Heliyon 10 (2024) e35561

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Identifying factors related to mortality of hospitalized COVID-19 patients using machine learning methods

Farzaneh Hamidi [a], Hadi Hamishehkar [b,c], Pedram Pirmad Azari Markid [d], Parvin Sarbakhsh [e,*]

[a] Department of Biostatistics, Faculty of Medical Sciences, Tarbiat Modares University, Tehran, Iran
[b] Clinical Research Development Unit of Imam Reza Hospital, Tabriz University of Medical Sciences, Tabriz, Iran
[c] Drug Applied Research Center, Tabriz University of Medical Sciences, Tabriz, Iran
[d] Department of Clinical Pharmacy, Tabriz University of Medical Sciences, Tabriz, Iran
[e] Health and Environment Research Center, Tabriz University of Medical Sciences, Tabriz, Iran

ARTICLE INFO

Keywords:
COVID-19
Mortality
Machine learning
LASSO
Elastic net
Artificial neural network
Feature selection

ABSTRACT

Background: The COVID-19 pandemic has had a profound impact globally, presenting significant social and economic challenges. This study aims to explore the factors affecting mortality among hospitalized COVID-19 patients and construct a machine learning-based model to predict the risk of mortality.

Methods: The study examined COVID-19 patients admitted to Imam Reza Hospital in Tabriz, Iran, between March 2020 and November 2021. The Elastic Net method was employed to identify and rank features associated with mortality risk. Subsequently, an artificial neural network (ANN) model was developed based on these features to predict mortality risk. The performance of the model was evaluated by receiver operating characteristic (ROC) curve analysis.

Results: The study included 706 patients with 96 features, out of them 26 features were identified as crucial predictors of mortality. The ANN model, utilizing 20 of these features, achieved an area under the ROC curve (AUC) of 98.8 %, effectively stratifying patients by mortality risk.

Conclusion: The developed model offers accurate and precipitous mortality risk predictions for COVID-19 patients, enhancing the responsiveness of healthcare systems to high-risk individuals.

1. Introduction

As of August 2023, approximately 769 million people globally have contracted the coronavirus, with over 6.95 million fatalities [1].

The COVID-19[1] pandemic has significantly strained healthcare systems worldwide, causing unprecedented challenges and resource shortages, including essential medical supplies like ventilators and personal protective equipment (PPE) [2].

The surge in COVID-19 cases has also led to staff shortages, burnout, and mental health issues due to increased workloads [3–5].

When dealing with electronic health records (EHR) data, researchers encounter extensive datasets with numerous features, posing a challenge to incorporate all potential predictors into a model. Traditional statistical methods often fall short in handling high-

* Corresponding author.
E-mail address: sarbakhshp@tbzmed.ac.ir (P. Sarbakhsh).

[1] Coronavirus disease.

https://doi.org/10.1016/j.heliyon.2024.e35561
Received 4 February 2024; Received in revised form 27 July 2024; Accepted 31 July 2024
Available online 5 August 2024

dimensional data because they cannot effectively manage the complexities inherent in datasets with many features [6,7].

Consequently, some features are left out because of insufficient sample sizes during multivariable model building, leading to erroneous fits and misleading conclusions. These limitations highlight the need for alternative methodologies like machine learning algorithms, regularization techniques, and dimensionality reduction methods, which are better equipped to handle complex high-dimensional datasets.

This study employs a range of machine learning techniques, including LASSO, Elastic Net, and Artificial Neural Networks (ANN), for robust analysis and predictive modeling. These methodologies provide a comprehensive toolkit for feature selection, regularization, and non-linear modeling. By leveraging the strengths of LASSO, Elastic Net, and ANN, this study aims to achieve robust predictive performance and enhance interpretability [8,9].

Developing a predictive model for COVID-19 patient mortality can significantly benefit patient outcomes by facilitating effective treatments and helping medical staff prioritize patients efficiently [10].

2. Literature review

Recent studies have employed a range of methodologies, from traditional logistic regression to advanced machine learning techniques, to predict COVID-19 mortality risk effectively. Fei Zhou and et al. [11] identified key risk factors such as older age, high SOFA scores, and elevated d-dimer levels as early indicators of poor prognosis. Xuedi Ma and et al. [12] utilized machine learning models like Random Forest and XGBoost to rank clinical features, highlighting LDH, CRP, and age as critical predictors of severe COVID-19 outcomes. Adam L. Booth and et al. [13] found that CRP, lactic acid, and serum calcium significantly influence mortality prediction models.

Ze Chen and et al. [14] identified eight factors, including oxygen saturation, blood urea nitrogen, respiratory rate, admission date, Age, Procalcitonin, C-reactive protein (CRP), and absolute Neutrophil counts, as significant mortality predictors in COVID-19 patients.

Similarly, Caizheng Yu and et al. [15] demonstrated that factors such as older age, male sex, diabetes history, lymphopenia, and increased procalcitonin levels were associated with higher mortality odds.

Roohallah Alizadehsani and et al. [16] observed significant associations between older age, history of heart disease, and cancer with COVID-19 mortality, along with symptoms like anosmia, dry cough, ageusia, fever, and anorexia.

This study aims to leverage the strengths of the LASSO[2] and Elastic Net to develop an Artificial Neural Network (ANN[3]) model that identifies the most crucial features for predicting mortality in COVID-19 patients at Imam Reza Hospital in Tabriz, Iran. This approach seeks to address the limitations of current methods, which often lack generalizability, explainability, and ease of deployment.

3. Materials and methods

3.1. Study population and data collection

This secondary study is based on data collected by Azari et al. (ethics No. TBZMED.REC.1399.1075) from a retrospective observational study conducted at Imam Reza Hospital, the largest medical education and treatment center in northwest Iran, located in Tabriz, from March 2020 to November 2021. The study included 706 hospitalized COVID-19 patients. Standardized data collection protocols were implemented, and necessary data were gathered using Kardex, clinical and electronic files available in the hospital system. Additional information was obtained from the nurses and doctors treating each patient when their files contained insufficient data. The collected data encompassed demographics, clinical data, experimental data, and outcomes, totaling 96 features. Patients lacking sufficient information for data collection were excluded. Anomalies detected through automated checks were manually reviewed and rectified. The study was approved by the Ethics Committee of Tabriz University of Medical Sciences (ethics No. IR.TBZMED.REC.1402.073).

3.2. Feature selection

In various applications, it is crucial not only to differentiate between categories but also to identify the most relevant predictors for this differentiation. Eliminating unnecessary predictors can lead to more accurate models. LASSO, a regression method using regularization techniques, achieves more precise predictions by penalizing predictor variables, reducing their impact on the output variable [17]. LASSO uses a shrinkage technique to move extreme values of variables towards central values, only retaining features that enhance the fit. This process can set the coefficients of less important variables to near zero, thereby automatically selecting important variables. The LASSO method optimizes the equation:

$$\sum_{i=1}^{n}\left(Y_i-\sum_{j}X_{ij}\beta_j\right)^2+\lambda\sum_{j=1}^{k}|\beta_j|$$

[2] Least absolute shrinkage and selection operator.
[3] Artificial neural network.

where λ is the amount of shrinkage. As λ increases, the shrinkage causes less significant variables to be discarded. LASSO combines shrinkage and predictor selection [17,18].

3.2.1. Elastic Net

Elastic Net combines ridge regression and LASSO, shrinking some small coefficients while setting others to exactly zero.

$$argmin_{\beta}\left(\frac{\| Y - X\beta\|_2^2}{n} + \lambda_2 \| \beta\|^2 + \lambda_1 \| \beta\|_1\right) \text{and } \lambda_1, \lambda_2 \geq 0, \lambda_1 + \lambda_2 = 1$$

This method is robust to extreme correlations among features and can lower mean squared errors more effectively than LASSO or ridge regression when predictor variables are correlated [19,20]. Moreover, the Elastic Net determines affecting features more correctly than LASSO and has a lower false-positive rate than ridge regression [18,21].

3.2.2. Artificial Neural Networks

Artificial neural networks (ANN), ANN models mimic the human brain's learning capacity by connecting predictors to various hidden layer combinations through weighted connections. ANN has been extensively used in medical research to predict outcomes. This study developed a robust predictive model using a one-layer ANN to balance model complexity and interpretability, mitigating overfitting risks [22]. In order to forecast the outcome, ANN links predictors to various hidden layer combinations with the necessary weights as inputs. In summary, ANN predicts the outcome by connecting predictors to many hidden layer combinations through assigning appropriate weights [23]. The primary objective of this study is to develop a robust predictive model for analyzing clinical data of hospitalized COVID-19 patients. The selection of a one-layer ANN model is driven by the need for a balance between model

Table 1
Demographic and clinical characteristics of the patients (N = 706).

Characteristic	Total patients	patient's condition	
		Alive (n = 513)	Dead (n = 193)
Age (≥60)[a]	383 (54.2)	239 (46.6 %)	60: 144 (74.6 %)
Sex (male)[a]	379 (53.7 %)	271 (52.8 %)	108 (56 %)
Smoker[a]	48 (6.8 %)	37 (7.2 %)	11 (5.7 %)
Pulse (bpm)[b]	92.5 ± 16.3, (17–150)	91.24 ± 14.85	96.02 ± 19.37
SBP (mmHg)[b]	123.6 ± 17.9, (70–200)	123.5 ± 16.27	123.86 ± 21.69
DBP (mmHg)[b]	76.5 ± 11.8, (35–150)	76.72 ± 10.28	76.10 ± 15.36
Respiratory rate (bpm)[b]	21.7 ± 7.7, (10–50)	20.74 ± 4.39	24.36 ± 12.87
Body temperature (C°)[b]	36.8 ± 1.5, (35–39)	36.85 ± 0.51	36.67 ± 2.82
Sings & symptoms (yes)[a]			
Shortness of breath	551 (78 %)	403 (78.6 %)	148 (76.6 %)
Cough	429 (60.8 %)	336 (65.6 %)	93 (48.2 %)
Myalgia	295 (41.8 %)	235 (45.8 %)	60 (31.3 %)
Fatigue	201 (28.5 %)	138 (26.9 %)	63 (32.6 %)
Nausea and vomiting	139 (19.7 %)	118 (23 %)	21 (10.9 %)
Anorexia	83 (11.8 %)	65 (12.7 %)	18 (9.3 %)
Headache	62 (8.8 %)	49 (9.6 %)	13 (6.7 %)
Diarrhea	44 (6.2 %)	36 (7 %)	8 (4.1 %)
Decreased consciousness	41 (5.8 %)	14 (2.7 %)	27 (14 %)
Sputum	39 (5.5 %)	34 (6.6 %)	5 (2.6 %)
Sore throat	27 (3.8 %)	20 (3.9 %)	7 (3.6 %)
Stomach cramp	27 (3.8 %)	14 (2.7 %)	4 (2.1 %)
Loss of smell and taste	6 (0.8 %)	5 (1.0 %)	1 (0.5 %)
Skin symptoms	5 (0.7 %)	2 (0.4 %)	3 (1.6 %)
Coexisting conditions (yes)[a]			
Hypertension	113 (32.9 %)	167 (32.6 %)	97 (50.3 %)
Diabetes	173 (24.5 %)	116 (22.6 %)	57 (29.5 %)
Cardiovascular disease	40 (11.7 %)	65 (12.7 %)	50 (25.9 %)
Dyslipidemia	62 (8.8 %)	44 (8.6 %)	18 (9.3 %)
Lung disease	52(7.4)	35 (6.8 %)	17(8.8 %)
kidney failure	31 (4.4 %)	16 (3.1 %)	15 (7.8 %)
Hypothyroidism	29 (4.1 %)	20 (3.9 %)	9 (4.7 %)
Cancer	25 (3.5 %)	13 (2.5 %)	12 (6.2 %)
Pregnancy	14 (2 %)	11 (2.1 %)	3 (1.6 %)
Alzheimer	12 (1.7 %)	7 (1.4 %)	5 (2.6 %)
Immune deficiency	8 (1.1 %)	5 (1.0 %)	3 (1.6 %)
Parkinson	8 (1.1 %)	2 (0.4 %)	6 (3.1 %)
liver failure	7 (1 %)	4 (0.8 %)	3 (1.6 %)

SBP: Systolic blood pressure.
DBP: Diastolic blood pressure.
[a] Number (%).
[b] mean ± SD, (min-max).

complexity and interpretability. Given the nature of clinical data, where overfitting is a common concern due to potentially high dimensionality and noisy features, a simpler architecture helps mitigate these risks, a one-layer ANN can effectively capture the essential patterns without the risk of overfitting.

3.3. Statistical analysis

In this study, we preprocessed the data by removing variables with more than 20 % missing values. For variables with less than 20 % missing data, we applied an imputation method. Given the dataset's diversity, which includes continuous variables (e.g., age, lab results), categorical variables (e.g., gender, comorbidities), and binary variables (e.g., survival status), a flexible and robust imputation method was necessary. We used the "mice" package, which implements multiple imputation by chained equations (MICE), allowing for a comprehensive approach to handling missing data. MICE treats each variable with missing values as a separate imputation model, iterating through these models to generate multiple plausible datasets. For binary variables, the MICE package used logistic regression, modeling the probability of the binary outcome and imputing accordingly. Before imputation, we converted categorical variables into numeric codes or factors as required. We employed multiple imputations by chained equations using the "mice" package [24].

The Elastic Net algorithm was utilized to pinpoint predictors of COVID-19 patient mortality. All relevant features were input into the Elastic Net model for variable selection, performed using the "glmnet" package in R [25]. We applied a 10-fold cross-validation test with the "lambda.1se" criterion to find the optimal λ value, which signified the most regularized model with an error within one standard error of the minimum. Features with nonzero coefficients were retained for model building. To avoid overfitting, feature selection and model training were initially conducted with the training set and later validated with the testing set. The data was divided into training and testing sets in a 7:3 ratio. We employed the random oversampling technique to balance the classes in the training dataset [26].

To verify the robustness of the selected features, we constructed an artificial neural network (ANN) model. The fitted ANN model was a three-layers standard feed-forward neural network: an input layer, a hidden layer, and an output layer. The hidden layer's activation function was logistic. The relative importance of individual predictors was assessed by analyzing the model weights. The "nnet" package in R was used to create the ANN model [27].

The performance of the ANN model was evaluated through ROC curve analysis and calculation of the AUC value in the testing sets. The ROC curve visually represents the model's sensitivity (true positive rate) and specificity (true negative rate) across various thresholds. The shape and AUC of the ROC curve indicate the reliability of the model's predictions, with a smoother, more concave curve and a higher AUC suggesting a more reliable model with consistent performance across different thresholds.

4. Results

The demographic and clinical characteristics of the patients in this study are summarized in Table 1. Continuous features were

Table 2
Laboratory characteristics of the patients at baseline.

Characteristic	Total patients	patient's condition	
		Alive (n = 513)	Dead (n = 193)
PCR[a] (positive)	508 (93.9 %)	375 (73.1 %)	133 (68.9 %)
WBC count[b]	9.1 ± 8.8	8.1 ± 7.1	11.6 ± 11.8
Lymphocyte count[b]	1.6 ± 6.1	1.43 ± 5.0	1.9 ± 8.2
Lymphocyte count (<0.8)[a]	236 (33.4 %)	153 (29.8 %)	83 (43 %)
Haemoglobin[b]	13.4 ± 6	13.5 ± 6.8	13.2 ± 3.2
CRP[b]	70.4 ± 60	64.4 ± 59.3	86.4 ± 70.7
CRP (>100)[a]	152 (21.5 %)	98 (19.1 %)	54 (28 %)
Platelet count[b]	211.9 ± 89.5	210.7 ± 85.3	214.7 ± 99.9
ALT[b]	51.2 ± 121.9	48.4 ± 65.1	44.2 ± 55.3
AST[b]	58 ± 140.7	49.8 ± 55.1	62.1 ± 51.0
Creatinine[b]	1.3 ± 1.11	1.3 ± 1.4	1.8 ± 1.7
Urea[b]	51.8 ± 38.6	43.2 ± 28.4	74 ± 50.9
LDH[b]	807.7 ± 570.5	720.4 ± 318.21	971.2 ± 574.5
LDH (>245)[a]	619 (87.7 %)	439 (85.6 %)	180 (93.3 %)
PT[b]	13.8 ± 3.8	13.7 ± 4.1	14.0 ± 2.7
INR[b]	1.1 ± 0.53	1.1 ± 0.5	1.2 ± 0.5
Na[b]	138 ± 4.2	137.7 ± 3.6	138.8 ± 5.3
K[b]	4.2 ± 0.6	4.2 ± 0.5	4.3 ± 0.6

PT: Prothrombin time.
INR: International normalized time.
ICU: Intensive Care Unit.
NIV: noninvasive ventilation.
ARDS: Acute respiratory distress syndrome.
[a] Number (%).
[b] mean ± SD.

presented as mean ± standard deviation (SD), while categorical variables are shown as numbers (percentage). The study included 706 COVID-19 patients, of whom 379 (53.7 %) were males and 383 (54.2 %) were aged 60 and older. Hypertension and diabetic were present in 113(32.9 %) and 173(24.5 %) patients, respectively (Table 1). The highest percentages in signs and symptoms were shortness of breath (78 %) and cough (60.8 %). Also, the highest percentages in the condition of patients were hypertension (32.9 %) and diabetes (24.5 %). The laboratory characteristics, clinical outcomes of patients, and the information of drugs received by hospitalized patients are presented in Table 2, Tables 3 and 4. There were 508 individuals with positive PCR (93.9 %), 193 (27.3 %) died in the hospital, and 251 (35.6 %) were admitted to the intensive care unit.

Using the Elastic Net with the "lambda.1se" criterion, 70 of 96 features were excluded (Fig. 1), leaving the following prognostic features (ELASTIC NET = 26) with nonzero coefficients. Almost all 26 features (except for age) were about clinical data of patients in the hospital including 23 categorical features and 3 continuous ones. After analyzed by Spearman's rank correlation and Pearson correlation and validated by clinical assessment, 20 features were eventually selected for modelling. Of these, 7 features showed a positive relationship with mortality (Age≥60, bromhexine utilization, lymphocyte count ≥0.8, hydrocortisone utilization, CRP, INR and urea plasma level), while 13 features were negatively correlated with mortality (ARDS, acute cardiac injury, lower level of SPO2 in the hospital admission, admission in ICU, receiving invasive and noninvasive ventilation, septic shock, heart failure, acute kidney injury, sepsis, decreased consciousness, HTN, kidney failure, respiratory failure).

Fig. 3 represents box and jitter plots indicating the distribution of continuous features included in the ANN model between survived patients (n = 513) and dead patients (n = 193). As shown in Fig. 3, INR, CRP, and urea in the deceased group were higher than in the living group, whereas SPO2 on the first day was lower in the dead group than in the living group.

4.1. Model performance

We utilized the ANN model to build a predictive model with features obtained from Elastic Net. In the optimized model obtained, ARDS and urea had high importance in the mortality rate (Table 5). The ROC curve for the artificial neural network based on selected features showed a high AUC (Fig. 2). The final values used for the model were size = 5 (number of hidden layers) and decay = 0.1, so the ANN model achieved an AUC of 98.8 % (95 % CI: 97.8–99.8) and an accuracy of 97.14 % (95 % CI: 93.89–98.94), sensitivity of 98.04 %, specificity of 94.74 %, positive predictive value (PPV) of 98.04 %, and negative predictive value (NPV) of 94.74 %.

5. Discussion

In this retrospective study, we developed an Artificial Neural Network (ANN) model, an ensemble approach combining the Elastic Net machine learning algorithm, which accurately predicted various clinical characteristics, outcomes, symptoms, coexisting conditions, laboratory results, and medication reports for 706 COVID-19 patients. The ANN model, utilizing 20 predictors, achieved a high classification rate of AUC = 0.98 and demonstrated exceptional discriminatory power. ARDS, acute cardiac injury, lower level of SPO2

Table 3
Drug information received by hospitalized patient. (N = 706).

Characteristic	Total patients	patient's condition	
		Alive (n = 513)	Dead (n = 193)
Remdesivir[a]	343(48.6 %)	240 (46.8 %)	103 (53.4 %)
Corticosteroids[a]	554 (78.5 %)	377 (73.5 %)	177 (91.7 %)
Dexamethasone[a]	433(61.3 %)	310 (60.4 %)	123 (63.7 %)
Methylprednisolone[a]	232 (32.9 %)	141 (27.5 %)	91 (47.2 %)
Hydrocortisone[a]	160 (22.7 %)	64 (12.5 %)	96 (49.7 %)
Prednisolone[a]	20 (2.8 %)	17 (3.3 %)	3 (1.6 %)
Bromhexine[a]	407 (57.6 %)	317 (61.8 %)	87 (45.1 %)
Interferon Beta(1 b,1a)[a]	324 (45.9 %)	168 (32.7 %)	70 (36.3 %)
Famotidine[a]	275 (39 %)	209 (40.7 %)	66 (34.2 %)
Vitamin C[a]	269 (38.1 %)	194 (37.8 %)	75 (38.9 %)
Vitamin D_3[a]	217 (30.7 %)	154 (30.0 %)	63 (32.6 %)
Zinc[a]	185 (26.2 %)	138 (26.9 %)	47 (24.4 %)
N-acetylcysteine[a]	176 (24.9 %)	125 (24.4 %)	51 (26.4 %)
Hydroxychloroquine[a]	163 (23.1 %)	126 (24.6 %)	37 (19.2 %)
Azithromycin[a]	125 (17.7 %)	88 (17.2 %)	37 (19.2 %)
Atorvastatin[a]	100 (14.2 %)	65 (12.7 %)	35 (18.1 %)
Favipiravir[a]	72 (10.2 %)	58 (11.3 %)	14 (7.3 %)
lopinavir-ritonavir (Kaletra)[a]	80 (11.3 %)	56 (10.9 %)	24 (12.4 %)
Melatonin[a]	43 (6.1 %)	24 (4.7 %)	19 (9.8 %)
Hemoperfusion[a]	26 (3.7 %)	11 (2.1 %)	15 (7.8 %)
Colchicine[a]	25 (3.5 %)	15 (2.9 %)	10 (5.2 %)
Tocilizumab[a]	23 (3.3 %)	9 (1.8 %)	14 (7.3 %)
Oseltamivir[a]	11 (1.6 %)	9 (1.8 %)	2 (1.0 %)
IVIG[a]	3 (0.4 %)	1 (0.2 %)	2 (1.0 %)

[a] Number (%).

Table 4
Clinical outcomes of patients in the hospital Treatments for COVID19 trial.

Characteristic	Total patients	patient's condition	
		Alive (n = 513)	Dead (n = 193)
In-hospital mortality[a]	193 (27.3 %)	513 (72.7 %)	193 (27.3 %)
Respiratory failure[a]	48 (6.8 %)	3 (0.6 %)	45 (23.3 %)
Admission to ICU[a]	251 (35.6 %)	106 (20.7 %)	145 (75.1 %)
Duration of ICU stay (day)[b]	3.7 ± 7.3 (1–53)	7.75 ± 5.7	11.6 ± 10.5
Receiving invasive or noninvasive ventilation[a]	234 (33.1 %)	141 (27.5 %)	184 (95.3 %)
Sepsis[a]	142 (20.1 %)	39 (7.6 %)	103 (53.4 %)
ARDS[a]	150 (21.2 %)	3 (0.6 %)	147 (76.2 %)
Heart failure[a]	46 (6.5 %)	7 (1.4 %)	39 (20.2 %)
Septic shock[a]	70 (9.9 %)	7 (1.4 %)	63 (32.6 %)
Coagulopathy[a]	112 (15.9 %)	59 (11.5 %)	53 (27.5 %)
Acute cardiac injury[a]	64 (9.1 %)	12 (8.0 %)	52 (26.9 %)
Acute kidney injury[a]	88 (12.5 %)	41 (8.0 %)	47 (24.4 %)
Secondary infection[a]	251 (35.6 %)	129 (25.1 %)	122 (63.2 %)
Hypo proteinemia[a]	39 (5.5 %)	16 (3.1 %)	23 (11.9 %)

[a] Number (%).
[b] mean ± SD, (min-max).

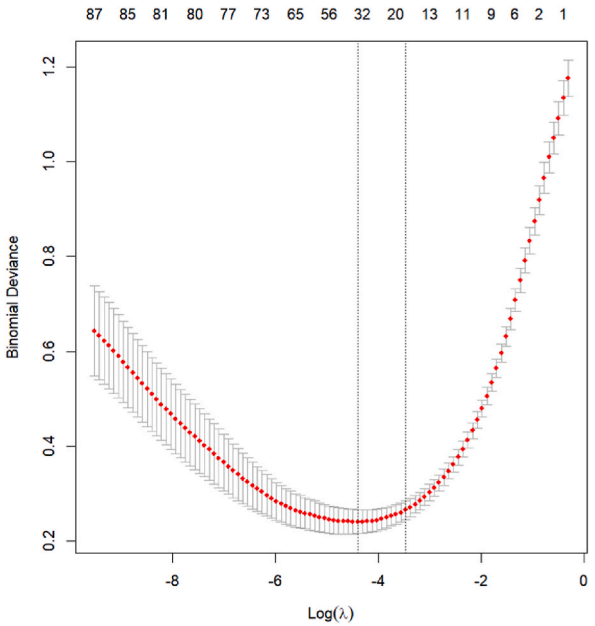

Fig. 1. Tuning parameter (lambda) selection in the Elastic Net model used 10-fold cross-validation based on "lambda.1se" criteria for COVID-19 prognosis.

in the hospital admission, admission to the ICU, receiving invasive and noninvasive ventilation, septic shock, heart failure, acute kidney injury, sepsis, decreased consciousness, HTN, kidney failure, respiratory failure, age ≥60, bromhexine utilization, lymphocyte count≥ 0.8, hydrocortisone utilization, CRP, INR, and urea plasma level were found to be essential factors for mortality in our study. All factors identified as predictor factors for mortality are expected, predictable, and reasonable, except bromhexine and hydrocortisone utilization, which need more discussion.

Coronaviruses enter host cells through interactions between the viral spike protein and host cell receptors, such as transmembrane serine protease 2 (TMPRSS2) or angiotensin-converting enzyme 2. Barzegar et al. [28] proposed that bromhexine can prevent viral entry into host cells by inhibiting TMPRSS2. Only a few numbers of studies with small sample sizes evaluated the effects of bromhexine on COVID-19 patients. Li et al. [29] reported that bromhexine improved chest CT findings, reduced the need for oxygen therapy, and increased discharge rates within 20 days. Ansarin et al. [30] found that early bromhexine use lowered ICU admission rates, intubation rates, and mortality. In contrast, the study by Tolouian et al. [31] indicates that bromhexine has not been shown to decrease rates of ICU admission, the need for oxygen therapy, or mortality. Based on the previous studies, the results regarding the use of bromhexine in COVID-19 patients are conflicting and large-scale studies are still warranted. Bromhexine has mild and self-limited adverse effects, including gastrointestinal side effects, transient elevation of serum aminotransferase, headache, vertigo, sweating, and rare allergic

Table 5
The rank of importance of features in the ANN model.

Predictors	Importance (%)
ARDS	100
Receiving invasive or noninvasive ventilation	96.2
Septic shock	76.7
Sepsis	58.6
Urea	48.2
SPO2 on first day	46.7
Respiratory failure	42.2
Acute cardiac injury	40.1
Acute kidney injury	36.3
Hydrocortisone	28.6
Age≥60	28.1
Decrease of consciousness	26.8
kidney failure	21.8
Bromhexine	21.8
Lymphocyte count ≥0.8	14.7
HTN	10.4
CRP	8.3
Heart failure	7.3
INR	5.4
Admission in ICU	0.9

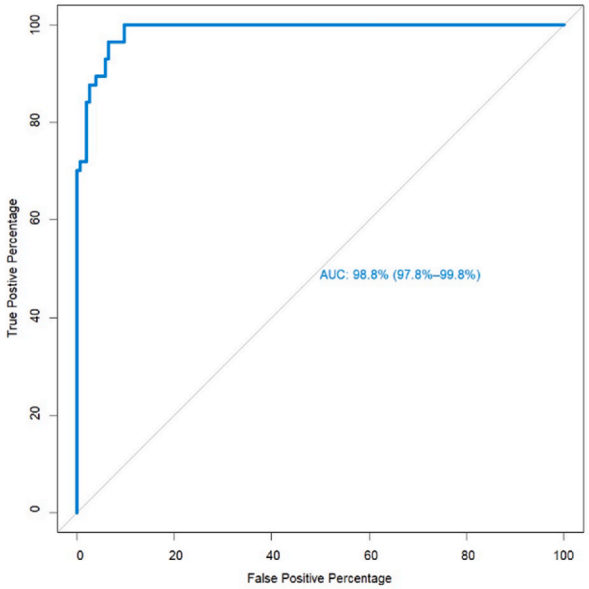

Fig. 2. ROC curve for the artificial neural network base on selected features.

reactions [32]. This safe over-the-counter medication is used in many various conditions without any expected major adverse event. The findings of our current retrospective observational study about Bromhexine do not align with previous clinical research. It may be due to the confounding effect of infection severity, which requires assessment in future clinical trials.

A network meta-analysis of 10544 patients from 19 trials revealed that the effects of glucocorticoids on COVID-19 patients depend on dosage, regimen, and type of glucocorticoids. Among different dosages and regimens, only pulse therapy with methylprednisolone is associated with a lower range of mortality, and evidence regarding the use of hydrocortisone and dexamethasone is still lacking [33]. Larger RCTs are highly recommended to evaluate the net effects of hydrocortisone in COVID-19 patients.

Sepsis guidelines advise the use of hydrocortisone to treat septic shock because it has a mineralocorticoid effect [34]. Therefore, hydrocortisone is used in critically ill patients with a higher probability of mortality. Then it seems that hydrocortisone may not be an independent risk factor for mortality, and it may not increase patients' mortality, possibly due to the severity of the patient's condition. Additionally, the attending physician prescribed several potentially effective drugs, so it does not appear that these drugs play a role in the patients' mortality.

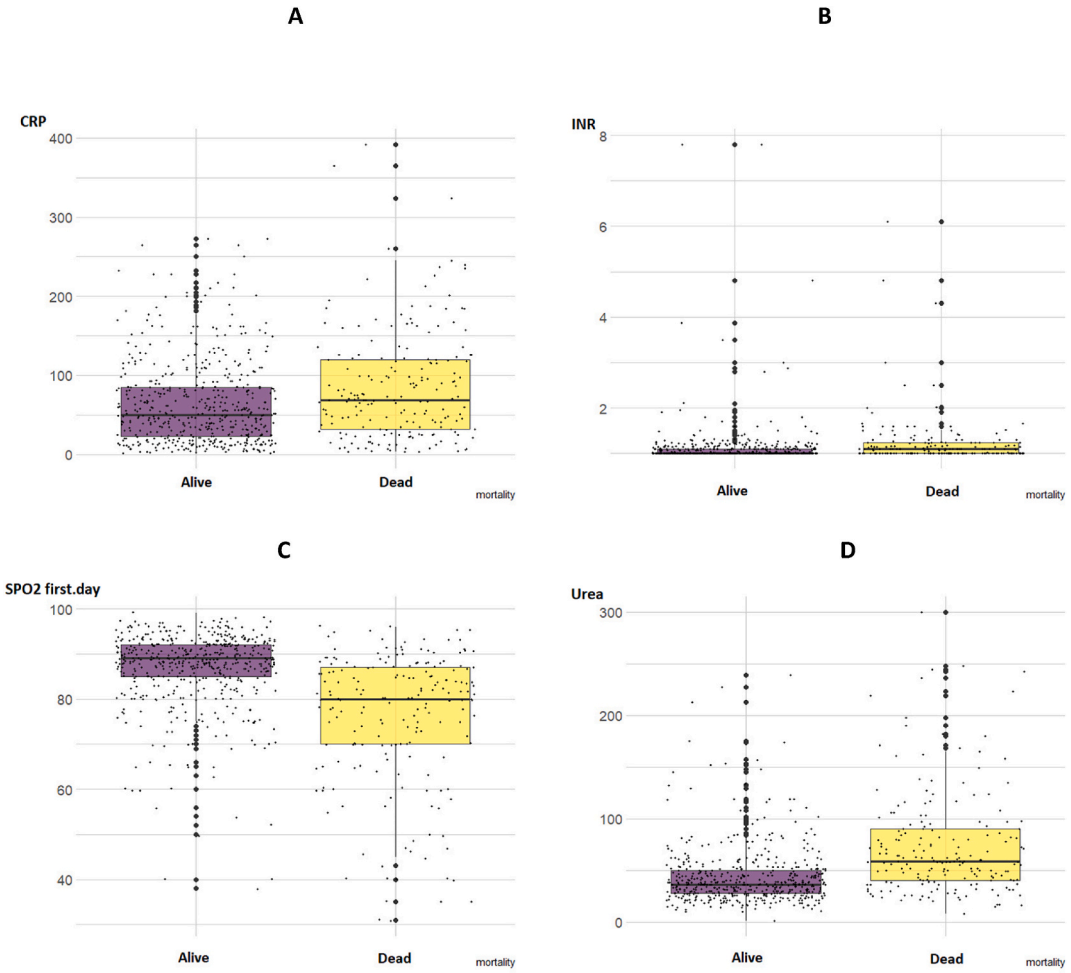

Fig. 3. Box and jitter plots of continues features (A: CRP; B: INR; C: SPO2 first day; D: Urea) according to the mortality outcome.

6. Conclusion

This study highlights the association of some variables with COVID-19 mortality, and the need for further research into the effects of bromhexine and hydrocortisone, given their unexpected association. The methodologies employed here are applicable for exploring additional predictors and refining predictive models. Insights from this study could enhance public health strategies, including resource allocation and management guidelines for hospitalized COVID-19 patients. Identifying high-risk patients early can improve hospital and ICU management. Future research should focus on demographic variability, new predictors, and integration into clinical workflows.

Strengths and limitations

A key strength of our study is the application of precise methods in developing and validating our classification and prediction model. We split the data into training and testing sets to avoid overfitting during model development. We utilized the Elastic Net algorithm and nonlinear ANN model to assess the selected features' performance in predicting patient mortality. However, our study had several limitations. Firstly, existing ML-based studies on prognosis prediction for COVID-19 patients face challenges such as limited sample sizes, restricted feature categories for prediction, short-term follow-up periods, and a lack of independent external validation. These factors could affect the generalizability of our findings.

Funding

This project was approved with the ethics code IR.TBZMED.REC.1402.073 and received financial support from the Vice-Chancellor for Research and Technology of Tabriz University of Medical Sciences under grant number 70419.

Ethics approval and consent to participate

This study was approved by the Ethics Committee of Tabriz University of Medical Sciences (ethics No. IR.TBZMED.REC.1402.073). Also, the primary study on which we carried out the current secondary analysis has ethics approval by the Ethics Committee of Tabriz University of Medical Sciences (ethics No. TBZMED.REC.1399.1075). All participants or their relatives had signed informed consent.

Consent for publication

Not applicable.

Data availability statement

The data supporting the findings of this study are available from the corresponding author upon reasonable request.

CRediT authorship contribution statement

Farzaneh Hamidi: Writing – original draft, Software, Methodology, Investigation, Formal analysis. **Hadi Hamishehkar:** Writing – original draft, Supervision, Project administration, Investigation, Data curation, Conceptualization. **Pedram Pirmad Azari Markid:** Writing – original draft, Investigation, Data curation. **Parvin Sarbakhsh:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Methodology, Investigation, Conceptualization.

Declaration of competing interest

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

Acknowledgments

We would like to thank the research consultation of “Clinical Research Development Unit of Imam Reza Hospital”, Tabriz University of Medical Sciences for kind support. We also extend our gratitude to everyone who participated in the primary research project and contributed to data collection. We are also very thankful to the “anonymous” reviewers.

References

[1] WHO Coronavirus Disease (COVID-19) Dashboard DlupC, https://covid19.who.int/. Available from: https://covid19.who.int/.
[2] Y. Shang, et al., Scoring systems for predicting mortality for severe patients with COVID-19, EClinicalMedicine 24 (2020).
[3] S. Amanullah, R. Ramesh Shankar, The impact of COVID-19 on physician burnout globally: a review, in: Healthcare, MDPI, 2020.
[4] C.J. Worby, H.-H. Chang, Face mask use in the general population and optimal resource allocation during the COVID-19 pandemic, Nat. Commun. 11 (1) (2020) 4049.
[5] A.T. Janke, et al., Analysis of hospital resource availability and COVID-19 mortality across the United States, J. Hosp. Med. 16 (4) (2021) 211–214.
[6] M. Lagadinou, et al., Prognosis of COVID-19: changes in laboratory parameters, Age (yrs) 62 (13.4) (2020) 47, 16.42.
[7] F. Zhou, et al., Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study, The lancet 395 (10229) (2020) 1054–1062.
[8] S. Ghosh, M. Alamgir Islam, Performance evaluation and comparison of heart disease prediction using machine learning methods with elastic net feature selection, Am. J. Appl. Math. Stat. 11 (2) (2023) 35–49.
[9] A.R. Patil, S. Kim, Combination of ensembles of regularized regression models with resampling-based lasso feature selection in high dimensional data, Mathematics 8 (1) (2020) 110.
[10] Z. Yang, et al., Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions, J. Thorac. Dis. 12 (3) (2020) 165.
[11] F. Zhou, et al., Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study, Lancet 395 (10229) (2020) 1054–1062.
[12] X. Ma, et al., Development and validation of prognosis model of mortality risk in patients with COVID-19, Epidemiol. Infect. 148 (2020) e168.
[13] A.L. Booth, E. Abels, P. McCaffrey, Development of a prognostic model for mortality in COVID-19 infection using machine learning, Mod. Pathol. 34 (3) (2021) 522–531.
[14] Z. Chen, et al., A risk score based on baseline risk factors for predicting mortality in COVID-19 patients, Curr. Med. Res. Opin. 37 (6) (2021) 917–927.
[15] C. Yu, et al., Clinical characteristics, associated factors, and predicting COVID-19 mortality risk: a retrospective study in Wuhan, China, Am. J. Prev. Med. 59 (2) (2020) 168–175.
[16] R. Alizadehsani, et al., Risk factors prediction, clinical outcomes, and mortality in COVID-19 patients, J. Med. Virol. 93 (4) (2021) 2307–2320.
[17] R. Tibshirani, Regression shrinkage and selection via the lasso, J. Roy. Stat. Soc. B 58 (1) (1996) 267–288.
[18] H. Zou, T. Hastie, Regularization and variable selection via the elastic net, J. Roy. Stat. Soc. B 67 (2) (2005) 301–320.
[19] L. Bornn, R. Gottardo, A. Doucet, Grouping Priors and the Bayesian Elastic Net, 2010 arXiv preprint arXiv:1001.4083.
[20] F. Hamidi, et al., Exploration of potential miRNA biomarkers and prediction for ovarian cancer using artificial intelligence, Front. Genet. 12 (2021).
[21] G. Tutz, J. Ulbricht, Penalized regression with correlation-based penalty, Stat. Comput. 19 (2009) 239–253.
[22] K. DeGregory, et al., A review of machine learning in obesity, Obes. Rev. 19 (5) (2018) 668–685.
[23] P.J. Lisboa, A.F. Taktak, The use of artificial neural networks in decision support in cancer: a systematic review, Neural Network. 19 (4) (2006) 408–415.
[24] S. van Buuren, K. Groothuis-Oudshoorn, mice: Multivariate Imputation by Chained Equations in R, J. Stat. Softw. 45 (3) (2011) 1–67.
[25] J. Friedman, et al., Package ‘glmnet’, J. Stat. Software 33 (1) (2010), 2022.
[26] N. Lunardon, G. Menardi, N. Torelli, ROSE: a package for binary imbalanced learning, R journal 6 (1) (2014).
[27] B. Ripley, W. Venables, M.B. Ripley, Package ‘nnet’, R package version 7 (3–12) (2016) 700.

[28] A. Barzegar, et al., New hope for treatment of respiratory involvement following COVID-19 by bromhexine, Pharmacol. Ther. 28 (2) (2014) 98–108.
[29] T. Li, et al., Bromhexine hydrochloride tablets for the treatment of moderate COVID-19: an open-label randomized controlled pilot study, Clinical and translational science 13 (6) (2020) 1096–1102.
[30] K. Ansarin, et al., Effect of bromhexine on clinical outcomes and mortality in COVID-19 patients: a randomized clinical trial, Bioimpacts: BI 10 (4) (2020) 209.
[31] R. Tolouian, et al., Effect of bromhexine in hospitalized patients with COVID-19, J. Investig. Med. 71 (7) (2023) 691–699.
[32] M.F. Law, R. Ho, K.W.T. Law, C.K.M. Cheung, Gastrointestinal and hepatic side effects of potential treatment for COVID-19 and vaccination in patients with chronic liver diseases, World J. Hepatol. 13 (12) (2021) 1850–1874.
[33] Q. He, et al., Efficacy and safety of glucocorticoids use in patients with COVID-19: a systematic review and network meta-analysis, BMC Infect. Dis. 23 (1) (2023) 896.
[34] L. Evans, et al., Surviving sepsis campaign: international guidelines for management of sepsis and septic shock 2021, Intensive Care Med. 47 (11) (2021) 1181–1247.