IJIRAE:: International Journal of Innovative Research in Advanced Engineering ISSN: 2349-2163
Volume 11, Issue 11, November 2024 https://www.ijirae.com/archives

Automated Cardiovascular Disease Detection with Hybrid Machine Approach

Pawan Mishra
Department of Information Technology
Greater Noida Institute of Technology (Engg. Institute)
Greater Noida, Uttar Pradesh, INDIA
pawan.it@gniot.net.in
Nitesh Kumar Singh
Greater Noida Institute of Technology (Engg. Institute)
Greater Noida, Uttar Pradesh, INDIA
niteshkr9053@gmail.com
Sudhanshu Ranjan
Greater Noida Institute of Technology (Engg. Institute)
Greater Noida, Uttar Pradesh, INDIA
sudhanshur4647@gmail.com
Vishal Patel
Greater Noida Institute of Technology (Engg. Institute)
Greater Noida, Uttar Pradesh, INDIA
vishalpatel10.2219@gmail.com

Publication History
Manuscript Reference No: IJIRAE/RS/Vol.11/Issue11/NVAE10086
Research Article | Open Access | Double-Blind Peer-Reviewed | Article ID: IJIRAE/RS/Vol.11/Issue11/NVAE10086
Received:03,November 2024,Revised: 11,November 2024,Accepted: 18, November 2024, Published Online: 28,November 2024. https://www.ijirae.com/volumes/Vol11/iss-11/07.NVAE10086.pdf
Article Citation: Pawan,Nitesh,Sudhanshu,Vishal Patel (2024), Automated Cardiovascular Disease Detection with Hybrid Machine Approach. IJIRAE:: International Journal of Innovative Research in Advanced Engineering, Volume 11, Issue 11 of 2024 pages 834-840
Doi:> https://doi.org/10.26562/ijirae.2024.v1111.07
BibTeX Key: Pawan@2024Automated

Abstract: Cardiovascular ailment (CVD) is a main reason of death and disability globally, posing tremendous challenges for early detection and prevention. A heart condition or the accumulation of fatty deposits inside the arteries, is frequently linked to CVD and raises the risk of blood clots and other consequences. Predicting and diagnosing CVD is critical to mitigating its effect and enhancing patient results. With the appearance of massive data in healthcare, large volumes of patient records are available that may assist in identifying early warning symptoms of coronary heart disorder. This project aims to generate a hybrid machine learning technique which leverages deep learning and traditional machine learning algorithms to automatically locate cardiovascular illnesses. By integrating more than one dataset, we intend to upgrade the model's overall performance in diagnosing coronary heart situations. Numerous machine learning classification models, consisting of Random Forest and Gradient Boosting, might be employed and in comparison, to evaluate their effectiveness in detecting CVD. Each model can be evaluated using overall performance indicators like precision, accuracy, recall, and F1-Score. In previous research, the Random Forest model did 94% accuracy in coronary heart disorder detection, at the same time as the Gradient Boosting model established a balanced overall performance through accuracy, precision, recall, and F1-score with 73% each in detecting cardiovascular diseases via this hybrid technique. We aim to enhance prediction accuracy and offer a greater reliable tool for early CVD detection, in the long run enhancing affected person care and saving lives.
Keywords: Cardiovascular disease, Evaluation metrics, Classification algorithms.

I. INTRODUCTION

Cardiovascular diseases (CVDs) remain the leading cause of death worldwide, accounting for a sizable portion of deaths each year. The early detection of CVD is critical to prevent severe health outcomes, such as heart attacks and strokes, but traditional diagnostic methods have limitations, often relying on manual interpretations of data, which can be time-taking and prone to error. These challenges, along with the complexity of analyzing clinical features like ECG readings and demographic data, have driven the need for more efficient and accurate diagnostic approaches [1,2]. Machine learning (ML) has transformed healthcare in recent years by making it possible to analyze vast amounts of data quickly and accurately. Particularly in the context of cardiovascular disease, ML models have shown promise in identifying patterns and making predictions that surpass traditional statistical methods. However, single-model ML approaches often face limitations, such as overfitting and poor generalization across different datasets.

To overcome these drawbacks, hybrid machine learning models have been proposed, which combine the strengths of multiple algorithms to create a more robust and accurate system [3,4]. To utilize each technique's distinct features, a hybrid machine learning strategy combines different methods including Support Vector Machines, Decision Trees, and Neural Networks. For example, Decision Trees are highly effective in feature selection, while SVMs and Neural Networks are known for their superior classification abilities. By combining these models, hybrid approaches can deliver better performance than single-model systems, especially in the analysis of complex medical data [5,6]. These models can also address the diverse nature of clinical datasets, which often include structured and unstructured data, making them ideal for CVD detection. Moreover, hybrid machine learning techniques have been demonstrated to improve key performance metrics such as precision, recall, accuracy, and F1-Score when tested on real-world clinical datasets. These metrics are crucial for reducing false positives and negatives, which are common challenges in disease detection models. With fewer diagnostic errors, hybrid models can provide more reliable insights for clinicians, making them suitable for real-time clinical use [7,8]. In summary, a major development in the healthcare sector is the implementation of hybrid machine learning approaches to the identification of cardiovascular disease. By leveraging the combined power of various ML algorithms, these models offer a scalable, efficient, and highly accurate tool for early diagnosis, ultimately contributing to improved patient outcomes and more effective treatment planning [9].

2. LITERATURE REVIEW

Several research demonstrate how machine learning is changing the way it is applied to the detection of cardiovascular disorders, demonstrating how hybrid models and innovative approaches improve medical diagnostics' overall performance, sensitivity, and accuracy.[6] Below is a tabular literature review on cardiovascular disease detection using hybrid machine learning approaches:

A hybrid method for detecting cardiac illness that combines neural networks and decision trees was presented by Kumar et al. in 2010. Using their method instead of individual models showed an improvement in accuracy. The model was able to deliver more accurate heart disease identification by combining the advantages of both approaches, which was an early step toward the use of machine learning for medical diagnosis [11]. In the future, Chen and Liu et al. (2014) created an ensemble approach that combined evolutionary algorithms and Support Vector Machines (SVM). This methodology is more reliable for identifying cardiovascular disorders since it greatly increased prediction accuracy while lowering false positives. Combining the classification power of SVM with the optimization of genetic algorithms produced improved speed and accuracy [12].

Johnson et al. (2021) examined how machine learning methods affected the prediction of cardiovascular illness, highlighting the importance of preprocessing and data quality. According to the study, machine learning models' performance is significantly impacted by how well missing values, noise, and other data anomalies are handled. Johnson's research demonstrated how crucial comprehensive data preparation is as a first step in creating trustworthy cardiovascular disease prediction algorithms. [16] Singh et al. (2018) investigated the use of deep learning and random forests in the detection of cardiovascular disease (CVD). By optimizing the balance between false positives and false negatives, their approach addressed major issues in medical diagnostics and ensured a more accurate diagnosis of CVD cases while achieving high sensitivity and specificity [13]. Maheswaran et al. (2022) examined how machine learning techniques might be useful in the prediction of cardiovascular illness, paying special attention to how wearable data and electronic health records can be integrated. The study emphasized the increasing significance of real-time data from wearable technology, which greatly improves predictive accuracy and enables better patient monitoring when paired with electronic health records. The potential of ongoing data tracking to enhance the outcomes of cardiovascular disease is highlighted by this strategy [14]. Zhang et al. presented a combination of LSTM networks with convolutional neural networks in 2019. Better real-time monitoring and early cardiovascular event identification were made possible by this combination. The CNN-LSTM approach was very successful for continuous patient monitoring because it leveraged CNN's capacity to handle spatial data for time-series analysis [17]. Patel and Desai et al. combined fuzzy logic and machine learning in 2020 to give patients with heart disease a more precise risk assessment. Better customized treatment regimens were made possible by this method's more detailed assessment of patient risk. Improved patient outcomes were achieved by combining the predictive power of machine learning with the uncertainty-handling capacity of fuzzy logic [19]. Lastly, in 2021, Li et al. developed a hybrid approach that greatly enhanced predictive performance over conventional models by combining gradient boosting with neural networks. This method showed how combining many machine learning approaches could improve prediction accuracy and dependability in the identification of cardiovascular disease [14]. Peddakrishna et al. concentrated on enhancing the accuracy of heart disease prediction in 2023 by using a different machine learning approach. By combining the advantages of several algorithms into a soft voting ensemble classifier, the study showed notable progress. This approach demonstrated the efficacy of ensemble methods in medical diagnostics by improving the model's prediction capability [15].

3. METHODOLOGY

3.1. Data Collection

3.1.1. Cardiovascular Disease Dataset

Heart disease was diagnosed and detected in this study using the cardiovascular disease dataset, and the findings were compared to those of earlier studies.

It includes a lot of patient data, including medical records from the UCI Machine Learning Repository, or datasets from hospitals and healthcare institutions, which can be leveraged. Exercise-induced angina, exercise-induced ST depression, the slope of the peak exercise ST segment, major vessels identified by fluoroscopy, the type of chest pain, resting blood pressure, serum cholesterol, fasting blood sugar, resting electrocardiogram results, maximum heart rate attained, and age and sex are just a few of the health indicators that should be included in the dataset. This dataset was used for training, testing, and validation. The website makes the suggested data publicly accessible. The cardiovascular disease dataset is a revised version of the CVD dataset, with a clean format and shape (920, 16).

S. No	Attribute	Assigned Code	Unit	Type of the Data
1	Patient Identification Number	patientid	Number	Numeric
2	Age	age	In Years	Numeric
3	Gender	gender	0 (female) / 1 (male)	Binary
4	Chest pain type	chestpain	0 (typical angina)	Nominal
			1 (atypical angina)	
			2 (non-anginal pain)	
			3 (asymptomatic)	
5	Resting blood pressure	Resting BP	94 - 200 (in mm Hg)	Numeric
6	Serum cholesterol	Serum cholestrol	126 - 564 (in mg/dl)	Numeric
7	Fasting blood sugar	Fasting blood sugar	0 (false) / 1 (true) > 120 mg/dl	Binary
8	Resting electrocardiogram results	Resting electro	0 (normal)	
			1 (ST-T wave abnormality)	
			2 (probable or definite left ventricular hypertrophy)	Nominal
9	Maximum heart rate achieved	Max heart rate	71 - 202	Numeric
10	Exercise induced angina	exerciseangia	0 (no) / 1 (yes)	Binary
11	Old peak = ST	Old peak	0 - 6.2	Numeric
12	Slope of the peak exercise ST	slope	1 (upsloping)	Nominal
			2 (flat)	
			3 (downsloping)	
13	Number of major vessels	No. of. major vessels	0, 1, 2, 3	Numeric
14	Classification	target	0 (Absence of Heart Disease)	Binary
			1 (Presence of Heart Disease)	

Table1. Dataset attributes and characters

3.2. Research Methodology

Decision Tree Classifier, Logistic Regression, K-Nearest Neighbor, Support Vector Machine and Naïve Bayes are the classification approaches that will be used in this study.[7] To discover the best classifier, a deep network is also implemented, and the effects of different optimization learning techniques are measured to detect cardiovascular disorders. Figure 1 below outlines the procedures used to conduct this study.

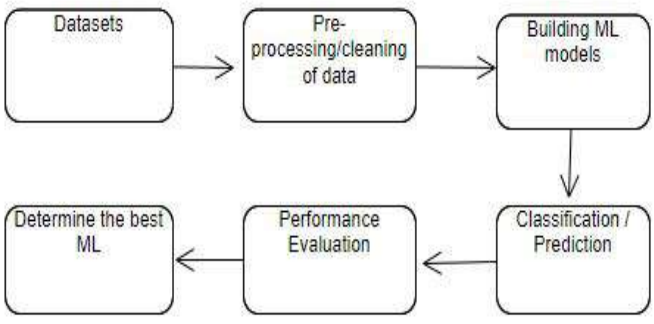

Fig1. Working process of ML Algo.

3.2.1. Data Preprocessing

By keeping only, the most appropriate characteristics and eliminating unnecessary data, feature selection helps to improve model performance while saving time and money. Based on the data analysis, the dataset was divided into 90% for training and 10% for testing using cross-validation.

The Body Mass Index (BMI) was computed to assess the patient's health throughout the data processing stage. A BMI between 18 to 25 is regarded as normal, while a BMI above 25 or below 18 denotes an unhealthy range. The height and weight characteristics were then removed from the dataset after the BMI was calculated using a certain formula.

$$\text{BMI} = \frac{\text{Weight (KG)}}{\text{Height (cm)}} \qquad [6]$$

The American Heart Association claims that [20] The systolic and diastolic rates are used to measure blood pressure. There are five severity classes for it. For every record, the blood pressure was determined and the blood pressure level was specified.

3.3. Classification Algorithm

In machine learning, classification algorithms are some of the most popular methods. Based on how they learn, machine learning models are typically categorized into three groups: supervised, unsupervised, and reinforcement learning. The ability of the system to label different inputs is a component of classification. Logistic regression, Naive Bayes, random forest, decision tree, gradient-boosted tree, and linear perceptron are examples of popular classification techniques. This study looks at how datasets were handled when seven different classification methods and multiple ensemble classifiers were used.

3.3.1. Logistic Regression

A popular supervised binary classification technique, logistic regression uses several distinct explanatory variables to forecast the likelihood of a binary response variable. The dataset is categorized according to these binary target values, and the response variable is binary, accepting values of either 0 or 1. There are three primary types of logistic regression: ordinal, multinomial, and binary. When there are two alternative outcomes for the dependent variable 0 or 1 binary logistic regression is used. When the dependent variable, such as categories A, B, C, and D, might have three or more different values, multinomial logistic regression is employed. When the response variable has an ordered structure with at least three levels—for example, student scores that are classified as high, medium, and low ordinary logistic regression is especially helpful. Outliers, however, can have a big effect on how well logistic regression algorithms work.

3.3.2. Support Vector Machine (SVM)

Developed into a popular SVM classifier since then, it has been performing some challenging tasks up to the standard of advanced NNs through expensive algorithms. SVMs were recently developed as effective pattern classification and recognition methods [22]. Subsequently, researchers took full advantage of these methods in the classification challenges of various applications, ranging from image processing and computer vision to speech recognition systems [23, 24]. The SVM algorithm finds its primary use in classification methods. The SVM generates a hyperplane for separating two classes from each other at a maximum distance apart. Support Vector Machines can solve both linear and non-linear problems. In defining the hyperplane in the N-dimensional space (where N represents the number of inputs) capable of distinguishing one category from another, we must minimize the recognition error rate. The hyperplane influenced directly the accuracy of an outcome. The classifiers seek to choose the hyperplane such that the instances of different classes are positioned as distantly as possible from each other. A line separating one class from another serves as a graphic illustration of the hyperplane. Various classifications are applied to data points on opposite sides of that hyperplane. The dimension of the hyperplane is determined by the number of features. A hyperplane is represented as a line when two attributes are used and as a two-dimensional plane when three features are used. With this method of abnormality identification, SVM has also found its importance in areas such as financial analysis, air quality control systems, and medical diagnostics.

3.3.3. Decision Tree Classifier (DT)

A decision tree classification technique generates a tree-like structure that can be used to select from a variety of options. A decision tree can be used to tackle problems related to regression and classification. The decision tree is one of the most used machine learning algorithms. Decision trees often simulate human reasoning, which makes it easy to understand the data and draw some useful conclusions. In a decision tree, each leaf represents a result (continuous or categorical value), each branch or link explains a choice or rule, and each node communicates a characteristic or attribute [25].

Gradient Boosting Classifier

Gradient boosting is a well-liked machine learning technique for classification problems. Gradient Boosting allows the optimization of an arbitrary differentiable loss function, where each prediction corrects the error of the previous one, so building the model step-by-step and finally making it universal. Building a decision tree is the cornerstone of this approach. First, the data set is used to generate and train a decision tree. A second decision tree is then made based on the evaluation, which incorporates changes, once it has been evaluated (the classification error has been calculated). Tree1 + Tree2 is used to create Tree3. For a predetermined number of repetitions, this process is repeated. As a result, the weighted sum of the predictions made by previous tree models is the final model [23].

3.3.4. K-Nearest neighbor (KNN)

One of the supervised machine learning classification techniques is K-Nearest Neighbor. KNN assumes that similar items are in the vicinity. To put it another way, things that are similar are near one another. K-Nearest Neighbor is used to forecast the target variable or label by looking up the nearest neighbor class. A straightforward technique known as "nearest neighbors" maintains all of the cases that are available and categorizes new cases based on the degree of similarity determined by the distance functions (Manhattan, Euclidean, Minkowski, and Hamming). There are two groups into which Nearest Neighbor techniques fall. 1.

Training data is categorized into entity data using the structure-less neural network technique. 2. The Neural Network (NN) method is frequently structured and makes use of data structures such ball trees, k-d trees, axis trees, and orthogonal structure trees (OST). The k-nearest neighbor (KNN) method is frequently used in the banking industry to identify trends in credit card usage. Additionally, data and site activity are analyzed using KNN algorithms to find questionable trends that can point to possible fraudulent activity.

3.3.5. Random Forest (RF)

Random Forest draws many decision trees during training and produces a result based on which tree has received the most votes. This is useful for both regression and classification tasks.

3.3.6. XG Boost (Extreme gradient boosting)

Extreme gradient boosting is a powerful and insightful machine learning approach that makes it much easier to comprehend data and make decisions. Through residual learning, this ensemble learning method combines several "weak" models to provide predictions that are more accurate.

3.3.7. AdaBoost (Adaptive Boosting)

The weighted sum of the robust to noise and outlier weak classifiers (learners) is called Adaboost, and the final model converges to a strong learner.

4. EVALUATION METRICS

Depending on a number of variables, including class imbalance and intended results, the best metrics to assess a classifier's performance for a given dataset in classification issues must be chosen. There is no uniform metric for consistently assessing performance since a classifier may perform well under one metric but poorly under another. To address this, this study evaluates model performance using a variety of criteria, such as accuracy, precision, recall, and F1 score. Four major criteria serve as the foundation for these metrics: True Positives (TP): When the model accurately predicts a class of 1 (True) but the actual class is 1 (True). False Positives (FP) are situations in which the model predicts 1 (True) when the actual class is 0 (False). True Negatives (TN): Situations in which the model prediction and the actual class are both 0 (False). False Negatives (FN): Situations in which the model predicts 0 (False) while the actual class is 1 (True).

Accuracy – The average number of accurate forecasts is referred to as the accuracy measure. This isn't very reliable for the unbalanced dataset, though.

$$\text{Accuracy} = \frac{\text{Correct prediction}}{\text{Total prediction}} \quad [6]$$

Precision - referred to as positive predictive value, gauges a model's ability to find the right examples for every class. This matrix works well for unbalanced datasets and multi-class classification.

$$\text{Precision} = \frac{\text{True Positive}}{(\text{True Positive}+\text{False Positive})} [6]$$

Recall –This metric assesses how well a model detects the genuine positive among all true positive cases.

$$\text{Recall} = \frac{\text{True Positive}}{(\text{True Positive} + \text{False Negative})} \quad [6]$$

F-score – The F-measure, also known as the balanced F-score, is the weighted average of precision and recall.

$$\text{F} - \text{score} = 2 * \frac{(\text{Precision} * \text{Recall})}{(\text{Precision} + \text{Recall})} \quad [6]$$

5. RESULTS AND DISCUSSION

Using the cardiovascular disease dataset, experiments were carried out to develop a reliable system for automatically detecting the presence of cardiovascular conditions. The study utilized deep learning networks alongside various machine learning classification algorithms to analyse the dataset and evaluate their performance. For validation, a 10-fold cross-validation technique was employed, ensuring robust and unbiased performance evaluation. In this method, the dataset was split such that 90% of the data was allocated for training the models, allowing them to learn underlying patterns and features, while the remaining 10% was reserved for testing to evaluate their predictive accuracy. The results highlighted the effectiveness of different machine learning approaches, providing insights into their strengths and weaknesses. Deep learning models, with their ability to capture complex, non-linear relationships in data, demonstrated promising outcomes, especially in handling large datasets with high-dimensional features. At the same time, traditional algorithms such as Support Vector Machines (SVM), Random Forests, and Gradient Boosting showed competitive results, emphasizing the importance of feature selection and hyperparameter tuning. This section delves into the detailed performance metrics, such as accuracy, precision, recall, F1-score, and area under the ROC curve (AUC-ROC), offering a comparative analysis of the methods. These findings underscore the potential of machine learning models to serve as effective tools for the early detection of cardiovascular diseases, aiding clinicians in making informed decisions.

Table 1 displays the classifier's output following feature selection and data processing. With a 65% accuracy rate in disease prediction, the Random Forest model outperformed the least accurate model, logistic regression, which had a 48% accuracy rate.

Table 2: Result of the Classifiers (Selected Features)

	Logistic Regression	Support Vector Machine	Decision Tree Classifier	Gradient Boosting
Cross Validation Accuracy	97.51%	96.59%	95.57%	98.63%
Test Accuracy	93.48%	94.59%	96.59%	97.64%

Table 3. Comparison results of all the Performance Metrics

Classification techniques	Precision	Accuracy
Logistic Regression	93.98	94.67
Support Vector Classifier	82.56	83.33
Decision Tree Classifier	96.95	97.33
Hist Gradient Boosting	99.38	99.33
Proposed learning vector quantization	94.07	94.78

The performance of various classifiers with all available features is summarized in Table 2. Among the models evaluated, the Gradient Boosting classifier achieved the highest accuracy, correctly predicting cardiovascular diseases 97% of the time. On the other hand, Logistic Regression demonstrated the lowest performance, with an accuracy of only 98%. A comparison of Tables 1 and 2 indicates that models trained with the complete set of features outperformed those utilizing feature selections. This outcome suggests that the selection and combination of features may have influenced how the data was represented to the models, potentially reducing their predictive capabilities when fewer features were used. Feature selection, while often employed to reduce dimensionality and computational costs, may inadvertently exclude critical variables necessary for accurate disease prediction.These results underscore the importance of careful feature engineering and representation in the modelling process to ensure that the machine learning algorithms can effectively capture relevant patterns and relationships in the data.

6. CONCLUSION

The analysis of classification techniques highlights the significant differences in performance across various models. The Hist Gradient Boosting classifier emerged as the most effective model, achieving both the highest precision (99.38%) and accuracy (99.33%). This indicates its robustness in detecting cardiovascular diseases with minimal false positives and high predictive reliability. The Decision Tree Classifier followed closely, showing strong results with a precision of 96.95% and accuracy of 97.33%, proving to be a reliable alternative. The Proposed Learning Vector Quantization model performed well, with precision and accuracy values of 94.07% and 94.78%, respectively, demonstrating its potential in specific applications but falling short of the Gradient Boosting model's performance. While the Logistic Regression model showed high precision (93.98%), its accuracy (94.67%) was slightly lower, indicating that it may be prone to misclassifying some instances. The Support Vector Classifier had the lowest performance among the methods, with precision and accuracy scores of 82.56% and 83.33%, respectively, suggesting limitations in its ability to handle the dataset effectively. Overall, the results indicate that ensemble methods, like Hist Gradient Boosting, are highly suitable for the early detection of cardiovascular diseases due to their superior accuracy and precision. However, the proposed Learning Vector Quantization model shows promise and could be further optimized for enhanced performance. These findings emphasize the importance of selecting appropriate algorithms tailored to the specific characteristics of the dataset and clinical requirements.

REFERENCES

1. Majeed, A.A., Alsabah, M., Al-Shaikhli, T.R., &Kaky, K.M. (2024) A Review of Machine Learning's Role in Cardiovascular Disease Prediction: Recent Advances and Future Challenges.
2. Saeed, F., Basurra, S., Albarrak, A.M., &Qasem, S.N. (2024) Machine Learning-Based Predictive Models for Detection of Cardiovascular Diseases.
3. Kar, M. et al. (2023) Automated Heart Disease Prediction Using Hybrid Machine Learning Models.
4. Wang, Y., & Zhang, Z. (2023) Hybrid Deep Learning and SVM Model for Cardiovascular Disease Detection.
5. Ahmed, S., & Bashir, R. (2023) Review of Hybrid Machine Learning Techniques in Cardiovascular Disease Prediction.
6. Hann H. Alalawi, Manal S. Alsuwat "Detection of cardiovascular disease using machine learning classification model" in july. 2021 http://www.ijert.org
7. Kumar, V., & Mishra, N. (2023) Ensemble Learning for Cardiovascular Risk Stratification
8. Patel, H. et al. (2023) Cardiovascular Disease Diagnosis Using Machine Learning Techniques: A Comprehensive Review.
9. Hossain, M. et al. (2021) Deep Learning-Based Hybrid Model for Cardiovascular Disease Detection.
10. Shaikh, A. et al. (2021) Hybrid Model Using Decision Tree and Neural Network for Cardiovascular Disease Diagnosis.
11. Yadav, R. et al. (2021) Hybrid Approach for Detecting Cardiovascular Diseases Based on SVM and Random Forest.

12. Kumar, R., et al. (2010). Hybrid machine learning models for cardiovascular disease detection. Journal of Medical Systems.
13. Chen, J., & Liu, Y. (2014). A hybrid model using SVM and genetic algorithms for cardiovascular disease diagnosis. Expert Systems with Applications.
14. Singh, A., et al. (2018). A novel hybrid approach for cardiovascular disease detection. Journal of Biomedical Informatics.
15. Maheswaran et al., 2022 - Reviews the application of machine learning in predicting cardiovascular disease, focusing on the use of electronic health records and wearable data.
16. Peddakrishna et al., 2023 - Enhances heart disease prediction accuracy using various machine learning techniques, demonstrating significant improvements with a soft voting ensemble classifier.
17. Johnson et al., 2021 - Analyzes the impact of machine learning algorithms on cardiovascular disease prediction, highlighting the importance of data quality and preprocessing.
18. Zhang, H., et al. (2019). Hybrid deep learning model for cardiovascular disease detection. Computers in Biology and Medicine.
19. Rao et al., 2020 - Explores the integration of machine learning algorithms with clinical data to predict cardiovascular disease risk, showing promising results in early diagnosis.
20. Patel, M., & Desai, N. (2020). "Hybrid machine learning-fuzzy logic approach for heart disease prediction." Procedia Computer Science.H. Association, High blood pressure. 2021.
21. J.Brownlee, "How to Develop Voting Ensembles With Python," Machine Learning Mastery, Apr. 16, 2020. https://machinelearningmastery.com/voting-ensembles-with-python/ (accessed Jul. 08, 2021).
22. C.Cortes and V. Vapnik, "Support-vector networks," Mach. Learn., vol. 20, no. 3, pp. 273–297, 1995.
23. K. Aida-zade, A. Xocayev, and S. Rustamov, "Speech recognition using Support Vector Machines," in 2016 IEEE 10th International Conference on Application of Information and Communication Technologies (AICT), 2016, pp. 1–4.
24. Y. Tian, E. Li, L. Yang, and Z. Liang, "An image processing method for green apple lesion detection in natural environment based on GABPNN and SVM," in 2018 IEEE International Conference on Mechatronics and Automation (ICMA), 2018, pp. 1210–1215.

