Journal of Information Technology Management
ISSN-OnLine: 2980-7972
University of Tehran

Predicting Heart Disease Using Automated Machine Learning Based on Genetic Algorithms

Ahmad Jafarnejad *

*Corresponding author, Prof., Department of Industrial Management, Faculty of Industrial Management and Technology, College of Management, University of Tehran, Tehran, Iran. E-mail: jafarnjd@ut.ac.ir

Arman Rezasoltani

Ph.D. Candidate, Department of Industrial Management, Faculty of Industrial Management and Technology, College of Management, University of Tehran, Tehran, Iran. E-mail: armanrezasoltani@ut.ac.ir

Amir Mohammad Khani

Ph.D. Candidate, Department of Industrial Management, Faculty of Industrial Management and Technology, College of Management, University of Tehran, Tehran, Iran. E-mail: amir.mo.khani@ut.ac.ir

Journal of Information Technology Management, 2025, Vol. 17, Issue 2, pp. 91-122
Published by the University of Tehran, College of Management
doi: https://doi.org/ 10.22059/jitm.2024.382556.3829
Article Type: Research Paper

Received: October 22, 2024
Received in revised form: December 16, 2024
Accepted: December 29, 2024
Published online: April 20, 2025

Abstract

This study aims to apply automatic machine-learning approaches using genetic algorithms to enhance heart disease prediction. Heart disease has remained the major cause of mortality in the world, necessitating an effective and timely diagnosis. Most current diagnostic and assessment processes are lengthy and expensive, relying heavily on clinical expert knowledge. To help address these issues, machine learning approaches, which derive their utility from examining substantial datasets for the recognition of patterns, have emerged as a potential solution, providing solutions beyond those achievable by human recognition alone. Genetic algorithms are also suited to addressing these issues as they mimic natural evolution to perfect high-caliber machine-learning models, feature selection, and parameter selection in machine-learning applications. This study examines the utilization of genetic algorithms working alongside AutoML frameworks to improve accuracy in heart disease predictions.

Reducing to the best combination of attributes and the optimum parameters for each attribute is a time-consuming task, so automating this aspect of the process allows for more accurate and prompt predictions, consequently reducing the manual work. The AutoML approach followed in this research is TPOT, which uses genetic algorithms to ascertain optimally designed machine-learning pipelines. The application of AutoML, together with genetic algorithms, is the most prominent finding that yields a significant improvement in the quality of the predictions for heart disease compared to the traditional assessment approaches, with an accuracy of 93.8%. This approach will enhance diagnostic accuracy and enable early diagnosis, thereby reducing the likelihood of misdiagnoses or ineffective treatments and ultimately lowering associated costs.

Keywords: Heart Disease Prediction, Automatic Machine Learning, Genetic Algorithms, TPOT Framework.

Introduction

Heart disease is known as one of the main causes of death worldwide, and its early and accurate diagnosis plays a critical role in reducing this statistic. Traditional methods of heart disease diagnosis, such as angiography and electrocardiogram, are usually expensive and time-consuming and may require clinical expertise. In this regard, new medical technologies have advanced to include machine learning as a powerful tool for disease prediction and diagnosis. This approach enables the analysis of a large volume of medical data, uncovers complex patterns extracted by humans, and generates appropriate predictions based on these patterns (Shah et al., 2020). Optimization algorithms are among the prominent approaches employed to increase the performance of machine learning models. Genetic algorithms evolved from principles in natural evolution and have been realized as a significantly robust optimization method applicable in machine learning. Using selection, mutation, and crossover processes, these algorithms help find the best combination of parameters to improve the accuracy of machine learning models. Since the prediction of heart diseases requires adequate accuracy, the combination of genetic algorithms with automatic machine learning can create more effective prediction models (Yadav et al., 2021).

Over the years, many studies showed that genetic algorithms improve the performance of machine learning models such as Artificial Neural Networks (ANN) and Support Vector Machines (SVM). For example, some research revealed that using genetic algorithms to optimize neural network parameters significantly improves the accuracy of predicting heart disease (Arroyo & Delima, 2022). Meanwhile, a new and efficient approach for early detection of such disease conditions involves an automatic learning machine powered by a genetic algorithm to predict heart diseases. These approaches increase the quality of medical care by reducing diagnostic errors and thus reducing treatment costs due to misdiagnosis (Verma et al., 2021). Cardiovascular diseases are one of the leading causes of death in the

world, especially in Iran. Early diagnosis and prediction of these diseases have become more critical due to the increasing prevalence of risk factors such as obesity, diabetes, high blood pressure, and unhealthy lifestyles in Iranian society. Although the traditional methods of diagnosing heart disease are widely used, they cannot provide accurate and reliable results due to the complexity of the factors affecting this disease. On the other hand, the growing volume of medical and clinical data has increased the need for more advanced methods to analyze and extract hidden patterns in these data.

In the meantime, AutoML algorithms based on genetic algorithms as a new approach for predicting heart diseases can raise the possibility of improving the accuracy of diagnosis and reducing treatment costs. On the other hand, due to financial resource constraints and a lack of treatment facilities in most regions of the country, there is a need for automatic and low-cost forecasting methods. In this regard, genetic algorithm-based AutoML can be an efficient tool for heart disease prediction and management in Iran due to its ability to analyze large volumes of data and optimize prediction models. This research can develop algorithms to enable high accuracy and early diagnosis of heart diseases at a lower cost, preventing an increase in the number of patients and treatment costs in the future. Cardiovascular diseases (CVDs) have been among the top-ranking causes of death in Iran. Based on the data obtained from research studies, the rate of deaths caused by CVDs in some regions of Iran is between 30% and 47% of all deaths (Fahimfar et al., 2018; Sarraf-Zadegan et al., 1999). Another study has also highlighted that in Iran, about 46% of the burden of diseases and deaths is due to cardiovascular diseases such as heart attacks and strokes (Sarrafzadegan & Mohammmadifard, 2019). On the other hand, the volume of medical and clinical data has grown, increasing the need for more advanced methods of analysis and extraction of hidden patterns in the data. Applying AutoML algorithms based on genetic algorithms as a new approach to the prediction of heart diseases offers an opportunity to enhance diagnosis accuracy while reducing the cost of treatment.

Although previous works have significantly improved heart disease prediction using machine learning and optimization algorithms, some setbacks still plague such research. First, there is a strong need for more mixed models to boost accuracy while reducing the overall errors. Most research is confined to stand-alone algorithms such as ANN, SVM, or random forests. However, none of these algorithms can show the best performance solely on their own, as each has its strengths and weaknesses. Besides, previous studies have not fully covered another key feature selection problem. Though the feature selection increased the accuracy of the results in some studies, such as Yadav et al. (2021) and Nazari and Jodki (2020), it was not addressed in many of them. While feature selection may help not only decrease the model complexity but also improve its quality, the best results can be achieved only by using an automated optimization method.

The current study tries to fill these knowledge gaps by linking AutoML with genetic algorithms. Using AutoML with the support of the genetic algorithm allows for the automatic selection of the best combination of parameters and features. Our proposed model could give more accurate predictions while reducing requirements for manual tuning. TPOT, in combination with other advanced AutoML tools whose basis is genetic algorithms, will offer an optimal and exact model for heart disease prediction. The study provides a valid contribution to enhancing early detection protocols related to heart diseases by giving improved results on accuracy, sensitivity, and predictive precision, especially when related to previously conducted studies. This enhancement in performance underscores the significance of automated optimization in medical applications and highlights its potential to become a standard practice within healthcare systems. The paper is structured as follows: the literature review section explores previous research on automated machine learning and genetic algorithms. The methodology section outlines the research method and implementation steps. Then, the results of the proposed model are presented, evaluated using different criteria, and compared to alternative models. Finally, the findings are analyzed in the discussion and conclusion sections, and recommendations are also offered for future research.

Literature Review

Automatic machine learning

AutoML is one of the most significant recent advancements in machine learning that intends to minimize or eliminate the aspects of human interaction in the different phases of the machine learning model creation. AutoML is one of the most significant recent advancements in machine learning that intends to minimize or eliminate the aspects of human interaction in the different phases of the machine learning model creation. AutoML helps users eliminate time-consuming data preprocessing, algorithm selection, hyperparameter tuning, and model assessment, enabling them to create accurate predictive models for tasks and datasets of their interest (Truong et al., 2019). Thus, AutoML makes it possible to speed up the prediction of outcomes and minimize the level of deep expertise in machine learning (Mueller et al., 2020).

Since the idea of AutoML came to the surface, it has been growing quite rapidly. Auto-WEKA, proposed by Kotov et al. in 2019, was among the first and successful systems to solve it. It also enabled the investigators to concentrate on creating even better models much faster (Nguyen et al., 2021). Several stages for AutoML should be considered. First, a brief overview of AutoML will be provided. It is administered as the first phase in the data preprocessing before being taken through the machine learning models. We ensure that the data go through the cleaning process, defects are rectified, and some features that cannot be used are formatted appropriately. AutoML tools automatically perform normalization and standardization for this purpose (Feurer et al., 2019). The next step is feature selection, in which the relevant features affecting the model results are determined automatically. At this

stage, metaheuristic methods, specifically genetic algorithms, are applied to acquire the chosen features for the model (Mohan et al., 2019).

In the modeling phase, several machine learning algorithms are applied to the input data, and AutoML chooses the right algorithm for prediction. One such software is Auto-sklearn, which selects the model by comparing different parameters such as accuracy and F1 score (Feurer et al., 2019). Another issue within traditional machine learning is how to set the hyperparameters. In AutoML, this process is carried out using random search and network search algorithms, which enhance both the accuracy and efficiency of the models (Bergstra & Bengio, 2012). AutoML has several advantages, such as cutting down the time of algorithm development, high model accuracy, and putting machine learning in the hands of more people. It also offers the convenience of saving time on the development of the models as it automates several operational processes. AutoML can also contribute to increasing the precision of models through selecting the best algorithms and using optimal hyperparameter searches (Mangalath Ravindran et al., 2022). A key objective of this technology is to make machine learning accessible to everyone, including individuals with little technical expertise in the field. Other tools like TPOT and H2O have aimed at making machine learning accessible in various fields (Olson & Moore, 2019). AutoML is thus a pivotal technology that simplifies the implementation of machine learning, making it more accessible and user-friendly. With its rapid advancements, it is expected to have more extensive contributions to driving broader applications across various industries, particularly in data analysis and prediction.

Genetic algorithm

A genetic algorithm is a robust optimization and search method inspired by biological processes such as natural selection, combination, and mutation. These algorithms create more optimal generations by starting from an initial population of random solutions and using a fitness function to evaluate the solutions. Genetic algorithms are widely used in optimization and complex-based problems because they have simple coding and high efficiency (Katoch et al., 2021). One of the most significant theories about the genetic algorithm is the "Schema Theorem" presented by John Holland, which states that parts of the chromosomes more compatible with the environment are preserved more in the next generations and lead to a gradual improvement in the performance of the population. This principle has been widely used to analyze the behavior and predict the performance of genetic algorithms (Whitley, 1994). Despite the extensive experimental successes of genetic algorithms in solving complex problems, there are still many theoretical debates surrounding the performance of these algorithms. Some researchers have concluded that genetic algorithms are theoretically less efficient than deterministic optimization algorithms, and the so-called "No Free Lunch Theorem" shows that no algorithm can perform best in all optimization problems (Salomon, 1997).

Other miscellaneous recent developments in the theory of genetic algorithms include new techniques such as multi-objective algorithms, parallel algorithms, and hybrid genetic algorithms. These techniques will enhance the algorithm's efficiency in solving problems like premature convergence and getting stuck in local optima (Oliveto et al., 2020). Genetic algorithms are still recognized as one of the most effective optimization tools, although there is a need for further development in areas such as the selection of genetic operators, convergence models, and performance of fitting functions to improve their performance and adapt them to specific problems (Katoch et al., 2021). Finally, the genetic algorithm, as one of the most robust optimization methods based on the simulation of biological processes, has been able to find wide applications in various fields, including the optimization of structures, complex modeling, and even bioinformatics problems (McCall, 2005).

Research background

Over the past few decades, the fields of machine learning and big data analytics have witnessed remarkable advancements, particularly in the medical domain. Among the most relevant medical challenges are those relating to heart disease prediction and its early diagnosis since their physiologic and clinical complications call for more precise diagnostic and preventive tools. The most excellent approaches for improving the accuracy of heart disease prediction involve machine learning with optimization techniques, including those supported by genetic algorithms. Genetic algorithms automatically choose the most appropriate features and models in the prediction of heart diseases by applying the idea of their generation mechanisms to the optimization process based on the principle of natural selection and evolution. Combining AutoML with these algorithms makes prediction systems more accurate than traditional models. Table 1 presents a summary of the research background related to using AutoML based on genetic algorithms for heart disease prediction.

Table 1. Research background summarizing studies on heart disease prediction using machine learning, genetic algorithms, AutoML frameworks, and hybrid approaches

Authors	Title of the article	Results	The model used	Focus Area
Maihami et al. (2016)	Designing an expert system to identify heart diseases using fuzzy systems	The proposed algorithm, tested on data from Tawheed Sanandaj Hospital, accurately predicted individuals at a 98% accuracy rate.	Fuzzy algorithms	Machine Learning
Mansouri and Dadvar (2017)	Diagnosing heart attacks using a model based on genetic algorithm and ensemble learning	The proposed weighted voting algorithm based on a genetic algorithm was better than other similar methods.	Genetic algorithm and ensemble learning	Genetic Algorithms
Rezaeenoor et al. (2019)	Prediction of Cardiovascular Diseases Using an Optimized Artificial Neural	The neural network with five middle layer neurons exhibited superior accuracy, predicting heart attack patients with 97.7%	Multilayer Perceptron Artificial Neural Network with Error Backpropagation Algorithm combined	Genetic Algorithms

	Network	accuracy.	with Genetic Algorithm.	
Veisi et al. (2021)	Improving the efficiency of machine learning algorithms in heart disease diagnosis by optimizing data and features	The SVM algorithm outperformed all others in terms of accuracy, with a 92.9% accuracy rate, while the multilayer perceptron neural network had the highest accuracy rate at 94.6%.	Decision tree, random forest, SVM, and XGBoost	Machine Learning
Nazari and Jodki (2020)	Using genetic algorithm and K-means clustering to improve the accuracy of support vector machine in the diagnosis of heart disease	The proposed method employed a genetic algorithm and a K-means algorithm for feature selection, achieving an accuracy rate of nearly 77%.	Genetic algorithm and K-means clustering	Genetic Algorithms
Shah et al. (2020)	Heart disease prediction using machine learning techniques	K-NN had the best performance in predicting heart disease with an accuracy of 90.78%.	Neural networks, KNN, decision tree, Naive Bayes, logistic regression, random forest, support vector machine	Machine Learning
Rajdhan et al. (2020)	Prediction of heart disease using machine learning	The random forest algorithm had the best performance in predicting heart disease, with an accuracy of 90.16%.	Random Forest, Support Vector Machine, Artificial Neural Networks, Decision Tree, and Logistic Regression	Machine Learning
Ali et al. (2021)	Heart disease prediction using supervised machine learning algorithms: Performance analysis and comparison.	The random forest algorithm achieved 100% accuracy in predicting heart disease.	Random Forest, KNN, Decision Tree	Machine Learning
Yadav et al. (2021)	Feature optimization-based heart disease prediction using machine learning.	The naive Bayes algorithm obtained 96% accuracy after optimizing features using a genetic algorithm.	Naive Bayes, Genetic Algorithm, SVM, KNN, Random Forest	Genetic Algorithms
Pandiaraj et al. (2021)	Effective heart disease prediction using hybrid machine learning.	The proposed hybrid model using SVM and genetic algorithm performed better than other methods in predicting heart disease.	SVM, genetic algorithm	Hybrid Approaches
Verma et al. (2021)	A genetic algorithm-based hybrid deep learning approach for heart disease prediction	The hybrid approach reached 98% accuracy by combining genetic algorithms and deep neural networks.	Genetic algorithms, deep neural networks	Hybrid Approaches
Maleki and Mehrjerdi (2022)	Diagnosis of coronary artery disease by Bat and Harris Hawk meta-heuristic optimization algorithms and machine learning methods	Based on the findings, feature selection using the Harris Hawks optimization algorithm combined with machine learning methods increased the accuracy of the results.	Harris Hawks optimization meta-heuristic algorithm and machine learning method, including decision tree and nearest neighbor	Genetic Algorithms
Tiwari et al.	An ensemble framework	92.34% accuracy was	Hybrid ensemble	Hybrid

(2022)	for predicting cardiovascular diseases	achieved using the combined method of ensemble learning.	learning	Approaches
Arroyo and Delima (2022)	An optimized neural network using a genetic algorithm for cardiovascular disease prediction.	The genetic algorithm improved the prediction accuracy of the neural network by 5.08%.	Artificial neural networks, genetic algorithm	Genetic Algorithms
Paladino et al. (2023)	Evaluating the performance of AutoML tools for heart disease diagnosis and prediction	AutoML tools like AutoGluon and PyCaret performed better, with an accuracy between 78% and 86% compared to conventional methods.	PyCaret, AutoGluon, AutoKeras	AutoML
Koshiga et al. (2023)	Prediction of heart disease based on machine learning algorithms	The model based on voting among algorithms had the highest accuracy of 98.36%.	Logistic regression, decision tree, random forest, support vector machine, voting algorithms	Hybrid Approaches
Bhatt et al. (2023)	Effective heart disease prediction using machine learning techniques	Bio-inspired algorithms performed better in predicting and accurately diagnosing heart diseases.	Bio-inspired algorithms, genetic algorithm	Genetic Algorithms
Yu (2023)	Analysis and prediction of heart disease based on machine learning algorithms.	Machine learning algorithms are widely used in improving the accuracy of heart disease prediction and helping reduce treatment costs.	Random forest, decision tree, support vector machine	Machine Learning
Akkur (2023)	Prediction of cardiovascular diseases based on voting ensemble model and SHAP analysis.	The proposed model based on cumulative voting showed the best performance in predicting heart disease with an accuracy of 93.34%.	Cumulative voting model, SHAP analysis	Hybrid Approaches
Wang et al. (2024)	Explainable coronary artery disease prediction model based on AutoGluon from AutoML framework	The AutoML model with appropriate data preprocessing performed between 87.41% and 92.3% in diagnosing heart disease.	Combining 13 base models with AutoML and using data normalization to improve performance	AutoML
Deepan et al. (2024)	FLAML-HDPS model: An efficient and intelligent AutoML approach for heart disease prediction	FLAML significantly improves heart disease prediction models, enhancing accuracy, precision, recall, and ROC-AUC, saving doctors and nurses valuable time and enabling more effective risk assessments and treatment.	Fast and Lightweight AutoML (FLAML)	AutoML
Reddy et al. (2024)	Optimizing heart disease prediction through ensemble and hybrid machine learning techniques	A 95.8% accuracy was achieved using advanced feature selection and hybrid ensemble methods.	Stacking classifier with Random Forest, Multi-Layer Perceptron, XGBoost, LightGBM, and Logistic Regression as metamodel.	Hybrid Approaches

Dorraki et al. (2024)	Improving cardiovascular disease prediction with machine learning using mental health data	Accuracy improved from 71.31% (traditional factors only) to 85.13% by including psychological factors.	Ensemble ML model: Decision Tree, Random Forest, XGBoost, Support Vector Machine, and Deep Neural Networks (DNNs).	Machine Learning
Jha et al. (2024)	An automated machine learning approach for detecting chronic ischemic heart disease	The diagnostic accuracy of chronic ischemic heart disease improved using the TPOT AutoML tool.	Tree-based Pipeline Optimization Tool (TPOT)	AutoML

Most of the earlier works on heart disease prediction focused on using different machine-learning techniques and optimization algorithms to further improve the accuracy of the test. The literature review indicated neural networks, random forests, and genetic algorithms among the most frequent and effective models introduced. Shah et al. (2020) illustrated that K-NN showed the best performance among others in predicting heart diseases, with an accuracy of 90.78%. In contrast, Ali et al. (2021) reported that the random forest algorithm had 100% accuracy in predicting heart diseases. Also, Yadav et al. (2021) stated that the Naive Bayes algorithm obtained an accuracy of 96% after optimizing the features using the genetic algorithm. Other research works have also handled hybrid models. Pandiaraj et al. (2021) employed a hybrid model using SVM and a genetic algorithm, which performed better than other methods. Verma et al. (2021) obtained 98% accuracy by combining genetic algorithms and deep neural networks. Further, Verma et al. (2021) showed a better performance of the genetic algorithm than BAT and BEE algorithms in predicting heart disease. Furthermore, Arroyo and Delima (2022) reported that the genetic algorithm improved the prediction accuracy of the neural network by 5.08%. Recently, Koshiga et al. (2023) achieved an accuracy of 98.36% using voting between algorithms.

However, previous research shows several gaps, which this study has sought to address. First, the insufficient use of hybrid models is one of the weaknesses of past studies, highlighting the need for more optimal combinations of machine learning and optimization algorithms. On the other hand, one of the main factors that can enhance accuracy is effective feature selection, which has yet to receive adequate attention. Although studies like Yadav et al. (2021) and Nazari and Jodki (2020) demonstrated that feature selection improves accuracy, this aspect remains largely underexplored in many other research works. Third, more recent and advanced algorithms, such as the Harris Hawks optimization algorithm introduced by Maleki and Mehrjerdi (2022), have yet to be thoroughly evaluated, compared, and explored in depth, warranting further investigation. Considering these gaps, the necessity of conducting the present research is clear. This research effort aims to come up with an enhanced heart disease prediction model by incorporating AutoML along with the genetic algorithm. Thus, the optimization of parameters and automatic feature selection will help in the diagnosis systems for heart disease, performing better than earlier research attempts.

The recent trend is to show several studies using AutoML frameworks like AutoGluon, PyCaret, and TPOT on top of state-of-the-art results in a wide variety of applications that also include heart disease and other medical diagnoses. For instance, in a recent paper, Paladino et al. (2023) studied the application of several AutoML tools for heart disease diagnosis using pre-existing feature sets only, without considering even advanced algorithms such as GAs for the optimization. On the other hand, Wang et al. (2024) proposed AutoML models with preprocessing techniques but without the integration of genetic algorithms in feature selection or parameter adjustment. While AutoML frameworks facilitate model development, studies of this nature prove that complementary optimization methods may help enhance their performance. On the other hand, hybrid model research, such as the work of Reddy et al. (2024) achieved the integration of ML techniques with ensemble learning but failed to automate the pipeline. In contrast, the present study bridges the gap by incorporating AutoML with Gas while also implementing model selection, feature selection, and automatic hyperparameter tuning within a scalable framework. It addresses the challenges of achieving higher accuracy, reducing manual effort, and maintaining cost-effectiveness, making it an ideal solution for resource-constrained healthcare systems.

Methodology

This study aims to systematically establish and optimize the machine learning model of heart disease prediction based on AutoML integrated with a genetic algorithm. This research starts with data collection and preprocessing for a comprehensive heart disease dataset to ensure the quality of the data and the relevance of each feature. The TPOT framework provides automation in the process of designing machine learning pipelines for optimization in model selection, feature engineering, and hyperparameter tuning through evolutionary algorithms. It performs comparative analysis in the form of benchmarking TPOT versus traditional and ensemble models like Logistic Regression, Decision Tree, SVM, XGBoost, Extra Trees, Naive Bayes, and KNN. In the evaluation of each model, key metrics such as accuracy, precision, recall, F1 Score, and specificity are employed to build a strong and valid prediction system.

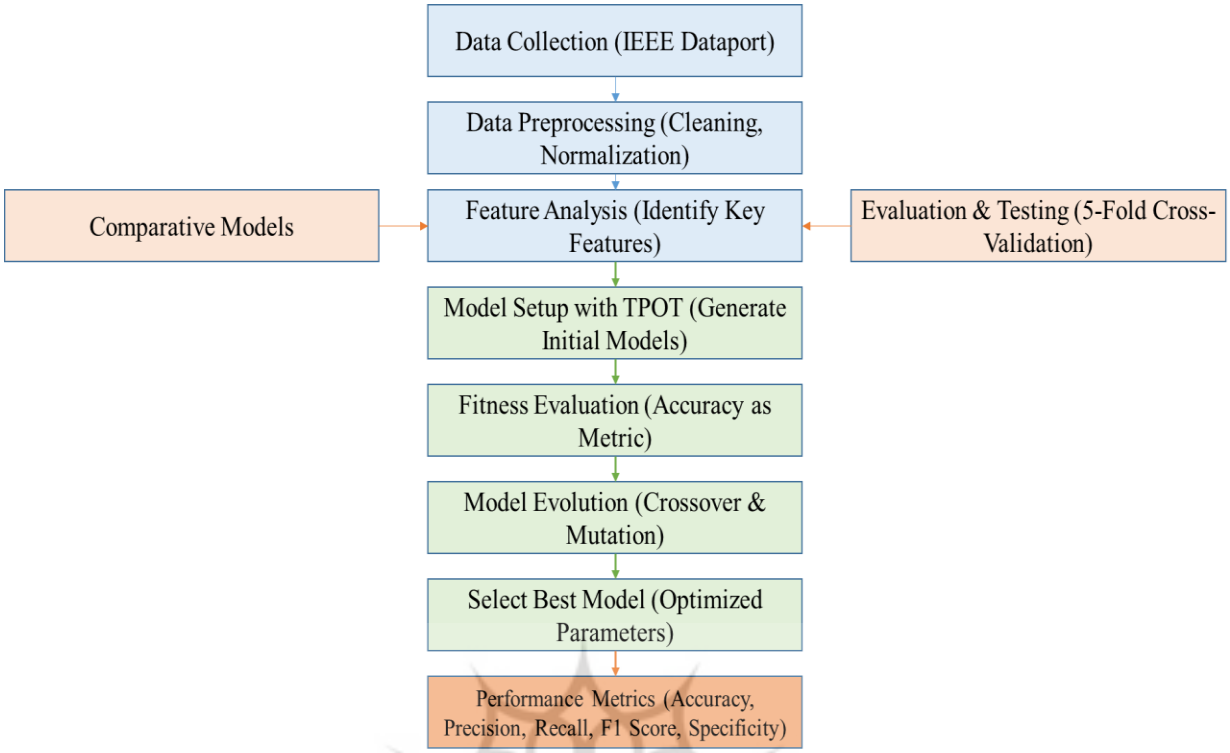

Figure 1. Steps of the methodology

Data collection

Using a valid and standard heart disease dataset, extracted via the IEEE Dataport platform, this research combines five famous and independent datasets from sources, including the Cleveland, Hungary, Switzerland, Long Beach, VA, and Statlog (Heart) Data Set. This dataset is compiled with 1190 records and 12 common features and is the biggest one that exists in the heart disease area (Wang et al., 2024). This dataset is located at https://ieee-dataport.org/open-access/heart-disease-dataset-comprehensive (Siddhartha, 2020). This study is realized using Python and relies on libraries like NumPy for numerical calculations, Pandas for data preprocessing, Matplotlib and Seaborn for visualizations, Scikit-learn for the development of machine learning models, and finally, TPOT for automated machine learning with genetic algorithms.

Dataset features

The dataset contains 12 features, each directly or indirectly affecting the prediction of heart disease. Table 2 provides a complete description of each feature:

Table 2. Description of Features in Heart Disease Dataset

Feature	Feature Description	Values
Age	Age of the patient	Numerical
Sex	Gender of the patient	0 = female, 1 = male
ChestPainType	Type of chest pain	1: stress-related pain, 2: unrelated pain, 3: non-cardiac pain, 4: no pain
RestingBpS	Blood pressure at rest	Numerical
Cholesterol	Blood cholesterol levels	Numerical
FastingBloodSugar	Fasting blood sugar level	0 = less than 120 mg/dL, 1 = greater than 120 mg
RestingECG	ECG results at rest	0 = normal, 1 = abnormal, 2 = left ventricular shrinkage
MaxHeartRate	Maximum heart rate	Numerical
ExerciseAngina	Chest pain during physical activity	0 = absence of pain, 1 = presence of pain
OldPeak ST	The amount of ST wave changes	Numerical
STSlope	The slope of ST wave changes	1 = bullish, 2 = flat, 3 = bearish
Target	Heart disease status	0 = no disease, 1 = presence of heart disease

Each of the features in this dataset is directly or indirectly related to heart disease risk factors (Darrab et al., 2024). Below is an explanation of the importance of each feature:

Table 3. Explanations of Features in Heart Disease Dataset

Feature	Explanation
Age	Age is a key factor in the occurrence of heart diseases; with age, vessel damage, and blood pressure increase.
Gender	Men are more at risk of heart disease than women up to a certain age.
Type of Chest Pain	Different types of chest pain can indicate heart disease; pain from activity or stress is often a warning sign.
Blood Pressure at Rest	High blood pressure is a major risk factor for cardiovascular diseases.
Cholesterol	Cholesterol is a crucial clinical parameter; abnormal levels (high or low) can indicate heart disease risk.
Fasting Blood Sugar	Elevated fasting blood sugar may indicate diabetes, which is directly related to heart disease.
Resting EKG	Abnormal EKG results can indicate heart failure or other heart-related issues.
Maximum Heart Rate	A higher-than-normal heart rate during physical activity may indicate heart function issues.
Pain During Exercise	Chest pain during exercise suggests potential blood flow problems to the heart.
ST Wave Changes	Changes in the ST wave may be signs of coronary artery blockage.
Slope of ST Wave Changes	The slope type of the ST wave provides detailed information about heart function.
Feature	Explanation

The target characteristic in this research is having or not having heart disease, which is predicted using 11 other features. Among the 1190 records in this research, 629 have heart disease (positive record), and 561 do not have heart disease (negative record).

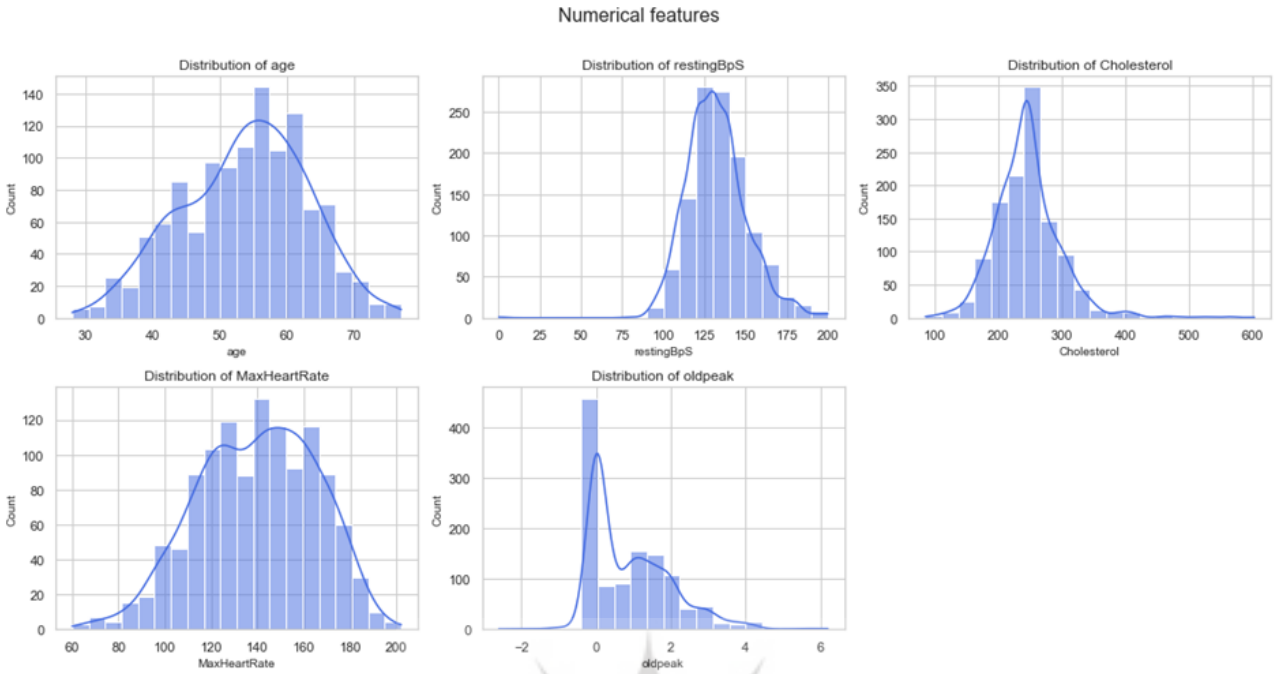

Figure 2. Distribution of Key Numerical Features in Cardiovascular Data

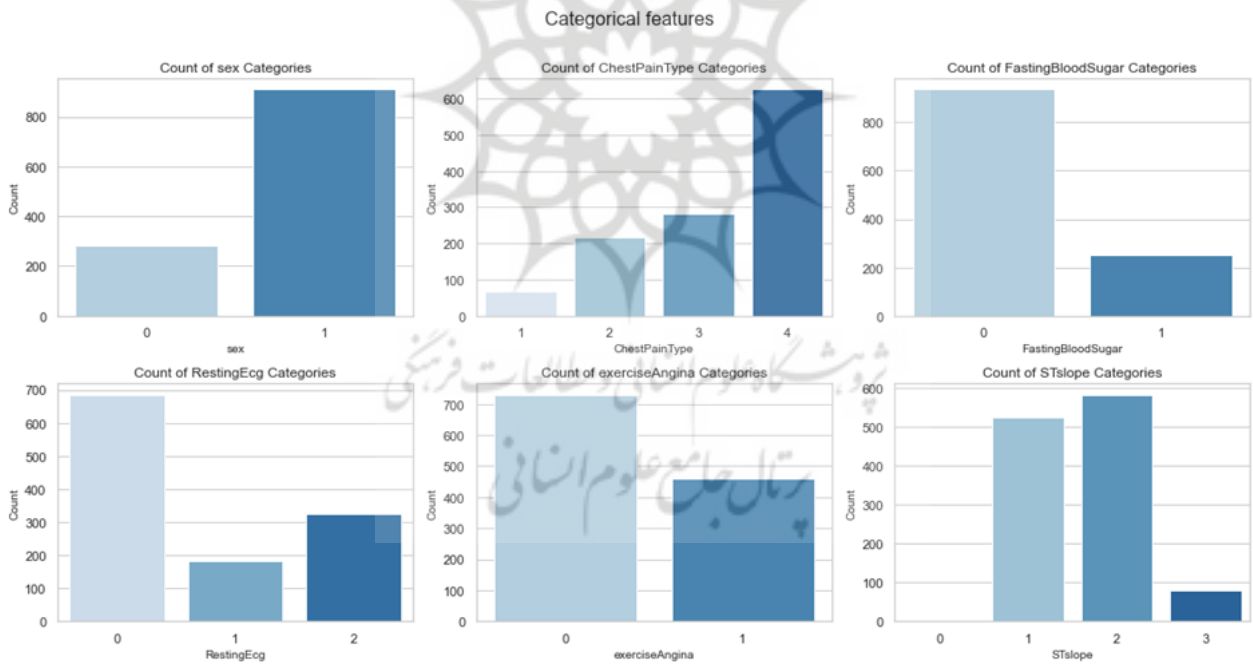

Figure 3. Distribution of Categorical Features in Cardiovascular Data

Based on the analysis of graphs depicting numerical and categorized characteristics, the following results can be inferred about the dataset (Akkur, 2023):

- This dataset includes five numerical and six categorical features.
- Numerical features in this dataset have a relatively normal distribution.
- The highest frequency among people with heart disease is related to men.

- The most common type of chest pain is asymptomatic or stress-free pain.
- Most people's fasting blood sugar level is below 120 mg/dL.
- ECG values (electrical activity of the heart) are within normal range.
- Most people do not have angina (chest pain caused by heart disease).
- Most people have a flat ST slope.

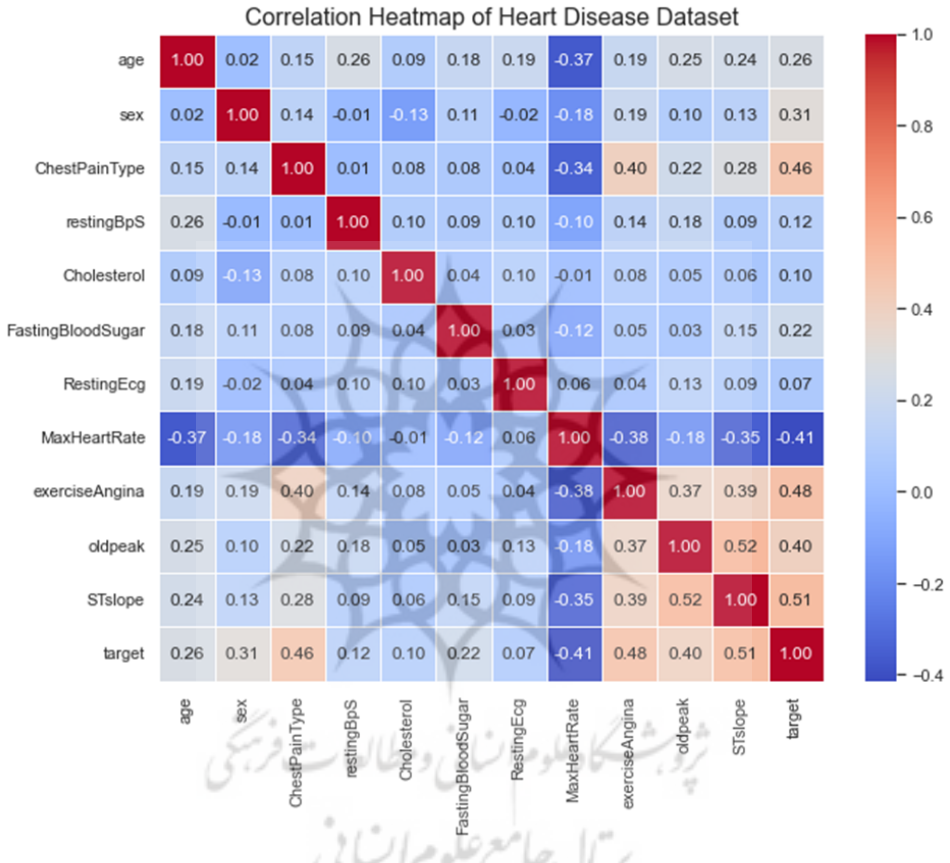

Figure 4. Correlation Heatmap of Features in the Heart Disease Dataset

Figure 4 depicts the correlation of various dataset features related to heart disease. In the analysis, the correlation coefficient ranges between -1 and 1 in numerical values, where 1 is for complete positive, -1 is for complete negative, and 0 represents no correlation (Khani et al., 2022). All the features in this data set are somehow correlated with heart disease (target variable) and other features, but the degree of these correlations is different. Some characteristics, such as exercise angina, ST slope, and chest pain type, show a higher correlation with heart disease and act as key factors in identifying heart patients. Specifically, a correlation of 0.48 between exercise angina and heart disease, a correlation of 0.51 for the ST slope, and a correlation of 0.46 for the type of chest pain indicate that these characteristics contribute significantly to diagnosing people with heart disease. Characteristics such as cholesterol and resting blood pressure are also associated with heart disease, but their

correlation is relatively weaker. For example, the correlation of cholesterol with heart disease is 0.10, which indicates that although this characteristic is associated with heart disease, it alone cannot be a strong indicator for diagnosis. Even features such as maximum heart rate and ST depression, which are inversely related to heart disease, have provided useful information about the condition of the patients. The inverse relationship between the maximum heart rate and heart disease is at a rate of -0.41, indicating that with a drop in heart rate, the possibility of the disease development increases, as evidenced among heart patients. Generally speaking, some features have a stronger association with heart diseases than others, but all are somehow related and can be useful in a diagnosis or a model of heart diseases.

Data preprocessing

The entire data collected for this study were obtained from the Comprehensive Heart Disease Dataset data set, which is standard and valid. This dataset was derived from other datasets on heart diseases and is pre-processed with quality and completeness assessment done. The final verifications of the operations executed throughout the study were conducted in this part of the research. Other quality control measures included checking for clean data and ensuring none or fewer missing or abnormal values. For the missing numerical values, the average of the available values was adopted to ensure that the data were all complete and right for analysis. Additionally, an outlier test was conducted on the entire features, confirming that there were no outliers in the given dataset. It also helps avoid instances where a particular model is influenced by some outlier data.

Python has been employed in this study in the various project phases. MinMaxScaler has also been used to apply additional feature scaling in all models. MinMaxScaler is preferred because specific models, such as SVM and KNN, are influenced by the scales of the data since their performance is based on distances between the data points and decision boundaries. MinMaxScaler fits all variables within a defined range (0-1), ensuring that even if responses are of a much broader scale than predictors or vice versa, the model's performance will not be influenced. The prediction results demonstrate that standardization of the data by MinMaxScaler would go along way in enhancing the accuracy and effectiveness of the models to perform better. The general equation of MinMaxScaler is as follows:

$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

In this formula:

xscaled is the normalized (standardized) value of the property.

x is the original value of the attribute.

xmin is the smallest feature value in the data.

xmax is the largest feature value in the data.

This formula sets the values of each feature within the range of 0 to 1, ensuring that the features are placed on the same scale (Akkur, 2023).

Data Preprocessing Steps

- Handling Missing Values: Missing numerical values have been filled with the mean value to ensure the dataset's completeness for analysis.
- Outlier Detection: An outlier analysis was conducted, confirming that no significant outliers were present to influence the results or compromise the analysis.
- Feature Scaling: Scaled numerical features in the range of [0, 1] were chosen using MinMaxScaler because:

1. It ensures that all features are on the same scale, which becomes crucial for models sensitive to feature scaling.
2. Consistency: MinMaxScaler has been consistently applied by other researchers using this dataset, enabling a fair and reliable comparison of results across various studies.

- Data Splitting: The dataset is split into a training set (70% of the total) and a test set (30%) using a random split with random_state = 42 for reproducibility. This step contributes critically to assessing the model's performance on unseen data.

Introducing TPOT and how to use it

TPOT is an automated machine learning tool using genetic algorithms to optimize machine learning pipelines. The most crucial goal of TPOT is to automatically find the best model and data preprocessing settings without manual interference. TPOT operates by using evolutionary algorithms to explore the space of machine learning models and their respective settings before evaluating each model using cross-validation and returning the best model (Orlenko et al., 2020). Instead of running only one simple model, TPOT optimizes a series of models and preprocessing as a pipeline. A pipeline can consist of several steps, including feature selection, data normalization, model selection, and hyperparameter tuning. TPOT works as an evolutionary algorithm, using the concepts of population and generations. TPOT optimizes in each generation using mutation and combination operations and keeps the best models (Gijsbers et al., 2017).

Main stages of TPOT

Initial population generation: First, TPOT generates a set of pipelines, including different machine learning models and preprocessing settings as the initial population. These models include various techniques such as logistic regression, decision trees, random forests, and other famous models (Olson & Moore, 2019).

- $P_0=\{M_1,M_2,\ldots,M_n\}$

Where P0 is the initial population, and Mi indicates the models in the initial population.

Evaluation of models (Fitness Function): In each generation, TPOT evaluates the performance of each model using a fitness function. Here, accuracy is usually used as the main evaluation criterion, but other criteria can also be used.

- $F_i(M_i)=\text{Accuracy}(M_i,X_{train},Y_{train})$

Where Fi is the fitness function for the model, and Xtrain and Ytrain are the training data (Kenny et al., 2023).

Selection of Top Models (Selection): TPOT selects the best-performing models from the current population. This selection is done using techniques such as roulette wheel selection or elitism selection.

- $S_i=\text{Select}(F_i(M_i))$

Where Si indicates the selected models in generation i (Olson & Moore, 2019).

Combination (Crossover): TPOT combines selected models to generate new models. This process is similar to combining the genes of two parents to produce a child. Blending involves combining components from one model with those of another.

- $M_{new}=\text{Crossover}(M_i,M_j)$

Where the new model is produced from Mi and Mj models (Gijsbers et al., 2017).

Mutation: In this step, TPOT creates new models by applying random changes to some models. These changes include changing hyperparameters or adding or removing preprocessing steps.

- $M_{mutated}=\text{Mutate}(M_i)$

The evaluation, selection, combination, and mutation steps are repeated for a specified number of generations (for example, 100 generations). In each generation, the models are improved, and TPOT preserves the best available models (Olson & Moore, 2019).

Final Model (Best Model): After completing the number of generations, TPOT selects the best model with the best performance based on the fitness function and proposes it as the final model.

- $M_{best}=Best(F_i(M_i))$

TPOT uses a combination of population-based optimization and cross-validation. For each pipeline, TPOT uses cross-validation to assess model accuracy. Specifically, each Mi pipeline is evaluated using k-partition cross-validation.

The fitness function for each model is defined as follows:

$$F(M_i) = \sum_{j=1}^{k} \frac{1}{k} Accuracy(M_i, X_{train}^{(j)}, y_{train}^{(j)}) \quad (2)$$

Where k is the number of validation sections.

In this study, the following settings were used for TPOT:

Number of generations: 100 generations

Population Size: 50 generations

Cross-validation: Cross-validation with 5 sections was used to evaluate the models.

Comparative models

In addition to the selected TPOT model, seven other models were selected for comparison:

- Logistic Regression: It is a machine learning algorithm used for binary classification, seeking to predict the occurrence or non-occurrence of an event (0 or 1). It assumes a linear relation between the input features and the output, after which it passes through the sigmoid function to map the values in a range of 0 to 1. This approach is rather simple yet very effective in predicting heart disease outcomes through probability assessment based on various features (Anshori & Haris, 2022).

- Decision Tree: A Decision Tree model divides the data into subsets based on the value of the input features. It organizes the splits into a tree-like structure, meaning that every internal node represents a feature, and every leaf node represents the final classification decision. Thus, this model is also very interpretable since the decision-making process is visually shown, highlighting its leading role in both classification problems and heart disease prediction (Liu et al., 2023).

- Support Vector Machine (SVM): SVM is a binary classification algorithm that identifies the best decision boundary or hyperplane to separate data points belonging to two classes. The algorithm, in effect, maximizes the margin, which is the distance between the hyperplane and the closest data points of each class. SVM is particularly effective in the case of high-dimensional data and is known for its robustness in classification tasks, which includes medical data analysis (Wang, 2023).

- XGBoost: XGBoost is the abbreviation of Extreme Gradient Boosting. One of the most robust ensemble learning techniques based on boosting involves constructing multiple weak learners sequentially, with each learner—often a decision tree—addressing and correcting the errors of its predecessor. This algorithm is all too efficient and scalable for structured data problems. XGBoost is applied to the prediction of heart diseases because of its superior performance in handling complicated data structures and the minimization of error (Jafarnejad Chaghoshi et al., 2024).

- Extra Trees: Extremely Randomized Trees, better known as Extra Trees, is an ensemble learning method that builds a large number of decision trees in a much more randomized manner than Random Forest. It selects features and thresholds for random data splitting, increasing diversity among the trees and often leading to improved generalization performance. The model is in high favor because of its speed and efficiency in handling large datasets (Geurts et al., 2006).

- Naive Bayes: A Naive Bayes classifier is based on the probabilistic approach of Bayes' theorem and assumes mutual independence among features. This is surprisingly good despite the simplicity in some data sets and is particularly effective when the features are categorical or there is a clear separation between classes. In heart disease prediction, Naive Bayes provides an extremely straightforward approach to dealing with noisy data and still results in reasonable performance (Yadav et al., 2021).

- K-Nearest Neighbors (KNN): KNN is a non-parametric algorithm that classifies a data point based on the majority votes of its k-nearest neighbors. It first calculates the distance between data points and then assigns the most frequent class among its neighbors to a test point. KNN is quite simple and intuitive but might get computationally expensive for large datasets. It is used in heart disease prediction to identify patterns based on the proximity of data points (Ali et al., 2021).

Different models were evaluated using the following indicators:

- Confusion Matrix: The confusion matrix is a table that summarizes the performance of a classification model (Amin & Mahmoud, 2022).

- Accuracy: The proportion of correctly classified samples out of the total samples (Bumm et al., 2023).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

Table 4. Confusion matrix

		Predicted class	
		Positive	Negative
Real class	positive	TP (True Positive): Number of samples .correctly classified as positive	FN (False Negative): Number of samples .incorrectly classified as negative
	negative	FP (False Positive): Number of samples .incorrectly classified as positive	TN (True Negative): Number of samples .correctly classified as negative

- Precision: The ratio of true positive predictions to the total number of positive predictions, measuring the model's ability to avoid false positives (Villmann et al., 2014).

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

- Recall (Sensitivity): The ratio of true positive predictions to the total number of actual positives, revealing the model's ability to detect all positive samples (Villmann et al., 2014).

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

- F1 Score: The harmonic mean of precision and recall, balancing both metrics when there is an uneven class distribution (Takahashi et al., 2021).

$$F1\ Score = \frac{Precision \times Recall}{Precision+Recall} \times 2 \quad (6)$$

- Specificity: The ratio of true negative predictions to the total number of actual negatives, reflecting the model's ability to identify negative samples (Gonzalez-Abril et al., 2017).

$$Specificity = \frac{TN}{TN+FP} \quad (7)$$

Results

This section investigates the results of different machine learning models employed to predict heart disease. All models were run on a system with an Intel Core i5-7200U processor, 8GB of RAM, and Python 3.12. The models used are TPOT (Gradient Boosting), Logistic Regression, Decision Tree, SVM, XGBoost, Extra Trees, Naive Bayes, and KNN. Also, all the parameters of the proposed model have been meticulously examined by TPOT.

- The model proposed by TPOT (Gradient Boosting)

After reviewing and optimizing the models, TPOT presented the Gradient Boosting model as the best with the following optimal parameters (Table 5):

Table 5. Hyperparameters for the Machine Learning Model

parameter	The value
Learning Rate	0.1
Max Depth	9
Max Features	0.1
Min Samples Leaf	2
Min Samples Split	10
Number of Estimators	100
Subsample	0.6000000000000001

This model is a robust data classification method that improves predictive power by sequentially combining multiple decision trees. Each tree tries to reduce the error of the previous trees, which leads to an increase in the model's accuracy. The above parameters control the complexity of the model, dictate how features are combined, and determine the number of decision trees that TPOT automatically optimizes.

- Confusion Matrix

Table 6 shows the results of the Confusion Matrix for each of the models on the test data:

Table 6. Confusion Matrix Results for Different Models

Model	TP	FN	FP	TN
TPOT (Gradient Boosting)	194	9	13	141
Logistic Regression	179	24	28	126
Decision Tree	179	24	23	131
SVM	183	20	23	131
XGBoost	189	14	18	136
Extra Trees	190	13	14	140
Naive Bayes	179	24	27	127
K-Nearest Neighbors	182	21	27	127

The Confusion Matrix table presents the performance of eight machine learning models in classifying positive and negative examples regarding the prediction of heart disease. The best performance for identifying people with heart disease is shown by TPOT (Gradient Boosting), classifying 194 samples as True Positive and presenting only 9 False Negative (FN) samples, meaning that TPOT rarely misclassifies a real patient as healthy. Moreover, this method has 141 True Negative (TN) samples and only 13 False Positives (FP), which means it is very accurate in correctly classifying healthy individuals (low false diagnoses). Furthermore, XGBoost and Extra Trees show good results, with the former indicating 189 True Positives and only 14 False Negatives, with 136 True Negatives and 18 False Positives. On the other hand, Extra Trees has a slightly improved performance, with 190 True Positives and only 13 False Negatives, while indicating 140 True Negatives and 14 False Positives. This method is among the best models in terms of accuracy and reduction of errors. The SVM model performs better than both Logistic Regression and Decision Tree, indicating 183 True Positives and 20 False Negatives. However, this is still much behind the model performance of TPOT, XGBoost, and Extra Trees, as it has 131 True Negatives and 23 False Positives. The

Decision Tree and Logistic Regression models—the former with 179 True Positives—contain more errors compared to the best models.

In particular, Logistic Regression shows 28 False Positives, meaning it would rather classify a healthy individual as a patient. The Decision Tree performs slightly better, with 23 False Positives, but still falls short compared to the performance of the ensemble models. The Naive Bayes model has roughly the same number of True Positives as Logistic Regression and Decision Tree at 179, but with 24 False Negatives and 27 False Positives, it shows less overall effectiveness. The KNN model, with 182 True Positives and 21 False Negatives, performs slightly better than Naive Bayes but still has 27 False Positives, indicating some room for improvement in identifying healthy cases. Overall, TPOT (Gradient Boosting), XGBoost, and Extra Trees showed better performances in terms of accuracy and reduction of errors, correctly classifying both groups—healthy and sick—with the lowest error rates. In particular, TPOT resulted in the most reliable model among the eight tested by drastically reducing the number of False Negatives and False Positives.

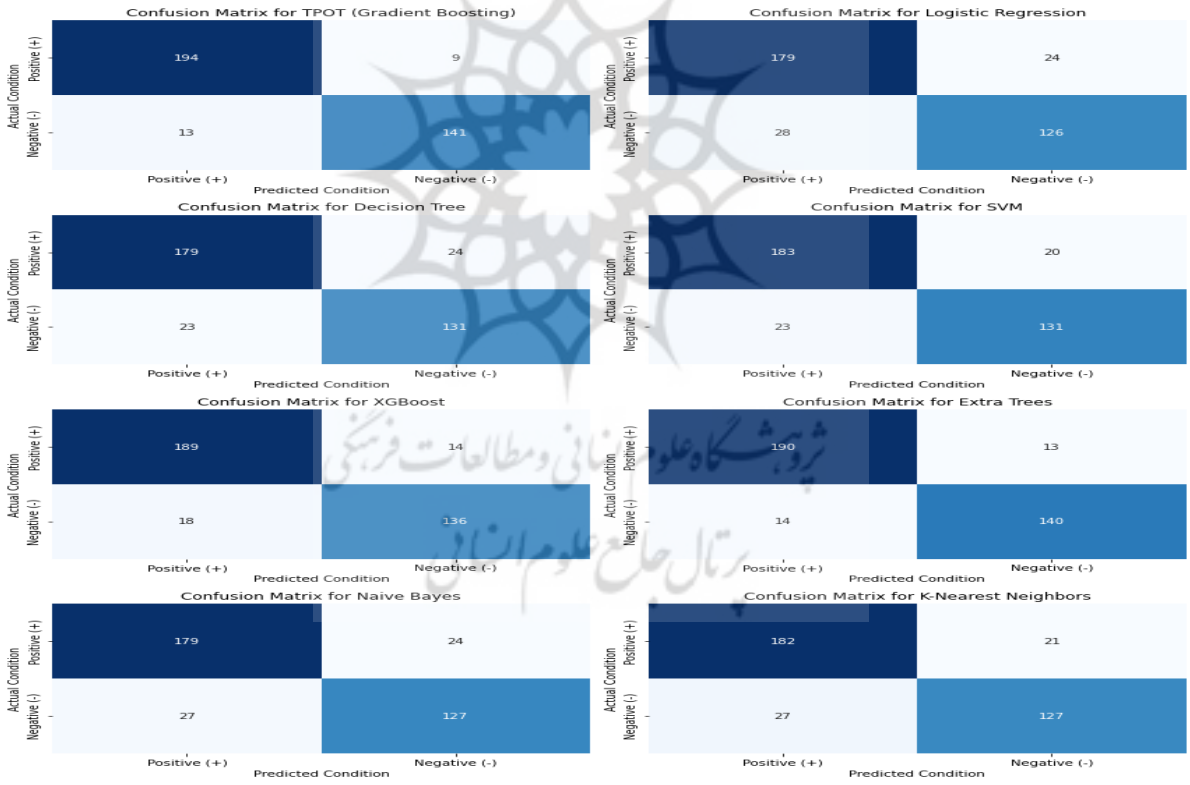

Figure 5. Visualization of Confusion Matrices for Various Machine Learning Models

Figure 5 compares TPOT-Gradient Boosting, Logistic Regression, Decision Tree, SVM, XGBoost, Extra Trees, Naive Bayes, and KNN. TPOT shows the highest accuracy due to Gradient Boosting, indicating the minimum number of errors in classifying an individual as sick or healthy. Second in providing the best performance will be those of XGBoost and Extra Trees, which also tend to minimize errors. For SVM, the performance is average, while in

Logistic Regression and Decision Tree, the misclassification rates are slightly higher, presenting more challenges in separating classes. Naive Bayes and KNN hold even more errors, underpinning the limits of predictability compared to the ensemble models. Overall, TPOT, together with XGBoost and Extra Trees, can be ranked as the most reliable in heart disease classification.

- Evaluation criteria of models

Table 7 shows the results of Accuracy, Precision, Recall, F1 Score, and Specificity criteria for each model:

Table 7. Performance Metrics Comparison of Classification Models

Model	Specificity	F1 Score	Recall	Precision	Accuracy
TPOT (Gradient Boosting)	0.915584416	0.946341463	0.955665025	0.937198068	0.93837535
Logistic Regression	0.818181818	0.873170732	0.881773399	0.8647343	0.854341737
Decision Tree	0.850649351	0.883950617	0.881773399	0.886138614	0.868347339
SVM	0.850649351	0.894865526	0.901477833	0.888349515	0.879551821
XGBoost	0.883116883	0.921951220	0.931034483	0.913043478	0.910364146
Extra Trees	0.909090909	0.933660934	0.935960591	0.931372549	0.924369748
Naive Bayes	0.824675325	0.875305623	0.881773399	0.868932039	0.857142857
K-Nearest Neighbors	0.824675325	0.883495146	0.896551724	0.870813397	0.865546218

Each model demonstrates distinct strengths and weaknesses in predicting heart disease, and the following analysis provides valuable insights into their comparative performance.

Accuracy: It is the percentage of samples correctly classified overall. TPOT stands for Gradient Boosting and has the highest accuracy, with a score of 0.938, correctly classifying approximately 93.8% of all the samples. Also, Extra Trees and XGBoost have good accuracies (0.924 and 0.910, respectively), while SVM has an accuracy of 0.880, and Decision Tree and K-Nearest Neighbors have moderate accuracies of 0.868 and 0.866, respectively. Baseline accuracy is 0.857 for Naive Bayes and 0.854 for Logistic Regression.

Precision: This refers to the model's ability to correctly identify positive samples. TPOT excels in this aspect with a precision score of 0.937, indicating exceptionally low error in identifying positive cases. It is closely followed by Extra Trees and XGBoost, which achieve respective precision scores of 0.931 and 0.913. SVM also showed great performance with a precision of 0.888, while Decision Tree and KNN had values of 0.886 and 0.871, respectively. Naive Bayes follows with a precision score of 0.869, and Logistic Regression ranks lowest at 0.86.

Recall measures the percentage of positive samples correctly identified. TPOT is far and away the best at Recall (0.956), which correctly identifies nearly all the positive samples. Extra Trees and XGBoost come next in the list, with recalls of 0.936 and 0.931, respectively. Not far behind, SVM has good performance in this respect, with 0.901, and then KNN with

0.897. Logistic Regression and Decision Tree both yield a recall of 0.882, while Naive Bayes falls just slightly below that at 0.882.

F1 Score: The F1 score is the harmonic mean between precision and recall, depicting the model's balance for both of these metrics. In the first position comes TPOT with an F1 score of 0.946, indicating a very good balance. Extra Trees and XGBoost follow not far behind, with F1 scores of 0.934 and 0.922, respectively. Next is SVM, achieving 0.895, followed by Decision Tree and KNN, with respective scores of 0.884 and 0.883. Logistic Regression and Naive Bayes show the lowest F1 scores, with respective values of 0.873 and 0.875.

Specificity represents the ratio of correctly identified negative samples by the model. TPOT leads with the highest specificity (0.916), significantly capable of differentiating patients from healthy individuals. Extra Trees and XGBoost deliver strong performances, with specificities of 0.909 and 0.883, respectively. Support Vector Machine and Decision Tree closely followed, both achieving a specificity of 0.851, while K Nearest Neighbours and Naive Bayes have poor specificities of 0.825. Logistic Regression ranks the lowest in this regard, with a specificity of 0.818.

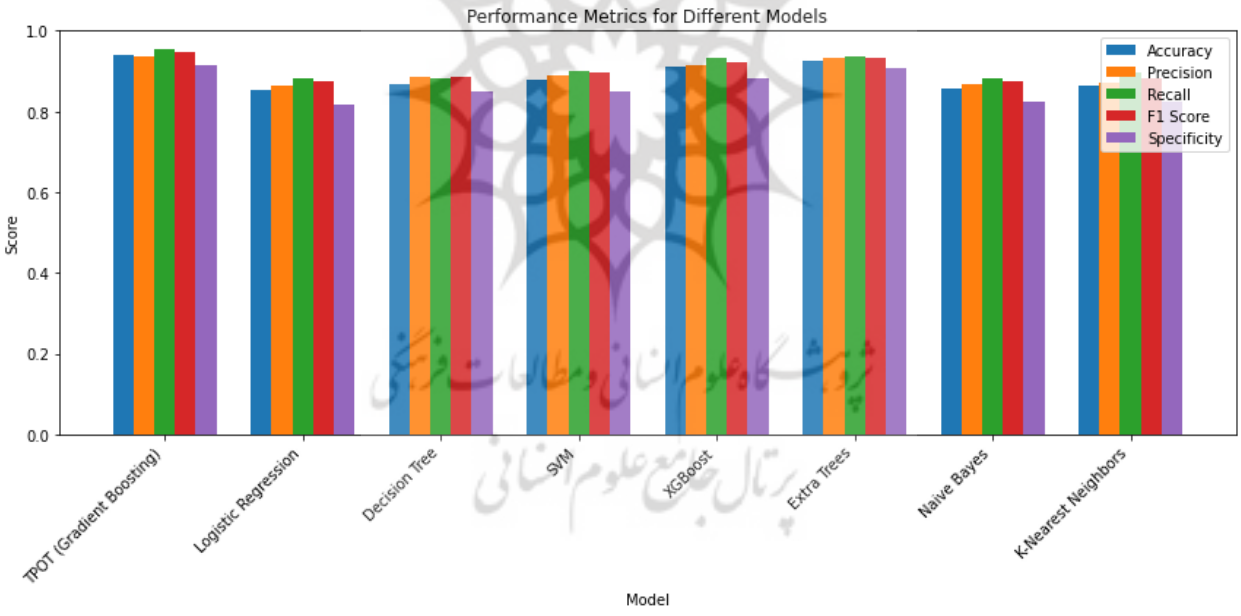

Figure 6. Visualization of Model Performance Across Evaluation Metrics

Figure 6 shows the performance of eight machine learning models, including TPOT-Gradient Boosting, Logistic Regression, Decision Tree, SVM, XGBoost, Extra Trees, Naive Bayes, and K-Nearest Neighbors, regarding the five main criteria of accuracy, precision, recall, F1 score, and specificity. TPOT with gradient boosting is the best, having had an extremely good balance on all measures and being outstanding for accuracy, sensitivity, and specificity. XGBoost and Extra Trees also show very good results, coming quite close to competing with TPOT, hence reliable when it comes to the performance results for all criteria.

Support Vector Machine performs quite well, yielding strong scores for precision and recall, though falling behind TPOT, XGBoost, and Extra Trees. Logistic Regression and Decision Tree show moderate performance since their decent scores are generally low across all metrics compared to the ensemble models. Naive Bayes and KNN do not perform well compared to the top models, especially in the specificity of the models, thereby showing more challenges in correctly predicting negative samples. Therefore, TPOT has emerged as one of the best models in this review by giving well-rounded and accurate predictions, though ensemble models like XGBoost and Extra Trees also have good performances. Although efficient, Logistic Regression, Decision Tree, Naive Bayes, and KNN fall short in both accuracy and reliability compared to the top-performing models.

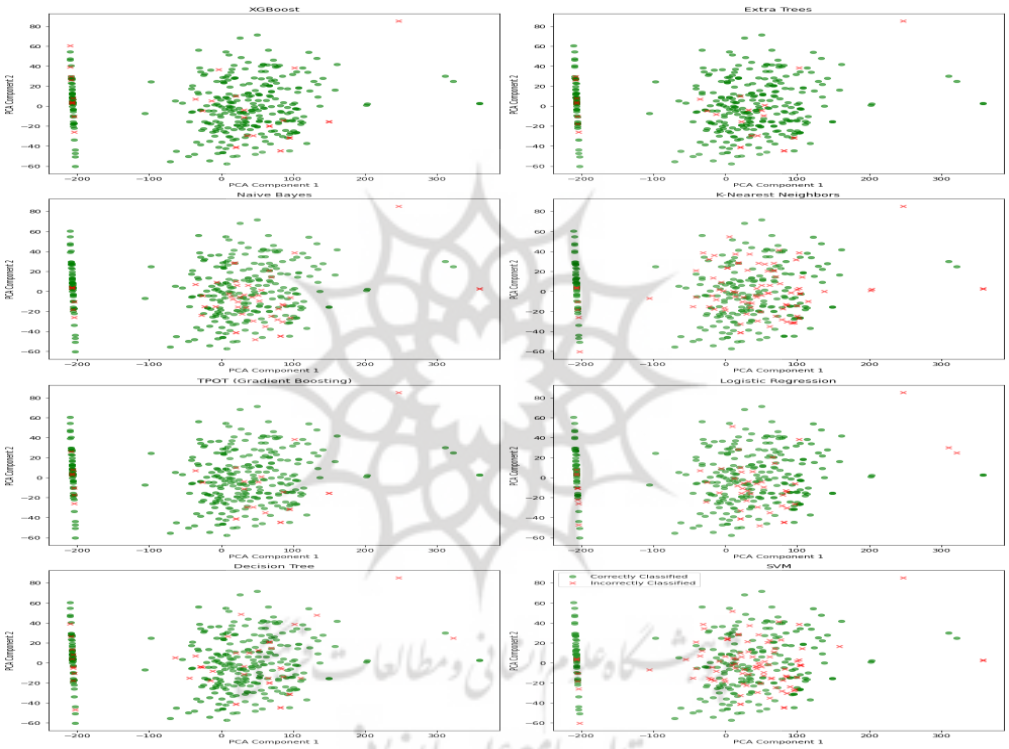

Figure 7. PCA-Based Visualization of Correct and Incorrect Classifications Across Models

Above (Figure 7) are the graphs showing the classification of samples by each of the eight different models: TPOT Gradient Boosting, Logistic Regression, Decision Tree, SVM, XGBoost, Extra Trees, Naive Bayes, and KNN. All these graphs reduce the dimensions of the data to two, using PCA to present the classification of the samples in a two-dimensional space. Correctly classified samples are presented by green dots, while red crosses ("X") show the misclassified ones. TPOT - again using Gradient Boosting - once again emerges as the top-performing method, as evidenced by the noticeably fewer red points indicating misclassified cases compared to other classification methods, indicating a better separation of the boundaries of the positives and negatives. The next best in terms of performance after TPOT would be XGBoost and Extra Trees, considering their relatively fewer misclassifications compared to the simpler models. Meanwhile, the Logistic Regression,

Decision Tree, and SVM models have more misclassified points, indicating that they have less precision in the classification task. The Naive Bayes and K-Nearest Neighbors exhibit higher error rates, particularly in areas that are challenging due to their class boundaries. Overall, TPOT performs better than other models in this comparative visual by yielding higher accuracy and having the least number of misclassifications. Table 8 provides the full comparison of the detailed results against accuracy, precision, recall sensitivity, and F1 score with values from the studies reviewed.

Table 8. Performance Metrics of Heart Disease Classification Models in Literature Using Combined or Independent Datasets from Cleveland, Hungary, Switzerland, Long Beach VA, and Statlog (Heart) Data Sources

Authors	Year	Accuracy	Precision	Recall	F1-Score	The model used
Shah et al	2020	0.9078	-	-	-	Neural networks, KNN, decision tree, Naive Bayes, logistic regression, random forest, support vector machine
Rajdhan et al	2020	0.9016	0.937	0.882	-	Random Forest, Support Vector Machine, Artificial Neural Networks, Decision Tree, and Logistic Regression
Tiwari et al	2022	0.9234	0.92	0.9349	0.9274	Hybrid ensemble learning
Verma et al	2022	0.9000	-	-	-	Genetic algorithm
Arroyo and the dilemma	2022	0.9300	-	-	-	Artificial neural networks, genetic algorithm
Yu	2023	0.9300	-	-	-	Random forest, decision tree, support vector machine
Akkur	2023	0.937	0.9459	0.9355	0.9407	Cumulative voting model, SHAP analysis
Our suggested model	-	0.93837535	0.937198068	0.955665025	0.946341463	AutoML is based on genetic algorithms.

Discussion and conclusion

This study sought to predict heart disease by employing AutoML, specifically using the TPOT model based on the principle of genetic algorithms. AutoML is a new methodology that chooses models and optimizes parameters, drastically reducing the time and effort required for manual setup while improving the accuracy of model performances. Medical research focused on disease diagnosis greatly benefits from AutoML's ability to improve prediction accuracy while reducing the occurrence of errors. TPOT is one of the popular AutoML tools that uses the power of genetic algorithms in model search. The genetic algorithm abstracts principles from natural evolution, with ongoing processes of mutation, selection, crossover, and reproduction to generate and further evaluate models for the most fitting according to predefined criteria. This is done by testing and refining multiple models to find the most effective combinations of features and parameters for accurate predictions.

Indeed, TPOT outperformed others in model selection and parameter tuning in this work, achieving substantial improvements across all metrics, such as accuracy, precision, recall, F1 score, and specificity, compared to manually crafted models. Logistic Regression, Decision Tree, and SVM showed adequate but inferior results compared with TPOT. TPOT allows improvement due not only to the power of gradient boosting algorithms but also to the capability of the genetic algorithm in its power to find the best settings. This underlines the fact that AutoML-based approaches, such as those of TPOT, are effective in medical applications that require high accuracy.

Compared to the identified models in this study, the models used in the previous studies were XGBoost, Extra Trees, Naïve Bayes, and KNN. From the analyses, TPOT has been found to perform better than all the models proposed in other studies, with an accuracy rate of 93.8%. For example, the KNN model proposed by Shah et al. (2020) had an accuracy of 90.78% compared to TPOT. Rajdhan et al. (2020) used Random Forest but attained an accuracy of only 90.16%, which is still lower than TPOT. Traditional models, such as Logistic Regression with an accuracy of 85.43% and Decision Tree with an accuracy of 86.83%, yielded comparatively lower performance, further highlighting the advantages of TPOT's automated optimization in achieving superior results. The other ensemble models, like XGBoost and Extra Trees, also proved to be efficient but were not better than TPOT. Regarding the accuracy, TPOT achieved the highest score at 93.72%, followed by SVM in second place and XGBoost in third. These models demonstrated superior performance, particularly in accurately diagnosing the positive cases. Naive Bayes and KNN were less accurate in terms of precision, indicating various misclassifications. In terms of Recall, TPOT gave a sensitivity of 95.7%, which was higher than all the models, hence minimizing what could be considered as false negatives. SVM and XGBoost achieved notably high recall scores in previous studies,, but TPOT was the most effective in identifying true positives.

The F1 score, indicating the balance of precision and recall, was highest for TPOT at 94.63%, confirming adequate robustness in model evaluation, followed by a low F1 score for SVM at 89.49%. While other ensemble models performed well, such as the Voting Ensemble in Akkur (2023), TPOG had a unique performance due to the automatic optimization of parameters. Overall, the best performance was provided by TPOT or Gradient Boosting with high accuracy and sensitivity, keeping false positives and false negatives as low as possible. These results show the importance of the automatic optimization of parameters in machine learning models. Other advanced AutoML tools that may be applied in future research to solve complex problems with deep neural networks include AutoKeras. In addition, other meta-heuristic methods, including PSO and other evolutionary algorithms, may further improve AutoML results. Ensemble techniques such as stacking and blending may result in better results for prediction not only in heart disease but also for other complex medical conditions.

Conflict of interest

The authors declare no potential conflict of interest regarding the publication of this work. In addition, the ethical issues including plagiarism, informed consent, misconduct, data fabrication and, or falsification, double publication and, or submission, and redundancy have been completely witnessed by the authors.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

Akkur, E. (2023). Prediction of Cardiovascular Disease Based on Voting Ensemble Model and SHAP Analysis. *Sakarya University Journal of Computer and Information Sciences*, *6*(3), 226-238. https://doi.org/https://doi.org/10.35377/saucis.1367326

Ali, M. M., Paul, B. K., Ahmed, K., Bui, F. M., Quinn, J. M., & Moni, M. A. (2021). Heart disease prediction using supervised machine learning algorithms: Performance analysis and comparison. *Computers in Biology and Medicine*, *136*, 104672. https://doi.org/https://doi.org/10.1016/j.compbiomed.2021.104672

Amin, F., & Mahmoud, M. (2022). Confusion matrix in binary classification problems: A step-by-step tutorial. *Journal of Engineering Research*, *6*(5), 1-12. https://doi.org/https://doi.org/10.21608/erjeng.2022.274526

Anshori, M., & Haris, M. S. (2022). Predicting heart disease using logistic regression. *Knowledge Engineering and Data Science (KEDS)*, *5*(2), 188-196. https://doi.org/https://doi.org/10.17977/um018v5i22022p188-196

Arroyo, J. C. T., & Delima, A. J. P. (2022). An optimized neural network using genetic algorithm for cardiovascular disease prediction. *Journal of Advances in Information Technology*, *13*(1), 95-99. https://doi.org/https://doi.org/10.12720/jait.13.1.95-99

Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *The journal of machine learning research*, *13*(1), 281-305.

Bhatt, C. M., Patel, P., Ghetia, T., & Mazzeo, P. L. (2023). Effective heart disease prediction using machine learning techniques. *Algorithms*, *16*(2), 88-101. https://doi.org/https://doi.org/10.3390/a16020088

Bumm, C. V., Wölfle, U. C., Keßler, A., Werner, N., & Folwaczny, M. (2023). Influence of decision-making algorithms on the diagnostic accuracy using the current classification of periodontal diseases—a randomized controlled trial. *Clinical Oral Investigations*, *27*(11), 6589-6596. https://doi.org/https://doi.org/10.1007/s00784-023-05264-z

Darrab, S., Broneske, D., & Saake, G. (2024). Exploring the predictive factors of heart disease using rare association rule mining. *Scientific Reports*, *14*(1), 18178. https://doi.org/https://doi.org/10.1038/s41598-024-69071-6

Deepan, P., Vidhya, R., Rajalingam, B., Santhoshkumar, R., & Arul, N. (2024). FLAML-HDPS Model: An Efficient and Intelligent AutoML Approach for Heart Disease Prediction. International Conference on Computer & Communication Technologies.

Dorraki, M., Liao, Z., Abbott, D., Psaltis, P. J., Baker, E., Bidargaddi, N., Wardill, H. R., van den Hengel, A., Narula, J., & Verjans, J. W. (2024). Improving cardiovascular disease prediction with machine learning using mental health data: a prospective UK Biobank study. *JACC: Advances*, *3*(9_Part_2), 101180. https://doi.org/https://doi.org/10.1016/j.jacadv.2024.101180

Fahimfar, N., Khalili, D., Sepanlou, S. G., Malekzadeh, R., Azizi, F., Mansournia, M. A., Roohafza, H., Emamian, M. H., Hadaegh, F., & Poustchi, H. (2018). Cardiovascular mortality in a Western Asian country: results from the Iran Cohort Consortium. *BMJ open*, *8*(7), e020303. https://doi.org/https://doi.org/10.1136/bmjopen-2017-020303

Feurer, M., Klein, A., Eggensperger, K., Springenberg, J., Blum, M., & Hutter, F. (2019). Efficient and robust automated machine learning. *Advances in neural information processing systems*, *28*, 113-134. https://doi.org/https://doi.org/10.1007/978-3-030-05318-5_6

Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine learning*, *63*, 3-42. https://doi.org/https://doi.org/10.1007/s10994-006-6226-1

Gijsbers, P., Vanschoren, J., & Olson, R. S. (2017). Layered TPOT: Speeding up tree-based pipeline optimization. *European Conference on Machine Learning*, 49-68. https://doi.org/http://ceur-ws.org/Vol-1998/paper_06.pdf

Gonzalez-Abril, L., Angulo, C., Nuñez, H., & Leal, Y. (2017). Handling binary classification problems with a priority class by using Support Vector Machines. *Applied Soft Computing*, *61*, 661-669. https://doi.org/https://doi.org/10.1016/j.asoc.2017.08.023

Jafarnejad Chaghoshi, A., Rezasoltani, A., & Khani, A. M. (2024). Unleashing the Power of Ensemble Learning: Predicting National Ranks in Iran's University Entrance Examination. *Industrial Management Journal*, *16*(3), 457-481. https://doi.org/https://doi: 10.22059/imj.2024.381521.1008178.

Jha, S., Vaithiyanathan, D., Verma, P., & Kaur, B. (2024). An Automated Machine Learning Approach for Detecting Chronic Ischemic Heart Disease. 2024 International Conference on Advances in Modern Age Technologies for Health and Engineering Science (AMATHE).

Katoch, S., Chauhan, S. S., & Kumar, V. (2021). A review on genetic algorithm: past, present, and future. *Multimedia tools and applications*, *80*(5), 8091-8126. https://doi.org/https://doi.org/10.1007/s11042-020-10139-6

Kenny, A., Ray, T., Limmer, S., Singh, H. K., Rodemann, T., & Olhofer, M. (2023). Hybridizing TPOT with Bayesian optimization. Proceedings of the Genetic and Evolutionary Computation Conference.

Khani, A. M., Kazazi, A., & Taqhavi Fard, M. T. (2022). Evaluating the quality of services of the cultural and social deputy of Tehran municipality in the field of culture and art. *Social Development & Welfare Planning*, *13*(50), 205-250. https://doi.org/https://doi.org/10.22054/qjsd.2021.58035.2110

Koshiga, N., Borugadda, P., & Shaprapawad, S. (2023). Prediction of heart disease based on machine learning algorithms. 2023 International Conference on Inventive Computation Technologies (ICICT).

Liu, X., Su, S., Wang, B., & Zhang, X. (2023). Prediction of heart disease based on logistic regression and random forest models. *Highlights in Science Engineering and Technology*, *49*, 489-495. https://doi.org/https://doi.org/10.54097/hset.v49i.8599

Maihami, V., Khormehr, A., & Rahimi, E. (2016). Designing an expert system for prediction of heart attack using fuzzy systems. *Scientific Journal of Kurdistan University of Medical Sciences*, *21*(4), 118-131. https://doi.org/https://civilica.com/doc/810306

Maleki, S., & Mehrjerdi, Y. Z. (2022). Diagnosis of coronary artery disease by Bat and Harris hawk meta-heuristic optimization algorithms and machine learning methods. *Journal of Health Administration*, *25*(1), 57-68.

Mangalath Ravindran, S., Moorakkal Bhaskaran, S. K., K. Ambat, S., Balakrishnan, K., & Manguttathil Gopalakrishnan, M. (2022). An automated machine learning methodology for the improved prediction of reference evapotranspiration using minimal input parameters. *Hydrological Processes*, *36*(5), e14571. https://doi.org/https://doi.org/10.1002/hyp.14571

Mansouri, M., & Dadvar, M. (2017). *Diagnosing heart attacks using a model based on genetic algorithm and ensemble learning* Fifth National Conference on Computer Science, Engineering and Information Technology, Babol. https://civilica.com/doc/810306

McCall, J. (2005). Genetic algorithms for modelling and optimisation. *Journal of computational and Applied Mathematics*, *184*(1), 205-222. https://doi.org/https://doi.org/10.1016/j.cam.2004.07.034

Mohan, S., Thirumalai, C., & Srivastava, G. (2019). Effective heart disease prediction using hybrid machine learning techniques. *IEEE access*, *7*, 81542-81554. https://doi.org/https://doi.org/10.1109/access.2019.2923707

Mueller, J., Shi, X., & Smola, A. (2020). Faster, simpler, more accurate: Practical automated machine learning with tabular, text, and image data. Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining.

Nazari, S., & Jodki, S. (2020). *Using genetic algorithm and K-means clustering to improve the accuracy of support vector machine in the diagnosis of heart disease* National Conference on Latest Achievements in Data Engineering and Knowledge and Soft Computing, Shahrekord.

Nguyen, T.-D., Musial, K., & Gabrys, B. (2021). AutoWeka4MCPS-AVATAR: Accelerating automated machine learning pipeline composition and optimisation. *Expert systems with applications*, *185*, 115643. https://doi.org/https://doi.org/10.1016/j.eswa.2021.115643

Oliveto, P. S., Auger, A., Chicano, F., & Fonseca, C. M. (2020). Guest Editorial Special Issue on Theoretical Foundations of Evolutionary Computation. *IEEE Transactions on Evolutionary Computation*, *24*(6), 993-994. https://doi.org/https://doi.org/10.1109/tevc.2020.3035225

Olson, R. S., & Moore, J. H. (2019). TPOT: A tree-based pipeline optimization tool for automating machine learning. Workshop on automatic machine learning.

Orlenko, A., Kofink, D., Lyytikäinen, L.-P., Nikus, K., Mishra, P., Kuukasjärvi, P., Karhunen, P. J., Kähönen, M., Laurikka, J. O., & Lehtimäki, T. (2020). Model selection for metabolomics: predicting diagnosis of coronary artery disease using automated machine learning. *Bioinformatics*, *36*(6), 1772-1778. https://doi.org/https://doi.org/10.1093/bioinformatics/btz796

Paladino, L. M., Hughes, A., Perera, A., Topsakal, O., & Akinci, T. C. (2023). Evaluating the performance of automated machine learning (AutoML) tools for heart disease diagnosis and prediction. *Ai*, *4*(4), 1036-1058. https://doi.org/https://doi.org/10.3390/ai4040053

Pandiaraj, A., Prakash, S. L., & Kanna, P. R. (2021). Effective heart disease prediction using hybridmachine learning. 2021 Third international conference on intelligent communication technologies and virtual mobile networks (ICICV).

Rajdhan, N. A., Agarwal, N. A., Sai, N. M., & Ghuli, N. D. R. D. P. (2020). Heart Disease Prediction using Machine Learning. *IJERT*, *09*(04). https://doi.org/https://doi.org/10.17577/ijertv9is040614

Reddy, N. N., Nipun, L., Baba, M. U., Rishindra, N., & Shilpa, T. (2024). Optimizing heart disease prediction through ensemble and hybrid machine learning techniques. *International Journal of Electrical and Computer Engineering (IJECE)*, *14*(5), 5744-5754. https://doi.org/https://doi.org/10.11591/ijece.v14i5.pp5744-5754

Rezaeenoor, J., Saadi, G., & Jahani, M. (2019). Prediction of Cardiovascular Diseases Using an Optimized Artificial Neural Network. *Journal of Ilam University of Medical Sciences*, *27*(5), 15-23. https://doi.org/http://dx.doi.org/10.29252/sjimu.27.5.15

Salomon, R. (1997). Raising theoretical questions about the utility of genetic algorithms. International Conference on Evolutionary Programming.

Sarraf-Zadegan, N., Boshtam, M., Malekafzali, H., Bashardoost, N., Sayed-Tabatabaei, F., Rafiei, M., Khalili, A., Mostafavi, S., Khami, M., & Hassanvand, R. (1999). Secular trends in cardiovascular mortality in Iran, with special reference to Isfahan. *Acta cardiologica*, *54*(6), 327-333. https://doi.org/https://pubmed.ncbi.nlm.nih.gov/10672288

Sarrafzadegan, N., & Mohammmadifard, N. (2019). ardiovascular disease in Iran in the last 40 years: prevalence, mortality, morbidity, challenges, and strategies for cardiovascular prevention. *C PubMed*, *22*(4), 204-210. https://doi.org/https://pubmed.ncbi.nlm.nih.gov/31126179

Shah, D., Patel, S., & Bharti, S. K. (2020). Heart Disease Prediction using Machine Learning Techniques. *SN Computer Science*, *1*(6), 345. https://doi.org/https://doi.org/10.1007/s42979-020-00365-y

Siddhartha, M. (2020). Heart disease dataset (comprehensive). *IEEE Dataport*, *10*. https://doi.org/https://dx.doi.org/10.21227/dz4t-cm36

Takahashi, K., Yamamoto, K., Kuchiba, A., & Koyama, T. (2021). Confidence interval for micro-averaged F1 and macro-averaged F1 scores. *Applied Intelligence*, *52*(5), 4961–4972. https://doi.org/https://doi.org/10.1007/s10489-021-02635-5

Tiwari, A., Chugh, A., & Sharma, A. (2022). Ensemble framework for cardiovascular disease prediction. *Computers in Biology and Medicine*, *146*, 105624. https://doi.org/https://doi.org/10.1016/j.compbiomed.2022.105624

Truong, A., Walters, A., Goodsitt, J., Hines, K., Bruss, C. B., & Farivar, R. (2019). *Towards Automated Machine Learning: Evaluation and Comparison of AutoML Approaches and Tools* 2019 IEEE 31st International Conference on Tools With Artificial Intelligence (ICTAI).

Veisi, H., Ghaedsharaf, H. R., & Ebrahimi, M. (2021). Improving the performance of machine learning algorithms for heart disease diagnosis by optimizing data and features. *Soft*

computing journal, *8*(1), 70-85. https://doi.org/https://dor.isc.ac/dor/20.1001.1.23223707.1398.8.1.6.2

Verma, K., Bartwal, A. S., & Thapliyal, M. P. (2021). A genetic algorithm based hybrid deep learning approach for heart disease prediction. *Journal of Mountain Research*, *16*(3), 179-187. https://doi.org/https://doi.org/10.51220/jmr.v16i3.19

Villmann, T., Kaden, M., Lange, M., Sturmer, P., & Hermann, W. (2014). *Precision-Recall-Optimization in Learning Vector Quantization Classifiers for Improved Medical Classification Systems* 2014 IEEE Symposium on Computational Intelligence and Data Mining (CIDM).

Wang, J., Xue, Q., Zhang, C. W. J., Wong, K. K. L., & Liu, Z. (2024). Explainable coronary artery disease prediction model based on AutoGluon from AutoML framework. *Frontiers in Cardiovascular Medicine*, *11*. https://doi.org/https://doi.org/10.3389/fcvm.2024.1360548

Wang, S. (2023). Research on the heart attack prediction based on logistic regression. *Highlights in Science Engineering and Technology*, *65*, 153-158. https://doi.org/https://doi.org/10.54097/hset.v65i.11357

Whitley, D. (1994). A genetic algorithm tutorial. . *Statistics and computing*, *4*(2), 65-85. https://doi.org/https://doi.org/10.1007/bf00175354

Yadav, D., Saini, P., & Mittal, P. (2021). *Feature Optimization Based Heart Disease Prediction using Machine Learning* 2021 5th International Conference on Information Systems and Computer Networks (ISCON).

Yu, H. (2023). *Analysis and Prediction of Heart Disease Based on Machine Learning Algorithms* 2023 8th International Conference on Intelligent Computing and Signal Processing (ICSP).

Bibliographic information of this paper for citing:

Jafarnejad, Ahmad; Rezasoltani, Arman & Khani, Amir Mohammad (2025). Predicting Heart Disease Using Automated Machine Learning Based on Genetic Algorithms. *Journal of Information Technology Management*, 17 (2), 91-122. https://doi.org/10.22059/jitm.2024.382556.3829