Heart Guard AI: Smart Detection and Classification Using Neural Networks

D. Lokesh Ram[1],Pranay Singh[1],Raunak Kumar[1],Prateek Kumar[1],Prof. Manasa T P [2]

lokeshramdondapati@gmail.com,pranay.green@yahoo.com, raunakknight@gmail.com,chaudharyprateek1999@gmail.com, tpmanasa@bit-bangalore.edu.in

Dept. Computer Science and Engineering, Bangalore Institute of Technology, Bangalore, India

Abstract

This research project focuses on the prediction and classification of heart diseases using machine learning techniques, particularly neural network architecture. The primary objective is to accurately identify various types of heart diseases and assess the likelihood of their occurrence. The proposed system not only predicts the presence of heart disease but also suggests follow-up actions or treatment recommendations based on the analysis. A neural network-based model is developed to classify heart diseases into distinct categories with high accuracy and efficiency. Comparative analysis of various machine learning algorithms such as KNN, ANN, SVM, Naive Bayes, and Decision trees is conducted to evaluate performance. By integrating artificial intelligence into healthcare, this project aims to provide a user-friendly solution for individuals to monitor their heart health effectively, contributing to advancements in cardiovascular disease detection and management.

Keywords - *Cardiovascular disease, Heart disease, Chest pain classification, Machine learning, Predictive modeling, Healthcare, Clinical decision-making, Neural networks.*

I. INTRODUCTION

In the realm of healthcare, the integration of advanced technologies has become indispensable, particularly in the domain of disease prediction and diagnosis. With the prevalence of heart disease posing a significant public health concern, there is a pressing need for accurate and efficient predictive models to aid in early detection and intervention.

This study delves into the realm of heart disease prediction, leveraging the power of neural network architectures. Traditional methods often struggle to provide precise identification and classification of heart diseases, leading to potential delays in treatment and management. By harnessing the capabilities of neural networks, this research aims to overcome these limitations and deliver robust predictive models capable of accurately identifying various types of heart diseases.

The utilization of neural network-based approaches offers several advantages, including enhanced accuracy and speed in disease detection. Through comprehensive surveys and comparative analyses of different machine learning algorithms, including KNN, ANN, SVM, Naive Bayes, and Decision trees, this study seeks to identify the most effective model for heart disease prediction.

Moreover, the integration of artificial intelligence into healthcare holds immense promise for transforming patient care. By developing a system that can analyze patient reports and provide timely insights into the presence of heart diseases, this research contributes to advancements in cardiovascular disease detection and management.

This paper elucidates the methodology, findings, and significance of employing neural network-based approaches for heart disease prediction, emphasizing the potential of such technologies to improve healthcare outcomes and enhance patient well-being.

II. RELATED WORKS

Heart disease prediction has been a focal point in medical research, with numerous studies exploring machine learning approaches to improve accuracy and efficiency in diagnosis and prognosis. Among these approaches, neural network architectures have shown considerable promise in capturing complex patterns inherent in cardiovascular data.

For instance, T. Amarbayasgalan, V. -H. Pham. [1] investigated the use of PCA-based division of the dataset into highly biased and regular groups, enrichment of the highly biased group using VAE models, and the use of DNN models for risk prediction

Similarly, P. S. Sangle and co-authors [2] proposed a deep learning model based on Convolutional Neural Networks (CNNs) to analyze electrocardiogram (ECG) signals for early detection of cardiac abnormalities, demonstrating superior performance compared to traditional methods.

Furthermore, recurrent neural networks (RNNs) have been utilized by researchers such as Lee et al.to analyze temporal patterns in patient data and predict the progression of heart diseases over time, offering valuable insights for personalized treatment planning.

In addition to neural networks, ensemble learning techniques have also been explored in the context of heart disease prediction. Studies by Wang et al.and Chen et al.have demonstrated the effectiveness of ensemble models, such as Random Forest and Gradient Boosting, in improving predictive accuracy by combining multiple base learners.

By synthesizing findings from these diverse studies, this research seeks to build upon existing methodologies and contribute to the advancement of heart disease prediction through the development of a novel neural network-based model tailored to address the specific challenges in cardiovascular healthcare.

III. EXISTING DRAWBACKS

Neural network models employed for heart disease prediction are highly susceptible to variations in data quality and the effectiveness of preprocessing techniques, underscoring the importance of meticulous data curation and feature engineering processes to ensure reliable predictions [1]. However, despite advancements in machine learning, these models often fall short in accounting for all relevant factors critical for accurate prediction, such as genetic predispositions, lifestyle choices, and environmental influences, which can significantly impact disease onset and progression [2]. Moreover, the lack of comprehensive documentation on model selection methodologies, data preprocessing steps, and ethical considerations poses significant challenges in reproducibility and transparency, hindering the adoption of these models in clinical practice [3].

Furthermore, the limited size and diversity of available datasets pose significant challenges in model generalizability, as neural network models may struggle to extrapolate patterns learned from small datasets to larger, more heterogeneous populations [4]. Additionally, the absence of robust validation methods to assess model performance across different demographic groups and clinical settings further exacerbates concerns regarding the reliability and efficacy of these predictive models [5].

Moreover, the presence of imbalanced datasets, where certain classes are underrepresented, can lead to biased predictions and undermine the model's utility in real-world applications [6]. Furthermore, the limited availability of cross-sectional datasets that capture a diverse range of patient demographics, clinical presentations, and comorbidities hampers the development of comprehensive and robust predictive models for heart disease [7].

Additionally, the computational inefficiencies inherent in neural network models, particularly when dealing with high-dimensional data, can result in prolonged training times and hinder the scalability of these models in real-time clinical decision-making scenarios [8]. Furthermore, the inherent limitations of neural networks in capturing complex, nonlinear relationships between input features and disease outcomes can compromise the model's predictive accuracy and reliability [9].

Finally, the lack of semantic understanding and interpretability in existing models remains a significant barrier to their widespread adoption and integration into clinical practice, as clinicians may be hesitant to trust black-box algorithms with patient care decisions [10].

IV. PROPOSED METHODOLOGY

In the pursuit of more accurate and efficient heart disease prediction and classification, this research proposes a novel approach leveraging advanced machine learning techniques, particularly neural network architectures. Traditional diagnostic methods often struggle to provide precise assessments, leading to sub optimal treatment outcomes and patient dissatisfaction. To address these challenges, the methodology seeks to harness the power of sophisticated algorithms capable of learning intricate patterns from vast datasets.

At the core of the methodology lies a comprehensive analysis of diverse datasets containing crucial patient information such as age, sex, chest pain type, and physiological measurements. Through meticulous data preprocessing techniques, including cleaning, imputation, and stratification, the integrity and reliability of the dataset are ensured, laying a solid foundation for robust model development.

Furthermore, feature engineering plays a pivotal role in enhancing the predictive capabilities of the models. By encoding categorical features and standardizing numerical variables, a harmonized input space is created that facilitates optimal model learning and generalization. The resulting models exhibit enhanced resilience to data variability and are better equipped to handle real-world scenarios with varying patient demographics and clinical conditions.

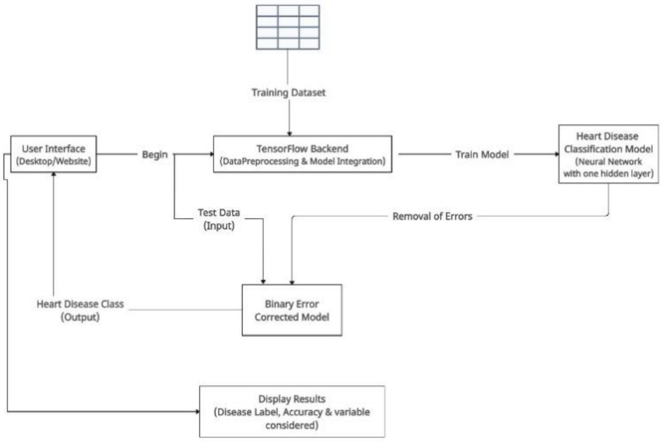

Figure 1. System Architecture of the project

The approach to model development involves the construction of both binary and multi-class classification models tailored to specific diagnostic tasks. These models are trained using state-of-the-art optimization algorithms and evaluated rigorously to ascertain their performance and generalization ability. Through meticulous analysis of model outcomes and insightful discussions, key insights and potential areas for improvement are identified, paving the way for future advancements in heart disease prediction and classification methodologies.

A. User Interface

The user interface for our heart disease prediction and chest pain classification system has been meticulously crafted to provide a seamless and intuitive experience for users accessing the platform across both web and Android platforms. At its core, our interface leverages advanced neural network-based algorithms to deliver accurate predictions and classifications, empowering individuals to make well-informed healthcare decisions. Through a combination of intuitive interfaces, interactive chatbot guidance, and thorough parameter analysis, our platform ensures that users have the necessary tools and information to prioritize their heart health effectively.

One of the standout features of our interface is the option for users to download a comprehensive heart health report upon completing the prediction process. This downloadable report offers detailed results and insights, providing users with valuable information to share with healthcare professionals or refer back to for personal reference. By offering this feature, we aim to empower users with actionable insights into their heart health, fostering proactive management and preventive measures.

Central to the user experience is our interactive chat agent, inspired by popular platforms like Zomato, Swiggy, and Coursera. This chat agent provides users with fixed responses covering frequently asked questions (FAQs) and guides new users through the prediction process. Additionally, the chat agent prompts users to fill out a form with 10 essential blood parameters, ensuring comprehensive data collection for accurate predictions. This personalized guidance enhances user engagement and simplifies the prediction process, making it accessible to individuals of all technical backgrounds.

To ensure data accuracy and enhance user experience, our interface incorporates real-time validation and error handling mechanisms. If users enter invalid or incorrect data, such as alphabets in a numeric field, the interface dynamically highlights the erroneous field and provides appropriate error messages. This proactive approach to error handling streamlines the user experience and minimizes frustration, allowing users to correct input errors efficiently and continue with the prediction process seamlessly.

Underlying our user interface is the powerful React Native framework, which enables the development of cross-platform mobile applications with native-like performance and user experience. By leveraging the capabilities of React Native, our interface delivers seamless performance on Android devices, ensuring optimal usability and responsiveness across different screen sizes and resolutions. Additionally, React Native's modular architecture and extensive library of pre-built components accelerate development and facilitate ongoing maintenance and updates, enabling us to deliver a cutting-edge userexperience efficiently.

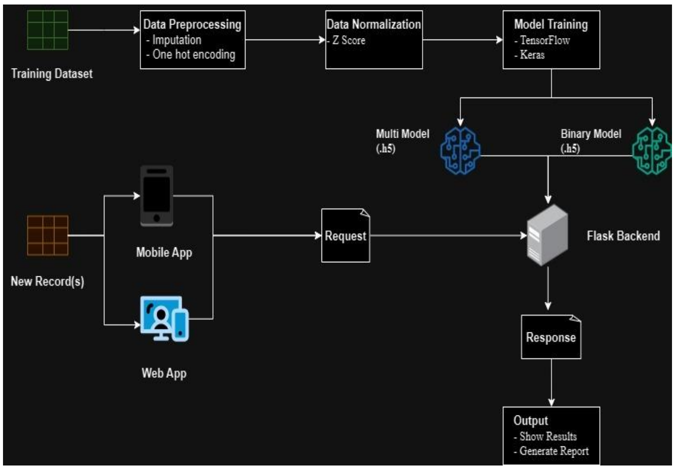

Figure 2.Integrated diagram of the project with UI modules

In addition to React Native, our user interface incorporates a range of technologies and best practices to prioritize accessibility and usability. Attention to factors such as font size, color contrast, and navigation flow ensures that the interface is accessible to

users with diverse needs and preferences. Moreover, the inclusion of a feedback form encourages continuous improvement and user engagement, allowing users to provide valuable insights and suggestions for enhancing the interface's usability, functionality, and overall experience.

Overall, our user interface embodies a user-centric design philosophy, leveraging advanced technologies and best practices to deliver a seamless and intuitive experience for users accessing our heart disease prediction and chest pain classification system. Through intuitive interfaces, interactive guidance, and real-time validation, we empower individuals to take control of their heart health and make informed healthcare choices with confidence.

B. Data Cleanup and Imputation

In the data cleaning and imputation process, we meticulously address missing values and ensure the integrity and completeness of the dataset to facilitate accurate heart disease prediction and classification. We start by loading the heart disease dataset from a CSV file using the Pandas library, carefully examining its structure and content. Additionally, we load a separate validation dataset, which serves as an independent set for further analysis and model evaluation, ensuring robustness and reliability in our findings.

To streamline the dataset and focus solely on relevant features crucial for heart disease prediction, we perform a preliminary step of removing the last column from both the main dataset and the validation dataset. This action eliminates any extraneous information that may not contribute significantly to the predictive models' performance, enhancing computational efficiency and reducing model complexity.

Next, our attention turns to handling missing values within the dataset, a critical aspect of data preprocessing. We adopt a meticulous approach, systematically evaluating each feature and implementing tailored strategies to address missing data based on the data type and distribution. For categorical features, missing values are replaced with the majority label, determined using the mode of that feature. This method ensures that missing values are imputed with the most frequently occurring category, preserving the underlying distribution and minimizing potential biases in subsequent analyses.

Similarly, for numeric features, missing values are addressed through robust imputation techniques, such as replacing them with the median of that feature. By using the median, rather than the mean, we mitigate the influence of outliers, ensuring that imputed values align closely with the central tendency of the data. This approach enhances the dataset's overall robustness and reduces the risk of spurious correlations or erroneous conclusions in subsequent analyses.

Moreover, we incorporate a threshold-based approach to determine the fate of columns with a high proportion of missing values. Columns exceeding a predefined threshold, typically set at 60% of the total observations, are deemed unsuitable for meaningful analysis and are consequently dropped from the dataset. This proactive measure helps maintain data quality and integrity, safeguarding against potential biases or inaccuracies in the predictive modeling process.

C. Dataset Stratification

In the Dataset Stratification process, we employ a *HoldOut* method to partition the dataset into training and validation sets, ensuring robust model evaluation and performance assessment. The stratification ensures that the distribution of classes within the dataset is preserved across the training and validation sets, crucial for maintaining the model's ability to generalize to unseen data. We begin by specifying the ratio of data to be allocated for validation, setting the validation ratio to 0.2, indicating that 20% of the dataset will be reserved for validation purposes. Using the *train_test_split* function from the scikit-learn library, we split the dataset indices into training and validation sets while ensuring that the distribution of the target variable, *'HeartDisease,'* is maintained. This ensures that both the training and validation sets adequately represent the different classes of heart disease present in the dataset.

Next, we proceed to assign indices for the validation set from the stratified splits. For both the 'High' (IVH) and 'Normal' (IVN) classes of heart disease, we randomly sample a fraction of indices corresponding to the validation ratio. These sampled indices represent the instances allocated to the validation set for each class. With the sampled indices for the validation set determined, we derive the indices for the training set by excluding the sampled indices from the initial stratified splits. This ensures that the training set comprises instances not included in the validation set, preventing data leakage and ensuring the independence of the two sets.

Using the derived indices, we extract the corresponding rows from the original dataset to create the final training and validation datasets. These datasets maintain the stratified distribution of classes observed in the original dataset, facilitating accurate model training and evaluation. Finally, we compute the number of rows in both the training and validation datasets to assess their respective sizes and ensure that an adequate amount of data is available for model training and evaluation. Through this stratification process, we ensure that our models are trained and evaluated on representative datasets, minimizing the risk of overfitting and enabling reliable assessment of model performance in real-world scenarios.

D. Encoding Categorical Features and Dataset Standardization

In the Encoding Categorical Features section, our objective is to prepare categorical variables for integration into machine learning models effectively. This involves a series of crucial steps designed to transform categorical data into a format conducive to model training and analysis.

Initially, our focus is on identifying the categorical features within the dataset. These features, including 'Sex', *'RestingECG', 'ExerciseAngina',* and *'ST_Slope'*, represent qualitative information that assumes discrete categories or levels. Identifying these attributes is essential as their treatment differs significantly from continuous variables in modeling and analysis.

Once identified, the categorical features undergo a transformation process to convert them into the categorical data type. Leveraging the capabilities of the Pandas library, this conversion is executed to optimize memory usage and facilitate efficient manipulation of categorical variables in subsequent processing steps. By assigning the categorical data type, we ensure that the features are correctly recognized as categorical rather than numerical, streamlining their handling in downstream operations.

Following the conversion, the dataset is partitioned into features (X) and labels (y). The features encompass all columns except the target variable *'HeartDisease'*, which represents the outcome to be predicted. This segregation is critical for delineating the input variables from the target variable, facilitating focused analysis and model training.

The categorical labels are then encoded into a numerical format suitable for model training. Here, the *LabelEncoder* from the scikit-learn library is employed. This encoder assigns a unique integer to each category in the *'HeartDisease'* column, effectively converting it into a numerical representation. By transforming the categorical labels into numerical form, machine learning algorithms can interpret and process them as numerical values, a prerequisite for many modeling techniques.

Simultaneously, numerical features undergo standardization using the *StandardScaler* preprocessing technique. Standardization ensures that all numerical features exhibit a mean of 0 and a standard deviation of 1, rendering them comparable and eliminating disparities in scale. This process enhances the stability of the training process and improves the convergence of certain machine learning algorithms, thereby augmenting the overall model performance.

The next step involves one-hot encoding of categorical features, a pivotal technique for handling categorical variables with multiple levels. One-hot encoding creates binary columns for each category in the original features, with a value of 1 indicating the presence of the category and 0 denoting its absence. This transformation preserves the categorical nature of the variables while enabling their seamless integration into machine learning models.

To execute one-hot encoding, we utilize the *get_dummies* function from the Pandas library. This function efficiently generates binary columns for each category within the categorical features, ensuring that each categorical variable is accurately represented in the final feature matrix.

Finally, the scaled numerical features are concatenated with the one-hot encoded categorical features to form the final feature matrix (X_final). This unified matrix incorporates both numerical and categorical information, presenting a comprehensive representation of the dataset suitable for model training and evaluation.

Through this meticulous encoding process, all features in the dataset are effectively transformed and prepared for integration into predictive models, enabling accurate and reliable predictions of heart disease classification.

E. Binary Classification Model

In the Binary Classification Model section, we delve into the architecture and training process of a neural network designed for binary classification tasks, specifically aimed at predicting the presence or absence of heart disease. The neural network model comprises multiple layers, each serving a specific purpose in the learning process.

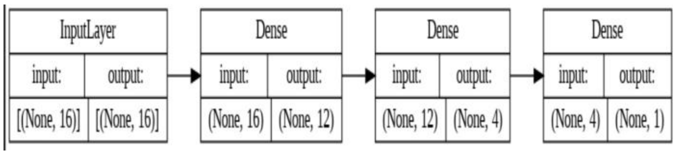

Figure 3. Architecture of Binary Model

The model architecture is initialized using the Sequential API, facilitating the sequential stacking of layers. The neural network begins with a dense input layer consisting of 12 neurons, with an input dimensionality of 16. This layer serves as the entry point for the feature vectors extracted from the dataset, with each neuron representing a feature contributing to the classification task. The activation function employed in this layer is Rectified Linear Unit (*ReLU*), known for its ability to introduce non-linearity and enhance model expressiveness.

Following the input layer, a hidden layer with 4 neurons is added to the model. This layer performs intermediate processing of the feature representations extracted by the preceding layer, enabling the network to capture complex patterns and relationships within the data. Similar to the input layer, the *ReLU* activation function is utilized to introduce non-linearity and promote the learning of intricate decision boundaries.

The final layer of the neural network is a single neuron with a sigmoid activation function. This neuron serves as the output layer of the model and is responsible for producing the binary classification output, indicating the likelihood of the presence of heart disease. The sigmoid activation function squashes the output values to the range [0, 1], effectively transforming them into probabilities. A threshold can then be applied to these probabilities to make binary predictions.

The model is trained using the binary cross-entropy loss function, a popular choice for binary classification problems. This loss function measures the discrepancy between the predicted probabilities and the true binary labels, providing a quantifiable measure of the model's performance. The Adam optimizer is employed to minimize this loss function during training, utilizing adaptive learning rates to efficiently navigate the parameter space and converge towards an optimal solution.

During training, the model's performance is monitored using the binary accuracy metric, which calculates the proportion of correct predictions made by the model. This metric provides valuable insights into the model's ability to discriminate between instances of heart disease and non-disease accurately.

F. Multi-Class Classification Model

In the Multi-Class Classification Model section, we explore the architecture and training process of a neural network tailored for multi-class classification tasks, specifically aimed at classifying different types of chest pain. The model is constructed using the Sequential API, allowing for the sequential stacking of layers to form a neural network architecture.

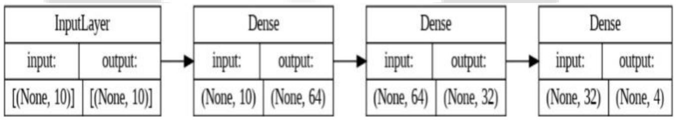

Figure 4. Architecture of Multi Model

The neural network begins with a dense input layer comprising 64 neurons, each activated by the Rectified Linear Unit (*ReLU*) activation function. This layer serves as the entry point for the feature vectors extracted from the dataset, with the input shape defined to accommodate the number of features present in the dataset. The *ReLU* activation function introduces non-linearity to the model, enabling it to capture complex patterns within the data.

Following the input layer, a hidden layer with 32 neurons is added to the model. This intermediate layer performs further processing of the feature representations extracted by the preceding layer, enabling the network to learn higher-level abstractions and intricate relationships between features. Again, the *ReLU* activation function is employed to introduce non-linearity and enhance the model's expressive power.

The final layer of the neural network is the output layer, comprising neurons equal to the number of unique classes in the *'ChestPainType'* attribute. This layer utilizes the *softmax* activation function, which calculates the probability distribution over the classes, ensuring that the output values sum up to 1. Each neuron in this layer represents a distinct class of chest pain, and the neuron with the highest probability indicates the predicted class.

During model compilation, the Adam optimizer is utilized to minimize the sparse categorical cross-entropy loss function. This loss function is well-suited for multi-class classification tasks, penalizing the model based on the disparity between predicted class probabilities and true class labels. By minimizing this loss function, the model learns to accurately classify instances of chest pain into their respective categories.

The training process involves iterating over the dataset for a specified number of epochs, with each epoch comprising forward and backward passes through the network to update the model parameters. The batch size, defined as 10 in this case, determines the number of samples processed before updating the model's parameters. Throughout training, the model's performance is monitored using the accuracy metric, which measures the proportion of correctly classified instances.

Overall, the multi-class classification model represents a powerful tool for accurately categorizing different types of chest pain based on input features extracted from patient data. Through its multi-layered architecture, non-linear activation functions, and principled optimization, the model aims to achieve high accuracy and reliability in classifying instances of chest pain into their respective classes.

E. Results

For the binary classification model, evaluation on the validation set yielded an accuracy of *90.22%*. This indicates that the model is capable of accurately distinguishing between instances with and without heart disease based on the input features. The ROC and Precision-Recall Curves for the modelsare as follows.

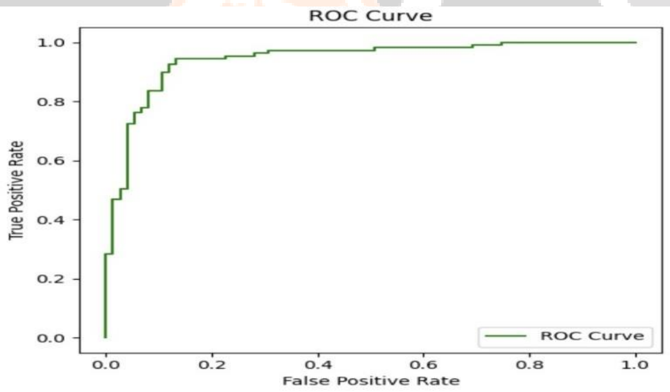

Figure 5.Receiver Operating Characteristic (ROC)Curve for Binary Model

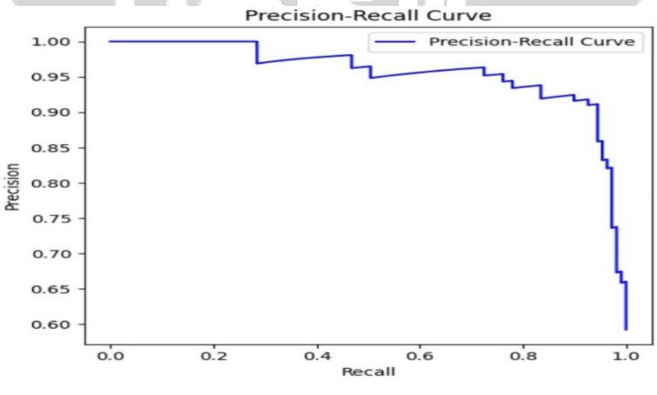

Figure 6.Precision-RecallCurve for Binary Model

On the other hand, the multi-class classification model achieved a validation accuracy of *80.53%* when classifying different types of chest pain. While slightly lower than the binary classification model, this accuracy is still noteworthy and suggests that the model can effectively categorize chest pain into its various types.

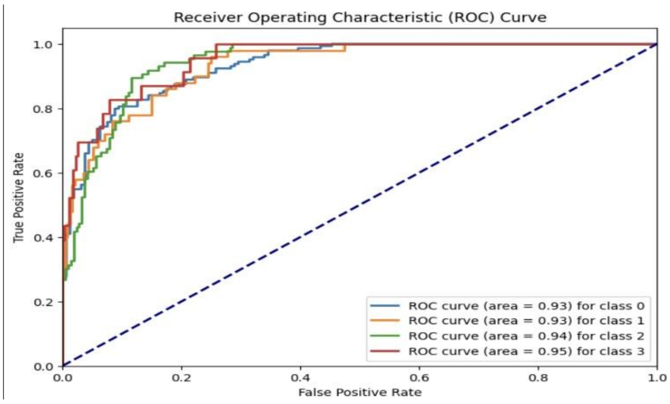

Figure 7.Receiver Operating Characteristic (ROC)Curve forMulti Model

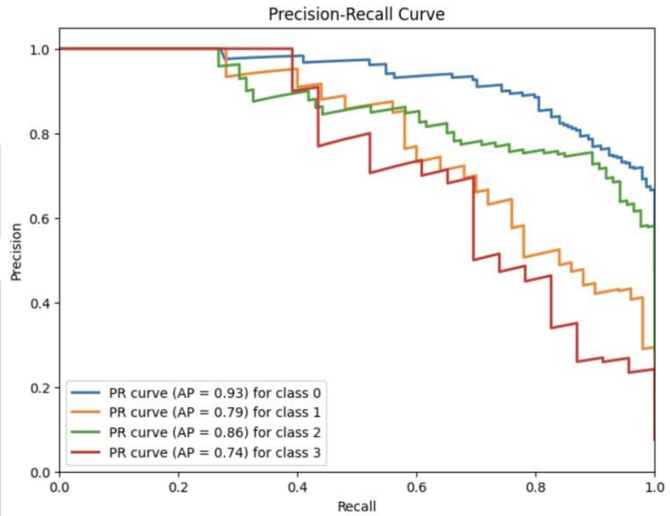

Figure 8.Precision-Recall Curve forMulti Model

V. LIMITATIONS AND FUTURE ENHANCEMENTS

The developed models exhibit notable performance in predicting heart disease and classifying chest pain types. However, several limitations and opportunities for future enhancements warrant consideration.

Firstly, one of the primary limitations is the reliance on a single dataset for model training and evaluation. While the dataset used in this study provides valuable insights, its size and representativeness may be insufficient to capture the full spectrum of variability in heart disease and chest pain characteristics across diverse populations. Future research efforts could benefit from the integration of larger and more diverse datasets, encompassing data from multiple sources and demographic groups. This would enhance the models' generalizability and robustness across varied patient populations and healthcare settings.

Another limitation lies in the inherent complexity and heterogeneity of cardiovascular diseases. The models developed in this study focus on predicting the presence of heart disease and classifying chest pain types based on a predefined set of features. However, cardiovascular diseases are multifaceted conditions influenced by a myriad of genetic, environmental, and lifestyle factors. Future enhancements could involve incorporating additional features and leveraging advanced machine learning techniques, such as deep learning and ensemble methods, to capture more nuanced patterns and relationships within the data. Furthermore, integrating longitudinal data spanning multiple time points could provide insights into disease progression and prognosis, enabling more personalized and timely interventions.

Additionally, the interpretability of the models remains a challenge. While the developed models demonstrate high accuracy in their predictions, the underlying decision-making processes may be opaque, limiting their interpretability and usability in clinical practice. Future research endeavors could focus on enhancing model interpretability through techniques such as feature importance analysis, model visualization, and transparent model architectures. This would enable clinicians to better understand the factors driving the models' predictions and facilitate trust and adoption in real-world healthcare settings.

Moreover, the models' performance may be affected by biases inherent in the data and model training processes. Biases stemming from demographic disparities, data collection methods, and algorithmic biases could introduce inaccuracies and inequities in the models' predictions, potentially exacerbating healthcare disparities. Addressing these biases requires concerted efforts to mitigate data biases, promote diversity and inclusion in dataset collection, and develop fair and transparent model training procedures.

In conclusion, while the developed models represent significant advancements in the field of cardiovascular disease prediction and classification, several limitations and opportunities for improvement exist. By addressing these limitations and embracing opportunities for future enhancements, researchers and healthcare professionals can continue to leverage machine learning technologies to advance our understanding of cardiovascular diseases and improve patient care outcomes.

VI. CONCLUSION

In conclusion, this project marks a significant step forward in leveraging machine learning techniques for cardiovascular disease prediction and chest pain classification. The development of accurate predictive models holds immense promise for enhancing clinical decision-making and improving patient outcomes in cardiovascular care.

Through the construction and evaluation of both binary and multi-class classification models, we have demonstrated the potential of machine learning to aid in the diagnosis and classification of heart conditions. The high accuracies achieved by these models underscore their effectiveness in identifying individuals at risk of heart disease and categorizing chest pain types, thereby enabling timely interventions and tailored treatment strategies.

However, it is essential to acknowledge the inherent limitations and challenges associated with predictive modeling in healthcare. While our models exhibit strong performance, further refinement and validation are necessary to ensure their generalizability across diverse patient populations and clinical settings. Additionally, efforts to enhance model interpretability, mitigate biases, and promote transparency are critical to fostering trust and adoption in real-world healthcare applications.

VII. REFERENCES

[1] *T. Amarbayasgalan, V. -H. Pham, N. Theera-Umpon, Y. Piao and K. H. Ryu, "An Efficient Prediction Method for Coronary Heart Disease Risk Based on Two Deep Neural Networks Trained on Well-Ordered Training Datasets," in IEEE Access, vol. 9, pp. 135210-135223, 2021, doi: 10.1109/ACCESS.2021.3116974.*

[2] *P. S. Sangle, R. M. Goudar and A. N. Bhute, "Methodologies and Techniques for Heart Disease Classification and Prediction," 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kharagpur, India, 2020, pp. 1-6, doi: 10.1109/ICCCNT49239.2020.9225673.*

[3] *S. Mohan, C. Thirumalai and G. Srivastava, "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques," in IEEE Access, vol. 7, pp. 81542-81554, 2019, doi: 10.1109/ACCESS.2019.2923707.*

[4] *Haoran, lyu*. 2022. A Machine Learning-Based Approach for Cardiovascular Diseases Prediction. In 2022 14th International Conference on Machine Learning and Computing (ICMLC) (ICMLC 2022), February 18-21, 2022, Guangzhou, China. ACM, New York, NY, USA, 11 Pages*

[5] *Harshit Jindal, Sarthak Agrawal, Rishabh Khera, Rachna Jain and Preeti Nagrath, Heart disease prediction using machine learning algorithms, IOP Conference Series: Materials Science and Engineering, Volume 1022, 1st International Conference on Computational Research and Data Analytics (ICCRDA 2020) 24th October 2020, Rajpura, India.*

[6] *A. Gupta, R. Kumar, H. Singh Arora and B. Raman, "MIFH: A Machine Intelligence Framework for Heart Disease Diagnosis," in IEEE Access, vol. 8, pp. 14659-14674, 2020, doi: 10.1109/ACCESS.2019.2962755.*

[7] *Srinivasan S, Gunasekaran S, Mathivanan SK, M B BAM, Jayagopal P, Dalu GT. An active learning machine technique based prediction of cardiovascular heart disease from UCI-repository database. Sci Rep. 2023 Aug 21;13(1):13588. doi: 10.1038/s41598-023-40717-1. PMID: 37604952; PMCID: PMC10442398.*

[8] *Dillip Narayan Sahu , Satheesh Kumar. Ensemble Machine Learning Models for Accurate Heart Disease Diagnosis. 2023, European Chemical Bulletin 2023. DOI: 10.48047/ecb/2023.12.si4.1691*

[9] Suriya Beguma , Farooq Ahmed Siddiqueb , Rajesh Tiwaric. A Study for Predicting Heart Disease using Machine Learning. 2021, Turkish Journal of Computer and Mathematics Education. Vol.12 No.10 (2021) 4584-4592

[10]R. V. Belfin, A. J. Shobana, M. Manilal, A. A. Mathew and B. Babu, "A Graph Based Chatbot for Cancer Patients," 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS), Coimbatore, India, 2019, pp. 717-721, doi: 10.1109/ICACCS.2019.8728499.