© 2024 IJNRD | Volume 9, Issue 4 April 2024| ISSN: 2456-4184 | IJNRD.ORG

IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

Effective Heart Disease Prediction Using Hybrid Machine Learning Technique

Tanaya Chaporkar
Department of Information Technology, Prof. Ram Meghe Institute of Technology & Research, Amravati, Maharashtra, India

Mayuresh Khanzode
Department of Information Technology, Prof. Ram Meghe Institute of Technology & Research, Amravati, Maharashtra, India

Hemant Malpani
Department of Information Technology, Prof. Ram Meghe Institute of Technology & Research, Amravati, Maharashtra, India

Tushar Joshi
Department of Information Technology, Prof. Ram Meghe Institute of Technology & Research, Amravati, Maharashtra, India

Prof. H. D. Misalkar
Assistant Professor
Department of Information Technology, Prof. Ram Meghe Institute of Technology & Research, Amravati, Maharashtra, India

Abstract— The diagnosis of heart disease is the most difficult task in the medical field. The diagnosis of heart disease is difficult as a decision relied on grouping of large clinical and pathological data. Due to this complication, the interest increased in a significant amount between the researchers and clinical professionals about the efficient and accurate heart disease prediction. In case of heart disease, the correct diagnosis in early stage is important as time is the very important factor. Heart disease is the principal source of deaths widespread, and the prediction of heart disease is significant at an untimely phase. Machine learning in recent years has been the evolving, reliable and supporting tools in medical domain and has provided the greatest support for predicting disease with correct case of training and testing. The main idea behind this work is to study diverse prediction models for the heart disease and selecting important heart disease feature using hybrid machine learning algorithm. Logistic regression and gradient boosting are Machine Learning algorithm which has the high accuracy compared to other Supervised Machine Learning algorithms. Byusing logistic regression and gradient boosting algorithm, we are going to predict if a person has heart disease or not.

Keywords— Machine Learning, Logistic Regression, Gradient Boosting Algorithm, Hybrid machine learning algorithms.

I. INTRODUCTION

Heart disease remains one of the leading causes of mortality worldwide, making early detection and effective prediction crucial for timely intervention and treatment. In recent years, machine learning algorithms have emerged as powerful tools in the field of healthcare, offering the potential to improve the accuracy and efficiency of diseaseprediction. Among these algorithms, logistic regression and gradient boosting have gainedprominence for their effectiveness in predicting complex

medical conditions.

Logistic regression is a widely used statistical technique for binary classification tasks, making it particularly suitable for predicting the presence or absence of heart disease basedon various risk factors and symptoms. It models the probability of a binary outcome usinga logistic function, allowing for the estimation of the likelihood of disease occurrence basedon input features.

On the other hand, gradient boosting is an ensemble learning technique that combines multiple weak learners, typically decision trees, to create a strong predictive model. Algorithms like XGBoost and LightGBM have demonstrated exceptional performance in a wide range of classification tasks, including heart disease prediction. By iteratively improving model accuracy through boosting, gradient boosting algorithms can effectivelycapture complex patterns and relationships within the data, leading to highly accurate predictions.

In this study, we aim to explore the application of logistic regression and gradient boostingmachine learning algorithms for heart disease prediction. By leveraging a dataset containing demographic information, medical history, and clinical measurements, we seekto develop predictive models capable of accurately identifying individuals at risk of heart disease. Through comprehensive data analysis, model training, and evaluation, we aim to assess the performance and effectiveness of these algorithms in predicting heart disease, ultimately contributing to improved patient outcomes and healthcare decision-making.

II. LITERATURE SURVEY

The literature survey for the research paper on effective heart disease prediction using hybrid machine learning techniques reveals a growing interest in leveraging the power of combining multiple algorithms to enhance predictive accuracy and improve patient outcomes. Rajkomar, A., Dean, J., Kohane, I. (2019) proposed “Machine Learning in Medicine”. This study explores machine learning and its potential to enhance clinical decision-making as a tool for safe value-based care. The authors discuss how machine learning can affect prognosis, diagnosis, treatment, clinician workflow, and access to expertise [1].

Krittanawong, C., Zhang, H., Wang, Z., Aydar, M., Kitai, T. (2018) proposed “Artificial Intelligence in Precision Cardiovascular Medicine”. This study explores the role of artificial intelligence, including machinelearning, in advancing precision medicine for cardiovascular diseases. It discusses recent developments in predictive modeling and risk stratification [2].

Dey, D., Slomka, P. J., Leeson, P., Comaniciu, D., Shrestha, S., Sengupta, P. P., & Marwick, T. H. (2020) proposed "Artificial intelligence in cardiovascular imaging". This study examines the applications of artificial intelligence in cardiovascular imaging modalities, such as echocardiography, cardiac MRI,and CT angiography

[3]. Al'Aref, S. J., Anchouche, K., Singh, G., Slomka, P. J., Kolli, K. K., Kumar, A., & Berman, D. S. (2019) propsed "Clinical applications of machine learning in cardiovasculardisease and its relevance to cardiac imaging". This study provides an overview of clinical applications of machine learning in cardiovascular disease, with a focus on cardiac imaging. It discusses the potential of ML algorithms to aid in risk prediction, diagnosis, and treatment planning [4].

Johnson, K. W., Torres Soto, J., Glicksberg, B. S., Shameer, K., Miotto, R., Ali, M., & Dudley, J. T. (2018) proposed "Artificial intelligence in cardiology" Journal of theAmerican College of Cardiology. This study provides a guide for clinicians on relevant aspects of artificial intelligence and machine learning, reviews selected applications of these methods in cardiology to date, and identifies how cardiovascular medicine could incorporate artificial intelligence in the future [5].

Weng, S. F., Reps, J., Kai, J., Garibaldi, J. M., Qureshi, N. (2017) proposed "Can machine-learning improve cardiovascular risk prediction using routine clinical data?". This study investigates the utility of machine learning algorithms for improving cardiovascularrisk prediction using routine clinical data records [6].

Attia, Z. I., Kapa, S., Lopez-Jimenez, F., McKie, P. M., Ladewig, D. J., Satam, G. & Friedman, P. A. (2019) proposed "Screening for cardiac contractile dysfunction using an artificial intelligence–enabled electrocardiogram". This study explores the use of artificial intelligence-enabled electrocardiogram (ECG) analysis for screening cardiac contractile dysfunction. It highlights the potential of ML algorithms to detect subtle changes in ECG patterns indicative of underlying cardiac abnormalities [7].

A. S. Abdullah and R. R. Rajalaxmi (2012) proposed "A Data Mining Model for Predictingthe Coronary Heart Disease Using Random Forest Classifier". This study presents a datamining model utilizing the Random Forest classifier to improve prediction accuracy and analyze CHD-related events [8]. A. H. Alkeshuosh, M. Z. Moghadam, I. Al Mansoori, and M. Abdar (2017) proposed "The Particle Swarm Optimization (PSO) calculation". This study presents that the work is based on the real-world CAD dataset and aims at the detection of this disease by producingthe accurate and effective rules [9].

N. Al-milli (2013) proposed "Backpropogating Neural Network for Prediction of heart disease." This study addressed that the focus lies on addressing the significant challenge of diagnosing heart disease through the development of intelligent medical decision support systems. The paper highlights the widespread utilization of neural networks for heart disease prediction due to their adaptability and learning capabilities. Specifically, the proposed framework utilizes 13 clinical properties to forecast coronary illness, demonstrating superior performance compared to existing methodologies. With an accuracy exceeding 85%, the system proves to be reliable, although it requires a substantial volume of training data to achieve such results [10].

P. K. Anooj (2012) proposed "A weighted fuzzy rule-based clinical decision support system (CDSS) is presented for the diagnosis of heart disease". This research introduces a ground-breaking approach to heart disease diagnosis through a Weighted Fuzzy Rule-Based Clinical Decision Support System (CDSS). Traditionally, clinical decision support systems relied on manual input from medical experts, which was time-consuming and potentially subjective. To overcome these limitations, Anooj employs machine learning techniques to automatically extract knowledge from patient data. By leveraging this innovative CDSS, the study aims to improve the accuracy and efficiency of heart disease diagnosis. This approach represents a significant advancement in medical decision support, allowing for more objective and data-driven diagnoses while reducing dependence on individual expert opinions [11].

C. A. Devi, S. P. Rajamhoana, K. Umamaheswari, R. Kiruba, K. Karunya, and R. Deepika proposed "Analysis of neural networksn based heart disease prediction system". This study aims to evaluate different machine learning and deep learning techniques for heart disease prediction and classification [12].

S. Srivastava and S. Mishra proposed "Machine learning techniques for heart disease prediction: A review". This study evaluates the efficacy of machine learning algorithms, including logistic regression, random forest classifier, and support vector machine (SVM), for heart disease prediction, utilizing a Kaggle dataset [13].

A. Gavhane, G. Kokkula, I. Pandya, and K. Devadkar proposed "Prediction of heart disease using machine learning". This study proposes the development of an application that can predict heart disease vulnerability based on basic symptoms such as age, sex, and pulse rate. They advocate for the utilization of neural networks, a machine learning algorithm known for its accuracy and reliability, in the proposed system [14].

H. A. Esfahani and M. Ghazanfari proposed "Cardiovascular disease detection using a new ensemble classifier". This study utilize data from cardiovascular patients obtained from the UCI Laboratory to employ pattern discovery algorithms such as Decision Trees, Neural Networks, Rough Sets, Support Vector Machines (SVM), and Naive Bayes. They compare the accuracy and predictive performance of these algorithms. Subsequently, the authors propose a hybrid algorithm aimed at enhancing the accuracy of these methods. Based on their findings, the proposed hybrid approach achieves an F-measure of 86.8%, surpassing the performance of other competing methods [15].

III. PROBLEM STATEMENT AND PROPOSED METHODLOGY

In this paper, we proposed Effective heart disease prediction using hybrid machine learning techniques. Cardiovascular diseases, including heart disease, are a significant global health concern, leading to substantial morbidity and mortality rates. Timely and accurate prediction of heart disease plays a critical role in preventing adverse outcomes and optimizing patient care. Traditional risk assessment methods often rely on single machine learning algorithms or statistical models, which may not fully capture the complexity and interplay of factors contributing to heart disease. Moreover, these approaches may lack the necessary accuracy and robustness required for early detection and intervention.
The problem at hand is to develop an advanced predictive model that effectively addresses the limitations of individual algorithms and traditional methods by harnessing the power of hybrid machine learning techniques. The primary objective is to enhance the accuracy, reliability, and interpretability of heart disease prediction, enabling healthcare professionals to make informed decisions and implement timely interventions.

The below figure shows the process flow diagram or proposed work. First, we have collected the Cleveland Heart Disease Database from UCI website then pre- processed thedataset and select 16 important features.

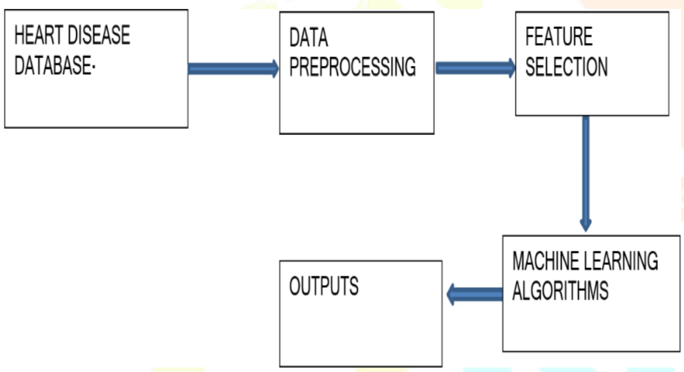


Fig.1 : System Architecture

For feature selection we used Recursive feature Elimination Algorithm using Chi2 methodand get 16 top features. After that applied ANN and Logistic algorithm individually and compute the accuracy. Finally, we used proposed Ensemble Voting method and compute best method for diagnosis of heart disease.
The entire work of this project is divided into 4 modules. They are:

- Data Pre-Processing
- Feature Extraction
- Classification
- Prediction

- **Data preprocessing:**
  This file contains all the pre-processing functions needed to process all input documents and texts. First, we read the train, test and validation data files then performed some preprocessing like tokenizing, stemming etc. There are some exploratory data analyses is performed like response variable distribution and data quality checks like null or missing values etc. Data preprocessing is the process of transforming raw data into an understandable format. It is also an important step in data mining as we cannot work withraw data. The quality of the data should be checked before applying machine learning or data mining algorithms. Preprocessing of data is mainly to check the data quality. The quality can be checked by the following-
    - Accuracy: To check whether the data entered is correct or not.
    - Completeness: To check whether the data is available or not.
    - Consistency: To check whether the same data is kept in all the places that do or do not match.
    - Timeliness: The data should be updated correctly.
    - Believability: The data should be trustable.
    - Interpretability: The understandability of the data.

- **Feature Extraction:**
  In this file we have performed feature extraction and selection methods from sci-kit learn python libraries. For feature selection, we have used methods like simple bag-of- words and n-grams and then term frequency like tf-idf weighting. We have also used word2vec and POS tagging to extract the features, though POS tagging and word2vec hasnot been used at this point in the project.

- **Bag of Words:**
  It's an algorithm that transforms the text into fixed-length vectors. This is possible by counting the number of times the word is present in a document. The word occurrences allow to compare different documents and evaluate their similarities for applications, suchas search, document classification, and topic modeling. The reason for its name, ―Bag- Of-Words‖, is due to the fact that it represents the sentence as a bag of terms. It doesn't consider the order and the structure of the words, but it only checks if the words appear inthe document.

- **N-grams**:
  N-grams are continuous sequences of words or symbols or tokens in a document. In technical terms, they can be defined as the neighboring sequences of items in a document.They come into play when we deal with text data in NLP (Natural Language Processing) tasks.

- **TF-IDF Weighting:**
  TF-IDF stands for term frequency-inverse document frequency and it is a measure, used inthe fields of information retrieval (IR) and machine learning, that can quantify theimportance or relevance of string representations (words, phrases, lemmas, etc.) in a document amongst a collection of documents (also known as a corpus).

- **Classification:**
  Here we have built all the classifiers for the heart disease prediction. The extracted featuresare fed into different classifiers. We have used Naive-Bayes, Logistic Regression, Linear SVM, Stochastic gradient decent and Random Forest classifiers from sklearn. Each of theextracted features was used in all the classifiers. Once fitting the model, we compared thef1 score and checked the confusion matrix. After fitting all the classifiers, 2 best performingmodels were selected as candidate models for heart diseases classification. We have performed parameter tuning by implementing GridSearchCV methods on these candidate models and chosen best performing parameters for these classifiers. Finally, selected model was used for heart disease detection with the probability of truth. In Addition to this, we have also extracted the top 50 features from our term-frequency tf-idf Vectorizer to see what words are most and important in each of the classes. We have also used Precision- Recall and learning curves to see how training and test set performs when we increase the amount of data in our classifiers.

- **Prediction:**
  Our finally selected and best performing classifier was algorithm which was then saved ondisk with name final_model.sav. Once you close this repository, this model will be copied to user's machine and will be

used by prediction.py file to classify the heart diseases. It takes a news article as input from user then model is used for final classification output thatis shown to user along with probability of truth.

- **Data Flow Diagram:**
The data flow diagram (DFD) is one of the most important tools used by system analysis. Data flow diagrams are made up of number of symbols, which represents system components. Most data flow modeling methods use four kinds of symbols: Processes, Datastores, Data flows and external entities. These symbols are used to represent four kindsof system components. Circles in DFD represent processes. Data Flow represented by a thin line in the DFD, and each data store has a unique name and square or rectangle represents external entities.

Level 0:

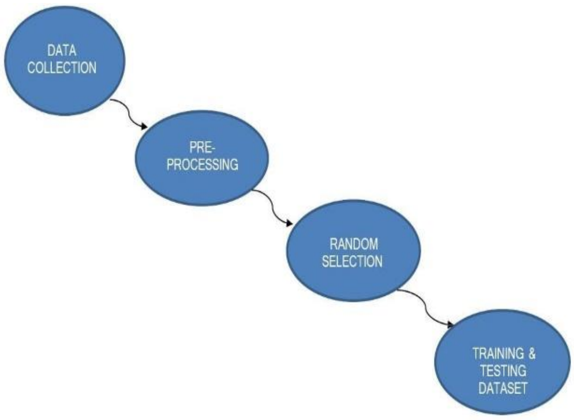


Fig 2: Data Flow Diagram Level 0

Level 1:

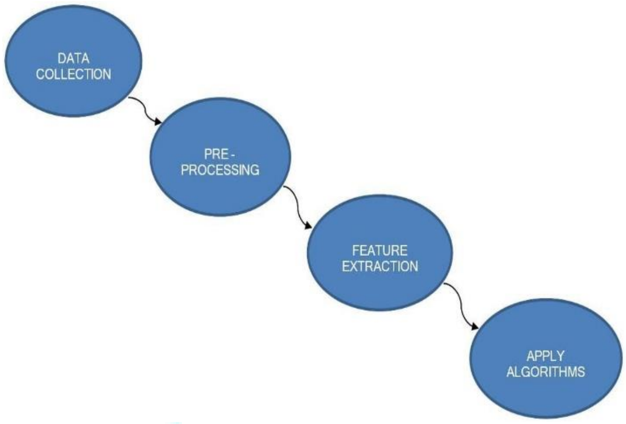


Fig 3: Data Flow Diagram Level 1

IV. IMPLEMENTATION

Implementing a system for heart disease prediction using hybrid machine learning techniques involves translating the system design into functional code. The steps which areused for implementation are given below.
Steps for Implementation:

- Install the required packages for building the _Passive Aggressive Classifiers:
  Install the necessary Python packages for building and implementing logistic regression and gradient boosting classifiers. This include scikit-learn, XGBoost.
- Load the libraries into the workspace from the packages:
  Import the required libraries into the Python workspace. This typically includes libraries such as pandas, NumPy, scikit-learn, and XGBoost.
- Read the input data set:
  Load the heart disease dataset into a pandas DataFrame.
- Normalize the given input dataset:
  Normalize the input features to ensure that they are on a similar scale, which can improve the performance of some machine learning algorithms.
- Divide this normalized data into two parts:
  Split the normalized dataset into training and testing sets.
- Train data:
  Here, X_train and y_train represent the features and target variable of the training set.
  Test data:
  X_test and y_test represent the features and
  target
  Variable of the testing set.

Algorithm Used:

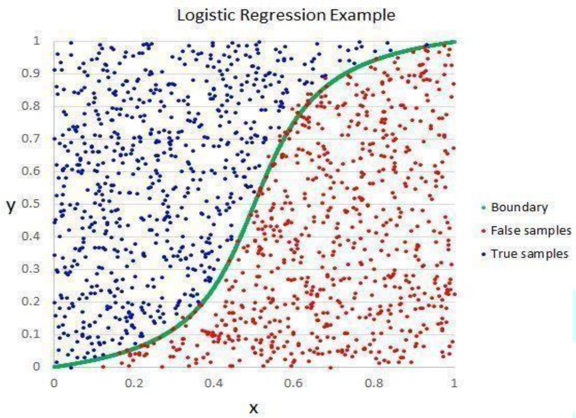
- **Logistic Regression:**

A popular statistical technique to predict binomial outcomes (y = 0 or 1) is Logistic Regression. Logistic regression predicts categorical outcomes (binomial / multinomial values of y). The predictions of Logistic Regression (henceforth, LogR in this article) are in the form of probabilities of an event occurring, i.e., the probability of y=1, given certainvalues of input variables x. Thus, the results of LogR range between 0-1. LogR models the data points using the standard logistic function, which is an S- shaped curve also called as sigmoidcurve and is given by the equation.
Logistic Regression Assumptions:

- Logistic regression requires the dependent variable to be binary.
- For a binary regression, the factor level 1 of the dependent variable should represent the desired outcome.
- Only the meaningful variables should be included.
- The independent variables should be independent of each other.
- Logistic regression requires quite large sample sizes.

- Even though, logistic (logit) regression is frequently used for binary variables (2 classes), it can be used for categorical dependent variables with more than 2classes.In this case It's called Multinomial Logistic Regression.

Fig 4: Logistic Regression



- **Gradient Boosting:**

Gradient Boosting is a powerful machine learning technique used for both regression and classification problems. It belongs to the ensemble learning methods, where multiple weaklearners are combined to form a strong learner. Gradient Boosting builds the model in a stage-wise fashion and generalizes them by optimizing a differentiable loss function. Here'sa breakdown of the Gradient Boosting algorithm:

Initialization:
Gradient Boosting starts with an initial prediction or model, often the mean of the target variable for regression or a constant for classification.

Fit Stage-wise:
In each iteration, a weak learner (usually a decision tree with shallow depth) is trained on the residuals of the previous predictions. The weak learner is fit to the negative gradient ofthe loss function with respect to the predictions made by the model so far. The weak learner's predictions are scaled by a factor called the learning rate (or shrinkage) before being added to the ensemble.

Update Model:
The weak learner's predictions are added to the ensemble of models built so far, updating the overall prediction of the model. The process continues for a predefined number of iterations or until a stopping criterion is met.

Regularization:
Regularization techniques like shrinkage (learning rate) and tree-related parameters (e.g., maximum depth, minimum samples per leaf) are used to prevent overfitting. Otherregularization techniques include subsampling (stochastic gradient boosting) and feature subsampling at each iteration.

Final Prediction:
The final prediction is made by summing up the predictions of all weak learners in the ensemble. For classification problems, a decision is made based on the aggregated predictions, often using a threshold or SoftMax function.
Loss Function:
Gradient Boosting minimizes a differentiable loss function, which is problem-specific (e.g., mean squared error for regression, cross-entropy loss for classification). Common loss functions include Least Squares Loss, Log Loss (Binary or Multiclass), and Huber Loss.

Gradient Descent:
Gradient Boosting optimizes the loss function using gradient descent (or a variant like stochastic gradient descent) to update the model parameters.

XGBoost, LightGBM, and CatBoost:
Variants of Gradient Boosting such as XGBoost, LightGBM, and CatBoost have been developed with optimizations and enhancements to improve efficiency, speed, and performance

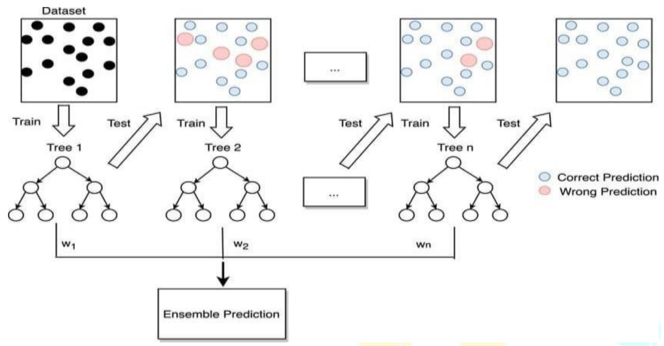


Fig 5: Gradient Boosting

V. RESULT

This section presents the outcomes derived from the executed system, demonstrating snapshots to elucidate the functionality and features of our system effectively.

- **Dataset Description:**

1. The dataset encompasses features related to heart disease, including age, sex, chest pain type, resting blood pressure, serum cholesterol level, fasting blood sugar, resting electrocardiographic results, maximum heart rate achieved, exercise-induced angina, ST depression induced by exercise, slope of the peak exercise ST segment, number of major vessels colored by fluoroscopy and thalassemia.
2. It comprises both categorical attributes (e.g., sex, chest pain type) and numerical features (e.g., age, resting blood pressure).
3. The label column ('target') delineates the diagnosis of heart disease, with values indicating the absence (0) or presence (1) of heart disease.

age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
63	1	1	145	233	1	2	150	0	2.3	3	0	6	0
67	1	4	160	286	0	2	108	1	1.5	2	3	3	1
67	1	4	120	229	0	2	129	1	2.6	2	2	7	1
37	1	3	130	250	0	0	187	0	3.5	3	0	3	0
41	0	2	130	204	0	2	172	0	1.4	1	0	3	0
56	1	2	120	236	0	0	178	0	0.8	1	0	3	0
62	0	4	140	268	0	2	160	0	3.6	3	2	3	1
57	0	4	120	354	0	0	163	1	0.6	1	0	3	0
63	1	4	130	254	0	2	147	0	1.4	2	1	7	1
53	1	4	140	203	1	2	155	1	3.1	3	0	7	1
57	1	4	140	192	0	0	148	0	0.4	2	0	6	0
56	0	2	140	294	0	2	153	0	1.3	2	0	3	0
56	1	3	130	256	1	2	142	1	0.6	2	1	6	1
44	1	2	120	263	0	0	173	0	0	1	0	7	0
49	1	2	130	266	0	0	171	0	0.6	1	0	3	0
64	1	1	110	211	0	2	144	1	1.8	2	0	3	0
58	0	1	150	283	1	2	162	0	1	1	0	3	0
58	1	2	120	284	0	2	160	0	1.8	2	0	3	1
58	1	3	132	224	0	2	173	0	3.2	1	2	7	1
60	1	4	130	206	0	2	132	1	2.4	2	2	7	1
50	0	3	120	219	0	0	158	0	1.6	2	0	3	0
58	0	3	120	340	0	0	172	0	0	1	0	3	0
66	0	1	150	226	0	0	114	0	2.6	3	0	3	0
43	1	4	150	247	0	0	171	0	1.5	1	0	3	0
40	1	4	110	167	0	2	114	1	2	2	0	7	1
69	0	1	140	239	0	0	151	0	1.8	1	2	3	0
60	1	4	117	230	1	0	160	1	1.4	1	2	7	1
64	1	3	140	335	0	0	158	0	0	1	0	3	1
59	1	4	135	234	0	0	161	0	0.5	2	0	7	0
44	1	3	130	233	0	0	179	1	0.4	1	0	3	0
42	1	4	140	226	0	0	178	0	0	1	0	3	0
43	1	4	120	177	0	2	120	1	2.5	2	0	7	1
57	1	4	150	276	0	2	112	1	0.6	2	1	6	1
55	1	4	132	353	0	0	132	1	1.2	2	1	7	1
61	1	3	150	243	1	0	137	1	1	2	0	3	0

Screenshot 1: Dataset

- **Preprocessing Steps:**
    - Handling Missing Values: Identify and address any missing values in the dataset. This may involve techniques such as imputation or removal of rows with missing data.
    - Feature Scaling: Normalize numerical features to ensure they have a similar scale, considering techniques like standardization or min-max scaling.
    - Encoding Categorical Variables: Convert categorical variables into a numerical format suitable for machine learning algorithms, using methods like one-hot encoding.
    - Dealing with Class Imbalance: Address any class imbalance issues in the dataset, ensuring representative samples for model training and evaluation.

- **Data Splitting:**
    - Train-Test Split: Divide the dataset into separate training and testing sets. The training set is used to train the machine learning models, while the testing set is used to evaluate their performance**.**
    - Stratified Sampling: Ensure that the distribution of fraudulent and legitimate transactions is maintained in both the training and testing sets. This helps prevent bias and ensures representative samples for model evaluation.
    - Cross-Validation (Optional): Consider employing techniques like k-fold cross-validation to further validate model performance. This involves splitting the dataset into multiple folds, training the model on different combinations of training and validation sets, and averaging the results.
- **Experimental Results**

User Registration:

Account Creation: Prior to accessing the heart disease detection system, users are required to create an account. This process involves providing necessary information such as username, email address, and password.

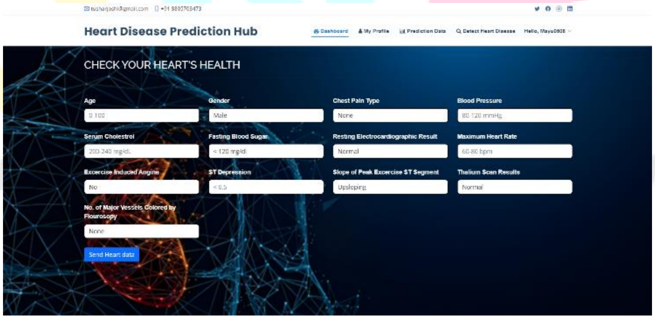
Heart Disease Detection:

Data Input: Upon accessing the heart disease detection page, users are prompted to input their heart-related details. This includes demographic information such as age and gender, as well as medical parameters like blood pressure, cholesterol levels, and any relevant symptoms.

Submission: After entering the required information, users submit the data to the heart disease detection system for analysis.

Model Prediction: The submitted data is processed by the heart disease prediction model, which employs machine learning algorithms trained on the provided dataset.

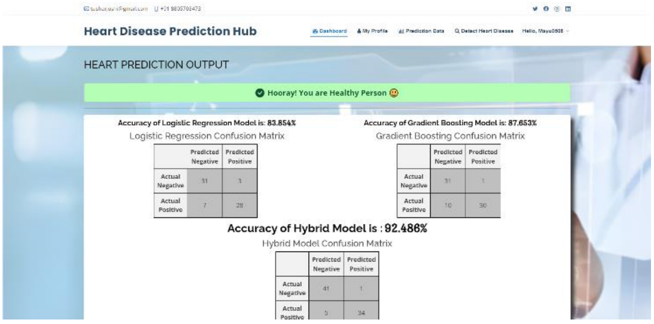
Outcome: Based on the analysis of the input data, the model generates a prediction regarding the presence or absence of heart disease in the user. This prediction is communicated to the user through the system interface.



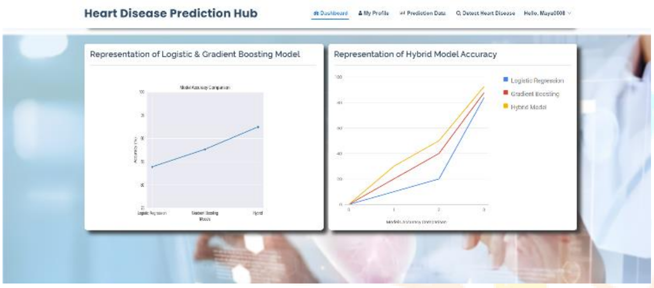
Screenshot 2: Heart Details Input Page

- **Results Snapshot:**

Below is a snapshot of the outcomes derived from the inputs given by user in heart details input page:



Screenshot 3: Report of Heart Disease Prediction



Screenshot 4: Graphical Representation of model

VI. Conclusion

Our study highlights the effectiveness of logistic regression and gradient boosting algorithms for heart disease prediction. Gradient boosting, in particular, shows superior performance. These findings underscore the potential of machine learning in improving heart disease diagnosis and patient care. Further research in this area can enhance predictive accuracy and clinical outcomes.

References

[1] Rajkomar, A., Dean, J., Kohane, I., "Machine Learning in Medicine", April 2019, TheNew-England Medical Review and Journal 380(14)

[2] Krittanawong, C., Zhang, H., Wang, Z., Aydar, M., Kitai, T. , "Artificial Intelligencein Precision Cardiovascular Medicine", May 2017, Journal of the American College of Cardiology 69(21):2657 – 64.

[3] Dey, D., Slomka, P. J., Leeson, P., Comaniciu, D., Shrestha, S., Sengupta, P. P., & Marwick, T. H., "Artificial intelligence in cardiovascular imaging": JACC state-of-the-artreview, Journal of the American College of Cardiology, Mar 2019, 73(11)

[4] Al'Aref, S. J., Anchouche, K., Singh, G., Slomka, P. J., Kolli, K. K., Kumar, A., ... & Berman, D. S., "Clinical applications of machine learning in cardiovascular disease and itsrelevance to cardiac imaging", European Heart Journal, Volume 40, Issue 24, June 2019.

[5] Johnson, K. W., Torres Soto, J., Glicksberg, B. S., Shameer, K., Miotto, R., Ali, M., ...& Dudley, J. T., "Artificial intelligence in cardiology", Journal of the American College ofCardiology, Volume 71, Issue 23, June 2018.

[6] Weng, S. F., Reps, J., Kai, J., Garibaldi, J. M., Qureshi, N., "Can machine-learningimprove cardiovascular risk prediction using routine clinical data?" PLoS One Journal,April 2017,12(4)

[7] Attia, Z. I., Kapa, S., Lopez-Jimenez, F., McKie, P. M., Ladewig, D. J., Satam, G., ... & Friedman, P. A., "Screening for cardiac contractile dysfunction using an artificial intelligence–enabled electrocardiogram", nature medicine, 2019 Jan,25(1)

[8] A. S. Abdullah and R. R. Rajalaxmi, "A Data Mining model has been created utilizingRandom Forest classifier", International Conference in Recent trends in Computational Methods, Communication and Controls,

April 2012, 22-25.

[9] A. H. Alkeshuosh, M. Z. Moghadam, I. Al Mansoori, and M. Abdar, A, "The ParticleSwarn optimization (PSO) Calculation", in Proc. Int. Conf. Comput. Appl. (ICCA), sep 2017.

[10] N. Al-milli (2013), "Backpropogation neural network for prediction of heart disease",J. Theor. Appl. Inf. Technol., vol. 56, no.1.

[11] P. K. Anooj (2012), "A weighted fuzzy rule-based clinical decision support system (CDSS) is presented for the diagnosis of heart disease, J. King Saud Univ.-Comput. Inf. Sci., vol. 24, no. 1, pp. 27–40, Jan 2012.

[12] 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.

[13] S. Srivastava and S. Mishra, "Machine learning techniques for heart disease prediction: A review", Fully refereed international journal, Volume 05, Issue 3, March 2023.

[14] 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.

[15] 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.