AITA
ARTIFICIAL INTELLIGENCE
THEORY AND APPLICATIONS

■ RESEARCH ARTICLE

# Heart Disease Diagnosis via Web Based Classification Software programmed with Julia Programming Language

**Hüseyin Kutlu** [a†], **Emek Güldoğan**[b], **Cemil Çolak**[b]

[a]Department of Computer Technologies, Adıyaman University, Adıyaman, Türkiye
[b]Department of Biostatistics and Medical Informatics, Inonu University, Malatya, Türkiye
[†] hkutlu@adiyaman.edu.tr, corresponding author

**RECEIVED** JULY 7, 2023
**ACCEPTED** NOVEMBER 1, 2023

**CITATION** Kutlu, H., Güldoğan, E., Çolak, C. (2024). Heart disease diagnosis via web based classification software programmed with Julia programming language. *Artificial Intelligence Theory and Applications, 4*(1), 1-10.

**Abstract**

In recent years, various tools and algorithms have been proposed and continue to be proposed by researchers to develop highly successful medical decision support systems. However, the clinical use of these algorithms is very limited due to various limitations. Making the necessary software installations to run the algorithm and lack of programming knowledge is some of these restrictions. In this study, a web-based classification software developed with the Julia programming language, which can be used by physicians in their medical research and clinical decisions, is introduced. Through this software, coronary artery disease detection was performed with the Cleveland heart disease database, which is a publicly accessible data set. The dataset was classified with eight different classifiers (KNN, SVM, DT, RF, AdaBoost, Gauss Naive Bayes, LDA, LR) supported by the software. The metrics obtained by 10-fold cross-validation of the data set are reported. The SVM classifier achieved the highest classification accuracy with 86.44%. The software proposed in this study may assist clinicians in research and patient identification.

**Keywords:** Julia; dash; classification; heart disease; web-based artificial intelligence software

## 1. Introduction

Computerized clinical decision support systems are an important part of healthcare today. A clinical decision support system takes patient information as input. With these inputs, it aims to improve medical decisions and improve health care delivery [1]. Clinical decision support systems support clinicians in their scientific studies and complex decision-making processes [2]. Advances in computer technologies have led to rapid development of clinical decision-making systems since the first use of clinical decision-making systems. Pharmaceutical databases, electronic records of patients, and international open databases have contributed to the development of these systems. Machine learning-based artificial intelligence algorithms are presented as a decision support system to help clinicians predict patient outcomes [3].

It can be said that there are various difficulties in the use of decision support systems that have emerged in recent years by clinicians. Decision support systems are usually



Artificial Intelligence Theory and Applications, ISSN: 2757-9778. ISBN: 978-605-69730-2-4 © 2024 İzmir Bakırçay University

connected to a computer system. This situation causes the system to be inaccessible from anywhere at all times. Every algorithm developed cannot always turn into a decision support system. In this case, programming knowledge is needed to code the algorithm. In order to code these programs, various software installations and data to train and test this software are needed. It is also inevitable to have a computer equipped with the equipment where these algorithms can be trained and run. In this study, web-based classification software implemented with the Julia programming language, which is a decision support system that can be run from any platform and which does not need to set up a program and have programming knowledge, is introduced. The software was developed by Inonu University, Department of Biostatistics and Medical Informatics.

In order to benefit from the speed of the Julia programming language, the web-based classification software was implemented in the Julia programming language [4]. Web-based classification software and the Cleveland heart disease database [5], a publicly accessible dataset, were used. With this database, the system has been trained and tested with the aim of detecting coronary artery disease.

Heart disease is often referred to as coronary artery disease. Coronary artery disease is a broad term that can refer to any condition that affects the heart. Coronary artery diseases can be defined as all kinds of disorders such as infections affecting the heart, genetic disorders, vascular disease, heart valve disease. Coronary artery disease is the most common form of cardiovascular disease and is the leading cause of heart attacks [6]. There are many factors that cause coronary artery diseases. While some of these factors cannot be changed (age, gender, family history), some of them can cease to be risk factors by changing their lifestyle (smoking, alcohol, physical activity, etc.). Coronary artery disease can be detected by symptoms of chest pain and fatigue while in some people, it shows no symptoms [7]. Therefore, it is necessary to monitor the symptoms that cause heart disease, and the risk status should be followed up by the physician and the patient. The primary method used to diagnose heart disease is angiography. However, it is very costly and requires technical experience [8]. Apart from this, various techniques such as blood pressure monitoring, echocardiogram, electrocardiography, electrophysiological examinations, myocardial perfusion scans and tilt table test are performed to diagnose heart disease [9].

Machine learning algorithms play a dominant role in diagnosing heart disease. Machine learning algorithms have the advantage of extracting necessary information from large amounts of data. Many studies have been carried out in the field of machine learning with the Cleveland heart disease database. Javeed et al. [10] developed an efficient and less complex model to improve coronary heart disease risk estimation using a random search algorithm and a Random Forest (RF) model. Using the 7-element subset of the features, they achieved an accuracy of 93.33%. The model showed a 3.3% improvement over standard RF. Pasha et al. [11] proposed a new feature reduction (NFR) model for effective heart disease risk estimation in Cleveland, Hungary, Statlog, and Switzerland datasets. They replaced the missing values with the average values of the column. Logistic Regression (LR), Random Forest (RF), Boosted Regression Tree (BRT), Stochastic Gradient Boosting (SGB) and Support Vector Machine (SVM) classification algorithms were used in the study. Classification metric values of the algorithms used were calculated. By comparing the metric values, the algorithm with the highest classification performance was determined. Using LR (9 features) on the Cleveland dataset, they reported an accuracy of 92.53% and an AUC of 0.9268. Saqlain et al. [12] proposed a feature subset selection method to improve cardiovascular risk prediction results using feature selection algorithms (MFSFSA), forward feature selection algorithm (FFSA), and reverse feature selection algorithm (RFSA) based on average fisherman

score. They classified the feature subsets with the RBF kernel-based SVM classifier. They tested the proposed model on Cleveland and different data sets. They achieved an accuracy of 81.19% on the Cleveland dataset with seven features. Muhammed et al. [13] proposed an intelligent prediction model for early detection of heart disease by training various machine learning classifiers on the best features of the Cleveland dataset using 10-fold cross validation. The researchers applied four feature selection algorithms: fast correlation-based filter (FCBF), minimum redundancy maximum relevance (mRMR), LASSO, and Relief to obtain key and more relevant features in the study. Researchers have achieved 94.41% accuracy with the Extra Tree classifier in the study. Ali et al. [14] used a 70:30 ratio validation for the training and test datasets to create an autonomous diagnostic system for heart disease identification using an enhanced deep neural network (DNN) and chi-square feature selection for classification in the Cleveland dataset. In the test dataset, they reported the accuracy of the proposed hybrid model as 93.33% and the AUC value as 0.94. Gupta et al. [15] obtained 92.30% classification accuracy using standardized data Logistic Regression in their study by dividing the Cleveland heart dataset for training and testing by a ratio of 70:30. They also obtained the best classification accuracy by testing the KNN classifier with a k value between 2 and 20 and 90.11% at k = 14.

In this study, web-based classification software coded with Julia programming language was used. The software supports eight different classifiers such as K - Nearest neighbors, SVM, Decision tree classifier, Random Forest Classifier, AdaBoost classifier, Gaussian Naive Bayes Classifier, LDA classifier, and Logistic Regression Classifier. The dataset was trained and classified with all the classifiers supported by the software. The software split the data set in a 10-fold cross-validation manner and presented the classification results to the user with tabular metrics.

## 2. Materials and Methods

Web-based classification software developed with the Julia programming language was used in the study. With the application in question, the heart disease classification model was trained and tested using the Cleveland heart disease database. The application contains eight different classification algorithms. In this section, the software, the dataset and the classifiers are introduced.

### 2.1. Dataset

The Cleveland heart disease database is an open access database [16]. The database contains 303 observations, 297 of which are complete observations and six observations with incomplete data. The database has 76 features. In studies conducted with the data set in the literature, missing observations are generally removed from the data set. In this way, the data set includes 137 patient (1) and 160 healthy (0) observations. In studies conducted in the literature with the data set, 13 features were generally used to increase the classification performance. In this study, in order to compare the classification performance of the web software with existing studies, missing observations were removed from the dataset and 13 features were used. The features used are described in Table 1.

Table 1. Details of the Cleveland heart disease database

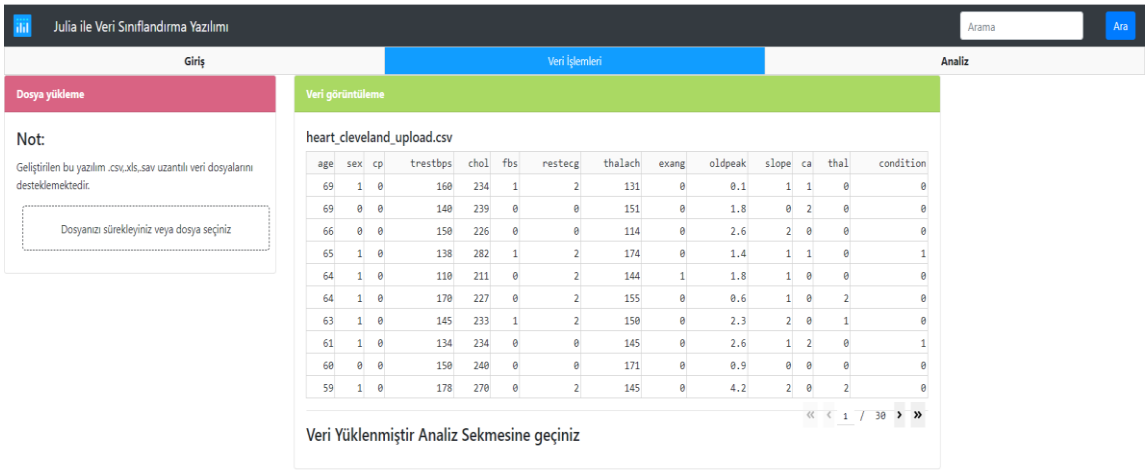
| Features | Explanation | State |
|---|---|---|
| **1. Age** | Numeric | input |
| **2. Sex** | 0: Female,<br>1: Male | input |
| **3. Cp:** chest pain | 0: typical angina,<br>1: atypical angina,<br>2: non-anginal pain<br>3: asymptomatic | input |
| **4. trestbps:** resting blood pressure (blood pressure in mm Hg at the time of admission to hospital) | Numeric | input |
| **5. kol:** Cholesterol value in mg / dl measured with the BMI sensor | Numeric | input |
| **6. fbs:** fasting blood sugar | 0: < 120 mg / dl<br>1: > 120 mg / dl | input |
| **7. restecg:** resting electrocardiographic results | 0: normal<br>1: Having ST-T wave abnormality (T wave inversions and/or ST elevation or depression > 0.05 mV)<br>2: Demonstration of probable or definite left ventricular hypertrophy according to Estes criteria | input |
| **8. thalach:** maximum heart rate | Numeric | input |
| **9. exang:** angina (compression) due to exercise | Numeric | input |
| **10. oldpeak:** Exercise-induced ST depression at rest | Numeric | input |
| **11. egim:** hill exercise ST segment slope | 1: upsloping<br>2: straight<br>3: sloping down | input |
| **12. ca:** number of major vessels colored by fluoroscopy (for calcification of vessels) | 0: No occluded vessel<br>1: 1 vessel<br>2: 2 vessel<br>3: 3 vessel | input |
| **13. thal:** results of nuclear stress test | 1: normal;<br>2: fixed defect;<br>3: reversible defect | input |
| **14. num:** Target variable representing the diagnosis of heart disease (angiographic disease status) in any major vessel | 0: < 50% diameter reduction<br>1: > 50% diameter reduction | output (target) |

## 2.2. Web Based Data Classification Software Programmed with Julia Programming Language

The interface of our data classification software [18], which is under the title of Julia software among Data Science and Artificial Intelligence Based Web Software [17], developed by our department is shown in Figure 1.


Figure 1. Interface of Interactive Web Software

Association Rules Mining, Data Classification Software, Cluster Analysis and Regression Analysis software are available in the Julia Web Software menu. Figure 2 shows the interface of the Classification Software.




| age | sex | cp | trestbps | chol | fbs | restecg | thalach | exang | oldpeak | slope | ca | thal | condition |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 69 | 1 | 0 | 160 | 234 | 1 | 2 | 131 | 0 | 0.1 | 1 | 1 | 0 | 0 |
| 69 | 0 | 0 | 140 | 239 | 0 | 0 | 151 | 0 | 1.8 | 0 | 2 | 0 | 0 |
| 66 | 0 | 0 | 150 | 226 | 0 | 0 | 114 | 0 | 2.6 | 2 | 0 | 0 | 0 |
| 65 | 1 | 0 | 138 | 282 | 1 | 2 | 174 | 0 | 1.4 | 1 | 1 | 0 | 1 |
| 64 | 1 | 0 | 110 | 211 | 0 | 2 | 144 | 1 | 1.8 | 1 | 0 | 0 | 0 |
| 64 | 1 | 0 | 170 | 227 | 0 | 2 | 155 | 0 | 0.6 | 1 | 0 | 2 | 0 |
| 63 | 1 | 0 | 145 | 233 | 1 | 2 | 150 | 0 | 2.3 | 2 | 0 | 1 | 0 |
| 61 | 1 | 0 | 134 | 234 | 0 | 0 | 145 | 0 | 2.6 | 1 | 2 | 0 | 1 |
| 60 | 0 | 0 | 150 | 240 | 0 | 0 | 171 | 0 | 0.9 | 0 | 0 | 0 | 0 |
| 59 | 1 | 0 | 178 | 270 | 0 | 2 | 145 | 0 | 4.2 | 2 | 0 | 2 | 0 |

Figure 2. Interface of web-based Data Classification Software programmed with the Julia programming language

The software consists of three menus. The software is introduced in the login menu. File (csv, xlsx, sav) loading and data display operations are performed in the data operations menu. The Analysis menu view of the software is shown in Figure 3.

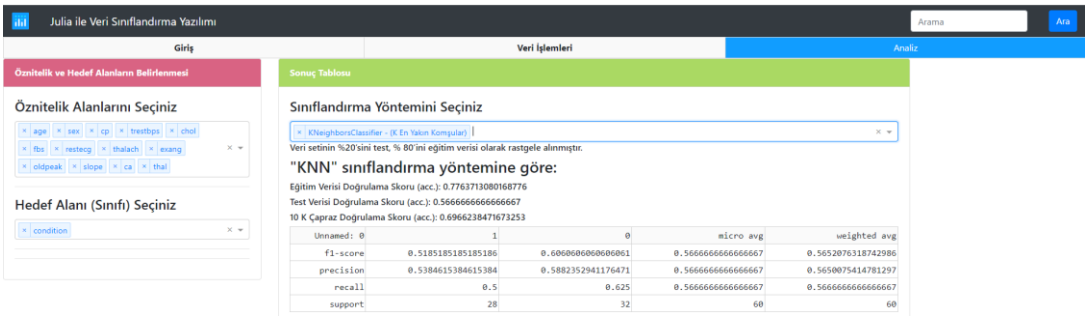



Figure 3. Analysis Menu View of web-based Data Classification Software programmed with Julia programming language

In the analysis menu shown in Figure 3, the feature fields in the data set can be selected. In this way, it can be determined how much which feature affects the model. Besides, the target area can also be selected. From the result table, the classification method (KNN, SVM, DT, RF, AdaBoost, Gauss Naive Bayes, LDA, LR) is selected. Then, classification accuracies of training and test datasets and classification metrics such as F1 Score, Precision, Recall of classes are presented to the user in tabular form.

**K-Nearest Neighbours (KNN) Classifiers:** KNN is based on estimating the class of the sample based on the information in which class the nearest neighbours of the vector formed by the independent variables are dense. The KNN algorithm makes predictions on two basic parameters; The Distance parameter represents the distance of the point to be estimated from other points. There are different distance calculation algorithms such as Euclid, Minkowski. The K (neighbourhood number) parameter is the parameter that tells how many nearest neighbours will be calculated over [19].

**Support Vector Machine (SVM):** SVM classifiers classify with the supervised learning method. It relies on drawing a line or hyperplane to separate points placed on a plane. It aims to have this line at the maximum distance for the points of both classes [20].

**Decision Tree (DT):** DT is one of the tree-based learning algorithms. It is among the supervised learning algorithms. They classify the dataset by dividing it into smaller sets by applying a set of decision rules.

**Random Forest (RF) algorithm:** RF is an algorithm that produces and classifies multiple decision trees by training each one on a different observation sample. The algorithm creates a decision tree for each sample, and the estimated value result of each decision tree is formed. Voting is performed for each value formed as a result of the prediction. Observation is assigned to the class with the most votes [22].

**AdaBoost Classifier:** AdaBoost Classifier is one of the Ensemble Learning methods. Boosting is to create a strong learner by bringing together many weak learners and training them cumulatively. In the Adaboost classification model, the training set is first trained with a weak learner. Learners who make incorrect predictions after training are important for the AdaBoost algorithm. In the next training, the incorrectly learned training data in the first estimation is retrained by giving more priority, that is, by increasing the weights. The results are combined by training the weak learner output to be the input to the other learner. In this way, it performs the classification process [23].

**Gaussian Naive Bayes Classifier:** It makes use of Bayes Theorem during the training phase. According to the conditional independence assumption, each feature is handled

independently. In this way, the number of parameters to be estimated is considerably reduced. Probability values are calculated for the algorithm to work. These are the probabilities of each class in the training dataset and the conditional probabilities of each input value given each class value. In the Gaussian Naive Bayes algorithm, in addition to the probabilities of each class, the mean and standard deviation values of each class are also calculated and classification is made.[24]

**Linear Discriminant Analysis (LDA):** The LDA Classification algorithm is based on developing a probability model per class based on the particular distribution of observations for each input variable. It works by calculating summary statistics for input properties by class label, such as mean and standard deviation. These statistics represent the model learned from the training data. In practice, linear algebra operations are used to efficiently calculate required quantities via matrix decomposition. Estimates are made by estimating the probability of a new instance of each class label based on the values of each input attribute. The class that results most likely is then assigned to the instance. Therefore, LDA can be thought of as a simple application of Bayes' Theorem for classification. LDA assumes that the input variables are numeric and normally distributed and have the same variance (spread). If this is not the case, it may be desirable to transform the data to have a Gaussian distribution and standardize or normalize the data prior to modeling [25].

**Logistic Regression Classifier (LR):** LR is a data analysis technique that uses mathematical calculations to find relationships between two attributes. LR then classifies using the mathematical relationship it establishes to estimate the value of the target variable.

## 3. Results

Eight classification algorithms supported by the developed software were trained and tested with the data set. The models were compared with the performance metrics supported by the software. The results are obtained with the 10-fold Cross validation method as default by the software.

### 3.1. Performance Metrics

Performance metrics supported by the software are explained and listed.

**Accuracy (Acc.):** Acc. is the ratio of all classification predictions to the number of successfully predicted data. FN and FP represent the number of incorrect predictions of classes with each other. TP and TN represent the number of observations for which classes were predicted correctly. With this information, Accuracy is calculated as shown in Equation (1).

$$Acc. = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

**Precision (P):** P is a probability measure that evaluates the probability of a positive prediction being correct, as shown in Equation (2).

$$P = \frac{TP}{TP + FP} \tag{2}$$

**Recall (R):** R is the proportion of positively correctly predicted samples belonging to the positive class and the formula is as stated in Equation (3).

$$R = \frac{TP}{TP + FN} \tag{3}$$

**F1 Score:** Often referred to as the F measure. The F1 score is a measure used to determine the accuracy of a test. Calculates the score taking into account the precision P of the test and the recall R. In order not to make an erroneous model selection in unequally distributed data sets, the F measure is sometimes used instead of Accuracy. The formula for measure F is as stated in Equation (4)

$$f_1 = 2 * \frac{P * R}{P + R} \tag{4}$$

### 3.2. Experimental Results

The presence of heart disease was estimated in this study. Estimation was carried out with different classification algorithms. The classification metrics of the algorithms were calculated. In Table 2, classification performance metrics are shown together with the classifier parameters.

Table 2. Classification performance metrics

| | Metrics | | | |
|---|---|---|---|---|
| **Methods** | **Acc. (%)** | **F1 Score (%)** | **P (%)** | **R (%)** |
| **KNN (K=5)** | 69,66 | 56,22 | 56,33 | 56,25 |
| **SVM (Kernel: Lineer C=0.025)** | 86,44 | 74,84 | 74,91 | 74,77 |
| **DT (MaxDepth:5)** | 73,37 | 64,75 | 64,81 | 64,73 |
| **RF (MaxDepth:5, n_estimators:10, max_features:1)** | 79,79 | 71,59 | 71,57 | 71,65 |
| **AdaBoost** | 82,31 | 69,96 | 70,00 | 70,08 |
| **Gaussian Naive Bayes** | 85,23 | 76,74 | 77,60 | 77,23 |
| **LDA** | 86,06 | 73,33 | 73,66 | 73,66 |
| **LR** | 86,02 | 73,30 | 73,33 | 73,44 |

As can be seen in Table 2, the SVM classifier (86.44%) achieved the highest classification accuracy. Figure 4 shows the graph comparing classifier performances.

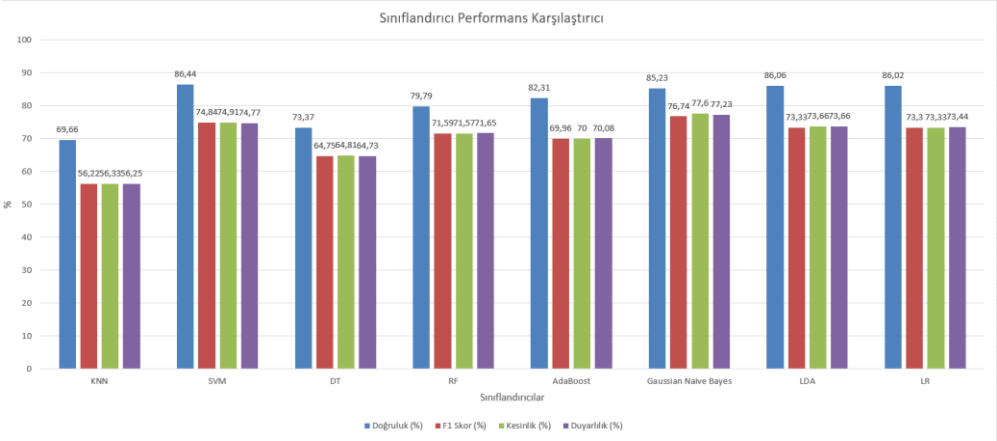



Figure 4 – Classifier performances

## 4. Conclusion

In this study, a software has been proposed for clinicians to use machine learning methods, which have been successfully presented in the literature, in their research and treatment. The software is a web-based software. It is programmed with the julia language. The software enables the use of classification algorithms without the need for program installation and programming knowledge. It aims to enable clinicians to analyze by simply uploading their data. In the study, the SVM classifier came to the fore with a classification accuracy of 86.44% in the detection of heart disease. The proposed software is under development and in the next versions, 10-fold cross-validation will be user-configurable. In the future, missing and excessive values in the data are planned to be detected by the software and removed from the user-approved data set. In addition, it is considered that the user-approved data set normalization phase will be carried out.

### References

[1] R. T. Sutton, D. Pincock, D. C. Baumgart, D. C. Sadowski, R. N. Fedorak, and K. I. Kroeker, “An overview of clinical decision support systems: benefits, risks, and strategies for success,” NPJ Digit Med, vol. 3, no. 1, p. 17, Feb. 2020, doi: 10.1038/s41746-020-0221-y.

[2] I. Sim et al., “Clinical Decision Support Systems for the Practice of Evidence-based Medicine,” Journal of the American Medical Informatics Association, vol. 8, no. 6, pp. 527–534, Nov. 2001, doi: 10.1136/jamia.2001.0080527.

[3] J. Amann et al., “To explain or not to explain?—Artificial intelligence explainability in clinical decision support systems,” PLOS Digital Health, vol. 1, no. 2, p. e0000016, Feb. 2022, doi: 10.1371/journal.pdig.0000016.

[4] K. Gao, G. Mei, F. Piccialli, S. Cuomo, J. Tu, and Z. Huo, “Julia language in machine learning: Algorithms, applications, and open issues,” Comput Sci Rev, vol. 37, p. 100254, Aug. 2020, doi: 10.1016/j.cosrev.2020.100254.

[5] D. W. Aha, “ Heart Disease Data Set,” https://archive.ics.uci.edu/ml/datasets/heart+disease, Jan. 26, 2023.

[6] R. Das, I. Turkoglu, and A. Sengur, “Effective diagnosis of heart disease through neural networks ensembles,” Expert Syst Appl, vol. 36, no. 4, pp. 7675–7680, May 2009, doi: 10.1016/j.eswa.2008.09.013.

[7] R. Das, I. Turkoglu, and A. Sengur, “Effective diagnosis of heart disease through neural networks ensembles,” Expert Syst Appl, vol. 36, no. 4, pp. 7675–7680, May 2009, doi: 10.1016/j.eswa.2008.09.013.

[8] N. Ghadiri Hedeshi and M. Saniee Abadeh, “Coronary Artery Disease Detection Using a Fuzzy-Boosting PSO Approach,” Comput Intell Neurosci, vol. 2014, pp. 1–12, 2014, doi: 10.1155/2014/783734.

[9] Heart tests, “https://www.heartfoundation.org.nz/your-heart/heart-tests,” https://www.heartfoundation.org.nz/your-heart/heart-tests, Jan. 25, 2023.

[10] A. Javeed, S. Zhou, L. Yongjian, I. Qasim, A. Noor, and R. Nour, “An Intelligent Learning System Based on Random Search Algorithm and Optimized Random Forest Model for Improved Heart Disease Detection,” IEEE Access, vol. 7, pp. 180235–180243, 2019, doi: 10.1109/ACCESS.2019.2952107.

[11] S. J. Pasha and E. S. Mohamed, “Novel Feature Reduction (NFR) Model With Machine Learning and Data Mining Algorithms for Effective Disease Risk Prediction,” IEEE Access, vol. 8, pp. 184087–184108, 2020, doi: 10.1109/ACCESS.2020.3028714.

[12] S. M. Saqlain et al., “Fisher score and Matthews correlation coefficient-based feature subset selection for heart disease diagnosis using support vector machines,” Knowl Inf Syst, vol. 58, no. 1, pp. 139–167, Jan. 2019, doi: 10.1007/s10115-018-1185-y.

[13] Y. Muhammad, M. Tahir, M. Hayat, and K. T. Chong, “Early and accurate detection and diagnosis of heart disease using intelligent computational model,” Sci Rep, vol. 10, no. 1, p. 19747, Nov. 2020, doi: 10.1038/s41598-020-76635-9.

[14] L. Ali, A. Rahman, A. Khan, M. Zhou, A. Javeed, and J. A. Khan, “An Automated Diagnostic System for Heart Disease Prediction Based on ${\chi^{2}}$ Statistical Model and Optimally Configured Deep Neural Network,” IEEE Access, vol. 7, pp. 34938–34945, 2019, doi: 10.1109/ACCESS.2019.2904800.

[15] C. Gupta, A. Saha, N. v Subba Reddy, and U. Dinesh Acharya, “Cardiac Disease Prediction using Supervised Machine Learning Techniques.,” J Phys Conf Ser, vol. 2161, no. 1, p. 012013, Jan. 2022, doi: 10.1088/1742-6596/2161/1/012013.

[16] Andras Janosi, William Steinbrunn, Matthias Pfisterer, and Robert Detrano, “ Heart Disease Data Set,” https://archive.ics.uci.edu/ml/datasets/heart+disease, Jan. 25, 2023.

[17] “Veri Bilimi ve Yapay Zekâ Tabanlı Web Yazılımları ,” http://biostatapps.inonu.edu.tr/.

[18] “Veri Sınıflandırma Yazılımı,” http://biostatapps.inonu.edu.tr/JVSY/.

[19] G. Guo, H. Wang, D. Bell, Y. Bi, and K. Greer, “KNN Model-Based Approach in Classification,” 2003, pp. 986–996. doi: 10.1007/978-3-540-39964-3_62.

[20] Y. I. A. Rejani and S. T. Selvi, “Early Detection of Breast Cancer using SVM Classifier Technique,” Dec. 2009.

[21] C. Kingsford and S. L. Salzberg, “What are decision trees?,” Nat Biotechnol, vol. 26, no. 9, pp. 1011–1013, Sep. 2008, doi: 10.1038/nbt0908-1011.

[22] M. Pal, “Random forest classifier for remote sensing classification,” Int J Remote Sens, vol. 26, no. 1, pp. 217–222, Jan. 2005, doi: 10.1080/01431160412331269698.

[23] R. E. Schapire, “Explaining AdaBoost,” in Empirical Inference, Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 37–52. doi: 10.1007/978-3-642-41136-6_5.

[24] H. Kamel, D. Abdulah, and J. M. Al-Tuwaijari, “Cancer Classification Using Gaussian Naive Bayes Algorithm,” in 2019 International Engineering Conference (IEC), Jun. 2019, pp. 165–170. doi: 10.1109/IEC47844.2019.8950650.

[25] Haoshi Zhang, Yaonan Zhao, Fuan Yao, Lisheng Xu, Peng Shang, and Guanglin Li, “An adaptation strategy of using LDA classifier for EMG pattern recognition,” in 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Jul. 2013, pp. 4267–4270. doi: 10.1109/EMBC.2013.6610488.

**Acknowledgement**

This article was published as a proceeding of 14th Turkish Congress of Medical Informatics Association in 2023.