Current Research in Translational Medicine 70 (2022) 103319

Available online at
ScienceDirect
www.sciencedirect.com

Elsevier Masson France
EM|consulte
www.em-consulte.com

Original article

Clinical prognosis evaluation of COVID-19 patients: An interpretable hybrid machine learning approach

Ozan Kocadagli[a,*], Arzu Baygul[b], Neslihan Gokmen[c], Said Incir[d], Cagdas Aktan[e]

[a] Department of Statistics, Mimar Sinan University, Silahsör Cad. No. 71, 34380 Bomonti/Sisli, Istanbul, Turkey
[b] Department of Biostatistics, Faculty of Medicine, Koc University, Turkey
[c] Neslihan Gokmen, Department of Basic Sciences, Istanbul Technical University, Turkey
[d] Said Incir, Department of Biochemistry, Koc University Hospital, Turkey
[e] Department of Medical Biology, Beykent University, Turkey

ARTICLE INFO

Article History:
Received 9 June 2021
Accepted 26 October 2021
Available online 30 October 2021

Keywords:
COVID-19 symptoms
Severity of COVID-19
Clinical prognosis
Artificial intelligence
Machine learning
Feature selection
ICOMP

ABSTRACT

This retrospective cohort study deals with evaluating severity of COVID-19 cases on the first symptoms and blood-test results of infected patients admitted to Emergency Department of Koc University Hospital (Istanbul, Turkey). To figure out remarkable hematological characteristics and risk factors in the prognosis evaluation of COVID-19 cases, the hybrid machine learning (ML) approaches integrated with feature selection procedure based Genetic Algorithms and information complexity were used in addition to the multivariate statistical analysis. Specifically, COVID-19 dataset includes demographic features, symptoms, blood test results and disease histories of total 166 inpatients with different age and gender groups. Analysis results point out that the hybrid ML methods has brought out potential risk factors on the severity of COVID-19 cases and their impacts on the prognosis evaluation, accurately.

© 2021 Elsevier Masson SAS. All rights reserved.

1. Introduction

Coronavirus 2019 (also named COVID-19) is a contagious respiratory disease that influences human beings and is induced by severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) [1]. The coronavirus that firstly emerged in Wuhan China in 2019 and has influenced human beings on a global level, defined as a worldwide epidemic by the World Health Organization in March 2020 [2]. For this reason, many studies related to the COVID-19 pandemic have focused on identifying some risk factors and assessing of disease severity and prognosis of infected patients [3–5]. For instances, Chen et al. [3] classified the COVID-19 clinical types (severe vs. non-severe) to reduce the risk of overloading the healthcare system. They used totally 52 features related to comorbidity, symptoms and laboratory testing results, and the proposed model were estimated by Random Forest (RF). The importance of features was determined by using Gini impurity. Top ten features were listed as age, hypertension, cardiovascular disease, gender, diabetes; D-Dimer, hsTNI, absolute neutrophil count, IL-6, and LDH, in descending order and the RF's accuracy is found above 99%. They emphasized that comorbidities and symptoms can be used as an initial screening tool to detect the disease's severity [3].

Assaf et al. [4] predicted patient risk for critical COVID-19 (severe and non-severe). Severity of the cases was defined as the patient who has low partial arterial oxygen pressure and low oxygen saturation. Artificial Neural Networks, Random Forests and Regression Trees were performed to predict the severity. According to the variable importance, APACHE II score, white blood cell count, time from symptoms to admission, oxygen saturation and blood lymphocytes count were found the most contributory variables to the models which have 92% accuracy.

Gong et al. [6] constructed model for early detection severe COVID-19 cases. To classify the severity, various ML methods such as Decision Tree, Logistic Regression, Random Forest and Support Vector Machine models were used. Besides, the importance of features was determined using Least Absolute Shrinkage and LASSO Regression methods. In analysis, Age, DBIL, LDH, CRP, RDW, ALB and BUN were found as important features. The validation cohort's AUC was calculated as 0.853 (95% CI, 0.790–0.916). As a result, prevailing epidemiological data showed that the patients' mortality rate related to severe COVID-19 is more than the non-severe patients.

Fernandes et al. [7] worked on "A multipurpose machine learning approach to predict COVID-19 negative prognosis in São Paulo, Brazil". To predict COVID-19 negative prognosis, they used some reputable ML

* Corresponding author.
E-mail addresses: ozan.kocadagli@msgsu.edu.tr (O. Kocadagli), abaygul@ku.edu.tr (A. Baygul), ngokmen@itu.edu.tr (N. Gokmen), sincir@ku.edu.tr (S. Incir), cagdasaktan@beykent.edu.tr (C. Aktan).

https://doi.org/10.1016/j.retram.2021.103319
2452-3186/© 2021 Elsevier Masson SAS. All rights reserved.

classifiers such as RF and XGBoost classifiers. In analysis, these classifiers provided approximately 0.96 AUC.

Cai et al. [8] utilized on CT quantification and ML models for assessment of disease severity and prognosis of COVID-19 patients, and they used deep learning based U-Net and RF to classify CT images in addition to demographic features, symptoms, blood test results and disease histories. In analysis, the proposed hybrid approach brought out 0.93 AUC.

Vaid et al. [9] made a cohort study in which mortality and critical events in COVID-19 positive in New York City were predicted using XGBoost and logistic regression. By means of these classifiers, the critical illness and mortality up to 10 days were evaluated. The critical events and mortality were recorded at daily periods such as 3, 5, 7 and 10 days. Specifically, the logistic regression model is conducted only for day 3. As a results, XGBoost showed superior performance than the logistic regression in terms of achieving higher AUC in the mortality evaluation.

Wong et al. [10] worked on uncovering the clinical risk factors and classifying severe/fatal infection of COVID-19 cases based on UK Biobank data using only XGboost trees with cross-validation. In this study, the feature importance was quantified by Shapley values and accuracy gain. In analysis, a total of 2386 severe and 477 fatal cases were identified, and among infected individuals (totally 7846), their approach provided AUCs of 0.723 (95% CI:0.711−0.736) and 0.814(CI: 0.791−0.838) for severe and fatal infections, respectively. According to analysis outputs, the top five important factors for severity are age, number of drugs taken, cystatin C, wait-hip ratio and Townsend Deprivation index. Besides, the top features related to the prediction of mortality are age, testosterone, drugs taken, waist circumference and red cell distribution width.

Chieregato et al. [11] estimated a COVID-19 severity predictive model from CT images and clinical data using a hybrid machine learning approach based CatBoost and CNN classifiers with feature selection and importance. In this hybrid approach, CNNs with PCA (for the feature extraction) are specifically used for classifying CT images, and achieves the best AUC score with the third validation fold as 0.889 (mean AUC in the ten folds is 0.806). Also, CatBoost classifiers produced 94% AUC score. Lastly, the selected features and their importance are discussed in the text.

Patel et al. [12] used various ML classifiers (such as RF, AdaBoost, SVM, MLP) to classify COVID-19 disease severity and mechanical ventilation need, separately. They considered the socio-demographic, clinical, and blood panel profile datasets at the time of initial presentation for predicting the need for intensive care and mechanical ventilation by means of ML classifiers. According to analysis result, RF classifier produced the best AUC = 0.80 for predicting ICU need, and AUC=0.82 for predicting the need for mechanical ventilation. Also, it is found that blood panel profile data have significant impact on increasing the AUC.

Alotaibi et al. [13] proposed SVM classifier with recursive feature elimination to develop and compare prognosis prediction machine learning models based on invasive laboratory and noninvasive clinical and demographic data from patients' day of admission. The models were estimated by means of these hybrid SVMs over the laboratory findings of 80 COVID-19-infected patients and their health histories. According to analysis results, the test accuracies of the joint, non-invasive, and invasive models estimated by SVMs are 0.80 ± 0.03, 0.77 ± 0.04, and 0.75 ± 0.4, respectively. Also, the prediction performance of invasive and non-invasive models was executed with different classification frameworks (such as an ensemble model of decision trees with adaptive logistic boosting).

Alballa and Al-Turaiki [14] carried out a review study related to the recent reports on ML algorithms used in diagnosis of COVID-19 and prediction of mortality risk and severity over readily available clinical and laboratory data. This review paper consists of detailed analysis on both ML classifiers used in model estimation process and their performance comparisons. Besides, important features obtained from the mentioned ML classifiers are given and discussed in the text.

In addition to findings in above studies, recent studies showing that while elderly people have a higher risk against COVID-19, and young adults are suffering from severe symptoms as well. For this reason, there is an urgent need for a comprehensive risk assessment based on personalized genetic and physiological characteristics. Due to raise of COVID-19 cases day by day, the triage of patients becomes more importance to facilitate the effective use of limited resources and more efficient treatments [15]. Therefore, early assessment of severe prognosis may provide controlled facilities for intensive care, intubation and ventilation. With this kind of early prognosis, patients can be also protected from known risks such as hypotension, cardiac arrest and hypoxia as well. In this context, developing hybrid AI-based support systems is inevitable to manage and under control the clinical treatments against such a pandemic.

To introduce a novel feature selection and model estimation procedure for ML classifiers that helps to find out the potential risk factors on the severity of COVID-19 case, this paper is structured as following. The motivation and overview of developing a novel procedure is explained in the Section 2. The framework of methodology is defined in Section 3. The analysis outputs obtained from the multivariate statistical and developed hybrid ML approaches are given in Section 4. Finally, the analysis results and conclusions are discussed in Section 4.

2. Motivation and overview

In the modeling with the non-linear and complex systems such as ML methods, the functional structure of approximated model and number of features in training dataset are directly related to the model complexity [16−18]. However, let's look at the previous studies in literature, then it can be seen that the model complexity is often overlooked by the analysts. As a result of this, the model complexity causes two type errors: the approximation and estimation known as variance and bias trade-off as well [19−21]. In the training procedure, the model complexity is controlled by means of various approaches such early stopping procedure with cross-validation, regularization, feature selection, information criteria, etc. Essentially, using information criteria in modeling by ML methods provides a crucial role to control the model complexity in terms of constructing an appropriate approximated model, reducing the feature size and choosing the best model in the alternatives [21, 22].

According to the statistical learning jargon, Occam's razor (also called as the law of parsimony) provides an inevitable framework to control the model complexity, so it is widely considered in regression, classification, clustering and time series problems. However, this rule doesn't work properly to evaluate the best fitting models in the classification tasks, because Akaike-type criteria (AIC, BIC, CAIC, DIC) do not deal with the discriminative information regarding to competing classes in the classification problems [16, 21]. For this reason, discriminative based information criteria such as ICOMP are more convenient for the classification task. For this reason, in our study, ML classifiers are hybridized with GAs and ICOMP. While GAs is adapted to feature selection procedure, ICOMP is used to both the fitting function in GAs and choosing the best model in the alternatives. Detailed information about ICOMP can be found in Bozdogan [16], Kocadagli et al. [19], Ilter et al. [21] and Akbilgic&Bozdogan [23].

The second handicap in the previous studies related to early prognosis COVID-19 severity is that sufficient number of ML classifiers have not been handled to make a detailed comparison of the analysis results. In this study, to make a fair comparison whether the performance criteria are satisfied or not, the analysis were executed with various ML classifiers that are widely used in the literature.

Lastly, another shortcoming in the previous studies is that they often focus on only assessing the severity of COVID-19 cases instead of the feature selection and interpreting their importance. In our study, alternatively both feature selection and interpretation are

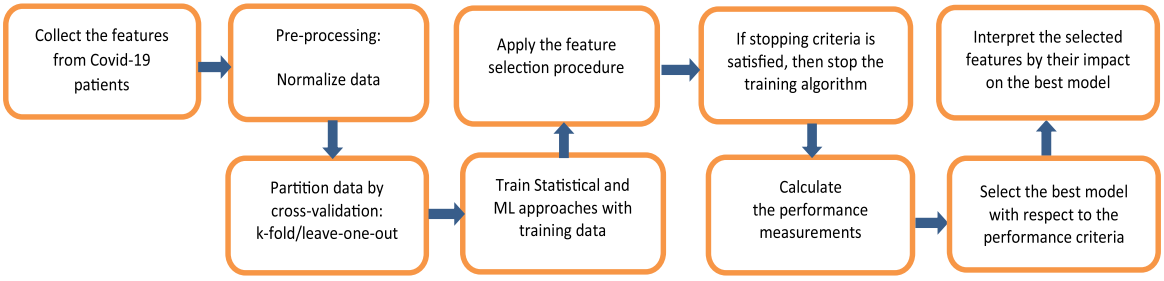


Fig. 1. Model estimation and selection framework.

handled together. Thus, the researchers not only take a notice about common features selected by ML classifiers, but also get an information regarding to their importance.

In light of the above, the main purpose of this study is to develop a prognostic machine-learning approach with a novel feature selection procedure. This approach helps to figure out important hematological characteristics and risk factors over the common clinical and laboratory parameters used in the prognosis evaluation of COVID-19 patients with hospital admission. Essentially, this framework provides a synchronized real time decision support system that allows the risk prediction about the hospital admission of the patients. Thus, evaluating severity and non-severity of COVID-19 cases might be executed by more robust ML models with selected and interpretable features. To develop this decision support system, firstly a COVID-19 dataset provided by Koc University Hospital, Istanbul, Turkey were collected. Specifically, this dataset includes demographic features, symptoms, blood test results, disease histories of total 166 inpatients with different age and gender groups.

3. Methodology

In this study, to figure out potential factors in blood-test, the first symptoms and demographic features that play an important role in evaluating severity and non-severity of COVID-19 cases, a novel training and feature selection procedure for ML classifiers are proposed. In this approach, three most popular ML classifiers; Artificial Neural Networks (ANNs), Support Vector Machines (SVMs) and Adapted Boosting (AdaBoost) were hybridized with a novel feature selection procedure based GAs and ICOMP. For multivariate statistical analysis, COVID-19 dataset were examined by means of discriminant analysis and logistic regression with the feature selection procedure integrated with the cross-validation and ICOMP. Alternatively, reducing the dimension of feature matrix was executed by PCA.

In literature, there exist many approaches in order to reduce the dimension of feature matrix such as all possible subset, forward selection, backward elimination, stepwise selections or some transformation techniques as PCA, ICA, factor analysis, fuzzy relations and their hybrid versions. However, these procedures have some shortcomings, and require some detailed statistical evaluation [16, 20, 21, 23]. For instance, all possible subset procedure needs 2^p iterations to determine the best feature subset where p is the number of features in the system. To overcome this computational burden, the hybrid feature selection procedures based evolutionary algorithms (such as GAs, PSO, Bee and Ant Colonies) and selection criteria are mostly preferred. In the literature, there exist remarkable studies in which GAs are used for feature selection procedure [21, 23-26]. For this reason, in this study, the feature selection procedure is based on GAs. Specifically, ICOMP is used as a fitting function in GAs. The framework of the proposed procedure is given in Fig. 1. The software of the proposed procedure is written in MATLAB 2020a The detailed information about ML and statistical methods can be found in Bishop [27, 28], Murphy [29] and Shahaid et al. [30].

The flowchart of the model estimation and selection procedure is exhibited in Fig. 2 as follow:

Before starting the analysis, firstly the dataset was normalized, and then the cross-validation type was chosen as k-fold or leave-one-out. After that, to train ML classifiers (ANNs, SVMs and AdaBoost) with integrated feature selection based GAs, initial tunings parameters must be defined.

To establish the structure of ANNs, the numbers of neurons and hidden layers are defined. To train ANNs with MSE or cross-entropy, the proposed approach provides various gradient-based algorithms: Gradient Descent with Momentum (GDwM), Scaled Conjugant Gradient (SCG), Levenberg Marquardt (LM) and BFGS. The more detailed information about the tuning parameters of these algorithms can be found in Bishop [27], Golden [31], Kocadagli [32] and MATLAB R2020 [33]. To control the model complexity in the training process, ANNs perform the cross-validation with an early stopping approach over-training and validation datasets. In training of SVMs, to obtain the robust models using the proposed estimation and feature selection procedure, various kernels in them such as linear, quadratic, cubic, radial and Gaussian were used. The more detailed information about these kernels can be found in Murphy [29] and MATLAB R 2020 [33].

4. Analysis

4.1. Data

In analysis, a dataset of Covid-19 patients that has been provided by Koç University Hospital, Istanbul, Turkey, was used. Specifically, this dataset consists of demographic characteristics, symptoms, blood test results and disease histories of totally 166 inpatients from different age and gender groups, and it was collected in compliance with the Turkish Ministry of Health. This study was approved by Koc University Ethics Committee, and informed consent was obtained from every participant. It was conducted in accordance with the ethical principles that have their origin in the Declaration of Helsinki. Besides, the research project regarding this study was approved by Koc University Ethics Committee (2020.269.IRB1.092) as well. Despite of existing many parameters regarding the hematological

Table 1
Dependent and independent features.

Independent Features		Dependent feature
Age (18–91)	Dyspnea (0/1)	Severity (0/1)
Spo2	Respiratory Rate	
Neutrophils	CRP	
GCS	Gender (0/1)	
Coronary Artery or Congestive Heart Disease (0/1)	HT (0/1)	
DM (0/1)	WBC	
First Symptom to Hospitalization (days)	Fever (0/1)	
Fatigue or Myalgia (0/1)		

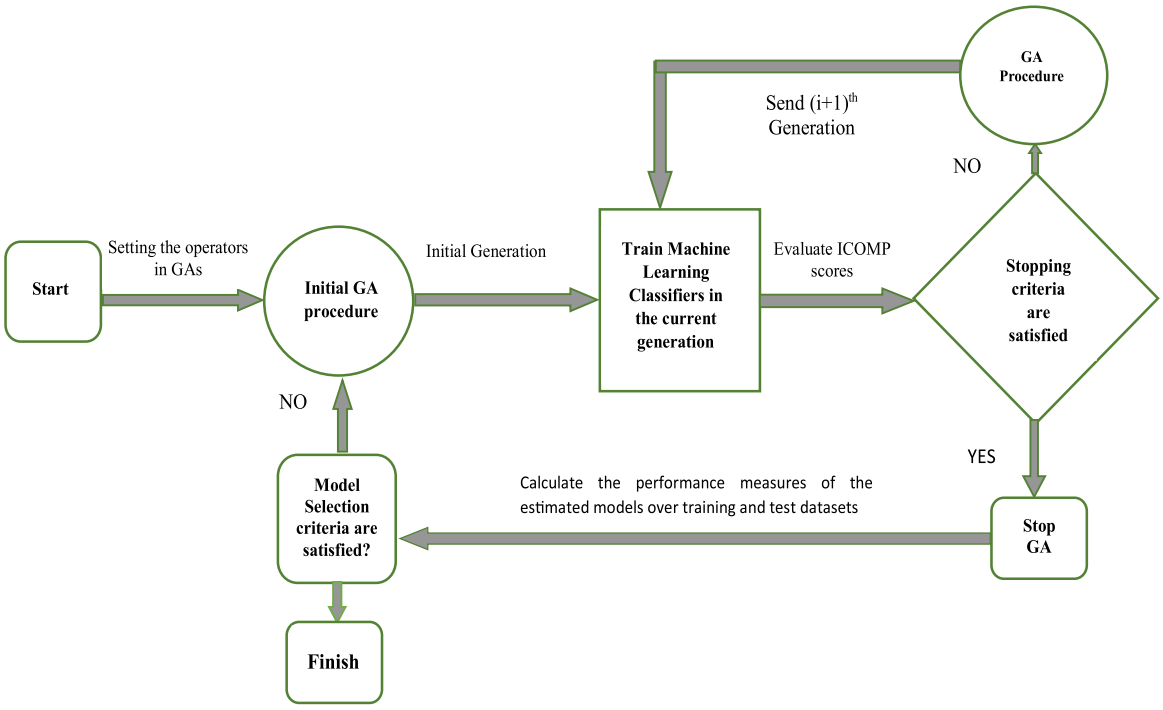


Fig. 2. The framework of the feature selection procedure.

measurements regarding COVID-19 and disease histories of patients, the features used in analysis were collected in accordance with the literature and expert knowledge [3–15]. The short description of these features are given in Table 1.

Specifically, training dataset includes totally 166 patients with the diagnosis of COVID-19. As seen from Table 1, this dataset includes 15 features, which are expected to be potential risk factors on the prognosis, were collected from the patients. The number of patients, whose prognosis are "severe" according to the need of ventilation is 52 (42.6). The patients with severe are older (67.3 ± 13.3) than the ones with "non-severe" (58.6 ± 14.5) patients (p = 0.001).

In the context of supervised learning, the model estimation procedure consists of two type features: dependent and independent. Dependent variable is the outcome, which is potentially effected by independent risk factors. In this study, while dependent feature corresponds to severe and non-severe cases of COVID-19 patients, the independent ones are clinical, treatment, blood-test measurements in addition to disease history and demographic characteristics.

4.2. Model estimation

After determining the cross-validation process, to start the model estimation process with the proposed approach, firstly some initial arrangements must be defined for ML classifiers and GAs before. Specifically, the feature selection procedure based GAs allows users to choose various crossover, mutation, immigration, elitism and selection operators. In analysis, outside of ICOMP, the other reputable IC's (AIC, CAIC BIC,) were defined as a fitness function in GAs as well. At the end of the training and feature selection procedure, the best models were selected with respect to some performance criteria such as ICOMP, classification accuracies, AUC, false positive and false negative ratios. The model performances of all the methods are given the following subsections.

Table 2
The classification performance of the ANN and SVM models.

Methods	Procedure	Number of Input	Accuracy ratios	AUC	FP	FN	Selected Features
ANNs	Feature Selection with GA	10	0.96	0.98	0.020	0.055	**Age, Dyspnea, Spo2, Respiratory Rate, Neutrophil, CRP, GCS, Coronary Artery or Congestive Heart Disease, First Symptom to Hospitalization, DM.**
	Full Model	15	0.87	0.85	0.180	0.100	All features in Table 1
SVMs	Feature Selection with GA	12	0.90	0.87	0.118	0.098	**Age, Dyspnea, Spo2, Respiratory Rate, Neutrophil, CRP, GCS, Coronary Artery or Congestive Heart Disease, First Symptom to Hospitalization, Fever**, Gender, HT.
	Full Model	15	0.81	0.79	0.216	0.169	All features in Table 1
AdaBoost (Weak learner: Decision Trees)	Feature Selection with GA	10	0.95	0.98	0.000	0.041	**Age, Dyspnea, Spo2, Respiratory Rate, Neutrophil, CRP, GCS, DM, Fever**, Fatigue or Myalgia.
	Full Model	15	0.93	0.94	0.020	0.041	All features in Table 1

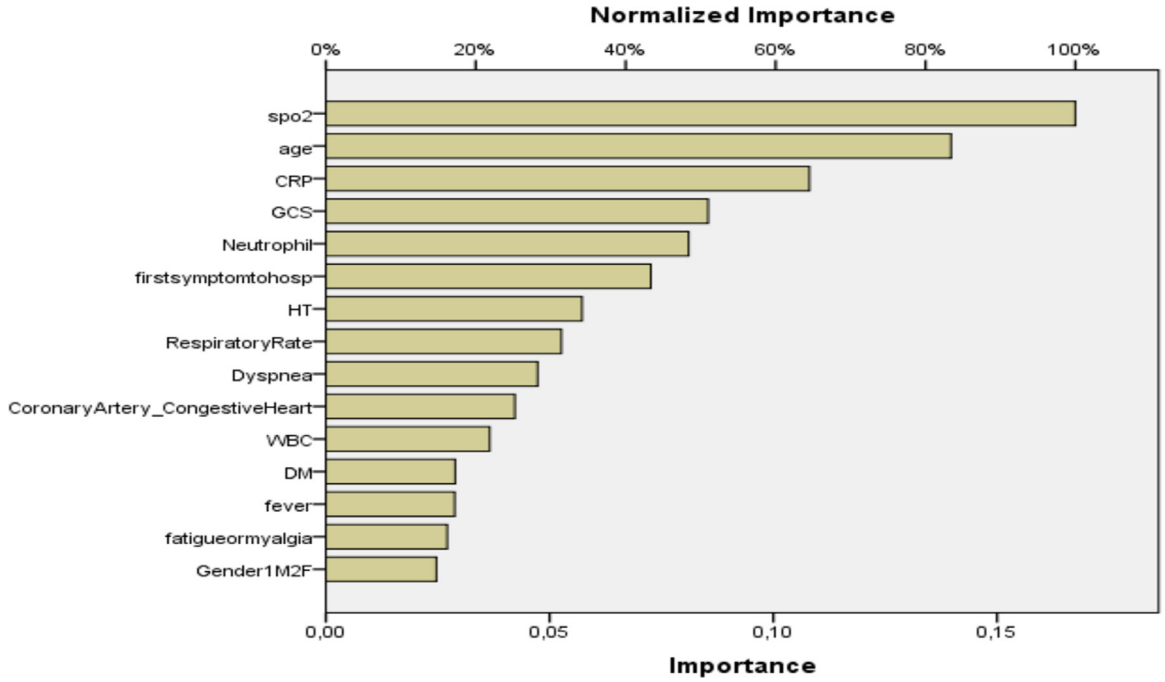


Fig. 3. Normalized Importance of explanatory features.

4.2.1. Analysis outputs of proposed approach

The performance criteria of best models obtained at end of the training of ANNs, SVMs and AdaBoost are given in Table 2. From Table 2, it can be inferred that ANNs shows better performance than SVMs with respect to all the performance criteria. Particularly, the best models with selected features are superior to the full models with all the explanatory ones according to all the performance measurements as well. In Table 2, the common explanatory features are included by all the estimated models are showed as bold characters: **Age, Dyspnea, Spo2, Respiratory Rate, Neutrophil, CRP, GCS, Coronary Artery or Congestive Heart Disease, First Symptom to Hospitalization, DM and Fever**.

To figure out importance of explanatory features on severity of COVID-19 cases, the estimated weights in the estimated ANN model were utilized. Normalized importance of explanatory features over the best full model and its ROC curve are given in Fig. 3 and Fig. 4, respectively.

4.2.2. Logistic regression analysis

In this section, to interpret the contribution of explanatory features with respect to the severity of COVID-19 cases individually, the logistic regression analysis was utilized. In analysis, the logistic regression models were estimated by the backward feature elimination integrated with k-fold and leave-one-out cross-validation procedures, separately. In addition, to control the model complexity and multi-collinearity, the size of dataset was reduced by using PCA. Specifically, the performances of estimated models are evaluated by various measures such as IC, accuracy ratio, false positive and false positive rates. The performances of estimated models are given in Table 3.

The selected features in Model 1, Model 2 and Model 3 are found statistically significant ($p = 0.001 < 0.05$), and two models except of Model 2 also suitable interpretations according to Hosmer-Lemeshow test statistics ($p \approx 0.96 > 0.05$). That is, Model 2 is not suitable interpretation according to Hosmer-Lemeshow test statistics ($p = 0.001 < 0.05$), even if the selected features are significant ($p = 0.001 < 0.05$). Model 3 has the highest accuracy (0.820) in addition to the lowest FP (0.157) and FN (0.211). The odds ratios of independent features related to the all the estimated models are given in Table 4.

4.2.3. Discriminant analysis

In this section, as an alternative to the logistic regression, the discriminant analysis was handled. After the analysis, the stepwise feature selection integrated with leave-one-out and k-fold validation approaches produced the best models given in Table 5. According to analysis result, the assumption of equality of variance-covariance matrices is provided (Box-M, $p < 0.001$) and the estimated parameters are found significant (Wilks' Lambda $p<0.001$). As seen from Table 5, both models consist of Age, Dyspnea, Spo2 and fatigue or myalgia. The first model (M1) has the highest accuracy (0.770) in addition to the lowest FP (0.186) and FN (0.289).

5. Results and conclusions

According to analysis results in Table 2, the proposed hybrid ANN classifiers brought out better performance than hybrid SVM and AdaBoost classifiers. All the ML classifiers integrated feature selection procedure based GAs and IC are able to reduce FP and FN rates at the reasonable levels much more than the full models in addition to providing the highest accuracy ratios. In other words, the proposed estimation and feature selection procedure helps to improve the all the performance criteria as well as controlling model complexity.

The best model estimated by ANNs (with these predictors: **Age, Dyspnea, Spo2, Respiratory Rate, Neutrophil, CRP, GCS, Coronary Artery or Congestive Heart Disease, First Symptom to Hospitalization, DM**) has better performance (ACC: 0.96, FN:0.020, FP: 0.055) than the full model with ACC:0.87, FN:0.180, FP: 0.100. Similarly, the best model estimated by SVMs (with these predictors: **Age, Dyspnea, Spo2, Respiratory Rate, Neutrophil, CRP, GCS, Coronary Artery or Congestive Heart Disease, First Symptom to Hospitalization, Fever, Gender, HT**) has better performance (ACC: 0.90, FN:0.118, FP: 0.098) than the full model with ACC:0.81, FN:0.216, FP: 0.169. Lastly, the best model estimated by AdaBoost (with these predictors: **Age,**

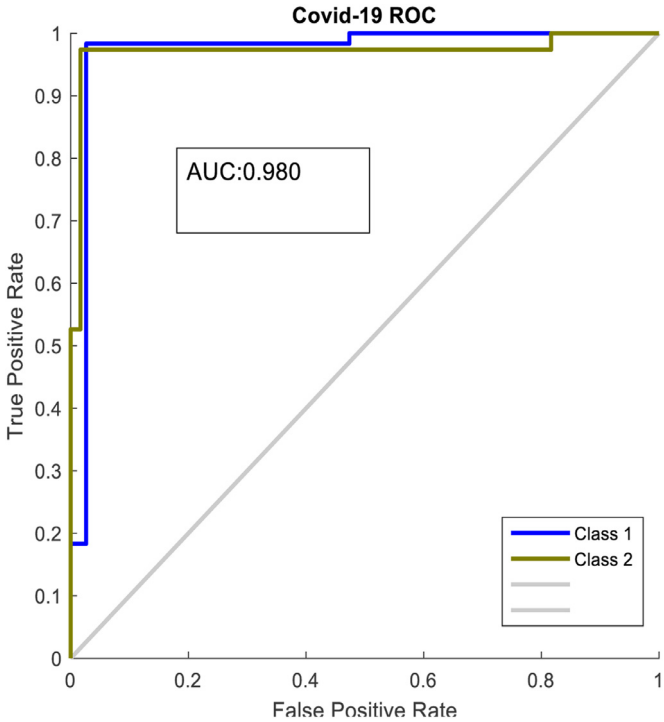


Fig. 4. ROC curve of hybrid ANN model.

Dyspnea, Spo2, Respiratory Rate, Neutrophil, CRP, GCS, DM, Fever, Fatigue or Myalgia) has better performance (ACC: 0.95, FN:0.000, FP: 0.041) than the full model with ACC:0.93, FN:0.02, FP: 0.041.

If all the best models estimated by the proposed approach are compared each other, then it can be seen that these models consist of the following common features: **Age, Dyspnea, Spo2, Respiratory Rate, Neutrophil, CRP, GCS, Coronary Artery or Congestive Heart Disease, First Symptom to Hospitalization, DM and Fever.** As seen from Fig. 3, let the explanatory features be sorted in descending order according to their importance over the best estimated ANN model, then the top ten important ones stand out as following: **Spo2, Age, CRP, GCS, Neutrophil, First Symptoms, Dyspnea, HT, Respiratory Rate, Coronary Artery or Congestive Heart Disease.**

In the logistic regression integrated backward elimination with leave-one-out, the best model (M3) consists of five features which can be considered as a potential risk factors on severity: **Age, Dyspnea, Spo2, Neutrophil and CRP**. According to the odds ratios of Model 3 given in Table 4, it can be inferred that while one unit changes in Spo2 decreases the severity by (1/0.705) 1.418 times; Dyspnea, age, CRP and Neutrophil increase it by 10.204, 1.064, 1.010, 1.000 times, respectively.

The best model estimated from discriminant analysis with the stepwise feature selection includes the following four features which can be considered as a potential risk factors on severity: **Age, Dyspnea, Spo2, Fatigue or Myalgia.** From the best models of the logistic regression and discriminant analysis, it can be seen that **Age, Dyspnea, Spo2** are common risk factors.

To examine the performances of the best models estimated by the proposed approach versus some reputed methods in the ML literature, the comparison of the best models is given in Table 6. To be fair, all the models were estimated by using same cross-validation procedure (k-fold =10).

According to the performances of the best models in Table 6, it can be inferred that the proposed approach (Hybrid ANNs) gives out superior performance to the other methods with respect to ACC, AUC, false positive and false negative rates. Actually, the proposed feature selection procedure based GAs and IC helps to control the model complexity, and reduce FP and FN rates, thus provides the highest accuracy ratios as well. RF with Hyper-parameter Bayesian Optimization has the second best performance according to related performance criteria as well. For this reason, to compare the first and second models each other in terms of consistency of the feature importance, the out of bag permuted predictor importance estimates of the best RF model are given in Fig. 5.

Let's take a look Fig. 3 and Fig. 5, it can be seen that **Age, Dyspnea, Spo2, Respiratory Rate, Neutrophil, CRP, GCS** and **First Symptom to Hospitalization** stand out similarly in the first and second best models in the context of feature importance. Apparently, these

Table 3
The performances of logistic regression models.

Models/Size Reduction	Input number	NSV	Accuracy Ratio	AUC	FP	FN	Selected Features
Model 1 Backward No cros-val.	15	6	0.803	0.800	0.171	0.231	**Age, Dyspnea, Spo2,** Neutrophil, CRP, fatigue or myalgia
Model 2 Backward with 10-fold	15	3	0.817	0.820	0.159	0.216	**Age, Dyspnea, Spo2**
Model 3 Backward with Leave-one-out	15	5	0.820	0.820	0.157	0.211	**Age, Dyspnea, Spo2,** Neutrophil, CRP
Model 4 PCA with Backward (10-fold)	5	3	0.80	0.80	0.153	0.222	–

NSV = Number of selected features; **FP:** False Positive; **FN:** False Negative.

Table 4
Odds ratios of independent features.

Model 1	OR (p)	Model 2	OR (p)	Model 3	OR (p)
Age	1.092 (<0.001)	Age	1.058 (0.014)	Age	1.064 (0.005)
Dyspnea	14.425 (<0.001)	Dyspnea	8.547 (0.001)	Dyspnea	10.204 (<0.001)
Spo2	0.946 (0.001)	Spo2	0.708 (0.004)	Spo2	0.705 (0.002)
Neutrophil	1.000 (0.027)			Neutrophil	1.000 (0.002)
CRP	1.013 (0.010)			CRP	1.010 (0.010)
Fatigue or myalgia	0.939 (0.064)				

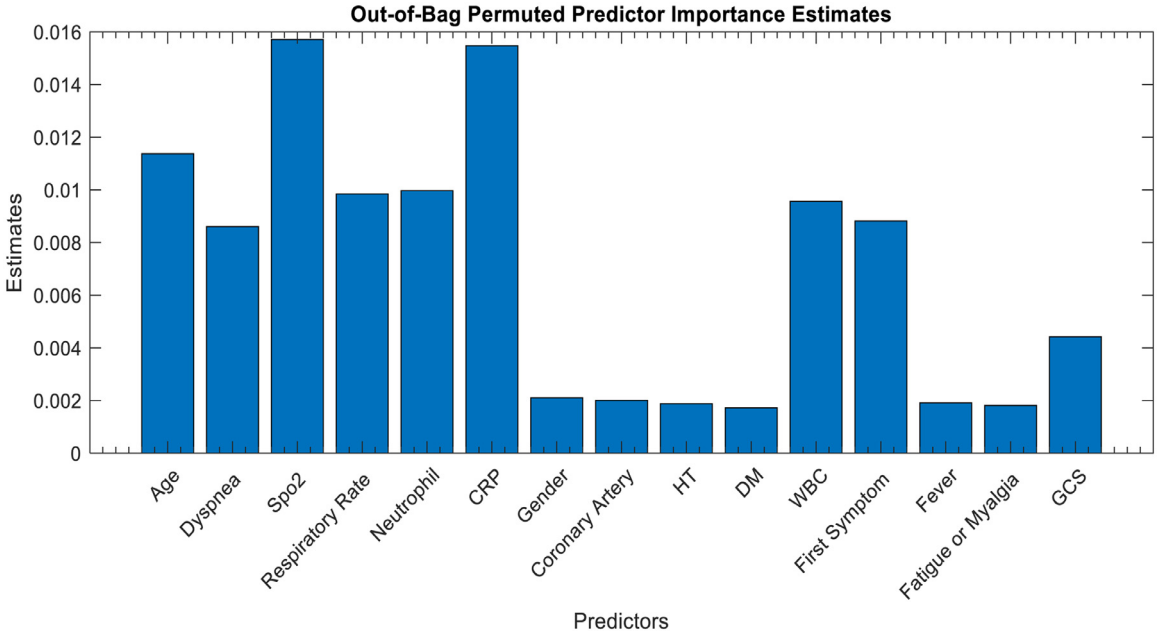


Fig. 5. Predictor importance estimates of the best RF model.

features are common risk factors in which are included in the models estimated by the proposed approach.

From analysis results given between Table 2 - Table 6, it can be apparently seen that ML classifiers produced superior performance than the classical statistical approach (Logistic and Discriminant) in terms of determining the severity of any COVID-19 case accurately. Besides, ML approaches utilizes much more risk factors than the classical statistical ones in order to classify severe and non-severe COVID-19 cases, accurately.

To enable the future comparisons our proposed procedure with existing methods in the literature, its MATLAB code and interface (COVID19_SEVERITY_GUI) will be shared with the readers of Journal. In addition, the initial setup steps of COVID19_SEVERITY_GUI are described in the Appendix I. In the future direction, we are planning to continue this research with much more comprehensive COVID-19 datasets and carry out more detailed analysis using various hybrid ML approaches.

The software availability statement

The software of the proposed approach will be provided by Authors (or Journal's Domain).

Table 5
The performance of discriminant analysis.

Models/ Size Reduction	Total Number of Features	Number of selected features	ACC	AUC	FP	FN	Selected Features
Model 1 (K-fold)	15	4	0.770	0.760	0.186	0.289	Age, Dyspnea, Spo2, fatigue or myalgia
Model 2 (Leave-one-out)	15	4	0.754	0.750	0.214	0.289	Age, Dyspnea, Spo2, fatigue or myalgia
Model 3 (PCA)	5	3	0.786	0.730	0.225	0.190	–

Table 6
The comparison of the best models.

Methods	ACC	AUC	FP	FN
Proposed Approach (Hybrid ANNs)	**0.96**	**0.98**	**0.02**	**0.05**
Random Forest with Hyper-parameter Bayesian Optimization	0.96	0.96	0.02	0.05
Proposed Approach (Hybrid AdaBoost)	0.95	0.98	0.00	0.04
KNN(Distance:Manhattan	0.93	0.94	0.06	0.08
Ensemble: Subspace KNN	0.90	0.93	0.06	0.08
Ensemble: Subspace Discriminant	0.90	0.95	0.00	0.31
KNN (Distance: Cosine	0.90	0.93	0.06	0.08
KNN (Distance: Euclidian	0.90	0.93	0.06	0.08
Proposed Approach (Hybrid SVMs)	0.90	0.88	0.11	0.09
Random Forest with Hyper-parameter Bayesian Optimization (without Cross-validation)	0.87	0.88	0.12	0.12
KNN (Distance: Cubic	0.86	0.92	0.07	0.08
KNN(Distance: Spearman	0.86	0.92	0.07	0.08
The Logistic Regression with Leave-one-out	0.820	0.820	0.157	0.211
Ensemble: RUS Boosted Trees	0.80	0.86	0.27	0.15
Fine Tree (Splitting: Gini's Index)	0.80	0.79	0.27	0.15
Coarse Tree (Splitting: Gini's Index)	0.80	0.79	0.27	0.15
Quadratic Discriminant	0.80	0.90	0.00	0.46
Linear Discriminant	0.77	0.76	0.18	0.28
Random Forest without hyper-parameter optimization	0.73	0.72	0.33	0.23

Conflict of Interest

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

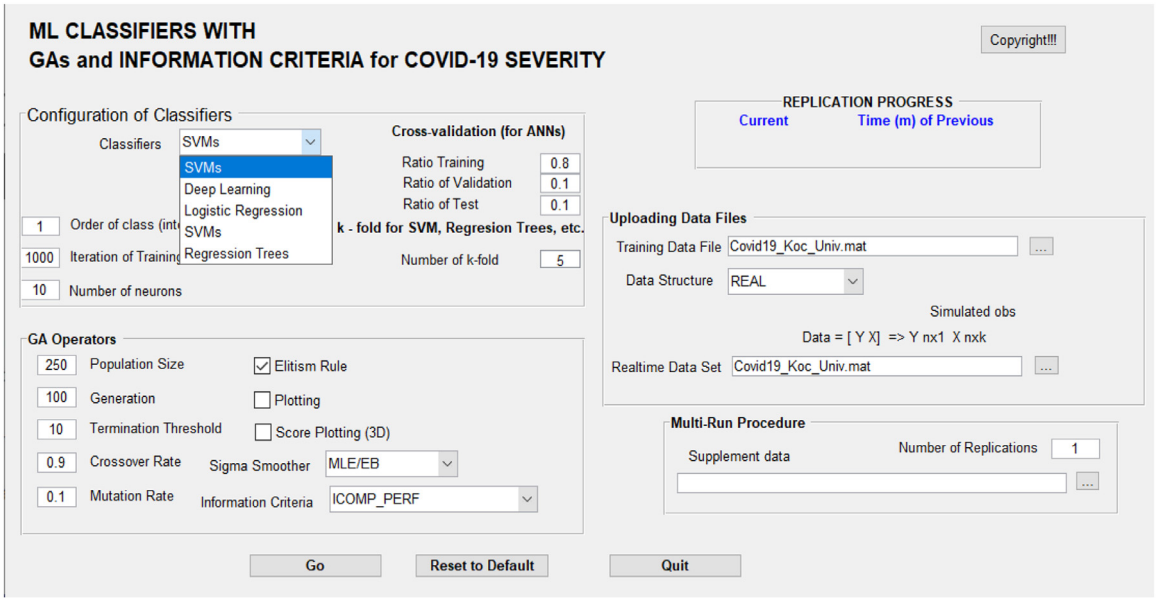
Acknowledgement

This research project was approved by Koc University Ethics Committee (2020.269.IRB1.092). COVID-19 dataset were provided by Koç University Hospital, Istanbul, TURKEY.

Appendix I. COVID19_SEVERITY_GUI

To run MATLAB code (COVID19_SEVERITY_GUI) of the proposed procedure, initial setup steps are summarized as follows:

1 Download runtime file (MyAppInstaller_mcr) from code provider website using the following link: https://www.mathworks.com/products/compiler/matlab-runtime.html
2 Copy setup files of COVID19_SEVERITY_GUI into C:\Program Files
3 In "for_redistribution_files_only" file, run "COVID19_SEVERITY_GUI.exe"
4 From the opened interface (COVID19_SEVERITY_GUI), define the setup parameters of the proposed procedure as follows:



1 Upload your Covid-19 data using “input file” segment (Data structure must be defined as “REAL”)
2 Select any classifier from “Classifier Segment”
3 Define Cross-validation parameters and GA operators
4 After click “go” button, then the training procedure will start automatically!

(Please read “Copyright Information” carefully!)

References

[1] WHO. Q&A on coronaviruses. Press Briefing; 2020.
[2] WHO. WHO director-general's opening remarks at the media briefing on COVID-19. Press Brief; 2020.
[3] Chen Y, et al. An interpretable machine learning framework for accurate severe vs non-severe COVID-19 clinical type classification. SSRN Electron J 2020. doi: 10.2139/Ssrn.3638427.
[4] Assaf D, et al. Utilization of machine-learning models to accurately predict the risk for critical COVID-19. Intern Emerg Med 2020;15(8):1435–43. doi: 10.1007/S11739-020-02475-0.
[5] Liao D, et al. Hematological characteristics and risk factors in the classification and prognosis evaluation of covid-19: a retrospective cohort study. Lancet Haematol 2020;7(9):E671–8. doi: 10.1016/S2352-3026(20)30217-9.
[6] Gong J, et al. A tool for early prediction of severe coronavirus disease 2019 (Covid-19): a multicenter study using the risk nomogram in Wuhan and Guangdong, China. Clin Infect Dis 2020;71(15):833–40. doi: 10.1093/Cid/Ciaa443.
[7] Fernandes FT, et al. A multipurpose machine learning approach to predict COVID-19 negative prognosis in São Paulo, Brazil. Sci Rep 2021;11:3343. doi: 10.1038/S41598-021-82885-Y.
[8] Cai W, et al. CT quantification and machine-learning models for assessment of disease severity and prognosis of COVID-19 patients. Acad Radiol 2020;27:1665–78.
[9] Vaid A, Somani S, Russak A, De Freitas J, Chaudhry F, Paranjpe I, Johnson K, Lee S, Glicksberg B. Machine learning to predict mortality and critical events in a cohort of patients with COVID-19in New York City: model development and validation. J Med Internet Res 2020;22(11). doi: 10.2196/24018.
[10] Kenneth CY, Xiang Y, So HC. Uncovering clinical risk factors and prediction of severe COVID-19: a machine learning approach based on UK biobank data. Medrxiv 2021 2020-09.
[11] M. Chieregato, F. Frangiamore, M. Morassi, C. Baresi, S. Nici, C. Bassetti & M. Galelli, "A hybrid machine learning/deep learning COVID-19 severity predictive model from CT images and clinical data", 2021, Arxiv Preprint Arxiv:2105.06141.
[12] Patel D, Kher V, Desai B, Lei X, Cen S, Nanda N, Oberai AA. Machine learning based predictors for COVID-19 disease severity. Sci Rep 2021;11(1):1–7.
[13] Alotaibi A, Shiblee M, Alshahrani A. Prediction of severity of COVID-19-infected patients using machine learning techniques. Computers 2021;10(3):31.
[14] Alballa N, Al-Turaiki I. Machine learning approaches in COVID-19 diagnosis, mortality, and severity risk prediction: *a review*. Informatics in Medicine Unlocked 2021:100564.
[15] Burdick H, et al. Prediction of respiratory decompensation in Covid-19 patients using machine learning: the ready trial. Comput Biol Med 2020;124:2020. doi: 10.1016/J.Compbiomed.2020.103949.
[16] Bozdogan H. Akaike's information criterion and recent developments in information complexity. J Math Psychol 2000;44(1):62–91. doi: 10.1006/Jmps.1999.1277.
[17] Kocadagli O. Hybrid Bayesian neural networks with genetic algorithms and fuzzy membership functions. Istanbul, Turkey: Department of Statistics, Mimar Sinan F. A. University; 2012 Phd Thesis.
[18] Kocadagli O, Asikgil B. Nonlinear time series forecasting with bayesian neural networks. Expert Syst Appl 2014;41(15):596–6610.
[19] Kocadagli O, Deniz E, Ilter D, D.. Artificial intelligence techniques and information criteria in statistical modeling, scientific research project. Mimar Sinan University; 2020 Grant Number: 2018-31.
[20] Kocadagli O, Langari R. Classification of EEG signals for epileptic seizures using hybrid artificial neural networks based wavelet transforms and fuzzy relations. Expert Syst Appl 2017;88. doi: 10.1016/J.Eswa.2017.07.020.
[21] Ilter D, Deniz E, Kocadagli O. Hybridized artificial neural network classifiers with a novel feature selection procedure based genetic algorithms and information complexity in credit scoring. Appl Stoch Model Bus Ind 2021;37(2):203–28. doi: 10.1002/Asmb.2614.
[22] Ilter D, Kocadagli O. Credit scoring by artificial neural networks based cross-entropy and fuzzy relations. Sigma J Natural Sci 2019;37(3):855–70 Yildiz Technical University.
[23] Akbilgic O, Bozdogan H. A new supervised classification of credit approval data via the hybridized rbf neural network model using information complexity. Studies in classification, data analysis, and knowledge organization, 48; 2015. p. 201513–27. doi: 10.1007/978-3-662-44983-7_2.
[24] Lia AD, Xueb B, Zhangb M. Multi-objective feature selection using hybridization of a genetic algorithm and direct multisearch for key quality characteristic selection. Inf Sci (Ny) 2020;523:245–65.
[25] Liberati C, Howe JA, Bozdogan H. Data adaptive simultaneous parameter and kernel selection in kernel discriminant analysis using information complexity. J. Pattern Recognit. Res. 2009;4(1):119–32.
[26] J.A. Howe and H. Bozdogan, “Regularized SVM classification with a new complexity-driven stochastic optimizer”, 2016, Vol. 9, No. 2, pp. 216−30.
[27] Bishop C. Pattern recognition and machine learning. Springer; 2006.
[28] Bishop C. Neural networks for pattern recognition. UK: Oxford University Press; 1996.
[29] Murphy KP. Machine learning: a probabilistic perspective. 2012 https://Www.Cs.Ubc.Ca/~Murphyk/Mlbook/(Ac).
[30] Shahid N, Rappon T, Berta W. Applications of artificial neural networks in health care organizational decision-making: a scoping review. PLoS ONE 2020;14(2).
[31] Golden R. Mathematical methods for neural network analysis and design. MIT Press; 1996.
[32] Kocadagli O. A novel hybrid learning algorithm for full bayesian approach of artificial neural networks. Appl Soft Comput 2015;35:1–958 Elsevier.
[33] “MATLAB R2020a.” https://Www.Mathworks.Com/Products/New_Products/Release2020a.html.