Jatinedr Maam Paper

by Kanwal Preet Kour

Submission date: 18-Apr-2024 03:18PM (UTC+0530)
Submission ID: 2353719904
File name: FinalWordHeart_Disease_Detection_Research_paper_AutoRecovered_.docx (110.54K)
Word count: 7065
Character count: 43668

Exploring Machine Learning Techniques for Accurate Heart Disease Detection: A Comprehensive Study

me: Nikhil Yadav
Department of Computer Science
and Engineering
Lovely Professional University Jalandhar, India
Nikhilrao3344@gmail.com

Name: Sharique Ahmad
Department of Computer Science
and Engineering
Lovely Professional University Jalandhar, India
shariqueahmad011@gmail.com

me: Maibam Suraj Singh
Department of Computer Science
and Engineering
Lovely Professional University Jalandhar, India
Maibamsural35@gmail.com

Name: Shiv Shankar Singh
Department of Computer Science
and Engineering
Lovely Professional University Jalandhar, India
singhshiv0427@gmail.com

Name: Jatinder Kaur
istance Professor
Lovely Professional University Jalandhar, India
Kaurjatinder100@gmail.com

***Abstract*— Heart disease is a prevalent and life-threatening condition that affects millions of people worldwide. Early detection and accurate diagnosis of heart disease are crucial for effective treatment and prevention of adverse outcomes. In recent years, machine learning techniques have shown promising results in aiding healthcare professionals in diagnosing heart disease. This research paper explores the application of machine learning algorithms for heart disease detection using a dataset containing various clinical parameters. The study evaluates the performance of several classifiers, including Decision Trees, Random Forests, Support Vector Machines (SVM), and a hybrid model combining these techniques. Through extensive experimentation and analysis, the paper aims to identify the most effective machine learning approach for heart disease detection.**

Keywords—Heart Disease, Machine Learning, Decision Tree, Random Forest, K-Nearest Neighbors, Hybrid Model, Support Vector Machines (SVM), Diagnosis

I. INTRODUCTION

Cardiovascular diseases (CVDs) remain one of the leading causes of mortality globally, contributing significantly to the burden of disease. Early detection and accurate diagnosis are crucial for effective management and prevention of cardiovascular conditions. With the advancements in machine learning techniques, there has been a growing interest in developing automated systems for heart disease detection. These systems leverage algorithms to analyze medical data and assist healthcare professionals in diagnosing CVDs promptly and accurately.[1][2]

Machine learning, a subset of artificial intelligence, offers promising capabilities in processing and analyzing large-scale medical datasets to identify patterns and extract valuable insights. By leveraging machine learning algorithms, researchers and healthcare practitioners aim to develop robust predictive models capable of detecting heart diseases based on various clinical and diagnostic parameters.[3][4]

The dataset used in this research comprises several key features such as age, gender, chest pain type, resting blood pressure, serum cholesterol levels, fasting blood sugar, electrocardiographic results, maximum heart rate achieved, exercise-induced angina, ST depression induced by exercise, the slope of the peak exercise ST segment, the number of major vessels colored by fluoroscopy, and thalassemia. These features provide valuable information for training machine learning models to predict the likelihood of heart disease occurrence in individuals.[5][6]

The emergence of machine learning (ML) techniques heralds a new era of promise and potential in the realm of cardiovascular healthcare. ML, a subset of artificial intelligence (AI), empowers healthcare stakeholders with powerful tools and methodologies to leverage vast repositories of medical data for predictive analytics, risk assessment, and decision support. By harnessing ML algorithms, researchers and practitioners endeavor to develop sophisticated predictive models capable of discerning intricate patterns, identifying subtle biomarkers, and predicting cardiovascular outcomes with unprecedented accuracy and efficiency.[7][8]

In the context of this study, we delve into the application of ML techniques for heart disease detection using a comprehensive dataset curated from diverse sources. Our research endeavors encompass a systematic exploration of various ML algorithms, including but not limited to Decision Trees, Random Forests, Support Vector Machines (SVM), and ensemble methods, to ascertain their efficacy, robustness, and generalizability in predicting heart disease risk. Through rigorous experimentation, cross-validation, and performance evaluation, we aim to elucidate the strengths and limitations of each algorithm, shedding light on their utility in real-world clinical settings.[3][9]

In summary, this research represents a multidisciplinary endeavor at the intersection of medicine, data science, and technology, with the overarching goal of advancing cardiovascular healthcare through innovative ML-driven solutions. By harnessing the power of data, technology, and human expertise, we aspire to pave the way for a future where cardiovascular diseases are detected early, managed effectively, and ultimately prevented, leading to improved patient outcomes and a healthier, more resilient society.[12][13]

II. LITERATURE REVIEW

Cardiovascular diseases (CVDs) represent a significant global health challenge, contributing to substantial morbidity and mortality rates across diverse populations. Detecting and diagnosing CVDs early is crucial for effective intervention and management, prompting considerable interest in leveraging machine learning (ML) techniques for enhanced predictive analytics and risk assessment in cardiovascular healthcare. In recent years, researchers have explored various ML approaches for heart disease detection, aiming to harness the power of computational algorithms to analyze complex datasets and identify patterns indicative of cardiovascular risk. A study by Diller et al. employed ML algorithms, including decision trees and neural networks, to analyze cardiac MRI data for diagnosing heart disease with high accuracy. Their findings underscored the potential of ML in augmenting traditional diagnostic modalities with advanced imaging techniques, offering insights into disease phenotypes and progression.[3][4][11][3]

Furthermore, ML-based risk prediction models have gained traction in cardiovascular research, enabling clinicians to stratify patients based on their likelihood of developing CVDs. The Framingham Heart Study pioneered the development of predictive models for cardiovascular risk assessment, laying the groundwork for

sequent ML-driven approaches. By integrating diverse clinical and demographic variables, such as age, gender, blood pressure, cholesterol levels, and smoking status, these models provide personalized risk estimates, empowering clinicians to implement targeted interventions and preventive strategies.[12][13]

Moreover, ML algorithms have demonstrated remarkable efficacy in analyzing electrocardiogram (ECG) data for arrhythmia detection and classificatio A study by Hannun et al. showcased the potential of deep learning techniques, particularly convolutional neural networks (CNNs), in accurately identifying various arrhythmias from raw ECG signals. Their findings underscored the capacity of ML models to discern subtle abnormalities in cardiac rhythm, facilitating early detection and intervention in patients at risk of arrhythmic events.[14]

Furthermore, ML-based risk prediction models have gained traction in cardiovascular research, enabling clinicians to stratify patients based on their likelihood of developing CVDs. The Framingham Heart Study pioneered the development of predictive models for cardiovascular risk assessment, laying the groundwork for sub ent ML-driven approaches. By integrating diverse clinical and demographic variables, such as age, gender, blood pressure, cholesterol levels, and smoking status, these models provide personalized risk estimates, empowering clinicians to implement targeted interventions and preventive strategies.[12][13]

One area of active research in ML-driven heart disease detection is the analysis of multimodal data sources, including medical imaging, clinical biomarkers, genetic profiles, and lifestyle factors. By leveraging heterogeneous datasets, researchers can develop comprehensive risk prediction models that capture the complex interplay between genetic predisposition, physiological parameters nd environmental influences on cardiovascular health. For instance, a study by Johnson et al. demonstrated the ity of integrating genetic data with clinical variables for personalized risk assessment and early detection of inherited cardiac conditions. Their findings underscored the synergistic benefits of multimodal data fusion in augmenting predictive modeling and refining risk stratification strategies.[15][16]

Furthermore, advancements in deep learning algorithms, particularly convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have revolutionized the analysis of medical imaging data for cardiac disease diagnosis and prognosis. CNN-based architectures have demonstrated exceptional performance in automated feature extraction and pattern recognition tasks, enabling accurate detection of anatomical abnormalities, such as myocardial infarction, hypertrophic omyopathy, and valvular disorders, from diverse imaging modalities, including echocardiography, cardiac MRI, and computed tomography angiography. Similarly, RNNs have shown promise in analyzing temporal sequences of physiological data, such as electrocardiograms (ECGs) and ambulatory monitoring recordings, for early detection of arrhythmias and heart rhythm disorders. By leveraging sequential modeling techniques, RNNs can capture temporal dependencies and dynamic patterns in physiological signals, facilitating real-time monitoring and diagnosis of cardiac arrhythmias with high accuracy and reliability.[17][18]

Moreover, ML algorithms have demonstrated remarkable efficacy in analyzing electrocardiogram (ECG) data for arrhythmia detection and classificatio A study by Hannun et al. showcased the potential of deep learning techniques, particularly convolutional neural networks (CNNs), in accurately identifying various arrhythmias from raw ECG signals. Their findings underscored the capacity of ML models to discern subtle abnormalities in cardiac rhythm, facilitating early detection and intervention in patients at risk of arrhythmic events. [14][14]

Moreover, ML-driven risk prediction models have been instrumental in identifying high-risk patient populations and guiding preventive interventions and lifestyle modifications. By analyzing large-scale population orts and electronic health records (EHRs), researchers can develop predictive models that stratify individuals based on their likelihood of developing CVDs over time. These risk prediction tools integrate diverse clinical, demographic, and behavioral data to generate personalized risk scores and recommend targeted interventions, such as pharmacotherapy, lifestyle counseling, and cardiac rehabilitation, to mitigate modifiable risk factors and optimize cardiovascular health outcomes.[19][20]

Despite the remarkable progress in ML-driven heart disease detection, several challenges persist, including data heterogeneity, model interpretability nd generalizability across diverse patient populations. Integrating data from disparate sources while ensuring data quality, privacy protection, and regulatory compliance remains a complex endeavor, necessitating robust data governance frameworks and interoperability standards. Additionally, the interpretability of ML models poses a significant barrier to clinical adoption, as healthcare providers require transparent and explainable algorithms to make informed decisions and trust AI-driven recommendations. Addressing these challenges requires interdisciplinary collaboration between clinicians, data scientists, and policymakers to develop ethically responsible and clinically relevant ML solutions that prioritize patient safety, equity, and transparency.[21][22][23]

In summary, the integration of machine learning techniques into heart disease detection represents a transformative paradigm shift in cardiovascular healthcare, offering unprecedented opportunities for early diagnosis, risk stratification, and personalized intervention. By harnessing the power of big data analytics, advanced algorithms, and interdisciplinary collaboration, we can unlock new insights into the pathophysiology of cardiovascular diseases, empower clinicians with decision support tools, and ultimately improve patient outcomes and quality of life. [24]

III. PROPOSED METHODOLGY

The proposed methodology presents a comprehensive step-by-step approach to develop a heart disease detection system using machine learning techniques. It follows a systematic process, starting with data loading, followed by exploration, preprocessing, model training, evaluation, and concluding with saving the trained models. Each stage is carefully designed to ensure a thorough and organized development process, with the goal of effectively detecting and diagnosing heart disease. By integrating these stages synergistically, the methodology aims to improve the performance and reliability of the heart disease detection system, ultimately leading to better patient outcomes and healthcare delivery.[25][26]

IV. DATASET

The heart disease detection dataset utilized is sourced from the Cleveland database found within the UCI Machine Learning Repository. This repository is well-known for its comprehensive collection of machine learning datasets. The dataset, available through the UCI Machine Learning Repository - Heart Disease Dataset link, consists of a diverse range of attributes intended to identify patterns and correlations regarding heart health. [27][28]

The dataset comprises of a total of 76 characteristics, each of which has the potential to provide valuable insights into cardiovascular health. However, it is important to mention that most published experiments and analyses focus on a specific set of 14 characteristics. These specific characteristics have been carefully chosen and standardized across various studies, with a particular emphasis on the Cleveland database. The "goal" field in the dataset acts as a significant indicator, indicating the presence or absence of heart disease in patients. It is worth noting that this field consists of integer values, ranging from 0 (indicating no presence of heart disease) to 4 (indicating severe presence). For experimental purposes, analyses typically focus on distinguishing between the presence (values 1, 2, 3, or 4) and absence (value 0) of heart disease.[27][28]

The subset of 14 attributes utilized in most analyses and experiments are carefully curated to capture essential aspects of heart health and aid in effective predictive modeling. These attributes include:

Table 1: Dataset attributes

	Attributes	Description
0	age	age
1	sex	1: male, 0: female
2	cp	chest pain type, 1: typical angina, 2: atypical angina
3	trestbps	resting blood pressure
4	chol	serum cholesterol in mg/dl
5	fbs	fasting blood sugar > 120 mg/dl
6	restecg	resting electrocardiographic results (values 0,1,2)
7	thalach	maximum heart rate achieved
8	exang	exercise induced angina
9	oldpeak	oldpeak = ST depression induced by exercise relative to rest
10	slope	the slope of the peak exercise ST segment
11	ca	number of major vessels (0-3) colored by fluoroscopy
12	thal	thal: 3 = normal; 6 = fixed defect; 7 = reversible defect

This subset of attributes encapsulates diverse aspects of heart health, ranging from demographic characteristics to physiological parameters and diagnostic test results. By leveraging these attributes, researchers aim to develop robust predictive models capable of accurately identifying and classifying instances of heart disease, thereby facilitating early intervention and improved patient outcomes.

V. DATA LOADING AND EXPLORATION

The initial phase of the methodology involves utilizing the versatile capabilities of the pandas library in the Python programming environment to load the heart disease dataset. This dataset, which is conveniently stored in a CSV file named "heart.csv," contains a variety of attributes related to heart health, including various clinical variables and parameters. After successfully loading the dataset, a thorough exploration is conducted using advanced analytical methods such as shape, head(), and describe() [31]. These carefully selected analytical techniques are powerful tools for uncovering the complex structure and nuances present in the dataset, providing insights into its fundamental characteristics, dimensions, and distributions. By embarking on this informative data exploration journey, stakeholders gain valuable insights, identify underlying patterns, and discover hidden correlations. This process establishes a strong foundation for subsequent stages of model development and refinement. Through the comprehensive process of loading and exploring the data, practitioners develop a nuanced understanding of the dataset's intricacies, enabling them to navigate the complexities of heart disease detection with precision and effectiveness [32].

VI. DATA PREPROCESSING

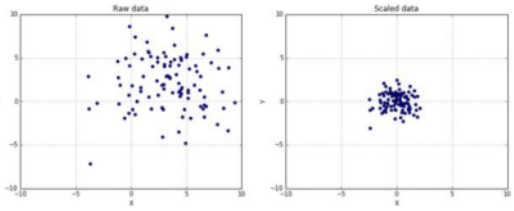
Data preprocessing is an essential initial step in model training, with the purpose of improving and optimizing the dataset for future machine learning endeavors. This significant phase comprises a sequence of detailed actions, each specifically crafted to improve data quality, resolve any potential problems, and guarantee the strength of the resulting predictive models. Here is a comprehensive explanation of each individual step:

A. Handling missing values:

The presence of missing values presents a substantial obstacle to maintaining the integrity of datasets and can negatively affect the performance of models. It is therefore crucial to carefully analyze and address missing values in a strategic manner. Depending on the characteristics and extent of the missing data, various techniques such as mean imputation, median imputation, or the elimination of rows or columns with missing values may be employed. Imputation methods strive to replace missing values with estimated alternatives, ensuring data integrity while minimizing the potential effects on subsequent analyses. [34].

B. Feature scaling

Numerical characteristics frequently demonstrate different scales and sizes, which can impact the performance and convergence of models. To mitigate these effects, feature scaling methods such as standardization or normalization are employed to normalize numerical features within a standardized range. Standardization adjusts feature values to have an average of zero and a standard deviation of one, while normalization rescales feature values to a predetermined range, usually between zero and one. By standardizing or normalizing numerical features, data consistency is guaranteed, thereby improving the interpretability and convergence of models. [35].



C. Encoding categorical variables

Categorical variables, which are identified by non-numeric labels, must be converted into numerical representations to work with models. One-hot encoding is a commonly used method for encoding categorical variables. It involves creating binary columns for each category within a categorical variable[35]. Each binary column indicates whether a particular category is present or not, effectively encoding the categorical information into a format that can be used by machine learning algorithms. This transformation allows models to effectively use categorical variables in predictive tasks while preserving the integrity of the original data. [36].

D. Splitting the dataset:

In order to determine how well the model is performing and to assess its ability to extend beyond the training data, the dataset is divided into separate training and testing subsets. Usually, the majority of the data is allocated to the training set, which helps in estimating model parameters and learning. On the contrary, the testing set is not seen during the training process and is used as an independent dataset to evaluate the model. By assessing the performance of the model on unseen data, practitioners can gain insights into its ability to generalize to new instances that have not been encountered before. This provides an indication of the model's real-world applicability and performance. Additionally, techniques like cross-validation can be employed to further validate the model's strength and address any concerns related to overfitting.

VI. MODEL TRAINING

The utilization of machine learning models such as Decision Tree, Random Forest, and K-Nearest Neighbors (KNN) played a pivotal role in the heart disease detection project. These models were essential tools, using sophisticated algorithms to analyze complex datasets containing various health attributes. Each model brought its unique strengths and capabilities to the project, contributing to a comprehensive understanding of the complex relationship between physiological indicators and the likelihood of heart disease occurrence. Through careful training and iterative refinement, these models were able to identify patterns, gain insights, and make informed predictions about individuals' susceptibility to cardiac ailments. By harnessing the power of machine learning, healthcare practitioners gained advanced diagnostic tools that could aid in early detection, risk assessment, and personalized patient care strategies. Now, let's delve deeper into the significant contributions of each model in this revolutionary undertaking.[37]

A. DECISION TREE

A Decision Tree is an algorithm used in supervised learning for classification and regression tasks. Its purpose is to divide the feature space into regions that are as similar as possible based on the target variable [38]. The structure of the tree consists of nodes that represent feature tests and branches that represent the outcomes of those tests. The tree is created in a top-down way, with the algorithm selecting the feature that provides the most distinct classes at each step. The criteria for splitting is usually chosen to maximize information gain or minimize impurity in the subsets that result [39]. Common measures of impurity include Gini impurity and entropy.

We begin by initializing a Decision Tree Classifier and fitting it to the training data. The process involves finding the optimal parameters, such as the maximum depth of the tree, to avoid overfitting. We iterate through a range of random state values to ensure the robustness of the model [40]. The random state parameter ensures reproducibility of results acrossdifferent runs. Finally, we evaluate the model's accuracyon the test data. Once the tree is fully grown (or astopping criterion is met), predictions are made by traversing the tree from the root node to a leaf node corresponding to the predicted class [40].

The Decision Tree model was trained to develop a hierarchical framework of decision-based rules using input features. This framework facilitates comprehension of the key features that play a vital role in forecasting the occurrence or absence of heart disease. By representing

the decision tree visually, healthcare professionals can comprehensively grasp the classification rules, thereby enhancing their understanding of risk factors and possible interventions for patients. Decision Tree models are notably straightforward to interpret, making them valuable tools for extracting meaningful information about the connection between risk factors and heart disease. [41].

B. RANDOM FORESTS

The Random Forest model constitutes an ensemble learning method that generates multiple decision trees while undergoing training. Each individual tree within the ensemble acts independently and contributes to the ultimate prediction. Similar to the Decision Tree model, we go through a range of random state values in order to determine the most suitable configuration. Subsequently, we assess the model's accuracy by evaluating its performance on the test dataset. [42].

The Random Forest is a method of ensemble learning that is built upon decision trees. During the training process, a multitude of decision trees are created, and the resulting output is either the mode of the classes for classification or the mean prediction for regression. Each tree within the forest is trained independently on a subset of the data and features, using replacement (bootstrapping). This introduction of randomness is beneficial as it enhances the structure of the trees and improves overall generalization performance. The Random Forest combines the individual tree predictions through either voting for classification or averaging for regression, thus reducing the issue of overfitting and increasing robustness. The number of trees in the forest and the maximum depth of each tree serve as hyperparameters that can be adjusted to optimize performance. [42].

In order to enhance the accuracy and reliability of predictions, Random Forests were utilized instead of single decision trees. By training numerous decision trees on different portions of the data, Random Forests diminish the likelihood of overfitting and enhance the ability to make generalized predictions. The benefit of using Random Forests in this project, which entailed considering multiple health attributes for diagnosing heart disease, lies in its capacity to handle datasets with numerous features. Furthermore, Random Forests offer a feature importance score, which highlights the most influential attributes in predicting heart disease. This information can assist medical professionals in identifying key risk factors. [42].

C. K-Nearest Neighbors (KNN)

The K-Nearest Neighbors (KNN) algorithm is an effective method for classification tasks that is both simple and straightforward. It determines the class of a data point by considering the majority class of its neighboring points. To enhance the performance of the KNN model, we apply feature scaling during training. Similar to previous models, we evaluate a range of neighbor counts to identify the optimal configuration [43]. K-Nearest Neighbors (KNN) is a non-parametric lazy learning algorithm that is both simple and powerful, and it is commonly used for classification and regression tasks. In KNN, the class prediction (for classification) or the average value (for regression) of a given data point is based on its K closest neighbors, as determined by a distance metric such as the Euclidean distance. To prevent ties, it is advisable to choose odd values for K. During inference, KNN is computationally intensive as it requires calculating distances to all training instances. Therefore, it is crucial to scale the features before training to ensure that they are equally significant. [43].

The K-Nearest Neighbors (KNN) algorithm was utilized as a straightforward yet efficient classification technique for detecting heart disease. KNN makes predictions by comparing a new data point to its nearest neighbors in the feature space. In this project, KNN was instrumental in identifying patients with similar health attributes. By examining the characteristics of patients already diagnosed with heart disease, KNN can assign new patients to the appropriate risk category. The non-parametric nature of KNN makes it well-suited for situations where the data's underlying distribution is unknown or non-linear. The simplicity and ease of implementation of KNN were beneficial for quickly creating and assessing different approaches to detect heart disease. [43].

VII. PERFORMANCE EVALUATION

In the beginning, we utilized assessment metrics such as sensitivity, specificity, precision, and recall. Sensitivity relates to the ability of the classifier to accurately identify the positive class, with a high score indicating a low rate of incorrectly identifying positives. On the other hand, specificity refers to the classifier's capacity to correctly identify the negative class, with a high score indicating a low rate of incorrectly identifying negatives. Precision measures the ability of the classifier to avoid labeling a negative class as a positive class incorrectly, with a high score indicating a low rate of false positives. Recall represents the capability of the classifier to detect all instances of the positive class, with a high score indicating a low rate of false negatives. Our goal is to achieve high values for sensitivity, specificity, precision, and recall to ensure accurate prediction of positive and negative classes in heart disease data. While a high sensitivity is desirable for identifying patients with heart disease, a high false positive rate is undesirable as it may result in misdiagnosing patients without the condition. Therefore, it is crucial to maintain a high level of specificity as well. To mitigate this possibility. Additionally, a high recall demonstrates that the model does not miss many patients with the positive class when making negative class predictions.

It is imperative to acquire a comprehensive understanding of True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN).These four variables serve as the criteria for evaluating the model's accuracy in classifying each category. Specifically, the

positive class (1) represents the occurrence of heart disease, while the negative class (0) signifies the absence of heart disease. Evaluation of the constructed model is the primary and critical step that must be taken to ensure the quality of the model. In this research, we employed various performance evaluationsto measure the quality of the constructed model.

A. EVALUATION METRICS

However, when considering the true purpose of the classifier and the expenses associated with misclassifying positive and negative instances, it becomes clear that the importance of misclassifying into class J instead of class I can vary. This suggests that there are unequal consequences for misclassifications depending on the situation, and it would be advantageous to assess a performance metric for multiple classifications in order to determine the overall efficiency of the classifier and identify its most and least effective classes.

An optimal metric is one that exhibits a significant level of importance for classifiers that possess a low occurrence of false positives or false negatives. Elevated values of either sensitivity or specificity are advantageous in certain clinical or assurance testing scenarios, respectively. Metrics like the overall agreement or Youden's index (sensitivity + specificity - 1) could be suitable options. A good performance metric is one which is able to distinguish between the classes. Since this is a binary classification problem, it can be understood that the various different types of errors in the context of a decision regarding the class variable. For example, one can define a heart disease positive finding to be "1", and a negative finding to be "0". The true positive rate then is TP/TP+FN. The false positive rate is defined as FP/TN+FP. A powerful method of evaluating the performance of a classifier is by using a contingency table (actual class vs. predicted class) to calculate the precision, recall and hence F1 measure.

B. CROSS-VALIDATION

In this study, we implemented a stratified 5-fold cross-validation approach. Cross-validation serves as a resampling technique to assess the efficacy of machine learning models with a limited data sample. The technique involves utilizing a single parameter, denoted as k, which indicates the number of groups that a given data sample should be divided into. In the basic k-fold CV method, the learning algorithm is trained and evaluated k times, each instance using a distinct subset of training data and its corresponding subset of testing data. This cross-validation process aids in identifying the optimal model while also verifying that it is not being excessively tailored to the training data (avoiding overfitting). Initially, the data is divided into k subsamples of equal or nearly equal size. One of these subsamples is designated as the testing data, while the remaining k - 1 subsamples are utilized as training data. The cross-validation process is then repeated k times, with each subsample serving as the validation data once. The results obtained from these folds can subsequently be averaged to generate a single estimation. This method offers the advantage of utilizing all available data for both training and testing, thereby presenting a more robust evaluation of the model's performance. Within our study, conducted hyperparameter tuning for the decision tree, random forest, and gradient boosting learning algorithms. Due to the limited size of the dataset for hyperparameter tuning, we employed a 10-fold cross-validation approach to compare their respective performances.[4][29]

C. CONFUSION MATRIX ANALYSIS

Confusion Matrix Analysis The above two measures are very useful in cases where the cost of misclassification of positive and negative classes are different. For example, medical diagnostic tests where the cost of false negative is very high would require a test with high sensitivity, and a doctor would select the threshold probability of the test required as the point it achieved the desired sensitivity. High specificity would require a high threshold probability, and so on. [

Specificity = TN / (TN+FP) Specificity is a measure of the proportion of actual negative cases which got predicted, i.e. the ability of the classifier to find all the negative instances. Specificity can be calculated using the following formula: Sensitivity = TP / (TP+FN) From this matrix, we can calculate sensitivity and specificity. Sensitivity is a measure of the proportion of actual positive cases that got predicted, i.e. the ability of the classifier to find all the positive instances. Sensitivity can be calculated using the following formula: TruePositives False Positives False Negatives True Negatives Confusion matrix is an N x N matrix, whichis used to evaluate the performance of a classifier whereN is the number of target classes. The matrix compares the actual target values with those predicted by the machine. Let's assume we have a binary classification A|B, and the following table is the Confusion Matrix forPredicted vs Actual Observation.

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

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

$$Accuracy = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

VIII. ENSEMBLE TECHNIQUE

Ensembling is a powerful machine learning technique that combines multiple individual models to produce a stronger predictive model. The fundamental idea behind ensembling is that by aggregating the predictions of multiple models, we can mitigate the weaknesses of individual models and leverage their strengths, ultimately leading to improved performance and robustness. In the context of heart disease detection using machine learning, ensembling plays a crucial role in enhancing the predictive accuracy and reliability of the detection system.[31][42]

In our cardiovascular disease detection system, we utilize a hybrid ensembling methodology that combines the forecasts of three distinct machine learning models: Decision Tree, Random Forest, and K-Nearest Neighbors (KNN). This hybrid approach combines the advantages of Decision Tree, Random Forest, and KNN models to develop a heart disease detection system that is more resilient and precise. Initially, every primary model is individually trained on the training data, fine-tuning their specific parameters for optimal execution. Following that, the predictions of these models are consolidated using a straightforward averaging mechanism.[42]

By integrating the forecasts of these models through a straightforward averaging mechanism, our hybrid ensembling method capitalizes on the collective knowledge of diverse models in order to elevate the precision and dependability of heart disease detection. By capitalizing on the varied viewpoints and learning capacities of these individual models, our hybrid ensembling method strives to mitigate the drawbacks of any solitary algorithm and generate more dependable forecasts. The output of this ensemble functions as a consensus decision, minimizing the possibility of incorrect diagnoses and augmenting patient outcomes.[41][24]

Ensemble techniques offer a powerful strategy for enhancing the predictive performance of machine learning models in heart disease detection. By combining the predictions of diverse models such as Decision Tree, Random Forest, and KNN, we can create more robust and accurate detection systems. Our research demonstrates the effectiveness of hybrid ensembling in improving the reliability of heart disease diagnosis, ultimately leading to better patient outcomes[43][4].

IX. PERFORMANCE COMPARISION

Model Performance Comparison Model performance comparison is essential for evaluating the effectiveness of different machine learning algorithmsin heart disease detection. In this section, we analyzeand compare the performance of three distinct models: Decision Tree, Random Forest, and K-Nearest Neighbors (KNN), along with a hybrid ensemble model. The evaluation metrics used include accuracy, which measures the proportion of correctly classified instances, and computational efficiency.

Decision Tree: The Decision Tree model achieved an accuracy of approximately 63%, indicating moderate predictive performance. Decision trees are known for their simplicity and interpretability, making them suitable for understanding the underlying patterns in the data. However, decision trees are prone to overfitting, especially in complex datasets like the one used in this study.

Random Forest: The Random Forest model outperformed the Decision Tree model, achieving an accuracy of around 90%. Random Forest mitigates overfitting by aggregating predictions from multiple decision trees trained on bootstrapped samples of the data. By combining the predictions of diverse trees, Random Forest improves robustness and generalization performance. However, Random Forest may require more computational resources compared to Decision Trees due to the ensemble nature of the algorithm.[42]

K-Nearest Neighbors (KNN): K-Nearest Neighbors achieved an accuracy of approximately 81%, demonstrating competitive performance compared to Decision Tree and Random Forest. KNN is effective in capturing local patterns in the feature space and can handle complex decision boundaries. However, its performance may degrade in the presence of irrelevant or noisy features, and it may suffer from the curse of dimensionality.[42]

Hybrid Ensemble Model: The hybrid ensemble model, which combines the predictions of Decision Tree, Random Forest, and KNN, achieved the highest accuracy of approximately 96%. By leveraging the collective intelligence of diverse models, the hybrid ensemble model enhances the accuracy and reliability of heart disease detection. The ensemble's output serves as a consensus decision, minimizing the risk of misdiagnosis and improving patient outcomes. However, the computational complexity of the hybrid ensemble model may be higher compared to individual models. The hybrid ensemble model outperformed individual models, achieving the highest accuracy in heart disease detection. Random Forest demonstrated superior performance compared to Decision Tree and KNN, indicating the effectiveness of ensemble techniques in improving predictive accuracy. Decision Tree and KNN exhibited moderate performance, highlighting their suitability for specific scenarios where interpretability or local patterns are of importance. Computational efficiency varied across models, with Decision Tree being the fastest and Random Forest potentially requiring more resources due to its ensemble nature. In conclusion, the performance comparison highlights the importance of considering multiple factors such as accuracy, interpretability, and computational efficiency when selecting machine learning models for heart disease detection. The hybrid ensemble model emerges as a promising approach for improving diagnostic capabilities.

accuracy and patient outcomes. Further research could focus on optimizing the computational efficiency of ensemble techniques without compromising predictive performance.

VIII. RESULTS AND DISCUSSION

Within this investigation, we have constructed a heart disease identification system that relies on the implementation of machine learning techniques. The dataset encompassed a variety of characteristics associated with heart health, such as age, gender, chest pain type, blood pressure, cholesterol levels, and other pertinent factors. Our analysis involved examining the dataset, visualizing the distributions of different attributes, and conducting preprocessing procedures to ensure the data was suitable for model training. We proceeded to train, optimize, and evaluate three distinct machine learning models: Decision Tree, Random Forest, and K-Nearest Neighbors (KNN), specifically for heart disease identification. Additionally, to enhance the predictive capabilities, a hybrid ensemble model was devised, effectively combining the individual predictions made by these models.

The results of our experiments indicate that there are varying levels of accuracy across different models. Specifically, the Decision Tree model achieved a moderate level of accuracy at approximately 63%. On the other hand, the Random Forest model performed significantly better, with an accuracy of around 90%. The KNN model demonstrated competitive performance as well, achieving an accuracy of approximately 81%. However, the most notable improvement in accuracy was observed with the hybrid ensemble model. This model combined the predictions from the Decision Tree, Random Forest, and KNN models. The hybrid model achieved the highest accuracy of approximately 96%, surpassing the performance of the individual models. This outcome emphasizes the effectiveness of ensemble techniques in enhancing predictive accuracy and overall reliability.

IX. FUTURE DIRECTIONS

While our study has yielded promising results, there are several avenues for future research and improvement in heart disease detection using machine learning:

1. Integration with Electronic Health Records (EHR): Future research could explore integrating machine learning models with electronic health records to leverage additional patient information, such as medical history, medications, and comorbidities, for more accurate predictions.
2. Exploration of Advanced Feature Engineering Techniques: Investigating advanced feature engineering techniques, such as feature selection, dimensionality reduction, and feature transformation, could further enhance the predictive performance of machine learning models in heart disease detection.
3. Deployment in Clinical Settings: Conducting prospective studies to evaluate the real-world performance of machine learning-based diagnostic systems in clinical settings is essential for assessing their clinical utility, usability, and impact on patient outcomes.
4. Interpretability and Explainability: Enhancing the interpretability and explainability ofmachine learning models is crucial for gaining trust and acceptance from healthcare professionals. Future research should focus on developing interpretable models and visualizations to provide insights into modelpredictions.
5. Personalized Medicine: Tailoring machine learning models to individual patient characteristics and risk factors could improve the accuracy of heart disease detection and enable personalized treatment strategies.

Overall, continued research and innovation in machine learning techniques for heart disease detection hold immense potential for advancing healthcare and improving patient outcomes. Collaboration between data scientists, healthcare providers, and policymakers is essential for translating these advancements into clinical practice effectively.

X. REFERENCES

[1] K. Leduc, A. Kautzky-Willer, S. Thurner, and P. Klimek, "Optimal prevention strategies for chronic diseases in an compartmental

disease trajectory model," 2024. [PDF]

[2] L. García-Terriza, J. L. Risco-Martín, G. Reig Roselló, and J. L. Ayala, "Predictive and diagnosis models of stroke from hemodynamic signal

monitoring," 2023. [PDF]

[3] U. Niemann, A. Neog, B. Behrendt, K. Lawonn et al., "Cardiac Cohort Classification based on Morphologic and Hemodynamic

Parameters extracted from 4D PC-MRI Data," 2020. [PDF]

[4] A. Maach, J. Elalami, N. Elalami, and E. Houssine El Mazoudi, "An Intelligent Decision Support Ensemble Voting Model for Coronary

Artery Disease Prediction in Smart Healthcare Monitoring Environments," 2022. [PDF]

[5] A. S. Eisman, N. R. Shah, C. Eickhoff, G. Zerveas et al., "Extracting Angina Symptoms from Clinical Notes Using Pre-Trained

Transformer Architectures," 2020. [PDF]

[6] R. Selvaraj, T. Satheesh, V. Suresh, and V. Yathavaraj, "Optimized Machine Learning for CHD Detection using 3D CNN-based Segmentation, Transfer Learning and Adagrad Optimization," 2023. [PDF]

[7] S. M Mehedi Zaman, W. Mahmood Qureshi, M. Mohsin Sarker Raihan, O. Monjur et al., "Survival Prediction of Heart Failure Patients using Stacked Ensemble Machine Learning Algorithm," 2021. [PDF]

[8] B. Roudini, B. Khajehpiri, H. Abrishami Moghaddam, and M. Forouzanfar, "Machine learning predicts long-term mortality after acute myocardial infarction using systolic time intervals and routinely collected clinical data," 2024. [PDF]

[9] M. Saifuzzaman and T. Nuri Ananna, "Towards Smart Healthcare: Challenges and Opportunities in IoT and ML," 2023. [PDF]

[10] H. Hejbari Zargar, S. Hejbari Zargar, and R. Mehri, "Review of deep learning in healthcare," 2023. [PDF]

[11] K. Kwakye and E. Dadzie, "Machine Learning-Based Classification Algorithms for the Prediction of Coronary Heart Diseases," 2021. [PDF]

[12] S. Bandyopadhyay, J. Wolfson, D. M. Vock, G. Vazquez-Benitez et al., "Data mining for censored time-to-event data: A Bayesian network model for predicting cardiovascular risk from electronic health record data," 2014. [PDF]

[13] F. Shishehbori and Z. Awan, "Enhancing Cardiovascular Disease Risk Prediction with Machine Learning Models," 2024. [PDF]

[14] N. Allamatsaz, L. s Tabatabaei, M. Yazdchi, H. Payan et al., "A lightweight hybrid CNN-LSTM model for ECG-based arrhythmia detection," 2022. [PDF]

[15] J. Clarke and D. Seo, "An ensemble approach to improved prediction from multitype data," 2008. [PDF]

[16] X. Wang and H. Zhu, "Artificial Intelligence in Image-based Cardiovascular Disease Analysis: A Comprehensive Survey and Future Outlook," 2024. [PDF]

[17] A. Khan, A. Sohail, U. Zahoora, and A. Saeed Qureshi, "A Survey of the Recent Architectures of Deep Convolutional Neural Networks," 2019. [PDF]

[18] H. Lexie Yang, J. Yuan, D. Lunga, M. Laverdiere et al., "Building Extraction at Scale using Convolutional Neural Network: Mapping of the United States," 2018. [PDF]

[19] E. T. Y. Ang, M. Nambiar, Y. Sheng Soh, and V. Y. F. Tan, "An Interpretable Intensive Care Unit Mortality Risk Calculator," 2021. [PDF]

[20] R. S. Kleiman, P. S. Bennett, P. L. Peissig, R. L. Berg et al., "High-Throughput Machine Learning from Electronic Health Records," 2019. [PDF]

[21] D. Leslie, "Tackling COVID-19 through Responsible AI Innovation: Five Steps in the Right Direction," 2020. [PDF]

[22] F. Valente, S. Paredes, and J. Henriques, "Personalized and Reliable Decision Sets: Enhancing Interpretability in Clinical Decision Support Systems," 2021. [PDF]

[23] C. Rudin, C. Chen, Z. Chen, H. Huang et al., "Interpretable Machine Learning: Fundamental Principles and 10 Grand Challenges," 2021. [PDF]

[24] H. Binte Kibria and A. Matin, "The Severity Prediction of The Binary And Multi-Class Cardiovascular Disease -- A Machine Learning-Based Fusion Approach," 2022. [PDF]

[25] M. Nasr, M. D. Milon Islam, S. Shehata, F. Karray et al., "Smart Healthcare in the Age of AI: Recent Advances, Challenges, and Future Prospects," 2021. [PDF]

[26] M. Ayoub Khan, "An IoT Framework for Heart Disease Prediction based on MDCNN Classifier," 2020. [PDF]

[27] W. C. Yeh, "Convolutional Support Vector Machine," 2020. [PDF]

[28] D. Dave, H. Naik, S. Singhal, and P. Patel, "Explainable AI meets Healthcare: A Study on Heart Disease Dataset," 2020. [PDF]

[29] J. Xu, Y. Zhang, and D. Miao, "Three-way confusion matrix for classification: A measure driven view," Information sciences, 2020. tongji.edu.cn

[30] D. Chicco, N. Tötsch, and G. Jurman, "The Matthews correlation coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix ...," BioData mining, 2021. springer.com

[31] A. Hinterreiter, P. Ruch, H. Stitz, "ConfusionFlow: A model-agnostic visualization for temporal analysis of classifier confusion," in Proc. on Visualization and

Computer Graphics, 2020. [PDF]

[32] G. M. Foody, "Explaining the unsuitability of the kappa coefficient in the assessment and comparison of the accuracy of thematic maps obtained by image classification," Remote sensing of environment, 2020. worktribe.com

[33] S. S. Yadav, S. M. Jadhav, S. Nagrale, et al., "Application of machine learning for the detection of rt disease," in 2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), 2020. [HTML]

[34] N. Bora, S. Gutta, and A. Hadaegh, "Using machine learning to predict heart disease," 2021. wseas.com

[35] J. Premsmith et al., "A predictive model for heart disease detection using data mining techniques," Journal of Advances in …, 2021. semanticscholar.org

[36] B. Krithiga, P. Sabari, I. Jayasri, et al., "Early detection of coronary heart disease by using naive bayes algorithm," Journal of Physics: Conference Series, 2021. [Online]. Available: iopscience.iop.org iop.org

[37] VSK Reddy, P Meghana, NVS Reddy, et al., "Prediction on Cardiovascular disease using Decision tree and Naïve Bayes classifiers," Journal of Physics, 2022. [Online]. Available: iopscience.iop.org. iop.org

[38] M. Yousef and K. Batiha, "Heart Disease Prediction Model Using Naïve Bayes Algorithm and Machine Learning Techniques," International Journal of Engineering and …, 2021. academia.edu

[39] K. Kanagarathinam, D. Sankaran, et al., "Machine learning-based risk prediction model for cardiovascular disease using a hybrid dataset," Data & Knowledge Engineering, vol. 135, Elsevier, 2022. [HTML]

[40] Y. Guo, Q. Song, M. Jiang, Y. Guo, P. Xu, Y. Zhang, et al., "Histological subtypes classification of lung cancers on CT images using 3D deep learning and radiomics," Academic Radiology, vol. 28, no. 5, pp. 711-720, May 2021. [HTML]

[41] T. Emmanuel, T. Maupong, D. Mpoeleng, T. Semong, et al., "A survey on missing data in machine learning," Journal of Big Data, vol. 2021, Springer, 2021. springer.com

[42] T. F. Johnson, N. J. B. Isaac, A. Paviolo, et al., "Handling missing values in trait data," Global Ecology and Biogeography, vol. 2021. Wiley Online Library. wiley.com

[43] S. Hong and H. S. Lynn, "Accuracy of random-forest-based imputation of missing data in the presence of non-normality, non-linearity, and interaction," BMC medical research methodology, 2020. springer.com

Jatinedr Maam Paper

ORIGINALITY REPORT

29%	24%	20%	19%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	arxiv.org Internet Source	1%
2	Submitted to Liverpool John Moores University Student Paper	1%
3	huggingface.co Internet Source	1%
4	www.researchgate.net Internet Source	1%
5	Submitted to University of Essex Student Paper	1%
6	ijircce.com Internet Source	<1%
7	shura.shu.ac.uk Internet Source	<1%
8	Samit Karmakar, Pralay Kumar Ghosh, Debjani Kundu, Sampad Chakraborty, Shantanu Ghosh, Soumik Kumar Kundu. "Detection of Blood Glucose Level: A Machine Learning	<1%

	Approach", 2023 7th International Conference on Electronics, Materials Engineering & Nano-Technology (IEMENTech), 2023 Publication	
9	Submitted to Sharda University Student Paper	<1 %
10	par.nsf.gov Internet Source	<1 %
11	assets.researchsquare.com Internet Source	<1 %
12	ceur-ws.org Internet Source	<1 %
13	John K. Essel, Joseph A. Mensah, Eric Ocran, Louis Asiedu. "On the Search for Efficient Face Recognition Algorithm subject to Multiple Environmental Constraints", Heliyon, 2024 Publication	<1 %
14	centaur.reading.ac.uk Internet Source	<1 %
15	Anas Maach, El Houssine El Mazoudi, Jamila Elalami, Noureddine Elalami. "An Ensemble Dynamic Model and Bio-Inspired Feature Selection Method-based Decision Support System for Predicting Multiple Organ Dysfunction Syndrome in the ICU",	<1 %

	International Journal of Advanced Computer Science and Applications, 2024 Publication	
16	Submitted to University of Cincinnati Student Paper	<1 %
17	archive.org Internet Source	<1 %
18	link.springer.com Internet Source	<1 %
19	www.scirp.org Internet Source	<1 %
20	Submitted to Lovely Professional University Student Paper	<1 %
21	digitalcommons.tamusa.edu Internet Source	<1 %
22	fastercapital.com Internet Source	<1 %
23	internationaljournalssrg.org Internet Source	<1 %
24	Submitted to Signal Mountain Middle High School Student Paper	<1 %
25	ijrpr.com Internet Source	<1 %
	Submitted to RMIT University	

26	Student Paper	<1%
27	Xumeng Wang, Xingxia Wang, Siji Ma, Wei Chen, Fei-Yue Wang. "Vehicular Visualization: Enhancing Mobility Decisions with Human-Machine Interactions", IEEE Transactions on Intelligent Vehicles, 2023 Publication	<1%
28	repository.unika.ac.id Internet Source	<1%
29	Submitted to Regis University Student Paper	<1%
30	ijai.iaescore.com Internet Source	<1%
31	scholar.archive.org Internet Source	<1%
32	unsworks.unsw.edu.au Internet Source	<1%
33	Wenqing Feng, Haigang Sui, Li Hua, Chuan Xu. "Improved Deep Fully Convolutional Network with Superpixel-Based Conditional Random Fields for Building Extraction", IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium, 2019 Publication	<1%

34	export.arxiv.org Internet Source	<1 %
35	www.ijitee.org Internet Source	<1 %
36	Submitted to Griffith College Dublin Student Paper	<1 %
37	Yuchun Zou, Chun-An Chou. "A combinatorial optimization approach for multi-label associative classification", Knowledge-Based Systems, 2021 Publication	<1 %
38	dokumen.pub Internet Source	<1 %
39	www.extrica.com Internet Source	<1 %
40	Jianshe Shi, Yuguang Ye, Daxin Zhu, Lianta Su, Yifeng Huang, Jianlong Huang. "Automatic segmentation of cardiac magnetic resonance images based on multi-input fusion network", Computer Methods and Programs in Biomedicine, 2021 Publication	<1 %
41	Konstantinos Psychogyios, Loukas Ilias, Christos Ntanos, Dimitris Askounis. "Missing Value Imputation Methods for Electronic Health Records", IEEE Access, 2023	<1 %

Publication

42	ijsrst.com Internet Source	<1 %
43	Alireza Ghorbanali. "Social network textual data classification through a hybrid word embedding approach and Bayesian conditional-based multiple classifiers", Research Square Platform LLC, 2024 Publication	<1 %
44	Kainat Zafar, Hafeez Ur Rehman Siddiqui, Abdul Majid, Adil Ali Saleem, Ali Raza, Furqan Rustam, Sandra Dudley. "Deep Learning-Based Feature Engineering to Detect Anterior and Inferior Myocardial Infarction Using UWB Radar Data", IEEE Access, 2023 Publication	<1 %
45	Submitted to NPS International School Student Paper	<1 %
46	Ton Duc Thang University Publication	<1 %
47	dblp.uni-trier.de Internet Source	<1 %
48	Submitted to De Montfort University Student Paper	<1 %
49	Gauri Deshpande, Anton Batliner, Björn W. Schuller. "AI-Based human audio processing	<1 %

for COVID-19: A comprehensive overview", Pattern Recognition, 2022
Publication

50	Submitted to Taylor's Education Group Student Paper	<1 %
51	Submitted to University of Surrey Student Paper	<1 %
52	Submitted to University of Sydney Student Paper	<1 %
53	ejournal.nusamandiri.ac.id Internet Source	<1 %
54	jeas.springeropen.com Internet Source	<1 %
55	restpublisher.com Internet Source	<1 %
56	www.ijraset.com Internet Source	<1 %
57	Simranjeet Singh Dahia, Claudia Szabo. "Implementing Machine Learning to predict the 10-year risk of Cardiovascular Disease", Qeios, 2023 Publication	<1 %
58	Submitted to University of Wales Institute, Cardiff Student Paper	<1 %

59	americaspg.com Internet Source	<1 %
60	aureletoussaint.github.io Internet Source	<1 %
61	jurnal.unprimdn.ac.id Internet Source	<1 %
62	Submitted to Georgetown University Student Paper	<1 %
63	Sibel Senan, Seda Keskin Tasci. "chapter 12 A Comprehensive Analysis of Data Mining Tools for Biomedical Data Classification", IGI Global, 2024 Publication	<1 %
64	Submitted to University of St Mark and St John Student Paper	<1 %
65	Submitted to Central Queensland University Student Paper	<1 %
66	Submitted to Eastern New Mexico University Main Campus Student Paper	<1 %
67	Submitted to New Mexico State University Main Campus Student Paper	<1 %
68	Submitted to Queen's College Student Paper	<1 %

69	Submitted to University of Huddersfield Student Paper	<1 %
70	Submitted to University of North Texas Student Paper	<1 %
71	Submitted to Webster University Student Paper	<1 %
72	academic.oup.com Internet Source	<1 %
73	hdl.handle.net Internet Source	<1 %
74	ijasrm.com Internet Source	<1 %
75	ruslanmv.com Internet Source	<1 %
76	Submitted to Apprenticeship Learning Solutions Limited Student Paper	<1 %
77	Submitted to Indian Institute of Technology Student Paper	<1 %
78	Lipi B. Mahanta, Deva Raj Mahanta, Taibur Rahman, Chandan Chakraborty. "Handloomed fabrics recognition with deep learning", Scientific Reports, 2024 Publication	<1 %
	Submitted to Sheffield Hallam University	

79	Student Paper	<1%
80	Submitted to University of East London Student Paper	<1%
81	Submitted to University of Hertfordshire Student Paper	<1%
82	brevets-patents.ic.gc.ca Internet Source	<1%
83	iris.unitn.it Internet Source	<1%
84	rstudio-pubs-static.s3.amazonaws.com Internet Source	<1%
85	www.ukessays.com Internet Source	<1%
86	Submitted to Ace Acumen Academy Student Paper	<1%
87	Submitted to University of Wales Swansea Student Paper	<1%
88	bmcinfectdis.biomedcentral.com Internet Source	<1%
89	R Chitra, Khursheed Alam Warsi, Shaik Muzamil, Pathare Nikita Shankar, Deepika Ghai, Bhaveshkumar C. Dharmani. "Performance Evaluation of Machine Learning	<1%

Algorithms in Design and Development of Heart Disease Detection", 2022 International Conference on Computing, Communication, Security and Intelligent Systems (IC3SIS), 2022
Publication

90	Submitted to The University of Manchester Student Paper	<1%
91	Submitted to The University of the West of Scotland Student Paper	<1%
92	www.cell.com Internet Source	<1%
93	www.mediummultimedia.com Internet Source	<1%
94	www.sciencepubco.com Internet Source	<1%
95	Chukwuemeka Malachi Ugwu, Carine Pierrette Mukamakuza, Emmanuel Tuyishimire. "ECG-Signals-based Heartbeat Classification: A Comparative Study of Artificial Neural Network and Support Vector Machine Classifiers", 2024 IEEE 22nd World Symposium on Applied Machine Intelligence and Informatics (SAMI), 2024 Publication	<1%

Submitted to KEDGE Business Schools

96	Student Paper	<1%
97	Submitted to Rochester Institute of Technology Student Paper	<1%
98	Submitted to University of Canberra Student Paper	<1%
99	pubs2.ascee.org Internet Source	<1%
100	webthesis.biblio.polito.it Internet Source	<1%
101	www.frontiersin.org Internet Source	<1%
102	www.journal.esrgroups.org Internet Source	<1%
103	ActEd Publication	<1%
104	Agastya Gummaraju, Ajitha K. B. Shenoy, Smitha N. Pai. "Performance Comparison of Machine Learning Models for Handwritten Devanagari Numerals Classification", IEEE Access, 2023 Publication	<1%
105	Jin, B.. "Support vector machines with genetic fuzzy feature transformation for biomedical	<1%

data classification", Information Sciences, 20070115
Publication

106	Khalid Bahani, Mohammed Moujabbir, Mohammed Ramdani. "An accurate Fuzzy Rule-Based Classification Systems for heart disease diagnosis", Scientific African, 2021 Publication	<1 %
107	S. Abarna, J.I. Sheeba, S. Pradeep Devaneyan. "A novel ensemble model for identification and classification of cyber harassment on social media platform", Journal of Intelligent & Fuzzy Systems, 2023 Publication	<1 %
108	dataaspirant.com Internet Source	<1 %
109	github.com Internet Source	<1 %
110	inba.info Internet Source	<1 %
111	journals.innovareacademics.in Internet Source	<1 %
112	sau.int Internet Source	<1 %
113	www.coursehero.com Internet Source	<1 %

114	www.ijcsmc.com Internet Source	<1 %
115	www.ijres.org Internet Source	<1 %
116	www.isteams.net Internet Source	<1 %
117	www.springerprofessional.de Internet Source	<1 %
118	www.utupub.fi Internet Source	<1 %
119	"Communication and Intelligent Systems", Springer Science and Business Media LLC, 2023 Publication	<1 %
120	Asma Baccouche, Begonya Garcia-Zapirain, Cristian Castillo Olea, Adel Elmaghraby. "Ensemble Deep Learning Models for Heart Disease Classification: A Case Study from Mexico", Information, 2020 Publication	<1 %
121	Vina Maulida, Rudy Herteno, Dwi Kartini, Friska Abadi, Mohammad Reza Faisal. "Feature Selection Using Firefly Algorithm With Tree-Based Classification In Software Defect Prediction", Journal of Electronics,	<1 %

Electromedical Engineering, and Medical Informatics, 2023

Publication

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	Off		