International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 14, Issue 1, 2026

Detection Of Cardiovascular Disease Utilizing Machine Learning And Optimal Feature Selection Techniques

Mrs.K.Amulya[1], Kondiboyina Doondi Sri Ram Hanuman[2],Bathula Anil Kumar[3], Vuyyuru Lakshmi Priya[4],Daliparthi Pushpa Veera Janakiram[5]
[1]Assistant Professor,Computer Science Engineering Department,Potti Sriramulu Chalavadi Mallikarjuna Rao College Of Engineering,One Town ,Vijayawada, India.
Kamulya@Pscmr.Ac.In
[2,3,4,5]Students;Computer Science Engineering Department,Potti Sriramulu Chalavadi Mallikarjuna Rao College Of Engineering,One Town ,Vijayawada, India
Mail Id; kamulya@pscmr.ac.in[1], sriramhanuman2004@gmail.com[2], anilkumar3031bathula@gmail.com[3], lakshmipriya21005@gmail.com[4], janakiramdaliparti18@gmail.com[5]

Abstract;
One of the main causes of death worldwide, cardiovascular disease (CVD), requires early and precise prognosis. In order to identify the most pertinent medical data features, this work presents an improved machine learning-based CVD diagnosis method that uses FCBF, MRMR, LASSO, ReliefF, and Particle Swarm Optimization. Prediction performance is enhanced by the classification techniques Logistic Regression, Random Forest, ExtraTree, and Gradient Boosting. Accuracy and robustness are increased by extending the suggested system with a Voting Classifier that incorporates AdaBoost, Decision Tree, and ExtraTree. A user-friendly interface for risk visualization, real-time prediction, and health advice is offered by a Flask-based online application with secure user authentication. According to experimental results, the technology is accurate and dependable, which makes it appropriate for practical healthcare applications and early identification of cardiovascular illness.

Keywords—*Crop Disease Detection, Deep Learning, Convolutional Neural Networks (CNN), Image Processing, Plant Leaf Classification, Agriculture, Artificial Intelligence, Computer Vision*

INTRODUCTION

Since cardiovascular disease is one of the main causes of death globally, survival depends on an early and precise diagnosis. According to recent studies, machine learning (ML) technology can predict cardiac sickness by accessing large amounts of medical data and identifying patterns that are invisible to traditional clinical practice [1], [2]. Prediction and clinical decision making are improved by the application of Random Forest, Decision Trees, Support Vector Machines, and Logistic Regression [1].ML-CVD detection models may be weakened by multidimensional and replication characteristics. The best predictive characteristics are found using algorithms based on correlation, mutual information, and chi-square [1], [5]. In order to reduce computational complexity and enhance classification performance, more sophisticated algorithms select the most productive features [3].In recent years, a large number of models in ensemble learning have improved forecast accuracy. Random Forest and hybrid ensemble classifiers are more reliable and generalizable than solo techniques [4]. Model efficiency and diagnostic accuracy are increased by feature selection and optimization techniques such as PSO [5].

An enhanced Voting Classifier in AdaBoost, Decision Tree, and ExtraTree as well as a more effective cardiovascular disease detection system in optimal feature selection will be provided by this research. A web application for safe, real-time prediction and visualization is built using Flask and user authentication. The suggested approach enhances validity, reliability, and utility when used in healthcare.

LITERATURE SURVEY

IoT is being used to collect ECG data for the diagnosis and prognosis of heart disease. The noise in the ECG data causes the diagnostic and prediction algorithm to be inaccurate. To improve diagnosis and prognosis, we eliminate ECG noise. To convert signals to digital, we employed a modified Sequential Recursive (SR) method. To lessen digital noise, the Revised Discrete Wavelet Transform (DWT) finds data peaks. Lastly, we find diagnostic and predictive traits using a feature dataset. Redundancies are removed with Fishers Linear Discriminant. A knowledge-base diagnostic feature was built using the MIT-BIH PhisioNet ECG dataset. We developed a proof-of-concept method to identify and predict heart disease using real ECG data [6].

Heart disease is a fatal condition that affects people all over the world. To preserve lives, this illness has to be detected early. The study creates a machine learning model for predicting cardiac disease. This model uses categorization techniques to provide reliable predictions. Cleveland data is used to train and evaluate a variety of classification techniques, such as Logistic Regression, Random Forest, Support Vector Machine, Gaussian Naïve Bayes, Gradient Boosting, K-nearest neighbors, Multinomial Naïve Bayes, and Decision trees. Additionally, important characteristics of the input data set are extracted using the chi square

technique, which enhances classifier performance and speed. The most accurate classifier is Random Forest. The Random Forest model can be used by medical experts, particularly cardiologists, to identify heart issues [7].

Researchers are interested in heart diseases because they have an impact on human health. Heart disease is one of the main causes of death. Data mining is often used by medical informatics computers. Medical data recognition is enhanced by a number of data mining methods.SVM, KNN, and Naive Bayes are used in this supervised machine learning study to predict cardiac disorders. Machine learning is implemented using the R programming language. Algorithm correctness is used to gauge performance. The functioning and outcomes of the algorithms were examined [8].

A correct diagnosis of cardiac disease can save lives, while an incorrect one can be fatal. The UCI Machine Learning Heart Disease dataset is used in this study to compare machine learning and deep learning results. 14 key dataset features are used in the study. The accuracy and confusion matrix confirms positive results. To eliminate superfluous features and standardize data for improved results, we employ Isolation Forest. Mobile devices and other multimedia technology could be used in this study. Deep learning yielded an accuracy of 94.2% [9].

Early identification is critical since the number of heart disease cases is continually increasing. This diagnosis is difficult and needs to be precise. The study investigates if health conditions increase the risk of heart disease. We developed a strategy for predicting heart disease based on medical history. We used KNN and logistic regression to predict and categorize individuals with heart disease. To control how the model enhances each person's heart attack prediction, a practical approach was applied. Unlike naïve bayes, the proposed model was appealing and used KNN and Logistic Regression to successfully predict heart disease in a given individual. A great deal of tension has been reduced by determining the classifier's accuracy in identifying heart disease. The Given heart disease prediction system enhances and reduces medical expenses. This research aids in the prediction of heart disease. The format is.pynb [10].

Fitness professionals can forecast sickness by identifying heart problems. Numerous research have made use of statistics and statistics mining. Heart disease is diagnosed using scientific datasets that contain parameters and outcomes from intricate testing. Patient data related to coronary heart disease are included in the collection. Coronary heart disease is predicted via class algorithms. The objective of determining the optimal classifier is to calculate classifier accuracy [11].

Cardiovascular diseases have always been serious. Heart disease is the leading cause of death among the top ten, according to the World Health Organization. Early identification is essential for treatment and rehabilitation. To identify heart problems, technology that anticipates cardiac irregularities is required. The goal of this article is to develop a machine learning-based medical system to assist physicians in diagnosing cardiac and cardiovascular diseases. Using a variety of data processing techniques, we correct missing and imbalanced data in the publicly available Framingham and UCI Heart Disease datasets. Additionally, we employ machine learning to select the optimal method for predicting cardiovascular disease. In terms of accuracy, sensitivity, F-measure, and precision, we discovered that our method performed better than earlier models [12].

METHODOLOGY

For the diagnosis of cardiovascular illness, a structured machine learning pipeline is used for data processing, model training, optimum feature selection, and real-time prediction. First, patient health parameters are imported into the system. To enhance quality, missing values, noise, and uninteresting features are eliminated during data preparation. While correlation analysis looks at feature correlations, label encoding transforms categorical input into numerical form for machine learning models.After preprocessing, the most significant features are determined using FCBF, MRMR, LASSO, ReliefF, and ANOVA. By choosing the best features, PSO reduces dimensionality and improves model performance. The optimized dataset is divided into training and testing sets for model development and assessment.

The features are used to train Logistic Regression, Random Forest, ExtraTree, and Gradient Boosting. Prediction performance is enhanced using an ensemble method that uses a Voting Classifier, AdaBoost, Decision Tree, and ExtraTree classifiers. The accuracy and robustness of this ensemble model are enhanced by combining predictions from other models.The trained model is provided by a Flask web application with a secure user interface and user authentication. Using the trained model and user input, the system computes risk percentage, accuracy, and health advice in real-time. This approach maximizes forecast accuracy, data processing, and the usability of healthcare apps.

A. Proposed Undertaking:

The suggested study uses sophisticated ensemble learning and a Voting Classifier using AdaBoost, Decision Tree, and ExtraTree algorithms to improve the cardiovascular disease prediction model. This hybrid ensemble model improves prediction accuracy by combining classifier strengths and lowering model constraints. The most significant dataset properties are identified using FCBF, MRMR, ReliefF, and PSO in order to enhance model performance and efficiency.A Flask-based real-time prediction web application with an easy-to-use user interface is part of the system. Only authorized users are able to access the system and make predictions thanks to secure user

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 14, Issue 1, 2026

authentication. Gauges and indicators clearly show health advice, accuracy, and risk percentage. The proposed study improves prediction accuracy, system usability, and healthcare data security.

B. **Design of the System*:***

The proposed cardiovascular disease detection system uses machine learning to evaluate medical data and produce precise predictions. First, patient health parameter datasets are gathered. During preparation, noise, irrelevant data, and missing values are eliminated to enhance data quality. Following preprocessing, data visualization reveals correlations and feature patterns. For machine learning algorithms, label encoding quantifies category input. The most significant heart disease prediction features are identified from processed data using MRMR, FCBF, LASSO, Relief, and ANOVA.To enhance model performance, PSO chooses the best feature subset. Optimal data is used by Logistic Regression, Random Forest, Extra Tree Classifier, Gradient Boosting, and Extended Voting Classifier. The Voting Classifier integrates model outputs to improve prediction robustness and accuracy. Accuracy, precision, sensitivity, specificity, MCC, and AUC from a confusion matrix are among the final performance metrics. Risk prediction is shown as the result on the web interface. This architecture maximizes accuracy, cardiovascular disease diagnosis, and data processing.

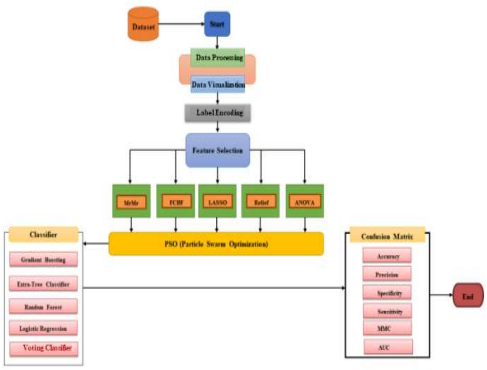

Fig.1. Proposed architecture

IMPLEMENTATION

A. MODULES:

1. Data-loading

Data is loaded by this module. To forecast cardiovascular illness, medical data is loaded into the relevant libraries.

2. Preprocessing Data:

The module removes extraneous properties, noise, and missing values from the dataset. By eliminating unnecessary columns, model training uses only significant and trustworthy data.

3. Visualizing Data:

In order to describe the dataset, this module creates correlation matrices and shows the findings. It locates relationships between feature patterns.

4. Label Encoding:

Label encoding is a method of numericalizing category data using numbers. This is necessary for data compliance with machine learning.

5. Choose Features:

In this session, the most important dataset attributes are identified via MRMR, FCBF, LASSO, Relief, and ANOVA. To make data simpler and increase model accuracy, redundant features are removed.

6. Train-and-Test Data Splitting:

Training and testing are included in the dataset. The testing set evaluates the model, whereas the training set constructs it.

7. Model Making:

Using the characteristics provided, the module creates machine learning models. We use Voting Classifier, ExtraTree, Random Forest, Logistic Regression, and Gradient Boosting. Evaluation metrics are computed, and features and PSO improve performance.

8. User Registration/Login:

This module protects data by securing user login and registration.

9. User Input:

Users enter medical variables to forecast in this module.

10. Prediction:

Using the learned model, the final module makes predictions. Heart disease risk, accuracy, and health suggestions are shown on the user-friendly interface.

ALGORITHMS:

1. Logistic Regression:

Logistic regression is a machine learning approach for binary classification. To determine the likelihood of cardiovascular illness, a sigmoidal function modifies input values between 0 and 1. The model is straightforward, effective, and comprehensible as it approximates the connection between the input feature and the target variable. In order to differentiate CVD-positive from CVD-negative cases, this endeavor utilizes it as a baseline.

2. Random Forest:

An ensemble learning technique called Random Forest generates a large number of decision trees during training and merges their results by majority vote. To lessen overfitting and boost variation, all trees are trained using random data and features. It makes more accurate predictions, particularly when dealing with high-dimensional medical data. For reliability and complex feature interaction, this system depends on Random Forest.

3. ExtraTree Classifier:

ExtraTree (Extremely Randomized Trees) is a more unpredictable ensemble method similar to Random Forest that uses random split points. It increases computing speed and model variation by constructing unpruned decision trees. Randomness lowers volatility and enhances generalization. ExtraTree was used in this work to enhance ensemble model classifiers and enhance prediction performance.

4. Gradient Boosting:
Gradient Boosting is a potent ensemble learning technique that builds on past mistakes by creating one model after another. Gradient descent optimization is used to minimize residual errors. This increases prediction accuracy, particularly when dealing with complex data. By gradually improving predictions using Gradient Boosting, this system maximizes diagnostic performance.
5. Voting Classifier (AdaBoost + Decision Tree + ExtraTree):
Several machine learning models are used by the Voting Classifier ensemble model to provide a majority vote result. AdaBoost, Decision Tree, and ExtraTree classifiers are used in this research to optimize their advantages. This combination method's accuracy, stability, and robustness make it more effective at identifying cardiovascular diseases.
6. ANOVA feature selection has the feature of PSO:
To ascertain the relevance of a characteristic, ANOVA (Analysis of Variance) compares class variance. Particle Swarm Optimization (PSO) improves feature selection to find the most pertinent attributes. By focusing only on significant attributes, this approach decreases the amount of data, removes redundant information, and enhances model performance.
7. MRMR Feature Selection using PSO:
The features with the least amount of overlap that are most pertinent to the target variable are chosen via MRMR (Minimum Redundancy Maximum Relevance). PSO is used to enhance the selection process in order to obtain the optimal feature subset. The combination increases model efficiency and prediction accuracy since only important and non-redundant information is employed.
8. LASSO Feature Selection with PSO:
Less important traits are penalized using the regression model LASSO (Least Absolute Shrinkage and Selection Operator), which sets their coefficients to zero. It may be used in conjunction with PSO to enhance the procedure and choose the optimal features. This enhances model generalization and lessens overfitting.

EXPERIMENTAL RESULTS

The cardiovascular disease detection system was evaluated using machine learning and optimized feature selection. Features were chosen from a preprocessed, labeled, encoded dataset using FCBF, MRMR, LASSO, ReliefF, and ANOVA using Particle Swarm Optimization. By eliminating superfluous data and improving features, these techniques improved model performance.Tests were conducted on classes including Gradient Boosting, Random Forest, ExtraTree, and Logistic Regression. In terms of accuracy and robustness, the extended ensemble Voting Classifier—which incorporates AdaBoost, Decision Tree, and ExtraTree—performed the best. The approach appears predictive with an accuracy of 88.5%.Real-time prediction using a Flask web interface. Based on user input, the method predicts cardiovascular disease with a high chance of 78.7%. Data comprehension is made easier using gauge charts and neatly arranged output panels. This method is superior since it provides individualized health recommendations.Accuracy, precision, sensitivity (recall), specificity, and AUC were the metrics used to validate the model. The findings demonstrate that optimal feature selection and ensemble learning outperform individual models in forecasting. For medical real-time cardiovascular disease risk assessment, the suggested method is dependable, efficient, and appropriate.

Accuracy: Evaluate test reliability using advantages and disadvantages. Mathematics comes next.

$$Accuracy = \frac{(TN + TP)}{T}$$

Precision: Positive examples or categorization accuracy are measured by precision. Accuracy is determined by applying these:

$$\text{Pre } cision = \frac{TP}{(TP + FP)}$$

Recall: Recall that a model's ability to identify every instance of the machine learning class is shown by the ratio of successfully predicted positive observations to total positives.

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

F1-Score: High F1 scores are obtained via accurate machine learning algorithms. Model accuracy is increased when precision and recall are combined. How well a model predicts data is known as its accuracy.

$$F1 = 2 \cdot \frac{(Recall \cdot \text{Pre } cision)}{(Recall + \text{Pre } cision)}$$

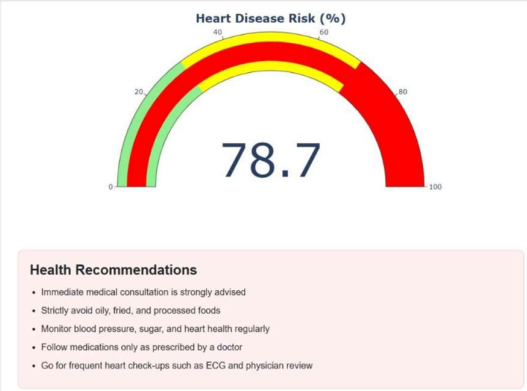

Fig. 2: Prediction Result Interface

The system output dashboard with "Heart Disease

Detected," risk percentage (78.75%), and model accuracy (88.5%) is displayed in this image. It also implies a significant level of danger.

Fig. 3: Heart Disease Risk Gauge Visualization

This image uses a gauge chart to display heart disease risk estimations. The system displays a risk value of 78.7%, with high danger indicated by green, yellow, and red zones.

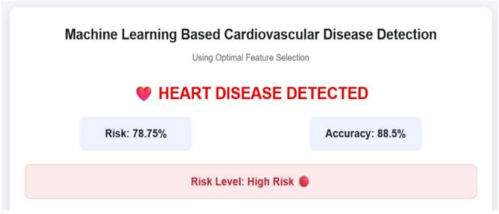

Fig. 4: Health Recommendations Panel

This figure offers health recommendations based on estimated risk. It recommends routine dietary control, medical consultation, monitoring, and follow-ups..

CONCLUSION

To enhance machine learning-based cardiovascular disease identification, this study incorporates an advanced Voting Classifier that combines AdaBoost, Decision Tree, and ExtraTree algorithms into the model. Compared to individual models, an ensemble approach increases forecast accuracy and resilience. To boost model performance, the feature set is enhanced utilizing Particle Swarm Optimization (PSO) using ideal feature selection techniques.

A user-authenticated Flask web application was developed for safe and simple real-time prediction. In an intuitive interface, it promptly presents the accuracy, heart disease risk percentage, and health suggestions. According to experimental data, the suggested method accurately and early diagnoses cardiovascular disease (88.5% accuracy). All things considered, the improved system improves security, accuracy, and usability for practical healthcare applications.

FUTURE SCOPE

In order to improve prediction accuracy, future research can employ deep learning models such as CNNs and RNNs to capture complex medical and ECG patterns. IoT-based wearables' real-time data collecting can enhance patient health monitoring.

To improve feature selection and classification, the model may also be refined using hybrid ensemble techniques and genetic algorithms. Generalization across groups is enhanced by the addition of diverse and extensive clinical data. Furthermore, the Flask-based application might be transformed into a cloud-based healthcare platform with mobile support for safe data interchange, scalability, and remote access for practical medical applications.

REFERENCES

[1] Prashant Maganlal Goad, Pramod J Deore, "Detect the Cardiovascular Diseases in Initial Phase using a Range of Feature Selection Techniques of ML", International Research Journal of Multidisciplinary Technovation, pp.171, 2024.

[2] Yogesh Kumar, Geet Kiran Kaur, Ranjit Singh, "Comprehensive Review of Machine Learning Applications in Heart Disease Prediction", International Journal of Innovative Science and Research Technology (IJISRT), pp.2805, 2024.

[3] N. A. Baghdadi, S. M. Farghaly Abdelaliem, A. Malki, I. Gad, A. Ewis, and E. Atlam, ‘‘Advanced machine learning techniques for cardiovascular disease early detection and diagnosis,’’ J. Big Data, vol. 10, no. 1, p. 144, Sep. 2023.

[4] M. Pal, S. Parija, G. Panda, K. Dhama, and R. K. Mohapatra, ‘‘Risk prediction of cardiovascular disease using machine learning classifiers,’’ Open Med., vol. 17, no. 1, pp. 1100–1113, Jun. 2022.

[5] P. Singh, G. K. Pal, and S. Gangwar, ‘‘Prediction of cardiovascular disease using feature selection techniques,’’ Int. J. Comput. Theory Eng., vol. 14, no. 3, pp. 97–103, 2022, doi: 10.7763/ijcte.2022.v14.1316.

[6] D. T. Thai, Q. T. Minh, and P. H. Phung, ‘‘Toward an IoT-based expert system for heart disease diagnosis,’’ in Proc. 28th Mod. Artif. Intell. Cogn. Sci. Conf. (MAICS), 2017, pp. 157–164.

[7] B. Padmaja, C. Srinidhi, K. Sindhu, K. Vanaja, N. M. Deepika, and E. K. R. Patro, ‘‘Early and accurate prediction of heart disease using machine learning model,’’ Turkish J. Comput. Math. Educ., vol. 12, no. 6, pp. 4516–4528, 2021.

[8] S. Anitha and N. Sridevi, Heart Disease Prediction Using Data Mining Techniques S Anitha, N Sridevi to Cite This Version, document HAL Id Hal02196156, 2019. [Online]. Available: https://hal.archives-ouvertes.fr/ hal-02196156/document

[9] R. Bharti, A. Khamparia, M. Shabaz, G. Dhiman, S. Pande, and P. Singh, ‘‘Prediction of heart disease using a combination of machine learning and deep learning,’’ Comput. Intell. Neurosci., vol. 2021, pp. 1–11, Jul. 2021, doi: 10.1155/2021/8387680.

[10] H. Jindal, S. Agrawal, R. Khera, R. Jain, and P. Nagrath, ‘‘Heart disease prediction using machine

learning algorithms,'' IOP Conf., Mater. Sci. Eng., vol. 1022, no. 1, Jan. 2021, Art. no. 012072, doi: 10.1088/1757- 899x/1022/1/012072.
[11] B. Pavithra and V. Rajalakshmi, ''Heart disease detection using machine learning algorithms,'' in Proc. Int. Conf. Emerg. Current Trends Comput. Expert Technol., vol. 35, 2020, pp. 1131–1137, doi: 10.1007/978-3-030- 32150-5_115.
[12] .Louridi, S. Douzi, and B. El Ouahidi, ''Machine learning-based identification of patients with a cardiovascular defect,'' J. Big Data, vol. 8, no. 1, pp. 1–5, Dec. 2021, doi: 10.1186/s40537-021-00524-9.
[13] M. Swathy and K. Saruladha, ''A comparative study of classification and prediction of cardio-vascular diseases (CVD) using machine learning and deep learning techniques,'' ICT Exp., vol. 8, no. 1, pp. 109–116, Mar. 2022, doi: 10.1016/j.icte.2021.08.021.
[14] D. Vaddella, C. Sruthi, B. K. Chowdary, and S.-R. Subbareddy, ''Prediction of heart disease using machine learning techniques,'' Restaur. Bus., vol. 118, no. 1, pp. 125–129, 2019, doi: 10.26643/rb.v118i1.7621.
[15] V. V. Ramalingam, A. Dandapath, and M. Karthik Raja, ''Heart disease prediction using machine learning techniques: A survey,'' Int. J. Eng. Technol., vol. 7, no. 2, p. 684, Mar. 2018, doi: 10.14419/ijet.v7i2.8.10557.
[16] J. P. Li, A. U. Haq, S. U. Din, J. Khan, A. Khan, and A. Saboor, ''Heart disease identification method using machine learning classification in E-healthcare,'' IEEE Access, vol. 8, pp. 107562–107582, 2020, doi: 10.1109/ACCESS.2020.3001149.
[17] P. Kalpana, S. S. Vignesh, L. M. P. Surya, and V. V. Prasad, ''Prediction of heart disease using machine learning,'' J. Phys., Conf. Ser., vol. 1916, no. 1, May 2021, Art. no. 012022, doi: 10.1088/1742-6596/1916/1/012022.
[18] A. Ed-Daoudy and K. Maalmi, ''Real-time machine learning for early detection of heart disease using big data approach,'' in Proc. Int. Conf. Wireless Technol., Embedded Intell. Syst. (WITS), Apr. 2019, pp. 1–5, doi: 10.1109/WITS.2019.8723839.
[19] I. D. Mienye, Y. Sun, and Z. Wang, ''An improved ensemble learning approach for the prediction of heart disease risk,'' Informat. Med. Unlocked, vol. 20, Jan. 2020, Art. no. 100402, doi: 10.1016/j.imu.2020.100402.
[20] A. Gupta, R. Kumar, H. S. Arora, and B. Raman, ''MIFH: A machine intelligence framework for heart disease diagnosis,'' IEEE Access, vol. 8, pp. 14659–14674, 2020, doi: 10.1109/ACCESS.2019.2962755.