

Computer Methods in Biomechanics and Biomedical Engineering

ISSN: (Print) (Online) Journal homepage: <https://www.tandfonline.com/loi/gcmb20>

Automated prediction of COVID-19 mortality outcome using clinical and laboratory data based on hierarchical feature selection and random forest classifier

Nasrin Amini, Mahdi Mahdavi, Hadi Choubdar, Atefeh Abedini, Ahmad Shalbfaf & Reza Lashgari

To cite this article: Nasrin Amini, Mahdi Mahdavi, Hadi Choubdar, Atefeh Abedini, Ahmad Shalbfaf & Reza Lashgari (2022): Automated prediction of COVID-19 mortality outcome using clinical and laboratory data based on hierarchical feature selection and random forest classifier, Computer Methods in Biomechanics and Biomedical Engineering, DOI: [10.1080/10255842.2022.2050906](https://doi.org/10.1080/10255842.2022.2050906)

To link to this article: <https://doi.org/10.1080/10255842.2022.2050906>



Published online: 17 Mar 2022.



Submit your article to this journal [↗](#)



View related articles [↗](#)



View Crossmark data [↗](#)



Automated prediction of COVID-19 mortality outcome using clinical and laboratory data based on hierarchical feature selection and random forest classifier

Nasrin Amini^a, Mahdi Mahdavi^{b,c}, Hadi Choubdar^{b,c}, Atefeh Abedini^d, Ahmad Shalbaf^a and Reza Lashgari^b

^aDepartment of Biomedical Engineering and Medical Physics, School of Medicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran; ^bInstitute of Medical Science and Technology (IMSAT), Shahid Beheshti University, Tehran, Iran; ^cSchool of Medicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran; ^dChronic Respiratory Diseases Research Center, National Research Institute of Tuberculosis and Lung Diseases (NRITLD), Shahid Beheshti University of Medical Sciences, Tehran, Iran

ABSTRACT

Early prediction of COVID-19 mortality outcome can decrease expiration risk by alerting health-care personnel to assure efficient resource allocation and treatment planning. This study introduces a machine learning framework for the prediction of COVID-19 mortality using demographics, vital signs, and laboratory blood tests (complete blood count (CBC), coagulation, kidney, liver, blood gas, and general). 41 features from 244 COVID-19 patients were recorded on the first day of admission. In this study, first, the features in each of the eight categories were investigated. Afterward, features that have an area under the receiver operating characteristic curve (AUC) above 0.6 and the *p*-value criterion from the Wilcoxon rank-sum test below 0.005 were used as selected features for further analysis. Then five feature reduction methods, Forward Feature selection, minimum Redundancy Maximum Relevance, Relief, Linear Discriminant Analysis, and Neighborhood Component Analysis were utilized to select the best combination of features. Finally, seven classifiers frameworks, random forest (RF), support vector machine, logistic regression (LR), K nearest neighbors, Artificial neural network, bagging, and boosting were used to predict the mortality outcome of COVID-19 patients. The results revealed that the combination of features in CBC and then vital signs had the highest mortality classification parameters, respectively. Furthermore, the RF classifier with hierarchical feature selection algorithms via Forward Feature selection had the highest classification power with an accuracy of 92.08 ± 2.56 . Therefore, our proposed method can be confidently used as a valuable assistant prognostic tool to sieve patients with high mortality risks.

ARTICLE HISTORY

Received 29 September 2021
Accepted 4 March 2022

KEYWORDS

COVID-19; random forest; mortality prediction; forward feature selection; laboratory features

1. Introduction

New coronavirus 2019 (COVID-19) disease has infected more than 180 million individuals all over the world until 10 July 2021, and has tremendously strained economic and healthcare systems worldwide (WHO Coronavirus Disease (COVID-19) Dashboard. <https://covid19.who.int>). The prevalence of the disease is significant, compelling governments to take strict social distancing and preventive measures to reduce infection transmission. However, the number of hospitalized patients is still very high, and the prevalence curve of the disease and the global mortality show little improvement (WHO Coronavirus Disease (COVID-19) Dashboard. <https://covid19.who.int>). As a result, more focus on the clinical aspect of the disease is needed to reduce patient mortality rates.

An essential clinical aspect of COVID-19 is the possible outcome of the disease. Despite the high workload of patients, exhausted medical personnel, and insufficient medical resources, rapid identification of patients that have high mortality risks becomes a key factor in reducing patient mortality. The high patient load of COVID-19 could quickly overload healthcare infrastructures and result in higher mortality rates due to inefficient management of limited medical resources and personnel (Quah et al. 2020). It is even possible for a patient to suddenly progress from a mild stage to severe or critical stages (Li et al. 2020). Targeted administration of proper treatment protocols based on the patient's condition is vital (Li et al. 2020), and critical patients will need timely intensive care unit (ICU) care and mechanical ventilators. A challenge during the current pandemic is the

efficient distribution of available resources. Toward this aim, a potential intervention avenue is the prediction of patient prognosis on the first day of admission. The first day of patients' admission is crucial for various clinical decisions and also for requesting additional care equipment. Most skilled physicians are often unable to accurately predict the prognosis of COVID-19 patients. Additionally, the course of COVID-19 can take unpredictable turns from a stable state to a critical stage. To provide an automatic, reliable, and objective estimation of the prognosis of COVID-19, machine learning models could be valuable assistants since they can detect complex patterns in large datasets (Mei et al. 2020).

Different methods are used for the prediction of the prognosis of COVID-19 patients in medicine. Some studies have quantified the lung CT images of the patients for assessment of the severity of COVID-19 lung involvement and also for predicting disease progression (Francone et al. 2020; Matos et al. 2020; Chassagnon et al. 2021; Ghosh et al. 2020; Amini and Shalbaf 2022; Shoeibi et al. 2020; Shan et al. 2020; Bai et al. 2020). But lung CT images in the initial disease phase do not contain enough information to predict the patient's condition and also have harmful radiation side effects. In addition to methods based on lung CT images, other researchers have considered patients' clinical information for the prediction of COVID-19 prognosis (Chao et al. 2021; Batllés et al. 2022; Yang et al. 2020b). In one study, COVID-19 patient's medical history and demographic information such as hypertension, diabetes, chronic kidney disease, cardiovascular disease, COPD, malignancies, cancer, and asthma were designated as features for increased disease severity risk and were used to predict patients who may need ICU (<https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-with-medical-conditions.html>). The patient's vital signs like max and min blood pressure, body temperature, and SPO2 have also been used in another article to monitor the patient situation (Du et al. 2020). Oxygen saturation less than 90, decreased level of consciousness, and hypotension is indicators for hospitalization of patients with COVID-19. Sankaranarayanan et al. combined the clinical features of COVID19 patients with deep learning methods for COVID-19 mortality prediction. The clinical data were obtained within 72 hours after the first positive test. They used GRU-D recurrent neural network and showed that clinical variables including age, charlson comorbidity index, minimum oxygen saturation, fibrinogen level, and serum iron level could be used for mortality prediction in COVID-19 patients

(Sankaranarayanan et al. 2021). Other studies have used laboratory findings for COVID-19 patients to predict the patient's prognosis (Pourbagheri-Sigaroodi et al. 2020; de Moraes Batista et al. 2020; Lee et al. 2020). Although several prognostic models have been proposed for COVID-19 mortality (Yan et al. 2020; Xie et al. 2020b; Yadaw et al. 2020; Liang et al. 2020; Pan et al. 2020), no comprehensive study has evaluated and compared the power of different clinical parameters to predict the prognosis of COVID-19 patients.

The novelty and contribution of this study were three-fold; first, to develop a mortality prediction machine learning model based on clinical data that are easily and routinely collected during the first day of patient admission to avoid missing high-risk patients; second, to investigate and compare the prediction power of eight different clinical features including demographic, vital signs, and laboratory data (CBC, coagulation, kidney, liver, blood gas, and general) for predicting COVID-19 mortality outcome as patient death or survived by statistical and machine learning methods. In other words, to provide a direct comparison of mortality prediction powers between these eight groups of features; third, to find a set of best discriminative features from clinical data by designing a new hierarchical feature selection method and then inserting these selected features to different advanced classification methods for the prediction of mortality outcomes. The ability of this novel system is evaluated with 224 COVID-19 patients. This rapid and automated triage system could identify the high-risk patients who may proceed to critical states and need intensive care later to provide support for decreasing mortality rates.

2. Material and methods

2.1. Dataset

Archived electronic medical records of 328 patients who were infected with the COVID-19 were acquired randomly. The patients were admitted to the Masih Daneshvari Hospital, the largest respiratory and pulmonary care center and a specialized referral hospital for all lung and respiratory diseases in Tehran, Iran. Afterward, 84 patients were excluded with exclusion criteria being age under 18, receiving radically different treatment protocols, and having more than 20% missing data. No patients with breastfeeding or pregnancy were present in the initially selected population. Data from 244 individuals (62.6% male, 37.4% female) were included for model development. To predict mortality prognosis, two outcome classes (survived and death) were defined; the survived group

consisted of COVID-19 patients that were discharged after the completion of their treatment and two consecutive negative PCR results. The death group consisted of patients who died at any point during their treatment course. In the dataset, 115 patients were labeled as having a 'dead' outcome (47.3%), and 129 patients were labeled as having a 'survived' outcome [53,7]. All steps of this study were in accordance with the ethical roles of Shahid Beheshti University of Medical Sciences (ethics code: IR.SBMU.MSP.REC.1399.210) and informed consents were acquired from participants upon admission. The data of this study is available for other researchers at '<https://github.com/nasrinam/clinical-and-laboratory-dataset.COVID19/>'.

The data which are recorded from the first 24 hours of patients' admission comprised 41 features. These features were categorized into eight different groups; demographic and patient history, vital signs, and six different groups of laboratory features extracted from the results of blood tests (CBC, coagulation, kidney, liver, blood gas, and general). The demographic features were age and sex. The past medical history enveloped the presence or absence of smoking and common comorbidities, hypertension (HTN), diabetes mellitus (DM), and cardiovascular disease. The vital signs consisted of systolic blood pressure (BP MAX), diastolic blood pressure (BP MIN), pulse rate (PR), respiratory rate (RR), temperature (T), and blood oxygen saturation level (SPO2). As mentioned above, laboratory features extracted from the results of blood tests were categorized into six different groups; CBC, coagulation, kidney, liver, blood gas, and general. The CBC consisted of white blood cell count (WBC), neutrophil (Neutr), lymphocyte count (Lymph), red blood cell count (RBC), hemoglobin (Hgb), hematocrit (HCT), mean corpuscular volume (MCV), mean corpuscular hemoglobin (MCH), mean corpuscular hemoglobin concentration (MCHC), red cell distribution width (RDW), and platelet count (Plt). The coagulation features were erythrocyte sedimentation rate (ESR), prothrombin time (PT), partial thromboplastin time (PTT), and international normalized ratio (INR). The kidney features were Urea and creatinine (Cr). The liver features consisted of aspartate aminotransferase (AST), alanine aminotransferase (ALT), alkaline phosphatase (ALKP), bilirubin (Bili), and albumin. The general features were lactate dehydrogenase (LDH) and creatine phosphokinase (CPK). Finally, the blood gas features enveloped PH, partial pressure of carbon dioxide (PCO2), partial pressure of oxygen (PO2), the concentration of hydrogen carbonate (HCO3), and

bicarbonate (BE). If a patient feature was not available, the mean value of this feature from the dataset in each class was used for imputation.

2.2. Statistical evaluation

One of the methods to evaluate the performance of binary classification algorithms is receiver operating characteristic (ROC) curve (Klawonn et al. 2011). In the ROC diagram, both sensitivity or true positive rate (TPR) and recall or false positive rate (FPR) as indicators for the performance of binary classification algorithms based on logistic regression (LR) are combined and displayed as a curve. The area under the ROC curve (AUC) is also used for the evaluation of the performance of binary classification algorithms based on given input features. AUC, as a very useful and easy-to-use framework, tells how much the model is capable of distinguishing between two classes and seeing the importance of given input features (Mamitsuka 2006). The numerical value of the AUC varies from zero to one, with numbers closer to one meaning the test method has good detection or accuracy. Finally, the Wilcoxon rank-sum test was used to evaluate the significance of the extracted features. The Wilcoxon rank-sum test is a non-parametric test for two groups whose samples are independent of each other (Fay and Proschan 2010). The probability value (p -value) of this test indicates the probability of error in accepting the validity of the observed results. Utilizing this non-parametric analysis is a common method for selecting predictive features for classification algorithms. In this study, to evaluate the performance of binary classification algorithms and select the best features for them, ROC, AUC, and p -value criteria were used.

2.3. Hierarchical feature selection methods

In this study, hierarchical feature selection algorithms were used. First, the statistical significance of the extracted features between dead vs. survived patients were evaluated by the Kruskal-Wallis test and AUC analysis, and insignificant features with AUC lower than 0.6 and $p > 0.005$ were omitted. Afterward, five feature selection methods were used to choose the best combination of features from the remaining features for classification. Forward feature selection (Ververidis and Kotropoulos 2005), minimum Redundancy Maximum Relevance (mRMR) (Peng et al. 2005), Relieff (Liu and Motoda 2007), Linear Discriminant Analysis (LDA) (Subasi and Gürsoy 2010) and Neighborhood Component Analysis (NCA)

(Tuncer et al. 2021) are five feature selection algorithms utilized in this study. In MRMR and Relief, the features are ranked based on their characteristics, while feature selection in Forward feature selection is dependent on the classifier results. The forward feature selection algorithm uses the learning method to evaluate the usefulness of each subset of features and aims to find a subset of features with the least amount of classification error. First, all features are given to the classifier one by one, and the best feature is selected. Then the combination of the first selected feature and the other remaining features are given to the classifier, and the best double feature combination is determined. The framework can be continued to identify informative feature groups of various sizes. The computational load of this method with all possible combinations of features is very high. Common dimensionality reduction methods can reduce the number of features before the implementation of this method (Ververidis and Kotropoulos 2005). Two other utilized feature selection methods were MRMR and Relief-f. These two methods rely on the importance of the general features of the data set and rank them based on the characteristics of the features (Chandrashekar and Sahin 2014). In the Relief-f method, a sample is selected randomly. If the same feature in the chosen sample differs from the similar feature in the neighboring sample of the same class, this feature's score reduces. Also, if the same feature in the chosen sample differs from the similar feature in the opposite class's neighboring sample, the score of this feature increases. The third feature selection method is the MRMR. This method evaluates the features based on the maximum relevance and minimum redundancy. Features with the maximum amount of mutual information between features and class labels have a maximum relevance. Also, features with minimum redundancy are identified based on the principle that if two features are interdependent and one of them is removed, the classification performance will not change much. Finally, after calculating the score values of each feature in these two methods, these features are arranged in descending order to select the best features. The LDA algorithm method aims to map the data to the new space in such a way that the variance between the data class in the new space is maximized and the variance within the class is minimized (Subasi and Gürsoy 2010). Finally, NCA feature selection is used as a non-parametric method with the aim of maximizing prediction accuracy in classification algorithms (Tuncer et al. 2021).

2.4. Classification methods

In this study, random forest (RF) (Breiman 2001), Support Vector Machine (SVM) (Wang 2005), Logistic Regression (LR) (Kleinbaum et al. 2002), bagging, boosting, Artificial neural networks (ANN), and K nearest neighbors (KNN) methods were used to classify patients into two outcome classes. RF is one of the most widely used machine learning algorithms for classification that often provides excellent results (Criminisi et al. 2012). As the RF name implies, this algorithm builds a forest of bagged decision trees in a random manner. The main idea of the bagging method is that a combination of learning models enhances the overall results of the model. In short, the RF merges several trees together to make more accurate and stable predictions. For the final classification, each tree votes for a specific class, and the class with the majority vote wins. It should be noted that each tree is constructed using different data samples and features, which are selected randomly. This leads to a great variety and ultimately a better model. Furthermore, the number of hyperparameters in this method is not high, and they are determined easily. For comparison, SVM and LR methods were also used for classification. SVM is a supervised classifier that determines the best classification and separation of data by benchmarking support vectors. Support vectors are a set of points in the n-dimensional space of data that define the boundaries of categories, and the data is bounded and categorized according to them. In an SVM classifier, when there exist overlapped features, the feature space maps into a higher dimension space by using a kernel function then an optimal discriminant hyperplane is constructed in that space. Common choices for the kernel function are linear, quadratic, cubic, and Gaussian kernels such as Radial Basis Function (RBF). The correct choice of kernels and parameters greatly affects the performance and final result (Kecman et al. 2005; Wang et al. 2007; Yeung et al. 2007). We have used the RBF kernel, which is the most common kernel based on the Euclidean Distance. Also, the RBF kernel has good performance due to the consideration of data distribution. The parameters are set in the optimization process by Bayesian optimization. The availability of advanced kernel function or other structured large margin classifiers are explained more in these references (Wang et al. 2007; Yeung et al. 2007). LR is widely used in all scientific fields, especially medical applications. This model can be considered as a generalized linear model that uses the logit function as a link function, and its error follows a polynomial distribution.

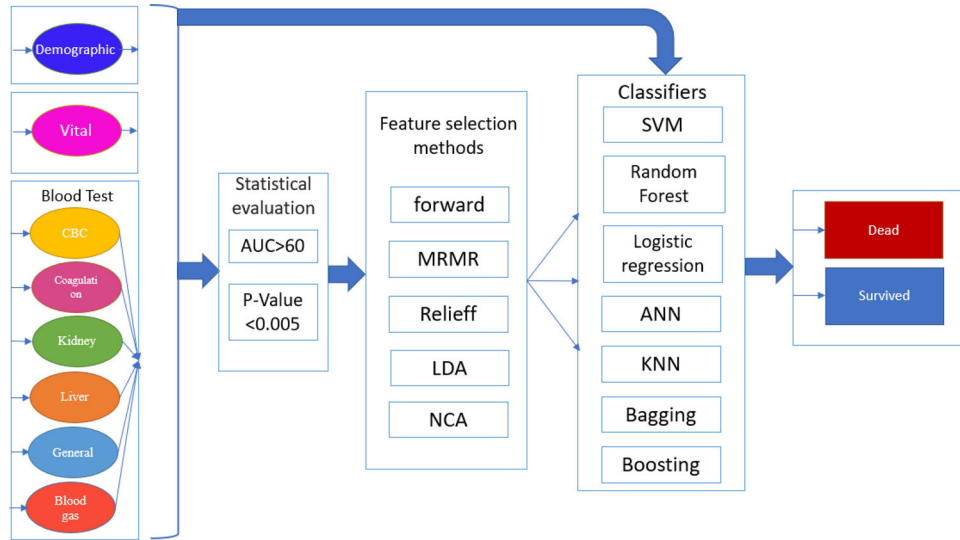


Figure 1. Framework of our proposed method for prediction of COVID-19 outcome.

Finally, we also have used KNN, feed-forward ANN, bagging and boosting classifiers for comparison. To generalize the results of model prediction to an independent data set, we performed 10-fold cross-validation on the patient data. For each run, 90% of data is used as train data, and 10% of data is used as test data. For each model, the accuracy, sensitivity, specificity, F-measure, and Kappa statistics (mean + standard division) were reported on the test data set. To evaluate the classifier results, each method ran ten times, and the mean was reported as final values.

2.5. Overview of the proposed method

Figure 1 shows a block diagram of the proposed method of this study. 41 features from 244 COVID-19 patients were recorded and categorized into eight different groups of features; demographic, vital signs, and six different groups of blood test features (CBC, coagulation, kidney, liver, blood gas, and general). First, to determine which of the 41 extracted features were more informative and also to compare the contribution of individual features for prediction of the mortality outcome, ROC curves were generated, and the amount of AUC and also p -value of these features between patient outcomes groups (survived and death) were calculated. Afterward, machine learning methods were used to provide an accurate mortality prediction system. First, the prediction power of a combination of features in eight groups was compared to determine which of them were more informative for the prediction of outcome situations. Toward this aim, the information of individual features in each of eight categories was combined, and mortality prediction was implemented using three different

classifiers (LR, SVM, RF) separately. By applying different feature groups separately, we try to find the best groups of the feature sets that are suitable to predict the COVID-19 prognosis. For example, one prediction model was utilized solely based on vital features, and another only based on CBC features. In the following step, to increase the prediction accuracy, hierarchical feature selection algorithms were used to find the best combination of different groups of features. First, the statistical significance of the extracted features between dead vs. survived patients was assessed by the Kruskal-Wallis test and AUC analysis. Then features from the original 41 extracted features with AUC above 0.6 and p -value below 0.005 were used as selected features for further analysis steps. Afterward, five feature selection methods were utilized to select the best combination of the 28 features from the previous step for classification. Forward Feature selection, MRMR, Relieff, LDA and NCA were five feature selection algorithms utilized. Finally, seven different advanced classifiers named: RF, SVM, LR, KNN, ANN, bagging, and boosting were used for predicting the prognosis outcome. All processes, including feature selection and classification algorithms, were implemented using MATLAB R2019b on a 4.5 GHz quad-core computer with an NVIDIA GeForce RTX 1660Ti graphics processing unit (GPU).

3. Result

Figure 2 displays the ROC plots of 41 total features in eight groups for the prediction of mortality outcomes. The AUC and p -value of these features are also calculated and shown in Table 1. Features with AUC close

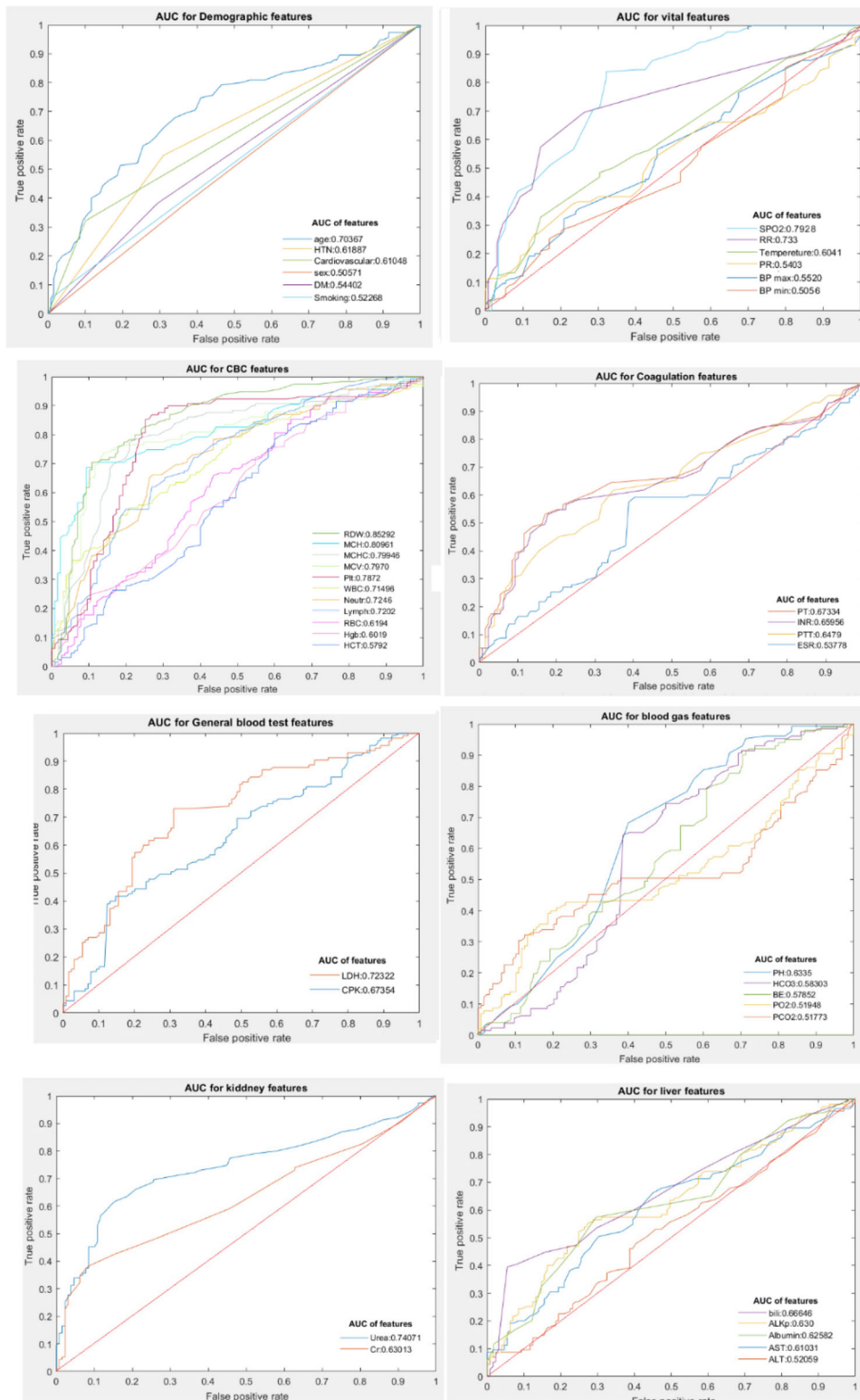


Figure 2. ROC curves of 41 features in eight groups of feature for prediction of mortality outcomes: (a) demographic, (b) vital, (c) CBC, (d) coagulation, (e) general, (f) blood gas, (g) kidney, (h) liver.

to one and a p -value below 0.005 have more contribution in the prediction of the final outcome. From Figure 2 and Table 1, RDW, MCH, MCHC, and MCV in the CBC feature group have the highest AUC with values of 85.29, 80.96, 79.94, and 79.70,

respectively. In the next spot, Spo2 in the vital signs group has a high AUC of 79.28. In the following, machine learning methods are used to provide an accurate mortality prediction system. First, we have investigated and compared the prediction power of a

Table 1. Amount of AUC and p -value of 41 extracted features between outcomes of two groups of COVID-19 patients (survived and death) using LR classifier.

Categories	Features	AUC%	p value
Demographic (6)	age	70.36	<0.005
	sex	50.57	0.855
	HTN	61.88	<0.005
	DM	54.42	0.147
	Cardiovascular	61.04	<0.005
	Smoking	52.26	0.060
Vital (6)	BP MAX	55.20	0.166
	BP MIN	50.56	0.597
	PR	54.03	0.255
	RR	73.30	<0.005
	T	60.41	<0.005
	SPO2	79.28	<0.005
CBC (11)	WBC	71.49	<0.005
	Neutr	72.46	<0.005
	Lymph	72.02	<0.005
	RBC	61.94	<0.005
	Hgb	60.19	<0.005
	HCT	57.92	0.031
	MCV	79.70	<0.005
	MCH	80.96	<0.005
	MCHC	79.94	<0.005
	RDW	85.29	<0.005
	Plt	78.72	<0.005
Coagulation (4)	ESR	53.77	0.425
	PT	67.33	<0.005
	PTT	64.79	<0.005
	INR	65.95	<0.005
Kidney (2)	Urea	74.07	<0.005
	Cr	63.01	<0.005
Liver (5)	AST	61.03	<0.005
	ALT	52.05	0.105
	ALKP	63.00	<0.005
	Bili	66.64	<0.005
	Albumin	62.58	<0.005
General (2)	LDH	72.32	<0.005
	CPK	67.35	<0.005
Blood gas (5)	PH	63.35	<0.005
	PCO2	51.77	0.126
	PO2	51.94	0.046
	HCO3	58.30	0.038
	BE	57.85	0.009

combination of features in eight groups to determine which of them are more informative. The results of classification accuracy, sensitivity, specificity, F-measure, and Kappa statistics (mean + standard division) for a combination of features in the eight mentioned groups using three different classifiers (LR, SVM, RF) are calculated and shown in Table 2. Results of the three classifiers revealed that the combination of features in CBC groups had the highest classification performance with an accuracy of 82.91 ± 5.53 via the SVM classifier, 79.79 ± 6.13 via the RF classifier, and 78.83 ± 3.57 via the LR classifier. After CBC, vital signs group had the highest classification performance with an accuracy of 75.83 ± 5.30 with the SVM classifier, 75.00 ± 3.67 with the RF classifier, and 72.62 ± 3.10 with the LR classifier.

Afterward, for increasing the prediction accuracy of the outcome of COVID-19 patients (survived and

death), hierarchical feature selection algorithms were used to find the best combination of different groups of features. First, 28 features from 41 extracted features that had AUC above 0.6 and p -value below 0.005 were used as selected features for further analysis. These 28 features were age, HTN, cardiovascular in demographic features; RR, T, Spo2 in vital features; WBC, Neutr, Lymph, RBC, Hgb, MCV, MCH, MCHC, RDW, Plt in CBC features; PT, PTT, INR in coagulation features; Urea and Cr in kidney features; AST, ALKP, Bili, Albumin in liver features; LDH, CPK in general features; and finally, PH in blood gas features. Box plots of these features with AUC > 0.6 and p -value < 0.005 for the two outcome classes are shown in Figure 3. Five feature selection methods, Forward Feature selection, MRMR, Relieff, LDA, and NCA were used to select the best combination of features. Finally, seven different classifiers, RF, SVM, LR, KNN, ANN, bagging, and boosting were used for predicting the prognosis outcome of COVID-19 patients. The results of classification accuracy, sensitivity, specificity, F-measure, and Kappa statistics (mean + standard division) of the seven different classifiers with different hierarchical feature selection algorithms (AUC > 60 and then Forward Feature selection, MRMR, Relieff, LDA, and NCA) are calculated and shown in Table 3. For comparison, the results of classification parameters with a combination of all features (Yeung et al. 2007) and features with AUC > 0.6 (28 features) with the seven classifiers are also calculated and shown in Table 3. It was revealed that RF classifier with hierarchical feature selection algorithms using features with AUC > 60 and then Forward Feature selection had the highest classification performance for COVID-19 outcome prediction with an accuracy of 92.08 ± 2.56 . The hierarchical features selection method increased the accuracy for the prediction of mortality outcomes from an accuracy of 82.91 ± 5.53 using the best groups of features with CBC groups to 92.08 ± 2.56 using the best combination of features. For selecting the best features in the best-mentioned method, the box plots for ten iterations run from one feature to thirteen features that were selected based on Forward feature selection and RF classifier are shown in Figure 4. It is shown in this figure that as the number of selected features increases, the performance of the RF classifier improves. The optimal number is where adding more features does not significantly improve the result. For best performance, the best-selected features was set to nine features which were: age in demographic features; RR, and Spo2 in vital features; Neutr, Lymph

Table 2. The results of classification accuracy, sensitivity, specificity, F1 score and Kappa (mean + standard division) of different classifiers (LR, SVM, RF) for combination of features in every of eight mentioned groups (demographic and patient history features, vital signs, CBC, coagulation, kidney, liver, blood gas and general), separately.

Run for 10 times						
	Classifier and feature selection method	Accuracy (mean \pm SD)	Sensitivity (mean \pm SD)	Specificity (mean \pm SD)	F1 score (mean \pm SD)	Kappa (mean)
1	SVM for demographic features	59.37 \pm 4.30	48.26 \pm 1.86	69.60 \pm 1.36	57.37 \pm 7.19	0.57
2	SVM for vital features	75.83 \pm 5.30	65.65 \pm 9.49	85.20 \pm 9.80	74.34 \pm 3.64	0.71
3	SVM for CBC features	82.91 \pm 5.53	77.39 \pm 6.41	88.00 \pm 7.54	81.91 \pm 7.26	0.80
4	SVM for coagulation features	68.54 \pm 4.75	46.95 \pm 8.88	88.40 \pm 7.64	66.77 \pm 5.45	0.65
5	SVM for kidney features	71.04 \pm 6.08	50.04 \pm 1.28	90.40 \pm 4.69	69.10 \pm 1.56	0.68
6	SVM for liver features	66.25 \pm 5.36	49.13 \pm 1.08	82.00 \pm 7.60	64.55 \pm 1.11	0.63
7	SVM for general features	63.12 \pm 8.89	34.34 \pm 2.63	89.60 \pm 1.01	61.03 \pm 1.03	0.60
8	SVM for blood gas features	66.45 \pm 5.41	40.86 \pm 1.28	90.00 \pm 5.07	64.02 \pm 1.02	0.63
1	RF for demographic features	62.29 \pm 5.59	57.82 \pm 7.40	66.40 \pm 8.68	60.81 \pm 4.65	0.59
2	RF for vital features	75.00 \pm 3.67	75.21 \pm 5.04	74.80 \pm 6.81	74.35 \pm 4.98	0.72
3	RF for CBC features	79.79 \pm 6.13	81.73 \pm 1.13	78.00 \pm 8.48	79.39 \pm 4.46	0.78
4	RF for coagulation features	69.16 \pm 3.37	66.95 \pm 6.54	71.20 \pm 1.12	68.32 \pm 4.54	0.66
5	RF for kidney features	63.95 \pm 7.41	58.87 \pm 7.38	69.27 \pm 1.78	61.43 \pm 4.08	0.61
6	RF for liver features	64.79 \pm 3.98	63.47 \pm 1.25	66.00 \pm 6.86	63.64 \pm 3.83	0.61
7	RF for general features	65.41 \pm 4.07	69.13 \pm 5.59	62.00 \pm 7.11	64.78 \pm 4.69	0.62
8	RF for blood gas features	60.23 \pm 7.19	63.47 \pm 7.72	57.20 \pm 8.65	59.39 \pm 2.38	0.58
1	LR for demographic features	64.11 \pm 4.75	53.91 \pm 5.49	69.39 \pm 5.54	63.79 \pm 3.61	0.62
2	LR for vital features	72.62 \pm 3.10	72.79 \pm 4.27	73.92 \pm 6.34	72.15 \pm 4.48	0.70
3	LR for CBC features	78.83 \pm 3.57	79.79 \pm 3.36	76.70 \pm 2.54	78.76 \pm 6.68	0.78
4	LR for coagulation features	69.06 \pm 6.02	65.20 \pm 5.26	66.72 \pm 4.40	66.13 \pm 4.91	0.66
5	LR for kidney features	66.44 \pm 4.47	65.47 \pm 4.54	63.87 \pm 7.38	65.94 \pm 3.58	0.63
6	LR for liver features	64.76 \pm 5.48	65.28 \pm 5.62	63.28 \pm 6.25	64.29 \pm 4.85	0.63
7	LR for general features	61.93 \pm 4.63	62.84 \pm 4.37	63.06 \pm 6.59	61.36 \pm 5.42	0.60
8	LR for blood gas features	60.65 \pm 6.19	61.06 \pm 6.32	58.29 \pm 6.38	60.17 \pm 2.79	0.60

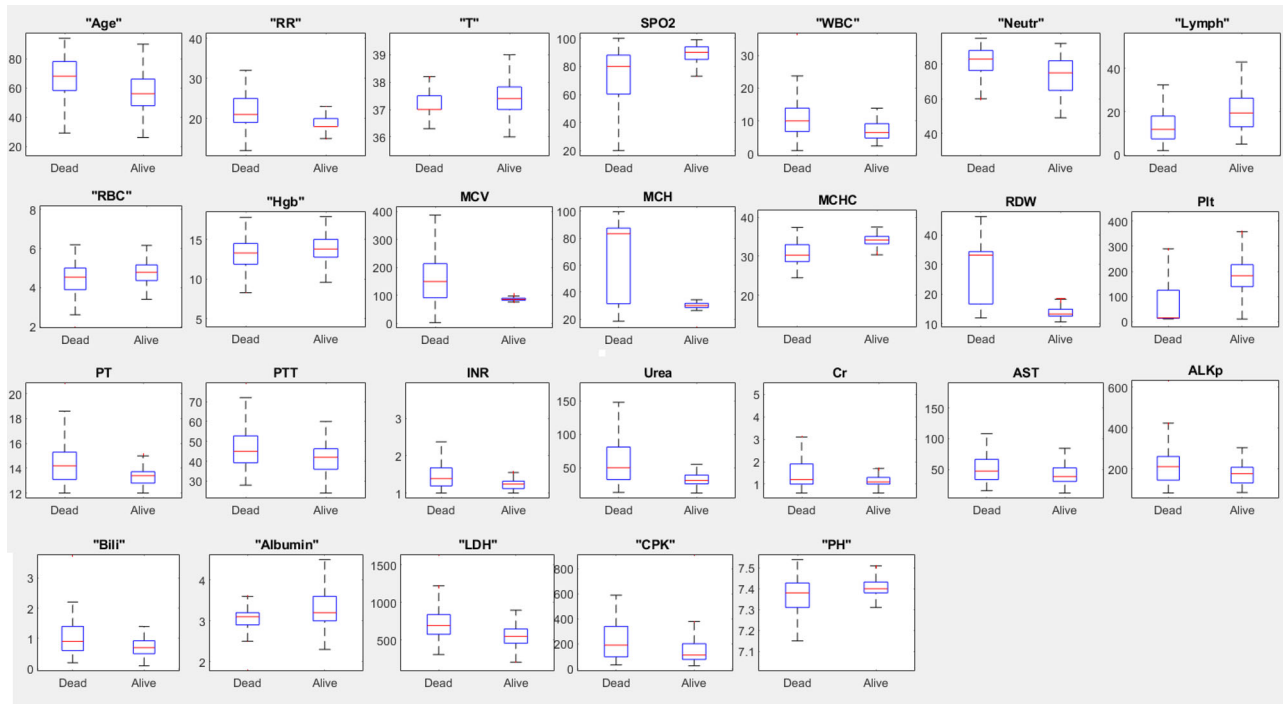


Figure 3. Box plot of features with AUC > 0.6 and p-value < 0.005 for two outcomes of COVID-19 patients (survived and death).

and MCH in CBC features; INR and PTT in coagulation features; and finally LDH in general features. The number of trees in RF is also a structural parameter and should be considered in an optimization process in which, for best performance, the best value was set to 30 trees. Consequently, our

proposed assessment system based on a hierarchical feature selection and RF classifier with an accuracy of 92.08% can be confidently used as an assistant prognostic triage tool for physicians to predict the patient outcome situation and sieve patients with high mortality risks.

Table 3. The results of classification accuracy, sensitivity specificity, F1 score and Kappa (mean + standard division) of different classifiers (LR, SVM, RF, ANN, KNN, Bagging, Boosting) with different hierarchical feature selection algorithms (Forward Feature selection, MRMR and Relieff, LDA, NCA).

Run for 10 times:						
	Classifier and feature selection method	Accuracy (mean \pm SD)	Sensitivity (mean \pm SD)	Specificity (mean \pm SD)	F1 score (mean \pm SD)	Kappa (mean)
1	SVM for all 41 features	80.20 \pm 4.83	75.65 \pm 1.04	84.40 \pm 5.48	80.40 \pm 2.39	0.78
2	SVM for all 28 features with AUC > 60	84.37 \pm 4.41	84.23 \pm 4.40	85.18 \pm 5.63	84.92 \pm 3.44	0.81
3	SVM with hierarchical feature selection (AUC > 60 and then forward feature selection)	89.58 \pm 2.94	89.56 \pm 2.96	90.54 \pm 3.94	89.98 \pm 2.32	0.87
4	SVM with hierarchical feature selection (AUC > 60 and then MRMR)	89.10 \pm 3.22	89.11 \pm 3.32	88.16 \pm 5.27	89.40 \pm 3.17	0.86
5	SVM with hierarchical feature selection (AUC > 60 and then Relieff)	88.04 \pm 3.93	88.04 \pm 4.04	87.12 \pm 4.23	87.40 \pm 2.76	0.88
6	SVM with hierarchical feature selection (AUC > 60 and then LDA)	89.33 \pm 4.14	88.21 \pm 3.42	90.44 \pm 4.16	89.04 \pm 4.57	0.86
7	SVM with hierarchical feature selection (AUC > 60 and then NCA)	87.26 \pm 4.64	85.37 \pm 5.28	88.46 \pm 2.78	87.29 \pm 3.15	0.85
1	RF for all 41 features	88.75 \pm 2.97	88.25 \pm 5.35	88.25 \pm 5.35	88.57 \pm 3.42	0.86
2	RF for all 28 features with AUC > 60	88.12 \pm 5.42	88.25 \pm 5.35	89.08 \pm 5.49	88.93 \pm 3.13	0.83
3	RF with hierarchical feature selection (AUC > 60 and then forward feature selection)	92.08 \pm 2.56	91.90 \pm 2.67	91.31 \pm 2.61	92.37 \pm 2.48	0.91
4	RF with hierarchical feature selection (AUC > 60 and then MRMR)	91.66 \pm 3.64	91.70 \pm 3.62	90.65 \pm 3.65	90.90 \pm 2.76	0.90
5	RF with hierarchical feature selection (AUC > 60 and then Relieff)	89.37 \pm 2.49	89.38 \pm 4.47	88.43 \pm 3.57	89.49 \pm 2.58	0.89
6	RF with hierarchical feature selection (AUC > 60 and then LDA)	87.50 \pm 4.28	89.56 \pm 7.44	85.60 \pm 5.71	87.22 \pm 4.59	0.86
7	RF with hierarchical feature selection (AUC > 60 and then NCA)	88.18 \pm 5.37	88.61 \pm 4.95	88.73 \pm 5.84	89.64 \pm 5.82	0.87
1	LR for all 41 features	87.95 \pm 4.61	87.20 \pm 2.74	87.25 \pm 5.35	88.27 \pm 3.32	0.86
2	LR for all 28 features with AUC > 60	88.95 \pm 4.61	88.86 \pm 4.74	87.85 \pm 4.72	87.98 \pm 2.95	0.88
3	LR with hierarchical feature selection (AUC > 60 and then forward feature selection)	90.20 \pm 3.41	90.14 \pm 3.47	89.16 \pm 3.43	89.31 \pm 2.46	0.90
4	LR with hierarchical feature selection (AUC > 60 and then MRMR)	89.16 \pm 4.25	89.14 \pm 4.18	90.10 \pm 4.35	89.25 \pm 3.62	0.89
5	LR with hierarchical feature selection (AUC > 60 and then Relieff)	89.58 \pm 3.67	89.49 \pm 3.79	88.51 \pm 4.73	89.79 \pm 2.85	0.88
6	LR with hierarchical feature selection (AUC > 60 and then LDA)	90.29 \pm 3.93	89.13 \pm 5.88	90.66 \pm 4.38	90.66 \pm 4.75	0.89
7	LR with hierarchical feature selection (AUC > 60 and then NCA)	88.36 \pm 4.64	89.39 \pm 4.12	87.41 \pm 6.73	88.84 \pm 3.90	0.87
1	ANN for all 41 features	59.54 \pm 4.68	60.71 \pm 5.13	60.44 \pm 3.96	59.68 \pm 4.75	0.58
2	ANN for all 28 features with AUC > 60	65.72 \pm 4.14	71.09 \pm 3.27	67.36 \pm 4.84	65.53 \pm 6.16	0.65
3	ANN with hierarchical feature selection (AUC > 60 and then forward feature selection)	67.48 \pm 6.38	74.93 \pm 5.74	69.28 \pm 3.26	67.37 \pm 5.98	0.67
4	ANN with hierarchical feature selection (AUC > 60 and then MRMR)	65.45 \pm 8.65	69.23 \pm 7.67	68.47 \pm 4.94	65.36 \pm 5.74	0.64
5	ANN with hierarchical feature selection (AUC > 60 and then Relieff)	66.68 \pm 3.77	64.23 \pm 6.36	67.38 \pm 3.33	65.64 \pm 4.28	0.63
6	ANN with hierarchical feature selection (AUC > 60 and then LDA)	64.26 \pm 4.39	65.46 \pm 4.83	69.71 \pm 4.79	64.27 \pm 3.46	0.62
7	ANN with hierarchical feature selection (AUC > 60 and then NCA)	64.68 \pm 6.65	63.65 \pm 4.48	66.62 \pm 5.32	64.22 \pm 5.49	0.62
1	KNN for all 41 features	83.41 \pm 4.05	82.60 \pm 7.16	84.00 \pm 5.97	82.92 \pm 4.80	0.80
2	KNN for all 28 features with AUC > 60	85.43 \pm 4.30	84.31 \pm 5.72	85.00 \pm 5.48	85.93 \pm 6.38	0.83
3	KNN with hierarchical feature selection (AUC > 60 and then forward feature selection)	87.50 \pm 5.53	86.95 \pm 8.56	88.00 \pm 6.65	86.63 \pm 5.90	0.86
4	KNN with hierarchical feature selection (AUC > 60 and then MRMR)	78.33 \pm 5.03	79.56 \pm 7.95	77.20 \pm 7.55	77.78 \pm 4.48	0.75
5	KNN with hierarchical feature selection (AUC > 60 and then Relieff)	76.04 \pm 2.02	73.91 \pm 8.76	83.60 \pm 5.82	75.55 \pm 3.86	0.71
6	KNN with hierarchical feature selection (AUC > 60 and then LDA)	88.33 \pm 3.82	89.65 \pm 6.16	86.80 \pm 6.53	88.13 \pm 3.60	0.85
7	KNN with hierarchical feature selection (AUC > 60 and then NCA)	75.62 \pm 4.63	77.39 \pm 6.47	78.63 \pm 5.19	76.63 \pm 3.45	0.73
1	Bagging for all 41 features	71.87 \pm 3.43	65.65 \pm 43.23	77.60 \pm 6.02	70.11 \pm 3.41	0.68
2	Bagging for all 28 features with AUC > 60	87.08 \pm 6.42	90.30 \pm 6.79	83.20 \pm 9.94	87.24 \pm 6.04	0.83
3	Bagging with hierarchical feature selection (AUC > 60 and then forward feature selection)	85.83 \pm 6.86	88.69 \pm 8.24	83.20 \pm 6.47	85.66 \pm 7.04	0.79
4	Bagging with hierarchical feature selection (AUC > 60 and then MRMR)	74.58 \pm 60.38	73.91 \pm 8.92	75.20 \pm 7.72	73.42 \pm 6.83	0.71
5	Bagging with hierarchical feature selection (AUC > 60 and then Relieff)	72.91 \pm 5.19	72.86 \pm 4.37	72.00 \pm 6.76	71.34 \pm 4.96	0.69
6	Bagging with hierarchical feature selection (AUC > 60 and then LDA)	77.91 \pm 66.75	79.13 \pm 7.33	76.80 \pm 8.59	77.46 \pm 6.69	0.76
7	Bagging with hierarchical feature selection (AUC > 60 and then NCA)	77.08 \pm 4.62	73.91 \pm 3.95	80.62 \pm 5.93	76.55 \pm 6.81	0.76
1	Boosting for all 41 features	81.25 \pm 3.47	86.96 \pm 4.38	76.00 \pm 6.75	81.63 \pm 2.94	0.78
2	Boosting for all 28 features with AUC > 60	87.91 \pm 3.22	87.39 \pm 6.30	88.40 \pm 3.97	87.32 \pm 3.49	0.83
3	Boosting with hierarchical feature selection (AUC > 60 and then forward feature selection)	89.16 \pm 4.89	88.69 \pm 6.54	89.75 \pm 4.91	88.60 \pm 8.68	0.87
4	Boosting with hierarchical feature selection (AUC > 60 and then MRMR)	84.16 \pm 7.49	81.73 \pm 1.04	86.40 \pm 7.35	83.03 \pm 8.30	0.81
5	Boosting with hierarchical feature selection (AUC > 60 and then Relieff)	81.87 \pm 6.51	80.43 \pm 9.22	83.23 \pm 8.28	80.96 \pm 6.35	0.79
6	Boosting with hierarchical feature selection (AUC > 60 and then LDA)	82.70 \pm 6.44	82.60 \pm 6.14	82.80 \pm 8.23	82.14 \pm 6.29	0.8
7	Boosting with hierarchical feature selection (AUC > 60 and then NCA)	81.25 \pm 5.14	78.26 \pm 8.50	80.05 \pm 5.20	80.65 \pm 6.38	0.78

4. Discussion

This study investigated the prediction power of demographic, vital characteristics, and blood test results that are routinely collected during the first day of patients' admission to predict the outcome situation of COVID-19 patients. We have compared several feature categories, and the results revealed that the combination of features in CBC and then vital sign groups had the highest classification performance. Furthermore, a robust machine learning approach based on hierarchical feature selection algorithms and RF classifier revealed that a combination of nine features, age, RR, Spo2, Neutr, Lymph, MCH, INR, PTT, and LDH, had the highest accuracy of 92.08 ± 2.56 to predict COVID-19 mortality outcome. Therefore, our automatic proposed model can be used as a valuable assistant prognostic tool to avoid missing high-risk patients and help provide on-time hospitalization and intensive care services to reduce mortality rates in Iranian patients.

An important weakness of routine rapid triage in a pandemic situation is the increased mortality rate due to missing high-risk patients (Liang et al. 2020). These patients might incorrectly be identified as mild and, without further workup, be advised to take a home-treatment approach. The disease has an unpredictable trajectory where the condition of some patients suddenly becomes critical, surprising even the most skilled physicians; this hampers physicians' performance by limiting their action time window. Furthermore, it has been shown that patients who later become critically ill carry significantly more viral loads even before their condition becomes critical (Siordia 2020). Thus, rapid isolation of high-risk patients is required to decrease infection spread. Our model using mentioned features could alleviate these problems by providing fast and accurate prognosis prediction after the first day of patient admission to support proper resource allocation and decision making and consequently avoid missing high-risk patients.

We have used widely known feature selection algorithms named Forward feature selection, mRMR and Relieff. The results displayed that the wrapper method, here forward feature selection, which includes a learning algorithm as a black box and uses its predictive performance to evaluate the usefulness of a subset of features, yielded better classification results. In wrapper methods, the feature selection part is combined with the training of a classifier and cannot be separated (Chandrashekar and Sahin 2014). This method uses a classifier during the feature selection phase, and this leads to better performance

results. On the other hand, filter methods, such as Relieff and mRMR, only consider the relevance of features with dependent classes using statistical measures (Chandrashekar and Sahin 2014). These methods rely on the importance of the general features and rank them based on the characteristics of the features and do not give the best combination of features. For these two groups of methods, we increased the number of features from highest rank to lowest rank until the best accuracy was achieved, while the best combination of features might be something else. Consequently, the results revealed that Forward feature selection, a wrapper method, led to better performance results compared with Relieff and mRMR filter methods.

The seven traditional classifiers algorithms used in the current study, RF, SVM, LR, KNN, ANN, bagging, and boosting are popular and have been among the most successful classification methods in the previous studies in various fields (Shalbfaf et al. 2020; Valizadeh et al. 2021). The advantages of these widely used methods, such as high performance, great classification rate, and high-speed calculation, make them applicable for medical classification problems. Specifically, the RF classifier has several superiorities compared to other statistical classifiers, namely the ability to model complex interactions, the participation of different predictors, higher generalization due to the inherent randomness in the feature and sample selection procedure in each iteration, and interpretability.

The best-selected features were nine features; age in demographic features; RR and Spo2 in vital features; Neutr, Lymph, and MCH in CBC features; INR and PTT in coagulation features; and finally LDH in general features. Older age is associated with more infection susceptibility and an atypical response to viral pathogens due to reduced expression of type I interferon-beta (Smits et al. 2010). Furthermore, age-related impairment of lymphocyte function along with abnormal expression of cytokines leads to prolonged pro-inflammatory responses and weakens the host's response to viral infection and inflammation control (Opal et al. 2005). SPO2 and RR were also significantly associated with patient mortality in this study. Dyspnea is not prominent in the initial stages of COVID-19. This is because carbon dioxide exchange is still present through alveoli at these stages. However, the oxygen exchange is distorted due to the alveolar collapse. This hypoxia leads to the progression of pneumonia in the absence of clinical symptoms (Teo 2020). Furthermore, hypoxia promotes the activity of the local inflammatory system

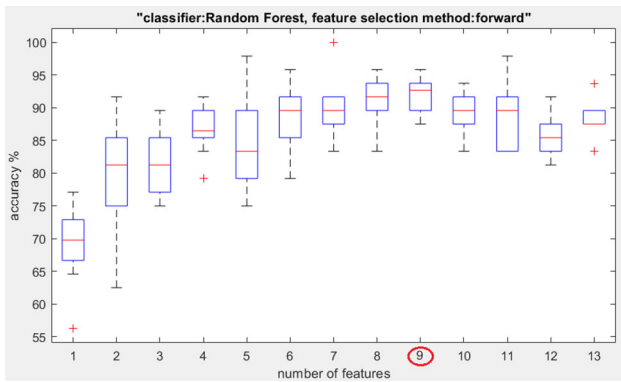


Figure 4. Box plot for 10 time run from one feature to thirteen features based forward feature selection method and RF classifier.

causing further damage to the lung tissue and higher hypoxia (Xie et al. 2020a). Another prominent association with patient COVID-19 mortality in this study was CBC results, specially Neutr, Lymph along with MCH. Previous studies have reported that platelets are affected through several possible routes in patients and that the rates of thrombocytopenia are higher in severe patients. COVID-19 patients are at an increased risk of thrombosis formation (Ratajczak and Kucia 2020); Furthermore, cytokine storm and other immune factors can destroy platelets (Humbert et al. 2020) or decrease their production (Xu et al. 2020). It also has been proposed that SARS-COV-2 can also distort the beta chain of hemoglobin, reducing its level and RBC population (Liu and Li 2020). White blood cells are also affected by COVID-19 infection. Lymphopenia is associated with a poor prognosis. The virus can kill lymphocytes via direct apoptosis or indirectly by cytokine storm (Tan et al. 2020). Furthermore, a reduction in the neutrophil population can be observed in severe patients that might be due to inflammation or secondary infection (Huang et al. 2020). INR and PTT in coagulation features are also significantly associated with patient mortality. The inflammatory response promoted by severe COVID-19 infection could cause endothelial damage, distortion of the coagulation cascade function, and coagulopathy. Therefore, levels of INR and PTT, as coagulation biomarkers, during COVID-19 infection can be informative of coagulopathy progression and disease severity (Connors and Levy 2020). The higher INR and PTT mean fibrin degradation products are increased and lead to life-threatening intravascular coagulation, necessitating rapid intervention (Terpos et al. 2020). Finally, elevated levels of LDH could reflect tissue injury caused by COVID-19 infection and concurrent lung fibrosis. Indeed, abnormal LDH

is commonly seen in idiopathic lung fibrosis (Yan et al. 2020; Giacobbe 2020). Furthermore, a robust immune response to COVID-19 infection and subsequent cytokine storm could cause multi-organ damage, which causes a further rise in the LDH level (Yang et al. 2020a).

The proposed machine learning system has some advantages compared with other studies. As opposed to imaging procedures such as CT scans, proposed clinical tests are common in medical centers, and exposure to X-ray is not needed. Furthermore, they are significantly cheaper, and their results are provided in a relatively short time. Vital signs and demographic data can be easily measured by simple questions and devices. These laboratory and non-laboratory variables can be suitable candidates for early triage of patients, especially in crowded regions or emergencies, such as pandemics. Another important point is that we have evaluated different groups of clinical features from each patient simultaneously, giving an advantage to our model from the perspective of available data in each feature. Finally, the RF classifier, which is used in this study, has several superiorities compared with other classifiers due to utilizing several decision trees built by a selected subset of features which improve the robustness of the classifier significantly.

Several limitations were present in this study. First, the study was conducted with a retrospective framework. Furthermore, several biomarkers, such as CRP, which are potentially linked to COVID-19 mortality, were not included due to the high rate of missing values resulting from the pandemic situation. The presence of other acute phase reactants and inflammatory markers, such as ESR, plt, and LDH, in this model, is likely to explain a significant portion of information explained by CRP. Moreover, patients in this study were admitted to the Masih Daneshvari hospital, which is a primary care center for COVID-19 patients. Thus, more severe patients were admitted, and the ratio of expired and discharged patients does not accurately mirror the whole society. Furthermore, the study population was relatively small and only contained Iranian patients. Finally, laboratory data requires invasive sampling, and many smaller health centers do not have access to laboratory equipment (Plebani 2015; Reeve and Twomey 2020).

Future studies should include larger and more diverse COVID-19 patient study groups to further investigate the utilization of machine learning models in clinical settings. Continuous data input from various hospitals could be used to develop and

incrementally train an online learning model to predict the prognosis of COVID-19 patients, giving increasingly precise and updated results to be used in clinical settings. Moreover, a more comprehensive archive of a person during hospitalization, from the first day to the last day of hospitalization, can be used for better prediction. Also, in this study, binary outcome (discharged and expired) was used for COVID-19 patients. Future studies can explore other outcomes, such as whether a patient was intubated or admitted to ICU as outcomes. Moreover, to devise specific prognostic models, future studies can focus on individual groups of comorbidities (e.g. cardiovascular) to develop separate models.

5. Conclusion

We have proposed a robust machine learning pipeline to predict the outcome of COVID-19 patients based on demographics, vital signs and CBC, coagulation, kidney, liver, blood gas, and general blood tests. We have compared eight different groups of features separately, and results revealed that the combination of features in CBC groups and then vital signs had the highest classification parameters with an accuracy of 82.91 ± 5.53 and 75.83 ± 5.30 , respectively. Finally, to find the best combination of different groups of features for increasing the prediction accuracy, a hierarchical feature selection algorithms using features which have AUC above 0.6 and p -value below 0.005 and then forward feature selection and finally an advanced classifier named RF have used with an accuracy of 92.08 ± 2.56 to classify COVID-19 patients as death or survived. The best-selected features are set to nine features which are: age in demographic features; RR, and Spo2 in vital features; Neutr, Lymph, and MCH in CBC features; INR and PTT in coagulation features; and finally LDH in general features. So, our proposed model can be confidently used as a valuable assistant diagnostic tool for the physician to predict the patient outcome of COVID-19 with high accuracy for clinical decision making and resource allocation and finally to sieve patients with high mortality risks.

Disclosure statement

The authors declare that they have no conflict of interest.

Ethical approval

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and

with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent

Informed consent was obtained from all individual participants included in the study.

Funding

The article is supported by Iran National Science Foundation: INSF (99009215) and Research Department of School of Medicine Shahid Beheshti University of Medical Sciences (grant no. 23255).

References

- Amini N, Shalbaf A. 2022. Automatic classification of severity of COVID 19 patients using texture feature and random forest based on computed tomography images. *Int J Imaging Syst Tech.* 32(1):102–110.
- Bai X, Fang C, Zhou Y, Bai S, Liu Z, Chen Q, Xu Y, Xia L, Gong S, Xie X, et al. 2020. Predicting COVID-19 malignant progression with AI techniques. *medRxiv*.
- Batlles PC, Cerdá-Alberich L, Fonfría-Esparcia C, Carreres-Ortega A, Muñoz-Núñez CF, Trilles-Olaso L, Martí-Bonmatí L. 2022. Development of severity and mortality prediction models for COVID-19 patients at emergency department including the chest X-ray. *Radiologia (English Edition)*.
- Breiman L. 2001. Random forests. *Mach Learn.* 45(1):5–32.
- Chandrashekar G, Sahin F. 2014. A survey on feature selection methods. *Comput Electr Eng.* 40(1):16–28.
- Chao H, Fang X, Zhang J, Homayounieh F, Arru CD, Digumarthy SR, Babaei R, Mobin HK, Mohseni I, Saba L, et al. 2021. Integrative analysis for COVID-19 patient outcome prediction. *Med Image Anal.* 67:101844.
- Chassagnon G, Vakalopoulou M, Battistella E, Christodoulidis S, Hoang-Thi T-N, Dangeard S, Deutsch E, Andre F, Guillo E, Halm N, et al. 2021. AI-driven quantification, staging and outcome prediction of COVID-19 pneumonia. *Med Image Anal.* 67:101860.
- Connors M, Levy JH. 2020. Thromboinflammation and the hypercoagulability of COVID-19. *J Thromb Haemost.* 18(7):1559–1561.
- Criminisi A, Shotton J, Konukoglu E. 2012. Decision forests: a unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning. *Found Trends Comput Graph Vis.* 7:81–227.
- de Moraes Batista AF, Miraglia JL, Donato THR, Chiavegatto Filho ADP. 2020. COVID-19 diagnosis prediction in emergency care patients: a machine learning approach. *medRxiv*.
- Du M, Zhao J, Yin X, Zhang N, Zheng G. 2020. The impact of vital signs on the death of patients with new coronavirus pneumonia: a systematic review and meta-analysis. *medRxiv*.
- Fay MP, Proschan MA. 2010. Wilcoxon-Mann-Whitney or t -test? On assumptions for hypothesis tests and multiple interpretations of decision rules. *Stat Surv.* 4:1–39.

- Francone M, Iafrate F, Masci GM, Coco S, Cilia F, Manganaro L, Panebianco V, Andreoli C, Colaiacomo MC, Zingaropoli MA, et al. 2020. Chest ct score in covid-19 patients: correlation with disease severity and short-term prognosis. *Eur Radiol*. 30(12):6808–6810.
- Ghosh B, Kumar N, Singh N, Sadhu AK, Ghosh N, Mitra P, Chatterjee J. 2020. A quantitative lung computed tomography image feature for multi-center severity assessment of COVID-19. *medRxiv*.
- Giacobbe DR. 2020. Clinical interpretation of an interpretable prognostic model for patients with COVID-19. *Nat Mach Intell*. 3(1):16.
- <https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-with-medical-conditions.html>
- Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Zhang L, Fan G, Xu J, Gu X, et al. 2020. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet*. 395(10223):497–506.
- Humbert S, Razanamahery J, Payet-Revest C, Bouiller K, Chirouze C. 2020. COVID-19 as a cause of immune thrombocytopenia. *Med Mal Infect*. 50(5):459–460.
- Kecman V, Huang T-M, Vogt M. 2005. Iterative single data algorithm for training kernel machines from huge data sets: theory and performance. In: Wang L, editor. *Support vector machines: theory and applications*. Berlin: Springer; p. 255–274.
- Klawonn F, Höppner F, May S, editors. 2011. *An alternative to ROC and AUC analysis of classifiers*. International Symposium on Intelligent Data Analysis; Springer.
- Kleinbaum DG, Dietz K, Gail M, Klein M, Klein M. 2002. *Logistic regression*. New York: Springer.
- Lee SG, Fralick M, Sholzberg M. 2020. Coagulopathy associated with COVID-19. *Can Med Assoc J*. 192(21):E583.
- Li Q, Cao Y, Chen L, Wu D, Yu J, Wang H, He W, Chen L, Dong F, Chen W, et al. 2020. Hematological features of persons with COVID-19. *Leukemia*. 34(8):2163–2172.
- Liang W, Yao J, Chen A, Lv Q, Zanin M, Liu J, Wong SS, Li Y, Lu J, Liang H, et al. 2020. Early triage of critically ill COVID-19 patients using deep learning. *Nat Commun*. 11(1):1–7.
- Liu H, Motoda H. 2007. *Computational methods of feature selection*. Boca Raton, FL: CRC Press.
- Liu W, Li H. 2020. COVID-19: attacks the 1-beta chain of hemoglobin and captures the porphyrin to inhibit heme metabolism. *ChemRxiv*. Cambridge: Cambridge Open Engage.
- Mamitsuka H. 2006. Selecting features in microarray classification using ROC curves. *Pattern Recognit*. 39(12):2393–2404.
- Matos J, Paparo F, Mussetto I, Bacigalupo L, Veneziano A, Bernardi SP, Biscaldi E, Melani E, Antonucci G, Cremonesi P, et al. 2020. Evaluation of novel coronavirus disease (covid-19) using quantitative lung ct and clinical data: prediction of short-term outcome. *Eur Radiol Exp*. 4(1):1–10.
- Mei X, Lee H-C, Diao K-y, Huang M, Lin B, Liu C, Xie Z, Ma Y, Robson PM, Chung M, et al. 2020. Artificial intelligence-enabled rapid diagnosis of patients with COVID-19. *Nat Med*. 26(8):1224–1225.
- Opal M, Girard TD, Ely EW. 2005. The immunopathogenesis of sepsis in elderly patients. *Clin Infect Dis*. 41(Supplement 7):S504–S512.
- Pan D, Cheng D, Cao Y, Hu C, Zou F, Yu W, Xu T. 2020. A predicting nomogram for mortality in patients with COVID-19. *Front Public Health*. 8:461.
- Peng H, Long F, Ding C. 2005. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans Pattern Anal Mach Intell*. 27(8):1226–1238.
- Plebani M. 2015. Diagnostic errors and laboratory medicine—causes and strategies. *eJIFCC*. 26(1):7–14.
- Pourbagheri-Sigaroodi A, Bashash D, Fateh F, Abolghasemi H. 2020. Laboratory findings in COVID-19 diagnosis and prognosis. *Clin Chim Acta*. 510:475–482.
- Quah P, Li A, Phua J. 2020. Mortality rates of patients with COVID-19 in the intensive care unit: a systematic review of the emerging literature. *Crit Care*. 24(1):1–4.
- Ratajczak MZ, Kucia MJL. 2020. SARS-CoV-2 infection and overactivation of Nlrp3 inflammasome as a trigger of cytokine ‘storm’ and risk factor for damage of hematopoietic stem cells. *Leukemia*. 34(7):1726–1729.
- Reeve JL, Twomey J. 2020. Consider laboratory aspects in developing patient prediction models. *Nat Mach Intell*. 3(1):18–11.
- Sankaranarayanan S, Balan J, Walsh JR, Wu Y, Minnich S, Piazza A, Osborne C, Oliver GR, Lesko J, Bates KL, et al. 2021. Covid-19 mortality prediction from deep learning in a large multistate electronic health record and laboratory information system data set: algorithm development and validation. *J Med Internet Res*. 23(9):e30157.
- Shalbf A, Shalbf R, Saffar M, Sleight J. 2020. Monitoring the level of hypnosis using a hierarchical SVM system. *J Clin Monit Comput*. 34(2):331–338.
- Shan F, Gao Y, Wang J, Shi W, Shi N, Han M, Xue Z, Shen D, Shi Y. 2020. Lung infection quantification of COVID-19 in CT images with deep learning. *arXiv Preprint arXiv:200304655*.
- Shoeibi A, Khodatars M, Alizadehsani R, Ghassemi N, Jafari M, Moridian P, Khadem A, Sadeghi D, Hussain S, Zare A, et al. 2020. Automated detection and forecasting of COVID-19 using deep learning techniques: a review. *arXiv Preprint arXiv:200710785*.
- Siordia A Jr. 2020. Epidemiology and clinical features of COVID-19: a review of current literature. *J Clin Virol*. 127:104357.
- Smits SL, de Lang A, van den Brand JMA, Leijten LM, van Ijcken WF, Eijkemans MJC, van Amerongen G, Kuiken T, Andeweg AC, Osterhaus ADME, et al. 2010. Exacerbated innate host response to SARS-CoV in aged non-human primates. *PLoS Pathog*. 6(2):e1000756.
- Subasi A, Gürsoy MI. 2010. Comparison of PCA, ICA and LDA in EEG signal classification using DWT and SVM. *Expert Syst Appl*. 37(12):8659–8666.
- Tan L, Wang Q, Zhang D, Ding J, Huang Q, Tang Y-Q, Wang Q, Miao H. 2020. Lymphopenia predicts disease severity of COVID-19: a descriptive and predictive study. *Signal Transduction and Targeted Therapy* 5(1):1–3.
- Teo J. 2020. Early detection of silent hypoxia in Covid-19 pneumonia using smartphone pulse oximetry. *J Med Syst*. 44(8):1–2.
- Terpos E, Ntanas-Stathopoulos I, Elalamy I, Kastiris E, Sergentanis TN, Politou M, Psaltopoulou T, Gerotziafas G, Dimopoulos MA. 2020. Hematological findings and

- complications of COVID-19. *Am J Hematol.* 95(7): 834–847.
- Tuncer T, Dogan S, Subasi A. 2021. EEG-based driving fatigue detection using multilevel feature extraction and iterative hybrid feature selection. *Biomed Signal Process Control.* 68:102591.
- Valizadeh G, Babapour Mofrad F, Shalbah A. 2021. Parametric-based feature selection via spherical harmonic coefficients for the left ventricle myocardial infarction screening. *Med Biol Eng Comput.* 59(6):1261–1283.
- Ververidis D, Kotropoulos C, editors. 2005. Sequential forward feature selection with low computational cost. 2005 13th European Signal Processing Conference; IEEE.
- Wang D, Yeung DS, Tsang ECC. 2007. Weighted mahalanobis distance kernels for support vector machines. *IEEE Trans Neural Netw.* 18(5):1453–1462.
- Wang L. 2005. Support vector machines: theory and applications. Berlin: Springer.
- WHO Coronavirus Disease (COVID-19) Dashboard. <https://covid19.who.int>.
- Xie J, Covassin N, Fan Z, Singh P, Gao W, Li G, Kara T, Somers VK. 2020a. Association between hypoxemia and mortality in patients with COVID-19. *Mayo Clin Proc.* 95:1138–1147.
- Xie J, Hungerford D, Chen H, Abrams ST, Li S, Wang G, Wang Y, Kang H, Bonnett L, Zheng R, Li X, et al. 2020b. Development and external validation of a prognostic multivariable model on admission for hospitalized patients with COVID-19. *medRxiv*.
- Xu P, Zhou Q, Xu J. 2020. Mechanism of thrombocytopenia in COVID-19 patients. *Ann Hematol.* 99(6):1205–1208.
- Yadaw S, Li Y-c, Bose S, Iyengar R, Bunyavanich S, Pandey G. 2020. Clinical predictors of COVID-19 mortality. *medRxiv*.
- Yan L, Zhang H-T, Goncalves J, Xiao Y, Wang M, Guo Y, Sun C, Tang X, Jing L, Zhang M, et al. 2020. An interpretable mortality prediction model for COVID-19 patients. *Nat Mach Intell.* 2:283–288.
- Yang L, Liu S, Liu J, Zhang Z, Wan X, Huang B, Chen Y, Zhang Y. 2020a. COVID-19: immunopathogenesis and immunotherapeutics. *Sig Transduct Target Ther.* 5(1):1–8.
- Yang W, Cao Q, Qin L, Wang X, Cheng Z, Pan A, Dai J, Sun Q, Zhao F, Qu J, et al. 2020b. Clinical characteristics and imaging manifestations of the 2019 novel coronavirus disease (COVID-19): a multi-center study in Wenzhou city, Zhejiang, China. *J Infect.* 80(4):388–393.
- Yeung DS, Wang D, Ng WWY, Tsang ECC, Wang X. 2007. Structured large margin machine: sensitive to data distribution. *Mach Learn.* 68(2):171–200.