IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

# Heart Disease Prediction

Matam Bhuvaneswar,Chennamsetty Charan Sai, Yandrapalli Laxmi Remanth Kumar Yedla Gnaneshwar Reddy

**Abstract**

The application of machine learning in the field of medical diagnosis is increasing gradually. This can be contributed primarily to the improvement in the classification and recognition systems used in disease diagnosis which is able to provide data that aids medical experts in early detection of fatal diseases and therefore, increase the survival rate of patients significantly. Heart disease is the Leading cause of death worldwide. With the rampant increase in the heart stroke rates at juvenile ages, we need to put a system in place to be able to detect the symptoms of a heart stroke at an early stage and thus prevent it. It is impractical for a common man to frequently undergo costly tests like the ECG and thus there needs to be a system in place which is handy and at the same time reliable, in predicting the chances of a heart disease. So we proposed a system with the help of machine learning techniques and algorithms like Logistic Regression, KNN, SVC, Random Forest ,Decision Tree , XGB Classifier and Naïve Bayes to predict Heart Disease based on different parameters entered by the user in the front end.

**Key words:** Heart disease prediction, Machine learning algorithms, Random Forest Classifier , Early detection , Preventive healthcare

## 1 Introduction

Due to busy schedule as well as routine assignments peoples are facing severe stress and anxiety. More over some other peoples are addicted with chronic habitual behaviour, like consumption of Cigars and Gutuka, those peoples are suffering from chronic diseases like, heart diseases, cancer, Liver problems, Kidney failures etc. To cure such persons with chronic disease is a big hurdle to well know doctors, is a current world issue. Regarding this new challenge, IT professionals are provided hand to hand support to predict such disease early and cure as well as recover the patients from the chronic disease.

In the present scenario each humans are so exceptional in his individual features and manners, but even though every humans may have different pulse rate as well as blood pressure ratings. Based on the history and generic evaluation of medical practitioners and researchers believed that, a healthy humans pulse rate is varied in between of 60 to 100 bpm and BP is varied in between of 120/80 to 140/90 (mm Hg), and these readings are proved by medical practitioners. Heart syndrome is one the vital abrupt death or accidental death of humans in this world, this is might be because of poor dieting as well as physical exercise and other activities like, consumption of alcoholic products, smoking etc. In this article, author is tried to predict and analysis the heart syndrome with

respect to many features like age, gender, blood pressure, heart rate, diabetes etc, but however actual prediction of heart syndrome is totally a critical task to the medical practitioners and analyst. In present market, health industries has many machine learning tools and techniques are used to predict various chronic diseases, but still researchers find some sort of flaws, so they expect some more effective and efficient predictive algorithms to find chronic diseases of humans in early stage itself, so that we can save the life of the patients.

So we proposed a system with the help of machine learning techniques and algorithms like Logistic Regression, KNN, SVC, Random Forest ,Decision Tree , XGB Classifier and Naïve Bayes to predict Heart Disease based on different parameters entered by the user in the front end.

## 2 Litreature Review

P. Sujatha and K. Mahalakshmi [1]In this research paper, the presence of heart disease is predicted by employing DecisionTree, NaïveBayes, RandomForest, SupportVectorMachine, KNearestNeighborandlogisticRegression algorithms. The performance of the algorithms was analyzed using parameters such as Accuracy, Precision, AUC and F1-score. From the experimental result, it is found that the Random Forest is more accurate for predicting the heart disease with accuracy of 83.52 Percent compared with other supervised machine learning algorithms. The F1- Score, AUC and precision score of Random forest classifiers are 84.21 Percent, 88.24 Percent and 88.89Percent respectively.

P. S. Kohli and S. Arora :[2]In this paper, we apply different classification algorithms, each with its own advantage on three separate databases of disease (Heart, Breast cancer, Diabetes) available in UCI repository for disease prediction. The feature selection for each dataset was accomplished by backward modeling using the p value test. The results of the study strengthen the idea of the application of machine learning in early detection of diseases.

Ed-Daoudy and K. Maalmi :[3] This paper propose a realtime heart disease prediction system based on apache Spark which stand as a strong large scale distributed computing platform that can be used successfully for streaming data event against machine learning through in-memory computations. The system consists of two main sub parts, namely streaming processing and data storage and visualization. The first uses Spark MLlib with Spark streaming and applies classification model on data events to predict heart disease. The seconds uses Apache Cassandra for storing the large volume of generated data.

Lakshmanarao, A. Srisaila and T. S. R. Kiran[4] : In this paper, they proposed a novel machine learning model for heart disease prediction. The proposed method was tested on two different datasets from Kaggle and UCI. We applied sampling techniques to the unbalanced dataset and feature selection techniques are used to find the best features. Later several classifier models were applied and achieved good accuracy with ensemble classifier. The experimentations on two datasets shown that the proposed model is effective for heart disease prediction. Python was used for all implementations.

Erdoğan and S. Güney [5]: Nowadays, one of the most important illness is heart disease which cause of mostly patients dead. Medical diagnosis of heart diseases is very difficult. While heart diseases are diagnosed medically, they can be confused with other diseases that show same symptoms such as chest pain, shortness of breath, palpitations and nausea. This makes it difficult to diagnose heart diseases medically. In this study, the presence of heart diseases was determined by using machine learning algorithms. In this study, the data obtained from the patients were weighted according to their effects on the success rate. In this study, a method is proposed for determine weight coefficient. According to proposed method's results, 86,90 Percent success was achieved with 13 different features obtained from the patients

S. Farzana and D. Veeraiah :[6] Health Care Field having enormous data, for processing these data we must use any advanced techniques which will be helpful to provide the effective results and making effective decisions on data and getting the appropriate results. Heart disease is the leading problem and one of the biggest causes for no. of deaths happening all over the world. In this paper, an effective Heart Disease Prediction framework is implemented using algorithms in Machine Learning such as Gaussian Naïve Bayes, Random Forest, K-Nearest Neighbour, Support Vector Machine The framework uses 13 features such as age, gender, blood pressure,

cholesterol, obesity, cp, etc. It is a user-friendly system where we are having some phases. In the first phase, we upload the dataset file and select the algorithm to perform on the selected dataset. Then the accuracy is predicted for each selected algorithm along with a graph, and the modal is generated for the one having highest frequency by training the dataset to it

## 3 Objective

The main aim of this project to predict the Heart Disease using machine learning techniques and algorithms like like Logistic Regression, KNN, SVC, Random Forest ,Decision Tree, XGB Classifier and Naïve Bayes based on different parameters entered by the user in the front end.

## 4 Methodology

A Heart Disease Dataset is taken.

1. The dataset is loaded and preprocessed with various machine learning techniques.
2. The preprocessed data is divided as training and testing data.
3. The prediction model is built using machine learning algorithms like Logistic Regression, KNN, SVC, Random Forest ,Decision Tree, Naïve Bayes and XGB Classifier.
4. The model is trained using training dataset and once the model has been trained successfully it has to be tested.
5. The trained model is tested using testing dataset and accuracy is calculated.
6. The algorithm which gives the best accuracy is taken as our final prediction model.
7. The finalized model is converted into pickle model (binary format data) and saved.
8. A Front End is developed with the help of Flask and HTML.
9. Now user will enter the various parameters required to predict the heart disease in the front end.
10. The collected parameters from the front end are given as input to our finalized algorithm to predict whether the person has the heart disease or not.
11. Finally the predicted output is displayed on the front end.

## 4.1 Preprocessing Data

The dataset undergoes thorough preprocessing to ensure its suitability for machine learning models. Steps include handling missing data, encoding categorical variables, standardizing numerical features, detecting and treating outliers, and performing feature engineering. The data is then split into training and testing sets for model evaluation.

## 4.2 Machine Learning Algorithms

Implementation of Machine Learning Models for Heart Disease Prediction

### 4.2.1 Logistic Regression

Logistic regression, a statistical technique adapted for binary classification problems, is employed. The logistic function transforms a linear combination of input features into a probability between 0 and 1. The model is trained on the preprocessed data to predict the probability of heart disease presence.

### 4.2.2 Random Forest Classifier

A Random Forest Classifier, a versatile algorithm for both classification and regression, is utilized. This ensemble method creates decision trees on randomly selected data samples, and the final prediction is based on the voting from individual trees. The model's flexibility and ability to handle missing values contribute to its effectiveness.

### 4.2.3 K-Nearest Neighbors

The K-Nearest Neighbors algorithm is implemented, a simple yet effective supervised learning method. It classifies a data point based on the majority class of its k-nearest neighbors. The algorithm's simplicity and adaptability make it suitable for our heart disease prediction task.

### 4.2.4 Decision Tree Classifier

A Decision Tree Classifier is introduced, providing a visual representation of decision rules. This algorithm recursively partitions the data based on feature values, making it interpretable and capable of capturing non-linear patterns.

### 4.2.5 Naïve Bayes Classifier

The Naïve Bayes Classifier, a probabilistic model based on Bayes' theorem, is applied. It assumes independence betweenfeatures, makingitcomputationallyefficient. Despiteitssimplicity, NaïveBayesperformswellinvarious applications.

### 4.2.6 Support Vector Machine

Support Vector Machine, a powerful algorithm for classification and regression, is employed. SVM aims to find the hyperplane that best separates the data points into classes. Kernel functions, such as linear, polynomial, and radial basis function, enhance its ability to handle complex relationships.

## 5 Model Evaluation and Selection

Each model is trained and tested on the preprocessed dataset, and accuracy metrics are calculated. The algorithm with the highest accuracy is selected as the final model for heart disease prediction.

# 6 Front-End Implementation

In the context of this research, the front-end development plays a pivotal role in creating a user interface for user registration, login, and heart disease prediction. The front-end is developed using Flask, PyMySQL for database interactions, and HTML for creating user interfaces.

## 6.1 Flask Application Setup

The Flask application is configured to handle user requests and serve HTML templates. The application maintains a connection to a local MySQL database using PyMySQL.

## 6.2 User Registration and Login Routes

Two primary routes are established for user registration and login. The registration route collects user-provided username and password, checks for existing usernames, and inserts new user data into the database. The login route verifies user credentials against the database.

## 6.3 Prediction Route

A dedicated route is created for accessing the heart disease prediction functionality. Users are directed to this route upon successful login. This route ensures that only authenticated users can access the prediction functionality.

## 6.4 Heart Disease Prediction Form Submission

A form is presented to the user for entering various parameters required for heart disease prediction. Upon form submission, theinputparametersareextracted, andthetrainedmachinelearningmodelisusedtomakepredictions. The prediction result is then displayed on the front-end.

## 6.5 HTML Templates

Two HTML templates are used in the front-end implementation. The ‘index.html‘ template serves as the user registration and login interface. Users can register or log in using this template. The ‘prediction.html‘ template provides the interface for entering parameters and displaying the prediction result.

This front-end implementation ensures a seamless user experience, guiding users through registration, login, and ultimately allowing them to interact with the heart disease prediction functionality. The HTML templates are designedtobeuser-friendly, facilitatinginputandresultdisplay. TheFlaskapplication, combinedwithPyMySQL, ensures smooth communication with the back-end machine learning model and the local database.

## 6.6 Backend

This back-end process ensures that the system is equipped with a well-trained machine learning model capable of accurately predicting heart disease based on user-input parameters. The use of Pickle serialization enhances the system’s efficiency by enabling quick access to the finalized model.

## 6.7 conclusion

In conclusion, the proposed system harnesses the power of various machine learning algorithms to predict heart disease based on user-entered parameters. The integration of Logistic Regression, KNN, SVC, Random Forest, Decision Tree, XGB Classifier, and Naïve Bayes contributes to a comprehensive and accurate prediction model.

The importance of early detection in mitigating the severity of heart disease cannot be overstated. Our system, implemented and tested, achieves a commendable accuracy of 91.80percent, particularly attributed to the Random Forest Classifier. This high accuracy underscores the efficacy of machine learning in healthcare applications.

The user-friendly front end, implemented using Flask, HTML, and pymysql, facilitates seamless interaction for data input and retrieval of predictions. The system also incorporates data preprocessing steps, including handling missing values, encoding categorical variables, and standardizing numerical features, ensuring the reliability of predictions.

In practical terms, the successful implementation of this system has the potential to revolutionize early diagnosis of heart disease, enabling individuals to take proactive measures for their well-being. The utilization of machine learning in healthcare underscores the transformative impact technology can have on preventive healthcare practices.

As we move forward, continuous refinement of the model and exploration of additional features can further enhance the predictive capabilities of the system. This research stands as a testament to the synergy between healthcare and machine learning, paving the way for more sophisticated and accurate predictive models in the field of cardiology.

## References

[1] P. Kola Sujatha and K. Mahalakshmi. Performance evaluation of supervised machine learning algorithms in prediction of heart disease. *2020 IEEE International Conference for Innovation in Technology (INOCON)*, pages 1–7, 2020.

[2] Pahulpreet Singh Kohli and Shriya Arora. Application of machine learning in disease prediction. *2018 4th International Conference on Computing Communication and Automation (ICCCA)*, pages 1–4, 2018.

[3] Abderrahmane Ed-Daoudy and Khalil Maalmi. Real-time machine learning for early detection of heart disease using big data approach. pages 1–5, 04 2019.

[4] A. Lakshmanarao, A. Srisaila, and Srinivasa Tummala. Heart disease prediction using feature selection and ensemble learning techniques. pages 994–998, 02 2021.

[5] Alperen Erdoğan and Selda Guney. Heart disease prediction by using machine learning algorithms. *2020 28th Signal Processing and Communications Applications Conference (SIU)*, pages 1–4, 2020.

[6] Shaik Farzana and Duggineni Veeraiah. Dynamic heart disease prediction using multi-machine learning techniques. *2020 5th International Conference on Computing, Communication and Security (ICCCS)*, pages 1–5, 2020.