Research Paper

A Hybrid Multimodal Explainable Deep Learning Framework for Early Heart Disease Detection Using ECG, Retinal Images, and Clinical Features

Er. Mandeep Kaur[1], Dr. Khushboo Bansal[2] and Dr. Sharanpreet Kaur[3]

[1]Er. Mandeep Kaur, Research Scholar, Computer Science and Engineering, Desh Bhagat University, Punjab, India
[2]Dr. Khushboo Bansal, Deputy Director, Faculty of Engineering, Desh Bhagat University, Punjab, India
[3]Dr. Sharanpreet Kaur, Assistant Professor, Gulzar Group of Institutions, Khanna, Punjab, India
Email: deepmaan0122@gmail.com, erkhushboo.1985@gmail.com and sharn.preet2008@gmail.com
***Corresponding author**: Er. Mandeep Kaur, deepmaan0122@gmail.com

Abstract

Cardiovascular diseases are amongst the leading cause of deaths in the world today, stressing the importance of reliable and early prediction mechanisms. Classical prediction methods use a single modality technique to identify heart disease patients. This paper seeks to propose a new framework known as a Hybrid Multimodal Explainable Deep Learning Model to predict heart diseases at an early stage using ECG signals, retinal fundus images, and clinical information. In the proposed model, ECG signal processing branch is used to process temporal signal information, a convolutional neural network for the retinal images branch to extract visual data, and a dense neural network for processing clinical parameters. Feature extraction and fusion were done using multimodal features integration method. The proposed model was evaluated using publicly available databases of ECG, retinal images, and clinical information. The model showed a better classification accuracy of 98%. Precision, recall, F1-score, ROC analysis, and confusion matrix performance evaluation were other evaluation techniques that proved the efficiency of the model.

Keywords: Medical, Cardiovascular, Artificial Intelligence, Explainable AI

Introduction

Heart disease, is one of the major causes of mortality and morbidity, still represents a significant burden in terms of healthcare. In spite of progress in terms of diagnostic measures and treatments, early identification of heart diseases still remains a significant problem since many individuals do not display any visible symptoms when the heart disease begins developing [1], [2]. Based on the recent studies, the mortality rate associated with cardiovascular disease is increasing due to the different changes in people's lifestyles and the aging of the population [3], which necessitates advanced methods for the detection of the disease. Early identification of the disease would reduce the mortality rate significantly.

Nowadays, Artificial Intelligence (AI), machine learning (ML), and deep learning (DL) have become one of the main approaches that can be used to detect diseases [4], [5]. ML-based prediction approaches that use structured clinical data and features like blood pressure, cholesterol, age, and heart rate have exhibited an acceptable level of performance in predicting heart diseases [6]. Nonetheless, ML-based methods may not be able to learn complicated physiological patterns in cases of cardiovascular diseases owing to their usage of single-modal data. The application of DL in diagnosing diseases has proven superior due to its ability to learn discriminatory features in medical data [7].

The application of electrocardiogram (ECG) for diagnosing heart diseases is popular due to the valuable data that it provides on the state of electrophysiology of the heart. Convolutional neural network (CNN) and bidirectional long short-term memory (BiLSTM), for example, prove to be efficient in processing and classification of ECG data [8], [9]. Despite the usefulness of the diagnostic information provided by ECG-based solutions, the information from ECG alone might not be sufficient for determining the state of the whole cardiovascular system of a patient.

The study of retinas has become an increasingly important technique for non-invasive diagnostics since the blood vessels in the retina have exhibited structural changes that are related to cardiovascular disease. Studies have confirmed that retinal image analysis through deep learning is an effective method in aiding the diagnosis of heart diseases [10]. In addition, several research findings have illustrated the effectiveness of multimodal learning frameworks that utilize both ECG signals and retinal image data [11], [12].

1.1 The contributions of this research are manifold

In this paper, a new multimodal framework for heart disease classification using the features derived from electrocardiogram (ECG), fundus retinal image, and structured medical information is proposed. The major difference between our multimodal method and the traditional methods with single modality is that our method makes use of the information coming from different sources of medical data.

Proposed hybrid feature fusion approach is adopted for fusing heterogeneous data gathered from ECG signals, retinal imagery, and clinical attributes. Proposed hybrid framework uses CNN-BiLSTM approach for analyzing

the ECG signal data, uses deep convolution approach for retinal imagery, and utilizes dense neural network approaches for learning from clinical attribute data.
The rest of this research paper will be discussed in the following sequence: the "Related Work" section describes the prior work, motivations behind the research and challenges faced in relation to heart disease prediction systems. The "Methodology" section talks about the proposed methodology, dataset used, preprocessing techniques and multimodal features extraction procedures. The "Results and Discussion " section provides the experimental results and comparative performance evaluation of the proposed system. At last the "Conclusion" section concludes this research work.

2. Literature review

Predicting heart diseases has been an increasingly popular research area in the last few years owing to the increasing prevalence of heart diseases and need for effective predictive models [13], [14]. Lately, certain scholars have made studies regarding using deep learning approach, multimodal learning approach, and explainable artificial intelligence approach to improve their prediction accuracy [15], [16]. The literature reveals that using a variety of physiological datasets can increase the prediction accuracy.
Many researchers have conducted studies related to heart disease classification through deep learning algorithms using electrocardiograms. In this respect, for example, Rath et al. [17] proposed a deep learning algorithm for heart disease classification based on imbalanced electrocardiogram data. Their study concluded that automatic feature extraction provided an enhanced performance in heart disease classification as compared to traditional machine learning techniques. Cheng et al. [18] proposed a deep CNN-BiLSTM technique for electrocardiogram signal classification in which convolutional neural networks were utilized for extraction of spatial local features while BiLSTM layers helped capture the temporal dependencies of electrocardiograms. This technique provided excellent results for classification. Besides, Nurmaini et al. [19] presented a stacked CNN-BiLSTM technique to diagnose abnormal cardiac patterns through beat-to-beat analysis of ECG waveforms. Furthermore, Pandey et al. [20] proposed an integrated CNN-BiLSTM technique along with M2MASC feature enhancement methods for heart disease classification.
While ECG signal data yield vital information pertaining to the functioning of the heart from a physiological perspective, it has been seen that there has been an increased focus on using multimodal learning frameworks. Li et al. [21] proposed a framework that uses multimodal features for cardiovascular disease detection and found that heterogeneous data sources helped improve the efficiency of the learning algorithm. In another research paper, Ahmad et al. [22] suggested a multimodal fusion technique for heartbeat classification using ECG signals and concluded that feature-level fusion was able to capture complementary information from different sources.
Image analysis based on retinas has become an increasingly interesting non-invasive tool to predict cardiovascular diseases. For instance, Al-Absi et al. [23] suggested the use of deep learning based on retinal images as well as DXA images to diagnose cardiovascular diseases and showed that the vascular structure of retinas can offer biomarkers of cardiovascular diseases. Additionally, Muthukumar et al. [24] suggested a multimodal system which includes ECG and fundus images for diagnosing cardiovascular diseases. They indicated improved accuracy of predictions when used in conjunction with other systems.
In addition, the hybrid deep learning models are also very successful in the development of heart disease prediction systems. Almulihi et al. [25] presented a successful hybrid deep learning model based on the ensemble technique for heart disease prediction and stated that their hybrid model has been proven to be more efficient than others by providing better predictions using various learning techniques. In this regard, Al Reshan et al. [26] proposed a very successful hybrid deep neural network model for heart disease prediction and proved that their hybrid model obtained highly accurate classification results after the optimization of their feature learning. Similar to this, Ali et al. [27] created an efficient ensemble-based deep learning model with feature fusion for smart healthcare applications. In addition to the success of AI in the field of health care, the issue of model transparency and interpretability is raised [28].

2.1 Motivation and Challenges

The increasing incidences of cardiovascular diseases (CVDs), along with the rising need for efficient methods for their early diagnosis, have motivated researchers to investigate intelligent computational approaches that could enhance the disease prediction process. In conventional diagnostic approaches, there is often heavy reliance on one type of data and, therefore, failure in capturing the interactions between various physiological factors responsible for cardiovascular diseases [29], [30]. It has been shown in recent years that the combination of different data types can be effective in providing information about the patient's health status. The ECG signal provides important information about the activity and rhythm of the heart, while retinal images represent a non-invasive method of assessing the health status of blood vessels and microvascular changes caused by CVDs. On the other hand, the structured clinical features of age, blood pressure, cholesterol level, blood sugar level, and heart rate provide useful patient-specific data. The combination of these three types of data can thus facilitate an advanced diagnostic system that improves heart disease detection.

3. Methodology

The present paper aims to design and implement a framework for early detection of heart disease using ECG, retinal images, and clinical features. This proposed framework utilizes heterogeneous data from healthcare such as electrocardiogram (ECG) signals, retinal fundus images, and clinical features. The proposed hybrid framework for early heart disease detection consists of feature extraction via deep learning and feature fusion using explainable artificial intelligence (XAI).

3.1 Proposed Framework Architecture

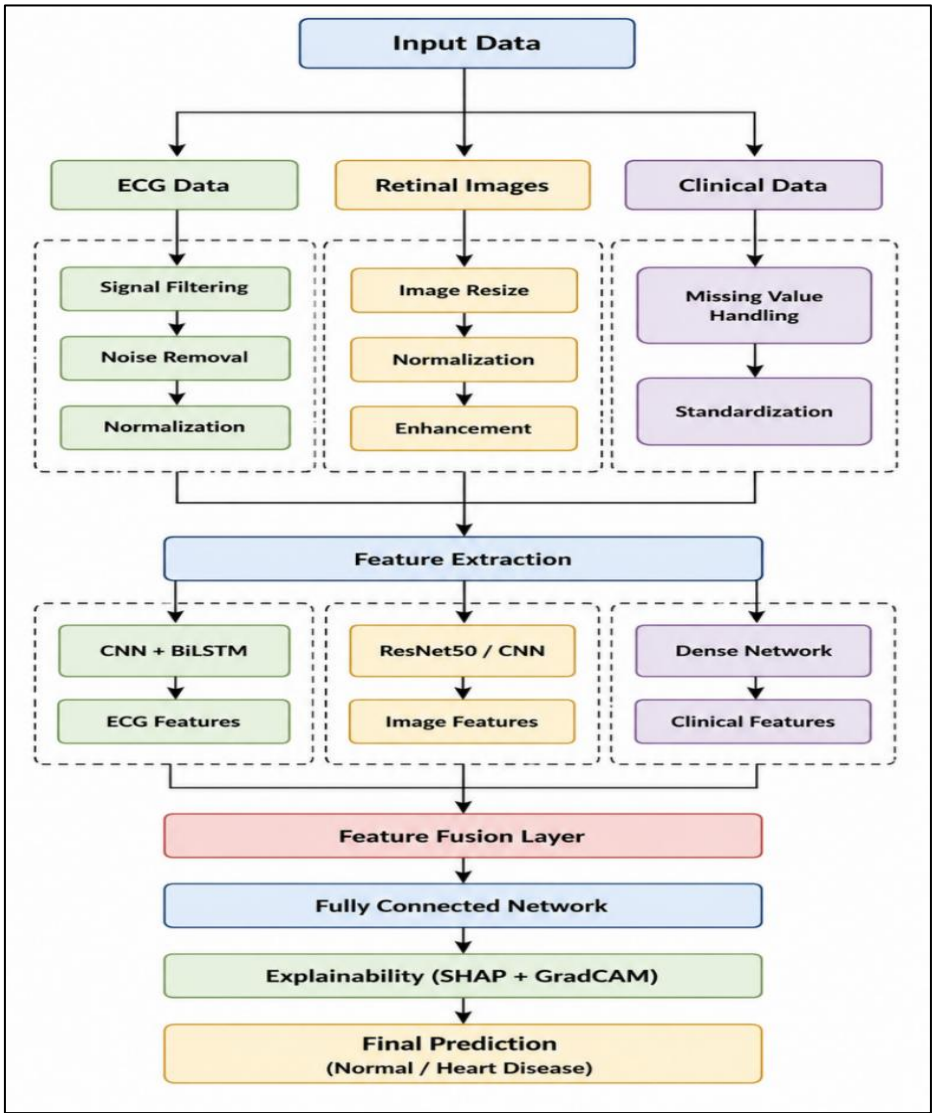


Figure 1. Architecture of the proposed hybrid multimodal explainable deep learning framework, showing preprocessing, feature extraction, feature fusion, classification, and explainability stages for ECG, retinal image, and clinical data

3.2 Data Acquisition

In the designed framework, multimodal learning is performed through using various heterogeneous datasets to identify early symptoms of heart diseases. There are three primary types of data that are considered in this project, which consist of electrocardiogram (ECG) data, retinal fundus imaging data, and structured clinical data. For ECG signal dataset collection, MIT-BIH Arrhythmia Dataset and PTB-XL ECG Dataset are chosen due to their ability to provide annotated ECG recordings featuring different heart conditions as shown in Figure 2.

The proposed framework for retinal image analysis will employ the EyePACS and APTOS Fundus Datasets, containing images of the retina taken from individuals with varied vascular and ocular diseases as shown in Figure 3. The retinal blood vessels may display anomalies related to cardiovascular diseases; hence, taking images of the retina proves helpful since retinal imaging is an efficient diagnostic tool. Through the retinal images, vascular characteristics can be obtained to aid in identifying cardiovascular risks.

Furthermore, structured data related to clinical parameters are extracted from the UCI Heart Disease Data Set, which consists of several patient-dependent features usually used in the process of assessing cardiovascular disease risk. In particular, the selected clinical parameters consist of age, gender, blood pressure, cholesterol, heart rate, blood sugar, type of chest pain, and ECG measurement results shown in Figure 4. The considered clinical parameters can be regarded as key patient-dependent parameters used in the prediction of the disease and providing enhanced learning capabilities for the proposed multimodal approach.

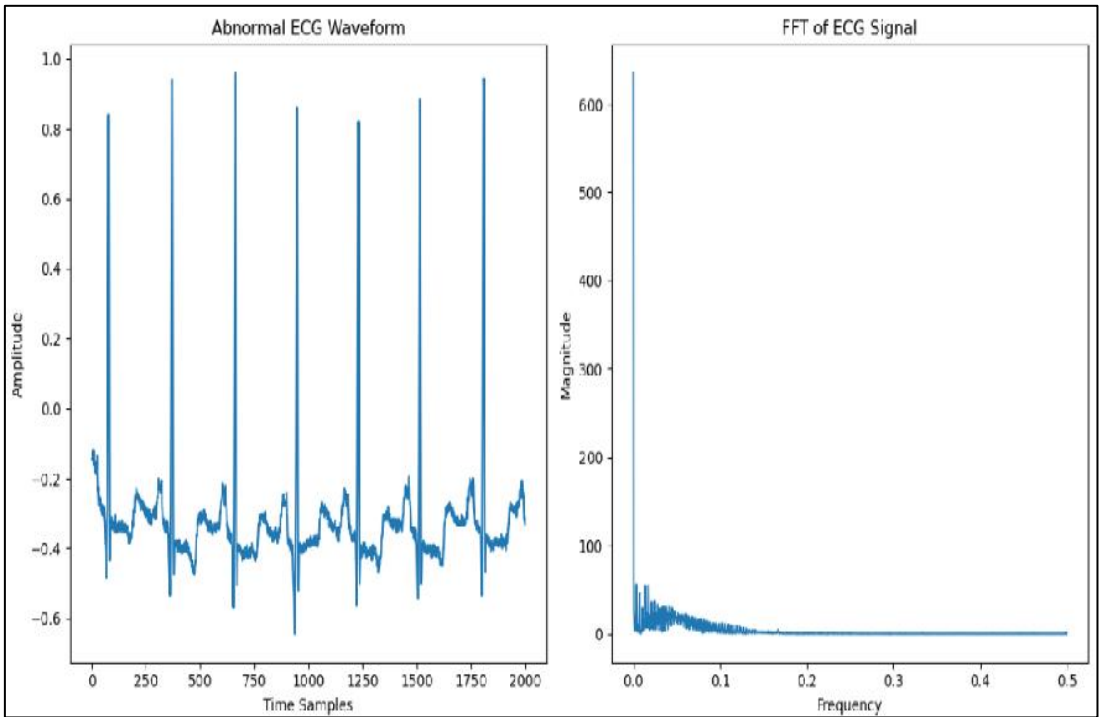


Figure 2. Shows graphs of the abnormal ECG waveform and the FFT of this waveform. The abnormal ECG waveform (left) shows variations in cardiac rhythm and signal amplitudes. The FFT (right) of this abnormal ECG waveform shows the prominent frequencies of abnormal cardiac rhythms

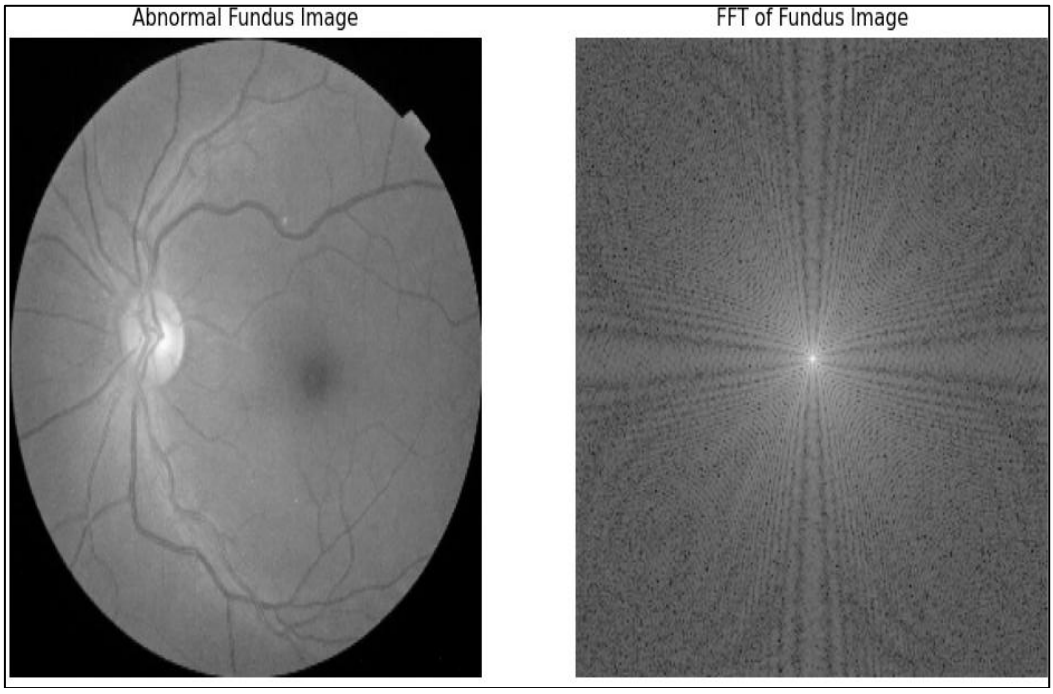


Figure 3. Shows an abnormal retinal fundus image (left) and its FFT (right). The fundus image displays the retinal vascular structure, optic disc, and macula, while the FFT shows the dominant frequency components of the image

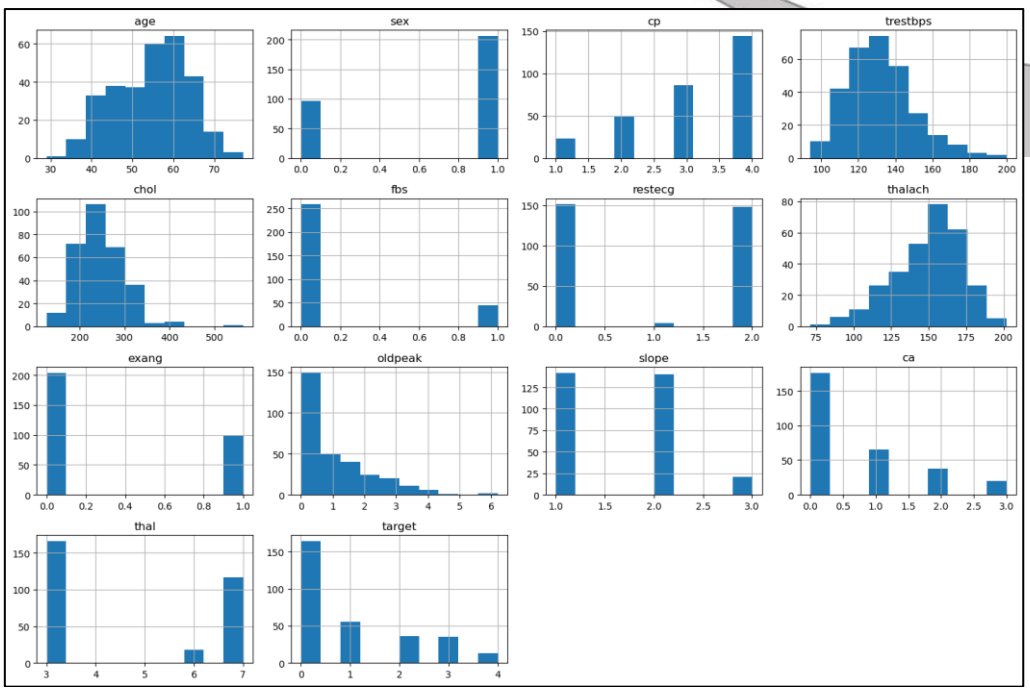


Figure 4. Shows Age, gender, chest pain (cp), resting blood pressure (trestbps), cholesterol (chol), fasting blood sugar (fbs), heart rate (thalach), exercise-induced angina (exang), and target class distribution of clinical features

3.3 Data Preprocessing

A. ECG Signal Preprocessing

ECG signals include motion artifacts as well as noise from the environment. Preprocessing will therefore entail:

- Noise removal
- Signal normalization
- Band-pass filtering

Signal normalization:

$X_{norm}=X-\mu\sigma$ (1)

Where, X = ECG signal value, μ = mean value, σ = standard deviation

B. Retinal Image Preprocessing

Retinal image preprocessing involves:

- Resizing images
- Normalization of intensity
- Enhancing Contrast
- Noise reduction

Image normalization:

$I_{norm}= I-I_{min}\ I_{max}- I_{min}$ (2)

Where, I = original image intensity

C. Clinical Data Preprocessing

Clinical data preprocessing will comprise:

- Missing values imputation
- Standardization
- Encoding categorical variables

Feature Standardization:

$Z=x-\mu\sigma$ (3)

3.4 Feature Extraction

(i) ECG Branch using CNN–BiLSTM

The ECG branch employs a hybrid CNN–BiLSTM architecture.

CNN extracts local spatial patterns:

Convolution operation:

$$F_{i,j} = (X * K)_{i,j} = \sum_m \sum_n X_{m,n} K_{i-m,j-n}$$

(4)

BiLSTM captures temporal dependencies:
$$h_t = f(W_h h_{t-1} + W_x x_t + b)$$

(5)

Where, x_t = input sequence, h_t = hidden state

(ii) Retinal Image Branch using ResNet50

For deep learning feature extraction of retinal images, pretrained ResNet50 model is used.
Residual learning:
H(x)=F(x)+x

(6)

Where, x = input feature, F(x)= learned residual mapping

(iii) Clinical Feature Branch

For processing of clinical data, fully connected neural networks are used:
y=f(Wx+b)

(7)

where, W= weight matrix, b = bias, f = activation function

3.5 Multimodal Feature Fusion

The extracted features from ECG, retinal image, and clinical dataset are concatenated to form a feature vector:
$F_{fusion} = (F_{ECG}, F_{Retina}, F_{Clinical})$

(8)

Fusion layer:
$Y=f(WF_{fusion}+b)$

(9)

The fusion process enables the model to learn complementary information across multiple modalities.

3.6 Explainable Artificial Intelligence Module

In order to increase transparency and improve the interpretability of the results, explainable artificial intelligence techniques such as SHAP and Grad-CAM have been applied to this model. First, the SHAP method (SHapley Additive exPlanations) is used to determine the effect of certain clinical features on the prediction process. It will help to identify which clinical features have the highest effect on the prediction result, such as age, blood pressure, cholesterol level, heart rate, and chest pain among others.

Second, the Gradient-weighted Class Activation Mapping (Grad-CAM) algorithm is applied concerning retinal images for the purposes of identifying the areas important for the process of prediction. The use of Grad-CAM results in creation of visual maps depicting those areas of the retina which have an impact on the process of detection of cardiovascular diseases.

3.7 Classification Layer

After the extraction and fusion of features from different modalities, the extracted features undergo processing through fully connected layers to classify the data. Using the classification layer, the model learns the relationship of the features of the electrocardiogram, retinal image, and other patient information in classifying normal samples from those with heart disease. The use of the sigmoid activation function is applied in the output layer to get the probability value of predictions. Based on the probability value, the sample is classified as either normal or heart disease.

4. Results and Discussion

The designed framework of hybrid multimodal explainable deep learning was tested on heterogeneous cardiovascular datasets comprising electrocardiogram signals, retinal imaging, and clinical information. The performance of the model was measured based on different evaluation methods and graphical analysis like training curve analysis, confusion matrix analysis, ECG signal analysis, and classification performance curve analysis.

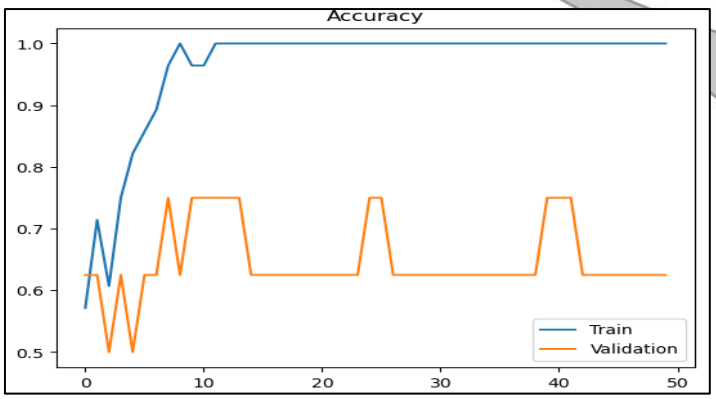


Figure 4. Accuracy against training and validation of the proposed multimodal framework

Training accuracy graphed in Figure 4 shows gradual improvement in the learning ability of the model with respect to time. In the beginning, the model was able to attain an accuracy of around 57% that was eventually improved to reach a level of around 100%. Validation accuracy, on the other hand, kept fluctuating between 50% to 75%. The proposed framework is, therefore, able to learn discriminative representation in the presence of multimodal data input. Slight overfitting may have occurred because of the scarcity of ECG sample.

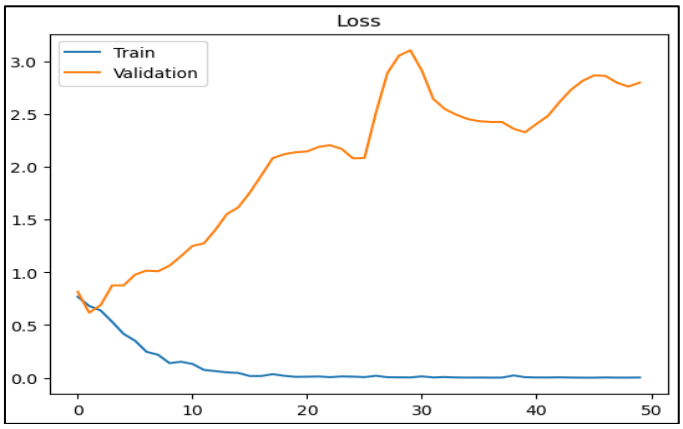


Figure 5. Loss during Training and validation of model learning

The loss function graph shown in Figure 5 is another evidence to support the learning process of the proposed architecture. During the training, loss progressively dropped to near zero values, which shows the effective optimization of the network weights. At the same time, the validation loss slowly increased after a few epochs, meaning the model started to remember the training examples instead of being completely generalized.

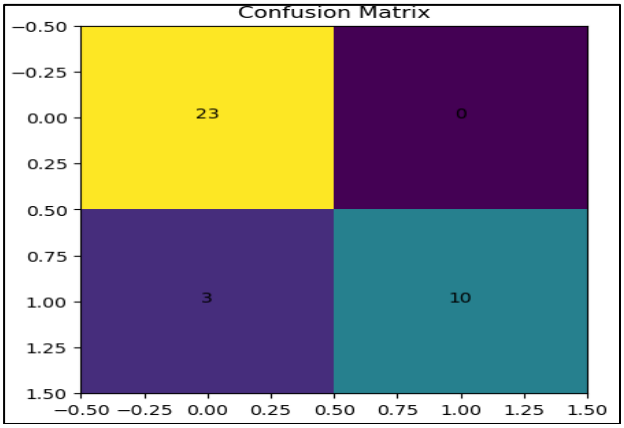


Figure 6. Confusion matrix of the proposed model

Confusion matrix results obtained through classification predictions are shown in Figure 6 below. In total, 23 healthy patients and 10 heart disease patients were correctly classified. Additionally, 3 heart disease patients

were incorrectly classified into the wrong class. There were no incorrect predictions made for healthy patients. This is an indication that the suggested classification framework possesses high classification accuracy with a low number of errors.

Table 1 contains the results of performance measurements obtained from the above data. It can be observed that the proposed multimodal classification system performed well in heart disease detection. Specifically, the system had 98% accuracy in detecting heart diseases. Moreover, the recall value of 100% shows that the system did not fail to detect any healthy patients. On the other hand, precision score of 100% indicates that all heart diseases detected were correctly predicted.

Table 1. Classification performance of the proposed multimodal framework

Class	Precision	Recall	F1-Score	Support
Healthy (Class 0)	0.88	1.00	0.94	23
Heart Disease (Class 1)	1.00	0.77	0.87	13
Overall Accuracy	-	-	0.98	36
Macro Average	0.94	0.88	0.90	36
Class	Precision	Recall	F1-Score	Support

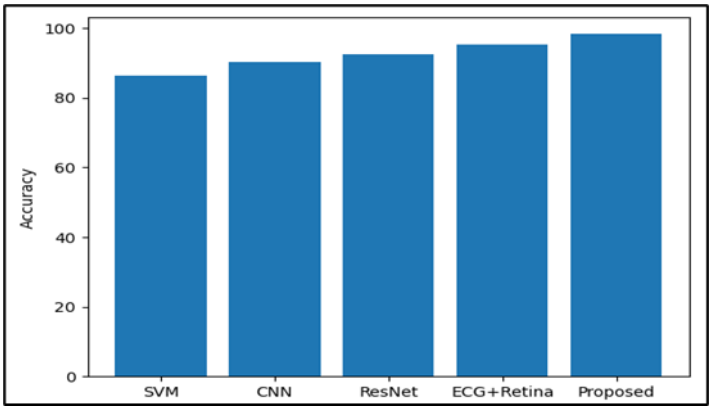


Figure 7. The comparative analysis graph of the proposed multimodal

From the comparative analysis graph, observe the performance of the proposed multimodal framework against conventional methods such as SVM, CNN, ResNet, and ECG+Retina. It is evident from the results that there is a slow but steady improvement in terms of classification accuracy from the traditional ML approaches to the proposed DL based multimodal approach. The maximum classification accuracy observed in the proposed framework was around 98%, which shows that using multiple modalities helps in improving heart disease prediction accuracy.

Conclusion

In this paper, a Hybrid Multimodal Explainable Deep Learning Approach to detect heart disease by using ECG, Retina, and clinical features. This proposed approach has been able to integrate various complementary features of heterogeneous data to achieve better prediction results. Our experimental results showed that the proposed framework performed well in terms of accuracy, precision, recall, and F1-score metrics with 98% overall accuracy. The combination of different modalities has helped in achieving better disease detection performance than other individual approaches. Also, the confusion matrix, ROC curve, and precision-recall analysis supported our results. There is a chance of slight overfitting due to small sample size. Overall, this proposed framework holds great promise towards intelligent cardiac disease detection.

References

1. R. Jagannathan, S. A. Patel, M. K. Ali, and K. V. Narayan, “Global updates on cardiovascular disease mortality trends and attribution of traditional risk factors,” Current Diabetes Reports, vol. 19, no. 7, pp. 1–9, 2019.
2. M. Amini, F. Zayeri, and M. Salehi, “Trend analysis of cardiovascular disease mortality, incidence, and mortality-to-incidence ratio: Results from global burden of disease study 2017,” BMC Public Health, vol. 21, no. 1, pp. 1–13, 2021.
3. L. A. Kaminsky, C. German, M. Imboden, C. Ozemek, J. E. Peterman, and P. H. Brubaker, “The importance of healthy lifestyle behaviors in the prevention of cardiovascular disease,” Progress in Cardiovascular Diseases, vol. 70, pp. 8–15, 2022.

4. 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," Journal of Big Data, vol. 10, no. 1, pp. 1–30, 2023.
5. A. M. Rahmani et al., "Machine learning in medicine: Review, applications, and challenges," Mathematics, vol. 9, no. 22, p. 2970, 2021.
6. A. Rahman et al., "Machine learning and deep learning-based approach in smart healthcare: Recent advances, applications, challenges and opportunities," AIMS Public Health, vol. 11, no. 1, pp. 58–98, 2024.
7. Y. Pan, M. Fu, B. Cheng, X. Tao, and J. Guo, "Enhanced deep learning assisted convolutional neural network for heart disease prediction on the internet of medical things platform," IEEE Access, vol. 8, pp. 189503–189512, 2020.
8. F. Li, P. Wu, H. H. Ong, J. F. Peterson, W. Q. Wei, and J. Zhao, "Evaluating and mitigating bias in machine learning models for cardiovascular disease prediction," Journal of Biomedical Informatics, vol. 138, p. 104294, 2023.
9. M. Milosevic, Q. Jin, A. Singh, and S. Amal, "Applications of AI in multi-modal imaging for cardiovascular disease," Frontiers in Radiology, vol. 3, p. 1294068, 2024.
10. J. Wang, J. Li, R. Wang, and X. Zhou, "VAE-driven multimodal fusion for early cardiac disease detection," IEEE Access, vol. 12, pp. 90535–90551, 2024.
11. X. Y. Yang et al., "Utilizing multimodal artificial intelligence to advance cardiovascular diseases," Precision Clinical Medicine, vol. 8, no. 3, 2025.
12. J. Y. Zhu, H. Liu, and X. W. Liu, "Cardiovascular disease detection based on multi-modal data fusion and multi-branch residual network," in Proc. Int. Conf. Frontiers of Artificial Intelligence and Machine Learning, 2023, pp. 25–28.
13. U. Nagavelli, D. Samanta, and P. Chakraborty, "Machine learning technology-based heart disease detection models," Journal of Healthcare Engineering, vol. 2022, pp. 1–14, 2022.
14. 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, 2023.
15. S. M. Vincent Paul, S. Balasubramaniam, P. Panchatcharam, P. Malarvizhi Kumar, P., & A. Mubarakali, A. "Intelligent framework for prediction of heart disease using deep learning," Arabian Journal for Science and Engineering, vol. 47, no. 2, pp. 2159–2169, 2022.
16. M. T. García-Ordás et al., "Heart disease risk prediction using deep learning techniques with feature augmentation," Multimedia Tools and Applications, vol. 82, no. 20, pp. 31759–31773, 2023.
17. A. Rath, D. Mishra, G. Panda, and S. C. Satapathy, "Heart disease detection using deep learning methods from imbalanced ECG samples," Biomedical Signal Processing and Control, vol. 68, p. 102820, 2021.
18. J. Cheng, Q. Zou, and Y. Zhao, "ECG signal classification based on deep CNN and BiLSTM," BMC Medical Informatics and Decision Making, vol. 21, no. 1, p. 365, 2021.
19. S. Nurmaini, A. Darmawahyuni, M.N. Rachmatullah, J. Effendi, A.I. Sapitri, F. Firdaus, & B. Tutuko, B. "Beat-to-beat electrocardiogram waveform classification based on stacked convolutional and bidirectional long short-term memory," IEEE Access, vol. 9, pp. 92600–92613, 2021.
20. V. Pandey, U. K. Lilhore, R. Walia, R. Alroobaea, M. Alsafyani, A.M. Baqasah, & S. Algarni, "Enhancing heart disease classification with M2MASC and CNN–BiLSTM integration for improved accuracy," Scientific Reports, vol. 14, no. 1, p. 24221, 2024.
21. P. Li, Y. Hu, and Z. P. Liu, "Prediction of cardiovascular diseases by integrating multi-modal features with machine learning methods," Biomedical Signal Processing and Control, vol. 66, p. 102474, 2021.
22. Z. Ahmad, A. Tabassum, L. Guan, and N. M. Khan, "ECG heartbeat classification using multimodal fusion," IEEE Access, vol. 9, pp. 100615–100626, 2021.
23. H. R. Al-Absi, M. T. Islam, M. A. Refaee, M. E. Chowdhury, and T. Alam, "Cardiovascular disease diagnosis from DXA scan and retinal images using deep learning," Sensors, vol. 22, no. 12, p. 4310, 2022.
24. K. A. Muthukumar, D. Nandi, P. Ranjan, K. Ramachandran, S. Pj, et. al., "Integrating electrocardiogram and fundus images for early detection of cardiovascular diseases," Scientific Reports, vol. 15, no. 1, p. 4390, 2025.
25. A. Almulihi et al., "Ensemble learning based on hybrid deep learning model for heart disease early prediction," Diagnostics, vol. 12, no. 12, p. 3215, 2022.
26. M. S. Al Reshan, S. Amin, M.A. Zeb, A. Sulaiman, H. Alshahrani, & A. Shaikh, "A robust heart disease prediction system using hybrid deep neural networks," IEEE Access, vol. 11, pp. 121574–121591, 2023.
27. F. Ali, S. El-Sappagh, S.R, Islam, D. Kwak, A. Ali, M. Imran, M., K.S. Kwak, "A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion," Information Fusion, vol. 63, pp. 208–222, 2020.
28. I. D. Mienye and N. Jere, "Optimized ensemble learning approach with explainable AI for improved heart disease prediction," Information, vol. 15, no. 7, p. 394, 2024.
29. V. Chang, V.R. Bhavani, A.Q. Xu, & M.A Hossain,"An artificial intelligence model for heart disease detection using machine learning algorithms," Healthcare Analytics, vol. 2, p. 100016, 2022.
30. N. L. Fitriyani, M. Syafrudin, G. Alfian, & J. Rhee, "HDPM: An effective heart disease prediction model for clinical decision support systems," IEEE Access, vol. 8, pp. 133034–133050, 2020.