Volume 14, Issue 2, Page 179-187, 2025 https://doi.org/10.46810/tdfd.1622670 Research Article

BİNGÖL UNIVERSITY

Türk Doğa ve Fen Dergisi
Turkish Journal of Nature and Science
www.dergipark.gov.tr/tdfd

Artificial Intelligence and Classification Algorithms In Heart Disease Data: Modern Approaches And Performance Comparison

Berat ELİAÇIK[1*], Ali Hakan IŞIK[1]

[1] Mehmet Akif Ersoy University, Engineering-Architecture Faculty, Computer Engineering Department, Burdur, Türkiye
Berat ELİAÇIK ORCID No: 0009-0008-7897-9200
Ali Hakan IŞIK ORCID No: 0000-0003-3561-9375

**Corresponding author: eliacikberat@gmail.com*

(Received: 20.01.2025, Accepted: 07.05.2025, Online Publication: 27.06.2025)

Keywords
Heart disease, Artificial intelligence, Performance comparison

Abstract: This study presents a data mining application aimed at investigating the prediction performance of classification algorithms on heart disease datasets. In this research, the likelihood of individuals having heart disease based on specific features was evaluated using various classification algorithms. The study utilised several classification algorithms, including regression, k- nearest neighbours (KNN), Naive Bayes, random forest, decision trees and support vector machines (SVM). All algorithms were implemented using the Python programming language and the Jupyter Notebook environment, and their classification performances were compared. The evaluation of success was based on metrics such as accuracy, sensitivity, specificity, and F1 score. According to the results, KNN, support vector machines and random forest algorithms achieved the highest performance with an accuracy rate of 86.79%, outperforming the other algorithms. This study highlights the potential of classification algorithms in the early diagnosis of heart disease, emphasising the significance of artificial intelligence and data mining applications in the healthcare field.

Kalp Hastalığı Verilerinde Yapay Zekâ ve Sınıflandırma Algoritmaları: Modern Yaklaşımlar ve Performans Karşılaştırması

Anahtar Kelimeler
Kalp hastalıkları, Yapay zekâ, Performans karşılaştırması

Öz: Bu çalışmada, sınıflandırma algoritmalarının kalp hastalıkları verilerindeki tahmin performansını incelemek amacıyla bir veri madenciliği uygulaması gerçekleştirilmiştir. Araştırma kapsamında, belirli özelliklere sahip bireylerin kalp hastalığı taşıma olasılığı, farklı sınıflandırma algoritmaları kullanılarak değerlendirilmiştir. Çalışmada regresyon, k-en yakın komşu (KNN), Naive Bayes, rastgele orman, karar ağaçları ve destek vektör makineleri (SVM) algoritmaları kullanılmıştır. Tüm algoritmalar Python programlama dili ve Jupyter Notebook ortamında uygulanmış ve sınıflandırma performansları karşılaştırılmıştır. Başarı değerlendirmesi doğruluk oranı, duyarlılık, özgüllük ve F1 skoru gibi ölçütler üzerinden yapılmıştır. Elde edilen sonuçlara göre, KNN, destek vektör makineleri ve rastgele orman algoritmaları %86,79 doğruluk oranıyla diğer algoritmalara kıyasla en yüksek başarıyı göstermiştir. Bu çalışma, kalp hastalıklarının erken teşhisinde sınıflandırma algoritmalarının potansiyelini ortaya koyarak, sağlık alanında yapay zekâ ve veri madenciliği uygulamalarının önemini vurgulamaktadır.

1. INTRODUCTION

Rapid developments in data storage technologies and the increasing impact of digitalisation have led to the rapid growth of large datasets. These developments have led to an increase in the number and capacity of databases, thus enabling data management to become more comprehensive and efficient. Systematic collection and analysis of data, especially health data, plays an important role in critical areas such as early diagnosis of diseases as well as transforming it into meaningful information. However, raw data are not meaningful on their own; they must be processed and analysed using the right methods and algorithms. In this context, data mining has an

important place as the process of extracting meaningful patterns from large data sets [13]. Data mining has a great potential in the field of health and the use of correct classification algorithms can contribute to critical decisions such as early diagnosis.

Heart diseases are one of the most common health problems in the world, and data-based analyses are of great importance for early diagnosis and correct treatment methods. In this study, the performance of different classification algorithms on data for early diagnosis of heart diseases is analysed. Data mining stands out as a powerful tool for the diagnosis of heart diseases and the main purpose of this study is to investigate how various classification algorithms can give successful results in this field.

This study was carried out on heart disease data created by John Moore's University in Liverpool, England and updated on 6 June 2020. The dataset consists of 1190 samples consisting of 11 features, and analyses were performed on these data with various classification algorithms. The algorithms used in the study include regression, k-nearest neighbour (KNN), Naive Bayes, random forest, decision trees and support vector machines (SVM). All algorithms were implemented in Python programming language and Jupyter Notebook environment and their classification performances were compared. Metrics such as accuracy rate, sensitivity, specificity and F1 score were used as success measures.

The findings revealed that KNN, support vector machines and random forest algorithms showed the highest success with an accuracy rate of 86.79%. These results reveal the potential of classification algorithms in the early diagnosis of heart diseases and emphasise the importance of artificial intelligence and data mining applications in the field of healthcare. This study constitutes a valuable example in terms of the applications of data mining and artificial intelligence techniques in the field of health and contributes to the scientific literature in this field.

2. LITERATURE REVIEW

Early diagnosis of heart diseases is of critical importance in the field of healthcare. In recent years, the use of machine learning and data mining techniques in this field has increased and many studies have been carried out on the performance of various classification algorithms.
In a study by Mustafa Coşar (2021), the use of machine learning algorithms in the detection of heart diseases was discussed. In this study, Random Forest algorithm gave the most successful results with an accuracy rate of 88%. Logistic Regression and k-nearest neighbour (kNN) algorithms showed lower performance with 85% and 70% accuracy rates, respectively.

Gamze Kaba and Seda Bağdatlı Kalkan (2022) compared machine learning classification algorithms for early diagnosis of cardiovascular diseases. In this study, Naive Bayes, Logistic Regression, Random Forest, k-nearest neighbour (kNN) and Support Vector Machines (SVM) algorithms were used. The results revealed the performance differences of different algorithms and contributed to the determination of the most successful algorithm.

These studies reveal the effectiveness of machine learning and data mining techniques and the performance of different algorithms in the early diagnosis of heart diseases. In our study, we aim to compare the performance of regression, k-nearest neighbour (KNN), Naive Bayes, Random Forest, decision trees and Support Vector Machines (SVM) algorithms on heart disease data, taking into account the findings in this literature.

3. FEATURES REQUIRED FOR COMPARISON OF CLASSIFICATION ALGORITHMS

Machine learning and artificial intelligence applications play a major role in data analysis and classification processes. This process consists of a series of steps such as data collection, data preparation, selection of appropriate algorithms and evaluation of model performance. In this section, the main steps used to evaluate the effectiveness of classification algorithms will be discussed.

3.1. Data Collection and Preparation

Data collection is the first stage of machine learning projects and this stage has a direct impact on the accuracy of the model. This process can be performed in one of two main ways: open data sources and closed data collection methods.

Open Data Sources: Open data collection is a method that is usually carried out with the permission of users. Users are subjected to a certain transparency and consent process when providing data to platforms. Such data can be collected from social media posts, publicly available data sets or open access surveys. These data sources offer great potential for analysis.

Closed Data Sources: Closed data collection focuses more on user interactions and behaviour. This process collects data by tracking users' online activities on digital platforms. For example, websites, mobile applications or e-commerce platforms use this data to deliver personalised content by tracking user behaviour.

3.2. Machine Learning Stages and Data Preparation

In the machine learning process, the correct processing of data is a critical factor that determines the success of the model. These stages are of great importance to ensure the accuracy of the data, increase the generalisation ability of the model and make sense of the results. These processes usually consist of the following steps:

3.2.1. Data selection and cleaning

Data selection aims to improve the accuracy and reliability of available data sources. Data cleaning ensures the elimination of missing or erroneous data that may adversely affect the accuracy of the model. In machine learning applications, eliminating erroneous data entries accelerates the learning process of the model and prevents unnecessary complexities.

3.2.2. Data integration

Data integration is the process of combining data from different sources. This stage ensures that information from multiple data sources is presented in a consistent and harmonised manner. In modern machine learning algorithms, data integration enables the model to produce more accurate and comprehensive results.

3.2.3. Data transformation and feature selection

In machine learning algorithms, it is important that the data is in a certain format. Data transformation is the process of making raw data processable and analysable. Feature selection ensures that only the variables that are meaningful for the model are selected. This process accelerates the learning process of the algorithm and enables the model to work more efficiently.

3.2.4. Model selection and training process

The correct selection of machine learning algorithms is critical to the success of the project. This stage includes the process of selecting the most appropriate algorithm from different algorithms such as classification or regression and training it on the data. In this study, classification of heart disease data was performed using algorithms such as decision trees, KNN, Naive Bayes and random forest.

3.2.5. Evaluation of model performance

The success of the model is measured by various evaluation metrics. These metrics include accuracy rate, sensitivity, specificity and F1 score. These metrics evaluate how well the model fits real-world data and its ability to make accurate predictions. In this study, the performance of the algorithms used in was evaluated with a comparative analysis and it was observed that KNN, SVM and random forest algorithms stand out with high accuracy rates.

3.2.6. Visualisation and presentation of results

Presenting the results obtained with visual tools is an important step to understand and share the success of the model. The results are presented in the form of user-friendly reports, graphs or tables using visualisation techniques. This stage enables the data obtained to be shared in a more understandable and effective way.

3.2.7. Translating results into practice

Translating results into practice demonstrates the capacity of machine learning models to solve real-world problems. This stage demonstrates how the results of the model contribute to real-world applications in the healthcare field, for example, early detection of heart disease.The results enable the creation of supportive tools for healthcare professionals.

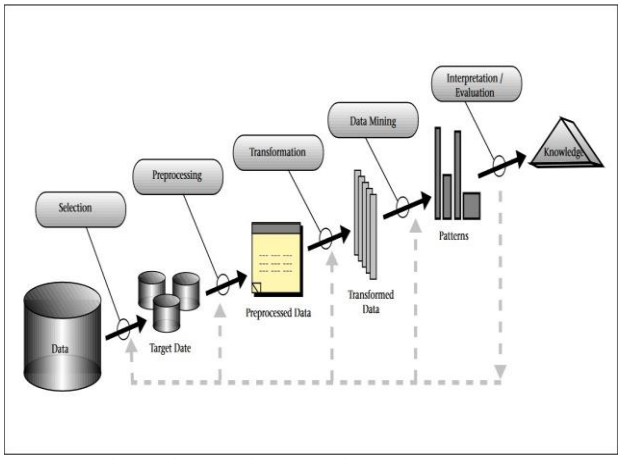

Figure 1. The Modern Machine Learning Process: From Raw Data to Actionable Insights

As illustrated in Figure 1 (The Modern Machine Learning Process: From Raw Data to Actionable Insights), the journey from raw data to valuable insights follows a structured and iterative process. This process begins with the careful selection of relevant data sources that align with the objectives of the study. Once selected, the data undergoes preprocessing steps such as cleaning, filtering, and standardization to ensure its quality and consistency. The next stage involves data transformation, where techniques like normalization, feature extraction, and dimensionality reduction are applied to prepare the data for analysis. At the core of the process lies the application of machine learning algorithms, whether supervised or unsupervised, which are used to uncover hidden patterns, make predictions, or classify data effectively. Finally, the outcomes are interpreted and evaluated to translate algorithmic results into actionable insights, assess model performance, and validate findings for practical use. This modern approach emphasizes not just extracting data, but actively learning from it using advanced machine learning methods to derive meaningful knowledge.

3.3. Machine Learning Models

Machine learning involves advanced analysis techniques used to discover meaningful patterns and relationships within large data sets. These techniques are supported by statistical models, machine learning algorithms, deep learning methods, and mathematical approaches such as decision trees, support vector machines, and neural networks. The models employed in this study enable accurate results by performing classification processes on the data, thus improving prediction capabilities and decision-making effectiveness.

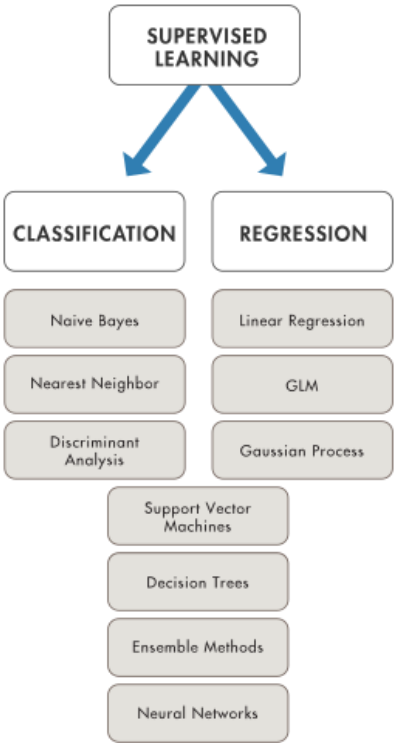

Figure 2. Supervised Learning Models

As shown in Figure 2 (Supervised Learning Models), supervised learning techniques are broadly divided into two main categories: classification and regression. Classification methods include algorithms such as Naive Bayes, Nearest Neighbor, Discriminant Analysis, Support Vector Machines, Decision Trees, Ensemble Methods, and Neural Networks, all of which are applied to categorical outcome predictions. On the other hand, regression methods, including Linear Regression, Generalized Linear Models (GLM), and Gaussian Processes, are designed to predict continuous outcomes.In the context of this study, the classification models outlined in Figure 2 were used to analyze heart disease data, aiming to distinguish between patients with and without heart disease. The selection of these models was based on their proven performance in medical data classification tasks, and their implementation has been thoroughly evaluated to ensure the robustness and accuracy of the results obtained.

3.3.1. Clustering

Clustering is one of the unsupervised learning methods and aims to group data with similar characteristics. This method is used to discover relationships in the data set and identify similarities between different data points. Clustering algorithms divide data points into natural clusters, allowing the features within each data group to be closer to each other. In analyses of health data such as heart disease, clustering methods are useful in identifying different groups of patients.

3.3.2. Association rules

Association rules are an approach to identifying relationships between variables in a data set. This method helps us to understand how a particular event can influence other events. Association rules can be classified into three main groups: multilevel association rules, multidimensional association rules and quantitative association rules. In machine learning and artificial intelligence applications, association rules are often used to discover possible relationships in health data. For example, it is possible to observe risk factors related to heart diseases together thanks to such analyses.

3.3.3. Classification

Classification is one of the supervised learning methods and is used to classify data into predefined classes. This method aims to predict which class each data point belongs to with various classifier algorithms. In this study, it is aimed to make accurate disease predictions by using various algorithms for the classification of heart disease data. Such classification problems offer an important application area for the analysis of health data.

3.4. Types of Classification Models

Classification can be performed using different algorithms and methods. These methods can be categorised according to the mathematical models and learning techniques used. Below are descriptions of the most common classification algorithms and how they are applied to health problems such as heart disease.

3.4.1. Regression

Regression is a method used to predict a continuous target variable and is often used in modelling linear relationships. Different types of regression, such as linear regression, multiple linear regression and non-linear regression, are used to better understand the relationship between data. In heart disease prediction, regression analysis can be used to predict possible disease development by performing risk analyses based on patients' health histories.

3.4.2. Naive Bayes

Naive Bayes is an algorithm for classification problems based on Bayes theorem. In particular, it gives effective results under the assumption of independence between data points. Naive Bayes classifier has the advantage of obtaining fast and accurate results,especially in large data sets. In the classification of medical data such as heart diseases, the Naive Bayes algorithm is widely used to estimate the probability of individuals carrying the disease.

3.4.3. K-nearest neighbours (knn)

K-Nearest Neighbours (KNN) is one of the lazy learning methods and does not create a trained model. Instead, it stores all the training data and performs classification by finding the nearest neighbours similar to the new incoming data. The KNN algorithm is widely used, especially to minimise classification errors. In heart disease data, the KNN algorithm can be effective as a

method of grouping people with similar characteristics to determine the risk status of individuals.

3.4.4. Decision tree

The decision tree is one of the most widely used algorithms for classification of data. This algorithm produces a set of rules that make decisions by sorting the data and is easy to visualise. Decision trees are particularly useful for making data more understandable and require low data preprocessing. In heart disease prediction, decision trees can be used to classify patients' possible disease states by considering different risk factors.

3.4.5. Random forest

Random forest is an algorithm that consists of a combination of many decision trees and combines the prediction of each tree to create a more powerful classifier. This algorithm is often used to improve accuracy and shows high performance on large data sets. In health data such as heart diseases, the random forest algorithm can improve the accuracy of disease predictions by evaluating various factors.

3.4.6. Support vector machine (svm)

Support Vector Machine (SVM) is a classification algorithm that gives effective results in high dimensional data sets and tries to separate data points in the best way. SVM is a preferred method to obtain accurate and generalisable results, especially in classification processes. In the classification of medical data such as heart diseases, the SVM algorithm has the ability to make accurate classifications according to the health status of patients.

3.5. Evaluation of Model Success

Various metrics are used to evaluate artificial intelligence and machine learning modelsuccess. Performance metrics such as accuracy, precision, sensitivity, F-score and AUC-ROC are basic tools for understanding the effectiveness and accuracy of the model.One of the most common methods used to evaluate model performance is the confusion matrix. The confusion matrix visually represents the performance of classification algorithms, allowing the model's predictions to be compared with the actual labels.

Table 1. Confusion Matrix

		Projected Status	
		Class = 1	Class = 0
Actual Situation	Class=1	TP (True Positive)	FN (False Negative)
	Class=0	FP (False Positive)	TN (True Negative)

3.5.1. Accuracy

Accuracy is a key performance metric that measures how often the model makes correct predictions. The accuracy of a model is calculated as the ratio of correct predictions to total predictions. Although high accuracy indicates the overall success level of the model, it can be misleading in unbalanced data sets. Accuracy is calculated by the following formula:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

In our study, although the accuracy of the model is high in the classification processes performed on heart disease data, it is necessary to pay attention to some class imbalances.

3.5.2. Precision

Precision indicates how many of the model's predictions of the positive class are correct. The precision metric is particularly useful when false positive predictions should be low. In health data, it is desirable to have fewer false positives (e.g. patients predicted to be healthy). Precision is calculated as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

In the study on heart disease prediction, we tried to keep the precision metric high in order to minimise the false positive rate.

3.5.3. Recall

Sensitivity measures the rate at which the true positive class is predicted. This metric is particularly critical in situations where false negatives (true positives that are incorrectly assigned to the negative class by the model) need to be minimised. In health data, minimising false negatives prevents patients from being missed. Sensitivity is calculated as follows:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

In our study, the sensitivity metric was prioritised especially for the detection of diseases, and it was aimed to identify the patients accurately.

3.5.4. F-Score

The F-score is used to evaluate the trade-off between precision and sensitivity. The F- score combines both metrics and presents the model performance as a single value. The F- score is maximised when the balance between precision and sensitivity is achieved. By taking both factors into account, this metric more accurately measures the overall success of the model. The F-score is calculated by the following formula:

$$\text{F Score} = \frac{2 \text{ x Precision x Sensitivity}}{\text{Precision} + \text{Sensitivity}}$$

In our study, a high F- score was targeted so that the model performs in a balanced way by reducing both false positives and false negatives.

4. APPLICATION

In this study, the data of individuals in Hungary, Switzerland, Cleveland, Ohio, USA and Long Beach, California were analysed. In this dataset, patients with chest pain complaints were analysed whether they had any of the heart diseases as a result of the examinations.Based on this data, a machine learning model was developed to predict whether an individual has heart disease or not. The success of the developed model was measured by the performance evaluation metrics used.

4.1. Dataset Used

The dataset used in this study has been compiled and merged to support research into the development of machine learning and artificial intelligence algorithms focused on Coronary Artery Disease (CAD). Last updated on 6 June 2020, the dataset includes 1190 examples with 13 key features, including age, sex, chest pain type (cp), resting blood pressure (trestbps), serum cholesterol (chol), fasting blood sugar (fbs), resting electrocardiographic results (restecg), maximum heart rate achieved (thalach), exercise-induced angina (exang), ST depression induced by exercise (oldpeak), slope of the peak exercise ST segment (slope), number of major vessels (ca), and thalassemia status (thal), along with the target variable indicating the presence of heart disease.

age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
57	0	0	120	354	0	1	163	1	0.6	2	0	2	1
...	...	...	...	...	...	...	...	...	...	...	...	...	...

Figure 3. Sample Rows Illustrating the Heart Disease Dataset

To enhance understanding of the dataset, example rows are provided in Figure 3, offering a snapshot of the data's structure and values. Sharing these sample data points helps clarify the nature of the features and supports a better grasp of the inputs used in model development.

4.2. Python and Data Mining Libraries

The operations on the data were performed using the Python programming language. Python is a software language that was released in 1990 and is widely used today for data science, machine learning and artificial intelligence applications. Python's simple syntax and ease of learning have made this language popular in both academic and industrial fields. Python, which contains the necessary libraries for data mining and machine learning, is an ideal choice for high performance computing. Within the scope of the study, data classification was performed using Jupyter Notebook, which is included in the Anaconda Navigator platform developed for Python. Jupyter Notebook is a powerful tool that facilitates interactive data analysis and modelling processes.

4.3. Data Preprocessing Steps

The dataset contains 76 attributes (features) and only 11 attributes have been analysed in existing studies. These attributes reflect the most important features of heart disease. In the data preprocessing stage, only these 11 attributes were evaluated in order to increase the success of the model. The selection of attributes is a step towards improving the accuracy and generalisability of the model. The data were cleaned and appropriately transformed to avoid any missing information or anomalies.

The dataset used in the study was created by John Moore's University in Liverpool, England. The data provides a wide range of information on the health status of individuals in the region. During the comparison of data mining algorithms, model successes were evaluated using the data of individuals in this region.

4.4. Partitioning the Dataset into Training and Test Sets

In order to accurately evaluate the performance of machine learning models, the data is divided into training and test sets. The training data set is used for the learning process of the model, while the test data set is used to evaluate the overall performance of the model. The data is divided as 80% training set and 20% test set. This data splitting technique is important to predict how the model will perform on new and unseen data. Splitting the data into training and test sets increases the ability of the model to generalise without overfitting. In this way, the probability of heart disease was estimated as a result of the tests of the model on the patient. While the model learnt to classify with the data used in the training process, the accuracy and generalisation power of the model were measured with the test set.

4.5. Performance Results of Classification Algorithms

In this study, the performance of different machine learning algorithms on heart disease prediction is evaluated. The results obtained with the sensitivity, precision, accuracy and F-score metrics of the algorithms used are presented below.

Table 2. Performance Comparison of Various Classification Algorithms

Algorithm	Recall	Precision	Accuracy	F- score
Logistic Regression	%85,18	%85,18	%86,88	%85,18
KNN	%85,18	%88,46	%88,52	%86,79
SVM	%85,18	%88,46	%88,52	%86,79
Naive Bayes	%85,18	%85,18	%86,88	%85,18
Decision Tree	%77,78	%70	%75,40	%73,68
Random Forest	%85,18	%88,46	%88,52	%86,79

Table 2 presents a comparative analysis of several classification algorithms based on key performance metrics, namely recall, precision, accuracy, and F- score. Logistic Regression shows a stable performance, with an accuracy of 86.88% and an F-measure of 85.18%,

reflecting its reliable predictive capabilities. In contrast, the K Nearest Neighbour (KNN), Support Vector Machine (SVM), and Random Forest algorithms display higher efficacy, with accuracy values of 88.52% and corresponding F- score scores of 86.79%. These results suggest that these algorithms are more effective in terms of balancing precision and recall. On the other hand, the Decision Tree algorithm exhibits a lower performance across all metrics, particularly in certainty and F- score, which suggests its relative inefficiency compared to the other models evaluated. This discrepancy highlights the Decision Tree's tendency to overfit or underperform in certain data scenarios, making it less suitable for tasks requiring high accuracy and certainty.

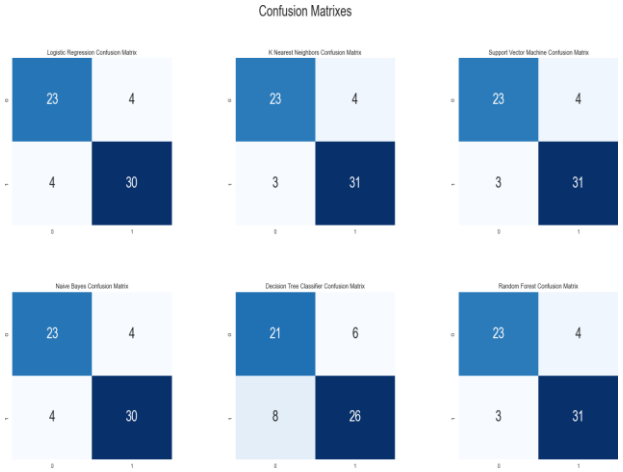

Figure 4. Confusion matrix images of Logistic Regression, KNN, SVM, Naive Bayes, Decision Tree and Random Forest algorithms.

The figure showing the effect of gender on the risk of heart disease can be interpreted as follows:

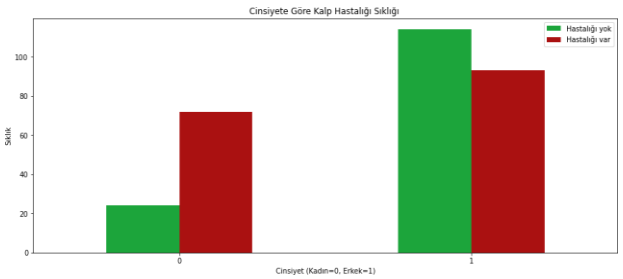

Figure 5. Frequency of Heart Disease by Gender

The graph shows the differences between female (0) and male (1) individuals with heart disease. The prevalence of heart disease is higher in men than in women. This finding reveals that men are in the higher risk group for heart disease. The frequency of heart disease according to age groups is shown in the graph below:

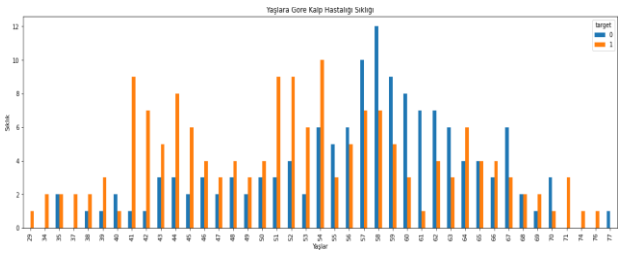

Figure 6. Frequency of Heart Disease by Age

When the age groups were analysed, it was observed that the risk of heart disease increased significantly in the 50-60 age range. This indicates that middle-aged individuals should be more careful about heart health. The relationship between the slope of the ST segment after exercise and the frequency of heart disease is shown in the graph below:

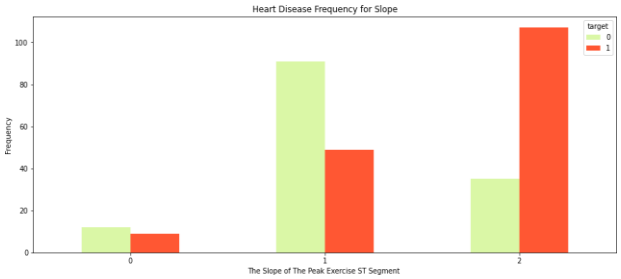

Figure 7. Exercise ST Segment Slope and Frequency of Heart Disease

This graph shows the distribution of individuals with heart disease at various levels of the slope (0, 1, 2). Individuals with a positive slope (2) were found to have a much higher risk of heart disease than the other groups. This supports that cardiac function after exercise may be an important predictor.

5. CONCLUSION AND EVALUATION

In this study, different machine learning classification algorithms were used to predict individuals who may have heart disease. In line with the results obtained, the accuracy, precision, sensitivity and F1-score measures of the algorithms are analysed in detail. In addition, the performance metric of each algorithm is supported by confusion matrix visualisations (Figure 3). According to the confusion matrix analyses, K-Nearest Neighbors (KNN), Support Vector Machine (SVM) and Random Forest algorithms showed the best performance in all measurement criteria. These three algorithms were the most successful algorithms with 88.52% accuracy, 88.46% precision and 86.79% F1-score. On the other hand, Logistic Regression and Naive Bayes algorithms achieved satisfactory results in terms of accuracy and precision with 86.88% and 85.18% respectively. On the other hand, the Decision Tree algorithm showed the lowest performance with an accuracy of 75.40%. Confusion matrix visualisations (Figure 3) present the classification success of the algorithms in more detail. For example, KNN, SVM and Random Forest algorithms managed to keep the false positive and false negative classification rates at a minimum level, while the Decision Tree algorithm was found to have higher rates. Logistic Regression and Naive Bayes algorithms provide satisfactory results in correct classification, but have some error margin in false positive and negative classifications.

It is an important finding of this study that sensitivity and precision criteria should be considered together. In particular, the sensitivity criteria of KNN, SVM and Random Forest algorithms were calculated as 85.18%. The high performance of these algorithms can be attributed to the balanced representation of the classes in the data set and the generalisation capabilities of the algorithms. On the other hand, the sensitivity of the

Decision Tree algorithm was significantly lower than the other algorithms with 77.78%.

In addition, in line with the studies in the literature, it should be emphasised that the performance of an algorithm may vary depending on the data set used (Alan & Karabatak, 2020). In this context, it has been observed that factors such as the size of the data set, class distribution and data preprocessing steps directly affect the performance of algorithms in data mining processes.In conclusion, in this study, KNN, SVM and Random Forest algorithms showed the highest success in heart disease prediction. However, algorithm selection in data mining studies should take into account the characteristics of the data set, the performance measures used and the requirements specific to the application area. The findings of this study provide an important guide for the selection of the right algorithm in medical diagnosis systems and other classification problems.

REFERENCES

[1] Alan A, Karabatak M. Evaluation of Factors Affecting Performance in Data Set - Classification Relationship. Fırat University Müh. Bil. Journal. 2020.

[2] Aydın S, Özkul AE. Data mining and an application in Anadolu University open education system. Journal of Education and Training Research. 2015.

[3] Berry MJ, Linoff GS. Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management. Wiley. 2004.

[4] Ergün K. Data Mining. Abdullah Baykal; [cited 2015 Dec 10]. Available from: http://abdullahbaykal.com.tr/verimaden.pdf.

[5] Fayyad U, Piatetsky-Shapiro G, Smyth AP. From Data Mining to Knowledge Discovery in Databases. American Association for Artificial Intelligence. 2008.

[6] Gürbüz F, Özbakır L, Yapıcı H. A Data Mining Application on Parts Removal Reports of an Airline Business in Turkey. Journal of Gazi University Faculty of Engineering and Architecture. 2013.

[7] Hungarian Institute of Cardiology. Heart Disease. Budapest; 2017.

[8] JavaPoint. Data Mining Techniques. [Internet]. 2011 [cited 2021 Mar 10]. Available from: https://www.javatpoint.com/data-mining-techniques.

[9] Kalıkov A. Data Mining and an E-Commerce Application. PegemA; 2006.

[10] Kaya M, Özel SA. Comparison of Open Source Data Mining Software. Academic Informatics. 2017.

[11] Özekes S. Data Mining Models and Application Areas. Istanbul Commerce University Journal. 2003.

[12] Park BH, Kargupta H. Distributed Data Mining: Algorithms, Systems, and Applications. CiteSeerx. 2002.

[13] Ramageri MB. Data Mining Techniques and Applications. Indian Journal of Computer Science and Engineering. 2010.

[14] Chollet F. Xception: Deep Learning with Depthwise Separable Convolutions. arXiv preprint arXiv:1610.02357. 2017.

[15] Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC. Mobilenetv2: Inverted residuals and linear bottlenecks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2018. p. 4510-20.

[16] Tan M, Le QV. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In: International Conference on Machine Learning; 2019. p. 6105-14.

[17] So-In C, Mongkonchai N, Aimtongkham P, Wijitsopon K, Rujirakul K. An evaluation of data mining classification models for network intrusion detection. In: Fourth International Conference on Digital Information and Communication Technology and its Applications (DICTAP); 2014.

[18] Şengür D, Tekin A. Prediction of Graduation Grades of Students with Data Mining Methods. Journal of Information Technologies. 2014.

[19] Rajak R. Deep Learning for Marble Defect Classification. Medium. 2022 Mar 20 [cited 2022 Mar 20]. Available from: https://rishirajak.medium.com/deep-learning-for-marble-defect-classification-f7928d077056.

[20] Selver MA, Akay O. Feature Extraction for Quantitative Classification of Marbles. ResearchGate. 2007.

[21] T.C. Ministry of Health. Turkish Society of Cardiology. [Internet]. 2015 [cited 2015 Oct 5]. Available from: https://tkd.org.tr/TKDData/Uploads/files/Turkiye-kalp-ve-damar-hastaliklari-onleme-ve-kontrol-programi.pdf.

[22] He K, Zhang X, Ren S, Sun J. Identity mappings in deep residual networks. In: European Conference on Computer Vision; 2016. p. 630-45.

[23] Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. The Journal of Machine Learning Research. 2014;15(1):1929-58.

[24] Iandola FN, Han S, Moskewicz MW, Ashraf K, Dally WJ, Keutzer K. SqueezeNet: AlexNet-level Accuracy with 50x Fewer Parameters and <0.5MB Model Size. arXiv preprint arXiv:1602.07360. 2016.

[25] Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, et al. Mobilenets: Efficient Convolutional Neural Networks for Mobile Vision Applications. arXiv preprint arXiv:1704.04861. 2017.

[26] Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the Inception Architecture for Computer Vision. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2016. p. 2818-26.

[27] He K, Zhang X, Ren S, Sun J. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. In: European Conference on Computer Vision; 2015. p. 346-61.

[28] Talend. Data Mining Techniques: The Complete List. [Internet]. 2020 [cited 2020 Jul 10]. Available

from: https://www.talend.com/resources/data-mining-techniques/.
[29] Taylor D. Data Mining Tutorial: What is | Process | Techniques & Examples. guru99. [Internet]. 2021 [cited 2021 Jan 8]. Available from: https://www.guru99.com/data-mining-tutorial.html#11.
[30] Tekerek A. Data Mining Processes and Open Source Data Mining Tools. Academic Informatics. 2011.
[31] Visa S, Ramsay B, Ralescu A, Knaap Ev. Confusion Matrix-based Feature Selection. Omnipress - Madison, WISCONSIN. 2011.
[32] Waikato. Weka 3: Machine Learning Software in Java. [Internet]. 2017 [cited 2017 Mar 25]. Available from: https://www.cs.waikato.ac.nz/ml/weka/.