Article

Influence of Optimal Hyperparameters on the Performance of Machine Learning Algorithms for Predicting Heart Disease

Ghulab Nabi Ahamad [1], Shafiullah [2,*], Hira Fatima [1], Imdadullah [3,*], S. M. Zakariya [3], Mohamed Abbas [4], Mohammed S. Alqahtani [5,6] and Mohammed Usman [4]

1   Institute of Applied Sciences, Mangalayatan University, Aligarh 202145, India
2   Department of Mathematics, K.C.T.C. College, Raxual, BRA, Bihar University, Muzaffarpur 842001, India
3   Electrical Engineering Section, University Polytechnic, Aligarh Muslim University, Aligarh 202002, India
4   Electrical Engineering Department, College of Engineering, King Khalid University, Abha 61421, Saudi Arabia
5   Radiological Sciences Department, College of Applied Medical Sciences, King Khalid University,
    Abha 61421, Saudi Arabia
6   BioImaging Unit, Space Research Center, Michael Atiyah Building, Univesity of Leicester,
    Leicester LE1 7RH, UK
*   Correspondence: shafi.stats@gmail.com (S.); imdadullah@zhcet.ac.in (I.)

Abstract: One of the most difficult challenges in medicine is predicting heart disease at an early stage. In this study, six machine learning (ML) algorithms, viz., logistic regression, K-nearest neighbor, support vector machine, decision tree, random forest classifier, and extreme gradient boosting, were used to analyze two heart disease datasets. One dataset was UCI Kaggle Cleveland and the other was the comprehensive UCI Kaggle Cleveland, Hungary, Switzerland, and Long Beach V. The performance results of the machine learning techniques were obtained. The support vector machine with tuned hyperparameters achieved the highest testing accuracy of 87.91% for dataset-I and the extreme gradient boosting classifier with tuned hyperparameters achieved the highest testing accuracy of 99.03% for the comprehensive dataset-II. The novelty of this work was the use of grid search cross-validation to enhance the performance in the form of training and testing. The ideal parameters for predicting heart disease were identified through experimental results. Comparative studies were also carried out with the existing studies focusing on the prediction of heart disease, where the approach used in this work significantly outperformed their results.

Keywords: heart disease prediction; UCI Kaggle dataset; machine learning algorithms; GridSearchCV; hyperparameters

1. Introduction

The most important factor in blood flow through veins is the heart [1]. The blood that circulates through our bodies and carries nutrients, oxygen, metals, and other essential substances is the most important part of our circulatory system. The faulty functioning of the heart can lead to serious health issues and even death [2]. Living an unhealthy lifestyle, using tobacco, drinking alcohol, and eating a lot of fat can all lead to heart disease [3,4]. The World Health Organization estimates that heart disease claims the lives of roughly 10 million people per year. Only a healthy lifestyle and early detection can stop circulatory system diseases [5,6]. Despite the fact that in recent years, cardiac issues have been identified as the main cause of death worldwide, they are still conditions that can be properly managed and controlled. How effectively an illness can be controlled overall depends on the exact timing of its detection. The recommended strategy tries to recognize certain cardiac abnormalities early in order to stop heart disease. Several researchers are utilizing statistical and data mining techniques to help identify heart illness [7]. The majority of the data in the medical database are discrete. Making decisions with these datasets is therefore extremely challenging [8–10].

In the healthcare sector, machine learning (ML) is an emerging topic that can help with the diagnosis of diseases, the discovery of new drugs, and the classification of images. It is particularly helpful for hospital administration and medical staff, including doctors and nurses, as well as residential treatment facilities. Early heart illness identification and prediction are more challenging without modern medical tools. By developing new models, ML algorithms are employed for early treatment and diagnosis. ML algorithms are essential for analyzing the presented data and finding hidden discrete patterns. Two heart disease datasets were examined for model performance using a variety of ML methods, namely, logistic regression, K-nearest neighbor, support vector machine, decision tree, random forest, and extreme gradient boosting. The goal of this research was to employ grid search cross-validation to enhance training and testing performance and to find the perfect ML algorithm parameters for heart disease prediction. The efficiency of heart disease prediction and patient survival can be successfully predicted using tuned hyperparameters for larger heart disease datasets, as well as for comprehensive datasets, but it is ineffective for small datasets. This improves the effectiveness and performance of ML algorithms for predicting heart disease in the form of training and testing statistical data [11–13]. To determine the efficiency of the machine learning algorithms, numerous reliable performance-measuring matrices were given. The major findings of the study are listed below:

- In the first step, common ML algorithms for predicting heart disease were used, as follows: logistic regression, K-nearest neighbor, support vector machine, decision tree, random forest classifier, and extreme gradient boosting.
- In the second step, a prediction system using six fundamental ML algorithms and a hyperparameter tuning technique was provided. Here, grid search cross-validation was also used to identify the appropriate hyperparameters for each method. The notations used to describe the analytical results of different tables are as follows: accuracy (A), precision (P), recall (R), F-1 score (F-1s), and support (S).
- Finally, a confusion matrix was used to compare the performances of the models for these two systems.

The following is the breakdown of the remaining text: An effort to review the literature is given in Section 2. The research methodology is described in Section 3, which also describes the actions that were done to conduct this study. The performance evaluation metrics, findings, and discussion for the experimental setups are presented in Section 4. All of the tests performed, their associated results, and a state-of-the-art comparison are discussed in Section 5. Finally, Section 6 concludes the work and gives suggestions for additional research.

2. Literature Review

The contributions of recent heart disease prediction systems are summarized in this section. To predict heart disease, researchers created numerous machine learning (ML) classification models.

2.1. Existing Models for Predicting Cardiovascular Diseases

To predict the risk of cardiovascular illness, well-known ML techniques, such as logistic regression, K-nearest neighbor, support vector machine, decision tree, and random forest classifier, were designed [14]. The classifier for early cardiovascular disease prediction using the artificial immune identification method (AIRS) with a fuzzy resource allocation mechanism is reported in [15]. For the diagnosis of cardiovascular disease without the use of intrusive diagnostic methods, a mixed model based on clinical data was developed [16].

2.2. Methodology for Detecting Heart Disease

On the other hand, various diseases or risk factors for smoking-related coronary disease were found, and features were found using entropy. A weighted methodology for converting bootstrap-aggregated ensemble learning to a weighted vote and comparing different averaging methods was built. This was successful at identifying heart disease,

with an accuracy of 89.30% using cluster-based decision tree learning and 76.70% using a random forest classifier [17,18].

By using clustering and classification approaches, heart disease was detected. The construction of a model for information-mining-based clinical mindfulness success and the function of rehabilitation specialists in clinical information mining were taken into consideration [19,20]. Heart disease prediction using a combination of deep learning and machine learning reported an accuracy of 100% [21].

Using the AI database at UCI, a system for predicting heart failure was created. It was based on the subtyping of ischemic stroke patients from a participant observation registry created by platform vascular psychiatrists. To reduce the computational complexity of the huge database, they employed a variety of feature selection approaches, including principle component analysis and extreme gradient boosting [22,23].

2.3. Levels of Heart Disease

The capacity of the heart disease application to predict risk levels for heart attacks was the subject of much investigation. For the purpose of predicting heart disease, 11 critical qualities were utilized, as well as fundamental data mining techniques, such as an NB, a J48 decision tree classifier, and bagging approaches [24,25].

The clinical summary report of the European Association of Comparative Cardiology, which details how to improve adherence to a medical therapy advised by recommendations in the lifestyle modification of heart disease, was produced. Deep hybrid learning and signal processing were used to detect early-stage heart disease using phonocardiogram and electrocardiogram signals [26,27].

Additionally, the study of clinical data for cardiovascular prediction also used ML algorithms. An extreme gradient boosting classifier was also employed, along with a successful model for predicting heart disease for a clinical decision support system [28,29].

3. Research Methodology

The datasets and recommendation mechanisms are discussed in this section.

3.1. The Objective of the Study

Today, heart disease is a major problem. Neither a more accurate automated technology nor a decrease in the effects of heart disease is currently available. Therefore, using machine learning algorithms to identify diseases from common symptoms would be a significant accomplishment. These machine learning techniques are useful for diagnosing and identifying dangerous diseases earlier. As a result of the medical information system using patient data, it is now possible to address new challenges in the healthcare industry through the application of an appropriate methodology and machine learning techniques. To build a model for data extraction and categorization, statistical and machine learning techniques that extrapolate knowledge from large and complex datasets are used.

The primary objective of this study was to create a model that can accurately identify issues with cardiac diseases.

3.2. Description of the Datasets

Dataset-I: UCI Kaggle Cleveland. The Medical Centre and the Cleveland Clinic Foundation released this dataset on heart disease, which may be found in the UCI repository, along with 2 classes, 14 features, and 303 instances.

Dataset-II: Comprehensive UCI Kaggle Cleveland, Hungary, Switzerland, and Long Beach V. This heart disease dataset, which has 14 features, 1025 instances, and 2 classes, is contributed to by the Medical Center, the Cleveland Clinical Foundation, the Hungarian Institute of Cardiology, Switzerland, and the Long Beach V Clinical Foundation. It may also be found in the UCI repository. Summary of dataset is given in Table 1.

Table 1. Descriptions of the datasets.

	Classes	Features	Instances
Dataset-I	2	14	303
Dataset-II	2	14	1025

3.3. Suggested Model

In this study, two stages of heart disease predictions were taken into account. Figure 1 shows the flow chart for diagnosing the traditional form of heart disease. This model was not described using a hyperparameter tuning technique for a classification algorithm [30]. Six distinct machine learning algorithms are displayed for both the conventional models and the suggested techniques depicted in Figure 1. These six suggested models were then used to examine the testing dataset and evaluate the accuracy of the results. One of the most crucial components in machine learning is hyperparameter tuning. If the hyperparameters are tuned, the ML algorithms will perform more efficiently. The ideal settings for the hyperparameters can be found by conducting a thorough search, such as GridSearchCV.

It is also capable of creating a model that generates and saves every possible model and parameter combination. Time and resources are saved by this search. The six different machine learning classifiers come from several machine learning applications, which include logistic regression classifier (LR) [31], K-nearest neighbors classifier (K-NN) [32], support vector machine classifier (SVM) [33], decision tree classifier (DT) [34], random forest classifier (RFC) [35], and extreme gradient boosting classifier (XGB) [36].

3.4. Problem Statement for the Study

Heart disease instances are rising quickly each day, and thus, it is crucial to predict any possible diseases in advance. The main challenge with heart disease is detecting it. There are various tools that can predict cardiac disease, but they must be calculated accurately and effectively. The mortality rate and total consequences can be reduced by early identification of aortic stenosis. Since it takes more intelligence, time, and knowledge, it is not always possible to accurately monitor patients every day, and a doctor cannot consult with a patient for a whole 24 h.

By using computer-assisted methods, such as machine learning, one may predict the patient's status quickly and more accurately while also drastically cutting costs. Machine learning is a highly broad and diverse field, and its application in healthcare is expanding daily. Today's internet contains a considerable amount of information. The data is therefore examined for hidden patterns using a variety of machine learning algorithms. In medical data, hidden patterns might be used for health diagnosis. By examining patient data that uses machine learning algorithms to identify whether a patient has heart disease, this effort seeks to predict possible heart disease.

3.5. Hyperparameter Tuning Optimization

The challenge of selecting a set of ideal hyperparameters for a learning algorithm is known as hyperparameter optimization or tuning in machine learning. A parameter whose value is utilized to regulate the learning process is known as a hyperparameter. There are several optimization methods, each with its benefits and drawbacks. The values of other parameters are often learned. Choosing the best hyperparameters has a significant influence on the performance model. Experiments on various optimization techniques were used to identify the best hyperparameter combination, which was consequently employed in these six machine learning algorithms: logistic regression classifier, K-nearest neighbor classifier, support vector machine classifier, decision tree classifier, random forest classifier, and extreme gradient boosting classifier.
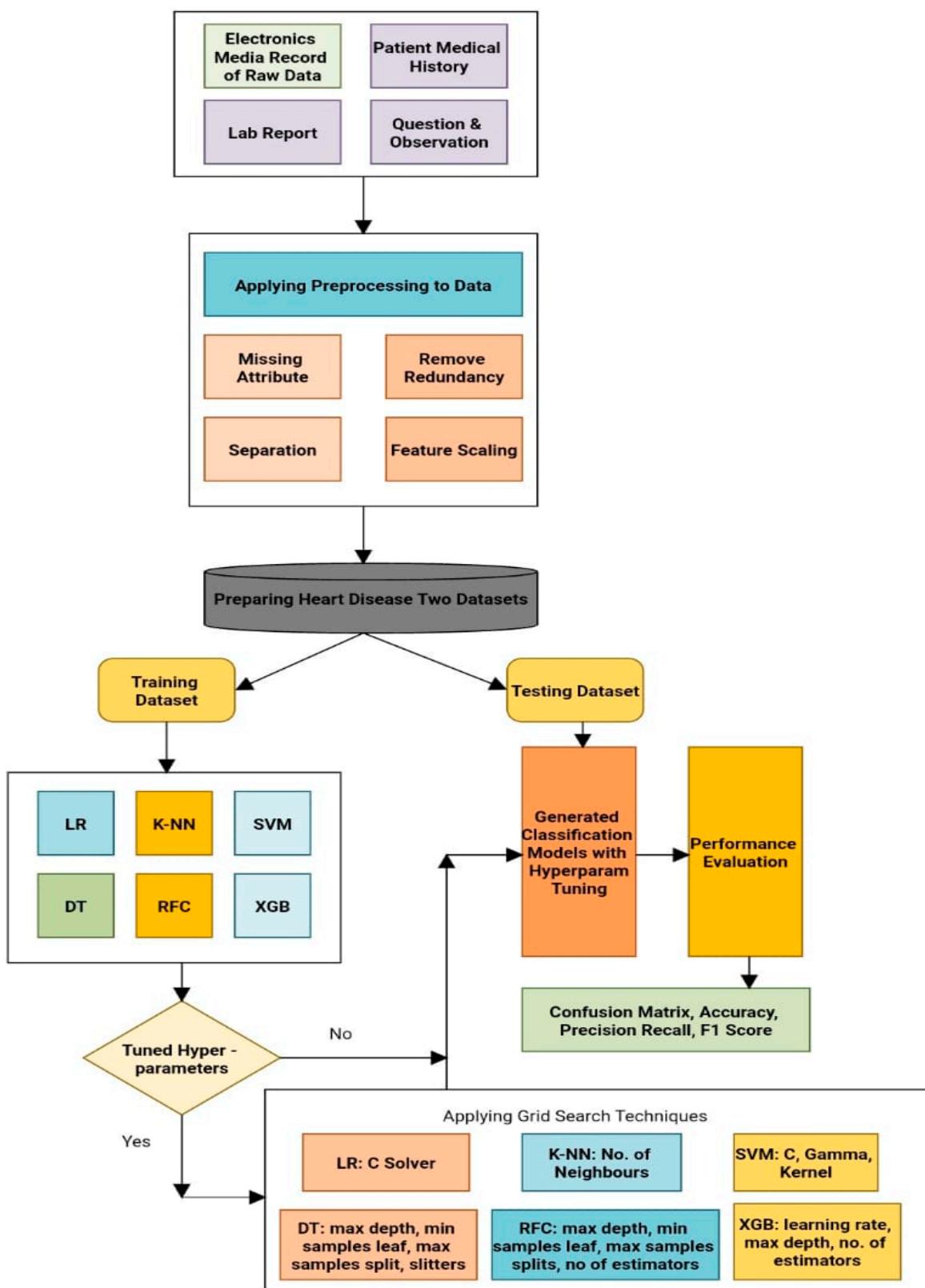
Figure 1. Suggested model with and without tuned hyperparameters.

The careful tuning of machine learning algorithms is one of the optimization challenges. The GridSearchCV approach is frequently used in hyperparameter optimization to overcome challenges and improve the accuracy of models. GridSearchCV is a time-tested

method that considers all hyperparameter combinations. The learning rate and layer count are used as hyperparameters in GridSearchCV.

4. Evaluation Metrics and Experimental Data Analysis

4.1. Metrics for Evaluation

To assess how well a statistical or machine learning system is performing, evaluation metrics are utilized. Each evaluation evaluates the machine learning algorithm. There are numerous assessment measures that can be used to test a model [37]. True negative (*TN*), true positive (*TP*), false positive (*FP*), and false negative (*FN*) are the parameters of the assessment metrics.

$$\text{accuracy} = \frac{\sum TP + \sum TN}{\sum TP + \sum TN + \sum FP + \sum FN} \tag{1}$$

$$\text{precision} = \frac{\sum TP}{\sum TP + \sum FP} \tag{2}$$

$$\text{recall} = \frac{\sum TP}{\sum TP + \sum FN} \tag{3}$$

$$\text{F} - 1\text{Score} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \tag{4}$$

4.2. Description of the Features of Heart Disease Datasets

The parameters for predicting heart disease are shown in Figure 2. These features (attributes) are found in both dataset-I and dataset-II. The descriptions are as follows:

Age: the age of the individual.

Sex: the gender of the individual using the following form: 1 = male and 0 = female.

Chest pain type (cp): the types of chest pain experienced by the individual using the following form: 1 = typical angina, 2 = atypical angina, 3 = non-anginal pain, and 4 = asymptotic.

Resting blood pressure (trestbps): The resting blood pressure value of an individual in mmHg (unit). The resting blood pressure (in mmHg) is often a reason for worry if it is between 130 and 140; a sick heart will stress more during exercise.

Serum cholesterol (chol): the serum cholesterol in mg/dL (unit); serum cholesterol is usually a cause for concern if it is 200 or higher.

Fasting blood sugar (fbs): Compares the fasting blood sugar value of an individual with 120 mg/dL. If fasting blood sugar > 120 mg/dL, then 1 (true); else, 0 (false).

Resting ECG (restecg): 0 = normal, 1 = having ST-T wave abnormality, and 2 = left ventricular hypertrophy.

Max heart rate achieved (thalach): the max heart rate achieved by an individual.

Exercise-induced angina (exang): angina caused by exercise according to the slope of the peak exercise ST segment; those with value 0 (no exercise-caused angina) had a higher risk of heart disease than those with value 1 (presence of exercise-induced angina).

ST depression induced by exercise relative to rest (oldpeak): the value, which is integer or float; peak exercise ST segment (slope): 1 = upsloping, 2 = flat, and 3 = downsloping.

Number of major vessels (0–3) colored using fluoroscopy (CA) is based on the principle that more vessels shown signify more blood movement.

The thalassemia (Thal): 3 = normal, 6 = fixed defect, and 7 = reversible defect.

Diagnosis of heart disease (target): displays whether the individual is suffering from heart disease or not: 0 = absence and 1 = present.
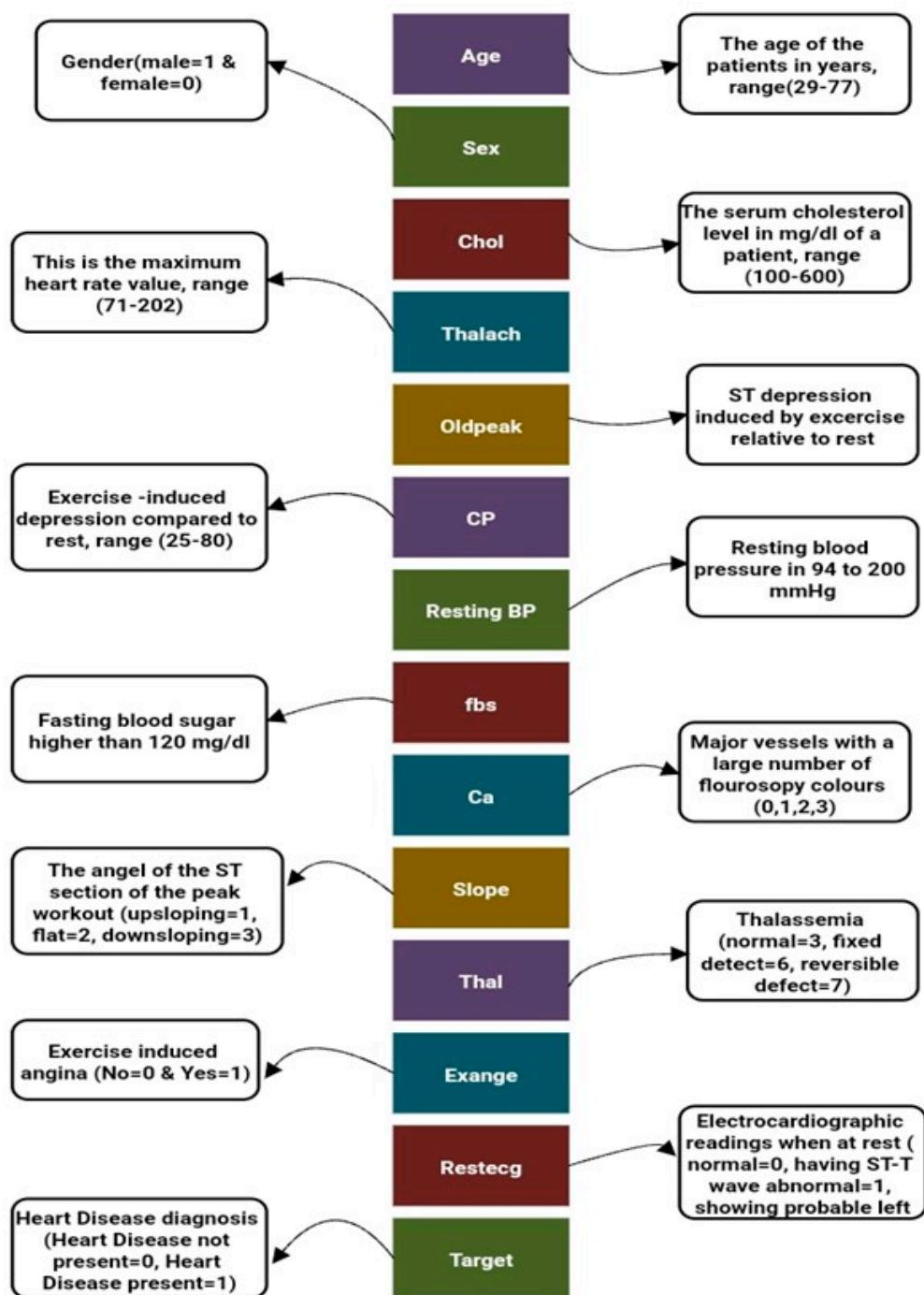
Figure 2. Heart disease prediction parameters.

4.3. Experimental Data Analysis

A total of 303 samples with 14 attributes make up dataset-I; 138 of the samples have heart disease, whereas 165 are healthy. Dataset-II comprises 1025 samples with 14 features, where 525 of the samples have heart disease and 499 do not. During the pre-processing stage, the statistical operation was completed to find and remove missing values, as well as to ascertain the maximum (max), minimum (min), mean, 25%, 50%, 75%, and standard deviation (std) of each feature set. Tables 2 and 3 display the results.

Table 2. The description of dataset-I for the count (303), minimum, maximum, mean, and standard deviation.

	Age	Sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
Mean	54.47	0.69	0.97	131.62	247.00	0.16	0.54	148.99	0.38	1.05	1.50	0.74	2.32	0.55
Std	9.09	0.48	1.04	17.55	52.77	0.37	0.54	22.57	0.48	1.17	0.63	1.03	0.62	0.51
Min	29.00	000	0.00	93	126	0.00	0.00	71	0.00	0.00	0.00	0.00	0.00	0.00
25%	47.00	0.00	0.00	120	211	0.00	0.00	133.50	0.00	0.00	1.00	0.00	2.00	0.00
50%	50	1.00	1.00	130	240	0.00	1.00	153	0.00	8.00	1.00	0.00	2.00	1.00
75%	61	1.00	2.00	140	274	0.00	1.00	166	1.00	1.60	2.00	1.00	3.00	1.00
Max	77	1.00	3.00	200	564	1.00	2.00	202	1.00	6.20	2.00	4.00	3.00	1.00

Table 3. The description of dataset-II for the count (1025), minimum, maximum, mean, and standard deviation.

	Age	Sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
Mean	54.43	0.70	0.94	131.61	245.06	0.15	0.53	149.11	0.34	1.07	1.39	0.75	2.32	0.51
Std	9.07	0.46	1.03	17.52	51.59	0.36	0.53	23.01	0.47	1.18	0.62	1.03	0.62	0.50
Min	29.00	0.00	0.00	94	126	0.00	0.00	71	0.00	0.00	0.00	0.00	0.00	0.00
25%	48.00	0.00	0.00	120	211	0.00	0.00	132	0.00	0.00	1.00	0.00	2.00	0.00
50%	56	1.00	1.00	130	240	0.00	1.00	152	0.00	8.00	1.00	0.00	2.00	1.00
75%	61	1.00	2.00	140	275	0.00	1.00	166	1.00	1.80	2.00	1.00	3.00	1.00
Max	77	1.00	3.00	200	564	1.00	2.00	202	1.00	6.20	2.00	4.00	3.00	1.00

According to Figure 2, people with "cp" 1, 2, or 3 on a resting ECG are more likely to develop heart disease than people with "cp" 0 (cp stands for chest pain type; value 1: typical angina, value 2: atypical angina, value 3: non-anginal pain, and value 4: asymptomatic). People with cp value 1 are more likely to have heart disease since it indicates an abnormal heartbeat, which is noticeable in issues ranging from trivial symptoms to major issues. Considering angina caused by exercise, according to the slope of the peak exercise ST segment, those with cp value 0 (no exercise-caused angina) had a higher risk of HD than those with cp value 1 (presence of exercise-induced angina).

The number of main blood vessels (0–3) colored using fluoroscopy was based on the principle that continuous blood circulation makes the heart better. Then, for simplicity and improved comprehension, the histogram of categorical and continuous characteristics is shown in Figure 3. For the structure and frequency range of continuous and categorical observations, histogram charts display the distribution of each characteristic value.

Figure 3 shows that those with a maximum heart rate of more than 140 were more likely to have heart disease. Resting blood pressure (in mmHg) is often a reason for worry if it is between 130 and 140, and serum cholesterol is usually a cause for concern if it is 200 or higher.

Figure 4 displays the heat map that illustrates how characteristics of the heart disease datasets were related to one another. Here, the values on the two-dimensional surface are shown using various hues. It is clear that qualities with a categorical value were more concentrated than those with a continuous value.
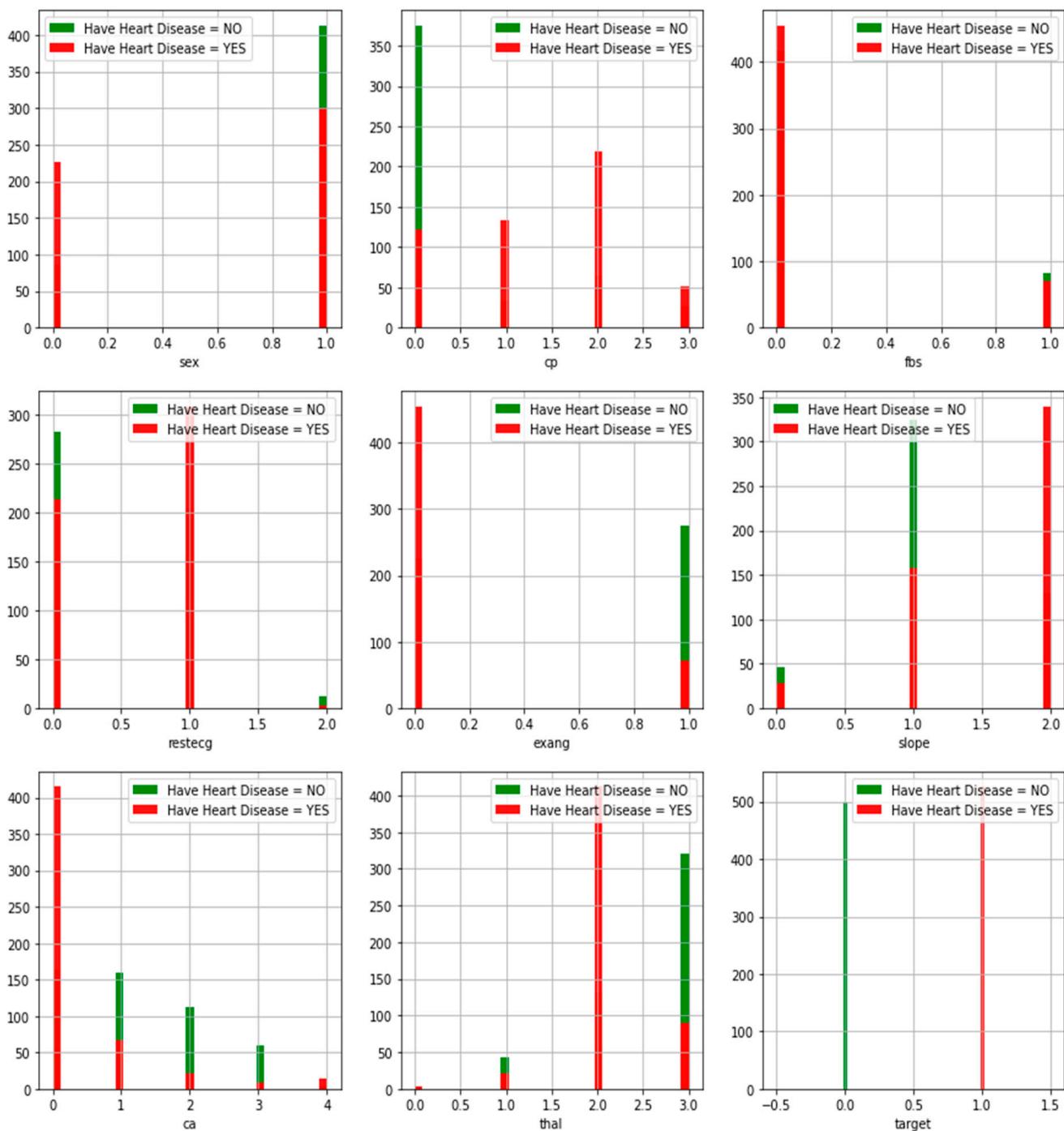
Figure 3. The histograms of characteristics with categorical values.
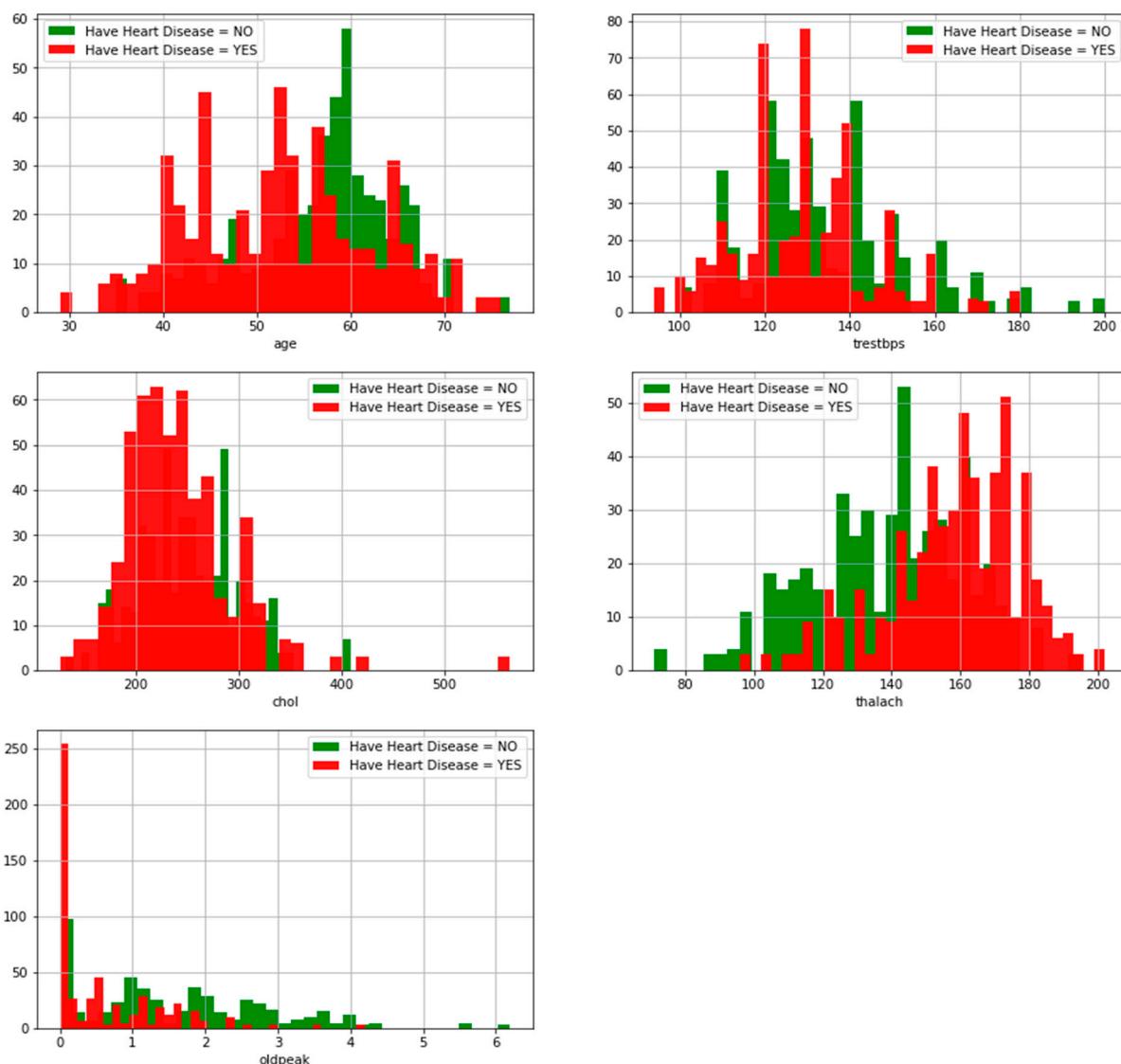
Figure 4. The histograms of continuous-valued attributes.

5. Discussion and Analysis of the Experiment Results

5.1. Data Preparation

Data preparation was used to identify null values; process corrupt, missing, disrespectful, and inaccurate values; and eliminate the duplication of particular characteristics. The standard data format was then identified through splitting, feature scaling, and normalization. The dataset was divided into a training dataset and a test dataset, with the training dataset containing 70% of the data and the test dataset containing 30% of the data. Pre-processing is a statistical technique that finds and removes missing data while also determining the maximum, minimum, mean, and standard deviation of each feature set.

5.2. Performance Evaluation and Comparison with a Traditional System for Dataset-I

The default settings were used when applying the ML algorithms in this experiment. The results of the system are shown in Figure 5. The accuracy, precision, recall, and F-1 score of the model were found to be 86.79%, 87%, 87%, and 87%, respectively, during the fitting and running phases of the LR training phase. The accuracy, precision, recall, and F-1 score for the test dataset that this LR predicted were 86.91%, 87%, 87%, and 87%, respectively. Again, during the training phase, the K-NN model was run with the parameters' "uniform" weights and the number of neighbors (K) = 5, and it produced an accuracy, precision, recall,

and F-1 score of 86.79%, 87%, 87%, and 87%, respectively. The accuracy, precision, recall, and F-1 score for this K-NN model, which predicted the test set, were 86.81%, 87%, 87%, and 87%, respectively. The performances of the SVM for training in terms of the accuracy, precision, recall, and F-1 score were 93.40%, 93%, 93%, and 93%, respectively, using the settings of kernel = RBF, gamma = 0.1, and C = 1.0. The accuracy, precision, recall, and F-1 score of the test dataset for the SVM were 87.91%, 88%, 88%, and 88%, respectively. The DT ran the model with random_state = 42 parameters for training datasets; the results indicated 100% accuracy, precision, recall, and F-1 score. The accuracy, precision, recall, and F-1 score for the test dataset produced using the DT were 78.02%, 78%, 78%, and 78%, respectively. The RFC ran the model with n_estimators = 1000 and random_state = 42 parameters for the training datasets; the results showed 100% accuracy, precision, recall, and F-1 score. The accuracy, precision, recall, and F-1 score for the test dataset produced using the RFC were 82.02%, 82%, 82%, and 82%, respectively.
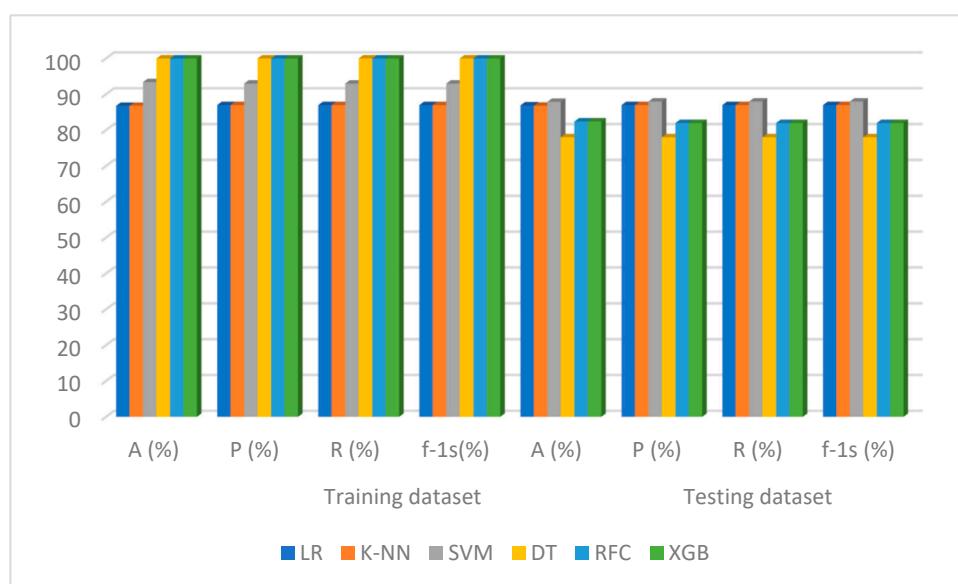
Figure 5. Graphical representation of the performance evaluation of the traditional system.

The model was run on training datasets using the XGB with level encoder = false parameters; the results indicated 100% accuracy, precision, recall, and F-1 score. The accuracy, precision, recall, and F-1 score for the test dataset produced using the XGB were 82.42%, 82%, 82%, and 82%, respectively.

The details of the Figure 6 result are listed in Appendix A, Table A1. The classification report of several types of training and testing results for traditional models, including precision, recall, F-1 score, and support, are shown in Table 4. The confusion matrix for the type-I error and type-II error of the traditional model is shown in Table 5. The total number of instances was 303.

5.3. Performance Evaluation with Tuned Hyperparameters for Dataset-I

The suggested method used GridSearchCV to find the ideal hyperparameters. After the hyperparameters were modified, the classifying models were constructed. For both the training and test datasets for each of the six machine learning methods, Figure 7 shows the outcomes of the suggested system with tuned hyperparameters in terms of the accuracy, precision, recall, and F-1 score. The details of the Figure 5 findings are given in Appendix A, Table A2.
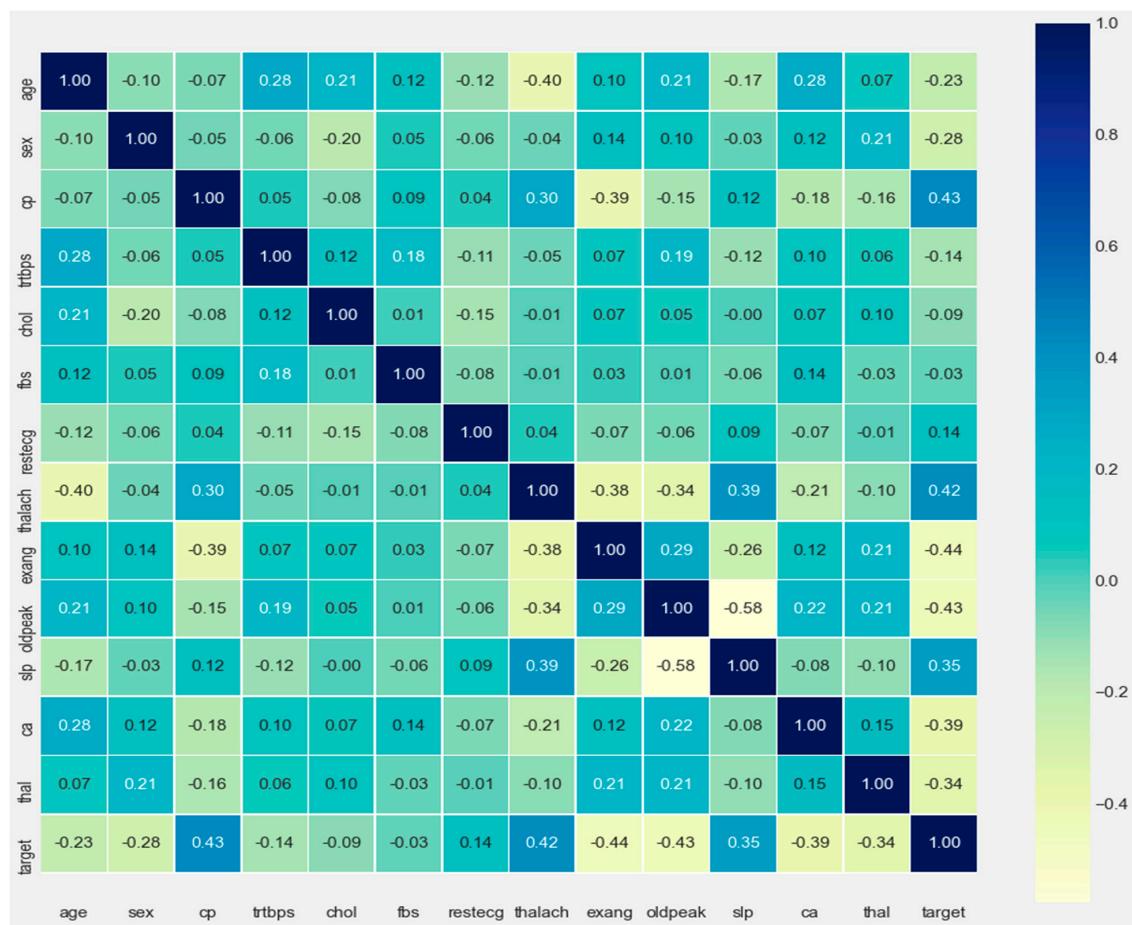
Figure 6. The heart disease datasets heat map for correlation characteristics.

Table 4. Analytical results of various types of training and testing datasets of traditional ML models.

	Traditional Models		Normal (0)	Abnormal (1)	A (%)	Macro Avg (%)	Weighted Avg (%)
LR	Training result	P (%)	88	86	87	87	87
		R (%)	82	90	87	86	87
		F-1s (%)	85	88	87	87	87
		S (%)	97	115	87	212	212
	Testing result	P (%)	87	87	87	87	87
		R (%)	83	90	86	86	87
		F-1s (%)	85	88	87	87	87
		S (%)	41	50	87	91	91
K-NN	Training result	P (%)	86	87	87	87	87
		R (%)	85	89	87	87	87
		F-1s (%)	85	88	87	87	87
		S (%)	97	115	87	212	212
	Testing result	P (%)	85	88	87	87	87
		R (%)	85	88	87	87	87
		F-1s (%)	85	88	87	87	87
		S (%)	41	50	87	91	91
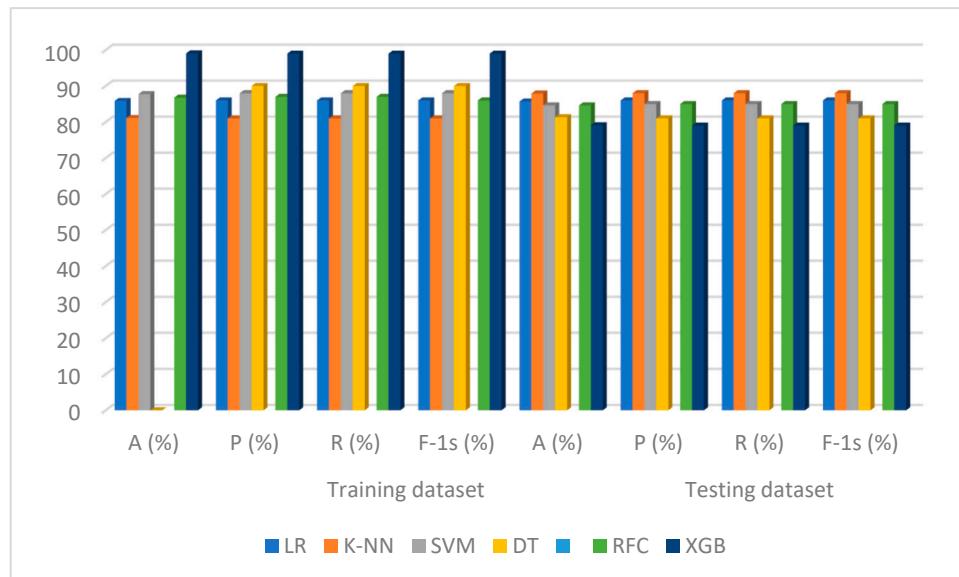
Table 4. *Cont.*

Traditional Models		Normal (0)	Abnormal (1)	A (%)	Macro Avg (%)	Weighted Avg (%)
SVM	Training result					
	P (%)	94	93	93	93	93
	R (%)	92	95	93	93	93
	F-1s (%)	93	94	93	93	93
	S (%)	97	115	93	212	212
	Testing result					
	P (%)	86	90	88	88	88
	R (%)	88	88	88	88	88
	F-1s (%)	87	90	88	88	88
	S (%)	41	50	88	91	91
DT	Training result					
	P (%)	100	100	100	100	100
	R (%)	100	100	100	100	100
	F-1s (%)	100	100	100	100	100
	S (%)	97	115	100	212	212
	Testing result					
	P (%)	73	84	78	88	79
	R (%)	83	74	78	88	78
	F-1s (%)	77	79	78	88	78
	S (%)	41	50	78	91	91
RFC	Training result					
	P (%)	100	100	100	100	100
	R (%)	100	100	100	100	100
	F-1s (%)	100	100	100	100	100
	S (%)	97	115	100	212	212
	Testing result					
	P (%)	80	84	82	82	82
	R (%)	80	84	82	82	82
	F-1s (%)	80	84	82	82	82
	S (%)	41	50	82	91	91
XGB	Training result					
	P (%)	100	100	100	100	100
	R (%)	00	100	100	100	100
	F-1s (%)	100	100	100	100	100
	S (%)	97	115	100	212	212
	Testing result					
	P (%)	80	84	82	82	82
	R (%)	80	84	82	82	82
	F-1s (%)	80	84	82	82	82
	S (%)	41	50	82	91	91

The classification report of several types of training and testing results for the suggested model with tuned hyperparameters, including precision, recall, F-1 score, and support, are shown in Table 6. The confusion matrix for the type-I error and type-II error of the model with tuned hyperparameters is shown in Table 7.

Table 5. Performance evaluation using a confusion matrix on the training and testing datasets.

		Training Dataset			Testing Dataset	
Models	**Confusion Matrix**	**Type-I Error**	**Type-II Error**	**Confusion Matrix**	**Type-I Error**	**Type-II Error**
LR	[[80 17] [11 104]]	184 (correct)	28 (incorrect)	[[34 7] [5 45]]	79 (correct)	12 (incorrect)
KNN	[[82 15] [13 102]]	184 (correct)	28 (incorrect)	[[35 6] [6 44]]	79 (correct)	12 (incorrect)
SVM	[[89 8] [6 109]]	198 (correct)	14 (incorrect)	[[36 5] [6 44]]	80 (correct)	11 (incorrect)
DT	[[97 0] [0 115]]	212 (correct)	0 (incorrect)	[[34 7] [13 37]]	71 (correct)	20 (incorrect)
RFC	[[97 0] [0 115]]	212 (correct)	0 (incorrect)	[[33 8] [8 42]]	75 (correct)	16 (incorrect)
XGB	[[97 0] [0 115]]	212 (correct)	0 (incorrect)	[[33 8] [8 42]]	75 (correct)	16 (incorrect)

Figure 7. The performance evaluation of the suggested system in graphical form.

Table 6. The classification report of models using the training and testing datasets compared with a hyperparameter tuning strategy.

Hyperparameter Tuning Classification Report		Normal (0)	Abnormal (1)	A (%)	Macro Avg	Weighted Avg
Tuned LR	Training result					
	P (%)	86	86	86	86	86
	R (%)	82	89	86	86	86
	F-1s (%)	84	87	86	86	86
	S (%)	97.00	115	86	212	212
	Testing result					
	P (%)	85	86	86	86	86
	R (%)	83	88	86	85	86
	F-1s (%)	84	87	86	86	86
	S (%)	41	50	86	91	91

Table 6. *Cont.*

Hyperparameter Tuning Classification Report		Normal (0)	Abnormal (1)	A (%)	Macro Avg	Weighted Avg
Tuned K-NN	Training result					
	P (%)	84	80	81	82	81
	R (%)	83	88	81	81	81
	F-1s (%)	78	83	81	81	81
	S (%)	97	115	81	212	212
	Testing result					
	P (%)	89	87	88	88	88
	R (%)	83	92	88	87	88
	F-1s (%)	86	89	88	88	88
	S (%)	41	50	88	91	91
Tuned SVM	Training result					
	P (%)	88	87	88	88	88
	R (%)	85	89	88	87	88
	F-1s (%)	86	90	88	88	88
	S (%)	97	115	88	212	212
	Testing result					
	P (%)	85	85	85	85	85
	R (%)	80	88	85	84	85
	F-1s (%)	83	86	85	84	85
	S (%)	41	50	85	91	91
Tuned DT	Training result					
	P (%)	94	87	90	90	90
	R (%)	82	96	90	89	90
	F-1s (%)	88	91	90	89	90
	S (%)	97	115	90	212	212
	Testing result					
	P (%)	80	82	81	81	81
	R (%)	78	84	81	81	81
	F-1s (%)	79	83	81	81	81
	S (%)	41	50	81	91	91
Tuned RFC	Training result					
	P (%)	89	85	87	87	87
	R (%)	81	91	87	86	87
	F-1s (%)	85	88	86	87	87
	S (%)	97	115	87	212	212
	Testing result					
	P (%)	85	85	85	85	85
	R (%)	80	88	85	84	85
	F-1s (%)	83	86	85	84	85
	S (%)	41	50	85	91	91
Tuned XGB	Training result					
	P (%)	100	98	99	99	99
	R (%)	98	100	99	99	99
	F-1s (%)	99	99	99	99	99
	S (%)	97	115	99	212	212
	Testing result					
	P (%)	76	82	79	79	79
	R (%)	78	80	79	79	79
	F-1s (%)	77	81	79	79	79
	S (%)	41	50	79	91	91

Table 7. The performance evaluation and comparison of the confusion matrix with hyperparameter tuning during the training and testing on dataset-I.

	Training Dataset			Testing Dataset		
Models	**Confusion Matrix**	**Type-I Error**	**Type-II Error**	**Confusion Matrix**	**Type-I Error**	**Type-II Error**
TLR	[[80 17] [13 102]]	182 (correct)	30 (incorrect)	[[34 7] [6 44]]	78 (correct)	13 (incorrect)
TK-NN	[[71 26] [14 102]]	173 (correct)	40 (incorrect)	[[34 7] [4 46]]	80 (correct)	11(incorrect)
TSVM	[[82 15] [11 106]]	188 (correct)	26 (incorrect)	[[33 8] [6 44]]	77 (correct)	14 (incorrect)
TDT	[[80 17] [5 110]]	190 (correct)	22 (incorrect)	[[32 9] [8 42]]	74 (correct)	17 (incorrect)
TRFC	[[79 18] [10 105]]	184 (correct)	28 (incorrect)	[[33 8] [6 44]]	77 (correct)	14 (incorrect)
TXGB	[[95 2] [0 115]]	210 (correct)	2 (incorrect)	[[32 9] [10 40]]	72 (correct)	19 (incorrect)

The performance of the suggested method was compared with that of the existing system in Figure 8. The details of the Figure 7 results are provided in Appendix A, Table A3.
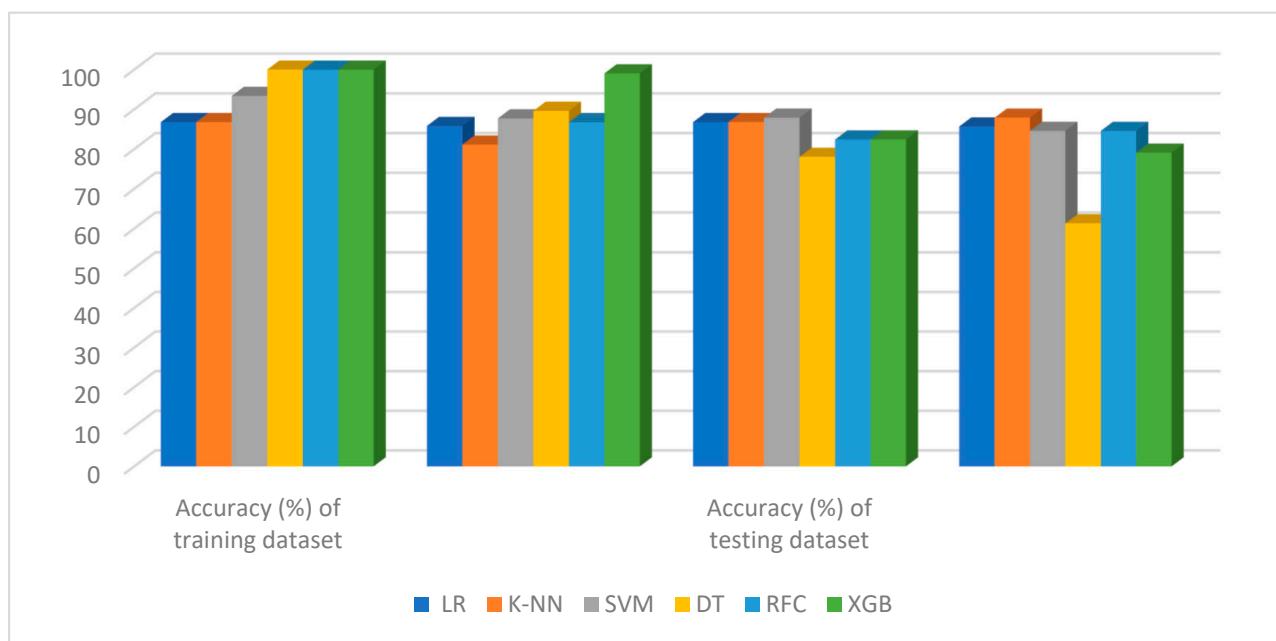

Figure 8. Graphical comparison of the accuracy.

The results from dataset-I did not seem to considerably improve after the hyperparameter tuning due to the small dataset.

5.4. Performance Evaluation and Comparison with the Traditional System for Dataset-II

For both the training and test datasets for each of the six machine learning methods, Table 8 shows the results of traditional system without tuned hyperparameters in terms of accuracy, precision, recall, and F-1 score. Table 8 is depicted graphically in Figure 8.

Table 8. For both the training and testing datasets, classification models were evaluated and compared after using dataset-II.

Traditional Method		Training Dataset				Testing Dataset			
Models	Parameters	A (%)	P (%)	R (%)	F-1s (%)	A (%)	P (%)	R (%)	F-1s (%)
LR	Solver = liblinear	89.54	90	90	90	81.82	82	82	82
K-NN	K = 5, weights = uniform	91.77	92	92	92	81.82	82	82	82
SVM	Kernel = "rbf", gamma = 0.1, C = 1.0	95.40	95	95	95	90.26	90	90	90
DT	Random_state = 42	100.00	100	100	100	97.08	97	97	97
RFC	n_estimators = 1000, random_state = 42	100.00	100	100	100	98.05	98	98	98
XGB	Label_encoder = false	100.00	100	100	100	99.03	99	99	99

The classification report of several types of training and testing results, namely, the precision, recall, F-1 score, and support, for the traditional system without tuned hyperparameters are shown in Table 9. The confusion matrix for the type-I error and type-II error of the model with tuned hyperparameters is shown in Table 10. The total number of instances was 1025.

Table 9. Classification report of various types of training and testing of traditional models.

Traditional Models			Normal (0)	Abnormal (1)	A (%)	Macro Avg	Weighted Avg
LR	Training result	P (%)	91	89	90	90	90
		R (%)	87	92	90	89	90
		F-1s (%)	89	90	90	89	90
		S (%)	340	377	90	717	717
	Testing result	P (%)	85	89	82	82	82
		R (%)	89	85	82	82	82
		F-1s (%)	82	82	82	82	82
		S (%)	159	149	0.82	308	308
K-NN	Training result	P (%)	91	92	92	92	92
		R (%)	91	92	92	92	92
		F-1s (%)	91	92	92	92	92
		S (%)	340	377	92	717	717
	Testing result	P (%)	86	78	82	82	82
		R (%)	87	87	82	82	82
		F-1s (%)	81	82	82	82	82
		S (%)	159	149	82	308	308
SVM	Training result	P (%)	97	94	95	96	95
		R (%)	93	97	95	95	95
		F-1s (%)	95	96	95	95	95
		S (%)	340	377	95	717	717
	Testing result	P (%)	94	87	90	90	91
		R (%)	86	95	90	90	90
		F-1s (%)	90	99	90	90	90
		S (%)	159	149	90	308	308

Table 9. *Cont.*

Traditional Models			Normal (0)	Abnormal (1)	A (%)	Macro Avg	Weighted Avg
DT	Training result	P (%)	100	100	100	100	100
		R (%)	100	100	100	100	100
		F-1s (%)	100	100	100	100	100
		S (%)	340	377	100	717	717
	Testing result	P (%)	95	100	97	88	97
		R (%)	100	94	97	88	97
		F-1s (%)	97	97	97	88	97
		S (%)	159	149	0.97	308	308
RFC	Training result	P (%)	100	100	100	100	100
		R (%)	100	100	100	100	100
		F-1s (%)	100	100	100	100	100
		S (%)	340	377	100	717	717
	Testing result	P (%)	96	100	98	98	98
		R (%)	100	96	98	98	98
		F-1s (%)	98	98	98	98	98
		S (%)	159	149	98	308	308
XGB	Training result	P (%)	100	100	100	100	100
		R (%)	100	100	100	100	100
		F-1s (%)	100	100	100	100	100
		S (%)	340	377	100	717	717
	Testing result	P (%)	96	100	98	98	98
		R (%)	100	96	98	98	98
		F-1s (%)	98	98	98	98	98
		S (%)	159	149	98	308	308

Table 10. Performance evaluation and comparison of the confusion matrix for the training and testing datasets.

Models	Training Dataset			Testing Dataset		
	Confusion Matrix	Type-I Error	Type-II Error	Confusion Matrix	Type-I Error	Type-II Error
LR	[[295 45] [30 347]]	642 (correct)	75 (incorrect)	[[125 34] [22 127]]	252 (correct)	56 (incorrect)
KNN	[[310 30] [29 348]]	658 (correct)	59 (incorrect)	[[123 36] [20 129]]	252 (correct)	56 (incorrect)
SVM	[[317 23] [10 367]]	784 (correct)	33 (incorrect)	[[137 22] [8 141]]	278 (correct)	30 (incorrect)
DT	[[340 0] [0 377]]	717 (correct)	0 (incorrect)	[[159 0] [9 143]]	302 (correct)	9 (incorrect)
RFC	[[340 0] [0 377]]	717 (correct)	0 (incorrect)	[[159 0] [6 143]]	302 (correct)	6 (incorrect)
XGB	[[340 0] [0 377]]	717 (correct)	0 (incorrect)	[[159 0] [6 143]]	302 (correct)	6 (incorrect)

5.5. Performance Evaluation with Tuned Hyperparameters for Dataset-II

A grid search was employed in this recommended method to locate the ideal hyperparameters. The classifying models were constructed once the hyperparameters had been adjusted. For both the training and test datasets for each of the six machine learning methods, Figure 9 shows the outcomes of the suggested system with tuned hyperparameters in terms of accuracy, precision, recall, and F-1 score. Table A4 of Appendix B contains detailed information about Figure 10.
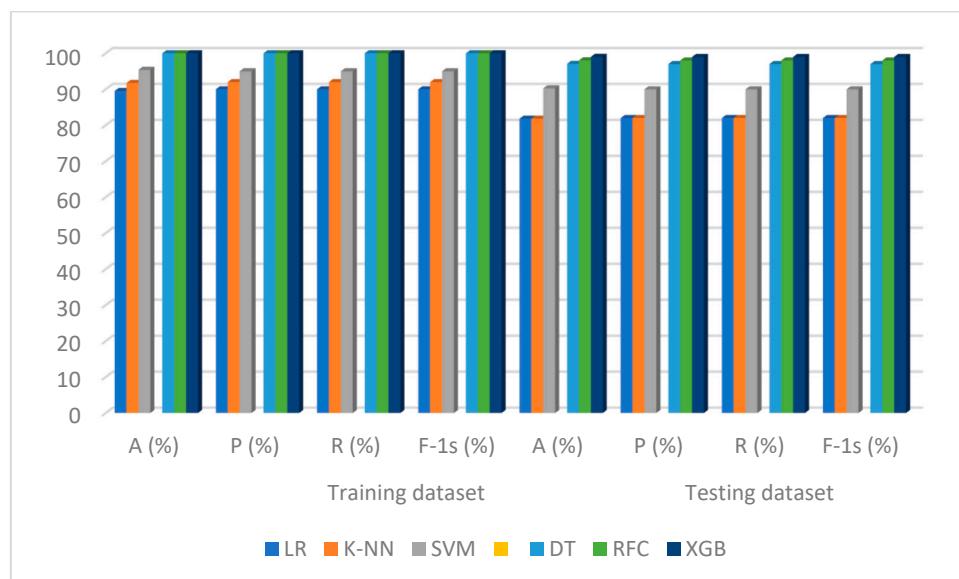
**Figure 9.** Graphical representation of the performance evaluation of the traditional system.
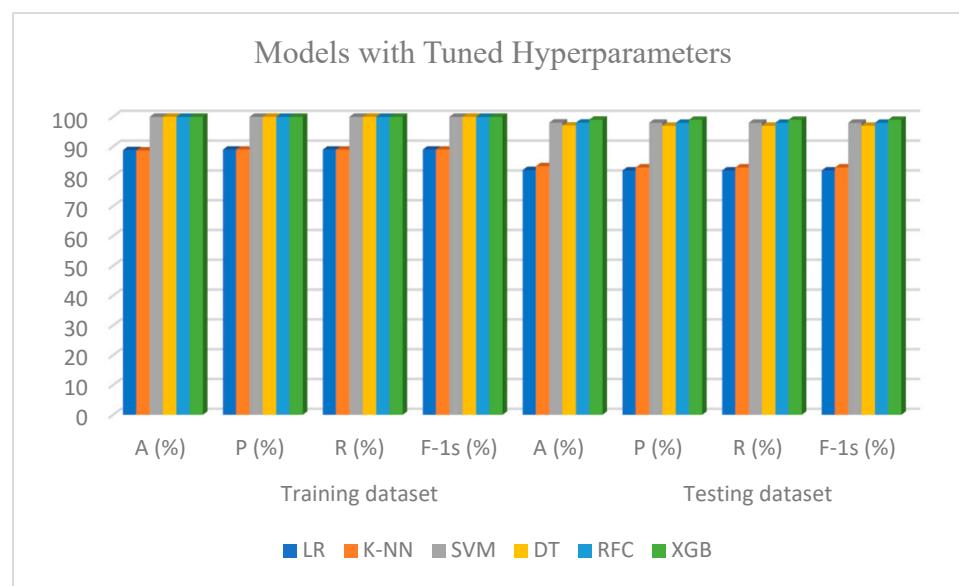
Figure 10. The bar plots comparing the accuracies of the training and testing datasets.

The classification report of several types of training and testing results, namely, precision, recall, F-1 score, and support, for the recommended model with tuned hyperparameters are shown in Table 11. The confusion matrix for the type-I error and type-II error of the suggested model with tuned hyperparameters is shown in Table 12. The total number of instances was 1025.

Table 11. For the training set and the test set, classification report models were evaluated and compared using a hyperparameter tuning strategy.

Hyper Parameter Tuning Classification Report			Normal (0)	Abnormal (1)	Macro Avg	Weighted Avg	A (%)
Tuned LR	Training result	P (%)	89	89	89	89	89
		R (%)	87	90	89	89	89
		F-1s (%)	88	90	89	89	89
		S (%)	340	377	717	717	89
	Testing result	P (%)	86	79	83	86	82
		R (%)	78	87	82	86	82
		F-1s (%)	82	82	82	86	82
		S (%)	159	149	308	308	82
Tuned K-NN	Training result	P (%)	90	88	89	89	89
		R (%)	86	92	89	89	89
		F-1s (%)	88	89	89	89	89
		S (%)	340	377	717	717	89
	Testing result	P (%)	90	78	84	84	83
		R (%)	77	91	84	83	83
		F-1s (%)	83	84	83	83	83
		S (%)	159	149	308	308	83
Tuned SVM	Training result	P (%)	100	100	100	100	100
		R (%)	100	100	100	100	100
		F-1s (%)	100	100	100	100	100
		S (%)	340	377	717	717	100
	Testing result	P (%)	96	100	98	98	98
		R (%)	100	96	98	98	98
		F-1s (%)	98	98	98	98	98
		S (%)	159	149	308	308	98
Tuned DT	Training result	P (%)	100	100	100	100	100
		R (%)	100	100	100	100	100
		F-1s (%)	100	100	100	100	100
		S (%)	340	377	717	717	100
	Testing result	P (%)	95	100	97	97	97
		R (%)	100	94	97	97	97
		F-1s (%)	97	97	97	97	97
		S (%)	159	149	717	717	97
Tuned RFC	Training result	P (%)	100	100	100	100	100
		R (%)	100	100	100	100	100
		F-1s (%)	100	100	100	100	100
		S (%)	340	377	717	717	100
	Testing result	P (%)	96	100	98	98	98
		R (%)	100	96	98	98	98
		F-1s (%)	98	98	98	98	98
		S (%)	159	149	308	308	98

Hyper Parameter Tuning Classification Report			Normal (0)	Abnormal (1)	Macro Avg	Weighted Avg	A (%)
Tuned XGB	Training result	P (%)	100	100	100	100	100
		R (%)	100	100	100	100	100
		F-1s (%)	100	100	100	100	100
		S (%)	340	377	717	717	100
	Testing result	P (%)	98	82	99	99	99
		R (%)	100	80	99	99	99
		F-1s (%)	99	81	99	99	99
		S (%)	159	149	308	308	99

Table 12. Performance evaluation and comparison of the confusion matrix after using a hyperparameter tuning approach on the training and testing datasets.

	Training Dataset			Testing Dataset		
Models	Confusion Matrix	Type-I Error (Correct)	Type-II Error (Incorrect)	Confusion Matrix	Type-I Error (Correct)	Type-II Error (Incorrect)
TLR	[[296 44] [36 341]]	637	80	[[124 35] [20 129]]	253	55
TKNN	[[291 49] [32 345]]	636	81	[[122 37] [14 135]]	257	51
TSVM	[[340 0] [0 377]]	717	0	[[159 0] [6 143]]	302	6
TDT	[[340 0] [0 377]]	717	0	[[159 0] [9 143]]	302	9
TRFC	[[340 0] [6 143]]	483	6	[[159 0] [6 143]]	302	6
TXGB	[[340 0] [0 377]]	717	0	[[159 0] [3 146]]	305	3

The performance of the suggested method was compared with that of the existing system in Figure 11. Figure 11 is extensively described in Table A5 of Appendix B.

5.6. Comparison of the Performance between Dataset-I and Dataset-II

ML algorithms were applied to the two datasets. Through the analysis described in the previous sections, the diagnostic systems were able to evaluate dataset-I with an accuracy that exceeded the evaluation of dataset-II during the testing phase. Figure 12 describes the analytical results to compare the performance of the ML algorithms on these two datasets. Table A6 of Appendix C contains information pertaining to Figure 12. The performance of the six used ML algorithms (LR, K-NN, SVM, DT, RFC, and XGB) with tuned hyperparameters on dataset-II during the training phase reached accuracies of 88.84%, 88.70%, 100%, 100%, 100%, and 100% respectively. In the testing phase, their accuracies were 82.14%, 83.44%, 98.05%, 97.08%, 98.05%, and 99.03%, respectively. The performances of these ML algorithms (LR, K-NN, SVM, DT, RFC, and XGB) with tuned hyperparameters on dataset-I during the training phase reached accuracies of 85.85%, 81.13%, 87.74%, 89.62%, 86.19%, and 99.06%, respectively. In the testing phase, their accuracies were 85.71%, 87.91%, 84.62%, 61.32%, 84.62%, and 79.12% respectively.
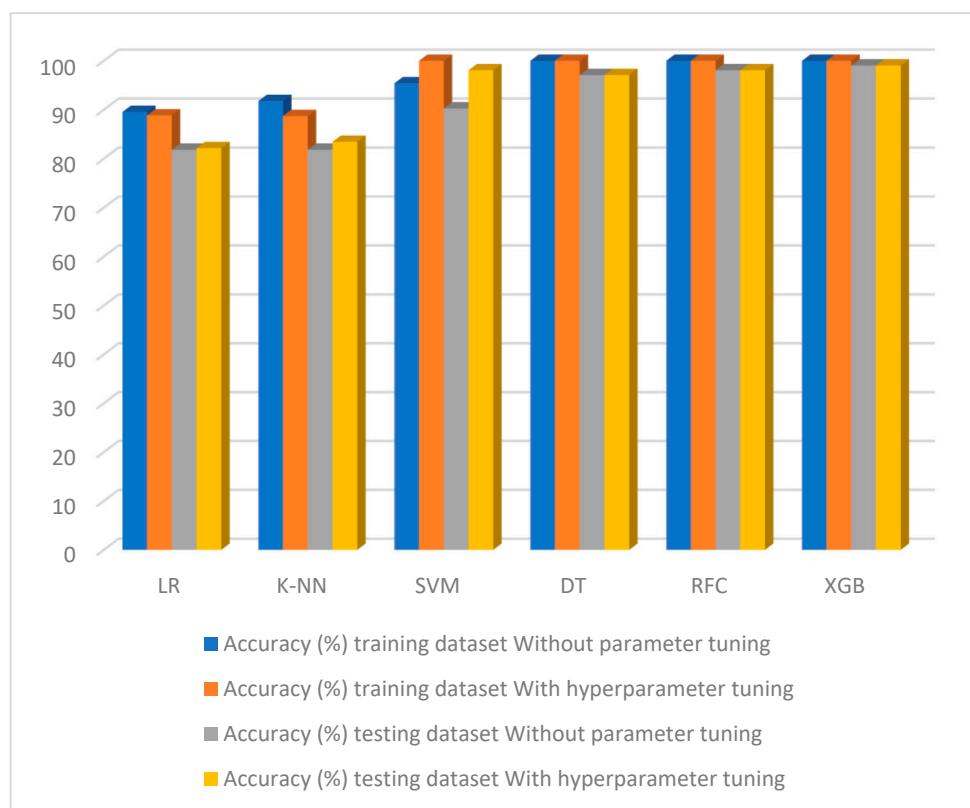
**Figure 11.** The graphical comparison of the accuracies of training and testing.
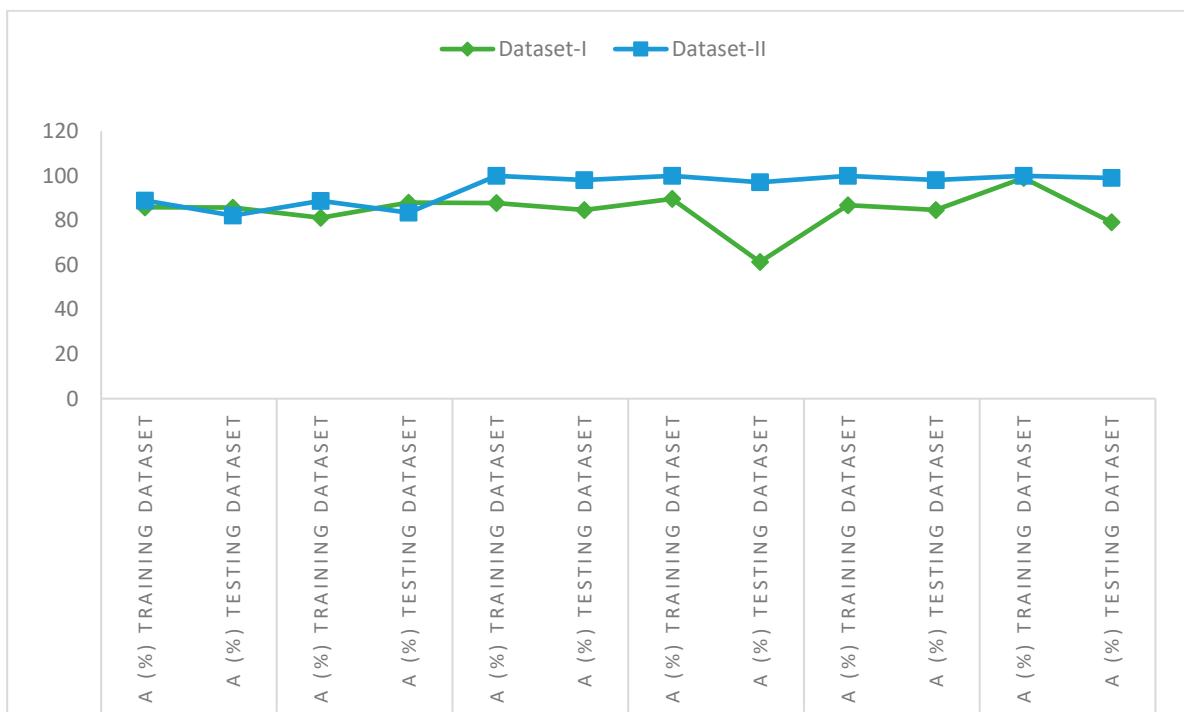
Figure 12. Comparison of the system performance regarding the diagnostic accuracies when using the two datasets.

5.7. Comparison with Previous Research

The evaluation of ML algorithms discussed on various criteria and compared with pertinent earlier works is described in Figure 13. Information relating to Figure 13 is

provided in Table A7 of Appendix C. It was mentioned that various criteria were used to evaluate earlier research. The accuracy of the recommended system was 100% using the training dataset and 99.03% using the testing dataset, while all prior research only managed to achieve accuracies ranging between 95% and 77.40%. The precision of the earlier investigations ranged between 97.62% and 78.15%, whereas the suggested method achieved 100% using the training dataset and 99% using the testing dataset.
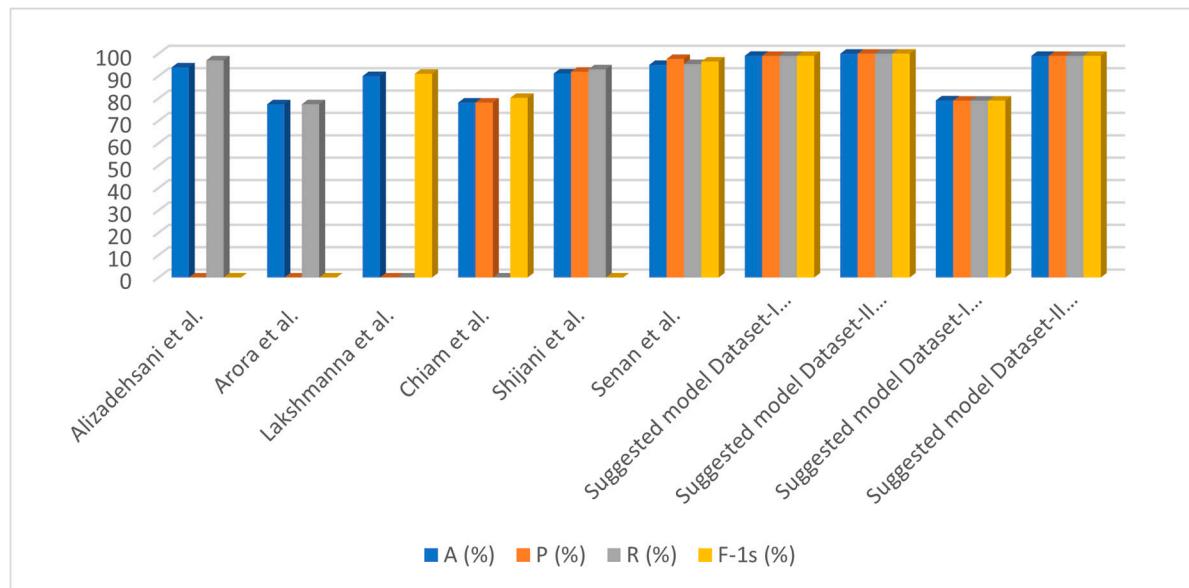
Figure 13. Graphical representation of the performance evaluation of the traditional system [35–40].

6. Conclusions

In this work, standard methods were used to predict heart disease from UCI Kaggle Cleveland datasets. In all situations, the heart disease prediction model was created using machine learning classifiers, namely, logistic regression, K-nearest neighbor (K-NN), support vector machine (SVM), decision tree, random forest, and extreme gradient boosting classifiers. These models primarily consist of six essential steps, although the suggested model was modified from the established model in terms of fine-tuning the hyperparameters. The accuracy rates when using the testing dataset for above used machine learning classifiers without hyperparameter tuning were found to be 86.91%, 86.81%, 87.91%, 78.02%, 82.42%, and 82.42, respectively, using dataset-I (UCI Kaggle Cleveland dataset). However, the accuracy rates with tuned hyperparameters for the same six classifiers were found to be 85.71%, 87.91%, 84.62%, 81.32%, 84.62%, and 79.12%, respectively, on the same dataset. The recommended model differs from the traditional model in terms of tuning the hyperparameters. The accuracy rates when using the testing dataset for the same six classifiers without tuned hyperparameters on dataset-II (Comprehensive UCI Kaggle Cleveland dataset) were found to be 81.82%, 81.82%, 90.26%, 97.08%, 98.05%, and 99.03%, respectively. However, the accuracy rates with tuned hyperparameters for the same six classifiers were found to be 82.14%, 83.44%, 98.05%, 97.08%, 98.05%, and 99.03%, respectively, on the same comprehensive dataset. Therefore, it was demonstrated through experimentation that the recommended models were more effective and may increase the accuracy of heart disease prediction. By developing a new model and a special model creation approach, the major goal of this work was to expand on prior work while making the model applicable to and simple to use in real-world circumstances.

The next phase of this study will involve creating a model using the feature selection strategy while utilizing various optimization techniques.

Appendix A. The Dataset-I Detailed Results

Table A1. For the training and test datasets, classification models were evaluated and compared.

Traditional Method		Training Dataset				Testing Dataset			
Models	Parameters	A (%)	P (%)	R (%)	F-1s (%)	A (%)	P (%)	R (%)	F-1s (%)
LR	Solver = liblinear	86.79	87	87	87	86.91	87	87	87
K-NN	K = 5, weights = uniform	86.79	87	87	87	86.81	87	87	87
SVM	Kernel = "rbf", gamma = 0.1, C = 1.0	93.40	93	93	93	87.91	88	88	88
DT	Random_state = 42	100	100	100	100	78.02	78	78	78
RFC	n_estimators = 1000, random_state = 42	100	100	100	100	82.42	82	82	82
XGB	Label_encoder = false	100	100	100	100	82.42	82	82	82

Table A2. The classification model results for the training and testing datasets using a hyperparameter tuning strategy.

Models with Hyper Parameters		Training Dataset				Testing Dataset			
MLAs	Tunned Hyperparameters	A (%)	P (%)	R (%)	F-1s (%)	A (%)	P (%)	R (%)	F-1s (%)
LR	Solver = liblinear, c = 0.234	85.85	86	86	86	85.71	86	86	86
K-NN	K = 27,	81.13	81	81	81	87.91	88	88	88
SVM	Kernel = rbf, gamma = 0.1, C = 5	87.74	88	88	88	84.62	85	85	85
DT	Criterion = entropy, max_depth = 5, min_samples_leaf = 2, splitter = 2	89.62	90	90	90	81.32	81	81	81
RFC	Max_depth = 2, max_features = auto, min_samples_leaf = 1, n_estimators = 1100	86.79	87	87	86	84.62	85	85	85
XGB	Learning_rate = 0.6427, max_depth = 3, n_estimators = 3	99.06	99	99	99	79.12	79	79	79

Table A3. Comparison of the accuracy of the training and testing datasets with and without tuned hyperparameters.

ML Classifiers	Accuracy (%) of Training Dataset		Accuracy (%) of Testing Dataset	
	Without Parameter Tuning	With Hyperparameter Tuning	Without Parameter Tuning	With Hyperparameter Tuning
LR	86.79	85.85	86.81	85.71
K-NN	86.79	81.13	86.81	87.91
SVM	93.40	87.74	87.91	84.62
DT	100.00	89.62	78.06	61.32
RFC	100.00	86.79	82.42	84.62
XGB	100.00	99.06	82.42	79.12

Appendix B. The Dataset-II Detailed Results

Table A4. For the training set and test set, classification models were evaluated and compared using a hyperparameter tuning strategy.

Models with Tuned Hyperparameters		Training Dataset				Testing Dataset			
		A (%)	P (%)	R (%)	F-1s (%)	A (%)	P (%)	R (%)	F-1s (%)
LR	Solver = liblinear, c = 0.088	88.84	89	89	89	82.14	82	82	82
K-NN	K = 27,	88.70	89	89	89	83.44	83	83	83
SVM	Kernel = rbf, gamma = 0.5, C = 2	100	100	100	100	98.05	98	98	98
DT	Criterion = entropy, max_depth = 11, min_samples_leaf = 1, splitter = best, min_samples_split = 2	100	100	100	100	97.08	97	97	97
RFC	Max_depth = 15, max_features = auto, min_samples_leaf = 1, min_samples_split = 2, n_estimators = 500	100	100	100	100	98.05	98	98	98
XGB	Learning_rate = 0.547, max_depth = 5, n_estimators = 338	100	100	100	100	99.03	99	99	99

Table A5. Comparison of the accuracy when using the training and testing datasets with and without tuned hyperparameters.

MLAs	Accuracy (%) Training Dataset		Accuracy (%) Testing Dataset	
	Without Parameter Tuning	With Hyperparameter Tuning	Without Parameter Tuning	With Hyperparameter Tuning
LR	89.54	88.84	81.82	82.14
K-NN	91.77	88.70	81.82	83.44
SVM	**95.40**	**100**	90.26	98.05
DT	100.00	100	97.08	97.08
RFC	100.00	100	98.05	98.05
XGB	100.00	100	**99.03**	**99.03**

Appendix C. Comparisons of Two Datasets and Previous Studies Details

Table A6. Accuracy (A) of the diagnosis of two datasets using six machine learning techniques.

Models with Hyperparameter		Dataset-I	Dataset-II
LR	A (%) training dataset	85.85	88.84
	A (%) testing dataset	85.71	82.14
K-NN	A (%) training dataset	81.13	88.70
	A (%) testing dataset	87.91	83.44
SVM	A (%) training dataset	87.74	100
	A (%) testing dataset	84.62	98.05
DT	A (%) training dataset	89.62	100
	A (%) testing dataset	61.32	97.08
RFC	A (%) training dataset	86.79	100
	A (%) testing dataset	84.62	98.05
XGB	A (%) training dataset	99.06	100
	A (%) testing dataset	79.12	**99.03**

Table A7. Comparison of the performances between the suggested system and previous studies.

Previous Studies	A (%)	P (%)	R (%)	F-1s (%)
Alizadehsani et al. [38]	93.85	00	97	00
Arora et al. [39]	77.4	00	77.4	0
Lakshmanna et al. [40]	90	00	00	91
Chiam et al. [41]	78.15	78.15	00	80.25
Shijani et al. [42]	91.14	91.90	93	00
Senan et al. [43]	95	97.62	95.35	96.47
Suggested model dataset-I training	**99.06**	**99**	**99**	**99**
Suggested model dataset-II training	**100**	**100**	**100**	**100**
Suggested model dataset-I testing	**79.12**	**79**	**79**	**79**
Suggested model dataset-II testing	**99.03**	**99**	**99**	**99**

References

1. Animesh, H.; Subrata, K.M.; Amit, G.; Arkomita, M.; Mukherje, A. Heart Disease Diagnosis and Prediction Using Machine Learning and Data Mining Techniques: A Review. *Adv. Comput. Sci. Technol.* **2017**, *10*, 2137–2159.
2. Buttar, H.S.; Li, T.; Ravi, N. Prevention of CVD: Role of exercise, dietary interventions, obesity and smoking cessation. *Exp. Clin. Cardiol.* **2005**, *10*, 229–249.
3. Ahmad, G.N.; Ullah, S.; Algethami, A.; Fatima, H.; Akhter, S.M.H. Comparative Study of Optimum Medical Diagnosis of Human Heart Disease Using ML Technique with and without Sequential Feature Selection. *IEEE Access* **2022**, *10*, 23808–23828. [CrossRef]
4. Nagamani, T.; Logeswari, S.; Gomathy, B. Heart Disease Prediction using Data Mining with Mapreduce Algorithm. *Int. J. Innov. Technol. Explor. Eng. (IJITEE)* **2019**, *8*, 137–140.
5. Nikhar, S.; Karandikar, A.M. Prediction of heart disease using machine learning algorithms. *Int. J. Adv. Eng. Manag. Sci.* **2016**, *2*, 617–621.
6. Franco, D.; Estefanía, L.V. Healing the Broken Hearts: A Glimpse on Next Generation Therapeutics. *Hearts* **2022**, *3*, 96–116. [CrossRef]
7. Gayathri, R.; Rani, S.U.; Čepová, L.; Rajesh, M.; Kalita, K. A Comparative Analysis of Machine Learning Models in Prediction of Mortar Compressive Strength. *Processes* **2022**, *10*, 1387. [CrossRef]
8. Brites, I.S.G.; da Silva, L.M.; Barbosa, J.L.V.; Rigo, S.J.; Correia, S.D.; Leithardt, V.R.Q. Machine Learning and IoT Applied to Cardiovascular Diseases Identification through Heart Sounds: A Literature Review. *Informatics* **2021**, *8*, 73. [CrossRef]

9. Reddy, K.V.V.; Elamvazuthi, I.; Aziz, A.A.; Paramasivam, S.; Chua, H.N.; Pranavanand, S. An Efficient Prediction System for Coronary Heart Disease Risk Using Selected Principal Components and Hyperparameter Optimization. *Appl. Sci.* **2023**, *13*, 118. [CrossRef]

10. Obaido, G.; Ogbuokiri, B.; Swart, T.G.; Ayawei, N.; Kasongo, S.M.; Aruleba, K.; Mienye, I.D.; Aruleba, I.; Chukwu, W.; Osaye, F.; et al. An interpretable machine learning approach for hepatitis b diagnosis. *Appl. Sci.* **2022**, *12*, 11127. [CrossRef]

11. UCI Machine Learning Repository: Heart Disease Dataset. Available online: https://www.kaggle.com/johnsmith88/heart-disease-dataset (accessed on 20 October 2022).

12. Detrano, R.; Janosi, A.; Steinbrunn, W.; Pfisterer, M.; Schmid, J.; Sandhu, S.; Guppy, K.; Lee, S.; Froelicher, V. International application of a new probability algorithm for the diagnosis of coronary artery disease. *Am. J. Cardiol.* **1989**, *64*, 304–310. [CrossRef] [PubMed]

13. Ebiaredoh-Mienye, S.A.; Swart, T.G.; Esenogho, E.; Mienye, I.D. A machine learning method with filter-based feature selection for improved prediction of chronic kidney disease. *Bioengineering* **2022**, *9*, 350. [CrossRef] [PubMed]

14. Mienye, I.D.; Sun, Y.; Wang, Z. An improved ensemble learning approach for the prediction of heart disease risk. *Inform. Med. Unlocked* **2020**, *20*, 100402. [CrossRef]

15. Polat, K.; Güneş, S. A hybrid approach to medical decision support systems: Combining feature selection, fuzzy weighted pre-processing and AIRS. *Comput. Methods Programs Biomed.* **2007**, *88*, 164–174. [CrossRef]

16. Alizadehsani, R.; Hosseini, M.J.; Khosravi, A.; Khozeimeh, F.; Roshanzamir, M.; Sarrafzadegan, N.; Nahavandi, S. Non-invasive detection of coronary artery disease in high-risk patients based on the stenosis prediction of separate coronary arteries. *Comput. Methods Programs Biomed.* **2018**, *162*, 119–127. [CrossRef]

17. Pham, H.; Olafsson, S. Bagged ensembles with tunable parameters. *Comput. Intell.* **2019**, *35*, 184–203. [CrossRef]

18. Magesh, G.; Swarnalatha, P. Optimal feature selection through a cluster-based DT learning (CDTL) in heart disease prediction. *Evol. Intell.* **2021**, *14*, 583–593. [CrossRef]

19. Wang, H.; Wang, S. Medical Knowledge Acquisition through Data Mining. In Proceedings of the IEEE International Symposium on IT in Medicine and Education, Xiamen, China, 12–14 December 2008; pp. 777–780. [CrossRef]

20. Singh, R.; Rajesh, E. Prediction of Heart Disease by Clustering and Classification Techniques. *Int. J. Comput. Sci. Eng.* **2019**, *7*, 861–866. [CrossRef]

21. Bharti, R.; Khamparia, A.; Shabaz, M.; Dhiman, G.; Pande, S.; Singh, P. Prediction of Heart Disease Using a Combination of Machine Learning and Deep Learning. *Comput. Intell. Neurosci.* **2021**, *2021*, 8387680. [CrossRef]

22. Manikandan, S. Heart Attack Prediction System. In Proceedings of the International Conference on Energy, Communication, Data Analytics & Soft Computing, Chennai, Tamil Nadu, 1–2 August 2017; pp. 817–820.

23. Garg, R.; Oh, E.; Naidech, A.; Kording, K.; Prabhakaran, S. Automating ischemic stroke subtype classification using machine learning and natural language processing. *J. Stroke Cerebrovasc. Dis.* **2019**, *28*, 2045–2051. [CrossRef]

24. Chourasia, V.; Pal, S. Data Mining Approach to Detect HDs. *Inter. J. Adv. Comput. Sci. Inf. Technol. (IJACSIT)* **2013**, *2*, 56–66.

25. Palaniappan, S.; Awang, R. Intelligent heart disease prediction system using data mining techniques. In Proceedings of the IEEE/ACS International Conference on Computer Systems and Applications, Doha, Qatar, 31 March–4 April 2008; pp. 108–115. [CrossRef]

26. Pedretti, R.F.E.; Hansen, D.; Ambrosetti, M.; Back, M.; Berger, T.; Ferreira, M.C.; Cornelissen, V.; Davos, C.H.; Doehner, W.; Zarzosa, C.D.P.Y.; et al. How to optimize the adherence to a guideline-directed medical therapy in the secondary prevention of cardiovascular diseases: A clinical consensus statement from the European Association of Preventive Cardiology. *Eur. J. Prev. Cardiol.* **2022**, *30*, 149–166. [CrossRef]

27. Chowdhury, M.T.H. Application of Signal Processing and Deep Hybrid Learning in Phonocardiogram and Electrocardiogram Signals to Detect Early-Stage Heart Diseases. Ph.D. Thesis. Faculty of the Computational Science Program Middle Tennessee State University: Murfreesboro, TN, USA, 2022.

28. Nadakinamani, R.G.; Reyana, A.; Kautish, S.; Vibith, A.S.; Gupta, Y.; Abdelwahab, S.F.; Mohamed, A.W. Clinical Data Analysis for Prediction of Cardiovascular Disease Using Machine Learning Techniques. *Comput. Intell. Neurosci.* **2022**, *2022*, 2973324. [CrossRef]

29. Fitriyani, N.L.; Syafrudin, M.; Alfian, G.; Rhee, J. HDPM: An Effective Heart Disease Prediction Model for a Clinical Decision Support System. *IEEE Access* **2020**, *8*, 133034–133050. [CrossRef]

30. Ali, Y.A.; Awwad, E.M.; Al-Razgan, M.; Maarouf, A. Hyperparameter Search for Machine Learning Algorithms for Optimizing the Computational Complexity. *Processes* **2023**, *11*, 349. [CrossRef]

31. Ambrish, G.; Ganesh, B.; Ganesh, A.; Srinivas, C.; Mensinkal, K. Logistic Regression Technique for Prediction of Cardiovascular Disease. *Glob. Transit. Proc.* **2022**, *4*, 127–130.

32. Zhang, C.; Zhong, P.; Liu, M.; Song, Q.; Liang, Z.; Wang, X. Hybrid Metric K-Nearest Neighbor Algorithm and Applications. *Math. Probl. Eng.* **2022**, *2022*, 8212546. [CrossRef]

33. Xue, T.; Jieru, Z. Application of Support Vector Machine Based on Particle Swarm Optimization in Classification and Prediction of Heart Disease. In Proceedings of the IEEE 7th Inter. Conference on Intelligent Computing and Signal Processing (ICSP), Virtual, 15–17 April 2022; pp. 857–860.

34. Vijaya Saraswathi, R.; Gajavelly, K.; Kousar Nikath, A.; Vasavi, R.; Reddy Anumasula, R. Heart Disease Prediction Using Decision Tree and SVM. In Proceedings of the Second International Conference on Advances in Computer Engineering and Communication Systems, Singapore, 11–12 August 2022; pp. 69–78.

35. Liu, Y.; Wang, Y.; Zhang. New machine learning algorithm: Random Forest. In Proceedings of the International Conference on Information Computing and Applications, Chengde, China, 14–16 September 2012; Springer: Berlin/Heidelberg, Germany, 2012; pp. 246–252.

36. Budholiya, K.; Shrivastava, S.K.; Sharma, V. An optimized XGBoost based diagnostic system for effective prediction of heart disease. *J. King Saud Univ. Comput. Inf. Sci.* **2020**, *34*, 4514–4523. [CrossRef]

37. Jeng, M.-Y.; Yeh, T.-M.; Pai, F.-Y. A Performance Evaluation Matrix for Measuring the Life Satisfaction of Older Adults Using eHealth Wearables. *Healthcare* **2022**, *10*, 605. [CrossRef]

38. Arabasadi, Z.; Alizadehsani, R.; Roshanzamir, M.; Moosaei, H.; Yarifard, A.A. Computer aided decision making for heart disease detection using hybrid neural network-Genetic algorithm. *Comput. Methods Programs Biomed.* **2017**, *141*, 19–26. [CrossRef]

39. Arora, S.; Maji, S. Decision tree algorithms for prediction of heart disease. In *Proceedings of the Information and Communication Technology for Competitive Strategies: Proceedings of Third International Conference on ICTCS 2017*; Springer: Singapore, 2019; pp. 447–454.

40. Lakshmanna, K.; Reddy, G.T.; Reddy, M.P.; Rajput, D.S.; Kaluri, R.; Srivastava, G. Hybrid genetic algorithm and a fuzzy logic classifier for heart disease diagnosis. *Evol. Intell.* **2020**, *13*, 185–196.

41. Chiam, Y.K.; Amin, M.S.; Varathan, K.D. Identification of significant features and data mining techniques in predicting heart disease. *Telemat. Inform.* **2019**, *36*, 82–93.

42. Feshki, M.G.; Shijani, O.S. Improving the heart disease diagnosis by evolutionary algorithm of PSO and Feed Forward Neural Network. In Proceedings of the 2016 Artificial Intelligence and Robotics (IRANOPEN), Qazvin, Iran, 9 April 2016; pp. 48–53.

43. Senan, E.M.; Abunadi, I.; Jadhav, M.E.; Fati, S.M. Score and Correlation Coefficient-Based Feature Selection for Predicting Heart Failure Diagnosis by Using Machine Learning Algorithms. *Comput. Math. Methods Med.* **2021**, *2021*, 8500314. [CrossRef] [PubMed]