Predicting Risk of Mortality in COVID-19 Hospitalized Patients using Hybrid Machine Learning Algorithms

Mohammad Reza Afrash[1], Mostafa Shanbehzadeh[2], Hadi Kazemi-Arpanahi[3,4]*

[1]PhD, Department of Artificial Intelligence, Smart University of Medical Sciences, Tehran, Iran
[2]PhD, Department of Health Information Technology, School of Paramedical, Ilam University of Medical Sciences, Ilam, Iran
[3]PhD, Department of Health Information Technology, Abadan University of Medical Sciences, Abadan, Iran
[4]PhD, Student Research Committee, Abadan University of Medical Sciences, Abadan, Iran

ABSTRACT

Background: Since hospitalized patients with COVID-19 are considered at high risk of death, the patients with the sever clinical condition should be identified. Despite the potential of machine learning (ML) techniques to predict the mortality of COVID-19 patients, high-dimensional data is considered a challenge, which can be addressed by metaheuristic and nature-inspired algorithms, such as genetic algorithm (GA).

Objective: This paper aimed to compare the efficiency of the GA with several ML techniques to predict COVID-19 in-hospital mortality.

Material and Methods: In this retrospective study, 1353 COVID-19 in-hospital patients were examined from February 9 to December 20, 2020. The GA technique was applied to select the important features, then using selected features several ML algorithms such as K-nearest-neighbor (K-NN), Decision Tree (DT), Support Vector Machines (SVM), and Artificial Neural Network (ANN) were trained to design predictive models. Finally, some evaluation metrics were used for the comparison of developed models.

Results: A total of 10 features out of 56 were selected, including length of stay (LOS), age, cough, respiratory intubation, dyspnea, cardiovascular diseases, leukocytosis, blood urea nitrogen (BUN), C-reactive protein, and pleural effusion by 10-independent execution of GA. The GA-SVM had the best performance with the accuracy and specificity of 9.5147e+01 and 9.5112e+01, respectively.

Conclusion: The hybrid ML models, especially the GA-SVM, can improve the treatment of COVID-19 patients, predict severe disease and mortality, and optimize the utilization of health resources based on the improvement of input features and the adaption of the structure of the models.

Citation: Afrash MR, Shanbehzadeh M, Kazemi-Arpanahi H. Predicting Risk of Mortality in COVID-19 Hospitalized Patients using Hybrid Machine Learning Algorithms. *J Biomed Phys Eng*. 2022;12(6):611-626. doi: 10.31661/jbpe.v0i0.2105-1334.

Keywords

Machine Learning; Artificial Intelligence; Coronavirus (COVID-19); Data Mining; Mortality

Introduction

In December 2019, a new coronavirus disease (COVID-19) outbreak appeared in Wuhan, China [1, 2]. Due to fast transmission, COVID-19 was known as a pandemic in a few months worldwide, affecting public health, economic, and social conditions [3, 4] with a wide range of clinical presentation and prognosis, such as the common cold, respiratory infections, multiple organ failure, and death [5, 6]. Also,

*Corresponding author:
Hadi Kazemi-Arpanahi
Department of Health Information Technology, Abadan University of Medical Sciences, Abadan, Iran
E-mail: h.kazemi@abadanums.ac.ir

Received: 13 May 2021
Accepted: 20 January 2022

fast exponential transmission and incremental mortality rate led to a tremendous panic in the world [7, 8]. Without fully licensed treatment or whole safe vaccination, some mitigation efforts were implemented to control the epidemic [9, 10]. In many low-and middle-income countries (LMICs), such as Iran, the public with low-health information followed fewer hygiene guidelines provided by the government for the protection from COVID-19, leading to the spread of the virus and broken the health systems, especially in LMICs [11, 12]. Therefore, plans based on the effective prognosis are most important for healthcare authorities to evaluate triage patients' conditions and manage limited medical resources adequately [13, 14].

Machine learning (ML), as a subgroup of Artificial Intelligence (AI), utilizes scientific algorithms to mine effective, previously unfamiliar, comprehensible and hidden patterns from huge raw datasets for predictions or decisions [15, 16]. The ML methods recognize tools for developing predictive models and extract valuable patterns from raw data [17]. In the earlier studies, some ML models were developed to predict and classify COVID-19 mortality, such as Artificial Neural Networks (ANNs) [18-25], Decision Trees (DT) [19, 22, 26], Support Vector Machine (SVM) [19, 22, 27], Random Forest (RF) [19, 22, 27, 28], and Naive Bayes (NB) [29]. On the other hand, a major challenge of ML algorithms is high-dimensional datasets leading to statistical or mathematical problems. Irrelevancy and redundancy in estimated variables and features can increase the misperception of ML algorithms and decrease learning accuracy. Accordingly, the elimination of these outlier variables and features is a great challenge that is particularly significant in the case of COVID-19, with many complexities and some unknown aspects [30, 31].

Considering the complexity and ambiguity of COVID-19, it is necessary to identify important features (predictors) to increase the predictability of the model and predict a specific outcome variable (e.g., the death of COVID-19 patients, their length of stay (LOS), and survival) [32]. The combination of some ML techniques usually can have better accuracy than just one ML algorithm [31, 33]. According to the Genetic Algorithm (GA), an attractive method is used to decrease the model's complexity by reducing the data dimensionality [34, 35]. This paper aimed to assess the performance of the GA paired with some ML algorithms to predict COVID-19 mortality at the initial hospitalization of the patients. The clinical variables with predictor roles in the mortality of COVID-19 were determined using the GA optimization procedures and also included in four ML algorithms K-nearest neighbor (KNN), DT, SVM, and ANN to construct the predictive models. Finally, the performance of each combination was measured using some evaluation criteria, including accuracy, sensitivity, specificity, and the area under the receiver operating characteristic curve (AUC).

Material and Methods

In this retrospective and single-center study, four ML algorithms were trained using optimized variables selected by the GA algorithm. Accordingly, these hybrid models were compared in terms of accuracy, precision, and specificity benchmarks. The GA was applied to identify and prioritize the best set of COVID-19 mortality affecting variables. Also, the best collection of COVID-19 mortality-affecting variables was inputted into ML algorithms to construct the prediction models. Finally, the results were evaluated using a 10-fold cross-validation method.

All the submitted models were coded using Python (version 3.7.7), and practical experiments were performed with a simulation environment, including a Core i7-4210U device with 6 GB of RAM.

The identity of the patients was hidden for their confidentiality during the data collection

process.

Dataset description

The dataset was obtained from the database registry at Ayatollah Taleqhani Hospital, affiliated with Abadan University of Medical Sciences, the main center for delivering COVID-19 specialized care and treatment in the southwest Khuzestan Province, Iran. A total of 12885 suspected COVID-19 cases were referred to this center, of whom 3350 cases were introduced as positive reverse transcription-polymerase chain reaction (RT-PCR) for COVID-19 from February 9 to December 20, 2020. Finally, only hospitalized patients, who met the inclusion criteria were involved in this study (Figure 1). The number of 56 features correlated in mortality prediction and a resultant feature as an output or the predicted variable is shown in Table 1.

Data Preprocessing

Data preprocessing is key in preparing an optimal dataset before training ML algorithms. In the present study, some preprocessing techniques were applied to the dataset after data collection. In this step, the rows of the dataset with missing values of greater than 70%, the noisy data, and outliers or inconsistent data were removed to enhance classification algorithms by two health-information management specialists and two infectious diseases experts.

Data balancing

The imbalanced data is one of the main obstacles to training ML algorithms due to the uncategorized classes. The dataset contains 955 cases related to alive individuals, while the death class has only 270 individuals. Accordingly, the developed models often deliver prejudiced results towards overriding class, and the ML models are much more likely to categorize new observations for the majority class. In this study, the number of individuals was balanced (equal to 955) for both alive and dead groups after using the synthetic minority

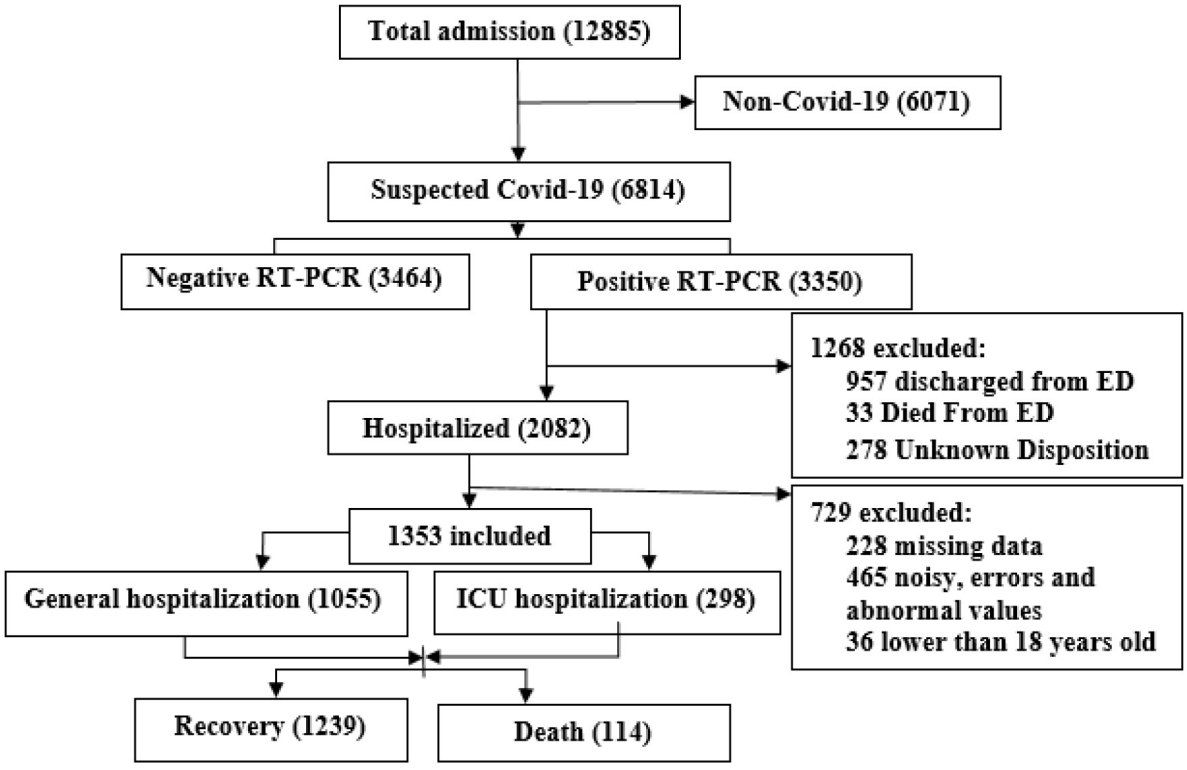

Figure 1: Flowchart describing patient selection

Table 1: Identifying initial list of variables affecting mortality in patients with COVID-19

Classes	Predictor variables	Outcome variable
Demographic	LOS, age, height, weight, blood type, gender;	Mortality status (alive/death)
Clinical manifestations	Cough, contusion, nausea, vomit, headache, GI symptoms, muscular pain, chill, fever, pneumonia, respiratory intubation, dyspnea, loss of taste, loss of smell, runny nose, sore throat;	
Comorbidities/ risk factors	Other underlying diseases, CVD, hypertension, diabetes, smoking, addiction, alcohol consumption;	
Laboratory tests	Creatinine, RBC count, WBC count, hematocrit, hemoglobin, platelet count, ALC, ANC, calcium, phosphorus, magnesium, sodium, potassium, BUN, total bilirubin, AST, ALT, albumin, glucose, LDH, activated PTT, PT, ALP, C-reactive protein, ESR, hypersensitive troponin, pleural effusion.	

LOS: Length of Stay, CVD: Cardiovasculary Diseases, RBC: Red Blood Cell, WBC: White Blood Cell, ALC: Absolute Lymphocyte Count, ANC: Absolute Neutrophil Count, BUN: Blood Urea Nitrogen, AST: Aspartate Aminotransferase, ALT: Alanine Aminotransferase, LDH: Lactate Dehydrogenase, PTT: Partial Thromboplastin Time, PT: Prothrombin Time, ALP: Alkaline Phosphatase, ESR: Erythrocyte Sedimentation Rate

over-sampling technique (SMOTE).

Selection of feature subsets with GA

Feature selection was based on removing unneeded variables from the original dataset without significantly lost information. Due to high-dimensional and complex data, feature selection was a crucial step in data mining and pattern recognition. Feature selection enhanced learning effectiveness and predictive recital and reduced the complication of learned results by input optimization [30, 33, 36]. The feature selection also determined the most optimal list of features and reduced the computational complexity of models. The GA as a feature selection method, which is based on the theory of natural selection or Darwin, can consider all possible connections between variables and identify the most proper combination of variables [30, 31, 34, 37, 38]. Therefore, GA iterations were implemented to select COVID-19 mortality predictors.

GA implementation

In the proposed hybrid models, the GA optimized the predictor variables, searching in the "candidate solution space" to find the best possible solution for a problem using "simulating" the process of evolution in nature. In the search process for the optimal solution, a set of initial solutions is firstly generated, and a set of modified solutions are produced in successive "generations," i.e., in each generation of the GA, specific changes are made in the genes of the chromosomes. The initial solutions are mostly changed so that the population of solutions "converges" towards the optimal solution in each generation [39, 40].

The process of GA is as follows:

1) Initializing population: the genetic algorithm starts by generating an initial population, including all the probable solutions to the given problem. The most popular technique for initialization is the use of random binary strings, generating an initial set of a and b values randomly (a and b values are named 'chromosomes'), 2) fitness function: the assessment of fitness in each chromosome by calculating an objective function, assigning a fitness score to every chromosome, which further determines the probability of reproduction, 3) selection: the best chromosome selected for the reproduction of offspring based on an individual's fitness value and passed on their genes to the next generation, 4) cross-

over: the genetic information of two parents is exchanged to produce a child, performed on pairs of parents that are randomly selected to create an offspring population of the same size as the parent population and 5) mutation: a random tweak in the chromosome to obtain a new solution and prevent premature convergence. When the operation of combination and reproduction are repeatedly used on strings or chromosomes in successive generations, the population of chromosomes or candidate solutions tends to become "homogeneous". The mutation operator helps the genetic algorithm to increase the "diversity" in the population of chromosomes or candidate solutions [33, 39, 41, 42].

ML algorithms

KNN: KNN is a simple and non-parametric algorithm for classifying objects based on closest training examples in the feature vector. K is a positive integer that refers to the number of nearest neighbors. If k=1, the KNN algorithm assigns the object to the class of its nearest neighbor [43, 44].

ANN: An ANN as a robust and flexible ML algorithm, which is based on the biological nervous systems, addresses unclear problems [45-47] with a mechanism as follows:

(1) Assigning weights to all the linkages to start the algorithm

(2) Using the inputs and linkages for the activation rate of hidden nodes

(3) Using the activation rate of hidden nodes and linkages to output, obtaining the activation rate of output nodes

(4) Obtaining the error rate at the output node and cascading down the error to hidden nodes

(5) Recalibrating the weights between the hidden nodes and the input nodes

(6) Repeating the process till the convergence

(7) Scoring the activation rate of the output nodes by the final linkage weights

SVM: The SVM classifier, based on the strategy of the maximal margin classifier, looks for the hyperplane that maximizes the border between those two classes with linear separation of two classes. For example, in Figure 2, the SVM classifier finds the best hyperplane (the red line) to maximize the distance between the nearest data samples of class A and class B [48]. This study used the SVM algorithm with the radial basis function (RBF) and linear kernels to predict mortality risk in hospitalized patients with COVID-19 [49].

DT: DT algorithm as a data mining algorithm with a top-down recursive method specifies the tree structure [35]. The flowchart-like structure of the DT algorithm includes nodes (root node and leaf node) and branches. Each

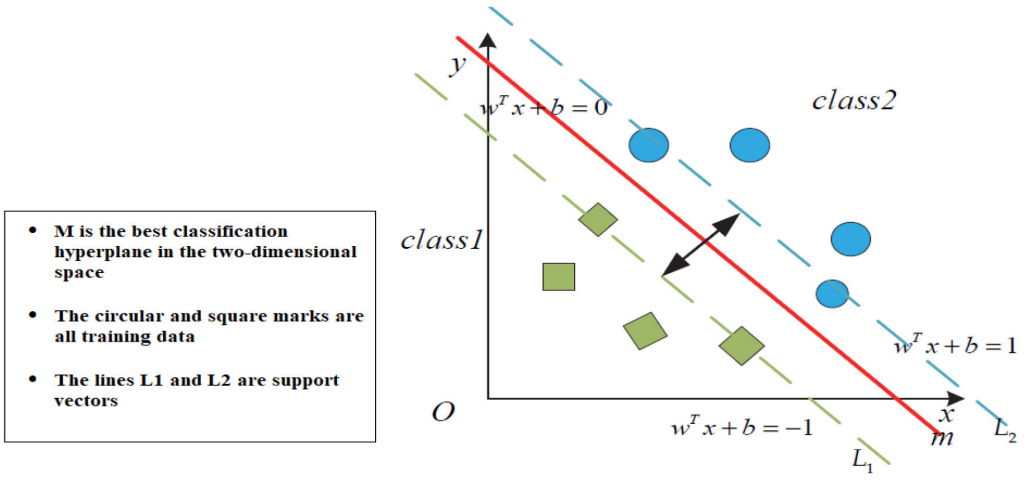

Figure 2: Classification of hyperplane in support vector machines (SVM) algorithm

node and branch indicate a feature and the value of the feature, respectively; however, the leave nodes indicate the classes.

Evaluation phase

The k-fold cross-validation method was used to evaluate and compare ML techniques for the prediction of COVID-19 mortality. Four evaluation metrics, including accuracy, sensitivity, specificity, and AUC, were used to compare ML models in predicting mortality in patients with COVID-19.

Results

Patient selection criteria

Information of 2082 patients was reviewed from the COVID-19 registry database of Ayatollah Taleghani Hospital, Abadan, Khuzestan, Iran, and 228 incomplete files with numerous missing data were removed from the analysis. Finally, the data of 1353 patients were studied (Figure 1).

Demographic and clinical characteristics of patients

In this study, 742 (54.85%) and 611 (45.15%) patients were male and female, respectively, with a median age of 57.25 (interquartile 18-100). Moreover, 298 (22.02%) were admitted to the Intensive Care Unit (ICU), and 1055 (77.98%) were hospitalized in general wards. A total of 1239 (91.57%) and 114 (8.43%) individuals were discharged in good condition and died, respectively. Tables 2 and 3 show descriptive statistics for the 1353 patients.

Simulation phase

The proposed hybrid ML techniques are investigated to classify and prioritize the clinical variables and mortality prediction. A total of 10 independent executions were on the dataset. The 10-fold cross-validation method was used to evaluate the classifiers. Adjusting parameters through the GA and other ML techniques are shown in Table 4; the used dataset

Table 2: Descriptive statistics of qualitative variables

Variable name	Frequencies (Values)
Blood type	27(A-); 552(A+)
	54(B-); 132(B+)
	49(O-); 421(O+)
	29(AB-); 89(AB+)
Gender	742(Male); 611(Female)
Cough	1058(+); 295(-)
Contusion	497(+); 856(-)
Nausea	459(+); 894(-)
Vomiting	396(+); 957(-)
Headache	340(+); 1013(-)
GI symptoms	300(+); 1153(-)
Muscular pain	661(+); 692(-)
Chill	666(+); 687(-)
Fever	706(+); 647(-)
Pneumonia	1135(+); 218(-)
Respiratory intubation	1122(+); 231(-)
Dyspnea	1178(+); 165(-)
Loss of taste	300(+); 1053(-)
Loss of smell	405(+); 948(-)
Runny noise	457(+); 196(-)
Sore throat	544(+); 809(-)
Other underlying diseases	763(+); 590(-)
CVD	406(+); 947(-)
Hypertension	495(+); 858(-)
Diabetes	1368(+); 985(-)
Smoking	69(+); 1284(-)
Alcohol consumption	139(+); 1214(-)
Addiction	37(+); 1316(-)
CRP	1163(+); 190(-)
Hypersensitive troponin	158(+); 1195(-)
Pleural effusion	514(+); 839(-)
Leukocytosis	610(+); 1743(-)
Mortality status (outcome)	114(+); 1239(-)

CVD: Cardiovascular disease, GI: Gastrointestinal, CRP: C-reactive protein

Table 3: Descriptive statistics of quantitative variables

Variable name	Range	Mean (SD)
Age (year)	18-100	57.25 (17.8)
Height (cm)	126-195	163.53 (7.5)
Weight (kg)	42-123	85.20 (11.3)
LOS	1-53	11(3.6)
Creatinine (mg/dL)	0.1-17.9	1.39 (1.4)
RBC count (mcL)	1.38-13.1	4.56 (0.9)
WBC count	1300-63000	8182.34 (4897.4)
Hematocrit	3.6-73.9	39.20 (6.7)
Hemoglobin	3.7-46	13.21 (2.4)
Platelet count	108000-691000	215493.66 (88380.1)
ALC	2-95	23.74 (11.8)
ANC	8-98	74.52 (12.3)
Calcium	0.9-14.1	9.68 (0.8)
Phosphorus	2-12.4	3.50 (0.5)
Magnesium	1.14-19.1	2.16 (0.6)
Sodium	37-157	137.94 (5.3)
Potassium	2.5-14.2	3.98 (0.7)
BUN	0.5-251	42.52 (31.7)
Total bilirubin	0.01-10	0.72 (0.7)
AST	3.8-924	44.45 (53.5)
ALT	2-672	38.29 (41.6)
Albumin	0.2-8.9	4.02 (0.5)
Glucose	18-994	136.09 (74.2)
LDH	4.6-6973	555.68 (339.0)
Activated PTT	1-120	28.56 (11.4)
PT	0.9-46.8	12.82 (1.9)
ALP	9.6-2846	213.12 (139.2)
ESR	2-258	40.65 (28.8)

LOS: Length of Stay, RBC: Red Blood Cell, WBC: White Blood Cell, ALC: Absolute Lymphocyte Count, ANC: Absolute Neutrophil Count, BUN: Blood Urea Nitrogen, AST: Aspartate Aminotransferase, ALT: Alanine Aminotransferase, LDH: Lactate Dehydrogenase, PTT: Partial Thromboplastin Time, PT: Prothrombin Time, ALP: Alkaline Phosphatase, ESR: Erythrocyte Sedimentation Rate, SD: Standard Deviation

contains 56 features in this study.

Results of feature selection

In this phase, the GA as a feature-selection method was used to identify the top predictors affecting the mortality of COVID-19 hospitalized patients. The GA algorithm was performed with different parameters in 10-independent iterative times on all the datasets. Some classification algorithms were used to measure the recital of each predictive model on the selected dataset. Finally, the most important predictors of mortality in COVID-19 patients were selected based on the comparison of the performance of several machine-learning techniques on the features selected

by the GA. Table 5 shows the most important variables for predicting mortality in patients with COVID-19.

Results of prediction models on selected features

In this phase, the features selected by the GA were tested on four prediction models with 10-fold cross-validation methods. Each model was repeated 10 iterations to better measure the performance of prediction models and the mean evaluation metrics: mean accuracy, mean specificity, and mean sensitivity.

Further, the mean, standard deviation, and

Table 4: Methods of adjusting parameters through genetic algorithm (GA) and other machine learning (ML) algorithms

Models	Parameters
GA	Population size=50, mutation probability rate (P_m)=0.3, crossover probability rate (P_c)=0.8, stop condition: maximum number of generations=100, number of independent executions=10
KNN	K=1, 3, 5
SVM	Kernel function = Gaussian, linear and RBF kernel
Decision tree	
ANN	57-10-5-2

GA: Genetic Algorithm, KNN: K-Nearest Neighbors Algorithm, SVM: Support Vector Machines, ANN: Artificial Neural Network

Table 5: Results of feature selection and 10-fold cross-validation prediction performance of used algorithms on selected features

Hybrid Classifier	Features selected		Accuracy (%)	Specificity (%)	Sensitivity (%)
GA-KNN	LOS, age, cough, respiratory intubation, dyspnea, CVD, leukocytosis, BUN, CRP, pleural effusion	**Mean±SD**	90.50±0.4	83.03±0.8	97.98±0.4
		MIN	89.99	81.98	96.86
		MAX	91.30	84.28	98.33
GA-DT	LOS, CVD, hypertension, hemoglobin, platelet count, ANC, pleural effusion	**Mean±SD**	82.6±0.5	81.17±0.9	84.17±0.7
		MIN	82.03	79.69	82.92
		MAX	84.02	82.92	85.64
GA-SVM	LOS, age, cough, respiratory intubation, dyspnea, CVD, leukocytosis, BUN, CRP, pleural effusion	**Mean±SD**	95.14±0.1	95.11±0.15	95.18 ±0.7
		MIN	94.03	93.09	94.23
		MAX	96.54	96.96	96.33
GA-NN	Age, CVD, hypertension, alcohol consumption hemoglobin, platelet count, ALC, ANC, BUN	**Mean±SD**	94.96±0.19	90.15±0.42	95.77±0.15
		MIN	94.70	89.51	92.42
		MAX	95.37	90.94	97.90
GA-Linear SVM	Age, CVD, dyspnea, platelet count, alcohol consumption, hemoglobin, ANC, CRP	**Mean±SD**	93.32±0.4	89.82±0.1	90.71±0.12
		MIN	87.14	86.35	89.75
		MAX	94.12	92.45	93.145
SVM-RBF	Age, CRP, pleural effusion, ALC, platelet count, and leukocytosis	**Mean±SD**	90.82±0.3	91.25 ±0.34	89.25±0.22
		MIN	86.92	91.47	93.74
		MAX	94.21	93.257	93.251

GA: Genetic Algorithm, KNN: K-Nearest Neighbor Algorithm, LOS: Length of Stay, CVD: Cardiovascular Disease, BUN: Blood Urea Nitrogen, CRP: C-Reactive Protein, DT: Decision Tree, ANC: Absolute Neutrophil Count, SVM: Support Vector Machines, NN: Neural Network, ALC: Absolute Lymphocyte Count, RBF: Radial Basis Function, SD: Standard Deviation

minimum and maximum values were measured in the selected dataset for accuracy, confusion matrix, and receiver operating characteristic (ROC) curve of ML models to predict mortality in the patients with COVID-19. Figure 3 illustrates the confusion matrix and ROC of all ML algorithms.

Ten features were selected based on the most positive correlation with the prediction of mortality in COVID-19 hospitalized patients. The results of feature selection and 10-fold cross-validation predictions are shown in Table 5.

Based on Table 5, when the selected features were included in the ML techniques in a total of 10 independent execution, the results show that the performance of the GA-SVM technique with the mean classification accuracy and mean specificity and mean sensitivity 95.14±0.1 and 95.11±0.15 and 95.18±0.7 had the best performance than that of other algorithms in predicting the mortality in COVID-19 hospitalized patients. The worst ML performance was observed for A total of 10 independent execution of the GA-DT hybrid with mean accuracy, mean specificity, and mean sensitivity, of 8.2674e+01, 8.1171e+01, and 8.4174e+0, respectively (Figure 4). The results of other algorithms in predicting the mortality in COVID-19 hospitalized patients on selected features are shown in Table 5.

Discussion

This study aimed to construct four ML-based prediction models for the prediction of mortality in COVID-19 hospitalized patients. The GA algorithm was used to optimize the best or most optimal subset of predictor variables. Four ML algorithms: KNN, DT, SVM, and ANN were trained based on selected features, and data balancing was performed by SMOTE over-sampling method. The findings show the SVM with the classification accuracy of 9.5147e+01 and specificity of 9.5112e+01 yielded the highest predictive performance among the developed ML techniques.

Feature selection is an important stage in preparing the data before training the model [48]. In the present study, 56 variables decreased to 10 by using GA. The selected features, include LOS, age, cough, respiratory intubation, dyspnea, CVD, leukocytosis, BUN, C-reactive protein, and pleural effusion.

Some studies are conducted on the application of hybrid ML methods in combination with

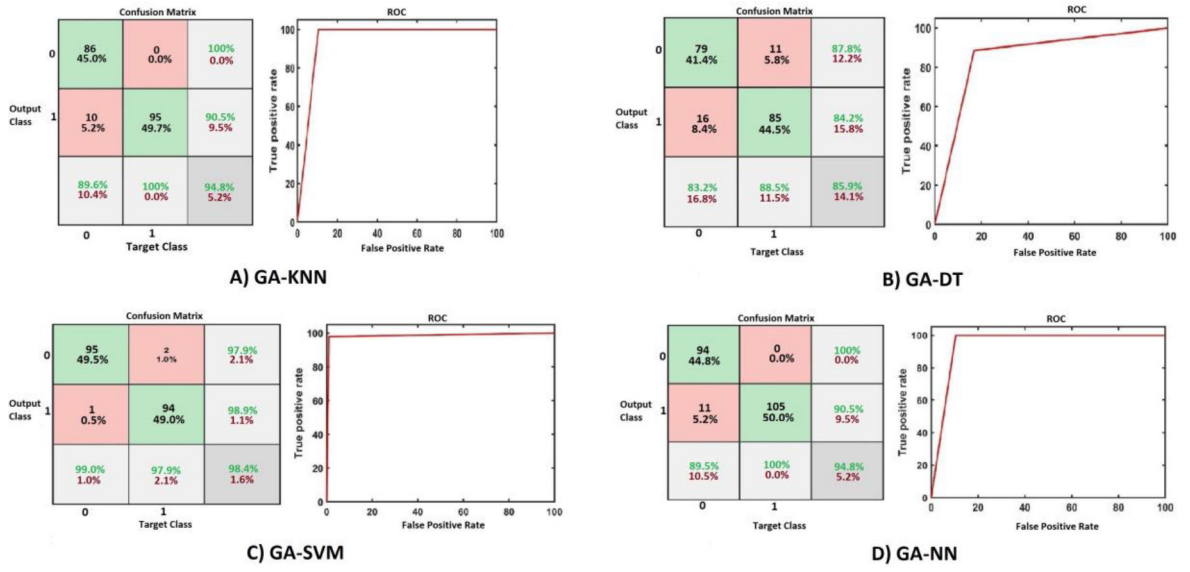

Figure 3: Confusion matrix and receiver operating characteristic (ROC) curve of four best implemented machine learning (ML) algorithms.

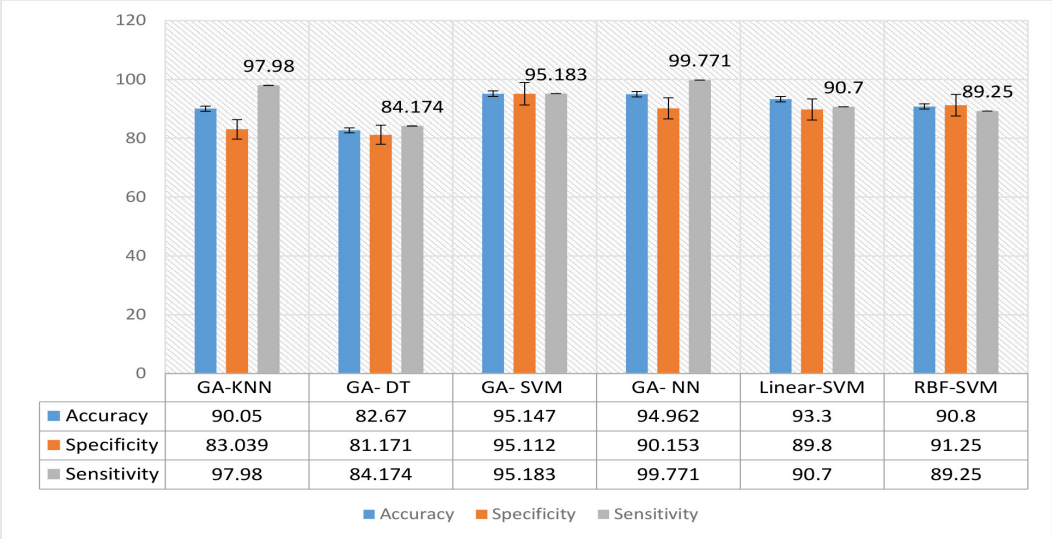

Figure 4: Comparing the performance of hybrid machine learning (ML) algorithms

GA to optimize the input variables, readjust the configuration of algorithms, and predict COVID-19-related outcomes. Monica et al. developed a hybrid ML-based model by using GA to find the optimal ensemble ANN configuration for COVID-19 prognosis and outcome prediction with 92% accuracy [34]. Sun et al. constructed a hybrid model using combined traditional backpropagation ANN and GA to optimize the input variables and improve the predictive performance effectively [49]. GA and convolutional neural network (CNN) were employed by Shukla et al. to design an automatic diagnostic model for predicting clinical deterioration and severity of the patients with COVID-19 based on chest X-ray images with good accuracy of 98.38% and 94.94% for training and testing, respectively [37]. Ghosh applied CNN-based models optimized by GA for diagnosing COVID-19 with optimal accuracy of 90.1% [50]. Albadr et al. used an optimized genetic algorithm-extreme learning machine (OGA-ELM) with three selection criteria, such as random, K-tournament, and roulette wheel to have prognoses of COVID-19 and predict severity and mortality risk using X-ray images with the accuracy of 100% [33]. Babukarthik et al. developed a hybrid model based on a genetic deep-learning convolutional neural network (GDCNN) for COVID-19 prediction with an accuracy of 98.84%, the precision of 93%, the sensitivity of 100%, and specificity of 97.0% [35]. Wang trained the two-hybrid intelligence models, including GA plus ANN and GA plus RF to classify clinical manifestations for COVID-19 severity prediction [51]. Shukla et al. proposed a COVID-19 diagnostic model based on multi-objective GA and CNN in chest X-ray images with an accuracy of 98.39% and 94.94% for training and testing, respectively [37]. Zivkovic proposed a new prediction model to predict the number of COVID-19 confirmed individuals based on the hybrid of adaptive neuro-fuzzy inference system and enhanced GA metaheuristics. Finally, they revealed that the suggested model outperformed other intelligent methods [52]. Doewes developed a COVID-19 analysis system using ensemble GA and ML classifiers with the accuracy, sensitivity, specificity, and AUC of 98.7%, 96.76%, 98.80%, and 92%, respectively [53].

As the above-reviewed studies showed the GA combination with the selected ML models can improve their performance. On the other hand, studies that used only ML models had

a performance lower than 90% in predicting the death of COVID-19 patients [54-59]. The present study also used some ML algorithms in combining with GA to predict mortality in COVID-19 hospitalized patients. The results showed that the GA-SVM algorithm was effective in the successful prediction of COVID-19 mortality with the accuracy of 9.5147e+01 and specificity of 9.5112e+01.

In the prior studies, the most important variables affecting COVID-19 mortality were extracted by ML-based [33-35, 37, 49, 51, 60] and clinical-based [4, 5, 32, 55, 61-66] techniques. The selected top variables in predicting COVID-19 mortality in the reviewed ML-based studies [33-35, 37, 49, 51, 60] optimized by GA were advanced age, longer LOS, decreased Oxygen saturation ($SPO_{2)}$ leukocytosis, raised C-reactive protein, and cardiovascular diseases.

On the other hand, many studies have been conducted to select the most significant variables for predicting COVID-19 mortality from a clinical perspective. In these studies, the top 10 predictors or effective factors for the mortality of COVID-19 patients are advanced age (older age) [2-6, 55, 63, 65, 67], longer LOS [1-3, 6, 65], mechanical ventilation [4, 7, 55, 61-63], fever [1, 2, 6, 55, 61, 62, 65], decreased SPO_2 (low oxygen saturation) [35, 49, 51, 60], elevated interlukin-6 [4, 5, 55, 61-65], high blood pressure [2, 4-6, 8, 55, 63, 64], leukocytosis [1, 4, 7, 8, 61, 63, 64], increased BUN [4, 5, 55, 61-65], cardiovascular [1, 2, 4-6, 8, 55, 61-65], and COPD [4, 6, 8, 61, 63-65]. The results of categorizing and ranking features in reviewed studies are consistent with those of 10 executions from the GA-SVM algorithm in the current study.

In the present study, the GA algorithm was utilized to address the optimization of the predictive variables and "the curse of dimensionality", which are considered one of the greatest challenges in ML models. According to the results, the GA as a powerful optimizer can select the best subset features in the ML algorithms.

The predictive models showed more promising performance than a single model by hybridizing different ML algorithms, constructing complex models, and extracting appropriate features. A valuable set of features leads to predicting the adequately acceptable performance of ML algorithms. However, the dataset is often insufficient or imbalanced in specific applications. Therefore, training algorithms and good results are vital based on the most relevant set of features.

The present study is important due to two reasons, as follows: 1) providing high-risk and important mortality predictors and 2) providing a simple and fast clinical screening tool to accurately predict the risk of death in COVID-19 patients. In the present study, the predictive models can support the treatment team's decision-making for the triage (prioritization) of COVID-19 patients based on the risk of death, without waiting for other clinical tests. Therefore, the proposed models can effectively triage (prioritize) patients in situations, in which time loss is important and in centers with limited resources.

The ML algorithms potentially have many advantages for the healthcare providers involved in the treatment of COVID-19 patients, and the trained ML methods can predict the death of COVID-19 patients with optimal performance [68, 69]. The developed models can help medical resources for deteriorating individuals, increasing the quality of care, and reducing medical faults due to exhaustion and working long hours in the ICU during the pandemic [70, 71]. Thus, ML-based prediction models can significantly contribute to triaging hazardous patients and allocating the limited hospital resources for mortality risk prediction [72, 73], resulting in reducing uncertainty by quantitative, objective, and evidence-based models for risk classification. Furthermore, the ML provides a better strategy for physicians to reduce complications and improve patient survival [74-77].

This study is conducted with some limitations, as follows: 1) training only four ML models, 2) disregarding imaging variables; more effective factors along with more ML models should be used to predict the mortality of COVID-19 patients, 3) dealing with a retrospective-single center dataset, and 4) the low quality (imbalanced, noisy, duplicates, and meaningless values), insufficient quantity (missing cells), and non-optimal generalizability of data in the selected database. In the current study, noises, duplicates, and meaningless records manually as much as possible from the dataset were firstly removed. The SMOTE method was used to minimize the bias by class balancing and address the problem of the unbalanced dataset. A dataset with a greater sample size should be applied in multi-center settings in future studies.

However, the predictability of ML models increased using a hybrid approach for accurate selection of the most effective features and conduction of an effective training process, the use of the proposed model is recommended for predictive analysis of sensitive, complex, and ambiguous conditions affecting public health, safety, and welfare, such as COVID-19. Due to the use of a precise approach for feature selection and data reduction, the proposed hybrid model can provide effectively predictive capabilities based on more data from multi-center settings during a longer period using training more ML algorithms.

Conclusion

In this study, a feature selection method was applied using GA to identify the key features affecting COVID-19 mortality. Further, this study aimed to investigate some predictive models for COVID-19 mortality in hospitalized patients and select the most important features via GA. In this study, diverse prediction models were evaluated, and experiments were performed to select the finest ML algorithms for the prediction of COVID-19 mortality. Four hybrid classifiers, i.e., GA-KNN, GA-DT, GA-SVM, and GA-ANN were used for prediction. The GA-SVM classifier performance had more predictive abilities than the other three hybrid ML techniques. Based on the GA feature selection, the most important attributes affect COVID-19 severity and mortality. The GA with prediction models improved the performances of the proposed models.

Acknowledgment

We thank the research deputy of Abadan University of Medical Sciences for financially supporting this project. We also would like to thank officials at Ayatollah Taleghani Hospital who assisted the research team in conducting this work.

Authors' Contribution

H. Kazemi-Arpanahi conceived the idea. The introduction section of the paper was written by H. Kazemi-Arpanahi and MR. Afrash. M. Shanbehzadeh gathered the images and the related literature as well as helped with writing the related works. The method was implemented by H. Kazemi-Arpanahi and M. Shanbehzadeh. Results and analyses were carried out by MR. Afrash. The research work was proofread and supervised by H. Kazemi-Arpanahi. All the authors read, modified, and approved the final version of the manuscript.

Ethical Approval

This study was approved by Abadan University of Medical Sciences with the code number: IR.ABADANUMS.REC.1400.017.

Informed consent

Before the study, all participants were informed about the aim of the study and signed the consent form. In addition, the confidentiality of the personal and research data was ensured.

Conflict of Interest

None

References

1. Erfannia L, Sharifian R, Yazdani A, Sarsarshahi A, Rahati R, Jahangiri S. Students' Satisfaction and e-

Learning Courses in Covid-19 Pandemic Era: A Case Study. *Stud Health Technol Inform*. 2022;**289**:180-3. doi: 10.3233/SHTI210889. PubMed PMID: 35062122.

2. Kashefizadeh A, Ohadi L, Golmohammadi M, Araghi F, Dadkhahfar S, Kiani A, et al. Clinical features and short-term outcomes of COVID-19 in Tehran, Iran: An analysis of mortality and hospital stay. *Acta Biomed*. 2020;**91**(4):e2020147. doi: 10.23750/abm.v91i4.10206. PubMed PMID: 33525218. PubMed PMCID: PMC7927529.
3. Muhammad R, Ogunti R, Ahmad B, Munawar M, Donaldson S, Sumon M, et al. Clinical Characteristics and Predictors of Mortality in Minority Patients Hospitalized with COVID-19 Infection. *J Racial Ethn Health Disparities*. 2022;**9**(1):335-45. doi: 10.1007/s40615-020-00961-x. PubMed PMID: 33538998. PubMed PMCID: PMC7861003.
4. Efeoglu Sacak M, Karacabey S, Sanri E, Omercikoglu S, Ünal E, Ecmel Onur Ö, et al. Variables Affecting Mortality Among COVID-19 Patients With Lung Involvement Admitted to the Emergency Department. *Cureus*. 2021;**13**(1):e12559. doi: 10.7759/cureus.12559. PubMed PMID: 33564550. PubMed PMCID: PMC7863081.
5. Aly MH, Rahman SS, Ahmed WA, Alghamedi MH, Al Shehri AA, Alkalkami AM, Hassan MH. Indicators of Critical Illness and Predictors of Mortality in COVID-19 Patients. *Infect Drug Resist*. 2020;**13**:1995-2000. doi: 10.2147/IDR.S261159. PubMed PMID: 32617010. PubMed PMCID: PMC7326396.
6. Bhargava A, Szpunar SM, Sharma M, Fukushima EA, Hoshi S, Levine M, et al. Clinical Features and Risk Factors for In-Hospital Mortality From COVID-19 Infection at a Tertiary Care Medical Center, at the Onset of the US COVID-19 Pandemic. *J Intensive Care Med*. 2021;**36**(6):711-8. doi: 10.1177/08850666211001799. PubMed PMID: 33759606. PubMed PMCID: PMC7992101.
7. Lai X, Liu J, Zhang T, Feng L, Jiang P, Kang L, et al. Clinical, laboratory and imaging predictors for critical illness and mortality in patients with COVID-19: protocol for a systematic review and meta-analysis. *BMJ Open*. 2020;**10**(12):e039813. doi: 10.1136/bmjopen-2020-039813. PubMed PMID: 33361074. PubMed PMCID: PMC7768616.
8. Moon SS, Lee K, Park J, Yun S, Lee YS, Lee DS. Clinical Characteristics and Mortality Predictors of COVID-19 Patients Hospitalized at Nationally-Designated Treatment Hospitals. *J Korean Med Sci*. 2020;**35**(36):e328. doi: 10.3346/jkms.2020.35.e328. PubMed PMID: 32924343. PubMed PMCID: PMC7490203.
9. Erfannia L, Amraei M, Arji G, Yazdani A, Sabzehgar M, Yaghoobi L. Reviewing and Content Analysis of Persian Language Mobile Health Apps for COVID-19 Management. Stud Health Technol Inform. 2022;**289**:106-9. doi: 10.3233/SHTI210870. PubMed PMID: 35062103.
10. Muhiyaddin R, Abd-Alrazaq AA, Househ M, Alam T, Shah Z. The Impact of Clinical Decision Support Systems (CDSS) on Physicians: A Scoping Review. *Stud Health Technol Inform*. 2020;**272**:470-3. doi: 10.3233/SHTI200597. PubMed PMID: 32604704.
11. Shahnazi H, Ahmadi-Livani M, Pahlavanzadeh B, Rajabi A, Hamrah MS, Charkazi A. Assessing preventive health behaviors from COVID-19: a cross sectional study with health belief model in Golestan Province, Northern of Iran. *Infect Dis Poverty*. 2020;**9**(1):157. doi: 10.1186/s40249-020-00776-2. PubMed PMID: 33203453. PubMed PMCID: PMC7671178.
12. Heydari MR, Joulaei H, Zarei N, Fararouei M, Gheibi Z. An Online Investigation of Knowledge and Preventive Practices in Regard to COVID-19 in Iran. *Health Lit Res Pract*. 2021;**5**(1):e15-23. doi: 10.3928/24748307-20201130-01. PubMed PMID: 33433629. PubMed PMCID: PMC7801262.
13. Yan L, Zhang H-T, Goncalves J, Xiao Y, Wang M, Guo Y, et al. An interpretable mortality prediction model for COVID-19 patients. *Nat Mach Intell*. 2020;**2**(5):283-88. doi: 10.1038/s42256-020-0180-7.
14. Hu H, Yao N, Qiu Y. Comparing Rapid Scoring Systems in Mortality Prediction of Critically Ill Patients With Novel Coronavirus Disease. *Acad Emerg Med*. 2020;**27**(6):461-8. doi: 10.1111/acem.13992. PubMed PMID: 32311790. PubMed PMCID: PMC7264631.
15. Rukmani P, Vergin Raja Sarobin M, Graceline Jasmine S, Jani Anbarasi L, Mishra P. Usage of artificial intelligence to prevent and regulate covid-19. *Int J Cur Res Rev*. 2021;**13**(6):S64-7. doi: 10.31782/IJCRR.2021.SP181.
16. Senthilraja M. Application of Artificial Intelligence to Address Issues Related to the COVID-19 Virus. *SLAS Technol*. 2021;**26**(2):123-6. doi: 10.1177/2472630320983813. PubMed PMID: 33390088. PubMed PMCID: PMC8958383.
17. Mei X, Lee HC, Diao KY, Huang M, Lin B, Liu C, et al. Artificial intelligence-enabled rapid diagnosis of patients with COVID-19. *Nat Med*. 2020;**26**(8):1224-8. doi: 10.1038/s41591-020-0931-3. PubMed PMID: 32427924. PubMed PMCID: PMC7446729.
18. Dhamodharavadhani S, Rathipriya R, Chatterjee JM. COVID-19 Mortality Rate Prediction for India Using Statistical Neural Network Models. *Front Public Health*. 2020;**8**:441. doi: 10.3389/fpubh.2020.00441. PubMed PMID: 32984242. PubMed PMCID: PMC7485390.
19. Gao Y, Cai GY, Fang W, Li HY, Wang SY, Chen L, et al. Machine learning based early warning system enables accurate mortality risk prediction for COVID-19. *Nat Commun*. 2020;**11**(1):5033. doi: 10.1038/s41467-020-18684-2. PubMed PMID: 33024092. PubMed PMCID: PMC7538910.
20. Tortajada-Goitia B, Morillo-Verdugo R, Margusino-

Framiñán L, Marcos JA, Fernández-Llamazares CM. Survey on the situation of telepharmacy as applied to the outpatient care in hospital pharmacy departments in Spain during the COVID-19 pandemic. *Farm Hosp*. 2020;**44**(4):135-40. doi: 10.7399/fh.11527. PubMed PMID: 32646343.

21. Guo Q, He Z. Prediction of the confirmed cases and deaths of global COVID-19 using artificial intelligence. *Environ Sci Pollut Res Int.* 2021;**28**(9):11672-82. doi: 10.1007/s11356-020-11930-6. PubMed PMID: 33415612. PubMed PMCID: PMC7789896.
22. Gupta VK, Gupta A, Kumar D, Sardana A. Prediction of COVID-19 confirmed, death, and cured cases in India using random forest model. *Big Data Mining and Analytics*. 2021;**4**(2):116-23. doi: 10.26599/BDMA.2020.9020016.
23. Jarndal A, Husain S, Zaatar O, Gumaei TA, Hamadeh A. GPR and ANN based Prediction Models for COVID-19 Death Cases. International Conference on Communications, Computing, Cybersecurity, and Informatics (CCCI); Sharjah, United Arab Emirates: IEEE; 2020. doi: 10.1109/CCCI49893.2020.9256564.
24. Yazdani A, Zahmatkeshan M, Ravangard R, Sharifian R, Shirdel M. Supervised Machine Learning Approach to COVID-19 Detection Based on Clinical Data. *Med J Islam Repub Iran.* 2022;**36**:110. doi: 10.47176/mjiri.36.110.
25. Rasjid ZE, Setiawan R, Effendi A. A Comparison: Prediction of Death and Infected COVID-19 Cases in Indonesia Using Time Series Smoothing and LSTM Neural Network. *Procedia Comput Sci*. 2021;**179**:982-8. doi: 10.1016/j.procs.2021.01.102. PubMed PMID: 33643496. PubMed PMCID: PMC7894084.
26. Li S, Lin Y, Zhu T, Fan M, Xu S, Qiu W, et al. Development and external evaluation of predictions models for mortality of COVID-19 patients using machine learning method. *Neural Comput Appl*. 2021:1-10. doi: 10.1007/s00521-020-05592-1. PubMed PMID: 33424133. PubMed PMCID: PMC7783503.
27. Sobrinho A, Queiroz ACDS, Da Silva LD, Costa EDB, Pinheiro ME, Perkusich A. Computer-aided diagnosis of chronic kidney disease in developing countries: A comparative analysis of machine learning techniques. *IEEE Access.* 2020;**8**:25407-19. doi: 10.1109/ACCESS.2020.2971208.
28. Ma X, Li A, Jiao M, Shi Q, An X, Feng Y, et al. Characteristic of 523 COVID-19 in Henan Province and a Death Prediction Model. *Front Public Health*. 2020;**8**:475. doi: 10.3389/fpubh.2020.00475. PubMed PMID: 33014973. PubMed PMCID: PMC7506160.
29. Kannan R, Wang IZW, Ong HB, Ramakrishnan K, Alamsyah A. COVID-19 impact: Customised economic stimulus package recommender system using machine learning techniques. *F1000Res.* 2021;**10**:932. doi: 10.12688/f1000research.72976.2. PubMed PMID: 34925768. PubMed PMCID: PMC8647044.
30. Sun L, Mo Z, Yan F, Xia L, Shan F, Ding Z, et al. Adaptive Feature Selection Guided Deep Forest for COVID-19 Classification With Chest CT. *IEEE J Biomed Health Inform*. 2020;**24**(10):2798-805. doi: 10.1109/JBHI.2020.3019505. PubMed PMID: 32845849.
31. Subramani P, Srinivas K, Kavitha Rani B, Sujatha R, Parameshachari BD. Prediction of muscular paralysis disease based on hybrid feature extraction with machine learning technique for COVID-19 and post-COVID-19 patients. *Pers Ubiquitous Comput.* 2021:1-14. doi: 10.1007/s00779-021-01531-6. PubMed PMID: 33679282. PubMed PMCID: PMC7926078.
32. Wan TK, Huang RX, Tulu TW, Liu JD, Vodencarevic A, et al. Identifying Predictors of COVID-19 Mortality Using Machine Learning. *Life (Basel).* 2022;**12**(4):547. doi: 10.3390/life12040547. PubMed PMID: 35455038. PubMed PMCID: PMC9028639.
33. Abdullah AS, Ramya C, Priyadharsini V, Reshma C, Selvakumar S. A survey on evolutionary techniques for feature selection. Conference on Emerging Devices and Smart Systems (ICEDSS); Mallasamudram, India: IEEE; 2017. p. 57-62. doi: 10.1109/ICEDSS.2017.8073659.
34. Mónica JC, Melin P, Sánchez D. Genetic Optimization of Ensemble Neural Network Architectures for Prediction of COVID-19 Confirmed and Death Cases. In: Castillo O, Melin P. Fuzzy Logic Hybrid Extensions of Neural and Optimization Algorithms: Theory and Applications. Springer. 2021. p. 85-98.
35. Babukarthik RG, Adiga VAK, Sambasivam G, Chandramohan D, Amudhavel J. Prediction of COVID-19 Using Genetic Deep Learning Convolutional Neural Network (GDCNN). *IEEE Access.* 2020;**8**:177647-177666. doi: 10.1109/ACCESS.2020.3025164. PubMed PMID: 34786292. PubMed PMCID: PMC8545287.
36. Pan P, Li Y, Xiao Y, Han B, Su L, Su M, et al. Prognostic Assessment of COVID-19 in the Intensive Care Unit by Machine Learning Methods: Model Development and Validation. *J Med Internet Res*. 2020;**22**(11):e23128. doi: 10.2196/23128. PubMed PMID: 33035175. PubMed PMCID: PMC7661105.
37. Shukla PK, Sandhu JK, Ahirwar A, Ghai D, Maheshwary P, Shukla PK. Multiobjective Genetic Algorithm and Convolutional Neural Network Based COVID-19 Identification in Chest X-Ray Images. *Math Probl Eng.* 2021;**2021**:1-9. doi: 10.1155/2021/7804540.
38. Aggarwal D, Bali V, Mittal S. An insight into machine learning techniques for Predictive Analysis and Feature Selection. *International Journal of Innovative Technology and Exploring Engineering (IJITEE).* 2019;**8**(9S):342-9. doi: 10.35940/ijitee.I1055.0789S19.
39. Curtis F, Li X, Rose T, Vázquez-Mayagoitia Á, Bhattacharya S, Ghiringhelli LM, Marom N. GAtor: A First-Principles Genetic Algorithm for Molecular Crystal Structure Prediction. *J Chem Theory Comput*. 2018;**14**(4):2246-64. doi: 10.1021/acs.jctc.7b01152.

PubMed PMID: 29481740.

40. Kim NH, Yang DW, Choi SH, Kang SW. Machine Learning to Predict Brain Amyloid Pathology in Pre-dementia Alzheimer's Disease Using QEEG Features and Genetic Algorithm Heuristic. *Front Comput Neurosci*. 2021;**15**:755499. doi: 10.3389/fncom.2021.755499. PubMed PMID: 34867252. PubMed PMCID: PMC8632633.
41. Yousefpour A, Jahanshahi H, Bekiros S. Optimal policies for control of the novel coronavirus disease (COVID-19) outbreak. *Chaos Solitons Fractals*. 2020;**136**:109883. doi: 10.1016/j.chaos.2020.109883. PubMed PMID: 32427205. PubMed PMCID: PMC7229919.
42. Guo L. Protection and Inheritance of Traditional Culture in Urbanization Construction Based on Genetic Algorithm under the Concept of Environmental Protection. *J Environ Public Health*. 2022;**2022**:5844732. doi: 10.1155/2022/5844732. PubMed PMID: 35755884. PubMed PMCID: PMC9225901.
43. Medjahed SA, Saadi TA, Benyettou A. Breast cancer diagnosis by using k-nearest neighbor with different distances and classification rules. *International Journal of Computer Applications*. 2013;**62**(1):1-5.
44. Odajima K, Pawlovsky AP. A detailed description of the use of the kNN method for breast cancer diagnosis. 7th International Conference on Biomedical Engineering and Informatics; Dalian, China: IEEE; 2014: doi: 10.1109/BMEI.2014.7002861.
45. Zhang Q, Deng D, Dai W, Li J, Jin X. Optimization of culture conditions for differentiation of melon based on artificial neural network and genetic algorithm. *Sci Rep*. 2020;**10**(1):3524. doi: 10.1038/s41598-020-60278-x. PubMed PMID: 32103071. PubMed PMCID: PMC7044330.
46. Afrash MR, Khalili M, Salekde MS. A comparison of data mining methods for diagnosis and prognosis of heart disease. *International Journal of Advanced Intelligence Paradigms*. 2020;**16**(1):88-97.
47. Zou J, Han Y, So SS. Overview of artificial neural networks. *Methods Mol Biol*. 2008;**458**:15-23. doi: 10.1007/978-1-60327-101-1_2. PubMed PMID: 19065803.
48. Vivekanandan T, Sriman Narayana Iyengar NC. Optimal feature selection using a modified differential evolution algorithm and its effectiveness for prediction of heart disease. *Comput Biol Med*. 2017;**90**:125-36. doi: 10.1016/j.compbiomed.2017.09.011. PubMed PMID: 28987988.
49. Sun F, Su J, Hu Y, Zhou L. The Prediction of New Medical Resources in China during COVID-19 Epidemic Period Based on Artificial Neural Network Model Optimized by Genetic Algorithm. *J Phys: Conf Ser*. 2021;**1815**:012033. doi: 10.1088/1742-6596/1815/1/012033.
50. Zeng Z, Wang B, Zhao Z. Research on CNN-based Models Optimized by Genetic Algorithm and Application in the Diagnosis of Pneumonia and COVID-19. *MedRxiv*. 2020. doi: 10.1101/2020.04.21.20072637.
51. Wang RY, Guo TQ, Li LG, Jiao JY, Wang LY, et al. Predictions of COVID-19 Infection Severity Based on Co-associations between the SNPs of Co-morbid Diseases and COVID-19 through Machine Learning of Genetic Data. 8th International Conference on Computer Science and Network Technology (ICCSNT); Dalian, China: IEEE; 2020. doi: 10.1109/ICCSNT50940.2020.9304990.
52. Zivkovic M, Bacanin N, Djordjevic A, Antonijevic M, Strumberger I, Rashid TA, editors. Hybrid genetic algorithm and machine learning method for covid-19 cases prediction. Proceedings of International Conference on Sustainable Expert Systems; Singapore: Springer; 2021. p. 169-84.
53. Doewes RI, Nair R, Sharma T. Diagnosis of COVID-19 through blood sample using ensemble genetic algorithms and machine learning classifier. *World Journal of Engineering*. 2021;**19**(2):175-82. doi: 10.1108/WJE-03-2021-0174.
54. Agieb R. Machine learning models for the prediction the necessity of resorting to icu of covid-19 patients. *IJATCSE*. 2020;**9**(5):6980-84.
55. Yadaw AS, Li YC, Bose S, Iyengar R, Bunyavanich S, Pandey G. Clinical features of COVID-19 mortality: development and validation of a clinical prediction model. *Lancet Digit Health*. 2020;**2**(10):e516-25. doi: 10.1016/S2589-7500(20)30217-X. PubMed PMID: 32984797. PubMed PMCID: PMC7508513.
56. Zhao Z, Chen A, Hou W, Graham JM, Li H, Richman PS, et al. Prediction model and risk scores of ICU admission and mortality in COVID-19. *PLoS One*. 2020;**15**(7):e0236618. doi: 10.1371/journal.pone.0236618. PubMed PMID: 32730358. PubMed PMCID: PMC7392248.
57. Vaid A, Jaladanki SK, Xu J, Teng S, Kumar A, Lee S, et al. Federated Learning of Electronic Health Records to Improve Mortality Prediction in Hospitalized Patients With COVID-19: Machine Learning Approach. *JMIR Med Inform*. 2021;**9**(1):e24207. doi: 10.2196/24207. PubMed PMID: 33400679. PubMed PMCID: PMC7842859.
58. Booth AL, Abels E, McCaffrey P. Development of a prognostic model for mortality in COVID-19 infection using machine learning. *Mod Pathol*. 2021;**34**(3):522-31. doi: 10.1038/s41379-020-00700-x. PubMed PMID: 33067522. PubMed PMCID: PMC7567420.
59. Parchure P, Joshi H, Dharmarajan K, Freeman R, Reich DL, Mazumdar M, et al. Development and validation of a machine learning-based prediction model for near-term in-hospital mortality among patients with COVID-19. *BMJ Support Palliat Care*. 2022;**12**:e424-31. doi: 10.1136/bmjspcare-2020-002602. PubMed PMID: 32963059. PubMed PMCID: PMC8049537.
60. Albadr MAA, Tiun S, Ayob M, Al-Dhief FT, Omar K, Hamzah FA. Optimised genetic algorithm-extreme

learning machine approach for automatic COVID-19 detection. *PLoS One.* 2020;**15**(12):e0242899. doi: 10.1371/journal.pone.0242899. PubMed PMID: 33320858. PubMed PMCID: PMC7737907.

61. Abraha HE, Gessesse Z, Gebrecherkos T, Kebede Y, Weldegiargis AW, Tequare MH, et al. Clinical features and risk factors associated with morbidity and mortality among patients with COVID-19 in northern Ethiopia. *Int J Infect Dis.* 2021;**105**:776-83. doi: 10.1016/j.ijid.2021.03.037. PubMed PMID: 33741488. PubMed PMCID: PMC7962557.
62. Chilimuri S, Sun H, Alemam A, Mantri N, Shehi E, Tejada J, Yugay A, Nayudu SK. Predictors of Mortality in Adults Admitted with COVID-19: Retrospective Cohort Study from New York City. *West J Emerg Med.* 2020;**21**(4):779-84. doi: 10.5811/westjem.2020.6.47919. PubMed PMID: 32726241. PubMed PMCID: PMC7390589.
63. Akcay M, Etiz D, Celik O. Prediction of Survival and Recurrence Patterns by Machine Learning in Gastric Cancer Cases Undergoing Radiation Therapy and Chemotherapy. *Adv Radiat Oncol.* 2020;**5**(6):1179-87. doi: 10.1016/j.adro.2020.07.007. PubMed PMID: 33305079. PubMed PMCID: PMC7718548.
64. Homayounieh F, Zhang EW, Babaei R, Karimi Mobin H, Sharifian M, Mohseni I, et al. Clinical and imaging features predict mortality in COVID-19 infection in Iran. *PLoS One.* 2020;**15**(9):e0239519. doi: 10.1371/journal.pone.0239519. PubMed PMID: 32970733. PubMed PMCID: PMC7514030.
65. Moledina SM, Maini AA, Gargan A, Harland W, Jenney H, Phillips G, Thomas K, Chauhan D, Fertleman M. Clinical Characteristics and Predictors of Mortality in Patients with COVID-19 Infection Outside Intensive Care. *Int J Gen Med*. 2020;**13**:1157-65. doi: 10.2147/IJGM.S271432. PubMed PMID: 33244256. PubMed PMCID: PMC7683500.
66. Van Lissa CJ, Stroebe W, VanDellen MR, Leander NP, Agostini M, Draws T, et al. Using machine learning to identify important predictors of COVID-19 infection prevention behaviors during the early phase of the pandemic. *Patterns (N Y).* 2022; **3**(4):100482. doi: 10.1016/j.patter.2022.100482. PubMed PMID: 35282654. PubMed PMCID: PMC8904175.
67. Gayam V, Chobufo MD, Merghani MA, Lamichhane S, Garlapati PR, Adler MK. Clinical characteristics and predictors of mortality in African-Americans with COVID-19 from an inner-city community teaching hospital in New York. *J Med Virol*. 2021;**93**(2):812-9. doi: 10.1002/jmv.26306. PubMed PMID: 32672844. PubMed PMCID: PMC7405133.
68. Ozdas A, Miller R. Care provider order entry (CPOE): a perspective on factors leading to success or to failure. *Yearb Med Inform*. 2007:128-37. PubMed PMID: 17700915.
69. Karthikeyan A, Garg A, Vinod PK, Priyakumar UD. Machine Learning Based Clinical Decision Support System for Early COVID-19 Mortality Prediction. *Front Public Health*. 2021;**9**:626697. doi: 10.3389/fpubh.2021.626697. PubMed PMID: 34055710. PubMed PMCID: PMC8149622.
70. Aljouie AF, Almazroa A, Bokhari Y, Alawad M, Mahmoud E, Alawad E, et al. Early Prediction of COVID-19 Ventilation Requirement and Mortality from Routinely Collected Baseline Chest Radiographs, Laboratory, and Clinical Data with Machine Learning. *J Multidiscip Healthc*. 2021;**14**:2017-33. doi: 10.2147/JMDH.S322431. PubMed PMID: 34354361. PubMed PMCID: PMC8331117.
71. Domínguez-Olmedo JL, Gragera-Martínez Á, Mata J, Pachón Álvarez V. Machine Learning Applied to Clinical Laboratory Data in Spain for COVID-19 Outcome Prediction: Model Development and Validation. *J Med Internet Res*. 2021;**23**(4):e26211. doi: 10.2196/26211. PubMed PMID: 33793407. PubMed PMCID: PMC8048712.
72. Burdick H, Lam C, Mataraso S, Siefkas A, Braden G, Dellinger RP, et al. Prediction of respiratory decompensation in Covid-19 patients using machine learning: The READY trial. *Comput Biol Med.* 2020;**124**:103949. doi: 10.1016/j.compbiomed.2020.103949. PubMed PMID: 32798922. PubMed PMCID: PMC7410013.
73. Cobre AF, Stremel DP, Noleto GR, Fachi MM, Surek M, Wiens A, et al. Diagnosis and prediction of COVID-19 severity: can biochemical tests and machine learning be used as prognostic indicators? *Comput Biol Med*. 2021;**134**:104531. doi: 10.1016/j.compbiomed.2021.104531. PubMed PMID: 34091385. PubMed PMCID: PMC8164361.
74. Lv H, Shi L, Berkenpas JW, Dao FY, Zulfiqar H, Ding H, et al. Application of artificial intelligence and machine learning for COVID-19 drug discovery and vaccine design. *Brief Bioinfo*m. 2021;**22**(6):bbab320. doi: 10.1093/bib/bbab320. PubMed PMID: 34410360. PubMed PMCID: PMC8511807.
75. Solanki YS, Chakrabarti P, Jasinski M, Leonowicz Z, Bolshev V, Vinogradov A, et al. A Hybrid Supervised Machine Learning Classifier System for Breast Cancer Prognosis Using Feature Selection and Data Imbalance Handling Approaches. *Electronics*. 2021;**10**(6):699-708. doi: 10.3390/electronics10060699.
76. Afrash MR, Erfanniya L, Amraei M, Mehrabi N, Jelvay S, Nopour R, Shanbehzadeh M. Machine Learning-Based Clinical Decision Support System for automatic diagnosis of COVID-19 based on the routine blood test. *JBE*. 2022;**8**(1):77-89. doi: 10.18502/jbe.v8i1.10407.
77. Afrash MR, Kazemi-Arpanahi H, Nopour R, Tabatabaei ES, Shanbehzadeh M. Proposing an Intelligent Monitoring System for Early Prediction of Need for Intubation among COVID-19 Hospitalized Patients. *Journal of Environmental Health and Sustainable Development*. 2022;**7**(3):1698-707. doi: 10.18502/jehsd.v7i3.10719.