J Appl Biotechnol Rep. 2023 June;10(2):1018-1024

Journal of
Applied Biotechnology Reports

doi 10.30491/JABR.2022.328558.1492

Original Article

Predicting COVID-19 Models for Death with Three Different Decision Algorithms: Analysis of 600 Hospitalized Patients

Masoud Arabfard [1], **Ali Najafi** [2*], **Ehsan Rezaei** [2*]

[1] Chemical Injuries Research Center, Systems Biology and Poisonings Institute, Baqiyatallah University of Medical Science, Tehran, Iran
[2] Molecular Biology Research Center, Systems Biology and Poisonings Institute, Baqiyatallah University of Medical Science, Tehran, Iran

Corresponding Authors: Ali Najafi, PhD, Associate Professor, Molecular Biology Research Center, Systems Biology and Poisonings Institute, Baqiyatallah University of Medical Science, Tehran, Iran. E-mail: najafi74@yahoo.com; Ehsan Rezaei, PhD, Molecular Biology Research Center, Systems Biology and Poisonings Institute, Baqiyatallah University of Medical Science, Tehran, Iran. E-mail: rezaie.ehs@gmail.com

Received February 7, 2022; **Accepted** July 12, 2022; **Online Published** June 18, 2023

Abstract

Introduction: COVID-19 virus has caused the biggest pandemic in a decade. The acute respiratory syndrome caused by this virus can lead to the death of patients. Death is very likely in people with severe forms of lung disease. Early identification of patients with severe disease can be very effective in the prevention of death outcomes with improves triage strategies and timely medical actions. The aim of this study was the prediction of COVID-19 models for death with three different decision algorithms with analysis of hospitalized patients.

Materials and Methods: In this study, in a retrospective analysis of 600 COVID-19 patients, we apply three decision tree algorithms including the C5.0, CRT, and CHAID using all related factors to the disease including demographic data, history of exposure, clinical signs, and symptoms, laboratory results, chest X-ray or computed tomography (CT) scans, underlying illness, treatment steps, and outcomes of each patient to build several models predicting the death of Covid-19 infection.

Results: The accuracy of the models was above 90%. Overall, in our retrospective analysis, age, hypertension, lung disease, O2Sat, diabetes, and body temperature, respectively are the most important factors that can affect the mortality rate of COVID-19 patients. Among them, age, and hypertension are common in our applied three models.

Conclusions: The design of such models and apply in hospitals can help to improve disease management and decrease the mortality rate spatially in about recent pandemic.

Keywords: Predicting model, COVID-19, Decision Tree, Data Mining, Decision Algorithms

Citation: Arabfard M, Najafi A, Rezaei E. Predicting COVID-19 Models for Death with Three Different Decision Algorithms: Analysis of 600 Hospitalized Patients. J Appl Biotechnol Rep. 2023;10(2):1018-24. doi:10.30491/JABR.2022.328558.1492

Introduction

Emerging the new type of corona (COVID-19) virus has become the most important for global health, now. Compared to the severe acute respiratory syndrome coronavirus (SARS) and Middle East respiratory syndrome (MERS), the COVID-19 virus overwhelmingly surpassed in terms of geographical prevalence, the number of infected people, and overall mortality rate.

Because so far there is no approved treatment for COVID-19, accurate and timely diagnosis of the disease is very important. Also, early identification of patients with severe disease is vital. According to a report by the China Centers for Disease Control and Prevention (CDC), among 44,500 people with COVID-19, 15.8% of them were in the acute or critical phase of the disease. So, most patients with COVID-19 have a mild course of the disease, so they experience a rapid decline, especially during the 7th to 14th day after the onset of symptoms.[1] However, in a part of COVID-19 patients the disease develops into severe respiratory distress, because of overreacting inflammatory response.[2] Current epidemiological data show that the mortality rate of severe COVID-19 patients is about 20 times higher than that of mild patients.[3] Such prominent statistics highlight the need to identify high-risk patients. Patients who may progress to the severe phase of the disease often require special medical equipment. Therefore, early identification of such patients can help reduce mortality rates by facilitating timely treatment decisions and providing equipment.

To design a valid predicting model, all related factors to the disease including demographic data, history of exposure, clinical signs and symptoms, laboratory results, chest X-ray or computed tomography (CT) scans, underlying illness, treatment steps, and outcomes of each patient should be placed as input factors. Using all data enhances early detection ability and improves triage strategies, creating a

balance between standard medical principles and limited resources. Now, an early prediction model to identification of at-risk COVID-19 patients is rare.[4] In this study, we apply three decision tree algorithms including the C5.0, CRT, and CHAID using all data to build several models predicting the death of COVID-19 infection.

Materials and Methods

Study Design and Data Collection

A retrospective analysis of 600 COVID-19 patients who were admitted to Baqiyatallah Hospital from February to March 2020 confirmed this study, which, affiliated with the University of Medical Sciences of Baqiyatallah, is one of Iran's largest hospitals dedicated to the care of COVID-19 patients. Cases contaminated with SARS-CoV-2 were confirmed in the throat and nasal swab samples by reverse polymerase chain reaction (RT-PCR). These patients' clinical outcomes were tracked in April 2020. The age range is from 22 years to 94 years. The gender composition was 80% male and 20% female. For the disease status, only having (1) or not having (2) and unknown (0) was considered. Values 1 and 2 were considered for all categorical data that have a yes and no answer, respectively, and the missing value was replaced by a value of 0.

The Baqiyatallah University of Medical Sciences, Iran (IR.BMSU.REC.1399.246) ethics committee reviewed and approved this report. Due to the retrospective nature of the study and the confidentiality of patient information, the ethics committee overlooked the need for informed written consent.

Information derived from electronic medical records is gathered from demographic data, history of exposure, clinical signs and symptoms, laboratory results, chest X-ray or computed tomography (CT) scans, underlying illness, treatment steps, and outcomes of each patient using standard forms. All knowledge obtained by an experienced team of physicians was checked and entered into a database of computers. Patients with incomplete details about the studied features or unknown medical records were removed. The disease's onset date was identified as the day the first sign or symptom appeared. After hospitalization, all patients' clinical findings were presented at the end of the study.

Data Mining Analysis

Bioinformatics tools include collecting, mining, analyzing, and finding valuable information are remarkably used in the different fields of medical biology.[5-9] Data Mining is a widely used technique in healthcare.[10] Here, we build several Models to predict the Death of COVID-19 infection. We apply three decision tree algorithms. Here, we briefly describe these algorithms.

The information is presented in raw in CSV format. All variables used in this study can be seen in the supplementary file 1. The data processing corresponds to the records on the epidemic (SARS-CoV-2) COVID-19. The treatment of the information is carried out through the application software Clementine SPSS (version 12.0) for data mining.

Algorithms

A decision tree offers a collection of rules to split data into various classes in order to make some sort of decision about them.[11] In Data Mining, these rules apply to data. As is evident from its name, the decision tree fits the form of a tree, but it is drawn upside down. At the top of the tree is the root. Then after applying rule 1, it is split into branches. At their ends, the branches have leaves. It is very likely that a leave will get into its final shape after applying the first rule and is not divided further. The method continues until a tree

Table 1. Evaluation Metrics for Binary Classification

Metric	Definition	Formula
False Positive Rate	Means the wrong detection rate for identifying the negative sample to the positive.	FPR=FP/(FP+TN)
False Negative Rate	Means the wrong detection rate for identifying a positive to negative.	FNR=FN/(FN+TP)
True Negative Rate	Alternatively, Specificity, Ratio of true negatives to total negatives in the data.	TNR=TN/(FP+TN)
Negative predictive value	Represents the probability that a person does not have a condition, given a negative test result.	NPV=TN/(TN+FN)
Accuracy	In general, accuracy means that the model accurately predicts the output. Looking carefully, you can immediately know if the model is trained correct or not and how it works in general.	ACC=(TN+TP)/(TN+FN+TP+FP)
Precision	The ratio of the sum of the correct items categorized by the algorithm of a given class is the total number of cases in which the algorithm is classified correctly or incorrectly in that class.	Precision=TP/(TP+FP)
Recall	Alternatively, Sensitivity, the ratio of the number of correct items categorized by an algorithm from a class to the number of items in the class.	Recall=TP/(TP+FN)
F-Measure (F1-Score)	Based on the calculations performed for the Precision and Recall metrics, we can calculate the F-Measure weighted value in this step. The F-Measure is a good parameter for assessing the quality of the classification.	F=2*(Precision*Recall)/(Precision + Recall)
Matthews correlation coefficient (MCC)	a more reliable statistical rate which produces a high score only if the prediction obtained good results in all of the four confusion matrix categories	MCC=TP*TN - FP*FN / sqrt((TP+FP)*(TP+FN)*(TN+FP)*(TN+FN))

type is drawn into the entire data collection. In terms of any goal measure, all the leave nodes are the decisions.[12,13] There are a variety of Decision Tree Implementation Algorithms, such as C5.0, CRT, CHAID, and so on, which we can use for data mining purposes. The evaluation metrics of the algorithms are mentioned in Table 1.

C5.0 Algorithm

The algorithm C5.0 is a C4.5 extension. A simple improvement in C5 over C4.5 is that, depending on the severity of their effect on the system, it handles all the errors with individual classification. It builds classifiers that help to reduce the cost of misclassification rather than the high toll of error. This C5 characteristic is known as variable costs of misclassification. An algorithm used to create a decision tree developed by Ross Quinlan is C4.5. C4.5 is an expansion of an earlier ID3 algorithm from Quinlan. For classification, the decision trees created by C4.5 can be used and, for this purpose, C4.5 is sometimes referred to as a statistical classifier.[14,15]

CART Algorithm (Gain Index)

Breiman et al., suggested the CART algorithm, and it is a method of binary segmentation.[16] The condition for splitting is determined according to the Gini index. Each process of splitting involves the information being separated into two subsets. Subsequently, to decide the next test attribute, each subset is further divided; the process of splitting is continued until the data can no longer be separated.

Pruning is performed after the CART algorithm is qualified. As the basis for pruning, the total error rate is used. The smallest tree provides the most effective classification (i.e., the tree with the least number of layers). For a target variable representing continuous and categorical data, the CART algorithm is applicable. If continuous data is represented by the goal variable, then the regression tree can be used. A classification tree may be used if the target variable includes categorical data.[17,18]

CHAID Algorithms (Chi-Square Test)

A chi-squared test ($\chi 2$, chi-square statistic) is implemented in the CHAID algorithm proposed by Kass et al., to evaluate the splitting condition.[19] The greater the value determined by $\chi 2$, the greater the degree of dependence and the probability value of the variable is primarily used to measure the degree of dependence between multiple variables. In addition, to estimate all the possible predictive variables, a probability value is used to determine whether to continue the splitting process in the CHAID algorithm.

Each node is branched on the basis of the chosen dependent variables in the CHAID branching process, and the chi-squared test is used as the branching norm. This means that the branching is carried out whether or not the classification attribute is essential. They are combined into the same division if the divisions have no major difference. Conversely, the branch is maintained and the splitting process is done on the next layer if the branches vary greatly.[20,21]

Results

The Clementine SPSS platform was used in our experiment.[22] It is well-known data mining software that supports a wide range of data mining algorithms with a friendly user graphical user interface. All models were built using Train and Test data (The data were divided into two categories: 80% training and 20% testing).

Here, we discuss the obtained prediction models. In the C5.0 decision tree model, the attribute demographic (age, hypertension and lung disease) and laboratory appears as the first splitting attribute. This indicates the importance of this information. The model can be interpreted as follows: If the patient suffers from other diseases, the model predicts death, otherwise recovery is predicted. According to this model, patients who have fewer problems with previous disease or whose test results are normal are more likely to survive COVID-19 infections. This could be due to the strength of their immune system.[8]

In the CHAID decision tree model, gender characteristics, age, hypertension, lung disease, Diabetes and laboratory data appear as the first important characteristics of division. Information such as O2Sat and Cough play an important role in laboratory data. Information such as travel and readmission are also important parameters in diagnosing death.

In the CRT decision tree model, laboratory data are more important in the model. Information such as O2Sat and Temperature body plays an important role in laboratory data. Also, demographic data (age and hypertension) and readmission are important parameters in diagnosing death. This shows the importance of this information. According to this model, patients who have more problems with laboratory results are more likely to Death from COVID-19 infections.

Information about the confusion matrix for all training and testing data can be seen in detail in Table 2.

Table 2. The Confusion Matrix for all Training and Test Data in Dataset

		CHAID		C5.0		CRT	
		Actual True	Actual False	Actual True	Actual False	Actual True	Actual False
Train	Predict True	26	8	24	0	21	3
	Predict False	9	450	11	458	15	455
Test	Predict True	5	4	4	0	5	1
	Predict False	1	95	2	99	1	98

Table 3. The Evaluation Metrics for Data Mining Models for Training Data

Measure	CHAID	C5.0	CRT
Sensitivity	0.7429	0.6857	0.5833
Specificity	0.9825	1	0.9934
Precision	0.7647	1	0.875
Negative Predictive Value	0.9804	0.9765	0.9681
False Positive Rate	0.0175	0	0.0066
False Negative Rate	0.2571	0.3143	0.4167
Accuracy	0.9655	0.9777	0.9636
F1-Score	0.7536	0.8136	0.7
Matthews Correlation Coefficient	0.7352	0.8183	0.6973

Table 4. The Evaluation Metrics for Data Mining Models for Test Data

Measure	CHAID	C5.0	CRT
Sensitivity	0.8333	0.6667	0.8333
Specificity	0.9596	1	0.9899
Precision	0.5556	1	0.8333
Negative Predictive Value	0.9896	0.9802	0.9899
False Positive Rate	0.0404	0	0.0101
False Negative Rate	0.1667	0.3333	0.1667
Accuracy	0.9524	0.981	0.981
F1-Score	0.6667	0.8	0.8333
Matthews Correlation Coefficient	0.6575	0.8084	0.8232

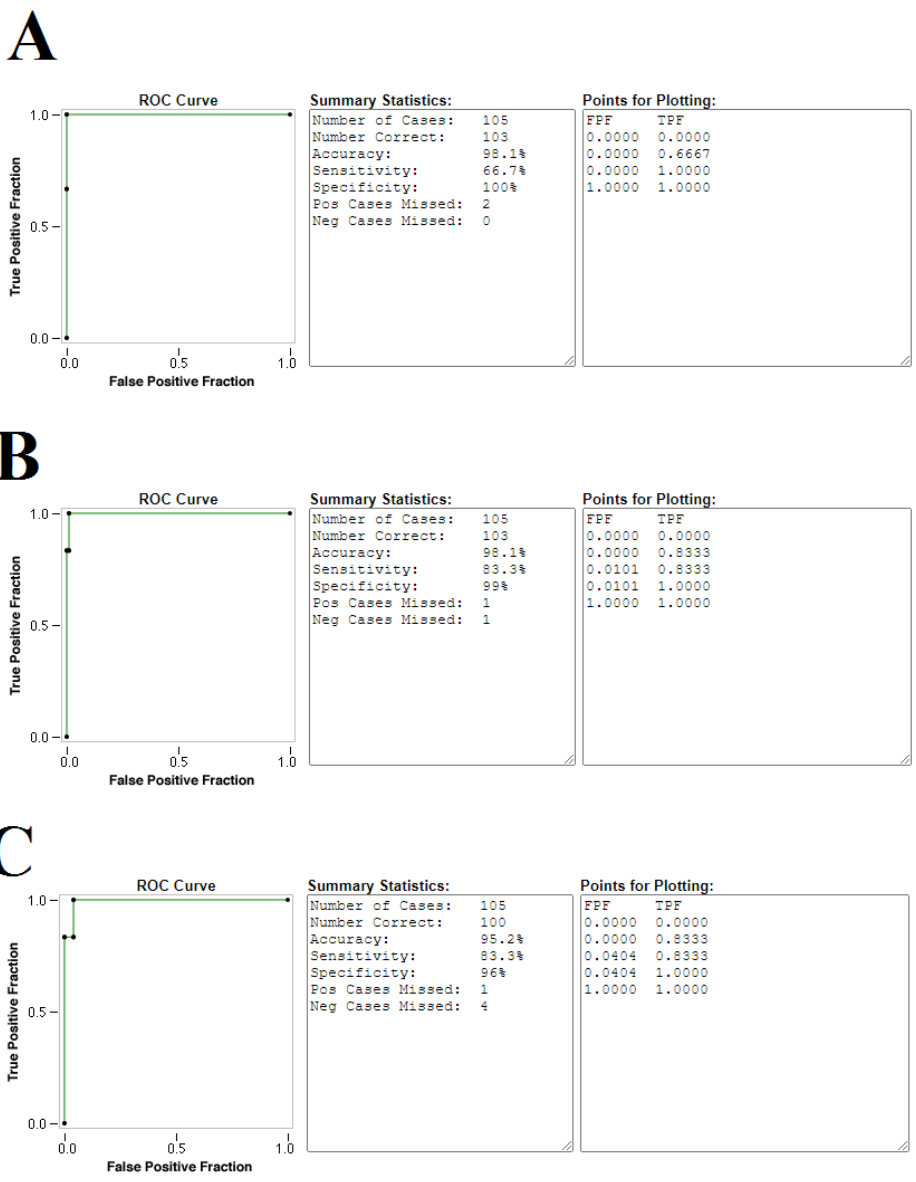

Figure 1. A) ROC Train and Test C5, **B)** ROC Train and Test CRT, and **C)** ROC Train and Test CHAID

Table 3 and Table 4 summarize the evaluation metrics for data mining models for training and testing data, respectively. In the training data, all three models are very similar in terms of accuracy, but the sensitivity of the CHAID model is higher than the others and the Specificity of the C5 model are higher than the others. In addition, the F1-Score is higher in C5 in terms of model fit among these three trees. However, in general, we see that all three models have high accuracy in Death diagnosis. All information related to training data can be seen in Table 3.

In the test data, the accuracy of CRT and C5 is equal and they are better than CHAID. The highest sensitivity is related to CHAID and CRT and the highest Specificity is related to C5. The F1-Score in terms of suitability introduces CRT, C5, and CHAID, respectively. All evaluation metrics related to test data can be seen in Table 4. A ROC curve is a graph representing the performance of a classifier.[23,24] Figure 1, 2, and 3 shows the ROC for the C5, CRT, and CHAID decision trees, respectively. Comparing ROC curves for all 3 models, we conclude that the selection of each model alone can be used to predict with appropriate accuracy.

We use a real dataset with three decision tree data mining algorithms in this research that are known to generate highly precise models. However, for real-world use the output of all models obtained from this data is satisfactory. The key constraint lies in the scale of the dataset for training. We assume that to strengthen predictions, data collection needs to be increased. Furthermore, more patient data (such as medical history) may be used.

Discussion

Severe respiratory illness develops in approximately 20% of COVID-19 patients with the overall mortality around 2.3%.[25] At now, there is no drug effectively targeting at SARS-CoV-2. Identifying people prone to severe forms of the disease for the prevention of high mortality rate is critical. There are few studies for determining the local risk factor association with Iranian COVID-19 patients.[26] However, several studies in the US, Europe and mainly china have been recently accomplished.[27-29] We evaluated the potential risk factors associated with severe cases at admission in a retrospective cohort of 600 COVID-19 patients in the Baqiyatallah Hospital. We used tree algorithms to build models predicting the death of COVID-19 infection including C5.0, CRT, and CHAID algorithms. The C5.0 decision tree model indicated the importance of pre-existing non-communicable comorbidities including diabetes, cancers, cardiovascular diseases, hypertension, chronic kidney diseases, chronic pulmonary diseases, and other chronic diseases. The C5.0 algorithm results are in agreement with many previous studies.[30-32] For example, Bai et al., developed models to find the mild patients who are easily go to the severe/critical cases using AI-based methods and logistic regression respectively.[33] They found that the older age and comorbidity with hypertension could be regarded to be the most important risk factors for malignant progression.[33] Similar what was mentioned in the Bai study, our C5.0 algorithm results suggested that the hypertension and lung disease as the most important comorbidities and the patient's age as the key demographic data are the risk factors association with severe respiratory illness. In the CHAID decision tree model, we found that diabetes and O2Sat also important parameters in diagnosing death. Previously, Akbariqomi et al in the study on the same target community with our study indicated that the diabetic COVID-19 patients are at a higher risk of complications and higher in-hospital mortality during hospitalization. They showed that from 595 hospitalized COVID-19 patients, 148 patients (24.9%) had diabetes. In comparison with non-diabetic patients, diabetic patients had more comorbidities (eg, hypertension [48.6% vs. 22.3%; $p<0.001$]); had higher levels of white blood cell count, neutrophil count, C-reactive protein, erythrocyte sedimentation rate and blood urea nitrogen, and had a higher proportion of patchy ground-glass opacity in chest computed tomography findings (52.7% vs. 25.7%; $p<0.001$). Finally, failure in the treatment and patient death was significantly higher in diabetic patients compared to non-diabetic patient (17.8% vs. 8.7%; $p = 0.003$).[34] The CRT decision tree model results confirmed the derived results from other two models so that the age, hypertension and O2Sat are important parameters in the treatment failure and death results. In addition, high temperature and readmission are the risk factors exclusively in this model. Overall, the age and hypertension are common in our applied three models. There are several case series in the different countries that have been demonstrated the effect of hypertension on COVID-19 severity. In a large case series from China, an overall case fatality rate of 2.3% (1023 of 44,672 confirmed cases) was found that increased to 6.0% for people with hypertension.[35] Also, in the largest epidemiological study to date on 17 million health records in England shows hypertension alone was associated with slightly increased risk.[36] The mechanism by which hypertension can increase the risk of death in COVID-19 patients is complex and may well relate to the age or underlying co-morbidity e.g. cardiovascular diseases.[37] Previous study on 70-79 year age group have been reported that the age can significantly increase both COVID-19 case fatality rates and hypertension prevalence to 8.0% and over 50% respectively.[38] Overall, in Iranian hospitalized COVID-19 patients, age, hypertension, lung disease, O2Sat, diabetes, and body temperature, respectively are the most important factor that can effects on mortality rate of COVID-19 patients. Co-morbidity to two or more mentioned parameters leads to increased risk of death from COVID-19 in our retrospective analysis.

Conclusion

We developed several models for predicting the death of COVID-19 infection in this paper. In the Clementine SPSS software, our models are constructed using the data mining algorithms of the CHAID, CRT, and C5.0 decision tree. The Model Decision Tree revealed that patients with no prior history of illness or laboratory testing issues were more likely to survive. In predicting patient stability, it was found that demographic characteristics are significant. Patients can be categorized after visiting the emergency department and people with age, hypertension, lung disease, O2Sat, diabetes, and body temperature, can be classified as high risk. By examining the ROC Curve for all three designed models, the results of all the models are very close to each other. Special cares can be taken for these people before they enter the severe phase of the disease. These measures include the type of medication prescribed or the priority of hospitalization in the intensive care unit. We assume that the efficiency of predictive models can be improved by using more patient data. In order to collect more information on patients with COVID-19 infection, we plan to add new patients directly to our data for future work. High accuracy in all models does not give superiority to a particular model. It is suggested that for further studies, these decision trees be evaluated on larger data.

Authors' Contributions

Conceptualization by ER and AN; Original draft preparation by ER and MA; Writing, review, and editing by MA, AN, and ER; Supervision by AN and ER; Interpreting the data and defining the method by MA. All authors have read and agreed to the published version of the manuscript.

Ethical Approval

For this type of study formal consent is not required. The Baqiyatallah University of Medical Sciences, Iran (IR.BMSU.REC.1399.246) ethics committee reviewed and approved this report.

Conflict of Interest Disclosures

The authors declare that they have no conflicts of interest.

References

1. Feng Z, Yu Q, Yao S, Luo L, Zhou W, Mao X, et al. Early prediction of disease progression in COVID-19 pneumonia patients with chest CT and clinical characteristics. Nat Commun. 2020;11(1):4968. doi:10.1038/s41467-020-18786-x
2. Sorci G, Faivre B, Morand S. Explaining among-country variation in COVID-19 case fatality rate. Sci Rep. 2020;10(1):18909. doi:10.1038/s41598-020-75848-2
3. Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. Lancet. 2020;395(10223):507-13. doi:10.1016/S0140-6736(20)30211-7
4. Gong J, Ou J, Qiu X, Jie Y, Chen Y, Yuan L, et al. A tool for early prediction of severe coronavirus disease 2019 (COVID-19): a multicenter study using the risk nomogram in Wuhan and Guangdong, China. Clin Infect Dis. 2020;71(15):833-40. doi:10.1093/cid/ciaa443
5. Keshtvarz M, Mahboobi M, Kieliszek M, Miecznikowski A, Sedighian H, Rezaei M, et al. Engineering of cytolethal distending toxin b by its reducing immunogenicity and maintaining stability as a new drug candidate for tumor therapy; an in silico study. Toxins. 2021;13(11):785. doi:10.3390/toxins13110785
6. Parvin S, Sedighian H, Sohrabi E, Mahboobi M, Rezaei M, Ghasemi D, et al. Prediction of genes involved in lung cancer with a systems biology approach based on comprehensive gene information. Biochem Genet. 2022;60:1253-73. doi:10.1007/s10528-021-10163-7
7. Rezaie E, Bidmeshki Pour A, Amani J, Mahmoodzadeh Hosseini H. Bioinformatics predictions, expression, purification and structural analysis of the PE38KDEL-scfv immunotoxin against EPHA2 receptor. Int J Pept Res Ther. 2020;26:979-96. doi:10.1007/s10989-019-09901-8
8. Rezaie E, Nekoie H, Miri A, Oulad G, Ahmadi A, Saadati M, et al. Different frequencies of memory B-cells induced by tetanus, botulinum, and heat-labile toxin binding domains. Microb Pathog. 2019;127:225-32. doi:10.1016/j.micpath.2018.12.003
9. Sohrabi E, Rezaie E, Heiat M, Sefidi-Heris Y. An integrated data analysis of mRNA, miRNA and signaling pathways in pancreatic cancer. Biochem Genet. 2021;59:1326-58. doi:10.1007/s10528-021-10062-x
10. Zamfir IC, Iordache AM. A review of data mining techniques in medicine. J Inf Syst Oper Manag. 2020; 14(1):93-106.
11. Alsagheer RH, Alharan AF, Al-Haboobi AS. Popular decision tree algorithms of data mining techniques: a review. Int J Comput Sci Mob Computing. 2017;6(6):133-42.
12. Dey A. Machine learning algorithms: a review. Int J Comput Sci Inf Technol. 2016;7(3):1174-9.
13. Mohamed WN, Salleh MN, Omar AH. A comparative study of reduced error pruning method in decision tree algorithms. 2012 IEEE International conference on control system, computing and engineering, IEEE. 2012. pp. 392-397. doi:10.1109/ICCSCE.2012.6487177
14. Hssina B, Merbouha A, Ezzikouri H, Erritali M. A comparative study of decision tree ID3 and C4. 5. Int J Adv Comput Sci Appl. 2014;4(2):13-9.
15. Pandya R, Pandya J. C5. 0 algorithm to improved decision tree with feature selection and reduced error pruning. Int J Comput Appl. 2015;117(16):18-21.
16. Breiman L, Friedman JH, Olshen RA, Stone CJ, editors. Classification and Regression Trees. 1983.
17. Loh WY. Classification and regression trees. Wiley interdisciplinary reviews: data mining and knowledge discovery. 2011;1(1):14-23. doi:10.1002/widm.8
18. Steinberg D. CART: classification and regression trees. The top ten algorithms in data mining: Chapman and Hall/CRC. 2009. pp. 193-216.
19. Kass GV. An exploratory technique for investigating large quantities of categorical data. J R Stat Soc C: Appl Stat. 1980;29(2):119-27. doi:10.2307/2986296
20. Milanović M, Stamenković M. CHAID decision tree: Methodological frame and application. Economic Themes. 2016;54(4):563-86. doi:10.1515/ethemes-2016-0029
21. Ritschard G. CHAID and earlier supervised tree methods. Contemporary Issues in Exploratory Data Mining in the Behavioral Sciences London: Routledge. 2013. pp.48-74.
22. Al Ghoson AM. Decision tree induction & clustering

techniques in SAS enterprise miner, SPSS clementine, and IBM intelligent miner a comparative analysis. Int J Manag Inf Syst. 2010;14(3). doi:10.19030/ijmis.v14i3.841
23. Yu CS, Lin YJ, Lin CH, Wang ST, Lin SY, Lin SH, et al. Predicting metabolic syndrome with machine learning models using a decision tree algorithm: retrospective cohort study. JMIR Med Inform. 2020;8(3):e17110. doi:10.2196/17110
24. Toraih EA, Elshazli RM, Hussein MH, Elgaml A, Amin M, El-Mowafy M, et al. Association of cardiac biomarkers and comorbidities with increased mortality, severity, and cardiac injury in COVID-19 patients: a meta-regression and decision tree analysis. J Med Virol. 2020;92(11): 2473-88. doi:10.1002/jmv.26166
25. Novel CPERE. The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19) in China. Zhonghua Liu Xing Bing Xue Za Zhi. 2020;41(2):145-51. doi:10.3760/cma.j.issn.0254-6450.2020.02.003
26. Zali A, Gholamzadeh S, Mohammadi G, Looha MA, Akrami F, Zarean E, et al. Baseline characteristics and associated factors of mortality in COVID-19 patients; an analysis of 16000 cases in Tehran, Iran. Arch Acad Emerg Med. 2020;8(1):e70.
27. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet. 2020;395(10223): 497-506. doi:10.1016/S0140-6736(20)30183-5
28. Lippi G, Mattiuzzi C, Sanchis-Gomar F, Henry BM. Clinical and demographic characteristics of patients dying from COVID-19 in Italy vs China. J Med Virol. 2020;92(10):1759-60. doi:10.1002/jmv.25860
29. Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, et al. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus–infected pneumonia in Wuhan, China. Jama. 2020;323(11):1061-9. doi:10.1001/jama.2020.1585
30. Caramelo F, Ferreira N, Oliveiros B. Estimation of risk factors for COVID-19 mortality-preliminary results. MedRxiv. 2020:2020-02. doi:10.1101/2020.02.24.20027268
31. Ssentongo P, Ssentongo AE, Heilbrunn ES, Ba DM, Chinchilli VM. Association of cardiovascular disease and 10 other pre-existing comorbidities with COVID-19 mortality: A systematic review and meta-analysis. PloS One. 2020;15(8):e0238215. doi:10.1371/journal.pone.0238215
32. Yang J, Zheng YA, Gou X, Pu K, Chen Z, Guo Q, et al. Prevalence of comorbidities and its effects in patients infected with SARS-CoV-2: a systematic review and meta-analysis. Int J Infect Dis. 2020;94:91-5. doi:10.1016/j.ijid.2020.03.017
33. Fang C, Bai S, Chen Q, Zhou Y, Xia L, Qin L, et al. Deep learning for predicting COVID-19 malignant progression. Med Image Anal. 2021;72:102096. doi:10.1016/j.media.2021.102096
34. Akbariqomi M, Hosseini MS, Rashidiani J, Sedighian H, Biganeh H, Heidari R, et al. Clinical characteristics and outcome of hospitalized COVID-19 patients with diabetes: A single-center, retrospective study in Iran. Diabetes Res Clin Pract. 2020;169:108467. doi:10.1016/j.diabres.2020.108467
35. Surveillances V. The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19)—China, 2020. China CDC weekly. 2020;2(8):113-22.
36. Williamson EJ, Walker AJ, Bhaskaran K, Bacon S, Bates C, Morton CE, et al. Factors associated with COVID-19-related death using OpenSAFELY. Nature. 2020;584 (7821):430-6. doi:10.1038/s41586-020-2521-4
37. Clark CE, McDonagh ST, McManus RJ, Martin U. COVID-19 and hypertension: risks and management. A scientific statement on behalf of the British and Irish Hypertension Society. J Hum Hypertens. 2021;35(4):304-7. doi:10.1038/s41371-020-00451-x
38. Kearney PM, Whelton M, Reynolds K, Muntner P, Whelton PK, He J. Global burden of hypertension: analysis of worldwide data. Lancet. 2005;365(9455):217-23. doi:10.1016/S0140-6736(05)17741-1