International Journal of Scientific Research & Engineering Trends
Volume 10, Issue 5, Sept-Oct-2024, ISSN (Online): 2395-566X

Enhancing Cardiovascular Disease Prediction with XAI Technique Using Machine Learning

P. Sravani, V.Charishma, N.Padmavathi, SK. Abdul Khadar, S.Rajeswari, Assistant Professor Dr.N.Chandrasekhar
Electronics and Communication Engineering
GMR Institute of Technology ,Rajam

Abstract- Globally, coronary diseases (CV) are several of the most significant causes of demise, improvements in predictive healthcare technologies are imperative. The goal of this study is to improve the predictability and interpretability of cardiovascular disease prediction models by combining machine learning methods with Explainable Artificial Intelligence (XAI). To create reliable predictive models, we investigate a range of machine learning algorithms, such as ensemble approaches, logistic regression, and XG-Boost. But while though precision is crucial, these predictions' interpretability is just as crucial for therapeutic use. Our goal is to make model procedures for making decisions concise and intelligible for physicians by utilising XAI techniques like SHAP (Shapley Additive Explanations) and LIME (Local Interpretable Model-agnostic Explanations). Using a real-world CVD dataset, our tests demonstrate that XAI-enhanced models do not not only increase the accuracy of predictions but also identify important variables affecting heart function. By providing a workable framework for using interpretable machine learning models in healthcare, this study advances the discipline and may result in better clinical judgements and more individualised patient care.The accuracy of the Random forest-CARDIO system is assessed against the Framingham heart disease dataset using the Colab Simulator. In the experiment, Random forest demonstrated a significant accuracy score of 91.38%, which is appreciably better than alternative techniques including, XGBoost (90.01%), RNN (85.02%), GRU (85.02%) and RNN+GRU (as a combined model) (86%).

Index Terms- Random Forest Classifier, Framingham Heart Disease Dataset, Model Accuracy Comparison, XGBoost, Recurrent Neural Network (RNN), Gated Recurrent Unit (GRU), Colab Simulator.

I. INTRODUCTION

The most significant cause of morbidity and mortality, responsible for over 70% of all deaths, is CARDIOVASCULAR disease (CVD). 43% of deaths are attributable to cardiovascular disease (CVD), per the 2017 Global Burden of Disease assessment. Based on projections from the World Health Organisation (WHO), 23.6 million deaths worldwide could be attributed to heart disease and stroke alone by 2030.Therefore, it's critical to detect heart disease early in order to minimise the psychological and physical harm that it may cause to people as well as organisations. ultimately, a complex It is necessary to have system to comfortably and effectively monitor cardiac health. With vital characteristics including sensing, connectivity, linearity, intelligence, and dependability, the Internet of Things (IoTs) has become more important for the healthcare industry. A range of wearable health monitoring devices can be employed to collect current time human body details, which can then be further processed into a health record and used to diagnose, treat, and resolve postoperative cardiovascular disorders. The use of IoT in healthcare has encouraged researchers worldwide to develop intelligent applications, such as mobile healthcare, recommendations, and healthcare systems in intelligent medical systems. By 2021, it is projected that the IoT healthcare market would grow to $136 billion. By 2026, this market is projected to increase at a compound annual growth rate of 23.4%. Patients can use it to gather health data, such as their heartbeat, blood pressure, and glucose levels. It is possible to keep the gathered data on a cloud server thanks to Internet of Things technology. Thus it renders feasible to view historical and current data remotely. Furthermore, a large number of low- and middle-income countries lack access to or cannot afford the diagnostic instruments needed to identify coronary heart disease, such as electrocardiograms and computed tomography (CT) scans. Obesity, sugar consumption, smoking, and excess body fat are heart disease risk factors that are more common in in wealthy nations; nevertheless, Chronic illness rates are also on the rise in middle-class and lower-class nations. Early detection and treatment of certain heart problems can be difficult, particularly in underdeveloped countries where resources such as qualified doctors and diagnostic facilities are scarce. This may affect the accurate prognosis for a patient. In part to the complexity of physically identifying cardiac anomalies, traditional techniques, which primarily include examining the patient Data are applied. Chaos and commotion surrounding the concept of the target class are brought on by unnecessary and superfluous characteristics. Managing has an impact on the classification's correctness. these characteristics requiring a lot of work. Furthermore, the basis of the cardiac diagnostic systems in use today is the application of conventional weighted methods. Absence of theoretical They increase mean square error (MSE) because of their base. additionally lower the performance of the prediction model. To get beyond these problems, a novel CVD prediction recommendation system DEEP-CARDIO, an IoT network, is suggested for diagnosis-giving community-based recommender system advice for nutrition and treatment for heart conditions.Increasing disease prevalence, coexisting a number of illnesses, an increase in the need for medical care, age-related impairment, and expanding healthcare costs are some of the major issues facing healthcare systems today. Still, heart illness belongs to the several ailments that particularly regarded as a significant public health issue since it affects millions of people worldwide. In particular, cardiovascular disease presents a social and economic challenge to healthcare systems in addition to a medical one. The main categories of cardiovascular disease and the misconceptions associated with each are listed in Table 1. Therefore, the signs of cardiovascular Illness might be lessened and the heart's operation could involve greatly enhanced with the appropriate treatment and early detection. Additionally, it would support individualized treatment plans and early intervention, improving healthcare systems. The prognosis for cardiovascular disease can be utilized to prevent surgery and thereby lower medical expenses. However, traditional techniques for predicting cardiovascular disease in humans are either ineffective or very expensive. Thus, it is now clear that intelligent and sophisticated healthcare systems are essential, underscoring the pressing need for their development. Physicians can monitor patients remotely with the help of smart healthcare systems, which makes it easier to monitor the course of a disease over time. These intelligent systems are also essential for the detection, diagnosis, classification, forecasting, prevention, and treatment of diseases. Consequently, healthcare systems can benefit from using a variety of computational intelligence (AI) methods, particularly strategies for machine learning can be implemented in healthcare systems, which will lower fatalities from cardiovascular disease.

II. LITERATURE SURVEY

Heart disease has been predicted in multiple studies using both conventional machine learning techniques and neural networks. An overview of several strategies will be given in this section.

- A hybrid classification-based organized learning-based CDP was proposed by Yaqoob et al. [11]. After evaluation, it was discovered that the MABC-SVM technique improved 1.5% forecast accuracy, attain a 1.6% decrease in classification error, and for maximum accuracy, 17.7% more loops are needed.
- A real-time smart big data prediction method for heart disease was proposed by Safa et al. [12]. When compared with cardiac data, the HCBDA method showed an overall accuracy of 96%. On the other hand, the suggested method has very low categorization accuracy.
- A machine learning technique for CVD prediction was proposed by Bizimana et al. [14]. The results showed a 96.7% total accuracy rate. However, the MLbPM technique has a very complicated algorithms.
- The most efficient CDP based on ML was proposed by Bhatt et al. [15]. When evaluated against real-world datasets, the proposed method's total accuracy was 87.28%. However, the recommended technique's durability is very low.
- A bidirectional gated recurrent unit with optimization support was proposed by Shukla et al. [16] for a big data healthcare monitoring system. The precision and MCC achieved by the proposed HMS approach are 0.8394 and 0.719442, respectively. However, the recommended HMS technique has very poor accuracy.
- A two-way GRU network-based Doppler radar was proposed by Lu et al. [17] for accurate heart rate monitoring. In terms of heart rate detection, the state-of-the-art method has an identifying F1 score of 95.62%, whereas the recommended strategy has a high F1 score of 98.06%. The recommended approach is highly computational, though.
- A bidirectional LSTM in a CNN was proposed by Alkhawaldeh et al. [18] for the classification of cardiac. The proposed model detects positive cases with 100% accuracy and 99.4% accuracy. But in order to produce results, the DeepResidualBiLSTM model needs more execution time.
- RNN with dual architecture was proposed by Islam et al. [19] as a method for heart rate classification. With the HARDC technique, the corresponding results are 99.60% recall, 98.21% F1 score, 97.66% precision, and 99.60% accuracy. But the processing time for the HARDC approach is high.
- A workable and useful deep learning method for diagnosing AFin long-term ECG was proposed by Zou et al. [20]. When compared with PhysioNet, the MIF-AFNet method produced a comparable AF detection accuracy of 98.63%. On the other hand, the MIF-AFNet approach performs very poorly.

- Heart disease prediction accuracy is poor due to the previously mentioned models' insufficient learning and optimization techniques. A DL-based and Internet of Things system is being built and tested in this study to estimate the risk level of cardiovascular disorders. The techniques listed above have low prediction rates, flexibility, and accuracy in classification [1, 21, 22, 23, 24]. Due to this, more research is needed to address the problems with the current models and improve the DL models' prediction accuracy. Additionally, the preceding issues strongly encourage the development of a DEEP-CARDIO technique to address this issue by using IoT devices to identify and classify CVD [25], [26], [27], and [28].
- A smart clothing designed by Sethuraman et al. tracks and gathers important health data, including heart rate variations, stress levels, and muscular activity, and then analyses the information on a centralised cloud server.
- To measure blood sugar levels in real time, Joshi et al. have presented iGLU 2.0, a non-invasive wearable consumer gadget. The iGLU 2.0 is a wearable consumer gadget for diabetics that uses the Internet of Medical Things (IoMT) to analyse and monitor blood glucose levels. It also incorporates short-range near-infrared spectroscopy. Faisal and associates have showcased a specially designed wearable telehealth monitoring system that incorporates a database of 70 healthy individuals' gaits.
- The objective was to identify knee joint and walking patterns based on age, sex, and BMI by extracting and utilising the most informative elements.
- Similar to this, Maji et al.'s iKardo smart healthcare system includes A electrocardiogram (ECG) gadget with intelligence for automatic critical beat recognition. Multi-layer perceptrons were used by the authors to identify patterns in the data.
- Tahir et al. proposed a model for COVID-19 (an infectious illness) forecasting using patient registration slips. ResNet-50, Mask R-CNN, and R-CNN were used for the evaluation.
- A potential approach combining wearable sensors and Deep Neural Networks (DNNs) has been proposed by Hassantabar et al. for assessing the COVID-19 infectious disease.
- A medical design was developed by Nandi et al. that uses a collection of sensors to monitor several health factors and has recognised the signs of infectious COVID-19 disease.
- A framework for COVID-19 healthcare monitoring has been developed by Adhikari et al. disorder prediction using machine learning in edge networks. Adhikari et al. have suggested an iCovidCare patient monitoring system that uses the Random Ensemble Forest method in another article.
- Using According to Nandy et al., the Bag-of-Neural Network model developed a unique the Internet of Medicine framework for tracking and evaluating EEG signals in real time. Instead of focussing on data analytics at the network's edge, the majority of current studies concentrate on remote health monitoring. With an eye towards addressing this weakness, we create a novel AI-powered A smart edge networking prototype for health surveillance that can analyse and forecast transmissible diseases.

III. METHODOLOGY

A novel, community-based framework is put out in this part to improve the prognosis of heart disease and offer focused health advice. In order to categorise heart illness across several cardiovascular categories, this system makes use of sophisticated machine learning models, particularly Random Forest and XGBoost, which demonstrated high predicted accuracies of 91% and 90%, respectively. Four biosensors are used to remotely gather physiological data from patients, including critical health markers. This ensures accurate, real-time input from the patient's surroundings. The Random Forest and XGBoost models' strong pattern recognition skills are used in the framework's construction to provide accurate diagnostics and classify cardiovascular illnesses. Following a categorisation, the system produces a number of suggestions depending on the unique cardiac profile of the patient. These guidelines address important facets of patient care, such as engaging in physical activity. dietary recommendations and lifestyle modifications according to each cardiovascular disease category.

Patients are then given simple access to actionable health advice and a customised care plan using a mobile application that presents the findings and insights. In addition to encouraging early diagnosis, the suggested paradigm gives patients the tools they need to take proactive control of their health under the direction of data-driven insights. The block diagram of this improved prediction and recommendation system, which shows the process from data collection to the provision of individualised care, is shown in Fig. 1.This version combines predictive analytics with user-centric recommendations, emphasising both the practical features of patient participation through mobile health advice and the technological sophistication of Random Forest and XGBoost models.

1. Data Acquistion

Wearable sensors are used to collect physiological data from patients. Medical sensors are used for the collection of physiological data. The patient's body is equipped with sensors for heart rate, blood pressure, blood sugar, and electrocardiogram. The patient's level of physical activity is a key predictor of heart disease. In order to identify and forecast

International Journal of Scientific Research & Engineering Trends
Volume 10, Issue 5, Sept-Oct-2024, ISSN (Online): 2395-566X

illnesses, IoT-based sensors transmit the collected data to the Arduino controller. The open-source hardware and software concept is the foundation of the Arduino electronics platform. The system consists of a microcontroller, programmable hardware, and an IDE, which is portable software for writing code and adding it to the board. The signal output of the ECG, pressure, pulse, and glucose sensors is shown in Fig. 2(a)–(c), which captures the electrical signals, pressure variation, and blood glucose level from the patient's heart to assess the patient's cardiac health.

Data Preprocessing

A vital phase in machine learning is data preprocessing, which converts unprocessed data into a format that can be used for modelling. It uses a number of crucial procedures to guarantee the consistency and quality of the data. One of the first steps is dealing with missing data, which can be done by deleting incomplete rows or columns or by using imputation techniques like substituting the mean, median, or mode for missing values. In order to make categorical variables accessible in algorithms, data encoding—also known as label encoding or one-hot encoding—converts them into numeric representations. Additionally, feature scaling is required, particularly for models that use distance measures, such as neural networks or KNN, where standardisation or normalisation guarantees that all features are on the same scale. The next step is data cleaning, which includes deleting duplicates and outliers and making sure every feature has the appropriate data type. While dimensionality reduction techniques like PCA can lower the number of input variables without sacrificing crucial information, feature engineering plays a vital role in improving model performance by either inventing new, useful features or choosing the most pertinent ones. The dataset is typically separated into training and testing sections in order to evaluate model performance sets before to modelling. Cross-validation is frequently used to avoid overfitting. Resampling methods or specialised algorithms can be used to address unequal class distributions in an unbalanced dataset. Lastly, skewed data distributions can be corrected by applying data transformations such log or power transformations. All things considered, efficient data preprocessing guarantees that the dataset is clear, consistent, and ideal for building trustworthy machine learning models.

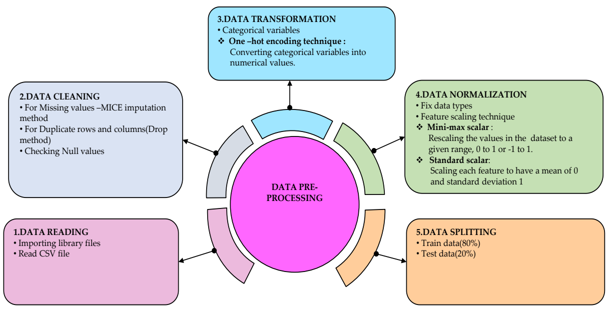


Figure 1: Stages of Data Pre-Processing

Block Diagram

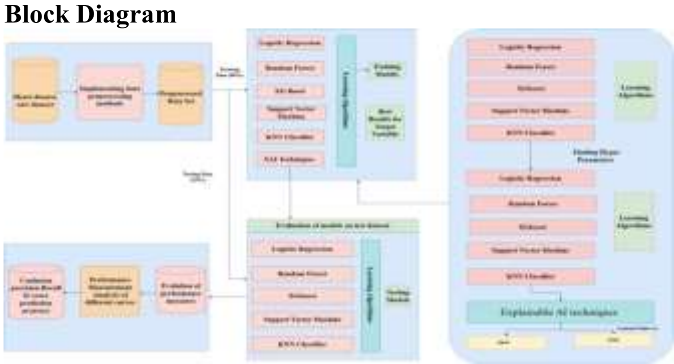


Figure 2: Block diagram of CVD Prediction

2. Cardio-Prediction Using Random Forest

A Random Forest method that can diagnose CVD is used in the implementation of a CVD prediction model. To enhance the precision of forecasts, Several decision trees are constructed during training and their outputs are combined in the Random Forest approach, an ensemble learning methodology. Using a random subset of the dataset's attributes to construct each decision tree in the forest, ensuring the model's robustness and minimizing overfitting. The input characteristics, decision trees (which serve as hidden layers), and the final aggregated output (classification) are the three main components of the model. The decision trees forecast the heart condition hazard determined by health parameters such age, gender, blood pressure, cholesterol, and smoking status using random subsets of data. At each decision point within the trees, the model evaluates the importance of the features and makes splits accordingly. The trees check the health data at time t t and compare it with other observations. Where the data is imported from health reports like $y\ t$ y t, where $t = (1,2,...,n)$ t=(1,2,...,n), represents the number of health metrics observed during a period, which are used as the input features for the Random Forest model. For each decision tree, the predicted output is set to $h\ t$ h t, while the model evaluates multiple trees to arrive at a final aggregated prediction using majority voting. By combining the results of each tree, the model achieves improved accuracy in diagnosing CVD, reaching a performance of 91% accuracy.

Random Forest Classifier: Using a set of decision trees, the Random Forest approach classifies or predicts data according to the input dataset's attributes. In the forest, every decision tree casts a vote for a classification; the class with the most votes becomes the final prediction.

Equations (1) and (2), used earlier for the update and reset gates, are now substituted by the ensemble learning process of Random Forest.

Bootstrap Aggregation: The Random Forest creates several decision trees using different bootstrapped subsets of the data. Each tree is trained independently on its respective subset.

International Journal of Scientific Research & Engineering Trends
Volume 10, Issue 5, Sept-Oct-2024, ISSN (Online): 2395-566X

Voting Mechanism

Once the forest is trained, each tree predicts the outcome of the input. The Random Forest's output for classification problems is the prediction mode (majority voting). The average prediction would apply to regression tasks.

The final prediction y^t\hat{y}_ty^t at time ttt is calculated as:

$$\hat{y}_t = 1/n \sum \quad Ti(xt)$$

Where:

- Ti(xt)T_i(x_t)Ti(xt) is the prediction from the iii-th decision tree,
- xtx_txt represents the input features at time ttt,
- nnn is the total number of trees in the forest.

Feature Importance

- Random Forest also provides the capacity to gauge the significance of features. This is done by looking at how much each feature decreases the impurity across the trees.

The importance IjI_jIj of feature jjj can be stated as:

$$Ij = i = 1 \sum n \Delta Gi(j)$$

Where:

- ΔGi(j)\Delta G_i(j)ΔGi(j) is the reduction in impurity in tree iii by splitting on feature jjj,
- nnn is the number of trees.

This allows for a clearer understanding of which features are most influential in the classification process.

Random Forest as an Ensemble Learning Method

The Random Forest classifier can be expressed as a combination of several individual trees, each contributing to the final classification by majority voting. This method improves generalization and reduces overfitting.The final classification report includes predictions for different cardiovascular classes, such as MI (Myocardial Infarction), ACS (Acute Coronary Syndrome), AF (Atrial Fibrillation), HTN (Hypertension), and LVH (Left Ventricular Hypertrophy).

Equations

- Equation 3: Prediction by majority voting.

y^t=1n∑i=1nTi(xt)\hat{y}_t = n1i=1∑nTi(xt)

Equation 4: Feature importance calculation.

$$Ij = i = 1 \sum n \Delta Gi(j)$$

Recommendation System

This part develops a sophisticated framework that uses both Machine Learning (ML) and Deep Learning (DL) approaches to improve the prediction of cardiac disease. The suggested approach incorporates Explainable Artificial Intelligence (XAI) to guarantee interpretability and transparency in decision-making, offering concise and intelligible suggestions for the treatment of cardiac disease. Patients' physiological data, including blood pressure, heart rate, and other vital signs, are gathered by wearable biosensors. To estimate the risk of cardiovascular disease (CVD), these data are processed using advanced machine learning (ML) and deep learning (DL) models, including Random Forest, XGBoost, and deep neural networks. Patients and healthcare professionals can comprehend how these algorithms make their predictions by using XAI. By offering thorough justifications of the model's results, such as feature significance and decision paths, The predictions are trusted and held accountable by the system. Following a diagnosis, the recommender system creates personalised recommendations for dietary guidelines, lifestyle changes, and treatment regimens based on each patient's unique cardiovascular risk. The explainability component of the system gives individuals the ability to actively manage their heart health and allows medical practitioners to make well-informed decisions. The block diagram of this XAI-enhanced prediction and recommendation system is shown in Fig. 1, emphasising the smooth transition from data collection to useful suggestions. With an emphasis on the combined strength of ML and DL approaches, this version emphasises the function of XAI in guaranteeing transparency in heart disease prediction while making sure that the suggestions are intelligible to patients and medical professionals alike.

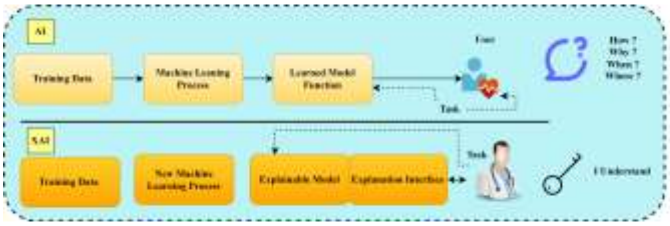


Figure 3: XAI Techniques

SHAP(SHapleyAdditiveexPlanations).

An explainable AI method based on cooperative game theory is called SHAP (SHapley Additive exPlanations). In order to equitably divide a model's prediction among its features according to their contributions, it makes use of Shapley values.

SHAP can offer both global and local interpretability by figuring out each feature's contribution, which enables users to comprehend both individual predictions and the behaviour of the entire model. This method works with a variety of models, even intricate ones like ensemble approaches and deep learning. In industries like marketing, healthcare, and finance where comprehending model behaviour, determining feature value, and maintaining fairness are essential, SHAP is particularly helpful.

LIME (Local Interpretable Model-agnostic Explanations)
The goal of LIME is to offer local justifications for the unique projections of every predictive machine learning model.It generates predictions for these samples after perturbing the input data to produce a dataset of modified samples.

The behaviour of the complex model is approximated locally around the prediction being explained by fitting a more straightforward, interpretable model (such as linear regression).

This enables LIME to show how each attribute contributes to that particular forecast. Because LIME is a model-agnostic technique, it may be used with any machine learning model, which makes it flexible and useful for comprehending particular predictions in applications like recommendation systems, fraud detection, and medical diagnostics.

IV. RESULTS AND DISCUSSION

1. Dataset Description
Framingham Dataset
A comprehensive resource with 4,241 observations across 13 variables, Data from the Framingham Heart Study is frequently used to investigate the hazards to cardiovascular health.

This dataset, which comes from the well-known Framingham Heart Study, contains important health and demographic characteristics like age, gender, education level, smoking status, BMI, and diabetes status. Along with important lifestyle indicators like smoking and alcohol consumption, it also monitors heart rate, systolic and diastolic blood pressure, total cholesterol, and glucose levels.

This information helps identify people who are at a higher risk of cardiovascular disease (CVD) throughout time and assists research and predictive modelling in this area. The dataset is widely used in statistical studies and machine learning, where it is an invaluable resource for researching risk variables, building and verifying risk prediction models, and suggestions for better cardiovascular results that are based on patient-centered health.

2. Performance-Metrics
The performance and usefulness of the developed CARDIO system were examined in order to evaluate the accuracy, precision, sensitivity, specificity, and measurement function in connection to the patient's risk of heart disease.

The Random forest models highest accuracy of 91% compared to the other Machine learning and Deep Learning models. 91.38% accuracy was attained using the Random Forest model in the project "Enhancing Heart Disease Prediction Using ML and DL." Based on the given images, the performance metrics are summarised as follows:

Confusion Matrix
According to the confusion matrix, the Random Forest model accurately detected 534 true positives (CHD) and 600 true negatives (No CHD). In detecting cases of heart disease, there was a balance between sensitivity and specificity, as seen by the 53 false positives and 54 false negatives.

ROC Curve
The model's capacity for class distinction is demonstrated by the Receiver Operating Characteristic (ROC) curve. The Random Forest model performs exceptionally well in differentiating between instances of CHD and those without CHD, approaching a perfect classifier with an AUC (Area Under Curve) of 0.97.

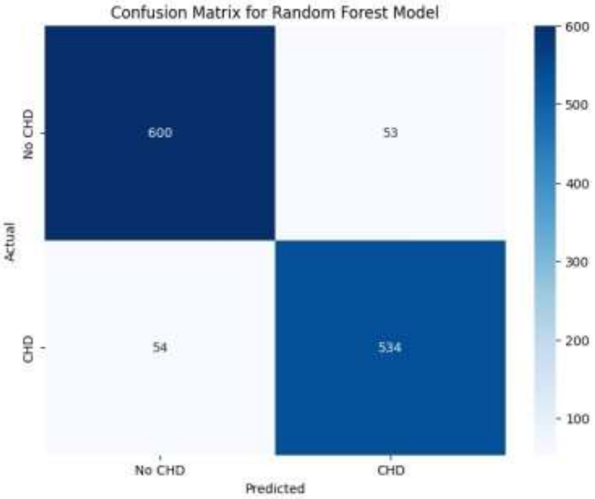
Precision-Recall Curve
The Random Forest model retains a high rate of correct positive predictions even as recall rises, as evidenced by the Precision-Recall curve, which displays good precision over the majority of recall levels. This quality is crucial for healthcare forecasting, as accuracy is necessary to lower false-positive diagnosis rates.

Learning Curve
Strong generalisation is indicated by the convergence of the training and validation curves in a learning curve for a Random Forest model with 91.38% accuracy. Usually, both curves plateau at almost the same accuracy, indicating a balanced model free from overfitting.

While the validation curve steadily improves and demonstrates stability with more data, the training curve may begin high. Ineffective learning without memorisation is indicated by a tiny gap between the curves. Overall, the model's robustness is confirmed by its high and steady accuracy on unseen data.

A thorough understanding of the Random Forest model's performance is offered by the combination of the confusion matrix, learning, precision-recall, and ROC curves.

Strong classification capacity is indicated by the ROC curve's high true positive rates, which are corroborated by the precision-recall curve's balanced handling of positive predictions. The convergence of the learning curve during training and validation points to efficient generalisation that avoids overfitting.



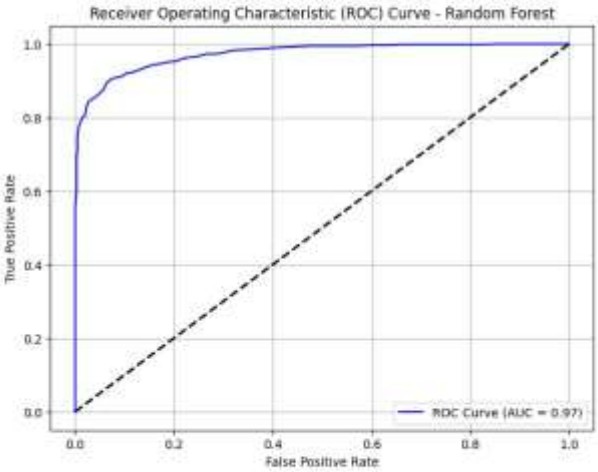
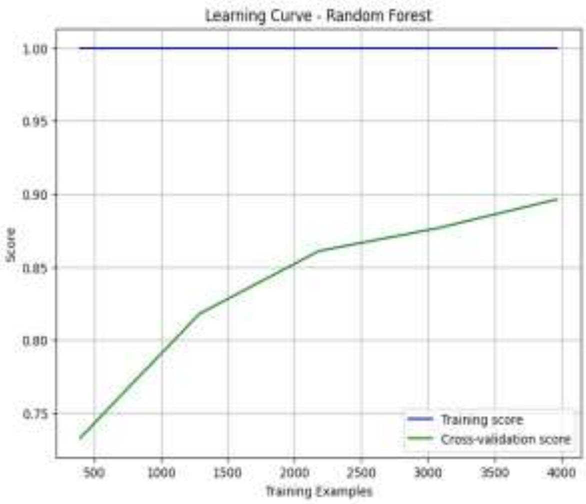


Fig 4: Confusion matrix and ROC curve for random forest model



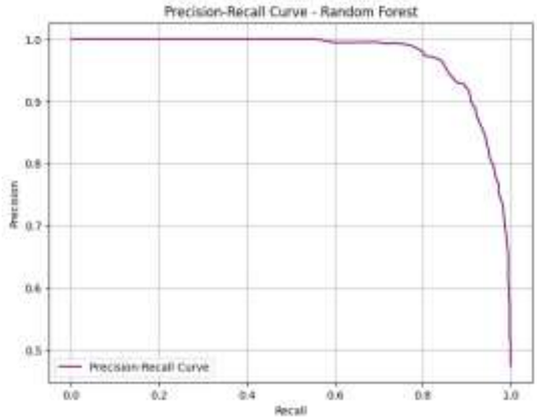
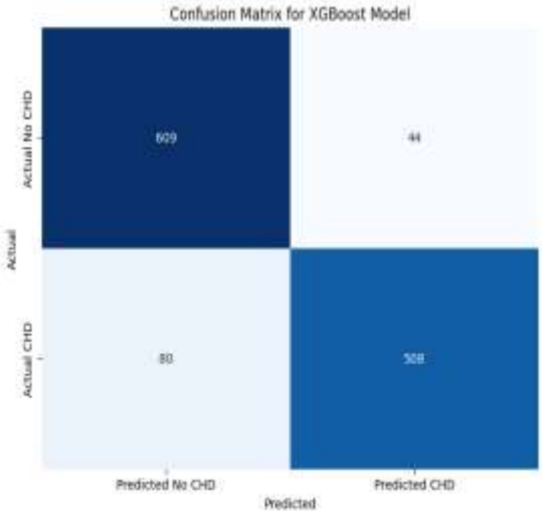


Fig 5: Learning curve and precision-recall curve for random forest model



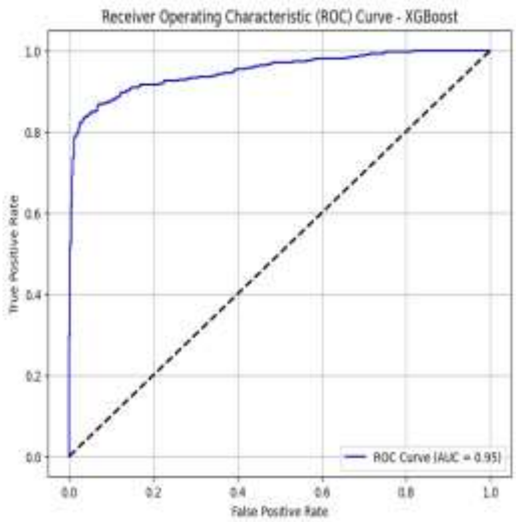


Fig 6: Confusion Matrix and ROC Curve for XGBoost Classifier

International Journal of Scientific Research & Engineering Trends
Volume 10, Issue 5, Sept-Oct-2024, ISSN (Online): 2395-566X

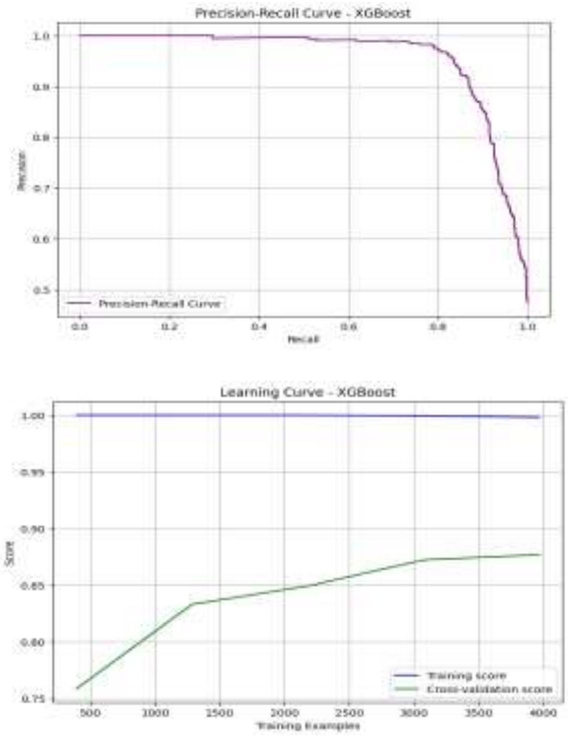


Fig 7: precision recall and learning curve for XG-Boost Classifier.

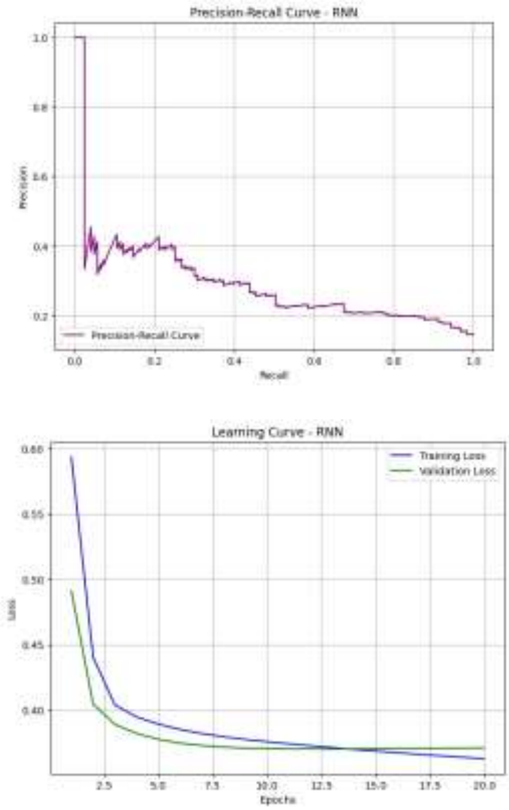


Fig 9: Precision-Recall Curve and Learning Curve for RNN Classifier

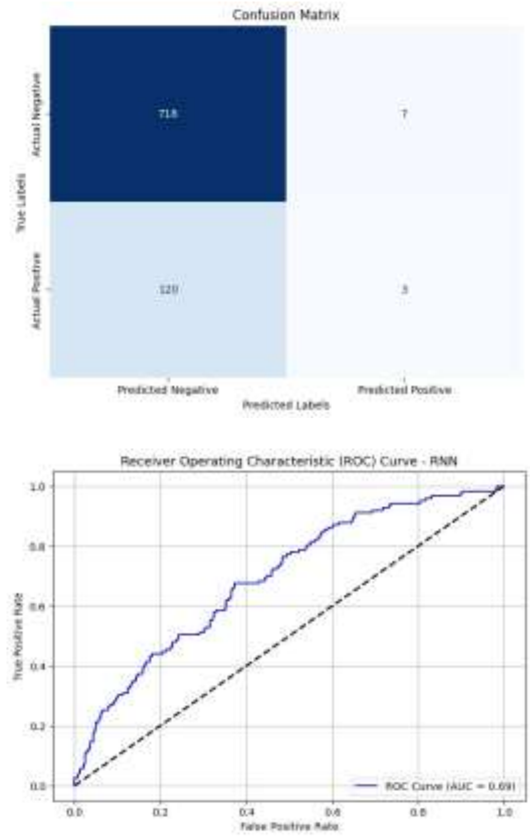


Fig 8: Confusion Matrix and ROC curve for RNN Classifier

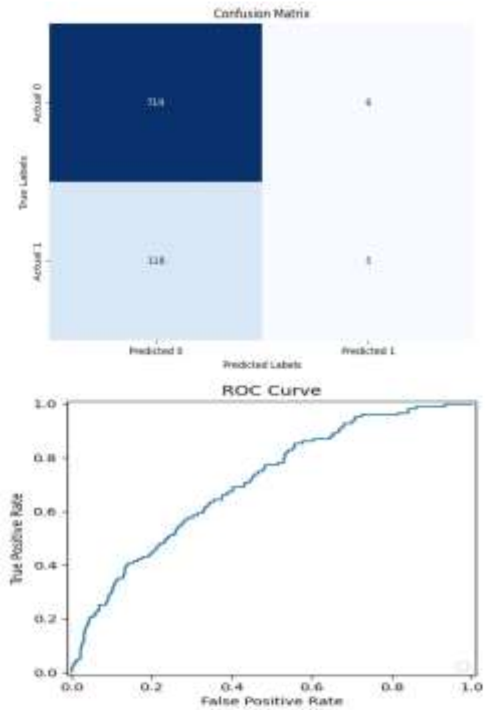


Fig 10: Confusion Matrix and ROC curve for Bi-GRU Classifier

International Journal of Scientific Research & Engineering Trends
Volume 10, Issue 5, Sept-Oct-2024, ISSN (Online): 2395-566X

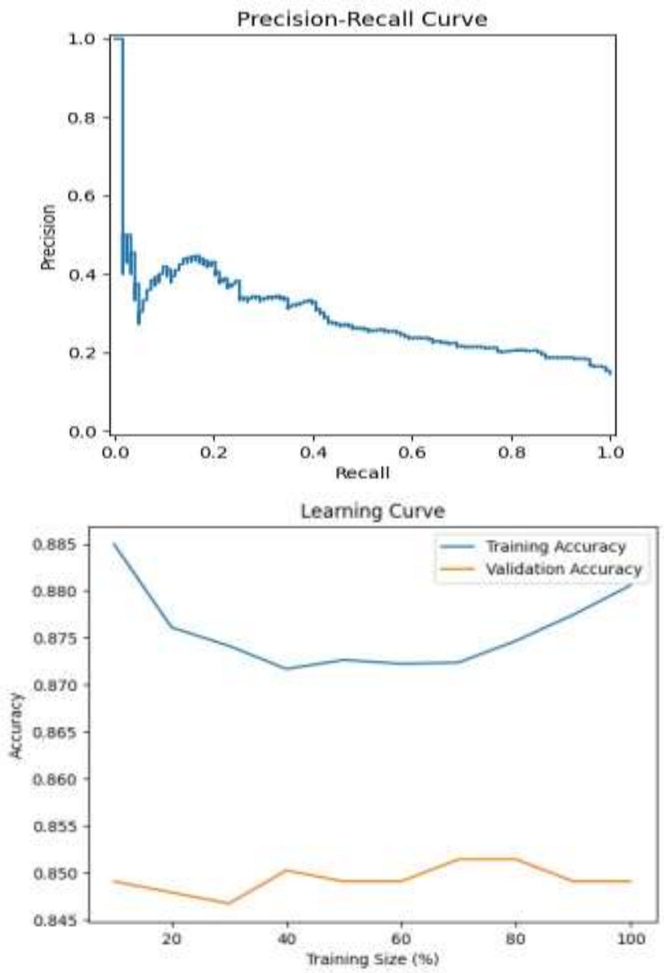


Fig 11: Precision-Recall Curve and Learning curve for Bi-GRU Classifier

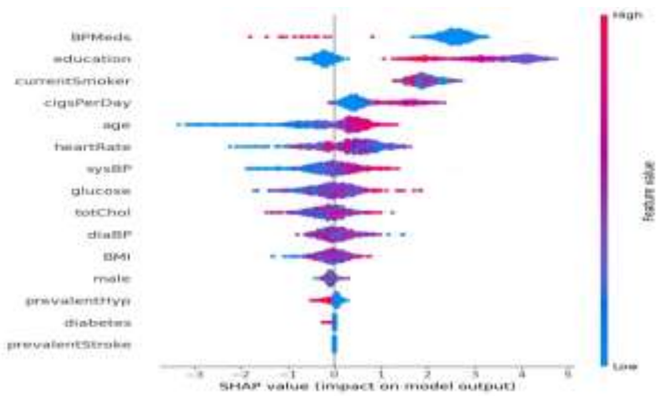


Fig 12: SHAP XAI Ensembling Techniques

Table 12: Accuracy Classification of ML and DL Models

Parameters	Random Forest Classifier	XG-BOOST Classifier	Logistic Regression Classifier	SVM Classifier	KNN Classifier	RNN Classifier	GRU Classifier	RNN+GRU Classifier	Bi-GRU Classifier
Precision	92%	88%	74%	80%	87%	86%	86%	86%	86%
Recall	92%	93%	68%	71%	66%	99%	99%	100%	99%
F1-score	92%	91%	71%	75%	75%	92%	92%	92%	92%
Accuracy	91.38%	90.01%	71%	75.26%	77.12%	85.02%	85.02%	86%	85.38%

V. CONCLUSION

The study concludes by showing how well machine learning (ML) and deep learning (DL) models can predict cardiac illness with high accuracy, especially when combined with an Internet of Things (IoT) sensor-powered web interface for real-time data collection. The Random Forest and XGBoost models demonstrate the strong performance of machine learning models in detecting patterns linked to heart disease, with 91% and 90.1% accuracy, respectively. Despite achieving somewhat lower accuracies (85.02% for RNN and GRU and 85.73% for the combined RNN+GRU), the deep learning models (RNN, GRU, and RNN+GRU) nevertheless offer insightful information by identifying temporal relationships in the data. Real-time monitoring and predictive analysis are made possible by the incorporation of these models into a web interface, which makes it possible to give patients and healthcare professionals early alerts and preventative insights. This method improves the practical applicability and scalability of cardiac disease prediction systems by increasing the prediction accuracy and utilising IoT-based sensors to collect continuous, real-world health data. All things considered, the use of both ML and DL models via an Internet of Things-enabled online platform presents a viable way to diagnose and treat cardiac disease early on, which could improve patient outcomes and lower medical expenses.

REFERENCES

1. D.Zhang, X. Liu, J. Xia, Z. Gao, H. Zhang, and V. H. C. de Albuquerque, "A physics-guided deep learning approach for functional assessment of cardiovascular disease in IoT-based smart health," IEEE Internet Things J., vol. 10, no. 21, pp. 18505–18516, 2023.
2. K.-P. Kresoja, M. Unterhuber, R. Wachter, H. Thiele, and P. Lurz,"A cardiologist's guide to machine learning in cardiovascular disease prognosis prediction," Basic Res. Cardiol., vol. 118, no. 1, p. 10,Mar. 2023.
3. S. Nandy, M. Adhikari, V. Balasubramanian, V. G. Menon, X. Li, and M. Zakarya, "An intelligent heart disease prediction system based on swarm-artificial neural network," Neural Comput. Appl., vol. 35, no. 20, pp. 14723–14737, Jul. 2023.
4. V. Chaurasia and A. Chaurasia, "Novel method of characterization of heart disease prediction using sequential feature selection based ensemble technique," Biomed. Mater. Devices, vol. 1, no. 2,pp. 932–941, 2023.
5. A. Dogan, Y. Li, C. Peter Odo, K. Sonawane, Y. Lin, and C. Liu,"A utility-based machine learning-driven personalized lifestyle recommendation for cardiovascular disease prevention," J. Biomed. Informat.,vol. 141, May 2023, Art. no. 104342.
6. N. Chandrasekhar and S. Peddakrishna, "Enhancing heart disease prediction accuracy through machine learning techniques and optimization,"Processes, vol. 11, no. 4, p. 1210, Apr. 2023.
7. X. Wei, C. Rao, X. Xiao, L. Chen, and M. Goh, "Risk assessment of cardiovascular disease based on SOLSSA-CatBoost model," Expert Syst.Appl., vol. 219, Jun. 2023, Art. no. 119648.
8. B. Paul and B. Karn, "Heart disease prediction using scaled conjugate gradient backpropagation of artificial neural network," Soft Comput.,vol. 27, no. 10, pp. 6687–6702, May 2023.
9. D. Hassan, H. I. Hussein, and M. M. Hassan, "Heart disease prediction based on pre-trained deep neural networks combined with principal component analysis," Biomed. Signal Process. Control, vol. 79, Jan. 2023, Art. no. 104019.
10. J. Kaur, B. S. Khehra, and A. Singh, "Back propagation artificial neural network for diagnose of the heart disease," J. Reliable Intell. Environments, vol. 9, no. 1, pp. 57–85, Mar. 2023.
11. M. M. Yaqoob, M. Nazir, M. A. Khan, S. Qureshi, and A. Al-Rasheed, "Hybrid classifier-based federated learning in health service providers for cardiovascular disease prediction," Appl. Sci., vol. 13, no. 3, p. 1911, Feb. 2023.
12. M. Safa, A. Pandian, H. L. Gururaj, V. Ravi, and M. Krichen, "Real time health care big data analytics model for improved QoS in cardiac disease prediction with IoT devices," Health Technol., vol. 13, no. 3, pp. 473–483, Jun. 2023.
13. M. Ozcan and S. Peker, "A classification and regression tree algorithm for heart disease modeling and prediction," Healthcare Anal., vol. 3,Nov. 2023, Art. no. 100130.
14. P. C. Bizimana, Z. Zhang, M. Asim, and A. A. A. El-Latif, "An effective machine learning-based model for an early heart disease prediction," BioMed Res. Int., vol. 2023, pp. 1–11, Apr. 2023.
15. C. M. Bhatt, P. Patel, T. Ghetia, and P. L. Mazzeo, "Effective heart disease prediction using machine learning techniques," Algorithms, +vol. 16, no. 2, p. 88, Feb. 2023.