AI-Driven Heart Disease Prediction using Machine Learning

Mrs.Subalakshmi R[1], Senthil vel S[2], Priyanka S[3], Dharun D[4], Vignesh N[5]

[1]Assistant Professor Computer Science Engineering, Dhanalakshmi Srinivasan college of Engineering
[2]UG Scholar, Computer Science Engineering, Dhanalakshmi Srinivasan college of Engineering
[3]UG Scholar, Computer Science Engineering, Dhanalakshmi Srinivasan college of Engineering
[4]UG Scholar, Computer Science Engineering, Dhanalakshmi Srinivasan college of Engineering
[5]UG Scholar, Computer Science Engineering, Dhanalakshmi Srinivasan college of Engineering

***Abstract*— Heart-disease (HD) is one of the most common diseases nowadays, and for people who provide health care, it is very necessary to work with them to take care of their patients' health and save their life. In this paper, different classifiers were analyzed by performance comparison to classify the Heart Disease dataset to classify it correctly and or to Predict Heart Disease cases with minimal attributes.**

Large amounts of data that contain some secret information were collected by the healthcare industries. This data collection is useful for making effective decisions. Some advanced data mining techniques are used to make proper results and making effective decisions on data. In this case, a Heart Disease Prediction System (HDPS) is developed using Logistic Regression, K Nearest Neighbor, Decision Tree, Random Forest Classifier, and Support Vector Machine algorithms to predict the heart disease risk level. The results reveal that the Random Forest Classifier and Support Vector Machine obtained the highest accuracy of 90.32%, whereas 87.09%, 70.96%, and 83.87% accuracy scores are obtained by logistic regression, KNN classifier, and decision tree respectively.

***Keywords*—**
Heart Disease Prediction, Machine Learning, Web-Based System, Random Forest, Support Vector Machine, Logistic Regression, K-Nearest Neighbor, Decision Tree, Doctor Appointment System, Healthcare Technology, Medical Data Analysis.

I. INTRODUCTION

With the rapid advancement of information technology, significant progress has been made in various scientific fields, particularly in medical science. Artificial Intelligence (AI) and Machine Learning (ML) have played a transformative role in healthcare, making it possible to predict diseases early and assist doctors in accurate diagnoses.

Cardiovascular diseases (CVDs) remain one of the leading causes of death worldwide, making early detection crucial in saving lives. Coronary Artery Disease (CAD), a major type of heart disease, affects a significant portion of the population without any prior symptoms, often leading to severe heart attacks and sudden fatalities.

Heart disease is a broad category of conditions affecting the heart and blood vessels, including coronary heart disease, hypertension, peripheral artery disease, rheumatic heart disease, congenital heart disease, and heart failure. Effective treatment of heart diseases requires timely intervention by cardiologists, thoracic surgeons, vascular surgeons, and interventional radiologists. The heart plays a vital role in pumping blood to all organs, and any inefficiency in its function can result in serious consequences such as organ failure and death. Traditional diagnostic methods often fail to detect heart diseases at an early stage, leading to late-stage interventions that may not always be successful. Therefore, there is a pressing need to develop a decision-support system for detecting heart disease at an early stage, thereby improving patient outcomes and reducing medical costs.

To address these challenges, a web-based heart disease prediction system has been developed using Machine Learning algorithms trained on the Kaggle UCI Heart Disease dataset. This system provides an interactive platform where users can check their heart health status and book an appointment with a doctor if needed. The website consists of two main components: the Admin Login and the User Login, ensuring a comprehensive healthcare solution. The Admin panel allows administrators to add doctors' details to the system, upload and manage datasets for machine learning predictions, and view user feedback to enhance the system. The User panel enables users to log in, enter their health parameters for heart disease prediction, and if diagnosed with heart disease, proceed to book an appointment

with a doctor. Additionally, users can submit feedback about their experience, which is visible to the admin for continuous improvement of the system.

The heart disease prediction model is built using various ML algorithms, including Random Forest Classifier, Support Vector Machine (SVM), Logistic Regression, K-Nearest Neighbor (KNN), and Decision Tree Classifier. Among these, the Random Forest Classifier and Support Vector Machine have achieved the highest accuracy in detecting heart disease, ensuring reliable predictions for users. By integrating machine learning with a user-friendly web-based platform, this system offers a comprehensive solution for early heart disease detection and medical consultation. The automation of heart disease diagnosis not only supports medical professionals but also empowers users to monitor their heart health conveniently. The platform enhances early diagnosis, reduces medical costs, and improves patient outcomes by enabling timely medical intervention. This system represents a significant step toward digitalized healthcare, offering an efficient and reliable method for heart disease prediction and management.

II. LITARATURE REVIEW

Vijeta Sharma proposed the paper titled "Heart Disease Prediction using Machine Learning Techniques". In this paper Machine Learning algorithms such as Random Forest, Support Vector Machine (SVM), Naive Bayes and Decision tree have been used for the development of model. Result shows that compared to other ML techniques, Random Forest gives more accuracy in less time for the prediction. By the end of the implementation part, we have discovered that Random Forest is giving the maximum accuracy level in our dataset which is 99% and Decision Tree is playing out the least with an accuracy level of 85%.

Noura Ajam recommended artificial neural network for heart disease diagnosis. Based on their ability, Feed forward Back propagation learning algorithms have used to test the model. By considering appropriate function, classification accuracy reached to 88% and 20 neurons in hidden layer. ANN shows result significantly for heart disease prediction.

Mr. Santhana Krishnan. J proposed the paper titled "Heart Disease Prediction using Machine Learning Algorithms". In this paper, two supervised data mining algorithm was applied on the dataset to predict the possibilities of having heart disease of a patient, were analysed with classification model namely Naive Bayes Classifier and Decision tree classification. These two algorithms are applied to the same dataset in order to analyse the best algorithm in terms of accuracy. The Decision tree model has predicted the heart disease patient with an accuracy level of 91% and Naive Bayes classifier has predicted heart disease patient with an accuracy level of 87%.

K. Polaraju et al, proposed Prediction of Heart Disease using Multiple Regression Model and it proves that Multiple Linear Regression is appropriate for predicting heart disease chance. The work is performed using training data set consists of 3000 instances with 13 different attributes which has mentioned earlier. The data set is divided into two parts that is 70% of the data are used for training and 30% used for testing. Based on the results, it is clear that the classification accuracy of Regression algorithm is better compared to other algorithms.

Prof. Kailas Devadkar (PhD) proposed the paper titled "Prediction of Heart Disease Using Machine Learning". In this paper they have used the neural network algorithm multi-layer perceptron (MLP) to train and test the dataset because of its efficiency and accuracy.

Sairabi H. Mujawar et al, [24] used k-means and naive bayes to predict heart disease. This paper is to build the system using historical heart database that gives diagnosis. 13 attributes have considered for building the system. To extract knowledge from database, data mining techniques such as clustering, classification methods can be used. 13 attributes with total of 300 records were used from the Cleveland Heart Database. This model is to predict whether the patient have heart disease or not based on the values of 13 attributes.

"A machine learning-based framework for personalized cardiovascular risk prediction using electronic health records" by Kwon et al. (2021). This study proposes a machine learning-based framework for personalized cardiovascular risk prediction using electronic health records, achieving high accuracy in predicting the risk of cardiovascular disease.

"A comparison of machine learning techniques for cardiovascular disease prediction" by Sathyanarayana et al. (2020). This study compares different machine learning algorithms for predicting cardiovascular disease, finding that the XG-Boost algorithm performed best.

"Risk prediction for cardiovascular disease using deep neural networks" by Lee et al. (2020). This study explores the use of deep neural networks for predicting cardiovascular disease, achieving high accuracy in risk prediction compared to traditional methods.

"Comparison of deep learning architectures for prediction of cardiovascular disease risk" by Attia et al. (2019). This study compares different deep learning architectures for predicting cardiovascular disease risk, finding that a convolutional neural network achieved the highest accuracy.

"Predicting cardiovascular risk using machine learning techniques: A systematic review" by Alizadeh et al. (2019). This systematic review evaluates the current state of machine learning-based cardiovascular risk prediction, highlighting the potential of these techniques for improving risk prediction accuracy.

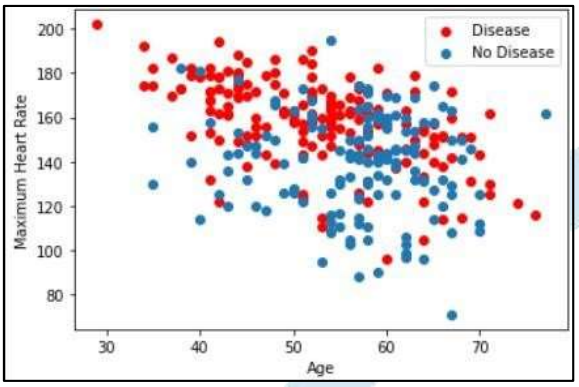
III. DATA SET INFORMATION

The name of the dataset is heart.csv. There are 1000 instances in this dataset, where the cases are either people having heart disease or they are healthy. Among 1000, 392 (39.2%) cases are people with heart disease and 608 (60.8%) are people without heart disease. The number of attributes is 16. There are no missing values in the data set nor any null values.

Features include age, sex, chest-pain type, rest BP, cholesterol, blood sugar level, ECG result, maximum heart rate achieved, exercise-induced angina, ST depression, the slope of peak

exercise ST segment, number of major vessels, and defect in heart as of 3-normal, 6-fixed defect and 7- reversible defect. Bar graph (Fig.1) showing the positive and negative cases (1=positive, 0=negative) Scatter plot (Fig.2) showing the positive and negative cases depending on age.

Fig.1 Positive and negative cases



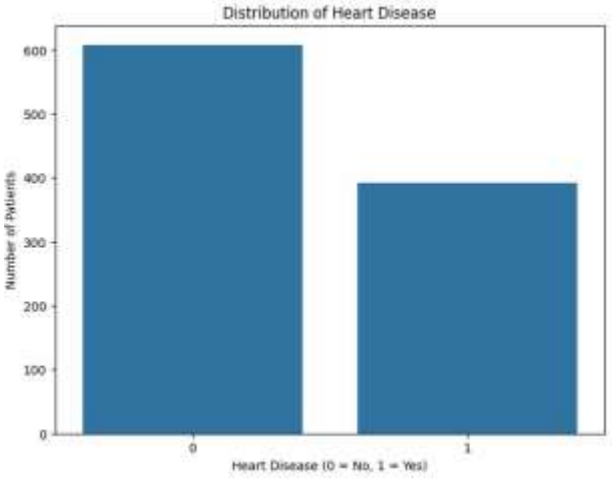


Fig.2 Positive and negative cases depending on age

	Age	Gender	Cholesterol	Blood Pressure	Heart Rate	Smoking
0	75	Female	228	119	66	Current
1	48	Male	204	165	62	Current
2	53	Male	234	91	67	Never
3	69	Female	192	90	72	Current
4	82	Female	172	163	83	Never
...	...	...	...	...	...	...
995	56	Female	269	111	86	Never
996	78	Female	134	145	70	Never
997	79	Male	151	179	81	Never
998	60	Female	326	151	68	Former
999	53	Male	226	116	82	Current

Fig.3 Data set description

IV. METHODOLOGY

A. *Data set information*

The main objective of this research is to develop a heart disease prediction system. The system can discover and extract hidden knowledge associated with diseases from a historical heart data set heart disease prediction system aims to exploit data mining techniques on medical data set to assist in the prediction of the heart diseases.

B. *Training and testing*

The training phase extracts the features (independent variables) from the dataset and the testing phase (containing dependent variables) is used to determine how the appropriate model behaves for prediction. We have divided the dataset into two sections. These are the training and testing phases. We have split the dataset into 90% training and 10% testing phase. And we have taken the random state as 1. For initializing the fixed internal random number generator, we use the random state parameter which will decide the splitting of data into train and test indices. Setting a random state will guarantee a fixed value that the same sequence of random numbers will be generated each time the code is being run. Setting random state, a fixed value will guarantee that the same sequence of random numbers is generated each time we run the code. Then we scaled the data using Standard scattered and fitted the training and testing data using 'fit. transform'.

```
from sklearn.model_selection import train_test_split
X = dataset.iloc[:, :-1].values
Y = dataset.iloc[:, -1].values
X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size= 0.1, random_state = 1)  #90% training, 10% testing
```

C. *Classsification used*

- Logistic regression

Logistic Regression is an analytical modeling technique. It is used for analyzing a dataset in which there are one or more independent variables that decide a result. Logistic Regression was imported with a random state of 0. And then the training model was fitted. The testing accuracy was 87.09%

- KNN Classifier

K-nearest neighbor algorithm is utilized for grouping and used in pattern recognition. It is widely used in predictive analysis. On the arrival of new data, the K-NN algorithm [8] identifies existing data points that are nearest to it. From 'sklearn.neighbors', 'KNeighbors Classifier' was imported with n_neighbors = 1. Then the training model was fitted. The testing accuracy was 70.96%

- Support vector machine

Support Vector Machine or SVM is one of the popular Supervised Learning algorithms in machine learning. The benefits of the SVM algorithm is that it creates the best suitable line or decision boundary that can separate a n-dimensional space into classes so that we can easily verify and put the new added data points in the correct category in the future. From 'sklearn', 'svm' was imported and we kept the kernel as linear and gamma as auto and C = 2.

And the training model was fitted. The testing accuracy was 90.32%.

- Random forest

  Random forest classifier is a powerful supervised classification tool. RF generates a forest of classification trees from a given dataset, rather than a single classification tree. Each of these trees produces a classification for a given set of attributes. From 'sklearn.ensemble', 'Random Forest Classifier' was imported. The n_estimators is kept at 10 and random state at 0. Then the training model was fitted. The testing accuracy was 90.32%.

- Decision Tree

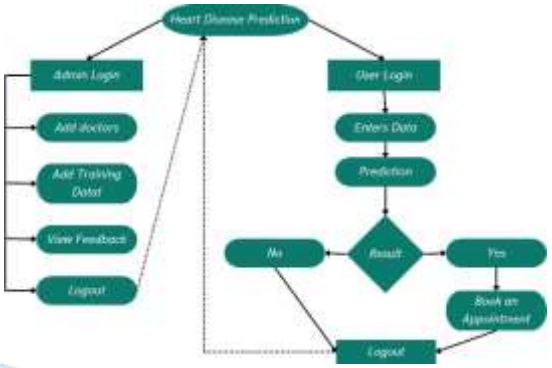
  The testing accuracy was 90.32%. A Decision tree is a tree shape-like diagram, where the internal nodes represent a test on an attribute, each branch denotes the outcome of the test, each leaf node denotes a class label. Decision Tree was imported where the random state was kept as 0 and then the training model was fitted. The testing accuracy was 83.87%. 6. Results Amongst all classification techniques, testing accuracy was best in the case of the random forest and SVM approach with an accuracy of 90.32%.

Table. 1 Comparison of performances difference classifier

Sl no	Algorithm	Testing accuracy
1	Logistic regression	87.09%
2	K nearest neighbour	70.96%
3	Random forest classifier	90.32%
4	Support vector machine	90.32%
5	Decision tree	83.87%

D. Web Development Process

The heart disease prediction website is designed to provide a user-friendly and efficient platform for both users and administrators. It integrates machine learning models with an interactive web interface, ensuring accurate predictions and streamlined doctor appointment booking. The web development process consists of several key components, including frontend development, backend development, database management, and API integration.



Frontend Development

The frontend of the website is built using **HTML, CSS, and JavaScript**, ensuring an intuitive and responsive user interface. The interface consists of two primary login sections: **Admin Login** and **User Login**. The **User Login** section allows individuals to enter their health parameters and receive a heart disease prediction. If a user is diagnosed with heart disease, they are given the option to **book an appointment** with a listed doctor. The **Admin Login** section provides a dashboard where administrators can **add training data**, **manage doctor information**, and **view user feedback** to improve the system's accuracy and effectiveness.

To enhance the user experience, frameworks such as **Bootstrap** are utilized to create a visually appealing and responsive design. **AJAX** is used to enable smooth interactions without the need for full page reloads, improving performance and usability.

Backend Development

The backend is developed using **Flask**, a Python-based web framework that connects the machine learning model to the web interface. Flask handles **user authentication**, **data processing**, and **prediction requests**, ensuring a seamless user experience. The backend consists of various APIs that allow the web application to interact with the heart disease prediction model, process user inputs, and return results.

For authentication, **Firebase Authentication** is used to securely manage user logins and admin access. Flask's built-in **JWT (JSON Web Token)** authentication ensures secure user sessions and access control.

Database Management

The website utilizes **Firebase Realtime Database** to store and manage data dynamically. The database structure includes:

- **User Data:** Stores registered users' information, login credentials, and prediction history.
- **Admin Data:** Stores admin credentials and permissions.
- **Doctor Information:** Contains details of doctors available for appointments.
- **Prediction Records:** Logs user predictions along with timestamps.
- **Feedback Data:** Stores user feedback to enhance the platform.

Machine Learning Model Integration

The **heart disease prediction model** is integrated into the backend using **Flask APIs**. The model, trained on the **Kaggle UCI Heart Disease Dataset**, processes user input in real-time and returns a prediction result. The backend API sends this result to the frontend, where it is displayed to the user. If a positive prediction is detected, the system automatically suggests booking an appointment with a listed doctor.

Doctor Appointment Booking System

The appointment booking system is an essential feature of the website. If a user is diagnosed with heart disease, they can schedule an appointment directly through the platform. The system fetches available doctors from the database, allowing users to select a doctor based on specialization and availability. Once an appointment is booked, confirmation details are sent to both the user and the doctor via **email notifications**.

Deployment and Hosting

The website is hosted using **Firebase Hosting** for the frontend and **Flask on a cloud server (Heroku or AWS)** for the backend. This ensures scalability and reliability. The system is optimized for **fast loading times**, **secure data transmission**, and **high availability**, making it accessible to users worldwide.

The heart disease prediction website integrates machine learning with an interactive web platform to provide an accessible, accurate, and efficient solution for early heart disease detection. With **user and admin authentication**, **real-time predictions**, **doctor appointment booking**, and **database management**, the platform aims to enhance healthcare accessibility and improve patient outcomes. By leveraging advanced web technologies, this system ensures seamless user interaction, data security, and a responsive interface, making it a valuable tool for both patients and healthcare providers.

V. RESULT

Amongst all classification techniques [Table.1], testing accuracy was best in the case of the random forest and SVM approach with an accuracy of 90.32%.

VI. CONCLUSION

Through this research, we have developed a comprehensive heart disease prediction system that integrates machine learning with a web-based platform to enhance early detection and accessibility to medical consultations. The system predicts whether an individual is at risk of heart disease based on various health attributes and provides an intuitive interface for users and administrators to manage predictions and doctor appointments.

The prediction model processes user inputs and outputs a result indicating the likelihood of heart disease, displaying either "Yes" or "No." If a user is identified as at risk, the system recommends consulting a cardiologist for further evaluation. The dataset used for model training and evaluation consists of records from the Kaggle UCI Heart Disease dataset, divided into training and testing subsets. We implemented multiple machine learning algorithms to assess accuracy, ensuring the most reliable prediction model is utilized within the application.

The web platform includes two primary user roles: **Admin** and **User.** In the **Admin Login**, administrators have the capability to add and update training data, manage doctor listings, and view user feedback, ensuring the platform is continually optimized. In the **User Login**, individuals can enter their health parameters and receive a heart disease prediction. If a positive result is detected, the system prompts the user to book an appointment with a listed doctor directly through the platform. This integration enhances accessibility to medical consultations and provides an efficient solution for early intervention.

The application is built using Flask for the backend, handling user authentication, data processing, and API interactions with the machine learning model. Firebase Authentication secures user credentials, and the database is managed using Firebase Realtime Database, storing user records, doctor information, and prediction history. The frontend, developed with HTML, CSS, and JavaScript, ensures a seamless user experience, incorporating responsive design elements with Bootstrap and dynamic content rendering using AJAX.

The deployment of the system utilizes Firebase Hosting for the frontend and Flask-based cloud hosting (Heroku or AWS) for the backend, ensuring a scalable and reliable infrastructure. The doctor appointment booking feature further strengthens the usability of the platform by directly connecting at-risk users with medical professionals, reducing delays in diagnosis and treatment. By providing real-time predictions, secure authentication, and a structured database, this platform serves as an essential tool in heart disease prevention and management, ultimately aiming to reduce mortality rates and improve healthcare accessibility.

VII. FUTURE SCOPE

Enhanced Machine Learning Model: The accuracy of heart disease prediction can be further improved by incorporating advanced deep learning techniques such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) to analyze ECG signals and other medical imaging data.

Integration with Wearable Devices: Future iterations of this system can integrate with wearable health monitoring devices such as smartwatches and fitness bands to collect real-time heart rate, blood pressure, and ECG data for continuous health monitoring.

Mobile Application Development: A dedicated mobile application can be developed to provide users with on-the-go access to heart disease predictions and doctor appointment scheduling, improving usability and accessibility.

Voice-Enabled and Chatbot Support: Implementing AI-driven voice assistants or chatbots can enhance user interaction by providing instant responses to health queries and guiding users through the prediction process.

Multi-Disease Prediction System: The platform can be extended to support the prediction of other chronic diseases such as diabetes, hypertension, and stroke, offering a comprehensive health monitoring system.

Multi-Language and Accessibility Features: To reach a wider audience, the platform can be enhanced with multi-language support, voice-to-text input, and accessibility features for visually impaired users.

VIII. REFERENCE

[1] V. Sharma, S. Yadav and M. Gupta, "Heart Disease Prediction using Machine Learning Techniques," 2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), Greater Noida, India, 2020, pp. 177-181, doi: 10.1109/ICACCCN51052.2020.9362842.

[2] M. S. Raja, M. Anurag, C. P. Reddy and N. R. Sirisala, "Machine Learning Based Heart Disease Prediction System," 2021 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, 2021, pp. 1-5, doi: 10.1109/ICCCI50826.2021.9402653.

[3] S. K. J. and G. S., "Prediction of Heart Disease Using Machine Learning Algorithms," 2019 1st International Conference on Innovations in Information and Communication Technology (ICIICT), Chennai, India, 2019, pp. 1-5, doi: 10.1109/ICIICT1.2019.8741465.

[4] R. Katarya and P. Srinivas, "Predicting Heart Disease at Early Stages using Machine Learning: A Survey," 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2020, pp. 302-305, doi: 10.1109/ICESC48915.2020.9155586.

[5] A. Gavhane, G. Kokkula, I. Pandya and K. Devadkar, "Prediction of Heart Disease Using Machine Learning," 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, 2018, pp. 1275-1278, doi: 10.1109/ICECA.2018.8474922.

[6] A. U. Haq, J. Li, M. H. Memon, M. Hunain Memon, J. Khan and S. M. Marium, "Heart Disease Prediction System Using Model Of Machine Learning and Sequential Backward Selection Algorithm for Features Selection," 2019 IEEE 5th International Conference for Convergence in Technology (I2CT), Bombay, India, 2019, pp. 1-4, doi: 10.1109/I2CT45611.2019.9033683.