Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques

SHAIK ARIFA[1], MR M.NARESH[2]

[1]PG Scholar, Dept. of C.S.E, Newton's Institute Of Engineering, Macherla, Palnadu (Dt), A.P

[2]Assoc Professor, Dept. of C.S.E, Newton's Institute Of Engineering, Macherla, Palnadu (Dt), A.P

Abstract: Heart disease remains one of the leading causes of mortality worldwide. Predicting cardiovascular conditions is a critical challenge in the field of medical data analysis. Artificial Intelligence (AI) has proven to be effective in supporting decision-making and making accurate predictions using the vast amount of data generated by the healthcare industry. Additionally, AI techniques have seen increased application in recent advancements across the Internet of Things (IoT). While many studies offer only a limited exploration of heart disease prediction using AI methods, this paper presents a novel approach aimed at identifying significant features through advanced AI techniques to improve prediction accuracy. The proposed model utilizes various combinations of features along with several established classification methods. By employing a Hybrid Random Forest with Linear Model (HRFLM), the system achieves an enhanced performance with an accuracy rate of 88.7% in predicting heart disease

Key Words: Heart Disease, WHO (World Health Organization)**,** Cleveland Heart Disease Database Random Forest, Hybrid algorithm, Machine learning.

1. Introduction

Heart disease is one of the major causes of human deaths. According to the World Health Organization (WHO), around one-third of all deaths worldwide, approximately 18 million people, are related to heart disease (1). Factors such as alcohol and tobacco use, poor diets, and insufficient exercise increase the likelihood of developing heart disease, which can manifest symptoms such as obesity and high blood pressure (2). Since these symptoms can overlap with those of other diseases, obtaining an accurate diagnosis is crucial to reduce life-threatening risks. Machine learning techniques offer a promising avenue for predicting and diagnosing heart disease.

Various methods exist for diagnosing heart problems. In this review, we focus on three main research areas for heart disease detection: detection based on standard clinical information, detection based on electrocardiogram (ECG) signals, The heart, one of the body's major organs, plays a critical role by pumping blood through the blood vessels of the circulatory system. This system is essential for delivering blood, oxygen, and other vital substances to various organs throughout the body. As the heart is the central component of the circulatory system, any malfunction can lead to severe health issues, including death. consequently, machine learning algorithms are essential for accurate heart disease prediction. Recent studies have focused on combining these strategies to create hybrid machine learning algorithms. Data pre-processing is used in the research proposal to remove noisy data, fill in blanks when necessary, fill in default values when appropriate, and categorize attributes for prediction and decision-making at multiple levels. To assess the efficacy of the treatment approach, techniques including classification, accuracy, sensitivity, and specificity analysis are performed. An accurate cardiovascular disease prediction model is demonstrated by comparing the levels of accuracy of applying rules to the outcome variables.

2. Literature Survey

Prediction of heart disease at early stage using data mining and big data analytics: A survey N. K. Salma Banu; Suma Swamy: This article looks at different improvement patterns of mining (dm) models for the recognizable proof of coronary contamination. Factual extraction has a few components to make a powerful model for clinical structures to figure out coronary (HD) tainting utilizing patient data records, which represents a gamble related with ischemic coronary illness. Logical specialists can help patients by analyzing coronary pollution before it happens. Numerous realities accessible from the clinical study were analyzed involving records in the mining office, and important information known as measurements were isolated. Digging is a method for looking at huge scope data game plans to get ideas that can be kept and right now exist in uncertain associations, and to record characters that will assist better with understanding clinical data to forestall coronary tainting. There are window and backing. vector machine (svm). A few investigations have been finished for a model of rising assumptions with a social methodology and a mix of no less than two strategies.

3. System Analysis

3.1 Existing System

In this approach, patient data such as age, blood pressure, cholesterol, and ECG results—is collected and pre-processed to remove inconsistencies and normalize values. Machine learning techniques, particularly Support Vector Machine (SVM), are then applied to predict ischemic heart disease. SVM is effective in classifying complex medical data by identifying the optimal boundary between healthy and at-risk patients. Data from the UCI heart disease dataset was used to compare SVM's performance with other models like Neural Networks, Decision Trees, and Naive Bayes. The results show that SVM offers high accuracy and reliability, making it suitable for early heart disease detection and clinical decision support.

3.2 Limitations of Existing System

Lower Precision: The existing system may have lower precision and accuracy in predicting heart diseases compared to more advanced machine learning techniques.

Lack of Scalability: Traditional models may struggle to scale effectively with large datasets or incorporate new data sources and features.

3.3 Proposed System

The hybrid model combining Random Forest with a Linear Model leverages the strengths of both ensemble learning and linear regression/classification techniques to improve heart disease prediction accuracy. The process begins with data collection from medical datasets containing patient information such as age, gender, blood pressure, cholesterol levels, ECG results, and other clinical attributes. Data pre-processing is then performed to handle missing values, normalize data, and encode categorical variables. The dataset is split into training and testing sets. In the hybrid approach, the Random Forest algorithm is first applied to identify the most important features and reduce dimensionality by eliminating irrelevant or less significant attributes. These selected features are then passed to a Linear Model such as Logistic Regression to perform the final prediction. This combination allows the Random Forest to handle complex feature interactions and non-linear patterns, while the Linear Model offers interpretability and simplicity in the final classification. The model is evaluated using metrics such as accuracy, precision, recall, F1-score, and AUC-ROC to assess its effectiveness in predicting the likelihood of heart disease. This hybrid approach improves robustness, reduces over fitting, and enhances the model’s ability to generalize well to new patient data.

4. System Architecture

The system architecture for heart disease prediction using a hybrid Random Forest model is a layered design that processes medical data efficiently for accurate outcomes. It starts with data collection from clinical or IoT sources, followed by pre-processing to clean and normalize the data.

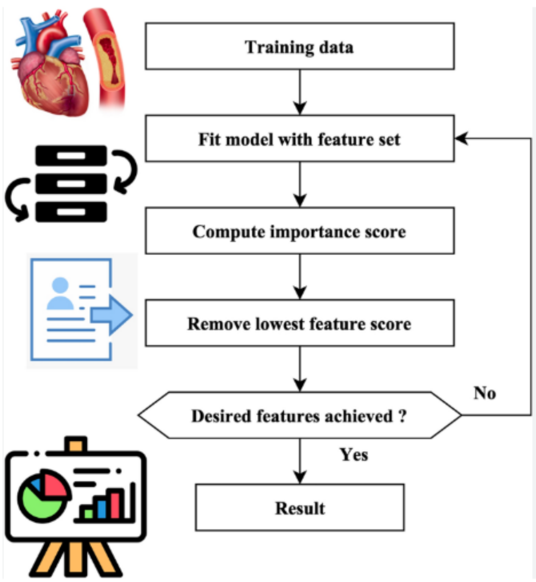

Fig 4: System Architecture

The Random Forest algorithm is used for feature selection, identifying key predictors. These features are passed to a hybrid prediction engine combining Random Forest and a Linear Model for robust and interpretable predictions. The system's performance is validated using metrics like accuracy and ROC-AUC, and results are delivered through a user-friendly interface for clinical use.

5. Modules

1) Data Collection: Assemble a comprehensive dataset, encompassing diverse health indicators such as age, blood sugar, and cholesterol, sourced from healthcare institutions, public repositories, and wearable's for a holistic representation.

2) Data Pre-processing: Cleanse and standardize the dataset, managing missing values and outliers, ensuring a robust foundation for subsequent analysis and modelling.

3) Feature Extraction and Selection: Feature extraction identifies important characteristics from the patient data that can be used for prediction. Feature selection determines which features are most relevant for the model, reducing dimensionality and enhancing model performance.

4) Machine Learning Model Training: This module involves training and fine-tuning machine learning models such as logistic regression, SVM, Multinomial Naive Bayes, and Decision Tree hybrid Random Forest model on historical patient data to learn patterns and relationships that can be used for prediction.

5) Model Evaluation Improvement: This module assesses the performance of the machine learning models using metrics like accuracy, precision, recall, and F1-score. It also facilitates model retraining with new data to adapt to changing health trends and improve predictive accuracy.

6. Result

The below pictures shows that , how the application measures whether the person has a heart disease or not based on some attributes like type of chest pain, blood pressure, cholesterol, blood sugar etc. Here, the output was that, person has a heart disease.

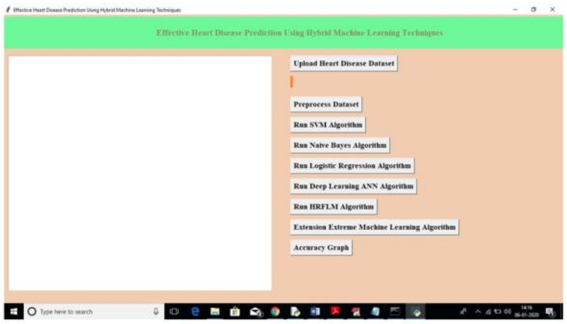

Fig 6.1: In above screen click on 'Upload Heart Disease Dataset' button to upload heart dataset

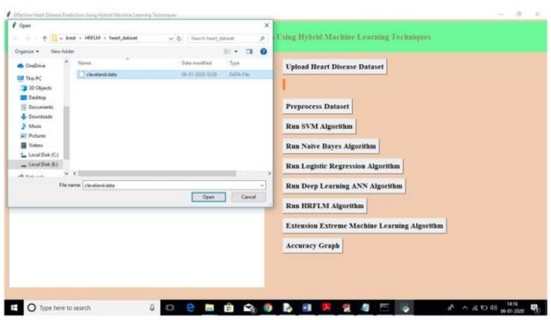

Fig 6.2: The next screen will appear when you upload the 'Cleveland. Data' dataset, which I am currently doing in the previous screen

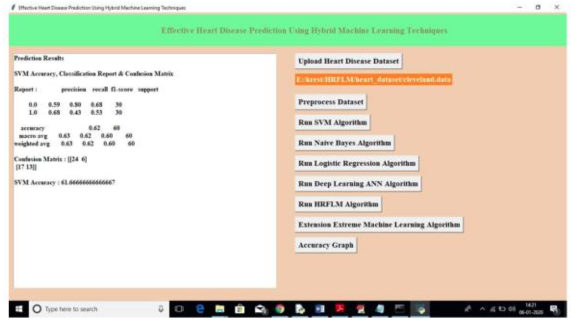

Fig 6.3: To see the accuracy of the 'Run Naïve Bayes Algorithm', which SVM achieved with 62% accuracy upon the previous screen, click upon that button.

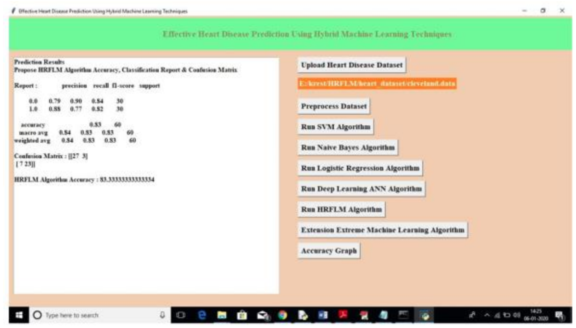

Fig 6.4: To see the EML extension's accuracy, select the "Extension Extreme Machine Learning Algorithm" button; the previous algorithm showed that HRFLM achieved 84% accuracy.

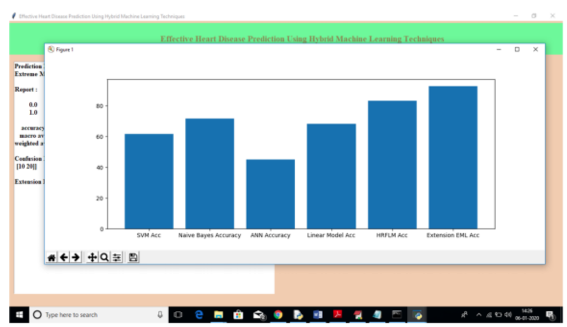

Fig 6.5: Accuracy of Heart disease Prediction

The x-axis in the following graph shows the names of the algorithms, while the y-axis shows their accuracy. The accuracy of the HRFLM plus extension algorithms was the highest of all the methods proposed.

7. Conclusion & Future Scope

Observing the handling of raw cardiac data may aid in lifesaving efforts and the long-term detection of earlier cardiovascular irregularities. In this study, artificial intelligence processes are implemented to handle raw data in order to provide a fresh and innovative solution to heart disease. In medicine, the anticipation of cardiac disease is both difficult and essential. No matter the cause, death rates may be significantly reduced with early diagnosis and prompt treatment. Focusing investigations on verifiable facts rather than speculative approaches and reenactments has been mainly emphasized throughout this study's subsequent expansion. The suggested hybrid HRFLM system is put into action, combining the features of a random forest area (RF) with an instant method (LM). Research has shown that HRFLM is a reliable predictor of coronary disease. With a different combination of AI methodologies to enhance estimation processes, that assessment's future direction should be doable. To get a better picture of the most important traits for increasing the likelihood of coronary disease, inventive ways of selecting them might be developed.

Future enhancements in heart disease prediction using Artificial Neural Networks (ANN) can focus on improving accuracy, adaptability, and real-time performance. Integrating deep learning architectures like CNNs or RNNs could help capture more complex patterns in patient data.

References

1. S. Abdullah and R. R. Rajalaxmi, ‘‘A data mining model for predicting the coronary heart disease using random forest classifier,’’ in Proc. Int. Conf. Recent Trends Comput. Methods, Commun. Controls, Apr. 2012, pp. 22–25.

2. A. H. Alkeshuosh, M. Z. Moghadam, I. Al Mansoori, and M. Abdar, ‘‘Using PSO algorithm for producing best rules in diagnosis of heart disease,’’ in Proc. Int. Conf. Comput. Appl. (ICCA), Sep. 2017, pp. 306– 311.

3. N. Al-milli, ‘‘Backpropogation neural network for prediction of heart disease,’’ J. Theor. Appl.Inf. Technol., vol. 56, no. 1, pp. 131–135, 2013.

4. C. A. Devi, S. P. Rajamhoana, K. Umamaheswari, R. Kiruba, K. Karunya, and R. Deepika, ‘‘Analysis of neural networks-based heart disease prediction system,’’ in

Proc. 11th Int. Conf. Hum. Syst. Interact. (HSI), Gdansk, Poland, Jul. 2018, pp. 233–239.

5. P. K. Anooj, ‘‘Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules,’’ J. King Saud Univ.-Comput. Inf. Sci., vol. 24, no. 1, pp. 27– 40, Jan. 2012. doi: 10.1016/j.jksuci.2011.09.002.

6. L. Baccour, ‘‘Amended fused TOPSIS-VIKOR for classification (ATOVIC) applied to some UCI data sets,’’ Expert Syst. Appl., vol. 99, pp. 115–125, Jun. 2018. doi: 10.1016/j.eswa.2018.01.025. 69

7. C.-A. Cheng and H.-W. Chiu, ‘‘An artificial neural network model for the evaluation of carotid artery stenting prognosis using a national-widedatabase,’’ in Proc. 39th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC), Jul. 2017, pp. 2566–25.

8. H. A. Esfahani and M. Ghazanfari, ‘‘Cardiovascular disease detection using a new ensemble classifier,’’ in Proc. IEEE 4th Int. Conf. Knowl.- Based Eng. Innov. (KBEI), Dec. 2017, pp. 1011–1014.

9. F. Dammak, L. Baccour, and A. M. Alimi, ‘‘The impact of criterion weights techniques in TOPSIS method of multi-criteria decision making in crisp and intuitionistic fuzzy domains,’’ in Proc. IEEE Int. Conf. Fuzzy Syst. (FUZZ-IEEE), vol. 9, Aug. 2015, pp. 1–8.

10. R. Das, I. Turkoglu, and A. Sengur, ‘‘Effective diagnosis of heart disease through neural networks ensembles,’’ Expert Syst. Appl., vol. 36, no. 4, pp. 7675–7680, May 2009. doi: 10.1016/j.eswa.2008.09.013.

11. M. Durairaj and V. Revathi, ‘‘Prediction of heart disease using back propagation MLP algorithm,’’ Int. J. Sci. Technol. Res., vol. 4, no. 8, pp. 235–239, 2015.

12. M. Gandhi and S. N. Singh, ‘‘Predictions in heart disease using techniques of data mining,’’ in Proc. Int. Conf. Futuristic Trends Comput. Anal. Knowl. Manage. (ABLAZE), Feb. 2015, pp. 520–525.

13. A. Gavhane, G. Kokkula, I. Pandya, and K. Devadkar, ‘‘Prediction of heart disease using machine learning,’’ in Proc. 2nd Int. Conf. Electron., Commun. Aerosp. Technol. (ICECA), Mar. 2018, pp. 1275–1278.

14. B. S. S. Rathnayakc and G. U. Ganegoda, ‘‘heart diseases prediction with data mining and neural network techniques,’’ in Proc. 3rd Int. Conf. Converg. Technol. (I2CT), Apr. 2018, pp. 1–6.