International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Heart Disease Prediction using Multimodal Data with Multi-Layer Perceptron

N. Revathi [1], P.M. Kavitha *[2], D. Narayani [3], S. Irin Sherly [4], M. Robinson Joel [5], P. Jose [6]

Submitted:13/03/2024 **Revised**: 28/04/2024 **Accepted**: 05/05/2024

Abstract: Cardiovascular diseases (CVD) continue to be an increasing worldwide health issue, requiring specialized diagnostic instruments for early identification and care. The proposed technique seeks to improve cardiac disease prediction by using a multi-modal approach that combines patient's demographic data with raw ECG signals. This approach combines signal processing, feature extraction, and selection methods to enhance the predicted accuracy of the system. Thus, here we develop an application which can predict the vulnerability of a heart disease by giving details like age, gender, height, weight, etc., along with 12 lead ECG signal images provided in the PTB-XL dataset. This system method employs advanced Fast Fourier Transform (FFT) feature extraction technique, to extract informative features from ECG signals. Additionally, random forest classification algorithm is utilized for feature selection to identify the most discriminative attributes for prediction. The extracted features are then inputted into a Multi-Layer Perceptron (MLP) model, which is trained on a comprehensive dataset comprising patient demographics and ECG signals.

Keywords: *Electrocardiogram, Multilayer Perceptron, Random Forest, Multi-Modal, PTB dataset.*

1. Introduction

Heart diseases include a wide range of disorders that impact the heart and blood arteries, such as arrhythmias, heart failure, and coronary artery disease. These conditions often manifest with diverse symptoms, making early detection challenging. Therefore, it is crucial to monitor heart attack episodes using electrocardiogram [15]. Predicting the presence or absence of heart disease in an individual using a multimodal approach involves integrating data from many types of data sources. In this particular context, the term "multimodal" refers to the use of diverse forms of data, including demographic details, medical records, and even data derived from 12 lead ECG signal images [4]. ECG signals provide significant diagnostic information. Consequently, there has been a consistent and thorough research effort dedicated to creating proficient and successful methods for processing and interpreting ECG data, specifically targeting the identification of vital and original diagnostic information [1]. Every electrical activity that occurs inside the heart is recognised and recorded by the ECG equipment. This explains the function of the intracardiac conducting tissue in the heart and uses its electrical characteristics to detect the existence of heart disease.

ECG signals, which record the complex electrical patterns of the heart's activity over a period of time, provide a valuable wealth of data that may be used by advanced deep learning models. This work is highly significant in improving the understanding of heart prediction. The research aims to provide detailed insights into the connection between demographic data and ECG patterns by using a multi-modal approach and utilizing Multi-Layer Perceptron (MLP).

2. Related Works

Although the ECG is thought to be the most crucial tool for identifying and treating cardiac issues, interpreting ECGs is time-consuming and requires trained experts with specialized expertise. Furthermore, the number of devices capable of recording ECG data is expanding quickly.

Adyasha Rath et al. [1] have addressed the problem of heart disease classification by using ECG samples from healthy people and heart disease patients as input data. The imbalanced dataset along with ML, DL and ensemble model were utilized to assess the patient's heart disease detection performance. Finally, the analysis of various performance standards shows that the GAN-LSTM model outperforms the three performance measures considered in the research.

[1] *Department of Computer Science and Applications, SRM Institute of Science and Technology, Ramapuram, Tamil Nadu, 600089, India*
ORCID ID : 0000-0001-8216-2246
[2] *Department of Computer Science and Applications, SRM Institute of Science and Technology, Ramapuram, Tamil Nadu, 600089, India.*
ORCID ID : 0000-0003-2709-2032
[3] *Department of Information Technology, School of Computing Sciences, VISTAS, Pallavaram., India*
[4] *Department of CSE, SIMATS Engineering, Saveetha Institute of Medical and Technical Sciences, Thandalam, Tamil Nadu, 602105, India*
ORCID ID : 0000-0002-4518-3213
[3] *Department of CSE, KCG College of Technology, Karapakkam, Chennai, Tamil Nadu 600097, India*
ORCID ID : 0000-0002-3030-8431
[6] *Department of CSE, Vel Tech Rangarajan Dr. Sagunthala R &D Institute of Science and Technology, Avadi, Chennai, India*
ORCID ID : 0000-0003-2340-6786
** Corresponding Author Email: pmkavimca@gmail.com*

Kusuma S et al. [2] has used the publicly available datasets for ECG based cardiovascular disease prediction. The proposed model consists of three key phases: pre-processing, feature extraction, and classification. Xinwen Liu et al. have examined research articles that methodically apply deep learning techniques to ECG diagnosis. The study offered deep learning techniques for categorising ECGs based on the appropriate algorithms. Then they provided a deep analysis of the models based on Generative Adversarial Nets, CNN, RNN, CNN-LSTM hybrid networks, Fuzzy deep neural networks (FDNN), Probabilistic Neural Networks (PNN), etc [3].

Mariya R. Kiladze et al. have introduced a multimodal neural network [4] designed to enhance the precision of identifying cardiac arrhythmias using patient data alongside ECG signals. They utilize a linear perceptron to process patient information and rely on an LSTM network for classifying ECG signals. However, a drawback of their system lies in the need for simultaneous access to both patient metadata and ECG signals.

Mohammed B. Abubaker et al. [5] have proposed a deep learning algorithm to forecast four primary cardiac abnormalities: irregular heartbeat, myocardial infarction, prior myocardial infarction, and normalcy, utilizing a publicly available dataset of ECG images from cardiac patients. Adyasha Rath et al. Two publicly available arrhythmia datasets, PTB-ECG and MIT-BIH, were utilized to train and validate the proposed models [6]. Furthermore, an ensemble classification model was created by integrating two of the best performing AE and SOM models using the majority voting approach.

Hasnain Ali Poonja et al. [7] has a proposed approach to classify cardiac disorders (17 - classes) 45 individuals in the MIT-BIH Arrhythmia database. The SLR research found that current approaches face a variety of open issues when dealing with imbalanced data, limiting their practical usability and functionality [8]. Hossain et al. used machine learning algorithms [9] for heart disease prediction of 1190 records from the UCI repository.

A medical framework technique for identifying abnormalities in ECG data associated with stroke illnesses was proposed by Anand Kumar et al. [10]. It introduced a paradigm for predicting stroke-related problems using ECG data and other features, based on the LSTM network. This study also showed that the model has a low efficiency overhead and is suitable for the early detection of stroke-related illnesses.

Attention-Based Convolutional Neural Networks (ABCNN), a novel deep learning model presented by Liu et al. [11], operate directly on raw ECG data and automatically identify significant correlations for efficient arrhythmia identification. It does this by utilising CNN and multi-head attention. The main objective is to differentiate arrhythmia from normal heartbeats and accurately detect heart diseases among the five types of arrhythmia. A model based on the χ 2 statistical model and Deep Neural Networks was proposed by P. Ramprakash et al. [12]. Overfitting and underfitting are no longer problems. On both test and training sets of data, this model performs better.

A framework for the diagnosis and prognosis of heart illness is provided by Sajja et al. The five algorithms that are evaluated are C4.5, ID3, Random Forest, KNN, and SVM [13]. The UCI Cleveland database was used for the analysis. The SVM was found to be the most reliable process after data mining methodologies for these algorithms were assessed for consistency and usefulness. Using a variety of reference data, Syam Ahmad Jamil et al. used the trained ANN to identify cardiac anomalies [14]. For cardiac issues, the electrocardiogram (ECG) signal's amplitude and duration are utilised as input parameters. The artificial neural network (ANN) structure used in this work is a multilayer perceptron and the training strategies are Bayesian Regularisation , Lavenberg Marquardt and Backpropagation.

Shinde et al., [15] demonstrated how real- time cardiac streaming or audio recordings of patients can be analyzed using deep learning approaches to forecast the patient's state of health. In order to attain higher prediction accuracy than currently published results, a hybrid deep neural network model that combines the inherent properties of convolutional and recurrent neural network models was constructed. Two different datasets from separate data science competitions were combined to create the dataset that the model was trained on. To improve the size and diversity of this merged dataset, data augmentation techniques have been applied.

Irin Sherly et al., [16] amalgamated four databases and input 14 clinical features into an Ensemble model. Across all experiments, the accuracy rate consistently reached 99 percent for the four datasets, surpassing other machine learning techniques and relevant academic studies.

Irin Sherly et al., [17] examined the diverse sources of noise that impact the quality of ECG signals, elucidating their origins. Mitigating these noise sources poses a significant challenge. Various filtering techniques are applied to attenuate noise artifacts in ECG data. The effectiveness of these filters is assessed through Signal-to-Noise Ratio (SNR) calculations and subsequent comparative analysis.

Irin Sherly et al., [18] introduced an HBA-FRCNN for predicting Congestive Heart Failure (CHF) with enhanced diagnostic accuracy. Authors [19] proposed leveraging machine learning models to aid in understanding the virus's exponential dynamics on a daily basis and forecasting its future spread. The system developed by the authors [20] proposed a novel approach to heart detection leveraging the synergy of a Convolutional Neural Network (CNN) with the Aquila Optimization Algorithm (AOA) to create a Hybrid Deep Convolutional Neural Network (DCNN). This hybrid model is adept at analyzing ECG images, traditionally used for cardiovascular disease prediction. By integrating AOA, the DCNN optimizes weight parameters crucial for image analysis, thus enhancing prediction accuracy. Additionally, to enrich prediction performance, clinical data is incorporated alongside pre-processed ECG images. This innovative method capitalized on both visual information from ECG images and quantitative data from clinical records, culminating in a robust prediction framework.

Authors [21] analyzed the extracted features to evaluate the effectiveness of RF in capturing relevant information from dengue datasets. Authors proposed a system which involved the use of cameras to capture images of wildlife intrusions. When an animal is detected near human habitats, a GSM notification and alarm are sent to forest officials, alerting them to the situation. Additionally, our system incorporated atmospheric monitoring to address the limitations of existing solutions. This refined prototype model enabled continuous detection and monitoring of wildlife intrusion [22].

Researchers [23] proposed a system which utilized OpenCV and face recognition technology to deliver auditory feedback, aiding blind individuals in recognizing people and text in real-time. By optimizing memory usage and processing time, the system achieved competitive performance, offering portability and user-friendliness as notable advantages. The authors [24] explored various deep learning methods for predicting CVD. Among these, convolutional neural networks (CNNs) stood out as the most effective, achieving a remarkable 94.2% accuracy in both training and testing phases. Deep learning constructs algorithms that enable artificial neural networks to make intelligent decisions based on datasets. ECG serves as a valuable specification for detecting cardiovascular diseases.

Irin Sherly et al., [25] offered an innovative method for extracting features from electrocardiogram (ECG) signals to forecast the likelihood of heart disease. Their approach blended wavelet transform and principal component analysis (PCA) to derive distinct features from ECG data. These features are then inputted into a machine learning model for predicting heart disease.

Vinston Raja et al., [26] used datasets containing heart-related information from five different countries namely Cleveland, Hungary, Switzerland, Long Beach, and Statlog was utilized. The accuracy of the MLR and RF models is assessed, and the most effective model is implemented in healthcare settings for diagnosing cardiac conditions.

Yogalakshmi et al., [27] introduced a novel approach involving a deep learning boosting machine for gradients. Utilizing the Kaggle dataset, the proposed method outperformed existing techniques, boasting an accuracy exceeding 97%. Specifically, it achieved an accuracy rate of 98.91%, a recall rate of 97.76%, an F1 score of 97.25%, and a kappa score of 0.96.

T.Judgi et al., [28] introduced an innovative distributed database designed to securely store voter information. Each voter's data is stored in the database along with a secret key and an electronic signature. They proposed a state-of-the-art electronic voting system built upon blockchain technology to address the shortcomings identified in the prior analysis.

Irin Sherly et al., [29][30] introduced a client abandonment prediction model, leveraging the Decision Tree algorithm, Random Forest algorithm, and Deep Learning algorithm. Specifically, the Deep Learning approach employed a Multilayer Perceptron Neural Network to develop the prediction model. The effectiveness of these algorithms are evaluated and compared based on their sensitivity.

The system integrated cutting-edge robotics and aerial technology with advanced machine learning methods, specifically utilizing the VGG-SVM model. A drone equipped with a high-resolution camera captures real-time images of plant leaves, which are then processed by the VGG-SVM model to detect diseases. The VGG model facilitates deep feature extraction, while SVM ensures precise classification, enabling the system to effectively distinguish between healthy and diseased leaves [31].

3. Methodology

3.1. Dataset Description

PTB is a freely available dataset on PhysioNet for electrocardiography (ECG) analysis research, including arrhythmia identification, heart disease diagnosis, and ECG waveform analysis. 21,799 12-lead ECG recordings from 18,869 individuals. With a median age of 62 years and an interquartile range of 22, patients range in age from 0 to 95 years. Approximately 52% are male and 48% are female. The dataset covers a wide range of diseases and a fraction of healthy control samples. Normal ECGs, myocardial infarctions, ST/T alterations, conduction abnormalities and hypertrophy are diagnosed. PTB-XL dataset is Ideal for developing and evaluating ECG analysis algorithms and models in cardiology, biomedical engineering, machine learning, and artificial intelligence.

3.2. Proposed System

The proposed technique as in Fig 1 seeks to improve cardiac disease prediction by using a multi-modal approach that combines patient's demographic data with electrocardiogram (ECG) signals. The method aims to get a full picture of cardiovascular health by integrating demographic information such as age, gender, height, and weight variables with precise analysis of ECG signals as in Fig 2.

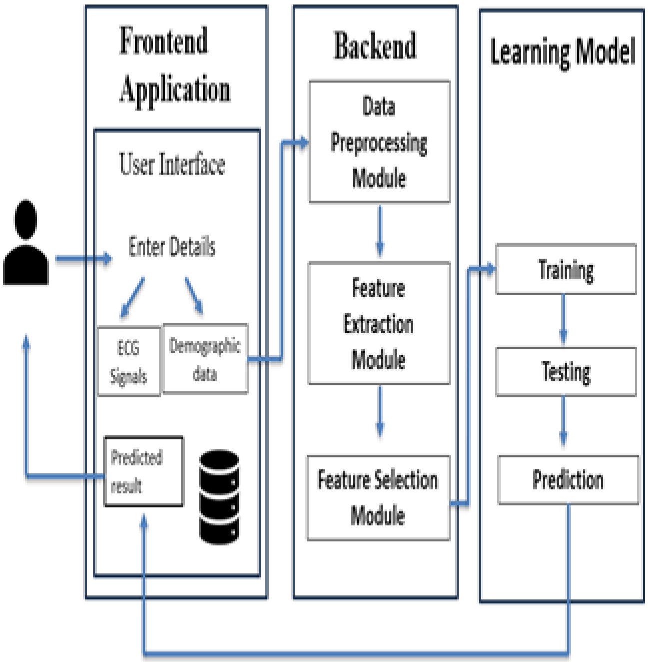

Fig.1. Architecture Diagram of prediction system

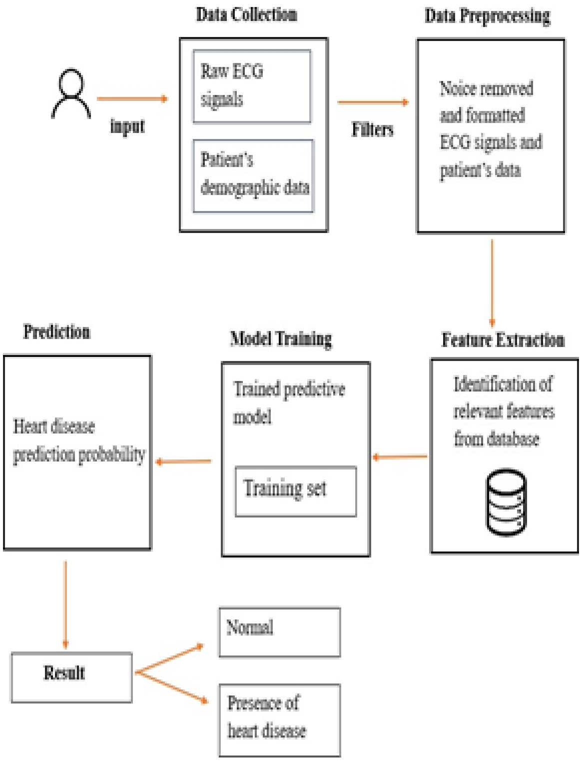

Fig.2. Workflow Diagram of the prediction approach

3.2.1. Preprocessing

At first, the dataset is obtained, which includes raw ECG signals and demographic information such as age, gender, height, weight, and ECG recordings from several leads. The ECG signals are subjected to thorough preprocessing, which includes the use of the Infinite Impulse Response (IIR) noise reduction approach as in Fig 4.

ecg_id	patient_ic	age	sex	height	weight	scp_codes
101	9086	0.528736	1	0.701613	0.326829	0
116	19695	0.54023	1	0.717742	0.336585	1
136	13445	0.574713	0	0.66129	0.326829	0
146	13447	0.494253	0	0.782258	0.414634	1
156	19547	0.54023	1	0.596774	0.317073	0
192	8340	0.367816	0	0.854839	0.390244	1
253	2096	0.632184	1	0.524194	0.312195	0
257	6078	0.896552	1	0.604839	0.239024	1
260	5471	0.528736	1	0.564516	0.195122	1
261	3810	0.586207	1	0.669355	0.336585	1
262	6501	0.678161	1	0.685484	0.404878	0
263	1595	0.528736	1	0.540323	0.243902	1
264	4441	0.735632	1	0.685484	0.321951	0
265	706	0.551724	1	0.645161	0.326829	0
266	641	0.908046	1	0.604839	0.341463	1
267	7662	0.678161	1	0.620968	0.434146	1
268	2533	0.632184	0	0.645161	0.326829	0
269	2282	0.977011	1	0.653226	0.317073	1
270	1480	0.551724	0	0.830645	0.429268	1

Fig.3. Pre-Processed Demographic Data

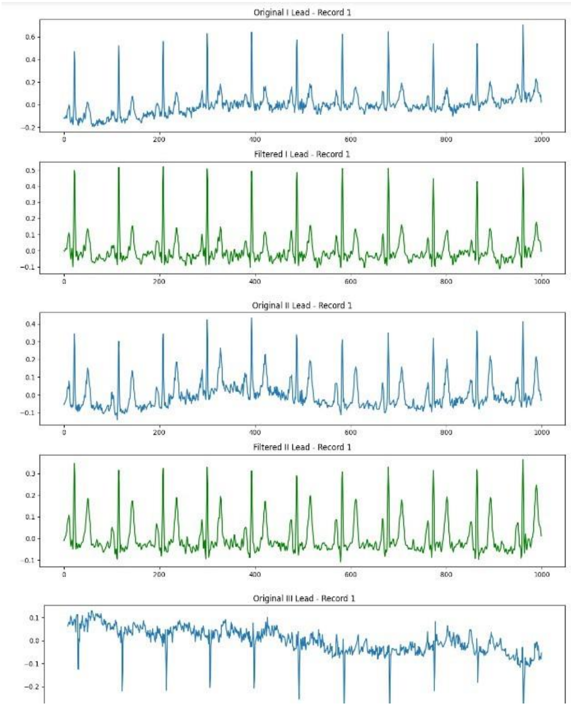

Fig.4. Signal Pre-Processing

The purpose of this process is to remove unwanted artifacts and improve the clarity of the signal. At the same time, demographic data is normalized using One hot encoder and min max normalization algorithm and scaled to guarantee consistency and reduce biases as in Fig 3. The preprocessing procedures lead to the creation of a refined dataset that is optimized for input into the MLP algorithm.

3.2.2. Feature extraction

The system utilizes the Fast Fourier Transform (FFT) approach for extracting important attributes from ECG. This process improves the model's ability to make accurate predictions. Signal processing applications like electrocardiography (ECG) analysis display signals as amplitude values over time. Complex signals may contain temporal data that is difficult to discern. The Fast Fourier Transform (FFT) mathematically converts a signal from time to frequency. It examines a signal and breaks it down into its frequencies to help comprehend its frequency components. Features extracted using Fast Fourier Transform (FFT) is fed into machine learning models for classification. The extracted features are mean_magnitude, mean_phase, std_magnitude, std_phase of all the leads along with patient's data such as age, gender, height, weight and scp_codes.

3.2.3. Feature selection

Our model development used the Random Forest method to improve heart disease prediction accuracy and dependability. A powerful ensemble learning technique called Random Forest combines many decision tree forecasts to get results that are more reliable and consistent as in Fig 5.

ead_0_fft	lead_9_fft	lead_5_fft	lead_1_fft	lead_1_fft	lead_4_fft	lead_4_fft	lead_8_ff
0.074703	0.131038	0.075799	0.156168	0.308475	0.097907	0.122471	0.060985
0.080375	0.111555	0.108711	0.101653	0.260153	0.082277	0.119739	0.079346
0.111736	0.096507	0.033201	0.168088	0.306602	0.13982	0.18297	0.060703
0.088151	0.276646	0.167922	0.148006	0.178254	0.164142	0.214692	0.1833
0.123273	0.085582	0.117323	0.191015	0.183104	0.128823	0.221033	0.064798
0.07509	0.120701	0.076215	0.137805	0.286563	0.100713	0.150334	0.065772
0.132399	0.109152	0.172301	0.196545	0.262682	0.099497	0.151887	0.078905
0.087601	0.127409	0.142784	0.123964	0.197506	0.064341	0.118739	0.076535
0.092784	0.18212	0.157916	0.150554	0.213236	0.093734	0.158129	0.101388
0.130243	0.218142	0.236651	0.19488	0.195482	0.09889	0.135559	0.150228
0.152095	0.106912	0.153167	0.226794	0.160022	0.146581	0.204103	0.073164
0.043398	0.146731	0.092316	0.099308	0.247973	0.066458	0.071649	0.067904
0.132534	0.094011	0.132691	0.182859	0.146156	0.12801	0.268188	0.060334
0.11119	0.122401	0.163903	0.170536	0.228517	0.085635	0.193892	0.076551
0.071654	0.281203	0.199727	0.119029	0.205414	0.115675	0.078259	0.157305
0.104887	0.151431	0.177516	0.171754	0.192454	0.082985	0.198042	0.102175
0.11044	0.142251	0.173944	0.163142	0.130306	0.079373	0.233521	0.104129
0.094362	0.24212	0.214979	0.141874	0.345639	0.085452	0.160856	0.180331
0.101893	0.129581	0.169363	0.149881	0.262323	0.061894	0.171261	0.092954
0.052635	0.068858	0.060961	0.095343	0.175454	0.062274	0.200574	0.045873
0.100131	0.129437	0.161583	0.15491	0.133906	0.068999	0.191839	0.08691
0.079039	0.130742	0.099377	0.167656	0.414267	0.100924	0.153099	0.067793
0.102322	0.263606	0.261988	0.157356	0.245535	0.068746	0.167034	0.17032
0.116348	0.070366	0.059155	0.160876	0.189077	0.135015	0.213123	0.047935
0.109788	0.150819	0.207106	0.204254	0.223244	0.076776	0.198195	0.08323
0.111271	0.138247	0.0714	0.1474	0.151821	0.158383	0.277161	0.106734

Fig.5. Selected Features Using Random Forest

Random Forest feature selection is used to identify the most important ECG signal dataset attributes. ECG signal characteristics included several heart function aspects. The information was fed to a Random Forest model that labelled each ECG signal occurrence with heart illnesses or anomalies. The Random Forest algorithm used ECG data input attributes to predict these classifications. Once trained, the Random Forest model produced feature significance scores for each input feature. Based on feature importance scores, the most significant qualities were selected. The top features with the highest importance ratings or qualities that exceed a relevance criterion is included in the selection criteria.

3.2.4. Multi-Layer Perceptron

In the proposed case the dataset model is partitioned into training and testing sets. 80% for training and 20% for testing. This stage is critical for constructing a reliable prediction model. The model is accessed for accuracy, precision, recall, F1 score, and specificity after training it with three hidden layers and activation functions. These measures showed how well the model classified cardiac illness and distinguished real positives from false positives. MLP can forecast and resist overfitting and noise, making it ideal for real-world applications with variable data quality.

4. Results and Discussion

PTB-XL dataset provides an extensive amount of electrocardiography (ECG) recordings from a varied patient group for machine learning algorithm training and validation. Over 21,000 ECG recordings from healthy people and patients with diverse heart diseases make up the PTB-XL dataset, which covers 12-lead ECG recordings.

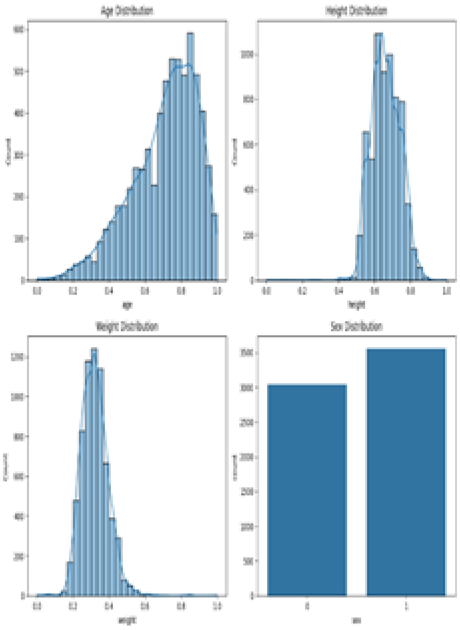

Fig.6. Distribution of PTB-XL Dataset Attributes

We used a subset of the PTB-XL dataset with clinical characteristics and ECG variables needed to predict cardiac disease. We carefully selected age, gender, blood pressure, cholesterol levels, and ECG waveforms to reflect the complexity of cardiovascular health and enhance our predictive model as in Fig 6.

The PTB-XL dataset's SCP codes column is essential for defining ECG recordings' cardiac states. ECG data show cardiac problems that these codes designate. Each code represents a heart anomaly or pathology, helping researchers and healthcare practitioners detect and study cardiovascular disease trends. In the proposed system if the spc_code is 1 then presence of heart disease otherwise normal heart. As shown in Fig 7, the system targets specific heart ailments or diseases by adding SCP codes to the dataset. This allows machine learning models and diagnostic algorithms to detect and categorize cardiac disease using SCP annotations.

Preprocessing the electrocardiogram (ECG) data entails applying filtering techniques and extracting relevant features in order to properly prepare the input for the multilayer perceptron (MLP) model. The MLP model then acquires the capability to categorize the data by utilizing its capacity to comprehend intricate patterns and connections within the data. The integration of these methods allows for precise and dependable identification of heart disease based on electrocardiogram (ECG) readings.

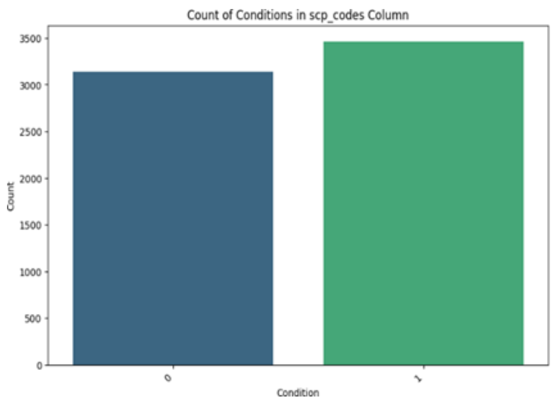

Fig.7. SCP_Codes

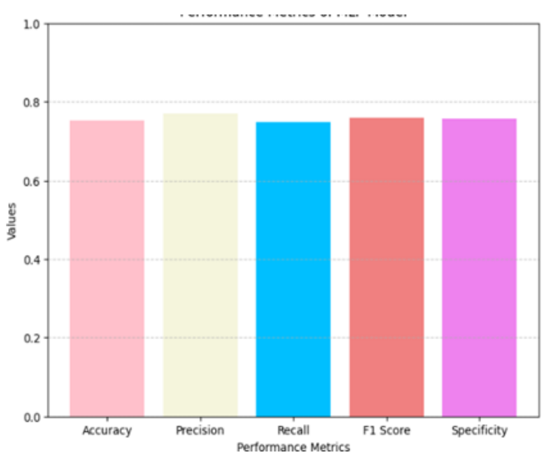

Fig.8. Performance Metrics of MLP Model

In proposed system the Multilayer Perceptron (MLP) model is trained on electrocardiography (ECG) dataset showed promise in heart disease classification. With 76% test accuracy, the model classified ECG samples well. Precision, is 79%. This indicates that the approach reduced false positive diagnoses procedures in Fig 8.

A recall (sensitivity) of 73% showed that the model could detect a considerable number of positive instances from the total positives. Medical diagnostics benefit from high recall to avoid missing essential patients. The F1 score, which balances accuracy and recall, was 76%, indicating a good balance between true and false positives.

The model's specificity was 79%, demonstrating a moderate ability to identify negative situations among all genuine negatives.

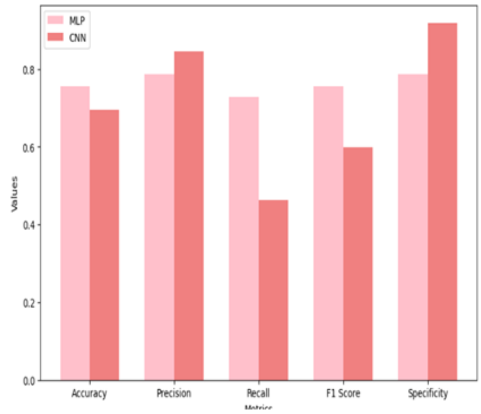

Fig.9. Performance Metrics of MLP Model

Using the same dataset as our model, we trained both the CNN and MLP models. The MLP model used a feedforward neural network to flatten and input dataset features into fully connected layers in Fig 9.

CNN used convolutional layers to extract ECG signals' sequential nature. Both models were trained and evaluated using the same dataset and a similar method for partitioning the data into training and testing sets. Both models were evaluated for accuracy, precision, recall, F1 score, and specificity.

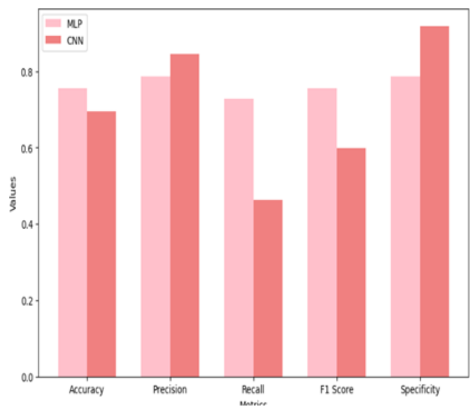

Fig.10. Comparison of Evaluation Metrics Between MLP and CNN

In Fig.11. the confusion matrix shows true negative (TN), false positive (FP), false negative (FN), and true positive (TP) predictions. 514 true negatives indicate valid forecasts of heart disease-free persons in this matrix. 154 false positives occur when the model predicts cardiac illness when it does not exist. 172 false negatives indicate that the model missed cardiac illness. Finally, 480 true positives indicate accurate heart disease forecasts.

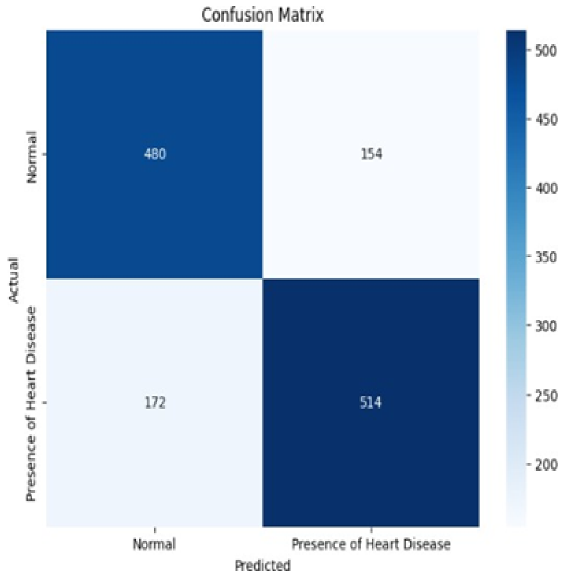

Fig.11. Confusion Matrix

Table.1. Confusion Matrix for Heart Disease Prediction

S.No	Prediction	Truth	n	Outcome
1	Normal	Normal	480	True Positive
2	Presence of Heart Disease	Normal	172	False Negative
3	Normal	Presence of Heart Disease	154	False Positive
4	Presence of Heart Disease	Presence of Heart Disease	514	True Negative

With a graphical user interface as in Fig 12, the model becomes more intuitive and user-friendly, making it easier for anyone to interact with it. This results in an improved user experience and seamless integration into current clinical procedures.Users may easily input patient data and electrocardiogram (ECG) readings using the graphical user interface (GUI), making the process of acquiring data and inputting it into the model simpler. This user-friendly interface removes the requirement for users to directly engage with the underlying code or algorithms, so allowing the model to be used by a wider range of people, including healthcare professionals with less programming knowledge.

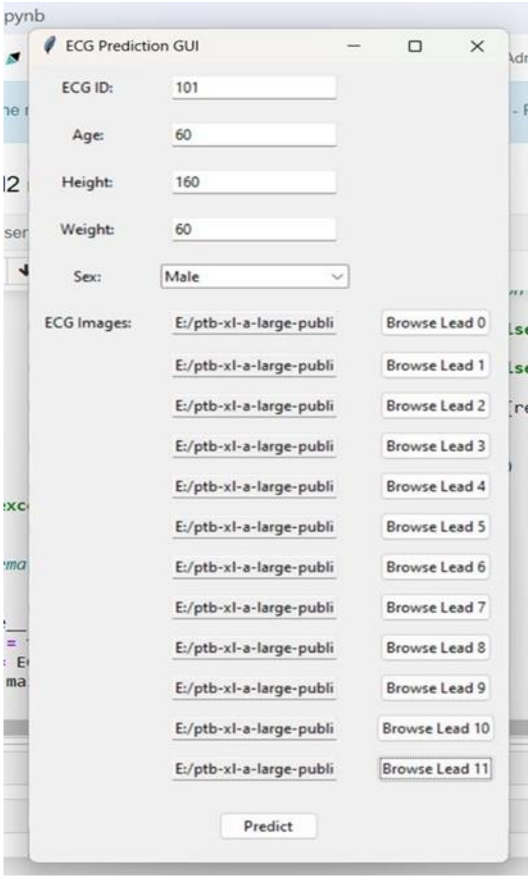

Fig.12. Graphical User Interface for the prediction system

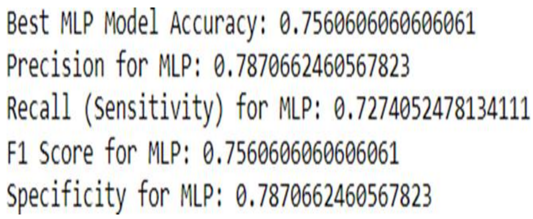

Fig.13. Performance metrics for MLP model

5. Conclusion

This study advances machine learning for heart disease prediction. We developed a Multilayer Perceptron (MLP) model with a user- friendly GUI to help healthcare professionals diagnose and treat cardiovascular diseases. Preprocessing the PTB-XL dataset included feature selection and data cleaning to assure data quality and relevance. The SCP codes column helped detect cardiac anomalies, guide feature selection, and improve model diagnostic accuracy. Using the pre-processed dataset, the MLP model was trained and evaluated with an emphasis on F1 score, accuracy, precision, and recall. The model showed good accuracy and balanced performance across many assessment criteria. A GUI transformed the user experience, giving healthcare practitioners an easy way to engage with the prediction model. The GUI enabled smooth data entry, real-time prediction, and performance measure display, enabling informed patient care decisions.

In the future, including additional datasets and characteristics improves the model's prediction power and allow it to detect more cardiovascular diseases along with improving the prediction model.

Author contributions

N. Revathi: Conceptualization, Methodology, Software, Field study, **P.M. Kavitha:** Data curation, Writing-Original draft preparation, Software, **D. Narayani:** Validation., Field study, **S. Irin Sherly:** Visualization, Investigation, **M. Robinson Joel:** Writing, **P. Jose:** Reviewing and Editing.

Conflicts of interest

The authors declare no conflicts of interest.

References

[1] Adyasha Rath, Debahuti Mishra, Ganapathi Panda and Suresh Chandra Sathyapathy, “Heart disease detection using deep learning methods from imbalanced ECG samples” Biomedical Signal Processing and Control Volume 68, July 2021.

[2] Xinwen Liu, Huan Wang, Zongjin Li, Lang Qin, Deep learning in ECG diagnosis: A review, Knowledge-Based Systems, Volume 227, 2021.

[3] Kusuma, Stalin & Udayan J, Divya. (2020). Analysis on Deep Learning methods for ECG based Cardiovascular Disease prediction. Scalable Computing: Practice and Experience. 21. 127-136.

[4] M. R. Kiladze, U. A. Lyakhova, P. A. Lyakhov, N. N. Nagornov and M. Vahabi, "Multimodal Neural Network for Recognition of Cardiac Arrhythmias Based on 12-Load Electrocardiogram Signals," in IEEE Access, vol. 11, pp. 133744-133754, 2023.

[5] M. B. Abubaker and B. Babayiğit, "Detection of Cardiovascular Diseases in ECG Images Using Machine Learning and Deep Learning Methods," in IEEE Transactions on Artificial Intelligence, vol. 4, no. 2, pp. 373-382, April 2023.

[6] Adyasha Rath, Debahuti Mishra, Ganapati Panda, Suresh Chandra Satapathy, Kaijian Xia, Improved heart disease detection from ECG signal using deep learning based ensemble model, Sustainable Computing: Informatics and Systems,Volume 35,2022.

[7] H. A. Poonja, M. Soleman Ali Shah, M. A. Shirazi and R. Uddin, "Evaluation of ECG based Recognition of Cardiac Abnormalities using Machine Learning and Deep Learning," 2021 International Conference on Robotics and Automation in Industry (ICRAI), Rawalpindi, Pakistan, 2021, pp. 1-4.

[8] Ehab Essa, Xianghua Xie, “An ensemble of deep learning-based multi-model for ECG heartbeats arrhythmia classification,” IEEE Access, vol 9, July 2021.

[9] A. I. Hossain, S. Sikder, A. Das and A. Dey, "Applying Machine Learning Classifiers on ECG Dataset for Predicting Heart Disease," 2021 International Conference on Automation, Control and Mechatronics for Industry 4.0 (ACMI), Rajshahi, Bangladesh, 2021, pp. 1-6,.

[10] M. Anand Kumar, N. Abirami, M. S. Guru Prasad and M. Mohankumar, "Stroke Disease Prediction based on ECG Signals using Deep Learning Techniques," 2022 International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES), Greater Noida,India, 2022, pp. 453-458

[11] Z. Liu and X. Zhang, "ECG-Based Heart Arrhythmia Diagnosis Through Attentional Convolutional Neural Networks," 2021 IEEE International Conference on Internet of Things and Intelligence Systems (IoTaIS), Bandung, Indonesia, 2021, pp. 156-162

[12] P. Ramprakash, R. Sarumathi, R. Mowriya and S. Nithyavishnupriya, "Heart Disease Prediction Using Deep Neural Network," 2020 International Conference on Inventive Computation Technologies (ICICT), Coimbatore, India, 2020, pp. 666-670.

[13] G. S. Sajja, M. Mustafa, K. Phasinam, K. Kaliyaperumal, R. J. M. Ventayen and T. Kassanuk, "Towards Application of Machine Learning in Classification and Prediction of Heart Disease," 2021 Second International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2021, pp. 1664-1669.

[14] A. H. -F. Syam Ahmad Jamil, J. A. Kadir, J. M. Jamil, F. R. Hashim, S. Shaharuddin and N. F. Makmor, "Multilayer Perceptron Optimization of ECG Peaks for Cardiac Abnormality Detection," 2022 IEEE 12th International Conference on Control System, Computing and Engineering (ICCSCE), Penang, Malaysia, 2022, pp. 37-41.

[15] Clement Virgeniya S., Ramaraj E.,A Novel Deep Learning based Gated Recurrent Unit with Extreme Learning Machine for Electrocardiogram (ECG) Signal Recognition, Biomedical Signal Processing and Control, Volume 68, 2021.

[16] Ya Li, Jing-hao Luo, Qing-yun Dai, Jason K.

Eshraghian, Bingo Wing-Kuen Ling, Ci-yan Zheng, Xiao-li Wang,"A deep learning approach to cardiovascular disease classification using empirical mode decomposition for ECG feature extraction", Biomedical Signal Processing and Control, Volume 79, Part 2,2023.

[17] Pengpai Li, Yongmei Hu, Zhi-Ping Liu, "Prediction of cardiovascular diseases by integrating multi-modal features with machine learning methods," Biomedical Signal Processing and Control, Volume 66, 2021.

[18] S. H. F. S. A. Jamil, J. A. Kadir, F. R. Hashim, B. Mustapha, N. S. Hasan and Y. Januar, "optimization of ECG Peaks for Cardiac Abnormality Detection using Multilayer Perceptron," 10th IEEE International Conference on Control System, Computing and Engineering (ICCSCE), 2020, pp. 169-173

[19] S. Shinde and J. C. Martinez-Ovando, "Heart Disease Detection with Deep Learning Using a Combination of Multiple Input Sources," 2021 IEEE Fifth Ecuador Technical Chapters Meeting (ETCM), Cuenca, Ecuador, 2021, pp. 1-3.

[20] S.Irin Sherly, G.Mathivanan, "An Ensemble based heart disease prediction using gradient boosting decision tree", Turkish Journal of Computer and Mathematics Education, Vol 12, No 10(2021), 3648-3660.

[21] S.Irin Sherly, G.Mathivanan, "ECG Signal Noises versus Filters for Signal Quality improvement", 2021 International Conference in Advances in Electrical, Computing, Communication and sustainable Technologies, 10.1109/ICAECT49130.2021.9392621.

[22] S. Irin Sherly, G. Mathivanan, "An efficient honey badger based Faster region CNN for chronic heart failure prediction", Biomedical Signal Processing and Control, 79, 2023.

[23] G.Revathy, Sophiya, M.P.Thiruvenkatasuresh, C.P Thamil, Selvi, ,Mrs S Irin Sherly, "Drugs For Corona Virus Prediction With Artificial Intelligence and Machine Learning", Nat. Volatiles & Essent. Oils, 2021; 8(5): 4583 – 4587.

[24] S. Irin Sherly, G. Mathivanan, "ECG based multi modal framework for heart disease detection", Journal of Theoretical and Applied Information Technology, 2023; 101(9): 3601 – 3614.

[25] Jackson Daniel, S. Irin Sherly, Veeralakshmi Ponnuramu, Devesh Pratap Singh, S.N. Netra, Wadi B. Alonazi, Khalid M.A. Almutairi, K.S.A. Priyan, and Yared Abera, "Recurrent Neural Networks for Feature Extraction from dengue Fever", Evidence-Based Complementary and Alternative Medicine, Volume 2022, Article ID 5669580, 2022.

[26] S.Irin Sherly, M.Haripushpam, B.Infantina Rinie, D.Jyotsna Devi, "Wildlife Intrusion and Early Fire Detection System Using IoT", International Journal of Innovative Research in Computer and Communication Engineering, Vol. 6, Issue 2, February 2018.

[27] S. Irin Sherly, P. M. Kavitha, S. Jayachandran, M. Robinson Joel, R. Jeena, "Empowering the Visually Impaired: Arduino Mega Enhanced Accessibility Device", International Journal on Recent and Innovation Trends in Computing and Communication, Vol. 11, Issue 8, August 2023.

[28] S. Irin Sherly, Prabu Sankar N, Ramaprabha Jayaram, C. Gnanaprakasam, Vinston Raja R, "Study of ECG Analysis based Cardiac Disease Prediction using Deep Learning Techniques", International Journal of Intelligent Systems and Applications in Engineering,Volume 11(4), 431–438, 2023.

[29] Deepak Kumar, N. Revathi, S. Irin Sherly, R. Lalitha, R. Vinston Raja, "Innovative Time Series-Based ECG Feature Extraction for Heart Disease Risk Assessment", Journal of Theoretical and Applied Information Technology, Vol.101. No 21, 2023.

[30] Irin Sherly S, Vinston Raja R, Deepak Kumar A, Prabu Sankar N, Chidambarathanu K, Thamarai I Krishnaraj M," Comparative Evaluation of Cardiovascular Disease using MLR And RF Algorithm with Semantic Equivalence", Journal of Theoretical and Applied Information Technology, Vol.101. No 18, 2023.

[31] Yogalakshmi V, Jenitha J. Merlin Mary, Chandravadhana A, S. Irin Sherly, Anitha G, "A Transparent Hybrid Deep Learning Framework to Assess German Measles Disease in Skin Images", IEEE Explorer, 2023.

[32] Srinivasa Rao, Kalaimagal Sivamuni, Chitra Devi D, Upendra Singh, S. Irin Sherly, Balaji, N, "An Architecture-based Self-Typing Service for Cloud Native Applications" IEEE Explorer, 2023.

[33] T.Judgi, R. Anitha, S.Padma, G.Nishanthi, S. Irin Sherly, "A Secured Framework Model to Design Electronic Voting System using BlockChain Methodology", in IEEE Explorer, 2024.

[34] Vinston Raja R, Deepak Kumar A, Prabu Sankar N, Gnanavel R, Krishnaraj M, Irin Sherly S, "Customer churn prediction in telecom industry using deep learning techniques", International Journal on Recent and Innovation Trends in Computing and Communication, 2023.

[35] M.S.Kavitha, R. Harikrishnan, S. Vijay Shankar, S. Krishnakumari, S. Irin Sherly, "A Smart Agriculture Friendly Intelligent Flying Robot Design to Identify Plant Leaf Diseases by using Digital Image Processing with Learning Approach", IEEE Explorer, 2023