

Research Paper

Predicting COVID-19 Mortality and Identifying Clinical Symptom Patterns in Hospitalized Patients: A Machine-learning Study



Nasrin Talkhi^{1,2}, Nooshin Akbari Sharak¹, Razieh Yousefi¹, Maryam Salari³, Seyed Masoud Sadati⁴, Mohammad Taghi Shakeri^{5*}

1. Department of Biostatistics, School of Health, Mashhad University of Medical Sciences, Mashhad, Iran.

2. International UNESCO Center for Health-Related Basic Sciences and Human Nutrition, Mashhad University of Medical Sciences, Mashhad, Iran.

3. Department of Biostatistics, Expert Management and Information Technology, School of Health, Mashhad University of Medical Sciences, Mashhad, Iran.

4. Center of Statistics and Information Technology Management, Imam Reza Hospital, Mashhad University of Medical Sciences, Mashhad, Iran.

5. Social Determinants of Health Research Center, Mashhad University of Medical Sciences, Mashhad, Iran.



Citation Talkhi N, Akbari Sharak N, Yousefi R, Salari M, Sadati SM, Shakeri MT. Predicting COVID-19 Mortality and Identifying Clinical Symptom Patterns in Hospitalized Patients: A Machine-learning Study. *Iranian Journal of Health Sciences*. 2024; 12(1):39-48. <http://dx.doi.org/10.32598/ijhs.12.1.952.1>

doi <http://dx.doi.org/10.32598/ijhs.12.1.952.1>



ABSTRACT

Background and Purpose: Identifying effective symptoms, demographic information, and underlying diseases to predict COVID-19 mortality is essential. We aimed to study the effective clinical and symptomatic characteristics of COVID-19 mortality in hospitalized patients with positive polymerase chain reaction (PCR) test results.

Materials and Methods: For this study, we prospectively collected complete data on 26867 hospitalized individuals who tested PCR positive for COVID-19 from February 20, 2020, to September 12, 2021, in the Khorasan Razavi Province, Iran. We analyzed the data using artificial neural networks (ANN) and logistic regression (LR) models.

Results: The accuracy of the ANN model was higher than the LR (90.27% versus 90.15%). The ten most important predictors that contributed to the prediction of death were decreasing consciousness level, cough, PO₂ level, age, chronic kidney disease, fever, headache, smoking status, chronic blood diseases, and diarrhea using the ANN model.

Conclusion: In conclusion, individuals suffering from chronic diseases such as cancer, kidney and blood diseases, as well as immunodeficiency are at a higher risk of mortality. This important finding can help decision-makers and medical professionals in their efforts to consider these conditions and provide effective preventative measures to reduce the risk of death.

Keywords: Machine learning, SARS-CoV-2, COVID-19 diagnostic testing, Logistic regression, Neural network

Article info:

Received: 29 Jul 2023

Accepted: 25 Aug 2023

Available Online: 01 Jan 2024

* Corresponding Author:

Mohammad Taghi Shakeri, Professor.

Address: Social Determinants of Health Research Center, Mashhad University of Medical Sciences, Mashhad, Iran.

Tel: +98 (915) 5151799

E-mail: ShakeriMT@mums.ac.ir

Introduction

In December 2019, a novel coronavirus was identified in Wuhan, Hubei Province, China [1, 2]. Due to the virus's high transmissibility, it quickly spread globally, resulting in a pandemic that significantly impacted daily life. The World Health Organization (WHO) has reported that the virus has now affected all countries, disrupting personal life, economy, and industry [3]. The first case of COVID-19 in Iran was confirmed to be PCR-positive on February 19, 2020, following the pandemic's spread to various regions worldwide [4].

According to research, the virus can remain viable on surfaces from several hours to a few days and spread rapidly from person to person [5]. Common symptoms of the infection include fatigue, weakness, chest pain, difficulty breathing, sore throat, fever, muscular pain, and acute respiratory distress, although some individuals may be asymptomatic [5, 6]. The mortality rate for high-risk groups, such as older people and those with preexisting conditions such as cardiovascular disease, diabetes, chronic respiratory disease, and hypertension, is notably higher than in healthy individuals [7]. There are four classifications of COVID-19: mild, moderate, severe, and critical. However, most infected individuals present with mild symptoms [8]. Early virus detection is crucial in ceasing its transmission and preventing complications and spread. Thus, timely diagnosis and treatment are crucial to controlling the disease [8].

Governments have implemented various interventions and strategies, such as maintaining social distancing, wearing masks, staying at home, and avoiding large gatherings to limit the spread of the pandemic [9]. Despite these efforts and several years of dealing with the disease, it continues to cause fatalities in many individuals [10]. Moreover, due to the virus's wide distribution, genetic recombination is possible, leading to the emergence of new coronavirus strains. Therefore, it is likely that this infectious disease will periodically resurface in humans [1].

In the current scenario, data analysts should use statistical, mathematical, and machine learning (ML) models to predict and diagnose fatalities. Providing earlier warnings to governments and policymakers about potential deaths can aid in making informed decisions. Statistical techniques and ML algorithms have widely been used in the medical field with successful outcomes [11, 12]. Studies have shown that methods such as artificial neural networks (ANN) and logistic regression (LR) can assist in disease diagnosis [13-16].

In this study, our objective was to examine the behavioral patterns of SARS-CoV-2 through LR and ANN. By analyzing these models, we could identify factors closely linked to the death status of individuals affected by the virus. We also assessed the significance of certain characteristics on the likelihood of dying from the virus. Based on these factors, we used this data to model and predict the disease mortality rate.

Materials and Methods

Study design and data collection

This retrospective study focused on hospitalized COVID-19 patients between February 20, 2021, and September 12, 2021, in Khorasan Razavi Province, northeast Iran (Figure 1). The study investigated various aspects of this patient population and employed a retrospective design. The study population comprised individuals admitted to hospitals within Khorasan Razavi Province due to COVID-19 symptoms. Patients were included based on their doctor's diagnosis and the necessity for hospitalization. The study focused on patients with positive polymerase chain reaction (PCR) test for COVID-19.

Sampling was conducted from the comprehensive records available in the Medical Care Monitoring Center (MCMC) database. The database contained data extracted from questionnaires filled out by nurses during patient care. The inclusion criteria consisted of individuals with confirmed COVID-19 diagnosis through PCR testing. It is important to note that Khorasan Razavi Province is a religious center that welcomes pilgrims from various parts of Iran and neighboring Muslim countries. Before analysis, a meticulous data-cleaning process was undertaken. The data were thoroughly reviewed to identify and remove inaccuracies, irrelevant entries, missing information, and incomplete records. As a result, a dataset comprising 23 variables and 26867 instances was compiled for subsequent analyses.

The independent variables in this study were demographic information, epidemiological data, and clinical symptoms. These variables include age, sex, fever, cough, muscular pain, respiratory distress, decreased consciousness, decreased sense of smell, reduced sense of taste, convulsions, headache, confusion, chest pain, skin inflammation, stomachache, nausea, vomiting, diarrhea, anorexia, smoking status, drug use, PO₂, cancer, chronic liver diseases, diabetes, chronic blood diseases, immunodeficiency, heart diseases, chronic kidney diseases, asthma, chronic neurological disorders, hypertension. The dependent variable, death status, was dichotomous with survivor or non-survivor status.

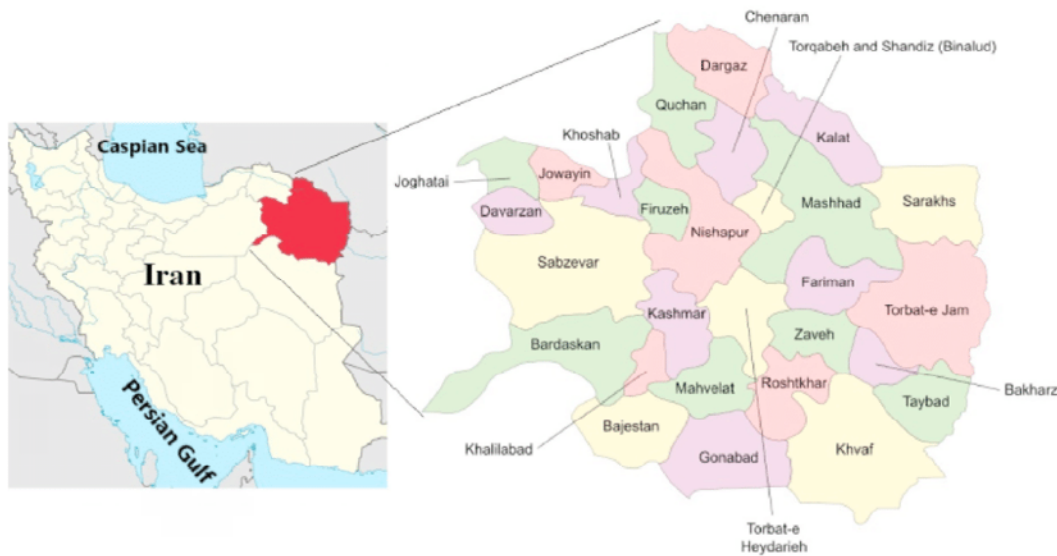


Figure 1. Map of Khorasan Razavi Province and its cities in Iran

Statistical analysis

To perform statistical and primary analysis in this study, the R software version 4.1.1 was used. All the analyses were performed in the PCR+ individuals. The descriptive statistics for age were reported as Mean±SD. The Kolmogorov-Smirnov test was used to check the normality of the age variable. The normality test showed that the age was distributed as normal. The Mean±SD death age in PCR+ individuals was 68.66±16.07 years. The age variable was categorized based on available information about age categories according to the WHO [17]. The Mann-Whitney U-test was applied to compare the mean age between two independent groups (survivor/non-survivor).

A body temperature higher than 37°C was considered to be a fever. The chi-square test was used to investigate the association between death status and qualitative variables. A P<0.05 was regarded as statistically significant.

Mortality modeling

LR is an appropriate statistical regression [18-20]. Recently, it has been commonly used in the medical and biological sciences [21]. The LR model is more popular than other methods among medical researchers because the exponentiated slope coefficient in the LR model can be interpreted as an odds ratio (OR) [22, 23]. The Hosmer-Lemeshow goodness of fit (GOF) test, Cox & Snell R², and Nagelkerke R² for the LR model were calculated. P>0.05 in Hosmer-Lemeshow GOF indicated a good fit for the model.

ANN has been one of the branches of artificial intelligence since the 1990s, with numerous applications in medical diagnosis. ANN models try to find a connection between the input and output data. In other words, they attempt to act similarly to the human brain. Also, ANN is an advanced intelligent tool for researchers and medical diagnosis fields such as cancer, heart diseases, hepatitis, and drug development [24]. To train and optimize an ANN model, the 10-fold cross-validation (CV) method was used in which 90% of records were randomly assigned to training data for constructing the models and the remaining 10% to testing data to evaluate the model's performance. The performance of the ANN and LR classifiers was assessed by the confusion matrix and the related performance evaluation criteria, such as sensitivity, specificity, accuracy, precision, and recall, as follows (Equations 1 and 2):

$$1. \text{ Sensitivity} = \frac{TP}{TP+FN} \quad \text{Specificity} = \frac{TN}{TN+FP}$$

$$2. \text{ Precision} = \frac{TP}{TP+FP} \quad \text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

, Where TP, FN, FP, and TN are true positive, false negative, false positive, and true negative, respectively [25-27]. In the end, a lollipop plot was used to depict the importance of predictor variables in the LR and ANN models.

Results

Descriptive statistics for the observed symptoms, underlying diseases, and other characteristics of the studied COVID-19 patients are reported in Table 1.

Table 1. Characteristics and symptoms of the study population

Variable [Reference]	Level	Mean±SD/No. (%)		P
		Survivor	Non-survivors	
Age (y)		52.67±19.09	68.66±16.07	<0.001
Age group (y) [<18]	Age group 2 [18-45]	7070(32.4)	393(7.8)	<0.001
	Age group 3 [45-65]	7901(36.2)	1325(26.4)	
	Age group 4 [65-75]	2980(13.6)	1177(23.4)	
	Age group 5 [≥75]	3213(14.7)	2089(41.6)	

Variables (Ref.)	No. (%)		P	Variables [Reference]	No. (%)		P
	Survivor	Non-survivors			Survivor	Non-survivors	
Fever*	8257(37.8)	1684(33.5)	<0.001	Anorexia*	1857(8.5)	424(8.4)	0.891
Cough*	10033(45.9)	1641(32.7)	<0.001	Smoking status*	386(1.8)	137(2.7)	<0.001
Muscular pain*	5368(24.6)	667(13.3)	<0.001	Drug use*	471(2.2)	212(4.2)	<0.001
Respiratory distress*	11942(54.7)	3962(78.9)	<0.001	Intubation*	509(2.3)	2247(44.7)	<0.001
Conscious*	590(2.7)	923(18.4)	<0.001	PO ₂ (>93 %)	10683(48.9)	4359(86.8)	<0.001
Decreased sense of smell*	358(1.6)	18(0.4)	<0.001	Cancer*	148(0.7)	162(3.2)	<0.001
Decreased sense of taste*	136(0.6)	15(0.3)	0.006	Liver disease*	88(0.4)	44(0.9)	<0.001
Convulsions*	79(0.4)	20(0.4)	0.700	Diabetes*	2357(10.8)	1069(21.3)	<0.001
Headache*	2333(10.7)	130(2.6)	<0.001	Blood diseases*	61(0.3)	42(0.8)	<0.001
Confusion*	578(2.6)	73(1.5)	<0.001	Immunodeficiency*	20(0.1)	13(0.3)	0.002
Chest pain*	693(3.2)	109(2.2)	<0.001	Heart disease*	1787(8.2)	969(19.3)	<0.001
Skin inflammation*	19(0.1)	4(0.1)	0.872	Kidney disease*	241(1.1)	206(4.1)	<0.001
Stomachache*	546(2.5)	58(1.2)	<0.001	Asthma*	360(1.6)	114(2.3)	0.003
Nausea*	1471(6.7)	211(4.2)	<0.001	Neurological disease*	243(1.1)	139(2.8)	<0.001
Vomit*	713(3.3)	112(2.2)	<0.001	Hypertension*	3669(16.8)	1504(29.9)	<0.001
Diarrhea*	586(2.7)	102(2.0)	0.008	Sex (female)	10677(48.9)	2866(57.1)	<0.001

*Negative. Note: The chi-square and Mann-Whitney U tests were used.

The results of the LR model are shown in Table 2. The Hosmer-Lemeshow GOF P, Cox & Snell R², and Nagelkerke R² for PCR+ individuals were obtained as 0.657, 0.646, and 0.892, respectively. The Pseudo R² (Cox & Snell and Nagelkerke) indicated a strong relationship between the predictors and the target variables. To achieve the optimal LR model, the non-significant predictors were removed from the model based on the highest P. The

remaining variables in the model were significant at the error level of 0.05. The obtained P>0.05 in the Hosmer-Lemeshow GOF indicated that the LR model is reliable. Accordingly, individuals over 75 had the highest effect in classifying the survivor and non-survivor with an OR equal to 5.45 [95% CI, 3.75%-8.17%]. After that, the consciousness status and PO₂<93% had high OR equal to 4.93 [4.32, 5.63] and 4.09 [3.71, 4.53], respectively.

Table 2. Results of fitting logistic regression model in PCR+ individuals

Variables	Estimate	SE	Z	P	Odds Ratio (95% CI)	
Intercept	-4.053	0.199	-20.332	<0.001	-	
Age group (y) [<18]	Age group 2 (18-45)	-0.034	0.203	-0.166	0.868	0.967 (0.659, 1.462)
	Age group 3 (45-65)	0.730	0.198	3.686	<0.001	2.075 (1.428, 3.110)
	Age group 4 (65-75)	1.335	0.199	6.691	<0.001	3.799 (2.607, 5.709)
	Age group 5 (≥75)	1.697	0.198	8.576	<0.001	5.456 (3.757, 8.174)
Sex (female)	0.319	0.038	8.363	<0.001	1.375 (1.276, 1.482)	
Cough*	-0.217	0.040	-5.453	<0.001	0.805 (0.744, 0.870)	
Muscular pain*	-0.308	0.053	-5.847	<0.001	0.735 (0.662, 0.814)	
Respiratory distress*	0.510	0.047	10.853	<0.001	1.665 (1.519, 1.827)	
Consciousness*	1.595	0.068	23.626	<0.001	4.930 (4.321, 5.631)	
Decreased sense of smell*	-0.644	0.280	-2.305	0.021	0.525 (0.292, 0.880)	
Convulsions*	0.736	0.353	2.082	0.037	2.087 (1.020, 4.090)	
Headache*	-0.552	0.103	-5.367	<0.001	0.576 (0.469, 0.702)	
Chest pain*	-0.476	0.124	-3.848	<0.001	0.621 (0.485, 0.788)	
Stomachache*	-0.368	0.164	-2.239	0.025	0.692 (0.497, 0.947)	
Nausea*	-0.279	0.089	-3.141	0.002	0.757 (0.634, 0.899)	
PO ₂ (>than 93%)	1.411	0.051	27.621	<0.001	4.098 (3.710, 4.533)	
Cancer*	1.302	0.144	9.036	<0.001	3.676 (2.772, 4.878)	
Diabetes*	0.357	0.050	7.213	<0.001	1.429 (1.297, 1.575)	
Blood diseases*	1.003	0.258	3.881	<0.001	2.727 (1.634, 4.511)	
Immunodeficiency*	1.012	0.442	2.291	0.022	2.752 (1.121, 6.436)	
Heart disease*	0.430	0.053	8.074	<0.001	1.537 (1.384, 1.706)	
Kidney disease*	1.092	0.119	9.162	<0.001	2.979 (2.358, 3.762)	
Neurological disease*	0.510	0.133	3.846	<0.001	1.665 (1.282, 2.156)	

* Negative. Note: A significant level was considered 0.05.

Cancer and age groups between 65 and 75 had an OR larger than 3. In addition, some variables such as chronic blood diseases, immunodeficiency, and chronic kidney diseases were identified as risk factors for mortality with OR larger than 2. More details are shown in [Table 2](#).

The extracted evaluation indices or classification metrics are sensitivity, specificity, area under the curve (AUC), recall, and precision ([Table 3](#)). In the prediction step, the accuracy of classification in the LR model was calculated at 90.15%.

The ANN model was trained with the entry of all studied predictor variables. Hyperparameters are size (hidden units) and decay (weight decay). Tuning range of and were used to choose the optimal values of size and decay, respectively. The decay weight and size were obtained at 0.4 and 3, respectively. In the optimization process, the number of iterations for converging the model was considered 100.

Table 3. The performance parameters of the LR and ANN model

Index	LR		ANN	
	Train	Test	Train	Test
Sensitivity (%)	82.91	78.30	83.63	78.89
Specificity (%)	92.00	91.72	92.35	92.51
Accuracy (%)	90.48	90.15	90.73	90.27
AUC (%)	85.64	78.47	88.03	80.60
Precision (%)	62.26	60.51	65.49	63.48

As Table 3 indicates, the accuracy of the fitted ANN model was 90.27%. The sensitivity of the ANN classifier on test data was 78.89%. The analyses of the training ANN model showed that the ANN model also had a high classification ability. Furthermore, the relative importance of independent variables in the ANN model was also computed and graphically shown in Figure 2-B. In addition, the architecture of the developed network can be seen in Figure 2-A.

Decreasing consciousness, cough, PO₂, age, chronic kidney disease, fever, headache, smoking status, chronic blood diseases, and diarrhea are the 10 most important predictors of death using the ANN model.

Discussion

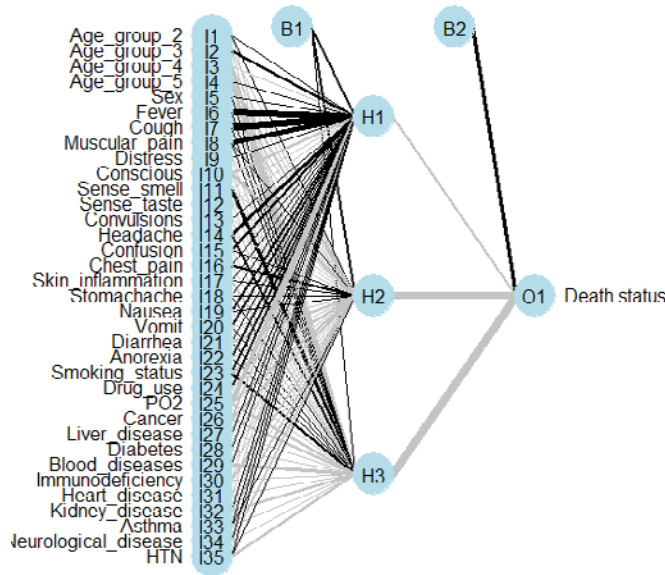
The primary objective of this study was to employ ANN and LR methods to predict the mortality risk associated with COVID-19 in Khorasan Razavi Province, Iran. Our analysis covered data from 26867 individuals, considering 23 variables, including demographic characteristics, symptoms, and underlying health conditions. The predictive models were evaluated based on sensitivity, specificity, AUC, recall, precision, and accuracy. The Hosmer-Lemeshow goodness of fit test confirmed the reliability of the LR model, showcasing its robustness in predicting patient outcomes.

Our findings demonstrated that the ANN and LR models could predict and classify patients effectively, though the ANN model outperforms the LR model. Remarkably, attributes such as decreased consciousness, cough, PO₂ levels, and age emerged as critical indicators in the ANN models. Among underlying health conditions, cancer, blood diseases, kidney diseases, and neurological diseases held the highest significance. Surprisingly, fever, a common symptom of COVID-19, did not prove to be a substantial predictor, demonstrating minimal influence

in distinguishing survivors from non-survivors. Similarly, variables such as asthma decreased taste sensation and sex, and individuals aged 18 to 45 showed relatively limited importance in classification.

A relevant study by Mohammadi et al. in Iran employed similar ANN and LR methods for COVID-19 patient diagnosis, corroborating our findings of the superior accuracy of ANN over LR in diagnostics. While they identified fever, cough, shortness of breath, fatigue, chills, and headache as common symptoms, only cough and headache aligned with our observations. Contrasts were also noted in prevalent underlying conditions: hypertension, diabetes, chronic obstructive pulmonary disease, and coronary heart disease in their study versus cancer, chronic kidney diseases, chronic blood diseases, and chronic neurological diseases [28]. Kaday et al. employed LR and multinomial Naive Bayes classifier approaches to detect COVID-19 and grouped clinical reports into various classes. Their study achieved remarkable accuracy (96.2%) with machine learning techniques, showcasing the potential of these methods in disease classification [29]. Wang et al. utilized logistic models and machine learning techniques to predict the COVID-19 epidemic's trajectory, underlining the versatility of machine learning in addressing multifaceted pandemic dynamics [30]. Similarly, Shaban et al. employed fuzzy inference engines and deep neural networks to detect COVID-19 infection, contributing to the expanding array of machine-learning applications [31]. Numerous other studies have used ANN and machine learning to forecast COVID-19 prevalence, mortality rates, new cases, and deaths, further confirming the consistency and applicability of our machine-learning approach [32-38].

A



B

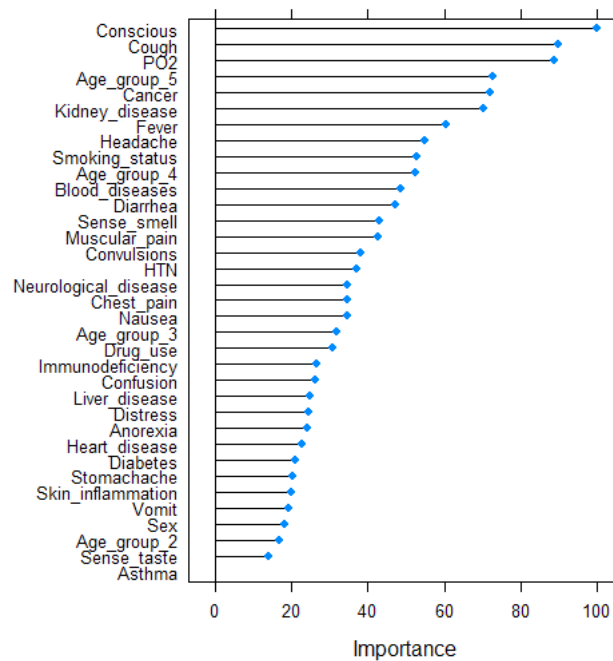


Figure 2. The visualization of network (A) and importance of variables (B) according to the ANN model

Conclusion

The current paper’s findings have presented the importance of some demographics, underlying diseases, and clinical symptoms in classifying survivors vs non-survivors of COVID-19 disease in one of the biggest provinces of Iran, namely Khorasan Razavi. The findings showed that decreasing consciousness, cough, PO₂, and age were the most important characteristics in ANN

models to predict mortality. The most common and significant underlying diseases were cancer, blood disease, kidney disease, and neurological disease. Although fever is one of the most common symptoms of COVID-19, it was not a significant predictor or an influential factor in classifying survivors vs non-survivors. Having asthma, decreased taste sense, sex, and age between 18 to 45 years had the least importance and contribution in classifying survivors vs non-survivors. Chronic blood dis-

eases, immunodeficiency, and chronic kidney diseases were identified as risk factors for mortality. In the end, we hope our results help the medical community.

Study limitations

Due to the elusive nature of the disease, its symptoms were subject to continuous evolution, with some variables added to the database at a later stage. Notably, gastrointestinal symptoms represent one such example. Another limitation is the study population, which is limited to hospitalized patients. The information collected related to underlying diseases relied on self-reporting from the individuals, thus introducing potential bias into the data.

Ethical Considerations

Compliance with ethical guidelines

The study was approved by the Ethics Committee of the [Mashhad University of Medical Sciences](#) in 2021 (Code: IR.MUMS.REC.1400.248).

Funding

This research did not receive any grant from funding agencies in the public, commercial, or non-profit sectors.

Authors contributions

Conceptualization, methodology, formal analysis and software: Nasrin Talkhi; Investigation: Nooshin Akbari Sharak; Data curation: Maryam Salari; Data management: Seyed Masoud Sadati; Writing the original draft: Nasrin Talkhi and Nooshin Akbari Sharak; Review and editing: Nooshin Akbari Sharak and Razieh Yousefi; Supervision and validation: Mohammad Taghi Shakeri.

Conflict of interest

The authors declared no conflict of interest.

Acknowledgements

The authors would like to acknowledge and thank all the experts who participated in the study.

References

- [1] Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A novel coronavirus from patients with pneumonia in China, 2019. *The New England Journal of Medicine*. 2020; 382(8):727-33. [DOI:10.1056/NEJMoa2001017] [PMID]
- [2] Esfahanian F, SeyedAlinaghi S A, Janfaza N, Tantuoyir M M. Prevalence of hypothyroidism among patients with covid-19 in Tehran City, Iran. *Iranian Journal of Health Sciences*. 2023; 11(2):87-92. [DOI:10.32598/ijhs.11.2.942.1]
- [3] Brem A, Viardot E, Nylund PA. Implications of the coronavirus (COVID-19) outbreak for innovation: Which technologies will improve our lives? *Technological Forecasting and Social Change*. 2021; 163:120451. [DOI:10.1016/j.techfore.2020.120451] [PMID]
- [4] Abdi M. Coronavirus disease 2019 (COVID-19) outbreak in Iran: Actions and problems. *Infection Control & Hospital Epidemiology*. 2020; 41(6):754-5. [DOI:10.1017/ice.2020.86] [PMID]
- [5] Yadav M, Perumal M, Srinivas M. Analysis on novel coronavirus (COVID-19) using machine learning methods. *Chaos, Solitons & Fractals*. 2020; 139:110050. [DOI:10.1016/j.chaos.2020.110050] [PMID]
- [6] Pontoh RS, Solichatus Z, Hidayat Y, Aldella R, Mahardika Jiwani N, Sukono. Covid-19 modelling in South Korea using A Time Series Approach. *International Journal of Advanced Science and Technology*. 2020; 29(7):1620-32. [Link]
- [7] Mohammadzadeh Rostami F, Nasr Esfahani BN, Ahadi AM, Shalibeik S. A review of novel coronavirus, severe acute respiratory syndrome coronavirus 2 (sars-cov-2). *Iranian Journal of Medical Microbiology*. 2020; 14(2):154-61. [DOI:10.30699/ijmm.14.2.154]
- [8] Wu Z, Huang Z, Wu Y, Jin Y, Wang Y, Zhao H, et al. Risk stratification for mortality in cardiovascular disease survivors: A survival conditional inference tree analysis. *Nutrition, Metabolism & Cardiovascular Diseases*. 2021; 31(2):420-8. [DOI:10.1016/j.numecd.2020.09.029] [PMID]
- [9] Belhadi A, Kamble S, Jabbour CJC, Gunasekaran A, Ndubisi NO, Venkatesh M. Manufacturing and service supply chain resilience to the COVID-19 outbreak: Lessons learned from the automobile and airline industries. *Technological Forecasting and Social Change*. 2021; 163:120447. [DOI:10.1016/j.techfore.2020.120447] [PMID]
- [10] Balli S. Data analysis of Covid-19 pandemic and short-term cumulative case forecasting using machine learning time series methods. *Chaos, Solitons, and Fractals*. 2021; 142:110512. [DOI:10.1016/j.chaos.2020.110512] [PMID]
- [11] Srinivas M, Lin YY, Liao HYM. Deep dictionary learning for fine-grained image classification. Paper presented at: 2017 IEEE International Conference on Image Processing (ICIP). 17-20 September 2017; Beijing, China. [DOI:10.1109/ICIP.2017.8296398]
- [12] Rabie AH, Saleh AI, Mansour NA. A Covid-19's integrated herd immunity (CIHI) based on classifying people vulnerability. *Computers in Biology and Medicine*. 2022; 140:105112. [DOI:10.1016/j.compbiomed.2021.105112] [PMID]
- [13] Ing EB, Miller NR, Nguyen A, Su W, Bursztyn LLCD, Poole M, et al. Neural network and logistic regression diagnostic prediction models for giant cell arteritis: Development and validation. *Clinical Ophthalmology (Auckland, NZ)*. 2019; 13:421-30. [DOI:10.2147/OPTH.S193460] [PMID]

- [14] Gour M, Jain S. Stacked convolutional neural network for diagnosis of covid-19 disease from X-ray images. Preprint from arXiv. 2020. [Link]
- [15] Battineni G, Sagaro GG, Chinatalapudi N, Amenta F. Applications of machine learning predictive models in the chronic disease diagnosis. *Journal of Personalized Medicine*. 2020; 10(2):21. [DOI:10.3390/jpm1002021] [PMID]
- [16] Selvi RT, Muthulakshmi I. An optimal artificial neural network based big data application for heart disease diagnosis and classification model. *Journal of Ambient Intelligence and Humanized Computing*. 2021; 12(6):6129-39. [DOI:10.1007/s12652-020-02181-x]
- [17] Worldometers. Age, sex, existing conditions of covid-19 cases and deaths [Internet]. 2020 [Updated 13 May 2020]. Available from: [Link]
- [18] Hosmer Jr DW, Lemeshow S, Sturdivant RX. *Applied logistic regression*. New Jersey: John Wiley & Sons; 2013. [DOI:10.1002/9781118548387]
- [19] Yee TW. *Categorical Data Analysis*. In: Yee TW, editor. *Vector generalized linear and additive models*. New York: Springer; 2015. [DOI:10.1007/978-1-4939-2818-7_14]
- [20] Brydon H, Blignaut R, Jacobs J. A weighted bootstrap approach to logistic regression modelling in identifying risk behaviours associated with sexual activity. *SAHARA: Journal of Social Aspects of HIV/AIDS Research Alliance*. 2019; 16(1):62-9. [DOI:10.1080/17290376.2019.1636708] [PMID]
- [21] Swaminathan S. Logistic regression - detailed overview. 2018. *Towards Data Science*. 2018; 1-3. [Link]
- [22] Schober P, Vetter TR. Logistic regression in medical research. *Anesthesia and Analgesia*. 2021; 132(2):365-6. [DOI:10.1213/ANE.0000000000005247] [PMID]
- [23] Bagley SC, White H, Golomb BA. Logistic regression in the medical literature: Standards for use and reporting, with particular attention to one medical domain. *Journal of Clinical Epidemiology*. 2001; 54(10):979-85. [DOI:10.1016/S0895-4356(01)00372-9] [PMID]
- [24] Moein S. *Artificial Neural Network for Medical Diagnosis*. In: *Medical Diagnosis Using Artificial Neural Networks*. Hershey, PA: IGI Global; 2014. [DOI:10.4018/978-1-4666-6146-2.ch007]
- [25] Nellore SB. Various performance measures in Binary classification- An Overview of ROC study. *IJSET-International Journal of Innovative Science, Engineering & Technology*. 2015; 2(9):596-605. [Link]
- [26] Tong Z, Liu Y, Ma H, Zhang J, Lin B, Bao X, et al. Development, validation and comparison of artificial neural network models and logistic regression models predicting survival of unresectable pancreatic cancer. *Frontiers in Bioengineering and Biotechnology*. 2020; 8:196. [DOI:10.3389/fbioe.2020.00196] [PMID]
- Fawcett T. An introduction to ROC analysis. *Pattern Recognition Letters*. 2006; 27(8):861-74. [DOI:10.1016/j.patrec.2005.10.010]
- [27] Mohammadi F, Pourzamani H, Karimi H, Mohammadi M, Mohammadi M, Ardalan N, et al. Artificial neural network and logistic regression modelling to characterize COVID-19 infected patients in local areas of Iran. *Biomedical Journal*. 2021; 44(3):304-16. [DOI:10.1016/j.bj.2021.02.006] [PMID]
- [28] Khanday AMUD, Rabani ST, Khan QR, Rouf N, Mohi Ud Din M. Machine learning based approaches for detecting COVID-19 using clinical text data. *International Journal of Information Technology*. 2020; 12(3):731-9. [DOI:10.1007/s41870-020-00495-9] [PMID]
- [29] Wang P, Zheng X, Li J, Zhu B. Prediction of epidemic trends in COVID-19 with logistic model and machine learning technics. *Chaos, Solitons & Fractals*. 2020; 139:110058. [DOI:10.1016/j.chaos.2020.110058] [PMID]
- [30] Shaban WM, Rabie AH, Saleh AI, Abo-Elsoud MA. Detecting COVID-19 patients based on fuzzy inference engine and Deep Neural Network. *Applied Soft Computing*. 2021; 99:106906. [DOI:10.1016/j.asoc.2020.106906] [PMID]
- [31] Toğa G, Atalay B, Toksari MD. COVID-19 prevalence forecasting using Autoregressive Integrated Moving Average (ARIMA) and Artificial Neural Networks (ANN): Case of Turkey. *Journal of Infection and Public Health*. 2021; 14(7):811-6. [DOI:10.1016/j.jiph.2021.04.015] [PMID]
- [32] Hasan N. A methodological approach for predicting COVID-19 epidemic using EEMD-ANN hybrid model. *Internet of Things*. 2020; 11:100228. [DOI:10.1016/j.iot.2020.100228]
- [33] Lin JK, Chien TW, Wang LY, Chou W. An artificial neural network model to predict the mortality of COVID-19 patients using routine blood samples at the time of hospital admission: Development and validation study. *Medicine*. 2021; 100(28):e26532. [DOI:10.1097/MD.00000000000026532] [PMID]
- [34] Dhamodharavadhani S, Rathipriya R. COVID-19 mortality rate prediction for India using statistical neural networks and gaussian process regression model. *African Health Sciences*. 2021; 21(1):194-206. [DOI:10.4314/ahs.v21i1.26] [PMID]
- [35] Tamang S, Singh P, Datta B. Forecasting of Covid-19 cases based on prediction using artificial neural network curve fitting technique. *Global Journal of Environmental Science and Management*. 2020; 6(Special Issue (Covid-19)):53-64. [Link]
- [36] Mollalo A, Rivera KM, Vahedi B. Artificial neural network modeling of novel coronavirus (COVID-19) incidence rates across the continental United States. *International Journal of Environmental Research and Public Health*. 2020; 17(12):4204. [DOI:10.3390/ijerph17124204] [PMID]
- [37] Shawaqfah M, Almomani F. Forecast of the outbreak of covid-19 using artificial neural network: Case study Qatar, Spain, and Italy. *Results in Physics*. 2021; 27:104484. [DOI:10.1016/j.rinp.2021.104484] [PMID]

This Page Intentionally Left Blank
