PREDICTION OF RISK IN CARDIOVASCULAR DISEASE USING MACHINE LEARNING ALGORITHMS

[1]Dr.M.Chandra Mohan Reddy [2]G.Deepika [3]K.Bhashitha Lakshmi [4]G.Anvitha [5]B.Haneesha [6]D.Chandana Sree

Department of ECE

Narayana Engineering College,Nellore, 524004, Andhra Pradesh, India

Abstract: *One of the hardest tasks in the medical field is predicting heart disease; thus the early detection of heart disease became a potential area for research to save the life of the patient. Due to poor prediction and delay in taking proper treatment at home, the number of cardiac arrest cases at home has shot up enormously in the pandemic period. Data in the health care industry operates on processing huge amount of data and for this purpose solution avail is machine learning. Data science processes this information to make intelligent decisions for health care, avoiding risk and alerting patients. We are using few attributes to predict the probability of heart disease using the comparative study of various machine learning classifiers based on dataset's. ML algorithms for heart disease prediction is an Support Vector Machine, Multi-Layer Perceptron, Logistic Regression and Random forest algorithm It also identifies the correlation between different attributes thus is able to use them effectively for predicting heart attack*

I. INTRODUCTION

However, cardiovascular diseases continue to be among the most significant causes of morbidity and mortality globally, impacting millions of people across various ages, genders, and regions. Moreover, adding to the strain on health care systems is the rising burden of risk factors like hypertension, diabetes, high cholesterol, smoking and sedentary lifestyles. This system executed several algorithms that analyzed key health indicators like blood pressure, cholesterol levels, smoking status, diabetic status, and occurrences of heart rate variations. This, when combined with other features, can lead to the recognition of subtle correlations that would be imperceptible to other evaluative methods. This approach also allows a single well-trained machine learning model to generalise well and adapt to changing trends in medical data, improving the quality of cardiovascular risk prediction over time.

II. FUNCTIONAL OVERVIEW

Considering the growing concern of heart diseases, the need of the hour is to have advanced predictive based systems that can identify risk level based on multiple clinical & lifestyle parameters. Moreover, the traditional diagnostic approaches commonly rely on subjective evaluations and broad criteria that may not always accurately reflect an individual's health status complexities. With the emergence of AI and ML, a more data-driven and precise modeling technique on risk prediction has opened. Machine learning can help make a difference by studying patient-specific parameters to extract hidden patterns among the data to get personalized insight for each individual, ensuring they receive timely medical care and preventive measures. More importantly though it improves survival and decreases the total healthcare burden.

By training predictive models on huge data sets, researchers are able to pull from hundreds or thousands of years' worth of historical medical records to find connections between particular risk factors and subsequent health conditions. Pet and state will produce the ranking of all people comparing to the parameters of each person such as cholesterol, blood pressure, smoking, diabetes, heart rate variations, etc. In contrast to traditional risk calculators that use static scoring systems, machine learning algorithms dynamically learn and evolve, enabling enhanced predictive performance as they are exposed to more data. They are especially helpful in the identification of high-risk individuals who may remain undiagnosed until symptoms develop.

Such a predictive system is implemented by multiple machine learning algorithms to enhance its validity and accuracy. MD, Random Forest, Multi-layer Perceptron, Logistic Regression and Support Vector Machine models can be used to classify patients into different risk categories. These algorithms have their advantages for feature selection, for classification, for generalization to new data, etc. These techniques enable reliable predictions with high sensitivity and specificity, enabling healthcare providers to make informed decisions about patient management. The effectiveness of these models is evaluated using performance metric such as accuracy, precision, recall, and F1score to check if the predictions are in the same line as real-world clinical outcomes.

III. EXISTING WORK

In traditional health care systems, cardiovascular risk assessment is largely conducted through established clinical guidelines and physician expert opinion. Even sophisticated simulation models used by medical professionals, such as Framingham Risk Score and the Risk Estimator, are based on population averages in estimating an individual’s heart risk. These models take into account age, cholesterol levels, blood pressure, smoking habits, and diabetes status to come up with a risk score. Hospital-based diagnostic techniques also help determine cardiovascular health, including electrocardiograms, stress tests, and blood tests. Such evaluations assist medical professionals in determining correct choices in patient treatment and lifestyle changes.

Hospitals and healthcare centers perform laboratory tests and imaging techniques to identify preexisting conditions of cardiovascular system. Examples include blood tests to assess metabolic health levels of cholesterol, triglycerides, or glucose as well as advanced imaging technologies like echocardiograms or coronary angiograms to diagnose blockages or abnormalities in blood flowing through arteries. These methods have been in broad use for generations and act as pivotal mechanisms for detecting higher-risk individuals. Doctors combine these diagnostic tools and a patient’s medical history to formulate treatment plans, which may involve taking medication

Disadvantages:

1. Old Approach: This approach involves manual risk assessment, expert opinion, and subjective assessments. This also raises the likelihood of human error and inconsistency in risk assessment.

2.Generalised Risk models: Many models are generalised and employ generalised statistical approaches rather than making personalized predictions. These approaches do not consider individual differences in lifestyle, genetics and health history.

3.Late diagnosis: Traditional diagnostic methods reveal cardiovascular risks at more advanced stages. This lag diminishes the prospects for early intervention and increases the risk of complications.
Yet little data is used in diagnosis, as traditional systems do not recognize patterns based on large-scale patient data. Due to the lack of data-driven insights, as a result, risk cannot always be predicted rapidly and healthcare is not recommended in a personalized Way.

IV. PROPOSEDWORK

The data is preprocessed before training to improve its quality and make it more suitable for training machine learning models. Missing values are handled via imputation techniques, while outliers are detected and removed to avoid biased predictions. Methods such as normalization and standardization are used for feature scaling so that different health indicators can have uniformity which will result in better models.

We explore different classification algorithms, Random Forest, Logistic Regression and Support Vector Machine. Random forests and boosted models are selected for their performance with structured healthcare data and robustness when it comes to making predictions. We perform the comparative analysis of various models to find the most optimal algorithm based on precision, recall and f1-score.

The preprocessed dataset is split into training and testing subsets used to train and evaluate models. You have to have information about the risk factors and you train the machine learning with this labeled data on the risk factors. Model hyperparameters tuning is done to optimize the models and assure good generalization to unseen patient data

Advantages:

1. Timely Detection of Risk: Cardiovascular risk is detected at an early stage and this helps the individuals to take preventive measures before they develop severe complications. With early interventions, this proactive approach can have a major impact on the potential risk of life-threatening health crises.

2. Data Is Key: The system enables data-based risk assessments, with machine learning algorithms processing medical data instead of subjective physician evaluations. When doctors use this information, they can make better treatment decisions and ultimately improve their patients' results.

3. Machine Learning: Data compared to October 2023 train Machine Learning and more accurate compared to statistics. Hence, in simple terms, optimized algorithms reduce errors and improve the classification

4. Easy to Use and Accessible: A web-based interface ensures that ease of use, making risk assessment widely accessible to patients and medical practitioners. Patients can assess their levels of risk without having to go to the hospital repeatedly, thus saving them significant time and effort.

5. Future Implications and Scalability: The system can be developed further to integrate real-time tracking via wearable technologies and mobile applications to monitor patients continuously.

Design:

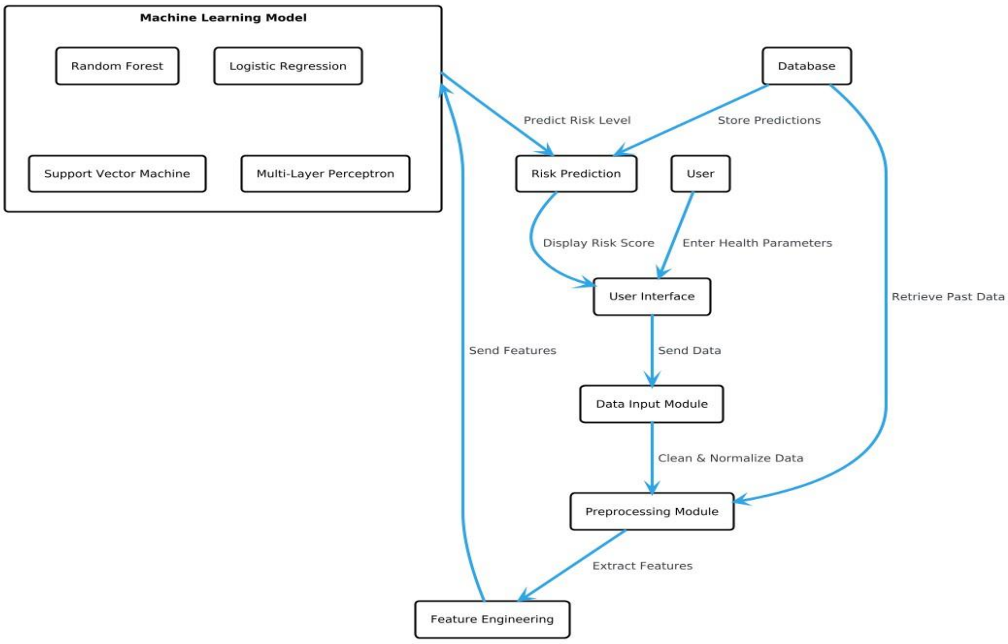

Fig 1: Data Flow Diagram

Data flow diagram depicting the sequential process of cardiovascular risk prediction and interrelation between the modules. It starts with the user entering the health parameters on the user interface that is then passed on to the average data input module. This collected data is sent to the preprocessing module, and here the data is cleaned and normalized. After completing preprocessing, feature engineering is the next step that retrieves the most relevant attributes and shape up the perfect data for ML models. The same mechanism continues with these extracted and refined features being fed into the risk prediction module, which applies machine learning algorithms— Random Forest, Logistic Regression, Support Vector Machine, and Multi-Layer Perceptron—to predict the user cardiovascular risk level.

The risk score once generated is shown on the UI, and therefore they can see their results in real-time. It is critical part of the system, since it is responsible for storing the past predictions or filtering the historic data in order to make further analysis. The risk prediction module also performs updates to the database with newly made predictions, allowing for continuous training and evaluation of model accuracy. The cross coupling is what makes sure that data flows accordingly-well between components, hence making the system more robust and accurate. By means of this well-integrated design users are given accurate cardiovascular risk assessments which help them make smart choices regarding their health.

V. EXPERIMENTAL RESULT

User Authentication Interface

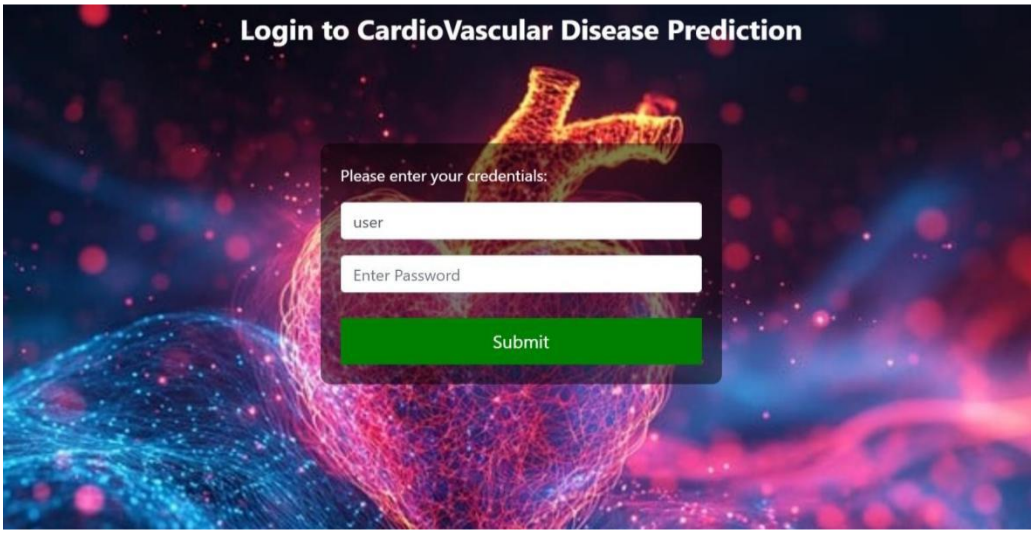

Fig 2: Login Page

Authorization Login Interface The login interface is the first point for users to access the Cardiovascular disease prediction system with secure authorization. It is capable of user authentication which means that it prompts users to enter their credentials and thereby only authentic users may use the system. The background has a rendered heart subtly emphasizing the medical theme The interface itself is a very basic and simple design with a username and password input field and submission button. The ECG waves in the background of the heart illustration help to nod toward a cardiovascular focused system without being too unprofessional. That's why the usage of the login page is so that noone in particular must have the data without any specific reason to determine the risk of the person to have a heart attack.

Welcome Screen and Navigation

Fig 3: Dashboard

Upon successful authentication, users will be redirected to the dashboard, which serves as the central hub of the system. It provides a bunch of options like Prediction, Classification Report, Accuracy, Logout, etc., so that you can use the application with ease. A background image with hands holding a heart (with ECG signal at it), indicate-working the system purpose which is cardiovascular health. Buttons highlighted in blue are clearer and easier to reach.

A little tooltip explaining each section would be a very handy way to improve user-experience for first time users. One of the most important features would be a user profile section that would allow you customize your profile based on your language preferences, theme selection. Line graphs can be generated from this data, creating an interactive experience and adding a competitive aspect, as it showcases all active users for a particular week or month.

Data Entry Interface

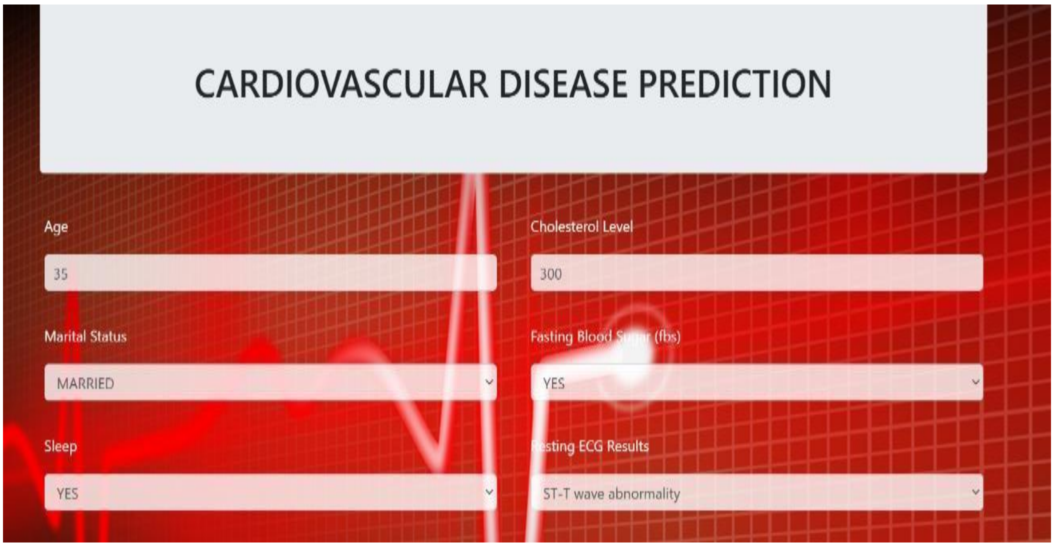

Fig 4: Data Input Interface

The input form data is suposed to an important component, which directly affects the accuracy and reliability of prediction. A good form is easy to use, providing a guided experience for users to enter structured data, with minimal errors. For example, validation checks can prevent submitting incorrect values (i.e., ensuring restrictions on normal ranges of medically relevant inputs). Furthermore, use of dropdown menus or radio buttons for categorical variables (smoking status, marital status) is a great way to speed up data entry and make it feel more intuitive.

Accuracy of Algorithms

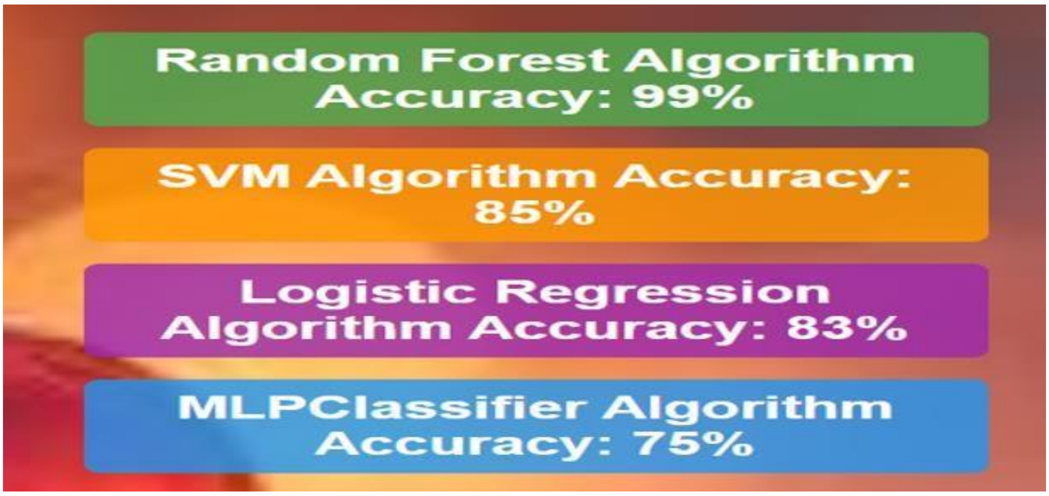

Fig 5: Model Accuracy Comparision

The comparison of the accuracy of various machine learning models holds significant importance in evaluating their effectiveness and dependability. Having performance reported side by side like this gives a good indication of which one to use for cardiovascular risk. While accuracy is an important metric, other metrics like sensitivity (true positive rate) and specificity (true negative rate) should also be a factor in evaluating the model's overall usefulness.

Bar charts or line graphs are helpful to visualize model performance to users for better interpretation of results. An interactive feature allowing the user to toggle between datasets or test cases for deeper insights into the strengths and weaknesses of each model would be a welcome addition. Also, metrics like training time and computational efficiency would help determine how feasible it is to deploy a given model iteratively in the real world.

Prediction of Risk

Fig 6: High Risk Prediction Result

This is the key result of the cardiovascular disease prediction system. It means that the user is at a risk of heart-related health issue as per the input data and analysis. This prediction is made by some more machine learning models that analyse number of clinical and lifestyle parameters. By communicating simply with the user in the result box, the user can instantly be made aware of a possible health risk. The image of a human heart and blood vessels in the background reminds us how important cardiovascular health is.

Such a prediction outcome is important in an aspect of preventive environment as it acts as a warning sign for high risk individuals. Heart attack gets a way of telling patients to seek treatment, improve lifestyle and go through the required treatment plan to avoid a heart attack. This assessment is thorough, accounting for not only cholesterol levels but also blood pressure, smoking habits and pre-existing conditions (diabetes). The system intends to encourage proactive health management by displaying the findings in a user-friendly schema.

The performance of the machine learning model, has to be such that it iteratively grows and improves its accuracy, and yes, the reliability is also based on it. The classifiers like Random Forest, Multi-layer Perceptron, and Support Vector Machine at the back-end make sure that predictions are data-driven and statistically sound. And not just an output in all sense as its what got printed after training and validating the model. In the mean time better data collection, as well as optimization of the models, could, at least in theory, make such predictions more accurate and thus even more valuable for medical decision -making process.

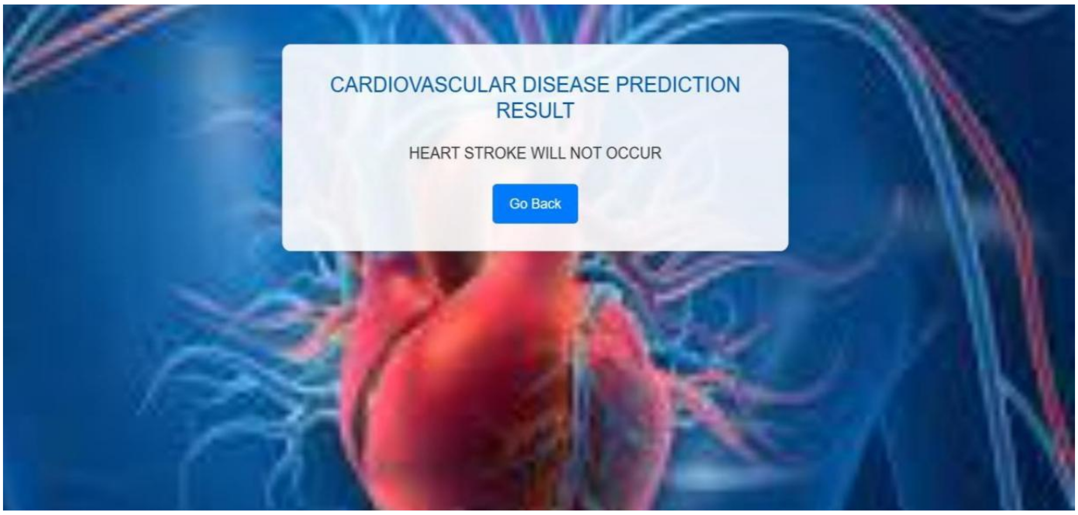

Fig 7: Low Risk Prediction Result

The example represents a good result for the cardiovascular disease prediction system whereby the subject is not at its risk of having a heart stroke. This is reassuring for the user and a reminder of the importance of a healthy lifestyle. Having analyzed numerous parameters, according to the machine learning algorithm, the individual is not subject to sufficient risk factors that could result in an emergency related to the heart. The order and aesthetics of the layout are such that the message is communicated effectively

VI. CONCLUSION

This system can successfully indicate the potential of machine learning in estimating the risk of heart diseases. Choosing the algorithms such as Random Forest Classifier, Multi-layer Perceptron, Logistic Regression, and Support Vector Machine, it offers an accurate and rapid approach to predicting an individual's probability of heart disease. This is further complemented by a user-friendly interface, enabling medical professionals and individuals alike to receive timely and accurate health assessments.

The implementation of this can yield one of the most important results, early detection and prevention of the disease. By taking into account important clinical and lifestyle variables like blood pressure, cholesterol, smoking status, diabetes status and exercise, the system delivers a complete risk assessment. This high level of reliability is achieved through the utilization of classic metrics such as precision, recall and F1-score analysis,

as well as confusion matrix analysis, all of which help to significantly reduce the misclassification risk. Such predictive elements can promote healthy living and seeking medical help when appropriate, thus reducing severe complications.

While the system has worked well enough, there remain ways to improve it further. Incorporating real-time information from wearable health devices would enable constant monitoring and more accurately predict risk. In addition, making the database more varied with larger patient samples would also contribute to greater generalization. Implementing deep learning algorithms to improve the accuracy of prediction, and working with medical institutions to validate the models, could also help increase the accuracy and trust of the system's predictions.

To conclude, this implementation is a remarkable milestone in AI powered health care and a pertinent resource in risk stratification. The impending potential of machine learning to enhance the medical landscape, especially in disease prevention by integrating real-time health monitoring into the research space, cannot be overstated.

VII. REFERENCES

[1] **https://professional.heart.org/en/guidelines-and-statements/prevent-calculator**

[2] **https://tools.acc.org/ascvd-risk-estimator-plus/**

[3] **https://www.nhs.uk/health-assessment-tools/calculate-your-heart-age**

[4] **https://www.paho.org/en/hearts-americas/cardiovascular-risk-calculator-app**

[5] **https://my.clevelandclinic.org/health/articles/17085-heart-risk-factor-calculators**