble learning methods. Therefore, we precipitate that the current research provides the first proteomics analyses of serum in COVID-19, stratiformed by the circulation of IL-6 levels using ensemble learning techniques.

Ultimately, the suggested model (Stacking ensemble learning model) realized the best prediction of disease severity based on the proteins comparatived to the other algorithms. The results indicate that alters in blood protein levels correlated with the disease's severity may be used in following the severity of COVID-19 disease and in early diagnosis and treatment.

Conflict of interests
The authors declare that they have no competing interests.

Financial Disclosure
All authors declare no financial support.

References

1. Ye Q, Wang B, Mao J. Cytokine storm in COVID-19 and treatment. J Infection. 2020;80:607-13.
2. Zhang ZL, Hou YL, Li DT, et al. Laboratory findings of COVID-19: a systematic review and meta-analysis. Scand J. 2020;80:441-7.
3. Aziz M, Fatima R, Assaly R. Elevated interleukin-6 and severe COVID-19: A meta-analysis. J Med Virol. 2020;1-3.
4. Herold T, Jurinovic V, Arnreich C, et al. Elevated levels of interleukin-6 and CRP predict the need for mechanical ventilation in COVID-19. J Allergy Clin Immunol. 2020;146:128-36.
5. Schwenker F. Ensemble methods: Foundations and algorithms [book review]. IEEE Comput Intell Mag.2013;8:77-9.
6. D'Alessandro A, Thomas T, Dzieciatkowska M, et all. Serum proteomics in COVID-19 patients: Altered coagulation and complement status as a function of IL-6 level. J Proteome Res. 2020;19:4417-27.
7. Brown ML, Kros JF. Data mining and the impact of missing data. IMDS.2003;103:611-21.
8. Rapidminer DA. RapidMiner 4.1 User Guide. In: Dortmund, 2008.
9. He Y. Missing data imputation for tree-based models. In: University of California, Los Angeles, 2006.
10. Chawla NV, Bowyer KW, Hall LO, et al. SMOTE: synthetic minority over-sampling technique. J Artif Intell Res. 2003;16:321-57.
11. Kursa MB, Rudnicki WR. Feature selection with the Boruta package. J Stat Softw.2010;36:1-13.
12. İnik Ö, Ülker E. Deep learning and deep learning models used in image analysis. Gaziosmanpasa J Sci Res. 2017;6:85-104.
13. Brijain M, Patel R, Kushik MR, et al. A survey on decision tree algorithm for classification. Int J Eng Sci Invention Res Dev. 2014;2:1-5.
14. Wyner AJ, Olson M, Bleich J, Mease D. Explaining the success of adaboost and random forests as interpolating classifiers. J Mach Learn Res. 2017;18:1558-90.
15. Krauss C, Do XA, Huck N. (2017). Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500. Eur J Oper Res. 2017;259:689-702.
16. Polikar R. Ensemble learning. In Ensemble machine learning. Springer, Boston, 2012;1-34.
17. Divina F, Gilson A, Goméz-Vela F, et al. Stacking ensemble learning for short-term electricity consumption forecasting. Energies. 2018;11:949.
18. Browne MW. Cross-validation methods. J Math. Psychol. 2000;44:108-32.
19. Sullivan GM, Feinn R. Using effect size—or why the P value is not enough. J Grad. Med. Educ. 2012;4:279-82.
20. Allaire J. RStudio: integrated development environment for R. MA. 2012;770:165-71.
21. Yasar Ş, Arslan AK, Colak C, et al. A Developed Interactive Web Application for Statistical Analysis: Statistical Analysis Software. MBSJHS. 2020;6:227-39.
22. Yasar Ş, Arslan AK, Yologlu S, et al. DTROC: Diagnostic Tests and ROC analysis Software [Web-tabanlı yazılım].2019.
23. Hofmann M, Klinkenberg R. RapidMiner: Data mining use cases and business analytics applications. In: CRC Press. 2016.
24. López-Úbeda P, Díaz-Galiano MC, Martín-Noguerol T, et al. COVID-19 detection in radiological text reports integrating entity recognition. Comput Biol Med.2020;127:104066.

25. Lalmuanawma S, Hussain J, Chhakchhuak L. Applications of machine learning and artificial intelligence for Covid-19 (SARS-CoV-2) pandemic: A review. Chaos, Solitons & Fractals. 2020;39:110059.

26. Wang X, Han T. Transformer fault diagnosis based on stacking ensemble learning. IEEJ Trans Electr Electron Eng. 2020;15:1734-9.

27. Hu X, Mei H, Zhang H, et al. Performance evaluation of ensemble learning techniques for landslide susceptibility mapping at the Jinping county, Southwest China. Natural Hazards. 2021;105:1663-89.

28. Tanaka T, Narazaki M, Kishimoto T. IL-6 in inflammation, immunity, and disease. Cold Spring Harb. Perspect. Biol. 2014;6:a016295.

29. Chen L, Liu HG, Liu W, et al. Analysis of clinical features of 29 patients with 2019 novel coronavirus pneumonia. CJTRD. 2020;43:E005.

30. Zumla A, Hui DS, Azhar EI, et al. Reducing mortality from 2019-nCoV: host-directed therapies should be an option. Lancet. 2020;395:35-6.

31. Flowers LO. COVID-19 and the Cytokine Storm. BJSTR. 2020;29:22644-7.

32. Mehta P, McAuley DF, Brown M, et al. COVID-19: consider cytokine storm syndromes and immunosuppression. Lancet, 2020;395:1033-4.

33. Zaim S, Chong JH, Sankaranarayanan V, et al. COVID-19 and multiorgan response. Curr Probl Cardiol. 2020;45:100618.

34. Kong Y, Han J, Wu X, et al. VEGF-D: a novel biomarker for detection of COVID-19 progression. Critical Care. 2020;24:1-4.

35. Ulhaq ZS, Soraya GV. Interleukin-6 as a potential biomarker of COVID-19 progression. Medecine et Maladies Infectieuses. 2020;50-382.

36. Montesarchio V, Parrella R, Iommelli C, et al. Outcomes and biomarker analyses among patients with COVID-19 treated with interleukin 6 (IL-6) receptor antagonist sarilumab at a single institution in Italy. JITC. 2020;8:e001089.

37. De Marco Verissimo C, Jewhurst HL, Tikhonova IG, et al. Fasciola hepatica serine protease inhibitor family (serpins): Purposely crafted for regulating host proteases. PLOS Negl Trop Dis. 2020;14:e0008510.

38. Shu T, Ning W, Wu D, et al. Plasma proteomics identify biomarkers and pathogenesis of COVID-19. Immunity, 2020;53:1108-22.

39. Overmyer KA, Shishkova E, Miller IJ, et al. Large-scale multi-omic analysis of COVID-19 severity. Cell systems. 2021;12:23-40.

40. Haidar A, Holloway L. Integration of deep and ensemble learning for detecting covid-19 in computed tomography images. Research Square. 2020;3:12-6.

41. Charran SR, Dubey RK. Deep learning based hybrid models for prediction of COVID-19 using Chest X-Ray. Tech Rxiv. 2020;45:2-10.