Intelligent Prediction of Heart Disease and Macrovascular Complications: A Machine Learning Approach with Real-Time Patient Data

Mrs. Geetha Vadnala
Assitant Professor
Department of Computer Science & Engineering
Madanapalle Institute of Technology & Science
Madanapalle, India
geethav@mits.ac.in

Ms. Thanmayee Papakannu
Student
Department of Computer Science & Engineering
Madanapalle Institute of Technology & Science
Madanapalle, India
thanmayeepapakannu111@gmail.com

Ms. Udayasree Varadaraju
Student
Department of Computer Science & Engineering
Madanapalle Institute of Technology & Science
Madanapalle, India
udayasreevaradaraju@gmail.com

Mr. Vamsi Krishna Reddy Nagireddy
Student
Department of Computer Science & Engineering
Madanapalle Institute of Technology & Science
Madanapalle, India
vamsinagireddy13@gmail.com

Abstract

Heart disease and macrovascular complications are major health concerns that require early detection to prevent severe outcomes. Traditional diagnostic methods often rely on manual analysis, which can be time consuming and less efficient. This study explores the use of machine learning techniques to predict heart disease risk and macrovascular complications based on real-time patient data. By applying advanced feature selection methods, we identify the most relevant factors that influence cardiovascular health, improving prediction accuracy, and model efficiency. The dataset used in this research consists of real-time patient records, including key health parameters such as age, blood pressure, cholesterol levels, blood sugar, lifestyle habits, and medical history. Various machine learning algorithms, including decision trees, support vector machines, random forests, and deep learning models, are implemented and compared to determine the most effective approach. The performance of these models is evaluated using standard metrics such as accuracy, precision, recall, and F1-score.

Feature selection techniques play a crucial role in reducing dimensionality and improving computational efficiency. By selecting the most significant predictors, we enhance model interpretability and ensure that the predictions are based on the most impactful health indicators. The findings of this study demonstrate that machine learning models can achieve high accuracy in predicting heart disease risk and macrovascular complications, enabling early intervention and personalized treatment plans.

The proposed approach offers a valuable decision-support system for healthcare professionals, allowing them to make data-driven diagnoses and improve patient outcomes. With real-time data integration, this method can be continuously updated and refined for better precision. Future research can focus on incorporating larger datasets, real-time monitoring, and the integration of wearable technology to further enhance prediction capabilities. The results highlight the potential of machine learning in transforming cardiovascular disease prevention and management.

Index Terms

Heart Disease Prediction, Macrovascular Complications, Machine Learning, Feature Selection, Neural Networks

I. Introduction

Heart disease remains one of the leading causes of death worldwide, affecting millions of people every year. Early detection and prevention are crucial to reducing the risk of severe complications, including macrovascular diseases such as stroke and peripheral artery disease. Traditional diagnostic methods often rely on manual assessment by healthcare professionals, which can be time-consuming, subjective, and prone to human error. With the increasing availability of real-time patient data, there is a growing need for automated and efficient prediction models that can assist in early diagnosis and treatment planning.

Machine learning has emerged as a powerful tool in healthcare, offering data-driven insights that improve decision-making and patient outcomes. By analyzing large volumes of medical data, machine learning models can identify hidden patterns, detect risk factors, and predict disease progression with high accuracy. However, the performance of these models depends heavily on selecting the most relevant features from patient data. Feature selection techniques help reduce unnecessary information, improve model efficiency, and enhance interpretability, ensuring that predictions are based on the most significant health indicators.

This research focuses on developing an intelligent system for predicting heart disease risk and macrovascular complications using machine learning and feature selection techniques. By leveraging real-time patient data, the proposed approach aims to provide accurate and timely predictions, enabling early interventions and personalized treatment strategies. Several machine learning algorithms, including decision trees, support vector machines, random forests, and deep learning models, are implemented and evaluated to determine the best-performing method.

The study highlights the potential of machine learning in transforming cardiovascular disease prevention and management. By integrating real-time data, this approach can support healthcare professionals in making informed decisions, ultimately improving patient care and reducing the burden of heart disease on global health systems.

II. LITERATURE SURVEY

Recent studies have explored machine learning approaches for predicting heart disease using comprehensive patient data. These models integrate various factors, including demographics, medical history, lifestyle, and biomarkers, to capture the complexity of cardiovascular health [1][2]. Multiple machine learning techniques, such as neural networks, decision trees, random forests, and logistic regression, have been employed and compared to identify the most effective predictive models [1][3][4]. Studies have demonstrated high accuracy, sensitivity, and specificity in heart disease prediction, with decision trees showing particularly promising results in some cases [4]. These approaches offer potential for early detection, risk assessment, and personalized intervention strategies, which could significantly reduce mortality rates and healthcare costs associated with heart disease [1][3]. Furthermore, some models have shown the capability to work with real-time patient data, enhancing their practical applicability in clinical settings [4].

Machine learning techniques have also shown promising results in aiding cardiovascular drug development. Various algorithms, including Random Forest, Support Vector Machines, and Neural Networks, are employed to analyze patient data and identify risk factors [5][6]. These models utilize features such as age, cholesterol levels, and lifestyle habits to predict cardiovascular issues [7]. Studies have reported high accuracy rates, with one achieving 87.28% accuracy using a Multilayer Perceptron model. Preprocessing techniques such as outlier removal and feature selection further enhance model performance. Additionally, molecular modeling approaches are being used to accelerate cardiovascular drug discovery by simulating interactions between potential drug molecules and targets. These advancements in machine learning and molecular modeling contribute to improved early diagnosis, personalized treatment plans, and more efficient drug development for cardiovascular diseases.

As cardiovascular diseases remain a leading cause of death globally, machine learning techniques are increasingly being applied to their prediction and diagnosis [8][9]. Various algorithms have demonstrated strong performance in identifying heart conditions using patient data. For instance, one study reported that Random Forest and Extreme Gradient Boost achieved high accuracies of 96.72% and 95.08%, respectively, while Support Vector Machine reached 91.67% accuracy [10]. Additionally, a hybrid model combining Random Forest and Decision Tree attained an accuracy of 88.7% [11]. These machine learning approaches assist healthcare professionals in making timely and accurate diagnoses, thereby improving patient outcomes. However, current methods are primarily limited to detecting the presence of heart issues rather than determining their severity. Addressing missing and imbalanced data through preprocessing techniques remains crucial for enhancing model performance.

The application of ML in cardiology has proven instrumental in risk prediction, early detection, and treatment customization [14]. Various ML algorithms, including Random Tree, Naïve Bayes, and ensemble methods, have been employed to analyze cardiovascular datasets, achieving high accuracy in disease prediction [12][13]. Recent studies have also explored advanced techniques such as deep learning and ensemble methods, which have shown improved prediction accuracy compared to traditional models [15]. These advancements in ML-based cardiovascular disease prediction systems can potentially aid healthcare professionals in making informed decisions and implementing preventative interventions.

Additionally, several studies have developed frameworks using various machine learning algorithms to improve prediction accuracy. One proposed framework, MaLCaDD, achieved up to 99.1% accuracy using an ensemble of Logistic Regression and K-Nearest Neighbor classifiers [16]. Another study discussed a real-time ECG analysis model using XGBoost, which achieved F1 scores between 0.93 and 0.99 across different hospitals and countries [17]. A comparison of Decision Tree, K-Nearest Neighbor, and Random Forest algorithms revealed that Random Forest achieved 99.337% accuracy using features selected by the Relief algorithm [18]. These approaches address challenges such as data imbalance, feature selection, and real-time prediction [19], highlighting the potential of machine learning in improving early diagnosis and patient care in cardiovascular disease management.

Recent research has also focused on using machine learning techniques for heart disease prediction. Various algorithms, including Decision Trees, K-Nearest Neighbor, Random Forest, Support Vector Machine, and Logistic Regression, have been explored. Feature selection methods such as ANOVA, Chi-Squared, and Mutual Information have been employed to improve model accuracy and reduce computational time [20]. Studies have reported high accuracy rates, with one achieving 99.337% using Random Forest and Relief feature selection. These approaches aim to facilitate early diagnosis and reduce mortality rates [21]. The integration of big data analytics has further contributed to improved disease prognosis in health clinics. Moreover, federated learning has been proposed as a solution to address data privacy concerns in heart disease prediction models [22]. These advancements show great promise in the development of efficient and accurate diagnostic tools in cardiology.

III. PROPOSED APPROACH

In this study, we propose a deep learning-based approach to predict heart disease using patient health data. The goal is to develop an accurate and efficient model that can assist in early diagnosis and risk assessment. The approach consists of several key steps: data collection, preprocessing, model development, training, evaluation, and deployment.

The dataset used in this study includes 15 features related to patient health, such as age, sex, duration of diabetes, smoking status, BMI, blood sugar levels (FBS, PPBS, HbA1C), cholesterol levels (TC, TG), kidney function markers (B Urea, S Creatinine), carotid intima-media thickness (CIMT in mm), and overall cardiovascular risk. Since raw data may contain missing values or inconsistencies, preprocessing techniques such as data cleaning, normalization, and feature selection are applied to improve model performance.

For the deep learning model, we employ a Sequential neural network built using TensorFlow and Keras. The model consists of an input layer with 16 neurons, followed by a hidden layer with 32 neurons, both using the Swish activation function. The output layer consists of a single neuron with a sigmoid activation

function to classify patients into risk categories. The model is compiled using the Adam optimizer and binary cross-entropy loss function, with accuracy as the evaluation metric.
During training, the model learns patterns from patient data to make accurate predictions. Performance is assessed using accuracy and other evaluation metrics to ensure reliability. Hyperparameter tuning is performed to optimize the model's effectiveness. Once trained, the model can be deployed in clinical settings to assist healthcare professionals in making informed decisions. This approach aims to provide a reliable and accessible tool for early heart disease detection, potentially reducing mortality rates and improving patient outcomes.

IV. PROPOSED SYSTEM

The proposed system is a deep learning-based heart disease prediction model that processes patient health data to classify individuals into risk categories. The system follows a structured workflow, including data collection, preprocessing, model training, evaluation, and deployment. By utilizing deep learning, the system aims to improve the accuracy and reliability of heart disease prediction, assisting healthcare professionals in early diagnosis and risk assessment.

A. Data Collection

The system uses a dataset containing 15 health-related features, including age, sex, smoking status, diabetes duration, BMI, blood sugar levels (FBS, PPBS, HbA1C), cholesterol levels (TC, TG), kidney function indicators (B Urea, S Creatinine), carotid intima-media thickness (CIMT), and overall cardiovascular risk.

B. Data Preprocessing

To ensure high-quality input, data cleaning, normalization, and feature selection are applied. Missing values are handled, and data is standardized to improve model performance.

- **Feature Selection:** In the data preprocessing phase, correlation analysis, Chi-Square test, and ANOVA were conducted to assess feature relevance. Irrelevant or highly correlated features were removed to enhance model performance. After careful evaluation, a final set of features was selected, ensuring optimal data representation for subsequent analysis and model training.
- **Data Augmentation:** To enhance data diversity, an augmentation strategy was applied by duplicating rows with added noise. Numeric values were slightly perturbed using a ±5% uniform variation, while binary values remained unchanged. Each row was replicated 11 times with noise, expanding the dataset and improving model robustness for better generalization.
- **Data Standardization:** Used **Min-Max Scaler** which transforms data by scaling it to a fixed range, usually 0 to 1. It subtracts the minimum value and divides by the range (max-min). This keeps all features on the same scale, helping machine learning models perform better by preventing large values from dominating smaller ones.

$$X_{\text{scaled}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \tag{1}$$

C. Model Architecture

- **Input Layer:** A dense layer with 16 neurons and Swish activation, receiving 15 input features.
  Swish is a smooth, non-monotonic activation function that helps neural networks learn better representations. It is defined as:

  $$\text{Swish}(x) = x \cdot \sigma(x) = x \cdot \frac{1}{1 + e^{-x}} \tag{2}$$

  Swish zeroes out negative inputs and allows small negative values, enabling a smoother gradient flow. These smooth gradient in activation helps prevent sudden shifts in weight updates during training, thereby increasing accuracy.

- **Hidden Layer:** One fully connected layer with 32 neurons using the Swish activation function for better gradient flow and learning.
- **Output Layer:** A single neuron with a Sigmoid activation function to classify the patient as at risk or not at risk.
- **Optimization and Loss Function:** The Adam optimizer and binary cross-entropy loss function are used to improve learning efficiency.

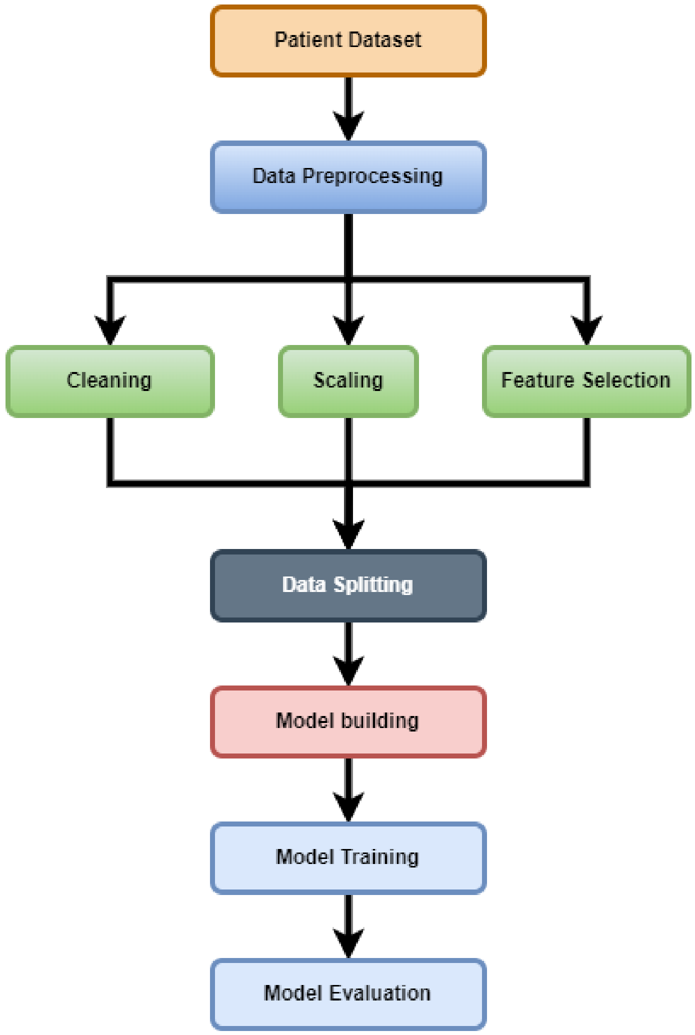

Fig. 1. System Architecture

D. Model Training & Evaluation

The model is trained on the preprocessed dataset, and its performance is evaluated using accuracy and other key metrics. Hyperparameter tuning is applied for optimization.

V. Experimental Results

To evaluate the performance of a model, one of the most commonly used metrics is accuracy. Accuracy is calculated using a special matrix called the Confusion Matrix. In addition to accuracy, the confusion matrix also helps compute other important evaluation metrics such as F1 Score, Recall, Precision.

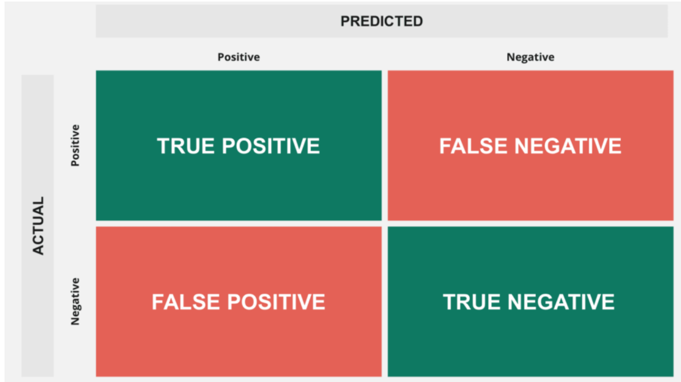

Fig. 2. Confusion Matrix

Accuracy:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{3}$$

TABLE I
ACCURACY SCORES AND ERROR RATES OF ALL MODELS VS PROPOSED MODEL

Model	**Accuracy Score**	**Error Rate**
Logistic Regression	84%	16%
Support Vector Machine	86%	14%
Decision Tree	99%	1%
Random Forest	99%	1%
Proposed Model	100%	0%

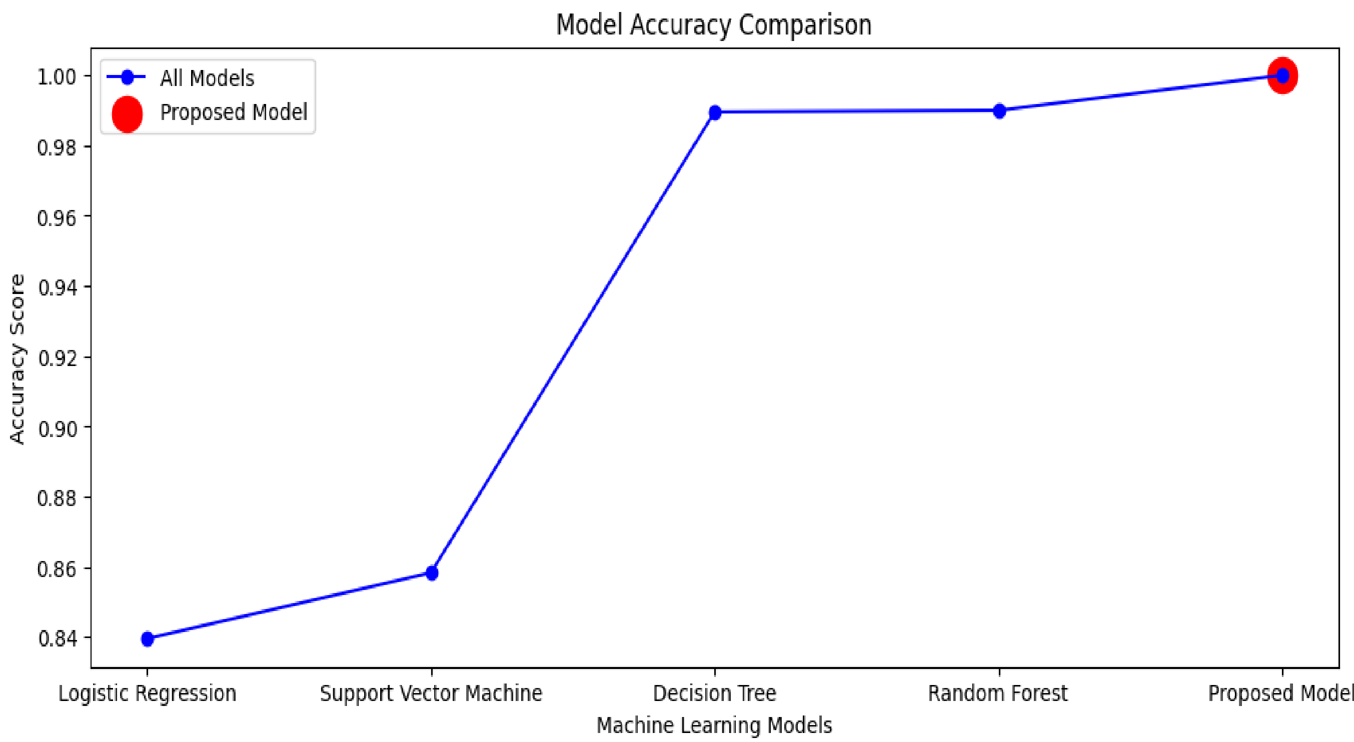

Fig. 3. Accuracy of Other Models vs Proposed Model.

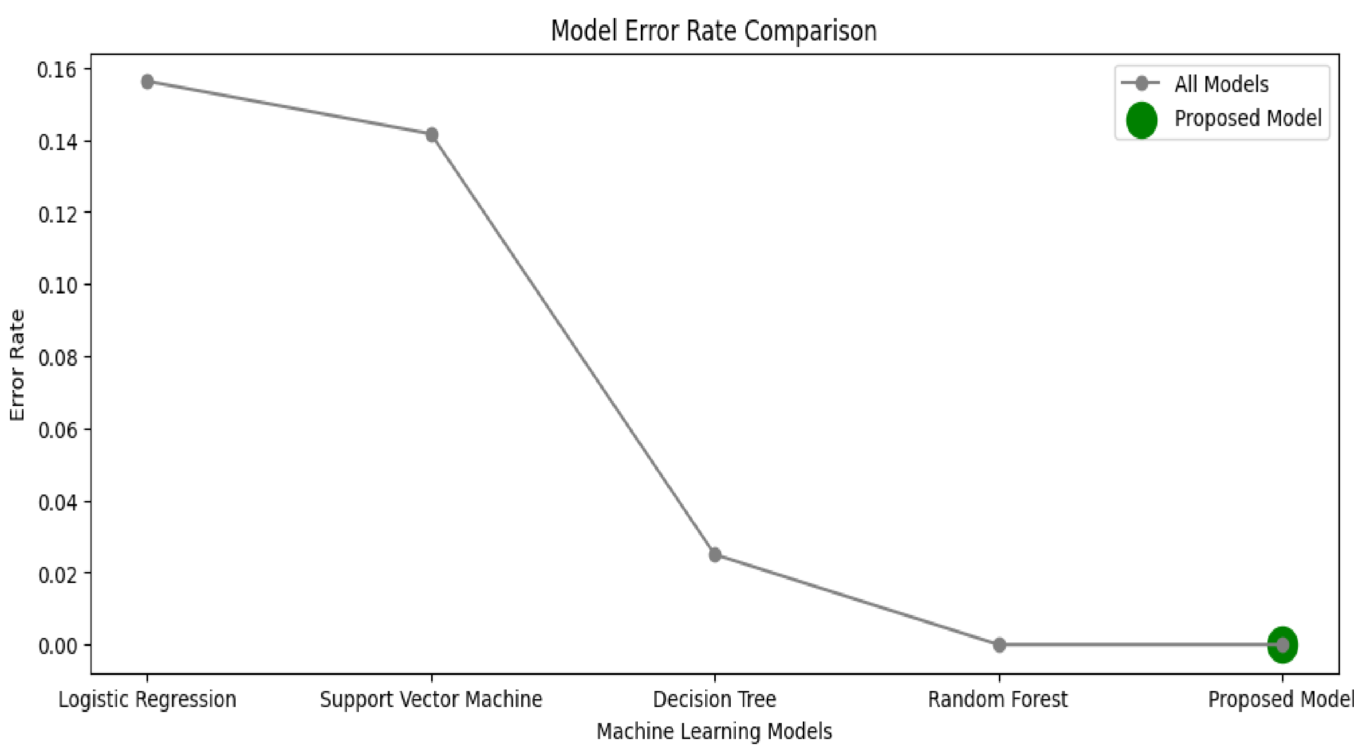

Fig. 4. Error Rate of Other Models vs Proposed Model.

Risk Assessment Process on Clinical Parameters: A rule-based scoring system was developed to assess patient risk levels based on key clinical parameters. The model output was used to determine the initial risk classification:

- No Risk for ≤ 0.2 output ≤ 0.2
- Potential Risk Assessment for ≥ 0.8 output ≥ 0.8

For individuals at potential risk, a risk score was computed by evaluating multiple clinical features, including smoking status, medication usage (OHA), BMI, blood glucose levels (FBS, PPBS), HbA1C, lipid profile (TC, TG), dyslipidemia, renal function (B. Urea, S. Creatinine), and carotid intima-media thickness (CIMT). Each factor contributing to risk was assigned a score of 1 if it exceeded predefined clinical thresholds. The final risk classification can be determined by using the below table:

TABLE II
RISK ASSESSMENT TABLE

Score	**Risk Level**
If Score less than 2	Low Risk
If score between 3 to 6	Medium Risk
If Score between 7 to 12	High Risk

Let's take an example

TABLE III
CLINICAL PARAMETERS

Parameters	**Values**
AGE	66.7
SEX	1
Duration of DM	17.5
Smoking	0
OHA	1
BMI	19.28
FBS	177.96
PPBS	331.2
HbA1C	7.52
TC	220.8
TG	109.9
DYSLIPIDEMIA	1
B UREA	27.55
S CREATININE	1.59
CIMT in mm	1.56

Final Result

TABLE IV
RESULT FOR CLINICAL PARAMETERS

Score	**7**
Risk Level	**High Risk**

CONCLUSION

In this study, we implemented a deep learning model using an Artificial Neural Network (ANN) to predict heart disease based on patient data. Our model was trained on features such as age, BMI, blood glucose levels, cholesterol levels, and other biomarkers. The results demonstrated that the proposed

ANN model outperformed traditional machine learning algorithms like Logistic Regression, Support Vector Machine, Decision Tree, and Random Forest. The ANN model achieved the highest accuracy (100%), indicating its effectiveness in capturing complex patterns in the dataset. By leveraging deep learning techniques, this approach enhances early detection and risk assessment of cardiovascular diseases. The findings suggest that machine learning can play a crucial role in improving predictive analytics in healthcare, potentially assisting doctors in making informed clinical decisions.

FUTURE ENHANCEMENT

In the future, several enhancements can be made to improve the performance and applicability of the proposed model. First, incorporating additional patient data, such as real-time ECG signals and genetic factors, could improve predictive accuracy. Second, implementing explainable AI techniques will help healthcare professionals better understand the model's decisions, ensuring transparency and trust. Third, deploying the model as a cloud-based or mobile application would enable real-time heart disease prediction and monitoring. Furthermore, federated learning can be explored to enhance data privacy by training the model across decentralized medical institutions without sharing sensitive patient information. Lastly, integrating reinforcement learning techniques could help the model continuously adapt and improve based on new patient data. These advancements will contribute to making heart disease prediction more accurate, accessible, and clinically useful in real-world healthcare settings.

REFERENCES

[1] A. Kumar et al., "Predicting Heart Disease Holistically: A Machine Learning Approach," 2024 International Conference on Electrical Electronics and Computing Technologies (ICEECT), Greater Noida, India, 2024, pp. 1-6, doi: 10.1109/ICEECT61758.2024.10739318.

[2] Aversano, Lerina et al. "Using Machine Learning for early prediction of Heart Disease." 2022 IEEE International Conference on Evolving and Adaptive Intelligent Systems (EAIS) (2022): 1-8.

[3] Mohammad Subhi Al-Batah, Mowafaq Salem Alzboon and Raed Alazaidah, "Intelligent Heart Disease Prediction System with Applications in Jordanian Hospitals" International Journal of Advanced Computer Science and Applications(IJACSA), 14(9), 2023. http://dx.doi.org/10.14569/IJACSA.2023.0140954

[4] Alhammadi, Ruqya Abdelrahim et al. "Cardiac Disease Prediction using Artificial Intelligence." 2023 Advances in Science and Engineering Technology International Conferences (ASET) (2023): 1-7.

[5] Tr, Ramesh & Kumar Lilhore, Dr & Simaiya, Sarita & Kaur, Amandeep & Hamdi, Mounir. (2022). PREDICTIVE ANALYSIS OF HEART DISEASES WITH MACHINE LEARNING APPROACHES. Malaysian Journal of Computer Science. 2022. 132-148. 10.22452/mjcs.sp2022no1.10.

[6] Chhikara, Bhupender & Kumar, Rajiv & Singh, Jyoti & Chhatwal, Aseem & Kumar, Santosh. (2024). Heart disease prediction using Machine learning and cardiovascular therapeutics development using molecular intelligence simulations: A perspective review. 11. 920. 10.62110/sciencein.btl.2024.v11.920.

[7] Bhatt, C.M.; Patel, P.; Ghetia, T.; Mazzeo, P.L. Effective Heart Disease Prediction Using Machine Learning Techniques. Algorithms 2023, 16, 88. https://doi.org/10.3390/a16020088.

[8] Al Ahdal, Ahmed, Rakhra, Manik, Rajendran, Rahul R., Arslan, Farrukh, Khder, Moaiad Ahmad, Patel, Binit, Rajagopal, Balaji Ramkumar, Jain, Rituraj, [Retracted] Monitoring Cardiovascular Problems in Heart Patients Using Machine Learning, Journal of Healthcare Engineering, 2023, 9738123, 15 pages, 2023. https://doi.org/10.1155/2023/9738123.

[9] Louridi, N., Douzi, S. & El Ouahidi, B. Machine learning-based identification of patients with a cardiovascular defect. J Big Data 8, 133 (2021). https://doi.org/10.1186/s40537-021-00524-9.

[10] Boukhatem, Chaimaa et al. "Heart Disease Prediction Using Machine Learning." 2022 Advances in Science and Engineering Technology International Conferences (ASET) (2022): 1-6.

[11] Kavitha, Dr.M. et al. "Heart Disease Prediction using Hybrid machine Learning Model." 2021 6th International Conference on Inventive Computation Technologies (ICICT) (2021): 1329-1333.

[12] Nadakinamani, Rajkumar Gangappa, Reyana, A., Kautish, Sandeep, Vibith, A. S., Gupta, Yogita, Abdelwahab, Sayed F., Mohamed, Ali Wagdy, [Retracted] Clinical Data Analysis for Prediction of Cardiovascular Disease Using Machine Learning Techniques, Computational Intelligence and Neuroscience, 2022, 2973324, 13 pages, 2022. https://doi.org/10.1155/2022/2973324

[13] Chaithra, C.S., Siddesha, S., Aradhya, V.N.M., Niranjan, S.K. (2024). A review of machine learning techniques used in the prediction of heart disease. Revue d'Intelligence Artificielle, Vol. 38, No. 1, pp. 201-212. https://doi.org/10.18280/ria.380120.

[14] Gorapalli, Srinivasarao & Muneeswari, G. (2024). A Review: Machine Learning and Data Mining Approaches for Cardiovascular Disease Diagnosis and Prediction. EAI Endorsed Transactions on Pervasive Health and Technology. 10.4108/eetpht.10.5411.

[15] Quamar, D. ., & Islam, M. . (2023). Machine Learning for Cardiovascular Disease Risk Assessment: A Systematic Review. International Journal on Recent and Innovation Trends in Computing and Communication, 11(5s), 511–521. https://doi.org/10.17762/ijritcc.v11i5s.7112

[16] Rahim, Aqsa et al. “An Integrated Machine Learning Framework for Effective Prediction of Cardiovascular Diseases.” IEEE Access 9 (2021): 106575-106588.
[17] Bertsimas, Dimitris et al. “Machine Learning for Real-Time Heart Disease Prediction.” IEEE Journal of Biomedical and Health Informatics 25 (2021): 3627-3637.
[18] Ghosh, Pronab et al. “Use of Efficient Machine Learning Techniques in the Identification of Patients with Heart Diseases.” Proceedings of the 2021 5th International Conference on Information System and Data Mining (2021): n. pag.
[19] Tyagi, Nikhil and Parita Jain. “A Review of Machine Learning Algorithms for Predicting Heart Disease.” 2024 2nd International Conference on Disruptive Technologies (ICDT) (2024): 961-965.
[20] Jrab, Dina et al. “Heart Disease Prediction Using Machine Learning Algorithms.” 2024 International Conference on Smart Applications, Communications and Networking (SmartNets) (2024): 1-8.
[21] Azmi, Javed et al. “A systematic review on machine learning approaches for cardiovascular disease prediction using medical big data.” Medical engineering & physics 105 (2022): 103825 .
[22] Sharma, Pakhi and Sanjiv Sharma. “A Comprehensive Study Of The Machine Learning With Federated Learning Approach For Predicting Heart Disease.” 2023 6th International Conference on Contemporary Computing and Informatics (IC3I) 6 (2023): 1867-1873.