IRAQI Academic Scientific Journals

العراقية المجلات الأكاديمية العلمية

University of Kut Journal

ISSN (E): 2616 - 7808 II ISSN (P): 2414 - 7419

js.uokut.edu.iq

k.u.c.j.sci@alkutcollege.edu.iq

Vol. 11 , Issue 1 , June 2026

Predicting & Diagnosing Heart Disease Patients Using Multinomial Logistic Regression

Hatem Abdul Hussein Ali Alquraishi [1] , Prof. Abbas Lafta Kneehr [2]

Abstract

Heart disease is one of the most common diseases and the principal cause of sudden death nowadays. According to the Iraqi Ministry of Health, 28.83% of deaths in Iraq are caused by some kind of heart disease. This is motivated by the worldwide increasing fatality of heart disease (HD) patients each year and the availability of statistical methods that can be used to extract convenient knowledge from data collected from patients, to aid physicians in diagnosing heart disease. A statistical model was developed using multinomial logistic regression (MLR) to predict heart diseases using a training dataset of 201 patients collected from Alkarama Teaching Hospital. The data were statistically analyzed to reduce bias and assess how much the attributes explain the diversity of the dependent variables, which resulted in 95.8%. The overall prediction model had an accuracy of 94.5%. The researchers concluded that the MLR model has high accuracy in predicting heart disease, and that echocardiographic findings of hypokinesia, chest pain, and electrocardiographic results are the top three attributes, in terms of accuracy, for diagnosing heart disease.

Keywords: heart disease, Diagnosis, Prediction models, Multinomial Logistic Regression

Affiliation of Authors

[1, 2] College of Administration & Economics, University of Wasit, Iraq, Wasit, Kut, 52001

[1] h_alqurashi@uowasit.edu.iq
[2] alafta@uowasit.edu.iq

[1] **Corresponding Author**

Paper Info.
Published: Jun. 2026

التنبؤ وتشخيص مرضى أمراض القلب باستخدام الانحدار اللوجستي المتعدد

حاتم عبد الحسين علي القريشي [1] د. عباس لفتة كنيهر [2]

المستخلص

يُعدّ مرض القلب من أكثر الأمراض شيوعاً وأحد الأسباب الرئيسة للوفاة المفاجئة في الوقت الحاضر. ووفقاً لوزارة الصحة العراقية، فإن 28.83% من الوفيات في العراق تعود إلى أحد أنواع أمراض القلب. تنبع أهمية هذه الدراسة من الزيادة العالمية المستمرة في معدلات وفيات مرضى أمراض القلب سنوياً، إضافةً إلى توافر الأساليب الإحصائية القادرة على استخلاص معرفة مفيدة من بيانات المرضى بما يساعد الأطباء على تشخيص أمراض القلب. تم بناء نموذج إحصائي باستخدام الانحدار اللوجستي متعدد الحدود للتنبؤ بأمراض القلب اعتماداً على مجموعة بيانات تدريبية مكوّنة من 201 مريض جُمعت من مستشفى الكرامة التعليمي. خضعت البيانات لتحليل إحصائي بهدف تقليل التحيزات وقياس مدى إسهام المتغيرات المستقلة في تفسير تباين المتغير التابع، حيث بلغت نسبة التفسير 95.8%. وبلغت دقة نموذج التنبؤ الكلية 94.5%. وخلص الباحثون إلى أن نموذج الانحدار اللوجستي المتعدد يتمتع بدقة عالية في التنبؤ بأمراض القلب، كما أظهرت النتائج أن نقص حركية عضلة القلب (Hypokinesia) وألم الصدر ونتائج تخطيط القلب الكهربائي هي أهم ثلاثة متغيرات من حيث الدقة في تشخيص أمراض القلب.

الكلمات المفتاحية: أمراض القلب، التشخيص، نماذج التنبؤ، الانحدار اللوجستي متعدد الحدود

انتساب الباحثين

[1، 2] كلية الادارة والاقتصاد، جامعة واسط، العراق، واسط، الكوت، 52001

[1] h_alqurashi@uowasit.edu.iq
[2] alafta@uowasit.edu.iq

[1] **المؤلف المراسل**

معلومات البحث
تأريخ النشر : حزيران 2026

Introduction

Heart disease is one of the most frequently occurring diseases and the major cause of sudden death in the world nowadays. Most people do not feel the signs of heart disease until it is too late. The World Health Organization (WHO) announced that heart disease is the leading cause of death in both high and low-income countries **[1]**. The Economic and Social Commission of Asia and the Pacific announced that 0.20% of the Asian region's population lives are lost to noncommunicable diseases (NCDs), similar to cardiovascular diseases, cancers, and diabetes **[2]**. The Iraqi Ministry of Health, in its annual report **[3],** reported 140,621 deaths in 2019, 28.83% of those deaths were due to some kind of heart disease, while cancer-related deaths were only 9.33%, and road traffic accidents accounted for 4.9%. This is motivated by the worldwide increasing mortality of heart disease patients each year and the availability of statistical methods that can be used to extract useful knowledge from data collected from patients, to aid physicians in diagnosing heart disease. Statistical analysis is a crucial step in knowledge discovery. It is the inspection of gigantic datasets to find concealed and previously unfamiliar patterns, relationships, and knowledge that are difficult to detect with the naked eye **[4].** The use of statistical analysis is exponentially growing in a lot of fields like analysis of organic compounds **[5]**, financial forecasting **[6]**, healthcare **[7],** and weather forecasting **[8]**. The use of statistical analysis in healthcare is quite new, and it is very useful in providing physicians with a prognosis, as well as a higher understanding of medical data. Medical data analysis tries to deal with real-world health issues in the diagnosis and treatment of diseases. Researchers are implementing statistical analysis techniques in the diagnosis of many diseases, such as diabetes **[9],** stroke **[10],** cancer **[11],** and heart disease**.** Different statistical analysis techniques used in the diagnosis of heart disease have different percentages of accuracy. Multinomial logistic regression [MLR] is being used in a lot of sectors. Like solving problems in risk assessment, it is also used in tourism to uncover a lot of knowledge about tourist behavior and what they like and what they don't. MLR has been used in trade and the stock market to identify factors that affect purchases of goods and stocks. In the healthcare sector, MLR has been used in risk assessment and diagnosis, e.g., the TREAT model (Thoracic Research Evaluation and Treatment model); it is used to calculate the danger of a lung nodule being cancer using patient characteristics that are available to the surgeons who are evaluating the patient, and it is being used in a thoracic surgery clinic in the US.

Related work

[12]: presented a detailed study on making a prediction model for diagnosing HD with the use of MLR, he formed his predictive model based on the data collected from patient's medicinal records, he also tested other statistical models to examine the accuracy of prediction of the models. The researchers have concluded that multinomial logistic regression yields the best accuracy in predicting heart disease, compared to the other models, with an accuracy of 75.60%, while Naive Bayes had an accuracy of 61.42%.

[13]: made an MLR model named Thoracic Research Evaluation and Treatment [TREAT], the TREAT model is an MLR model for lung cancer using logistic regression, which was developed on a sample of 492 patients in Vanderbilt University with 72% lung cancer prevalence. The model results were quite promising: the space under the receiver operating curve was 0.87 with 95% confidence interval, while the Brier score was 0.12. The model has a very high accuracy ($p < 0.001$).

[14]: presented a study on the application of the multinomial logistic regression model for categorical data analysis, and he has concluded that the logistic regression model is a sufficient model for a variety of data where the response variable has more than two categories. He also concluded that MLR does not have any restrictions on the explanatory variables, and it can be used in a variety of disciplines, ranging from social, educational, health, behavioral, and even scientific experiments. The model that he has proved its ability to predict with an accuracy of 80.7%..

[15]: gave a thorough study comparing logistic regression (LR) and boosted trees in predicting patient mortality from giant medical datasets. The c-statistic results of logistic regression were 0.884 with 95% confidence interval. The researchers have concluded that in large datasets of electronic medical records, logistic regression is able to predict patient mortality with high accuracy.

Research method

The researchers began by collecting historical medical data from patients from Alkarama Teaching Hospital in Wasit province, Iraq. In a span of 8 weeks, the researchers were able to collect 98 samples of patients with 3 types of heart diseases, and 103 control samples, which were diagnosed as normal in the cardiologist's eyes. The data was collected through an on-site questionnaire and was supervised by Dr. Abbas Alzirgany, PhD in internal medicine. The data collected consisted of 14 attributes and was statistically analyzed first by making sure that the dependent variable is measured at a nominal level, then the independent variables need to be continuous (scale), ordinal, or nominal. Also, making sure that the dependent variable is mutually exclusive and has exhaustive categories. The data was then processed in SPSS to detect the absence of multicollinearity. Finally, detecting the absence of outliers was done. After making sure that the data was sufficient to the logistic regression standards, likelihood ratio test was made to assert the significance of the 14 attributes which led to having only four significant attributes, the second likelihood ratio test was made for the resulted four attributes which lead to three significant attributes, the third likelihood test was made on the 10 non-significant attributes, which resulted in three significant attributes; The fourth likelihood test was made on the significant attributes from the second and third test which resulted in five significant attributes. The next test was made to analyze the impact of each independent variable on the logistic regression prediction accuracy. At the end, the experiment was conducted in SPSS to assess the accuracy of the predictive logistic regression model.

Multinomial Logistic Regression (MLR)

Regression methods are an essential component of any data analysis process that deals with finding the relation between a response variable and explanatory variables. The logistic regression model is one of the most commonly used regression models for data analysis **[16].** Logistic regression [LR] is considered a special case of linear statistical methods that is used for modeling and forecasting multivariate datasets. The reason why the logistic regression model is so commonly used is because of its ability to handle multiple variables and measurement scales. One of the common branches of logistic regression is the multinomial LR; it extends binary logistic regression in order to encompass categorical dependent variables with multiple levels. Multinomial LR is quite an effective tool for modeling and forecasting datasets that have multiple dependent variables. The equation of the MLR model is shown in the equation below:

$$\hat{Y} = \frac{e^{\hat{\beta}_0 + \sum_{i=1}^{k} \hat{\beta}_i x_i}}{1 + e^{\hat{\beta}_0 + \sum_{i=1}^{k} \hat{\beta}_i x_i}}$$

or $\hat{Y} = \frac{1}{1+e^{-(\hat{\beta}_0+\sum_{i=1}^{k}\hat{\beta}_i x_i)}}$ (1)

Where Ŷ is the dependent variable with nominal scale and more classes than two, xi is the ith autonomous variable with categorical data or continuous data, $\hat{\beta}_0$ is the constant value of the equation, $\hat{\beta}_i$ is the ith coefficient value of the independent variable Xi.

Dataset

A large number of relevant inputs must be considered during diagnosis to accurately diagnose heart disease. Physicians usually rely on all recorded symptoms, patient medical history, medical examination, and laboratory results, as well as the physician's own experience .

The Cleveland database from UCI Machine Learning Repository includes sufficient and carefully chosen attributes for heart disease diagnosis; thus, the data gathered from 201 patients in Alkarama Teaching Hospital was based on those medical attributes with the cardiologist's diagnosis of them as either being normal, having coronary heart disease, having arrhythmias, or having congestive heart failure. But the researchers had to make some changes to the attributes with the help of a cardiologist, due to the difficulty of obtaining them. The new dataset with the modified attributes is laid out in Table 1 below.

Table (1): The 14 attributes and their data type

1	**Age**	**Continuous Data**			
2	Sex	Female “0”		Male “1”	
3	Smoking	Yes “0”		No “1”	
4	Previous family history of HD	Yes “0”		No “1”	
5	Previous attack of angina	Yes (positive) “0”		No (negative) “1”	
6	Fasting blood sugar is> 120 mg/dl	Yes “0”		No “1”	
7	Serum cholesterol in mg/dl.	Normal “0”		Abnormal “1”	
8	Systolic blood pressure	Continuous Data			
9	Diastolic blood pressure	Continuous Data			
10	HR: heart rate achieved.	Continuous Data			
11	Chest pain	Typical angina “0”	Atypical angina “1”	Non-angina pain “2”	Asymptomatic “3”
12	Electrocardiographic results	Normal “0”	ST-T wave abnormality "1"	probable or definite left ventricular hypertrophy "2"	
13	The angina is exercise-induced	Yes“0”		No“1”	
14	Echo finding for hypokinesia	Yes “0”		No “1”	

Results and discussion

Before the testing began, the data set had been normalized with Z-Score normalization to eliminate some of the basis of large numbers in the data, such as the age and the heart rate, which maxes at 145, compared to more important attributes such as chest pain, which only maxes at 3 **[17]**. The Z-Score normalization process was made using SPSS, and the results (Descriptive statistics) are shown in Table 2.

Table (2): The Distributive's statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Age	**201**	**31**	**95**	**58.95**	**14.957**
systolic blood pressure	**201**	**40**	**200**	**120.95**	**23.272**
diastolic blood pressure	**201**	**30**	**100**	**73.11**	**13.496**
heart rate achieved	**201**	**32**	**145**	**88.53**	**16.896**

As seen in Table 2, age, for example, has a minimum value of 31 and a maximum value of 95 with a mean of 58.95 and a standard deviation of 14.957. Keeping the data without normalization will grant a bigger weight to age than it should, compared to an attribute like smoking, which is a binary variable. After the date normalization, the age had a max of 2.41020 and a min of -1.86868. Systolic blood pressure had a max of 3.39694 and a min of -3.47817, diastolic blood pressure had a max of 1.99217 and a min of -3.19470, and heart rate had a max of 3.34198 and a max of -3.34581.

As mentioned above, the dataset attributes were selected based on the Cleveland dataset, and some of the attributes were altered by a cardiologist due to the difficulty of obtaining them. In order to test the validity of the chosen attributes and their effect on the model, the researchers did a likelihood ratio test four times on the attributes.

In the first likelihood ratio test, all 14 attributes were tested, and it resulted in only four significant attributes, namely Chest pain, Electrocardiographic result, angina is exercise-induced, and echo finding for hypokinesia, as shown in Table 3.

Table (3): Significance results of the 14 independent variables

	Likelihood Ratio Tests			
Effect	**(-2) Log Likelihood about Reduced Model**	**Chi-Square**	**df**	**Sig**
Age	68.602^{b}	2.510	3	0.473
systolic blood pressure	67.622^{b}	1530	3	0.675
diastolic blood pressure	66.301^{b}	0.209	3	0.976
heart rate achieved	67.582^{b}	1.490	3	0.685

Sex	68.351[b]	2.259	3	0.520
Smoking	67.830[b]	1.737	3	0.629
Previous family history of heart disease	69.169[b]	3.077	3	0.380
Previous attack of angina	67.987[b]	1.895	3	0.594
fasting blood sugar	67.061[b]	0.969	3	0.809
Serum cholesterol	67.110[b]	1.018	3	0.797
Chest pain	103.515[b]	37.423	9	0.000
electrocardiographic result	128.641	62.549	6	0.000
The angina is exercise-induced	73.127[b]	7.035	3	0.071
Echo finding for hypokinesia	98.401[b]	32.309	3	0.000

In the second likelihood ratio test, the four significant attributes from the first test were tested against each other; the test resulted in only three significant attributes: Chest pain, Electrocardiographic result, and echo finding of hypokinesia, as shown in Table 4.

Table (4): Significance results of the 4 independent variables

Likelihood Ratio Tests				
Effect	**(-2) Log Likelihood about Reduced Model**	**Chi-Square**	**df**	**Sig**
Chest pain	91.477	53.474	9	0.000
Electrocardiographic result	115.338	77.335	6	0.000
The angina is exercise-induced	44.117[b]	6.114	3	0.106
Echo finding for hypokinesia	76.674[b]	38.671	3	0.000

In the third likelihood ratio test, the 10 non-significant attributes from the first test were tested against each other, yielding three significant attributes: diastolic blood pressure, Previous attack of angina, and serum cholesterol, as shown in Table 5.

Table (5): Significance results of the 10 independent variables

Likelihood Ratio Tests				
Effect	**(-2) Log Likelihood about Reduced Model**	**Chi-Square**	**df**	**Sig**
systolic blood pressure	299.730	1.785	3	0.618
diastolic blood pressure	307.736	9.791	3	0.020

heart rate achieved	303.925	5.980	3	0.113
Age	303.656	5.711	3	0.127
Sex	301.752	3.808	3	0.283
Smoking	299.904	1.959	3	0.581
Previous family history of heart disease	303.513	5.568	3	0.135
Previous attack of angina	375.404	77.460	3	0.000
fasting blood sugar	299.818	1.873	3	0.599
Serum cholesterol	324.736	26.791	3	0.000

On the fourth and final likelihood ratio test, the three significant attributes from the second test and the three significant attributes from the third test were tested against each other, and it resulted in five significant attributes, which are Diastolic blood pressure, Chest pain, Electrocardiographic result, Echo finding for hypokinesia, and previous attack of angina, as shown in Table 6.

Table (6): Significance Experiment results for the second and third Experiment variables

	Likelihood Ratio Tests			
Effect	**(-2) Log Likelihood about Reduced Model**	**Chi-Square**	**df**	**Sig**
Diastolic blood pressure	76.051[b]	13.391	3	0.004
Chest pain	97.502[b]	34.842	9	0.000
Electrocardiographic result	129.266	66.607	6	0.000
Echo finding for hypokinesia	107.798[b]	45.139	3	0.000
Previous attack of angina	72.084[b]	9.424	3	0.024
Serum cholesterol	63.638[b]	0.978	3	0.807

From the four likelihood tests combined, the researchers were able to determine that each and every attribute they chose is valid and has an effect on the model, based on the results of Pseudo R-Square; the results of the Nagelkerke coefficient value were 0.958 for the first test with all 14 attributes. In other words, the 14 attributes explain 95.8% of the diversity of the dependent variables, and only 4.2% in the model is explained by variables outside the 14 chosen attributes. In the second test, when the researchers took only the four significant attributes, the Nagelkerke coefficient value dropped to 0.930 or 93%, the third test, which was on the 10 non-significant attributes, the Nagelkerke coefficient was 0.629 or 62.9%. The final test that combined the significant attributes from the (2) and (3) tests, the Nagelkerke coefficient was 0.943 or 94.3%. The results show a pattern: whenever we remove some attributes, the Nagelkerke coefficient drops and the data start to explain less of the diversity. In the fourth test, when the researchers selected only the most significant attributes, the Nagelkerke coefficient value increased, but not by much compared with including all 14 attributes.

After confirming the validity of the attributes using the Nagelkerke coefficient, the researchers conducted another analysis to assess the model's accuracy in predicting heart disease for each of the four previous experiments. Table 7 shows the results.

Table (7): Model accuracy for each experiment

Experiment	Accuracy (%)
Experiment 1 (AllVariables)	94.5%
Experiment 2 (4 IndependentVariables)	85.6%
Experiment 3 (Ten IndependentVariables)	73.6%
Experiment 4 (The combination Experiment)	88.6%

From the results obtained, Experiment 1 which included all of the 14 attributes had the highest accuracy in predicting heart disease which was 94.5%, the second Experiment which only took the 4 significant attributes had an accuracy of 85.6% a very sharp drop in accuracy due to eliminating 10 of the attributes, and the same pattern happens in the other tests, each time we reduce the number of attributes the accuracy drops until it became 73.6% on the third Experiment. This test confirms the results of Pseudo R-Square .

A final analysis was made by the researchers where they eliminated attributes one by one to see the effect of each one of them on the accuracy of the model as a whole, and the results showed that for attributes like (age, sex, previous family history of heart disease, previous attack of angina, fasting blood sugar, and systolic blood pressure) the model accuracy drops by 0.5% if one of them is eliminated. On the other hand, for attributes like (heart rate achieved, serum cholesterol, the angina is exercise-induced, and diastolic blood pressure), eliminating one of them will decrease the model accuracy by 1%. Three attributes appear to have the highest impact on the model accuracy which are Echo finding for hypokinesia which eliminating it decreases the model accuracy by 4% and Chest pain which decreases the model accuracy by 4.5% if eliminated, and the main contributor the model accuracy appears to be an Electrocardiographic result which has an effect of 5.9% on the model accuracy, as shown in Table (8) below.

Table (8): Model accuracy results for each attribute if eliminated

Eliminated Attributes	Accuracy	Eliminated Attributes	Accuracy
Age	94.0%	**systolic blood pressure**	94.0%
Sex	94.0%	**diastolic blood pressure**	93.5%
smoking	94.5%	**heart rate achieved**	93.5%
Previous family history of heart	94.0%	**Chest pain**	90.0%

disease			
Previous attack of angina	94.0%	electrocardiographic result	88.6%
fasting blood sugar	94.0%	The angina is exercise-induced	93.5%
Serum cholesterol	93.5%	Echo finding for hypokinesia	90.5%

After ensuring the attributes' validity across multiple tests and selecting all 14 attributes without elimination, the multinomial logistic regression model was tested on the full dataset to predict heart disease among 201 patients from Iraqi hospitals, as shown in Table 9.

Table (9): Overall model accuracy

	Predicted				
Observed	**Normal**	**Coronary heart disease**	**Congestive heart disease**	**Arrythmias**	**Percent Correct**
Normal	103	0	0	0	100.0%
Coronary heart disease	0	55	2	1	94.8%
Congestive heart disease	0	7	15	0	68.2%
Arrythmias	0	1	0	17	94.4%
Overall Percentage	51.2%	31.3%	8.5%	9.0%	94.5%

For patients who do not have any heart diseases, the (MLR) model was able to diagnosis 100% of the sample as normal without any errors or false positives. For patients with Coronary heart disease (CHD), the model had an accuracy of 94.8% in diagnosis. Meanwhile, the model suffered in diagnosing patients with Congestive heart disease with an accuracy of 68.2% as the model made 7 false positive predictions, flagging patients with Congestive heart disease as having Coronary heart disease. Patients with Arrhythmias were easy to diagnose in the system, as they were the largest amount of training data in the dataset, with only 18 patients, but the model accuracy was 94.4%. The overall model percentage was 94.5%.

Conclusions

Multinomial logistic regression can diagnose heart disease with high accuracy, and the 14 attributes selected by the researchers explained 95.8% of the diversity of the dependent variable. Based on the data collected, coronary heart disease is one of the most common heart diseases in Iraqi hospitals, and echo findings for hypokinesia, chest pain, and electrocardiographic results are the top three attributes, accuracy-wise, in diagnosing heart disease using the MLR model. Consequently, the model can be used in Iraqi hospitals to help physicians make informed decisions in diagnosing patients with heart disease.

Recommendations

The researchers recommend testing the model with a larger dataset and more diverse attributes, as well as designing an easy-to-use user interface so physicians can interact with the model easily. In addition, implementing a medical reporting interface is recommended to allow more training data to be fed into the model over time.

References

[1] World Health Organization. Fact sheets [Internet]. 2021 [cited 2025]. Available from: https://www.who.int/news-room/fact-sheets.

[2] United Nations Economic and Social Commission for Asia and the Pacific (UNESCAP). Statistics [Internet]. 2021 [cited 2025]. Available from: https://www.unescap.org/our-work/statistics.

[3] Ministry of Health-Iraq. Annual statistical report [Internet]. 2019 [cited 2025]. Available from: https://moh.gov.iq/.

[4] Goebel M, Gruenwald L. A survey of data mining and knowledge discovery software tools. ACM SIGKDD Explor Newsl. 1999;1(1):20-33.

[5] Shimono E, Inoue K, Kurita T, Ichiraku Y. Logistic regression analysis for the material design of chiral crystals. Chem Lett. 2018;47(5):611-612.

[6] Ali SS, Mubeen M, Lal I, Hussain A. Prediction of stock performance by using logistic regression model: evidence from Pakistan Stock Exchange (PSX). Asian J Empir Res. 2018;8(7):247-258.

[7] Babiker S, Eltayeb Y, Sayed-Ahmed N, Abdelhafeez S, AlDien MS, Nasir O. Logit model in prospective coronary heart disease (CHD) risk factors prediction in the Saudi population. Saudi J Biol Sci. 2021;28(12):7027-7036.

[8] Sharma A. Weather forecasting application using linear and logistic regression algorithms. In: Proc IOE Graduate Conf. 2019;7.

[9] Rajendra P, Latifi S. Prediction of diabetes using logistic regression and ensemble techniques. Comput Methods Programs Biomed Update. 2021;1:100032.

[10] Ottenbacher KJ, Smith PM, Illig SB, Linn RT, Fiedler RC, Granger CV. Comparison of logistic regression and neural networks to predict rehospitalization in patients with stroke. J Clin Epidemiol. 2001;54(11):1159-1165.

[11] Zhou X, Liu KY, Wong ST. Cancer classification and prediction using logistic regression with Bayesian gene selection. J Biomed Inform. 2004;37(4):249-259.

[12] Munandar TB Ai, Sumiati S, Rosalina V. A predictive model for heart disease diagnosis based on multinomial logistic regression. Inform Technol Control. 2021;50(2):308–318. doi:10.5755/j01.itc.50.2.27672.

[13] Deppen SA, Blume JD, Aldrich MC, Fletcher SA, Massion PP, Walker RC, Chen HC, Speroff T, Degesys CA, Pinkerman R. Predicting lung cancer prior to surgical resection in patients with lung nodules. J Thorac Oncol. 2014;9(10):1477-1484.

[14] El-Habil AM. An Application on Multinomial Logistic Regression Model. Pakistan Journal of Statistics and Operations Research. 2012;8(2):271–291. doi:10.18187/pjsor.v8i2.234.

[15] Cowling TE, Cromwell DA, Bellot A, Sharples LD, van der Meulen J. Logistic regression and machine learning predicted patient mortality from large sets of diagnosis codes comparably. J Clin Epidemiol. 2021;133:43–52. doi:10.1016/j.jclinepi.2020.12.018.

[16] Hosmer DW Jr., Lemeshow S, Sturdivant RX. Applied Logistic Regression. 3rd ed. Hoboken (NJ): Wiley; 2013.

[17] Han J, Kamber M, Pei J. Data mining: concepts and techniques. 3rd ed. Waltham (MA): Morgan Kaufmann; 2012.