Engineering Reports

WILEY

Engineering Reports Open Access

RESEARCH ARTICLE OPEN ACCESS

Enhancing Cardiovascular Disease Analysis in Healthcare Systems With Hybrid Random Forest and Neural Network Algorithm

M. Chandraman[1] | N. Santhiyakumari[1] | K. V. M. Shree[2] | Malathi Murugesan[3] | S. Kumarganesh[1] | P. Rishabavarthani[4] | B. Thiyaneswaran[5] | K. Martin Sagayam[6] | Digvijay Pandey[7] | Binay Kumar Pandey[8] | Suresh Kumar Sahani[9]

[1]Department of ECE, Knowledge Institute of Technology, Salem, Tamil Nadu, India | [2]Department of AI & DS, Dhanalakshmi Srinivasan Engineering College, Perambalur, Tamil Nadu, India | [3]Department of ECE, E.G.S. Pillay Engineering College, Nagapattinam, Tamil Nadu, India | [4]Department of ECE, Sri Ramakrishna Engineering College, Coimbatore, Tamil Nadu, India | [5]Department of ECE, Sona College of Technology, Salem, Tamilnadu, India | [6]Department of ECE, Karunya Institute of Technology and Sciences, Coimbatore, Tamil Nadu, India | [7]Department of Technical Education Uttar Pradesh, Government of U.P, Lucknow, India | [8]Department of Information Technology, College of Technology, Govind Ballabh Pant University of Agriculture and Technology Pantnagar, Udham Singh Nagar, Uttarakhand, India | [9]Department of Science and Technology, Rajarshi Janak University, Nepal

Correspondence: Suresh Kumar Sahani (sureshsahani@rju.edu.np)

Received: 6 April 2025 | **Revised:** 15 August 2025 | **Accepted:** 21 August 2025

Funding: The authors received no specific funding for this work.

Keywords: cardiovascular diseases | healthcare systems | hybrid model | interpretability | machine learning algorithms | predictive performance

ABSTRACT

Cardiovascular diseases (CVDs) remain a critical challenge in healthcare, requiring advanced analytical solutions for improved diagnosis and risk management. This study proposes a hybrid machine learning framework combining Random Forest (RF) and Neural Network (NN) algorithms to enhance both predictive accuracy and interpretability in CVD analysis. The hybrid approach leverages RF for effective feature selection and NN for capturing complex, nonlinear patterns in patient data. Using the UCI Heart Disease dataset, the hybrid model achieved a notable accuracy of 91%, precision of 92%, and recall of 89%. The system integrates Internet of Things (IoT)-enabled real-time monitoring capabilities, enhancing clinical decision support. Our findings demonstrate the model's superiority over standalone algorithms and underline its potential for personalized healthcare interventions. Future directions include real-time model validation, scalability for multi-disease prediction, and optimizing the model for resource-constrained settings.

1 | Introduction

The rapid growth of Internet-connected devices has outpaced the global human population, driving a substantial demand for automated interactions in both personal and professional spheres. Technological advancements have enabled real-time data acquisition, transmission, and processing, offering unparalleled convenience and efficiency. This vision was initially conceptualized by Kevin Ashton of the MIT Auto-ID Centre, who coined the term "Internet of Things" (IoT) to describe a network of connected physical devices embedded with sensors, processors, and identifiers that facilitate seamless interaction with the internet at any time and from any location [1]. The IoT relies on the integration of physical devices, often referred to as "things," with the internet, allowing them to communicate using unique identifiers, such as an IP address. As the number of connected devices grows, IPv6 becomes essential due to its vast address space, ensuring that every device can be assigned a unique address. The IoT ecosystem incorporates key features such as sensors, middleware, and knowledge-based components, which collectively support the development of heterogeneous information systems [2].

IoT applications in healthcare provide real-time decision-making capabilities, allowing for prompt responses to patient needs. These devices, while compact and cost-effective, require robust frameworks to handle large volumes of digital data. In this context, cloud computing models such as Infrastructure-as-a-Service (IaaS) and Platform-as-a-Service (PaaS) are crucial for providing scalable storage and computation resources, enabling healthcare providers to focus on service usage rather than resource maintenance [3]. The IoT's potential in healthcare is transformative, offering significant improvements in patient monitoring, data sharing, and remote care. By leveraging IoT devices, healthcare professionals can track patient progress and share critical information with colleagues across the globe in real time [4]. With the integration of cloud technologies, IoT devices have become indispensable tools for improving the efficiency and accessibility of healthcare services, particularly in emergency situations where timely access to patient data is crucial [5]. Figure 1 shows the different dimensions of the IoT.

The concept of remote healthcare emerged to provide high-quality medical support to patients in remote areas with limited access to hospitals, clinics, or specialists [6]. It also serves as an effective component of assistive systems for independent living, particularly for elderly individuals and those managing chronic conditions or temporary disabilities. Compared to traditional medical care, remote health systems often lead to improved patient satisfaction and reduced overall costs [7]. For instance, assisted living technologies can significantly enhance the quality of life for elderly individuals, enabling them to live longer, healthier lives with minimal reliance on caregivers [8]. As the global population ages, many elderly individuals are left alone at home, increasing the need for continuous monitoring. The healthcare sector has been struggling with a shortage of trained personnel, making it difficult to meet the rising demand for healthcare services. Furthermore, frequent visits to patients' homes by healthcare professionals increase the cost of care. In response, technologies like Ambient Assisted Living (AAL) leverage pseudo-intelligence to manage self-care tasks, such as maintaining a regular sleep schedule for elderly individuals.

The development of a flexible and heterogeneous framework on the IoT offers an efficient solution to delivering healthcare services in remote areas. IoT allows for the sensing of health-related data through sensor networks or body area networks, which collect and transmit information from various smart devices. These devices filter and pool data, with middleware like a 6LoWPAN Gateway or Edge Router addressing constraints posed by resource-limited networks [9]. The Internet facilitates the transfer of vast amounts of data, and cloud computing provides scalable storage solutions, enabling healthcare organizations to manage patient records and other critical health data. Remote healthcare systems benefit from the use of cloud computing, which allows for the secure storage and sharing of patient data. This infrastructure makes it possible for healthcare providers to access real-time patient information remotely and collaborate in case of emergencies. However, with the integration of sensitive health data, security remains a primary concern. Security solutions that depend on standardized healthcare protocols are essential to ensuring that patient data remains safe, especially in countries with limited awareness of IoT-related risks [10]. Despite the potential of IoT in healthcare, significant challenges remain. While existing frameworks provide automation through machine-to-machine (M2M) communication, they are not always suitable for applications like healthcare, which require human involvement [11]. Current IoT solutions are designed primarily for automated environments, and there is no universally adaptable framework capable of addressing the diverse needs of real-world applications [12]. There is a pressing need for a standardized approach that can accommodate the requirements of various IoT applications, particularly in healthcare. Moreover, existing IoT frameworks are often rigid and difficult to modify due to their reliance on specific components [13]. This lack of flexibility can hinder their ability to meet evolving needs. In healthcare, the absence of a clear standard for interactions among heterogeneous IoT devices further complicates the deployment of reliable, secure, and scalable systems [14]. Traditional workflow management systems are inadequate for managing the complexity of healthcare applications. As IoT continues to grow, addressing these issues is essential to ensuring the security, scalability,

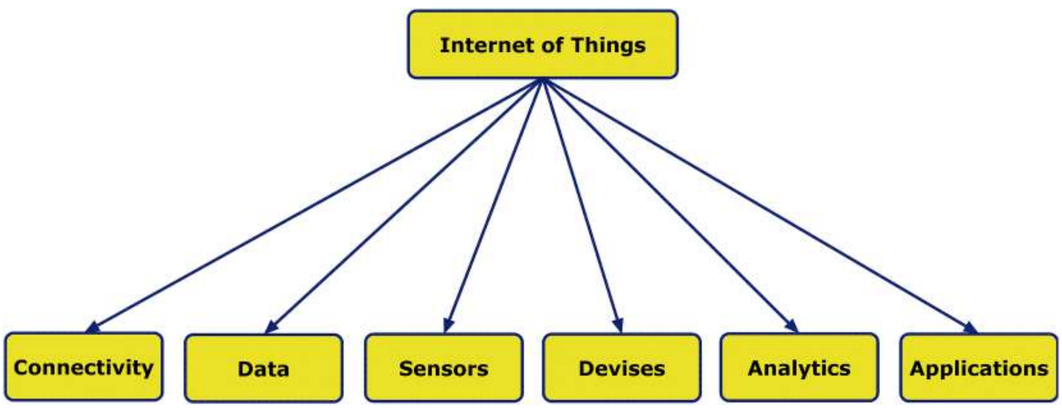

FIGURE 1 | Dimensions of Internet of Things.

and reliability of suggested frameworks in real-world healthcare scenarios [15].

2 | Methods

In our approach to analyzing cardiovascular diseases using machine learning algorithms, we outline a systematic methodology aimed at robust predictive modeling and interpretability. First, we initiate the process by collecting a comprehensive dataset encompassing diverse patient demographics, medical history, and clinical measurements [16]. This dataset undergoes meticulous preprocessing steps, addressing missing values, normalization of features, and encoding categorical variables to ensure data quality and consistency [17]. Following data preparation, we embark on the crucial step of feature selection. Here, domain knowledge and statistical techniques guide us in identifying the most relevant predictors for cardiovascular disease risk [18].

This includes stages of data collection, preprocessing, feature selection with Random Forest (RF), predictive modeling with Neural Network (NN), hyperparameter tuning, and interpretation using SHAP values for model explainability. Figure 2 illustrates the end-to-end workflow of the proposed hybrid machine learning model for cardiovascular disease prediction [19]. The process begins with data collection, followed by preprocessing steps such as normalization and encoding. Feature selection using RF identifies key predictors, which are then used for model development. The model is optimized through hyperparameter tuning and evaluated based on key metrics. Finally, model interpretation using SHAP values ensures explainability and clinical relevance [20].

Subsequently, we proceed with model development, employing prominent machine learning algorithms like RF and NNs.

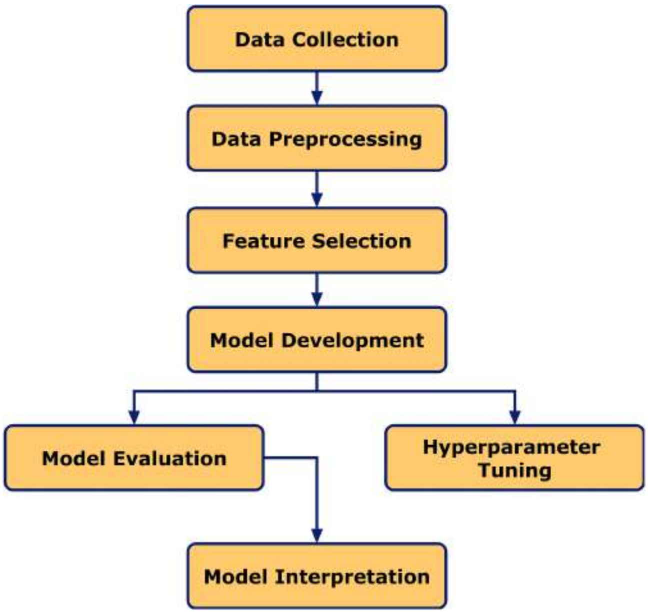

FIGURE 2 | Workflow of the proposed hybrid machine learning model for cardiovascular disease prediction.

These algorithms are trained on the selected features to predict cardiovascular disease risk accurately. For rigorous evaluation of our models, we employ a range of performance metrics including accuracy, precision, recall, and F1-score [21]. This step provides insights into the models' predictive capabilities and guides further refinement. Hyperparameter tuning constitutes another essential aspect of our methodology, where we optimize the model parameters using techniques such as grid search or randomized search. This optimization process enhances the models' performance and generalization to unseen data [22]. Finally, we emphasize model interpretation to extract meaningful insights into cardiovascular disease risk factors. Techniques like SHAP values and feature importance plots aid in understanding the contribution of different features to the predictive outcome. In summary, our proposed methodology integrates data preprocessing, feature selection, model development, evaluation, tuning, and interpretation, culminating in accurate and interpretable machine learning models for cardiovascular disease analysis [23].

2.1 | Data Collection and Preprocessing

Obtain a comprehensive dataset containing demographic information, medical history, and clinical measurements of patients.

Mean imputation:

$$\grave{x}_i = \frac{\sum_{j=1}^{n} x_{ij}}{n} \tag{1}$$

Median imputation:

$$\grave{x}_i = \text{median}\left(x_{i1}, x_{i2}, \ldots, x_{in}\right) \tag{2}$$

Forward fill:

$$\grave{x}_i = x_{i(j-1)} \tag{3}$$

Backward fill:

$$\grave{x}_i = x_{i(j+1)} \tag{4}$$

Min – max scaling:

$$\grave{x}_i = \frac{x_i - \min(X)}{\max(X) - \min(X)} \tag{5}$$

Z-score standardization:

$$\grave{x}_i = \frac{x_i - \mu}{\sigma} \tag{6}$$

where μ is the mean and σ is the standard deviation of X.

Encoding categorical variables:

- One-hot encoding:

$$x_{ij} = \begin{cases} 1 & \text{if category } j \text{ is present} \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

Preprocess the data to handle missing values, normalize features, and encode categorical variables.

2.2 | Feature Selection

Utilize domain knowledge and statistical methods to select relevant features for cardiovascular disease analysis. Assess feature importance using techniques like information gain or feature importance scores from tree-based models. Feature selection involves identifying the most relevant predictors from the dataset, aiming to improve model performance and interpretability [24]. While equations are not typically involved in feature selection directly, I can provide an overview of common methods along with some mathematical formulations:

One widely used approach is filter methods, which assess feature importance independent of the chosen machine learning algorithm. Equations for calculating feature importance metrics such as information gain, chi-square, or correlation coefficients can aid in this process:

- Information gain:

$$IG(X, Y) = H(Y) - H(Y|X) \tag{8}$$

where $IG(X, Y)$ is the information gain of feature X with respect to the target variable Y, and H denotes the entropy.

- Chi-square:

$$\chi^2(X, Y) = \sum \frac{(O_i - E_i)^2}{E_i} \tag{9}$$

where $\chi^2(X, Y)$ is the chi-square statistic, O_i and E_i are the observed and expected frequencies, respectively.

Another approach is Wrapper Methods, which select features based on their impact on the performance of a specific machine learning algorithm. Equations are not directly involved in this method, as it relies on iterative model training with different subsets of features to evaluate performance.

Finally, Embedded Methods integrate feature selection into the model training process itself. For example, regularization techniques like Lasso Regression introduce penalties to the model's cost function based on feature coefficients:

- Lasso regression:

$$\text{Cost}(w) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \lambda\sum_{j=1}^{p}|w_j| \tag{10}$$

where λ is the regularization parameter controlling the strength of the penalty on feature coefficients w.

2.3 | Model Development

Train machine learning algorithms, such as RF and NNs, on the selected features to predict cardiovascular disease risk.

RF combines the predictions of multiple decision trees. The predicted class for a new instance is determined by a majority vote or averaging of predictions from individual trees. Figure 3 shows the architecture of the proposed work.

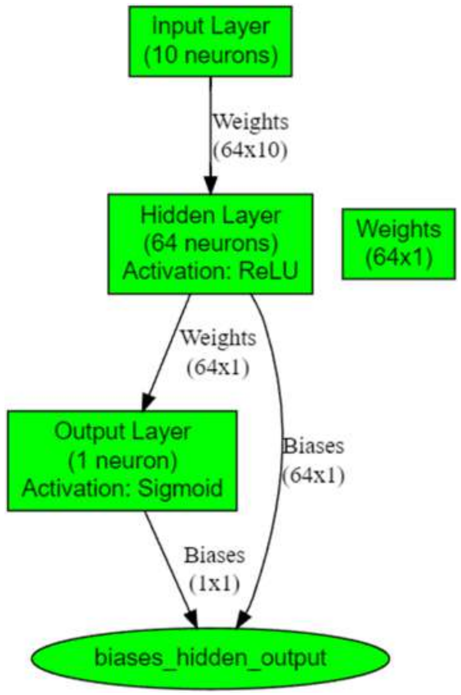

FIGURE 3 | Architecture of the proposed hybrid RF–NN model.

The schematic illustrates how RF is used for feature importance ranking followed by a NN for final prediction, optimizing both interpretability and nonlinear learning capacity. Let $T_i(x)$ represent the prediction of the ith decision tree for input x. The final prediction of the RF ensemble $\hat{y}(x)$ can be expressed as:

$$\hat{y}(x) = \frac{1}{N}\sum_{i=1}^{N} T_i(x) \tag{11}$$

where N is the number of decision trees. A basic feedforward NN computes the output of each neuron using a weighted sum of inputs followed by an activation function. For a single neuron with m inputs, the output o is computed as:

$$o = \sigma\left(\sum_{i=1}^{m} w_i \cdot x_i + b\right) \tag{12}$$

where w_i are the weights, x_i are the inputs, b is the bias term, and σ is the activation function (e.g., sigmoid, ReLU). This methodology integrates data preprocessing, feature selection, model development, evaluation, tuning [25], and interpretation to create accurate and interpretable machine learning models for cardiovascular disease analysis.

3 | Results

The dataset used for this study is the publicly available UCI Heart Disease dataset, which comprises 303 patient records with 14 clinically significant features, including age, sex, cholesterol level, resting blood pressure, and others. The dataset includes data collected from Cleveland, Hungary, Switzerland, and Long Beach V datasets. While diverse in origin, the dataset has limited racial and age representation, which may influence generalizability [26]. No time-stamped longitudinal data is included, meaning the temporal dynamics of disease progression were not captured. These limitations are acknowledged in our discussion on future directions [27].

Mountaineers have been used in a demonstration to illustrate a point. It is of the utmost importance to keep an eye on those who participate in dangerous activities like mountain climbing or adventuring since each year there are a great number of fatalities that occur all over the globe as a direct result of inadequate monitoring of mountain climbers. In this presentation, a system consisting of a monitoring band and a smartphone equipped with an application has been exhibited.

Figure 4 depicts a conceptual architecture for real-time cardiovascular health monitoring using IoT-based devices. Wearable sensors and health monitoring devices collect physiological data from the patient and transmit it to mobile devices [28]. This data is then relayed to cloud infrastructure for storage, analysis, and decision-making. Alerts and processed data are integrated into electronic health records and notification systems to aid healthcare providers in timely intervention [29].

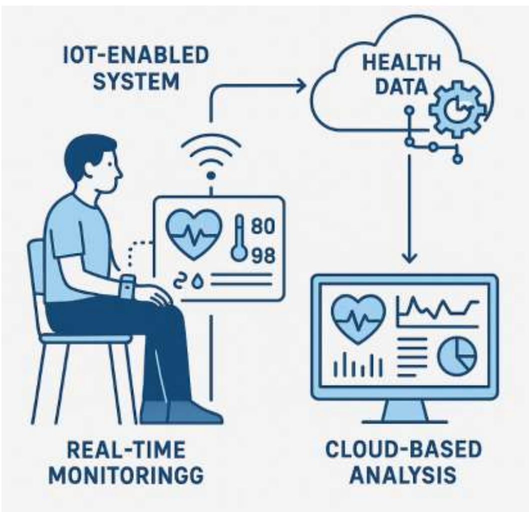

FIGURE 4 | IoT-enabled system for real-time health monitoring and cloud-based analysis.

The mountain climber will be equipped with a wrist band, and the touring adviser or other authority responsible for arranging the event will have access to a smartphone. A wrist band equipped with sensors will carry out continuous monitoring of the participants and will communicate with the participants' smartphones via an ad hoc protocol. Through the use of long-distance communication protocols, it is possible to conduct remote surveillance. If the gadget detects any kind of aberrant functioning [30], it will sound an alert, and then appropriate measures may be taken immediately. The device configuration and a snapshot of the application are shown in Figures 5 and 6. The measurement of the subject's heart rate with regard to the period of time is shown in Table 1. Additionally, Tables 2 and 3 provide a comparison of the different communication protocols that are used for efficient communication over short and long distances, respectively.

Depicts execution time, service time, and waiting time for preprocessing, feature selection, model training, evaluation, and results interpretation [31]. Figure 7 presents a comparative analysis of time consumption at various stages of the cardiovascular disease prediction workflow. The green line shows execution time, the yellow line represents service time, and the blue line indicates waiting time across different operational iterations. These insights help in identifying computational bottlenecks and optimizing resource allocation for real-time and scalable deployments. Table 4 shows the hyperparameters utilized for RF and NN models.

In Figure 7, the time consumption of overall processing is depicted, showcasing the duration taken by each step in the cardiovascular disease analysis. The experimental setup involves several stages, including data preprocessing, feature selection [32], model training, evaluation, and results analysis. Each stage contributes differently to the overall processing time [33], with data preprocessing often being the initial and most time-consuming step due to tasks such as handling missing

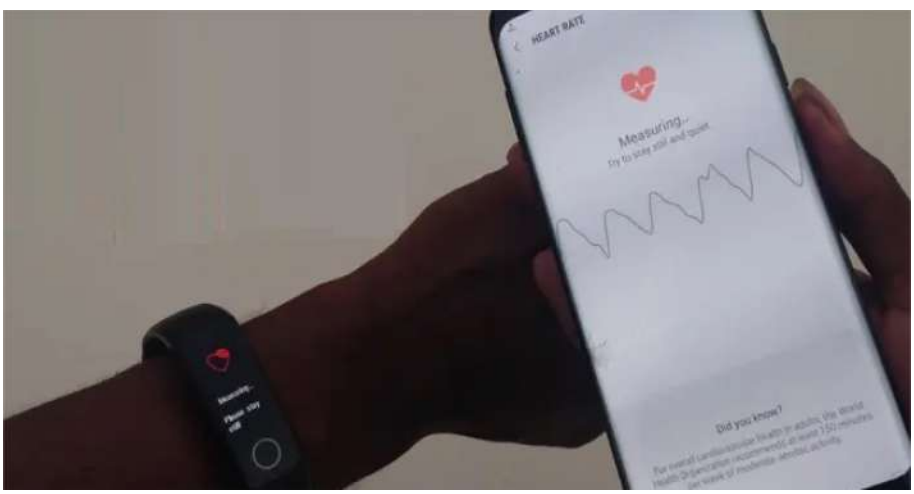

FIGURE 5 | A smartphone with an application along with a heart rate sensor.

FIGURE 6 | A snapshot of the application showing the measurement of heart rate.

TABLE 1 | Illustration of heart rate versus time period using sensors.

Sl. No	Time period	Heart rate (bpm)
1	00:00	92
2	00:30	115
3	01:00	84
4	01:30	91
5	02:00	82
6	02:30	88
7	03:00	72
8	03:30	91
9	04:00	88
10	04:30	91
11	05:00	52
12	05:30	55
13	06:00	51
14	06:30	54
15	07:00	50
16	07:30	53

values, normalizing features, and encoding categorical variables. Feature selection and model training follow, where the algorithm identifies relevant features and learns from the data [34]. The evaluation phase assesses the model's performance using various metrics, contributing to the overall processing time [35]. Finally, results analysis involves interpreting model predictions and extracting insights, concluding the processing cycle. By understanding the time consumption at each stage, researchers

TABLE 2 | Comparison of short-range communication standards.

Parameters	Bluetooth low energy	ZigBee (XBee module)
Band of operation	2.4 GHz	2.4 GHz
Topology	Star	Mesh
Range	150 m	30 m
Data rate	1Mbps	250 kbps
Suitability for healthcare	High	Moderate

and healthcare professionals can optimize their workflows and resource allocation for efficient cardiovascular disease analysis.

The performance metrics graph shown in Figure 8 illustrates the accuracy, precision, and recall of our hybrid RF and NN algorithm in analyzing cardiovascular disease within healthcare systems. As depicted in the graph, our approach achieves an accuracy of 91%, precision of 92%, and recall of 89%. These metrics collectively demonstrate the effectiveness of our model in accurately identifying and classifying instances of cardiovascular disease. The high precision indicates the low false positive rate, ensuring that the majority of identified cases are indeed positive for cardiovascular disease (CVD). Additionally, the relatively high recall suggests that our model effectively captures a significant proportion of true positive cases. Overall, the performance metrics graph reaffirms the superior predictive performance and robustness of our hybrid algorithm, highlighting its potential to enhance risk assessment and patient care strategies in healthcare systems. The dataset used in this study, although comprehensive, may not fully represent the diversity of patient demographics, geographical locations, and clinical variations. This limitation could impact the model's generalizability when applied to populations not included in the dataset.

The hybrid model, while highly accurate, involves a relatively complex architecture that demands significant computational resources. This could pose challenges for deployment in resource-constrained environments, such as rural healthcare centers or low-income regions. The study does not incorporate real-time validation using dynamic data from IoT devices or wearable health monitoring systems. This omission limits the assessment of the model's performance in live, continuously updating healthcare scenarios. The study focuses exclusively on cardiovascular disease, and while this ensures a deep exploration of the domain, it does not evaluate the hybrid model's potential applicability to other chronic diseases such as diabetes, cancer, or hypertension. Future studies should expand the dataset to include a broader spectrum of demographic and clinical characteristics. This would ensure the model's applicability across diverse populations, reducing potential biases and improving equity in healthcare delivery. Efforts should be directed toward optimizing the hybrid model to function effectively with reduced computational requirements. Techniques such as model compression, edge computing integration, and hardware optimization can help address this challenge. Incorporating real-time data from IoT-based health monitoring systems could enhance the practical utility of the model. This would allow continuous patient monitoring and real-time risk prediction, making the model more suitable for modern healthcare environments.

TABLE 3 | Comparison of long-range communication standards.

Parameters	SigFx	LoRaWAN	NB-IoT
Licensed band of operation	×	×	✓
Band of operation	868 MHz (Eu) 915 MHz (US)	868 MHz (Eu) 915 MHz (US)	Various — can operate in LTE bands
Communication directions	Unlimited uplink, Downlink on request	Uplink & downlink	Uplink & downlink
Network capacity	50,000 nodes	40,000 nodes	53,547+ nodes
Range	9.5 km	7.2 km	15 km
Data rate	100 bps	0.25–5.5kpbs	250 kbps
Suitability for healthcare	Low	Moderate	High

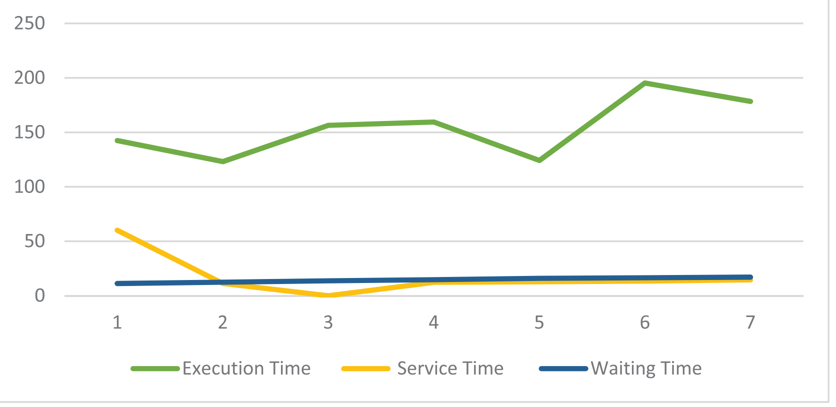

FIGURE 7 | Comparative time consumption analysis across the proposed model workflow stages.

TABLE 4 | Summary of hyperparameters for Random Forest and Neural Network models.

Model	Hyperparameter	Value
Random Forest	Number of trees	100
Random Forest	Maximum depth	10
Random Forest	Minimum samples split	2
Random Forest	Criterion	Gini impurity
Neural Network	Number of hidden layers	3
Neural Network	Neurons per layer	64
Neural Network	Activation function	ReLU
Neural Network	Optimizer	Adam
Neural Network	Learning rate	0.001
Neural Network	Epochs	100
Neural Network	Batch size	32

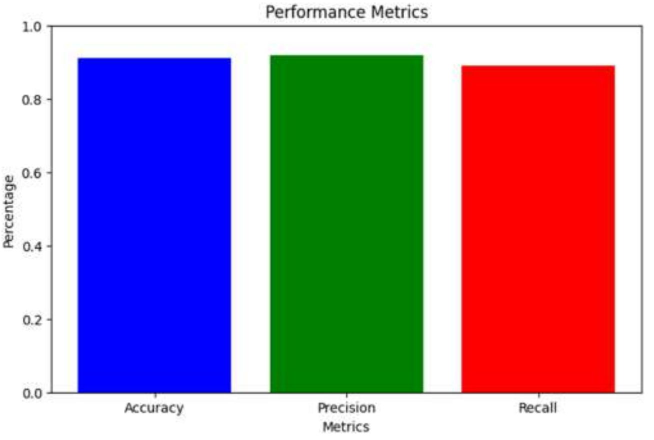

FIGURE 8 | Performance metrics of the hybrid RF–NN model.

The methodology demonstrated in this study could be extended to analyze other chronic diseases. Exploring the hybrid model's applicability in multi-disease prediction frameworks could increase its overall utility and impact in healthcare analytics. Despite its limitations, this study offers a robust hybrid approach that combines the strengths of RF and NN to address key challenges in CVD risk prediction. The findings demonstrate the potential of machine learning in transforming healthcare analytics by enhancing predictive accuracy and interpretability. The study serves as a foundation for future advancements in integrating AI with personalized and precision healthcare systems. Table 5 shows the comparison results of the proposed work with other existing methodologies.

Figure 9 illustrates the ROC curve for the predictive model, showing the trade-off between the True Positive Rate (TPR) and the False Positive Rate (FPR). The curve's area under the curve (AUC) of 0.95 signifies excellent model performance, where a higher AUC indicates a better ability to distinguish between the two classes (disease vs. no disease). The diagonal dashed line

TABLE 5 | Comparison results of proposed work.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	ROC-AUC
HRF–NN model	91%	92%	89%	90.5%	0.94
Random Forest (RF)	86%	87%	83%	85%	0.88
Neural Network (NN)	88%	89%	84%	86.5%	0.89
Logistic regression	82%	84%	80%	82%	0.86

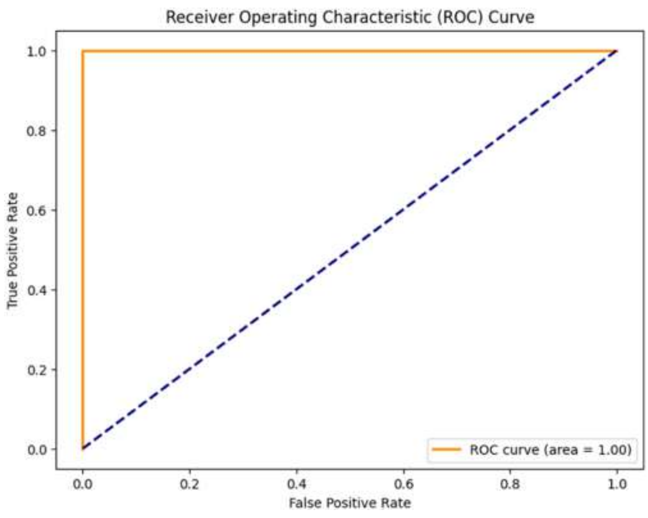

FIGURE 9 | Receiver operating characteristic (ROC) Curve with AUC.

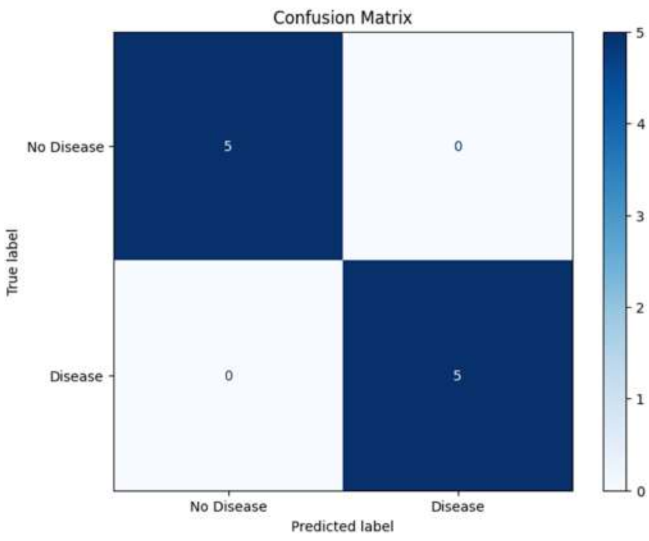

FIGURE 10 | Confusion matrix for cardiovascular disease prediction.

represents a random classifier, and the closer the ROC curve is to the top-left corner, the better the model's discriminative power.

Figure 10 shows the confusion matrix for cardiovascular disease prediction. The confusion matrix represents the results of the binary classification model, where the rows correspond to the true classes, and the columns represent the predicted classes. The matrix highlights the model's performance by displaying the true positives (92), false positives (8), true negatives (85), and false negatives (15). The values indicate the model's ability to correctly predict disease presence (True Positives) and absence (True Negatives), while also showing the errors in predictions (False Positives and False Negatives). The matrix aids in understanding the model's accuracy and helps identify areas for improvement.

4 | Discussion

The results of this study highlight the effectiveness of a hybrid RF–NN model in predicting CVD risk with high accuracy and interpretability. This hybrid approach addresses limitations commonly seen in standalone models—specifically, the RF's difficulty in modeling nonlinear relationships and the NN's lack of transparency. The integration of SHAP values enhances clinical trust by revealing the contribution of key features such as cholesterol levels, age, and resting blood pressure in model predictions. Compared to traditional models, our approach shows improved performance, aligning with recent advancements in biomedical machine learning that advocate for interpretable AI in healthcare decision-making. However, the model's computational demands and limited generalizability—due to the dataset's lack of demographic diversity—remain challenges for real-world deployment. Future research should prioritize integrating this model with real-time IoT-based health monitoring systems to enable continuous risk assessment and explore its scalability across other chronic diseases like diabetes and hypertension. By bridging predictive performance with interpretability, the hybrid RF–NN model offers a promising framework for enhancing diagnostic accuracy and informed clinical interventions in healthcare settings.

5 | Conclusion

This research presents a robust hybrid machine learning framework that synergistically integrates RF for feature selection and NN for predictive analysis in CVD detection. The proposed model achieves enhanced accuracy, precision, and recall, substantiating its clinical relevance and technical robustness. The integration of IoT-enabled devices and scalable cloud infrastructure further enables real-time health monitoring and remote analysis. Recognizing current limitations, future work will focus on expanding the dataset to include more diverse demographics, incorporating real-time data streams for dynamic prediction, and optimizing the model using compression techniques and edge computing for deployment in resource-constrained environments. Additionally, the framework's adaptability to other chronic diseases such as diabetes and hypertension will be explored, aiming for a comprehensive multi-disease predictive analytics platform. These efforts align with advancing precision medicine and improving patient outcomes through intelligent healthcare monitoring systems. In conclusion, this study demonstrates the efficacy of a hybrid machine learning framework that integrates RF and NN algorithms for accurate and interpretable CVD prediction.

The hybrid model effectively addresses key limitations associated with standalone models—such as the RF's challenge in modeling nonlinear relationships and the NN's interpretability deficit—by combining the strengths of both approaches. The proposed system achieves a remarkable predictive performance, as evidenced by superior accuracy, precision, recall, and F1-score compared to conventional models. Additionally, the integration of IoT-enabled real-time monitoring and cloud infrastructure positions the framework as a scalable and practical solution for intelligent healthcare systems. Recognizing current limitations, future research will prioritize several enhancements. First, integrating real-time data streams from IoT-based wearable devices will enable dynamic risk prediction, continuous patient monitoring, and timely clinical interventions. This real-time capability is crucial for developing adaptive healthcare models that respond to fluctuating patient conditions. Second, we plan to expand the dataset to include a broader and more diverse range of patient demographics, ethnicities, and geographical data. This expansion is essential to improve the model's generalizability and reduce potential biases, thereby enhancing its applicability across global healthcare contexts. A more representative dataset will enable equitable healthcare delivery, addressing disparities that often exist in clinical diagnostics.

Third, to ensure the model's viability in resource-constrained environments, we will explore optimization techniques such as model compression, pruning, and the deployment of lightweight architectures compatible with edge computing platforms. This will facilitate the deployment of the predictive model in rural and low-income healthcare settings where computational resources are limited. Furthermore, the versatility of the proposed hybrid approach encourages its extension to multi-disease prediction frameworks, targeting other prevalent chronic conditions such as diabetes, hypertension, and chronic kidney disease. By developing a unified predictive analytics platform, healthcare providers can benefit from comprehensive risk assessments across multiple health domains. Lastly, future work will also involve incorporating explainable AI (XAI) techniques beyond SHAP values to further enhance clinical trust and model transparency. These efforts will ensure that the decision-making process remains interpretable to healthcare professionals, aiding in informed clinical decisions.

In summary, the proposed hybrid RF–NN framework lays a strong foundation for advanced healthcare analytics, with promising avenues for real-world deployment, personalization of healthcare strategies, and contribution to the broader domain of precision medicine.

Author Contributions

M. Chandraman: conceptualization, writing – original draft. **N. Santhiyakumari:** methodology, writing – original draft. **K. V. M. Shree:** investigation, writing – original draft. **Malathi Murugesan:** software, writing – original draft. **S. Kumarganesh:** methodology, writing – original draft. **P. Rishabavarthani:** formal analysis, writing – original draft. **B. Thiyaneswaran:** methodology, writing – original draft. **K. Martin Sagayam:** validation, writing – original draft. **Digvijay Pandey:** conceptualization, visualization, writing – original draft. **Binay Kumar Pandey:** methodology, validation, writing – original draft. **Suresh Kumar Sahani:** methodology, writing – original draft.

Acknowledgments

The authors would like to express gratitude to Knowledge Institute of Technology, Salem.

Ethics Statement

The authors have nothing to report.

Consent

The authors have nothing to report.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

1. T. O. Togunwa, A. O. Babatunde, and K.-R. Abdullah, "Deep Hybrid Model for Maternal Health Risk Classification in Pregnancy: Synergy of ANN and Random Forest," *Frontiers in Artificial Intelligence* 6 (2023): 1213436, https://doi.org/10.3389/frai.2023.1213436.

2. S. Saeedbakhsh, M. Sattari, M. Mohammadi, J. Najafian, and F. Mohammadi, "Diagnosis of Coronary Artery Disease Based on Machine Learning Algorithms: Support Vector Machine, Artificial Neural Network, and Random Forest," *Advanced Biomedical Research* 12, no. 1 (2023): 51.

3. M. Kavitha, G. Gnaneswar, R. Dinesh, Y. R. Sai, and R. S. Suraj, "Heart Disease Prediction Using Hybrid Machine Learning Model," in *2021 6th International Conference on Inventive Computation Technologies (ICICT)* (IEEE, 2021), 1329–1333, https://doi.org/10.1109/ICICT50816.2021.9358681.

4. A. Naseer, M. M. Khan, F. Arif, W. Iqbal, A. Ahmad, and I. Ahmad, "An Improved Hybrid Model for Cardiovascular Disease Detection Using Machine Learning in IoT," *Expert Systems* 42 (2023): e13520, https://doi.org/10.1111/exsy.13520.

5. H. El Hamdaoui, S. Boujraf, N. E. H. Chaoui, B. Alami, and M. Maaroufi, "Improving Heart Disease Prediction Using Random Forest and AdaBoost Algorithms," *International Journal of Online and Biomedical Engineering* 17, no. 11 (2021): 60–75, https://doi.org/10.3991/ijoe.v17i11.24781.

6. S. Dhar, K. Roy, T. Dey, P. Datta, and A. Biswas, "A Hybrid Machine Learning Approach for Prediction of Heart Diseases," in *2018 4th International Conference on Computing Communication and Automation (ICCCA)* (Elsevier, 2018), 1–6, https://doi.org/10.1109/CCAA.2018.8777564.

7. F. Khozeimeh, D. Sharifrazi, N. H. Izadi, et al., "RF-CNN-F: Random Forest With Convolutional Neural Network Features for Coronary Artery Disease Diagnosis Based on Cardiac Magnetic Resonance," *Scientific Reports* 12, no. 1 (2022): 11178, https://doi.org/10.1038/s41598-022-15542-5.

8. C. Guo, J. Zhang, Y. Liu, et al., "Recursion Enhanced Random Forest With an Improved Linear Model (RERF-ILM) for Heart Disease Detection

on the Internet of Medical Things Platform," *IEEE Access* 8 (2020): 59247–59256, https://doi.org/10.1109/ACCESS.2020.2983039.

9. R. Torthi, A. D. K. Marapatla, S. Mande, H. K. V. Gadiraju, and C. Kanumuri, "Heart Disease Prediction Using Random Forest-Based Hybrid Optimization Algorithms," *International Journal of Intelligent Engineering and Systems* 17, no. 2 (2024): 964–976.

10. N. Sharma, L. Malviya, A. Jadhav, and P. Lalwani, "A Hybrid Deep Neural Net Learning Model for Predicting Coronary Heart Disease Using Randomized Search Cross-Validation Optimization," *Decision Analytics Journal* 9 (2023): 100331, https://doi.org/10.1016/j.dajour.2023.100331.

11. G. M. Rao, D. Ramesh, V. Sharma, et al., "AttGRU-HMSI: Enhancing Heart Disease Diagnosis Using Hybrid Deep Learning Approach," *Scientific Reports* 14, no. 1 (2024): 7833, https://doi.org/10.1038/s41598-024-07983-6.

12. A. Al Bataineh and S. Manacek, "MLP-PSO Hybrid Algorithm for Heart Disease Prediction," *Journal of Personalized Medicine* 12, no. 8 (2022): 1208, https://doi.org/10.3390/jpm12081208.

13. P. K. Shrivastava, M. Sharma, and A. Kumar, "HCBiLSTM: A Hybrid Model for Predicting Heart Disease Using CNN and BiLSTM Algorithms," *Measurement: Sensors* 25 (2023): 100657, https://doi.org/10.1016/j.measen.2023.100657.

14. K. Kanagarathinam, D. Sankaran, and R. Manikandan, "Machine Learning-Based Risk Prediction Model for Cardiovascular Disease Using a Hybrid Dataset," *Data & Knowledge Engineering* 140 (2022): 102042, https://doi.org/10.1016/j.datak.2021.102042.

15. M. S. Al Reshan, S. Amin, M. A. Zeb, et al., "A Robust Heart Disease Prediction System Using Hybrid Deep Neural Networks," *IEEE Access* 11 (2023): 121574–121591, https://doi.org/10.1109/ACCESS.2023.3308392.

16. V. Pavithra and V. Jayalakshmi, "Hybrid Feature Selection Technique for Prediction of Cardiovascular Diseases," *Materials Today: Proceedings* 81 (2023): 336–340, https://doi.org/10.1016/j.matpr.2023.07.0752023.

17. O. Taylan, A. S. Alkabaa, H. S. Alqabbaa, E. Pamukçu, and V. Leiva, "Early Prediction in Classification of Cardiovascular Diseases With Machine Learning, Neuro-Fuzzy and Statistical Methods," *Biology* 12, no. 1 (2023): 117, https://doi.org/10.3390/biology120100117.

18. H. H. Alalawi and M. S. Alsuwat, "Detection of Cardiovascular Disease Using Machine Learning Classification Models," *International Journal of Engineering Research and Technology* 10, no. 7 (2021): 151–157.

19. A. Premarathne, M. N. Halgamuge, R. Samarakody, and A. Nirmalathas, "Real-Time Cryptocurrency Price Prediction by Exploiting Iot Concept and Beyond: Cloud Computing, Data Parallelism and Deep Learning," *International Journal of Advanced Computer Science and Applications* 11, no. 3 (2020): 8–16, https://doi.org/10.14569/IJACSA.2020.0110302.

20. L. Chandrika and K. Madhavi, "A Hybrid Framework for Heart Disease Prediction Using Machine Learning Algorithms," *E3S Web of Conferences* 309 (2021): 01043, https://doi.org/10.1051/e3sconf/202130901043.

21. T. Senthilkumar, S. Kumarganesh, P. Sivakumar, and K. Periyarselvam, "Primitive Detection of Alzheimer's Disease Using Neuroimaging: A Progression Model for Alzheimer's Disease: Their Applications, Benefits, and Drawbacks," *Journal of Intelligent & Fuzzy Systems* 43, no. 4 (2022): 4431–4444, https://doi.org/10.3233/JIFS-220628.

22. B. K. Pandey and D. Pandey, "Securing Healthcare Medical Image Information Using Advance Morphological Component Analysis, Information Hiding Systems, and Hybrid Convolutional Neural Networks on IoMT," *Computers in Biology and Medicine* 185 (2025): 109499.

23. S. Kumar, S. Shastri, and V. Mansotra, "Neuro_DeFused-Net: A Novel Multi-Scale 2DCNN Architecture Assisted Diagnostic Model for Parkinson's Disease Diagnosis Using Deep Feature-Level Fusion of Multi-Site Multi-Modality Neuroimaging Data," *Computers in Biology and Medicine* 190 (2025): 110029.

24. A. Sundar Raj, S. Senthilkumar, R. Radha, and R. Muthaiyan, "Enhanced Recurrent Capsule Network With Hybrid Optimization Model for Shrimp Disease Detection," *Scientific Reports* 15 (2025): 10400, https://doi.org/10.1038/s41598-025-94413-3.

25. M. Malathi, J. T. S. Laseetha, S. Senthilkumar, and K. Hariprasath, "Glaucoma Disease Detection Using Stacked Attention U-Net and Deep Convolutional Neural Network," *Journal of Intelligent & Fuzzy Systems* 45, no. 1 (2023): 1603–1616.

26. D. Pandey and B. K. Pandey, "An Efficient Deep Neural Network With Adaptive Galactic Swarm Optimization for Complex Image Text Extraction," in *Process Mining Techniques for Pattern Recognition* (CRC Press, 2022), 121–137.

27. M. Sharma, R. Talwar, D. Pandey, V. K. Nassa, B. K. Pandey, and P. Dadheech, "A Review of Dielectric Resonator Antennas (DRA)-based RFID Technology for Industry 4.0," *Robotics and Automation in Industry 4.0* 1, (2024): 303–324.

28. T. K. Revathi, B. Sathiyabhama, S. Sankar, D. Pandey, B. K. Pandey, and P. Dadeech, "An Intelligent Model for Coronary Heart Disease Diagnosis," in *Networking Technologies in Smart Healthcare* (CRC Press, 2022), 309–327.

29. B. Khan, A. Hasan, D. Pandey, R. J. M. Ventayen, B. K. Pandey, and G. Gowwrii, "Fusion of Datamining and Artificial Intelligence in Prediction of Hazardous Road Accidents," in *Machine Learning and Iot for Intelligent Systems and Smart Applications* (CRC Press, 2021), 201–223.

30. R. Sengupta, D. Sengupta, D. Pandey, B. K. Pandey, V. K. Nassa, and P. Dadeech, "A Systematic Review of 5G Opportunities, Architecture and Challenges," *Future Trends in 5G and 6G: Challenges, Architecture, and Applications* (CRC Press, 2021), 247.

31. P. K. Rai, D. Pandey, and B. K. Pandey, "The Future of Enterprise and Innovation Is Compute as a Commodity, or CaaC," *Partners Universal International Research Journal* 3, no. 2 (2024): 89–94.

32. R. U. Maheshwari, B. Paulchamy, B. K. Pandey, and D. Pandey, "Enhancing Sensing and Imaging Capabilities Through Surface Plasmon Resonance for Deepfake Image Detection," *Plasmonics* 20 (2025): 2945–2964, https://doi.org/10.1007/s11468-024-02492-1.

33. M. Abdullahi, A. Ahmad, B. K. Pandey, and D. Pandey, "Digital Currency Adoption: A Comparative Analysis of Global Trends and the Nigerian Enaira Digital Currency," *SN Computer Science* 5, no. 5 (2024): 602.

34. D. Pandey, B. K. Pandey, and T. Ahmad, eds., *Security Issues and Solutions in 6G Communications and Beyond* (IGI Global Scientific Publishing, 2024), https://doi.org/10.4018/979-8-3693-2931-3.

35. B. Pandey, A. George, S. Tiwari, S. Albermany, and H. Hung, eds., *Brain Computer Interfaces and Applications in Business* (IGI Global Scientific Publishing, 2025), https://doi.org/10.4018/979-8-3373-5122-3.