THE JOURNAL OF
COGNITIVE SYSTEMS
Journal Homepage: https://dergipark.org.tr/en/pub/jcs

Vol.8, No.1, 2023

Evaluation of Performance Metrics in Heart Disease by Machine Learning Techniques

Abdulvahap Pınar[1] Cemil Çolak[2] Esra Gültürk[3]

[1]Department of Biostatistics, and Medical Informatics, Faculty of Medicine, Inonu University, Malatya 44100, Turkey (e-mail: abdulvahappinar7@gmail.com).
[2]Department of Biostatistics, and Medical Informatics, Faculty of Medicine, Inonu University, Malatya 44100, Turkey (e-mail: cemil.colak@inonu.edu.tr)
[3]Department of Biostatistics, Faculty of Medicine, Cumhuriyet University,Sivas, 58030,Turkey (e-mail: esra0709.eg@gmail.com)

ARTICLE INFO

Received: March.,19.2023
Revised: April,29.2023
Accepted: June.,01.2023

Keywords:
Hearth Disease
Machine Learning
Evaluation
ANN

Corresponding author: Abdulvahap pınar
✉ abdulvahappinar7@gmail.com
☎ +90 5380395132

ISSN: 2548-0650

DOI: https://doi.org/10.52876/jcs.1276688

ABSTRACT

Aim: In addition to affecting the individual sociologically and psychologically, heart disease also poses important problems in health systems. Evaluation of heart disease performances has gained great importance in terms of machine learning method. In the study, performances were compared with the machine learning method for risk methods that classify heart illness.
Materials and Methods: The categorization process Throughout the research made use of the "Heart Disease Dataset," an open access dataset. F1-score, sensitivity, selectivity, accuracy, balanced accuracy, negative and positive predictive values were used to assess the performance of the categorisation model using the machine learning approach. Random forest method, one of the variable selection methods, was used.
Results: According to the relational classification model's classification findings for heart disease, the accuracy, balanced accuracy, sensitivity, selectivity, positive predictive value, negative predictive value, and F1-score values were observed to be 0.997, 0.997, 0.995, 1, 1, and 0.995, respectively.
Conclusion: The relational classification model proposed in the analysis obtained in the web-based open access dataset yielded distinctively successful results in classifying heart disease according to performance criteria.
Keywords: heart disease, classification, relational classification.

1. INTRODUCTION

One of the fields of artificial intelligence known as machine learning allows users to get data on previously collected information without actively programming the system [1].

Heart disease affects both men and women equally across the world. As a result, people must take into account the risk factors for heart disease and adjust their lives accordingly. Heart disease risk factors include several elements including smoking, gender, age, family history, radiation therapy, high blood pressure, obesity, diabetes, stress, elevated blood sugar levels, and unsanitary environments [2].

Artificial Neural Networks (ANN), Deep Learning (DL), Decision Trees (DT), Classification and Regression models (CART, Logistic Regression-LR, K-Nearest Neighbors-KNN), Random Forest (methods like RF), Fuzzy Logic (FL), Genetic Algorithms (GA), and Support Vector Machines are some examples of machine learning techniques (SVM), and Expert Systems (US) are all just a few of the technics used in machine learning algorithms.

These techniques are employed throughout a wide range of the healthcare industry, according to the literature. The WBCD dataset was used to detect breast cancer in patients using methods including SVM, kNN, and NaviBayes. The study's results revealed also that NaviBayes method had a rate of success of 0.702, 0.737 for kNN, and 0.776 for SVM [3].

The use of the variable selection approach in machine learning techniques has been attempted to identify liver failure. Disease identification was carried out with a success rate of Light Gradient Enhancement Machine Classifier-LGBM 82.12%, Multilayer Sensor-MLP 81.13%, DT 81.13%, SVM 77.87%, and Logistic Regression-LR 77.80% in their application on the Indian Liver Patient Dataset (ILPD) [4].

Data mining is best understood as a science that unifies areas that are connected to many other disciplines, such as statistically establishing patterns and using database technology to display data (5).

In terms of resemblance to artificial neural networks, data mining can be described as a method that provides knowledge about the connections between the hidden layers of the variables and how to manage the pattern discovered by these hidden layers in the next phase. focuses on the procedures needed to create an accurate and practical model [6].

2. MATERIAL AND METHODS

2.1. Data set

The whole dataset for cardiac illness used in this study was obtained from the IEEEDataPort database at https://ieee-dataport.org/open-access/. A total of 1190 samples, comprising 629 (52.9%) patients with cardiac disease and 5 61 (47.1%) patients with healthy conditions, were processed in this open access data collection. Table 1 lists the variables from the relevant data set along with their descriptive characteristics.

TABLE I
THE VARIABLES IN THE DATA COLLECTION, TOGETHER WITH A DESCRIPTION OF EACH

Variable	Variable Description	Variable Type	Variable Role
Age	Patient's Age	Numerical	Independent/Predictive
Sex	Patient's Gender	Qualitative	Independent/Predictive
chest pain type	Chest Pain Type	Qualitative	Independent/Predictive
resting bp s	Blood pressure	Numerical	Independent/Predictive
cholesterol	Cholesterol	Numerical	Independent/Predictive
fasting blood sugar	Fasting Blood Sugar	Qualitative	Independent/Predictive
resting ecg	ECG results	Qualitative	Independent/Predictive
max heart rate	Maximum Heart Rate Reached	Numerical	Independent/Predictive
exercise angina	Exercise Induced Angina	Qualitative	Independent/Predictive
oldpeak	oldpeak	Numerical	Independent/Predictive
STslope	STslope	Qualitative	Independent/Predictive
Target	Target	Qualitative	Dependent/Target

2.2. Support Vector Machines

Support vector machines are a machine learning approach that are extensively used in the literature for both classification and regression analysis procedures, even though they are employed for data set classification. Support vector machines are built on a learning model utilised in kernel functions in addition to the supervised learning model, taking into account the type of data being used to test the method. In light of this, it is a strategy that is employed in both linear and nonlinear classification procedures.

Although every data is classed with a hyperplane in classification processes if it is fully decomposable, if the data is in an undifferentiated structure, it is not classified with a single plane. several kernel functions are therefore utilised [7].

2.3. Naive Bayes

The Naive Bayes approach assesses the likelihood of identical situations having the same effects on the outcome.

The Naive Bayes algorithm determines the frequency of occurrence of each output in the training set. A priori probability refers to the result that is taken into account. The total of the calculated probability is one. It displays the result's class in accordance with the highest value attained within the a priori probability [8].

The simplest way to describe naive Bayes is as the sum of all conditional probabilities. according to;

Likelihood Class Prior Probability

$$P(c\mid x)=\frac{P(x\mid c)P(c)}{P(x)}$$

Posterior Probability Predictor Prior Probability

$$P(c\mid X)=P(x_1\mid c)\times P(x_2\mid c)\times\cdots\times P(x_n\mid c)\times P(c)$$

2.4. Logistic Regression

It is a technique used to look at the groups to which the data set's observations are categorised. In logistic regression analysis, a classification model is developed by taking into account the information that each class is aware of. New observations or observations that need to be added to the data set are categorised with the aid of the logistic model that was produced. The employment of logistic models is a common strategy in healthcare applications. Risk is variable or processes that show whether the case is visible or not. Input variables. Early diagnosis and disease-causing variables are crucial for variable detection [9].

It is a technique used when the dependant variable is qualitative and contains two or more categories.

Dependent Variable	*Group*
Two-Category (Binominal)	died-living, successful-failed
More than two categories, unorganised (Multinominal)	unemployed-retired-employed, numerical-verbal-equal weight
More than two categories in the sequence (Ordinal)	low-medium-high, ineffective-effective-very effective

The logistic distribution function may be calculated by formulating the distribution function using two-category output variables.

$$P_i = E\left(Y = {}^{1}/_{X_i}\right) = \frac{1}{1+e^{-(\beta_0+\beta_1 X_i)}} \quad (1)$$

$Z_i = \beta_0 + \beta_1 X_i$ in case of conversion;

$$P_i = \frac{1}{1+e^{-Z_i}} \quad (2)$$

The Z_i value is a function known as the logistic distribution function. $-\infty < Z_i < \infty$ takes values in the range and P_i *if its value* $0 < P_i < 1$ takes value between. Z_i equation, so X_i *with* P_i There doesn't seem to be a direct connection between This gap in the linear probability model will be filled if these two relations are offered [10].

2.5. Performance metrics

The performance metrics utilised to compare classification performances were, in order, sensitivity, selectivity, accuracy, balanced accuracy, negative predictive value, positive predictive value, and F_1-score. Table 2 contains the matrix for categorising performance criteria.

TABLE II
PERFORMANCE CRITERION CLASSIFICATION MATRIX

		Actual Value Positive	Actual Value Negative	Total
Estimated Value	**Positive**	True Positive (TP)	False Negative (FN)	**TP+FN**
	Negative	False Positive (FP)	True Negative (TN)	**FP+TN**
	Total	**TP+FP**	**FN+TN**	**TP+TN+FP +FN**

Sensitivity = TP/[TP+FP]
Selectivity = TN/[TN+FN]
Accuracy = [TP+TN]/[TP+TN+FP+FN]
Balanced accuracy = [[TP/(TP+FP))]+[TN/(TN+FN)]]/2
Negative predictive value =TN/[TN+FP]
Pozitive predictive value = TP/[TP+FN]
F_1 – score = [2*TP]/[2*TP+FP+FN]

2.6. Data analysis

When the Shapiro-Wilk test was used to determine if quantitative data were normal, it was found that the data were not symmetrical. The effect size, which was determined by utilizing the median and interquartile range, provided evidence of this. The Mann-Whitney U test and the t-test for independent samples were used to compare the categories of the dependent/target variable, "Normal," and "Heart disease," in order to determine whether there was a difference that was statistically significant (Heart disease). At $p < 0.05$, obtained probability values were deemed statistically significant. The Biostatistics and Medical Informatics Department of inönü University built the open-access web-based http://biostatapps.inonu.edu.tr/ address from which all of the analyses were retrieved.

3. FINDINGS

In Table 3, descriptive statistics are shown for the input variables covered in this study. Age, Blood pressure, Cholesterol, Heart Rate Reached, and old peak variables were statistically different across the groups of the dependent variable ($p < 0.005$).

TABLE III
DESCRIPTIVE STATISTICS FOR QUANTITATIVE VARIABLES

Variable	Normal Median(IQR)	Heart Disease Median(IQR)	p-value	ES
Age	51(13)	57(11)	<0.001	0.077
Blood pressure	130(20)	132(25)	<0.001	0.278
Cholesterol	232(67)	226(148)	0.003	0.364
Heart Rate Reached	154(32)	128(34)	<0.001	0.434
oldpeak	0(0.8)	1.2(1.9)	<0.001	0.375

IQR: Interquartile Range; ES: effect size. ***: Mann Whitney U test, **: Independent samples t-test**

A heart illness dataset was used in the study, and the accuracy values of the classification performances of Support Vector Machines, Naive Bayes, and Logistic Regression models were examined. The model with the best performance was chosen. One of the methods used in machine learning, Support Vector Machines was used to classify the dataset related to heart disease. The classification matrix is seen in Table 4.

TABLE IV
CLASSIFICATION MATRIX FOR THE MODEL

Prediction	Reference Normal	Reference Heart disease
Normal	558	0
Heart disease	3	629

Table 5 provides statistics on the relational classification model's performance in terms of classification. The model's accuracy was 0.997, as were its balanced accuracy, sensitivity, specificity, positive predictive value, negative predictive value, and F1-score. Its negative predictive value was 0.995, and both its positive and negative predictive values were 1.

TABLE V
STATISTICS ON CLASSIFICATION PERFORMANCES OF THE MODEL

Metric	Value
Accuracy	0.997
Balanced Accuracy	0.997
Sensitivity	0.995
Selectivity	1
Positive predictive value	1
Negative predictive value	0.995
F1-score	0.997

The classification matrix utilized in the classification utilizing the dataset for heart disease and the Naive Bayes algorithm, one of the machine learning techniques, is shown in Table 6.

TABLE VI
CLASSIFICATION MATRIX FOR THE MODEL

Prediction	*Reference* *Normal*	*Reference* *Heart disease*
Normal	460	81
Heart disease	101	548

Table 7 provides statistics on the relational classification model's performance in terms of classification. The model's accuracy was 0.847, the balanced accuracy was 0.846, the sensitivity was 0.82, the selectivity was 0.871, the positive predictive value was 0.85, the negative predictive value was 0.844, and the F1-score was 0.835.

TABLE VII
STATISTICS ON THE MODEL'S PERFORMANCE IN CLASSIFYING OBJECTS

Metric	Value
Accuracy	0.847
Balanced Accuracy	0.846
Sensitivity	0.82
Selectivity	0.871
Positive predictive value	0.85
Negative predictive value	0.844
F1-score	0.835

The classification matrix used for the classification utilizing the heart disease dataset and the Logistic Regression model using machine learning techniques is shown in Table 8.

TABLE VIII
CLASSIFICATION MATRIX FOR THE MODEL

Prediction	Reference	
	Normal	Heart disease
Normal	462	82
Heart disease	99	547

Table 9 provides statistics on the model's classification results in relational classification. The model's accuracy was 0.848, along with balanced accuracy of 0.847, the sensitivity of 0.824, specificity of 0.87, positive predictive value of 0.849, negative predictive value of 0.847, and F1-score of 0.836.

TABLE IX
STATISTICS ON CLASSIFICATION PERFORMANCES OF THE MODEL

Metric	Value
Accuracy	0.848
Balanced Accuracy	0.847
Sensitivity	0.824
Selectivity	0.87
Positive predictive value	0.849
Negative predictive value	0.847
F1-score	0.836

4. DISCUSSION AND CONCLUSION

Heart disease, vascular blockage, and the inability to pump blood are some of the symptoms of this particular disease kind. In terms of treatment simplicity, early detection and prevention of heart disease and its many forms become more important. There are numerous drawbacks to technology, which is a part of daily life, as well as novel benefits that influence it. A person's health can be negatively impacted by variables like stabilization, stress, and a bad diet [11].

The extensive use of data mining techniques has been made possible by the rise in the risk of mortality in people with heart disease. To stop people from dying in hospitals after receiving an early diagnosis, serious investigations have been recommended [12].

Data mining is a method that enables information to be extracted from high-dimensional data using a variety of statistical approaches and in circumstances when there is a very low likelihood of guessing the correlations between variables [13]. It is a strategy for analysis that uses data mining to categorize and summarise variables while avoiding information ambiguity. Data preparation is used throughout the data analysis stage to find outliers and extremes as well as to establish the link between the variables. The choice to be made and the definitions that must be made can be guided by it [14].

The prediction of cardiac problems has been the subject of several research in the literature. The maximum success performance was reached with J48 as accuracy (83,732%) in research that compared the performance of approaches used for the identification of heart illness such as J48, K, Nearest Neighbor (KNN) algorithm, Decision tree, and Naive Bayes (NB) [15].

The accuracy performance ratings achieved using machine learning approaches were compared in a different investigation utilizing the same data set. The accuracy rate of Support vector machines (SVM) for classifying heart disease performance, per the study's findings, was 0.897. The SVM accuracy rate for this investigation was determined to be 0.997 [16].

This study made use of the "Heart Disease Dataset," an open-access online resource. The linked data set was subjected to the use of three data mining techniques: Support vector machines, Naive Bayes, and a Logistic Regression Classification Model. The diagnosis of heart illness (the dependent variable) and other relevant parameters (the independent variables) were used to create statistics based on the classification performance of the models of all three data mining techniques.

Sensitivity, selectivity, positive and negative predictive values, accuracy, balanced accuracy, and F1- The score was estimated as 0.997 after taking into account the statistics on classification performances by the analysis results obtained from the open access data set. Estimates with a high degree of accuracy were produced using the Naive Bayes and Logistic Regression approaches.

REFERENCES

[1] Hurwitz, J., et al., Augmented intelligence: the business power of human–machine collaboration. 2019: CRC Press.
[2] Widmer, R.J., et al. Digital health interventions for the prevention of cardiovascular disease: a systematic review and meta-analysis. in Mayo Clinic Proceedings. 2015. Elsevier.
[3] Fan, C.-Y., et al., A hybrid model combining case-based reasoning and fuzzy decision tree for medical data classification. Applied Soft Computing, 2011. 11(1): p. 632-644.
[4] Saygin, E. and M. Baykara, Karaciğer Yetmezliği Teşhisinde Özellik Seçimi Kullanarak Makine Öğrenmesi Yöntemlerinin Başarılarının Ölçülmesi. Fırat Üniversitesi Mühendislik Bilimleri Dergisi, 2021. 33(2): p. 367-377.
[5] Özekes, S., Veri madenciliği modelleri ve uygulama alanları. 2003.
[6] Pujari, A.K., Data mining techniques. 2001: Universities press.
[7] Chakraborty, K., et al., Sentiment Analysis of COVID-19 tweets by Deep Learning Classifiers—A study to show how popularity is affecting accuracy in social media. Applied Soft Computing, 2020. 97: p. 106754.
[8] Akbulut, S., Veri madenciliği teknikleri ile bir kozmetik markanın ayrılan müşteri analizi ve müşteri segmentasyonu. Gazi Üniversitesi Fen Bilimleri Enstitüsü, Ankara, 2006.
[9] Hosmer Jr, D.W., S. Lemeshow, and R.X. Sturdivant, Applied logistic regression. Vol. 398. 2013: John Wiley & Sons.
[10] GUJARATİ, D.N., Temel Ekonometri (Çev. Ümit Senesen). Literatür Yayıncılık, İstanbul, 2005.
[11] Macphee, C.H. and A. Daugherty, Cardiovascular diseases. Drug Discovery Today: Therapeutic Strategies, 2008. 1(5): p. 1-3.
[12] Patel, J., D. TejalUpadhyay, and S. Patel, Heart disease prediction using machine learning and data mining technique. Heart Disease, 2015. 7(1): p. 129-137.
[13] Feyyad, U., Data mining and knowledge discovery: Making sense out of data. IEEE expert, 1996. 11(5): p. 20-25.
[14] Kumar, A.S. and R. Wahidabanu, Data Mining Association Rules for Making Knowledgeable Decisions, in Data Mining Applications for Empowering Knowledge Societies. 2009, IGI Global. p. 43-53.
[15] Bahrami, B. and M.H. Shirvani, Prediction and diagnosis of heart disease by data mining techniques. Journal of Multidisciplinary Engineering Science and Technology (JMEST), 2015. 2(2): p. 164-168.
[16] YILMAZ, R. and F.H. YAĞIN, Early detection of coronary heart disease based on machine learning methods. Medical Records, 2022. 4(1): p. 1-6.

BIOGRAPHIES

Abdulvahap PINAR obtained her BSc. degree in Statistics from Fırat Üniversitesi in 2015. He received MSc. degree in Statistics from the Fırat Üniversitesi in 2015-2019. biostatistics and medical informatics from the Inonu University in 2021. He currently continues Ph.D.

Cemil ÇOLAK obtained her BSc. degree in Ondokuz May University in Turkey, where she studied science, literature, and statistics. Turkish University of Inönü, Institute of Health Sciences, Biostatistics, 1994–1999 MSc Ph.D. in biostatistics, Ankara University, Institute of Health Sciences, Turkey, 1999-2001, 2001-2005 Medical science, basic medical sciences, biostatistics and medical informatics, epidemiology, information security and reliability, computer vision, bioinformatics, artificial intelligence, computer learning and pattern recognition, pattern recognition and image processing, and other related fields.

ESRA GÜLTÜRK obtained her BSc. degree in Ondokuz May University in Turkey, where she studied science, literature, and statistics. Turkish University of republic, Institute of Health Sciences, Biostatistics, 2005 MSc Ph.D. in biostatistics, republic University, Institute of Health Sciences, Turkey, 2009, I graduated from Cumhuriyet University, Faculty of Economics and Administrative Sciences, Department of Numerical Methods in 2016. Since 2018, I have been working as a PhD lecturer in the field of biostatistics at Cumhuriyet University.