Analysis of Cardiovascular Disease using Traditional Machine Learning

[1]Teesta Chanda , [2]Gyalgen Lepcha, [3]Tanbir Alam,[4]Vivek Thapa

[1,2,3]School of Information Technology, [4]Assistant Professor
The ICFAI University Sikkim, Gangtok, India
[1]11teestachanda@gmail.com, [2]gyalgen22@gmail.com,
[3]tanbiralam321@gmail.com,[4]vivekthapa@iusikkim.edu.in

***Abstract*—The heart is a critical organ in the human body, and any deficiency in its functioning can lead to life-threatening consequences. Cardiovascular diseases are the leading cause of fatality worldwide. Early detection and diagnosis can prevent the risk of life and reduce treatment costs. This paper attempts to investigate a well-defined and utilitarian heart disease research program model based on traditional machine learning models, utilizing datasets from Kaggle and UCI Repository, and determining variables including hypertension, cholesterol level, smoking, sedentary lifestyle, etc. This study uses supervised learning algorithms, including Decision Tree, Random Forest, Logistic Regression, and Support Vector Machine (SVM). The algorithms are evaluated based on their respective performance. Overall, the study showcases the potential of Machine Learning as a decision-making service in health care with timely diagnosis and minimized cost. This study also tests for feature importance, in order to efficacy which clinical and lifestyle aspects would matter throughout the risk of heart disease. If health service organizations can identify which features provide the best information when modelling risk, then they would have the potential to facilitate a priority for patient screening while developing the best individualized treatments.**

***Index Terms*—Cardiovascular Disease (CVD), Machine Learning(ML), Supervised learning, Decision Tree, Random Forest, Logistic Regression, Support Vector Machine (SVM)**

I. INTRODUCTION

The human heart is a vital organ responsible for pumping blood throughout the body, supplying oxygen and essential nutrients to tissues, and removing waste products like carbon dioxide. A healthy heart can develop heart diseases over time due to a combination of lifestyle choices, medical conditions, and environmental or genetic factors. [1]It results in early symptoms, such as an irregular heartbeat, shortness of breath, chest discomfort, sudden dizziness, nausea, swollen feet, and a cold sweat. [2]The symptoms are common to other illnesses and might be confused with indicators of aging, making diagnosis difficult for practitioners. There are many more factors like unhealthy diet, lack of physical activity, smoking, excessive alcohol consumption, etc. [3] Heart disease, also referred to as cardiovascular illness, is a general term for conditions that affect the heart and blood vessels, contributing to millions of deaths annually. [4] Heart disease is an appalling illness that puts myriad lives in jeopardy, and the best way to fight it is to identify it in its early stages before it rots in the victim's body, as it is much easier to cure.

According to WHF, by 2021, deaths from CVDs increased to 20.5 million, a 60% increase from 12.1 million in 1990. According to BHF, CHD, which is a type of CVD, is a major contributor, causing around 9 million deaths annually. In 2021, CHD was recorded as the world's leading single cause of death annually. In 2021, CHD was recorded as the world's single leading cause of death, responsible for 1 in 7 deaths globally. [5]According to WHO, CVDs are the number one cause of death globally: more people die annually from CVDs than from any other cause. [6]The World Health Organization (WHO) announced that 17.9 million people died in 2019 because of it, 85% of total CVD deaths are caused by heart attacks and strokes, and around 1.28 billion adults globally die due to high blood pressure. Globally, poorer countries suffer more due to their high death rates and the high cost of treatment and surgeries.

A well-built and accurate forecasting model can help in effectively pinpointing individuals more likely to develop heart diseases. With this insight, timely actions and proper medical attention aid in reducing health risk and cut down treatment costs. [7]According to Saini (2023), Machine learning algorithms are gradually revolutionizing heart disease prediction since they can handle complex, multi-dimensional data sets.

[8]In recent times, computer technology and machine learning techniques are being used to make medical aid software as a support system for the early diagnosis of heart disease. Identification of any initial heart-related condition can significantly reduce the risk of fatal outcomes. [9]Therefore, it is crucial to use data mining and machine learning techniques to predict the likelihood of developing heart disease in order to save lives and reduce the economic burden on society.[10] Machine learning (ML), which falls under the umbrella of AI, utilizes mathematical models of varying complexity to facilitate non-instructional learning in computers. In the medical field, a vast amount of data is generated daily using data mining techniques, and we can find hidden patterns that can be used for clinical diagnosis. Machine learning is a way of manipulating and extracting implicit, previously unknown, or known, and potentially useful information about data. Machine learning is a broad and diverse field with an expanding scope and implementation. Machine learning incorporates various classifiers of Supervised, Unsupervised, Reinforcement, and Ensemble learning, which are used to predict and find the accuracy of the given dataset. [11] In supervised machine learning, training data is provided to the machine, which works as a supervisor that teaches the machine to predict the output accurately. The model learns patterns from historical data and makes predictions based on new inputs. Common algorithms used in supervised learning include:

- Decision Tree is a type of supervised learning that recursively splits data based on feature values to create a tree-like model for classification or regression.

- Logistic Regression is used for binary classification, such as predicting whether a patient has heart disease (Yes/No).
- Support Vector Machine (SVM) finds the optimal decision boundary to separate data points into different classes.
- Random Forest is an ensemble technique that uses multiple decision trees to improve prediction accuracy.

Unlike supervised learning, unsupervised learning deals with unlabeled data, meaning the algorithm identifies patterns and relationships without predefined output labels. This technique is useful for discovering hidden structures within medical data. Common unsupervised learning techniques include: Clustering (e.g., K-Means, DBSCAN), Principal Component Analysis (PCA), and Autoencoders.

The objective of this research paper is to develop an effective and accurate analysis model for predicting heart disease using traditional machine-learning techniques. By analysing different risk factors, it aims to identify people who may develop heart problems early, allowing them to take preventive measures and get timely medical help. This study explores various machine learning techniques to improve accuracy and find hidden patterns in medical data. By using data mining, it can process large healthcare datasets to support better diagnosis, lower treatment costs, and help doctors make more informed decisions.

II. Related Work

[12] This research paper examines the application of machine learning models in predicting heart disease with emphasis on early diagnosis to reduce mortality rates. The authors contrast different machine learning models, including k-nearest neighbor (KNN), decision tree, linear regression, and support vector machine (SVM), using the UCI heart disease dataset. This research indicates that the accuracy of the heart disease prediction model is 95.7%, which testifies to the effectiveness of machine learning in medical diagnosis.

[13] Explore heart disease prediction with machine learning, implementing Logistic Regression, Support Vector Machine (SVM), and Naive Bayes on a dataset containing 12 features. The outcome reveals SVM as the most accurate (86%), followed by Logistic Regression (83%) and Naive Bayes (79%). The research demonstrates the efficacy of machine learning in disease detection at an early stage. It is, however, constrained by dataset size.

[14] It investigates the forecasting of heart disease using machine learning algorithms, in which Logistic Regression, Decision Tree, Random Forest, and Support Vector Machine are applied to the UCI Heart Disease Dataset. The accuracy is highest for Random Forest, closely followed by Support Vector Machine. The study shows the potential of machine learning in early disease diagnosis and improved clinical decision-making. The study is, however, limited by the dataset size.

[15] Explored cardiovascular disease forecasting using hybrid machine learning models that combine ensemble learning methods like Gradient Boosting and Random Forest on the heart disease datasets from Cleveland and IEEE Dataport. The findings show that ensemble learning can be used to improve the accuracy and predictability of the predictions relative to single models. The study points to the benefits of using combined machine-learning models to enhance diagnostic ability for disease. It is, however, faced with the constraint of datasets.

[16] This paper built a machine learning model using Random Forest, SVM, Naïve Bayes, and Decision Tree on the UCI Heart Disease dataset, which has 14 features. The study compared them and found that Random Forest was the most accurate at 99%, followed by SVM at 98%, and the least accurate was Decision Tree. The model is designed to help physicians diagnose heart disease. This study uses only one dataset.

III. Objective

The objective of this project is to develop a robust and easy-to-use desktop application for analysing and predicting cardiovascular disease, leveraging conventional machine learning methods. The system aims to enhance early diagnosis and awareness by integrating comprehensive data analysis, comparative algorithm evaluation, and intuitive user interface design. By supporting data-driven insights and presenting clear, interpretable predictions, the application seeks to aid healthcare practitioners and individuals in understanding potential cardiovascular risks. Through systematic investigation, the project will:

- Conduct a comprehensive survey and analysis of various cardiovascular diseases, examining their occurrence, associated risk factors, and healthcare burden.
- Identify and evaluate diverse data sources relevant to heart disease prediction and diagnosis.
- Perform a comparative study of traditional machine learning algorithms to assess their accuracy, efficiency, and interpretability in predicting cardiovascular conditions.
- Design and develop an accessible desktop application that effectively communicates prediction results to end users, supporting informed health decisions.

IV. METHODOLOGY

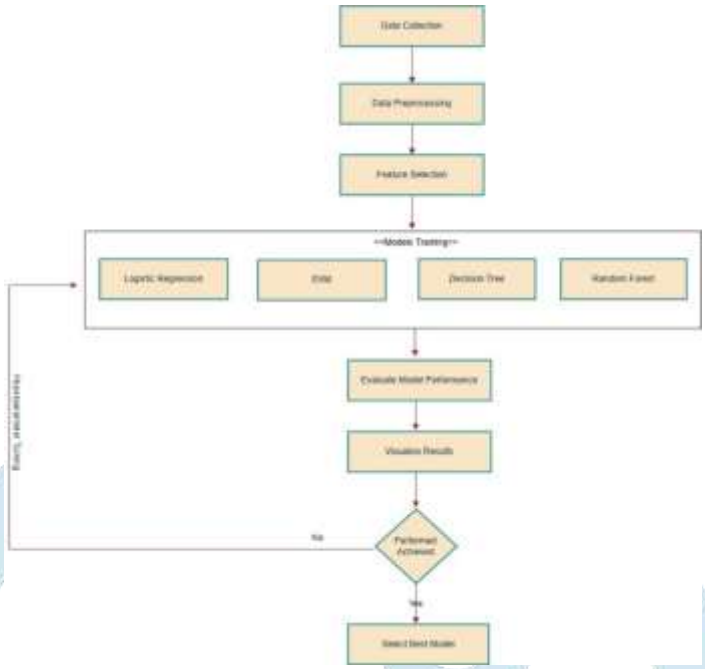

Fig. 1. Training Model

The steps of the proposed training methodology are as follows:

- To apply machine learning algorithms in cardiovascular disease prediction, first of all, import important libraries like Pandas, NumPy, Matplotlib, and Scikit-learn.
- Next is to collect the raw data and import the dataset of cardiovascular diseases.
- Preprocessing and data cleaning are then performed. This also involves handling missing values, removing noise, and normalizing data.
- After feature selection, select the most useful variables from pre-processed data to enhance the efficiency and performance of the models.
- Training the Models of Machine Learning, including Random Forests, Decision Trees, SVM, and Logistic Regression using the chosen features.
- It assesses the model's performance to determine which one delivers the best results, using evaluation metrics such as accuracy or precision.
- Visualization, which displays model evaluation results in a visually meaningful format such as graphs or charts.
- The condition will apply if the model's accuracy is greater than 80%; this means the model performs well enough and can be considered for further deployment if less than 80% revisits the earlier steps.

Lastly, select the best models after evaluation, which are optimized and prepared for deployment in a real-world scenario.

Fig. 2. Prediction Model

The steps of the proposed predictive methodology are as follows:

- First, gather raw data and enter it into the system.
- After entering the data is cleaned and made ready for processing, including deleting duplicates and normalizing data.
- Feature selection, which is the selection of the most useful variables from the pre-processed data, so that the models become more efficient and effective.
- After the application, a trained machine learning model is used by the system to predict from the data.
- Lastly, presenting the model's predictions or findings in a clear manner.

Abbreviations and Acronyms

Table 1 : Abbreviations and Acronyms

S. No.	Abbreviations	Meaning
1.	**AEHOM**	Autoencoder-based Hyperparameter Optimization Model
2.	**AI**	Artificial Intelligence
3.	**ANN**	Artificial Neural Network
4.	**BHF**	British Heart Foundation
5.	**BP**	Blood Pressure
6.	**CHD**	Coronary Heart Disease
7.	**CVD**	Cardio Vascular Disease
8.	**DT**	Decision Tree
9.	**ECG**	Electrocardiogram
10.	**KNN**	K-Nearest Neighbors
11.	**ML**	Machine Learning
12.	**MLDCNN**	Machine Learning Deep Convolutional Neural Network
13.	**PCA**	Principal Component Analysis
14.	**RFEM**	Random Forest Ensemble Model
15.	**SVM**	Support Vector Machine
16.	**UCI**	University of California, Irvine
17.	**USCOM**	Ultrasound Cardiac Output Monitoring
18.	**WHF**	World Heart Federation
19.	**WHO**	World Health Organisation
20.	**XGBoost**	Extreme Gradient Boosting

V. Results And Discussion

By using Random Forest, Decision Tree, Logistic Regression, and Support Vector Machine (SVM) algorithms on two datasets—a large dataset of 70,000 records with 98% accuracy and a smaller dataset of 304 records with 83% accuracy—predicting heart disease model is developed.16 features, including cold sweats or nausea, high blood pressure, high cholesterol, diabetes, smoking, family history, dizziness, and chest pain, were important features included in both datasets.

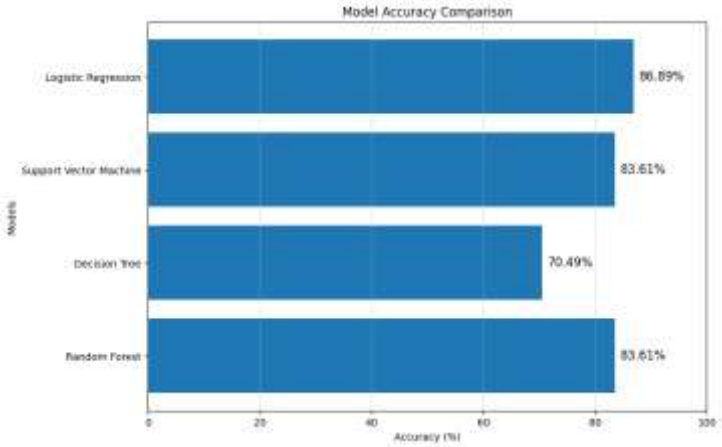

Fig.4: Accuracy trained Model

After a comprehensive analysis, the best model was saved with joblib python module and integrated into an intuitive desktop application created with Tkinter. This application allows users to enter patient data using the 16 features to assess heart disease risk (0-No, 1-Yes), facilitating effective clinical decision-making. A bar chart in Fig. 4 displays the results, illustrating the model's accuracy and predictions as shown in Fig. 5.

Fig.5: Predicted heart disease on the Desktop Application

VI. Acknowledgment

The authors would like to thank The ICFAI University, Sikkim, for providing the infrastructure and support essential to carry out this research. The authors appreciate their mentor, Mr. Vivek Thapa, Assistant Professor, School of Information Technology, for providing the constant encouragement, expert guidance, and valuable insights essential for this research success. The authors extend gratitude to all the researchers whose work inspired and informed our methodology, as well as to our peers and reviewers for their constructive feedback, which greatly helped improve the quality of this paper.

VII. Conclusion

This study aims to bridge serious gaps in the early detection of cardiovascular diseases (CVDs) through the combination of traditional machine learning techniques. The wide-ranging survey and studies of CVDs have laid a solid basis for thoughtfully understanding the intricate relationship between countless risk factors and disease progression. From the comprehensive analysis of open-source data hosted on Kaggle and the Cleveland UCI Repository, the research has established important attributes to build stable prediction models. Early research shows that machine learning models under supervision, namely, Random Forest, Logistic Regression, Decision Tree, and Support Vector Machine (SVM), show tremendous promise in accurately detecting heart diseases. Comparative analysis of the models is performed to identify the most efficient strategies, achieving a balance between predictive performance and computational expenses.

This desktop application, with its emphasis on simplicity, clear visualization, and interpretability, will not only improve diagnostic assistance but also support proactive health care management.

In summary, through data-driven information and real-world technological advancement, this research hopes to make a significant contribution to healthcare analytics, ultimately for earlier intervention, improved patient outcomes, and relief of the burden of heart diseases globally.

References

[1] K. V. V. Reddy, I. Elamvazuthi, A. A. Aziz, S. Paramasivam, H. N. Chua, and S. Pranavanand, “Heart disease risk prediction using machine learning classifiers with attribute evaluators,” Applied Sciences (Switzerland), vol. 11, no. 18, Sep. 2021, doi: 10.3390/app11188352.

[2] C. and M. Methods in Medicine, “Retracted: Implementation of a Heart Disease Risk Prediction Model Using Machine Learning,” Comput Math Methods Med, vol. 2023, no. 1, Jan. 2023, doi: 10.1155/2023/9764021.

[3] A. A. Almazroi, E. A. Aldhahri, S. Bashir, and S. Ashfaq, “A Clinical Decision Support System for Heart Disease Prediction Using Deep Learning,” IEEE Access, vol. 11, pp. 61646–61659, 2023, doi: 10.1109/ACCESS.2023.3285247.

[4] D. Salama AbdElminaam, N. Mohamed, H. Wael, A. Khaled, and A. Moataz, “MLHeartDisPrediction: Heart Disease Prediction using Machine Learning,” 2023.

[5] S. Jiang, "Heart Disease Prediction Using Machine Learning Algorithms," 2020.

[6] S. Ibrahim, N. Salhab, and A. El Falou, "Heart disease Prediction using Machine Learning," in 1st International Conference in Advanced Innovation on Smart City, ICAISC 2023 - Proceedings, Institute of Electrical and Electronics Engineers Inc., 2023. doi: 10.1109/ICAISC56366.2023.10085522.

[7] P. Kumar Bhowmik et al., "British Journal of Nursing Studies Advancing Heart Disease Prediction through Machine Learning: Techniques and Insights for Improved Cardiovascular Health," 2024, doi: 10.32996/bjns.

[8] V. Sharma, S. Yadav, and M. Gupta, "Heart Disease Prediction using Machine Learning Techniques," in Proceedings - IEEE 2020 2nd International Conference on Advances in Computing, Communication Control and Networking, ICACCCN 2020, Institute of Electrical and Electronics Engineers Inc., Dec. 2020, pp. 177–181. doi: 10.1109/ICACCCN51052.2020.9362842.

[9] C. M. Bhatt, P. Patel, T. Ghetia, and P. L. Mazzeo, "Effective Heart Disease Prediction Using Machine Learning Techniques," Algorithms, vol. 16, no. 2, Feb. 2023, doi: 10.3390/a16020088.

[10] M. Ahmed and I. Husien, "Heart Disease Prediction Using Hybrid Machine Learning: A Brief Review," 2024, Department of Agribusiness, Universitas Muhammadiyah Yogyakarta. doi: 10.18196/jrc.v5i3.21606.

[11] R. Bhalla and G. Ganesan, "Heart Disease Prediction using Machine Learning Method-A Review Article."

[12] H. Kamal et al., "Heart Disease Prediction Using Machine Learning," in 2024 2nd DMIHER International Conference on Artificial Intelligence in Healthcare, Education and Industry (IDICAIEI), IEEE, Nov. 2024, pp. 1–6. doi: 10.1109/IDICAIEI61867.2024.10842908.

[13] P. M Paithane, "Heart Disease Prediction Using Multiple Machine Learning Algorithms," Advances in Robotic Technology, vol. 2, no. 1, pp. 1–5, Jan. 2024, doi: 10.23880/art-16000114.

[14] M. M. Rakshana, "International Journal of Research Publication and Reviews Heart Disease Prediction Using Machine Learning Techniques," 2024. [Online]. Available: www.ijrpr.com

[15] M. Ahmed and I. Husien, "Heart Disease Prediction Using Hybrid Machine Learning: A Brief Review," 2024, Department of Agribusiness, Universitas Muhammadiyah Yogyakarta. doi: 10.18196/jrc.v5i3.21606.

[16] S. Singh, S. Vikram Singh, and Y. Aggarwal, "Heart Disease Prediction using Machine Learning Technologies," 2024. [Online]. Available: www.ijnrd.org

[17] N. Chandrasekhar and S. Peddakrishna, "Enhancing Heart Disease Prediction Accuracy through Machine Learning Techniques and Optimization," Processes, vol. 11, no. 4, Apr. 2023, doi: 10.3390/pr11041210.

[18] C. M. Bhatt, P. Patel, T. Ghetia, and P. L. Mazzeo, "Effective Heart Disease Prediction Using Machine Learning Techniques," Algorithms, vol. 16, no. 2, Feb. 2023, doi: 10.3390/a16020088.

[19] T. Venu Gopal and A. Sneha, "HEART DISEASE PREDICTION USING MACHINE LEARNING TECHNIQUES EMAIL: drtv.gopal@jntuh.ac.in," vol. 13, p. 7, 2023.

[20] Z. Noroozi, A. Orooji, and L. Erfannia, "Analyzing the impact of feature selection methods on machine learning algorithms for heart disease prediction," Sci Rep, vol. 13, no. 1, Dec. 2023, doi: 10.1038/s41598-023-49962-w.

[21] S. Ibrahim, N. Salhab, and A. El Falou, "Heart disease Prediction using Machine Learning," in 1st International Conference in Advanced Innovation on Smart City, ICAISC 2023 - Proceedings, Institute of Electrical and Electronics Engineers Inc., 2023. doi: 10.1109/ICAISC56366.2023.10085522.

[22] A. M. Qadri, A. Raza, K. Munir, and M. S. Almutairi, "Effective Feature Engineering Technique for Heart Disease Prediction with Machine Learning," IEEE Access, vol. 11, pp. 56214–56224, 2023, doi: 10.1109/ACCESS.2023.3281484.

[23] A. L. Yadav, K. Soni, and S. Khare, "HEART DISEASES PREDICTION USING MACHINE LEARNING."

[24] T. R. Ramesh, U. K. Lilhore, M. Poongodi, S. Simaiya, A. Kaur, and M. Hamdi, "PREDICTIVE ANALYSIS OF HEART DISEASES WITH MACHINE LEARNING APPROACHES," Malaysian Journal of Computer Science, vol. 2022, no. Special Issue 1, pp. 132–148, 2022, doi: 10.22452/mjcs.sp2022no1.10.

[25] C. and M. Methods in Medicine, "Retracted: Implementation of a Heart Disease Risk Prediction Model Using Machine Learning," Comput Math Methods Med, vol. 2023, no. 1, Jan. 2

[26] S. Jiang, "Heart Disease Prediction Using Machine Learning Algorithms," 2020.I. Koo, U. Zaman, H. Ha, and S. Nawaz, "Assessing the interplay of trust dynamics, personalization, ethical AI practices, and tourist behavior in the adoption of AI-driven smart tourism technologies," J. Open Innov. Technol. Mark. Complex., vol. 11, no. 1, p. 100455, 2025, doi: 10.1016/j.joitmc.2024.100455.