**PREDICTION AND DIAGNOSIS OF VARIOUS HEART DISEASES**

**AND CHANCES OF HEART ATTACK**

|  |  |  |  |
| --- | --- | --- | --- |
| Mayank Tyagi  *Department of IT*  Ajay Kumar Garg Engineering College  Ghaziabad, India | Mohit Arora  *Department of IT*  Ajay Kumar Garg Engineering College  Ghaziabad, India | Pragati Gupta *Department of IT*  Ajay Kumar Garg Engineering College  Ghaziabad, India | Radhika Bhatt *Department of IT*  Ajay Kumar Garg Engineering College  Ghaziabad, India |

**Abstract**

The prevalence of cardiovascular diseases remains a major global health concern, finding innovative approaches for early detection, diagnosis, and preventive health management. This project introduces an integrated system comprising predictive analytics for heart diseases, a diagnostic module for identifying specific conditions, and a health advisory chatbot for personalized health guidance. The system utilizes machine learning algorithms to analyse health data, assess the risk of heart diseases, and provide timely insights for preventive measures.

This project focuses on developing the heart disease detection based on the patient health factors and developing the chat bot which will advise the patient based on their heart disease condition. The system features an intelligent and user-friendly GUI implemented using Machine learning, boosting algorithms and the Flask framework. Various machine learning algorithms such as Random Forest (RF), K-Nearest Neighbour (KNN), Logistic Regression (LR) and Artificial Neural Network (ANN) used for heart disease classification and the results were discussed. The proposed project aims to develop and implement an AI-driven heart disease detection and health advisory system. This system, potentially in the form of a conversational bot, should address the following key objectives such as Early detection, Continuous monitoring, User-Friendly Interface.

Keywords: Heart Disease Detection, Smart Health Monitoring, Machine Learning, Conversational Bot

1. **INTRODUCTION**

Globally, cardiovascular disease (CVD) stands as the foremost cause of mortality, posing a considerable public health challenge. This widespread issue has led to significant socioeconomic burdens on patients, their families, and the respective governments. Identification of individuals at elevated risk for CVD through predictive models employing risk stratification is a crucial step. Subsequently, implementing targeted measures, such as dietary modifications and the administration of statins, can effectively mitigate the risk in this high-risk group, thereby contributing to primary prevention of cardiovascular disease

Cardiovascular diseases (CVDs) continue to be a primary contributor to global morbidity and mortality. Effectively addressing the impact of heart diseases necessitates early identification, precise diagnosis, and tailored health management. Unfortunately, current healthcare systems frequently lack effective resources for forecasting individual risks, pinpointing specific conditions, and offering ongoing health advice. This absence of a comprehensive solution results in deficiencies in preventive care, resulting in delayed interventions and less than optimal health results.

Various guidelines focusing on the assessment and handling of cardiovascular disease (CVD) advocate for the utilization of predictive models to pinpoint individuals at elevated risk and aid in clinical decision-making. Notably, the Pooled Cohort Equations and the Framingham Cardiovascular Risk Equation have undergone independent evaluations across diverse populations. However, the outcomes revealed that both equations exhibited only modest discriminatory power and inadequate calibration levels.

Heart disease manifests through a diverse array of symptoms, making swift and accurate diagnosis a challenging task. Treating databases of heart disease patients as practical cases involves assigning greater weight to attributes significantly impacting disease prediction. Leveraging the collective knowledge of various specialists stored in databases appears to be a sensible approach to enhance the diagnostic process. It not only provides an additional source of information for healthcare personnel but also contributes to decision-making.

The heart, as the body's primary organ, faces risk factors categorized as controllable and uncontrollable. Uncontrollable factors, supported by clinical evidence, elevate the likelihood of developing heart disease, specifically cardiovascular disease. These include a family history of cardiovascular disease, elevated levels of LDL (bad) cholesterol, diminished levels of HDL (good) cholesterol, hypertension, and adherence to a high-fat diet, among others.

Factors within human control play a pivotal role in preventing heart disease, encompassing lifestyle choices like smoking, alcohol consumption, weight management, blood pressure regulation, and cholesterol levels. Diagnosis typically involves analysing a patient's current health status, reviewing historical diagnoses of patients with similar conditions, and considering various criteria, allowing doctors to make informed decisions about their diagnoses.

The various heart disease that are medically diagnosed are,

1. Coronary Heart Diseases: Refers to the damage or infection affecting the primary blood vessels supplying the heart.
2. Cardiomyopathy: A condition, either genetic or acquired, that impacts the heart muscles.
3. Ischemic Heart Disease: Occurs when constrictions in the heart arteries impede the adequate flow of blood and oxygen to the heart muscles, leading to heart problems.
4. Heart Failure: A chronic condition where the heart is incapable of pumping blood at the necessary rate.
5. Hypertensive Heart Disease: Linked to high blood pressure, resulting in complications and issues related to the heart.
6. Inflammatory Heart Disease: Refers to heart diseases or problems caused by bacterial or viral infections.
7. Valvular Heart Disease: Involves damage to or dysfunction of a heart valve, impairing its normal function.

Even though, various heart disease each are containing their own risk factors. Some of the activities and habits are leading to the heart disease and those can be well analysed and can be use for the effective health care applications when we move forward towards the digital era.

Researchers are increasingly turning to artificial intelligence (AI) to extract novel medical insights, empowering clinicians to gain a deeper understanding of disease symptoms and make more informed decisions for their patients. This trend aligns with the growing prevalence of data from the Internet of Things (IoT) within healthcare systems. To explore previously unidentified risk factors, ongoing initiatives aim to standardize medical data and organize national health screening data. Standardizing medical data is the initial step, as these risk variables may be correlated with disease occurrences, offering valuable insights into underlying disease mechanisms. Moreover, the development of accurate prediction models for disease incidence requires the analysis of extensive datasets. The integration of AI and substantial data in cardiovascular disease (CVD) prediction models is increasingly becoming a common practice.

In this project a comprehensive analysis utilizing Random Forest (RF), k-Nearest Neighbors (KNN), Logistic Regression, and Artificial Neural Network (ANN) models was conducted to detect and compare performance on heart disease data. Additionally, a health chatbot was developed and assessed in parallel. The findings from this multifaceted investigation contribute valuable insights into the effectiveness of diverse machine learning algorithms in heart disease detection and the functionality of the health chatbot. These results pave the way for informed decision-making in the realm of cardiovascular health and highlight the potential of AI-driven solutions in the healthcare domain.

1. **LITERATURE REVIEW**

Md. Imam Hossain et al ., proposed Heart disease prediction using distinct artificial intelligence techniques: performance analysis and comparison The article explores the use of distinct artificial intelligence techniques for heart disease prediction. The authors analysed patient data to identify the most significant factors for accurate diagnosis. The article includes information on author contributions, funding, data availability, conflict of interest, and ethical approval. The study's findings could have implications for improving patient management opportunities.

Chintann M. Bhatt et.al., proposed Effective Heart Disease Prediction Using Machine Learning Techniques. This work Utilizing machine learning for the classification of cardiovascular disease occurrence can assist diagnosticians in minimizing misdiagnosis. A model has been developed in this research that can accurately predict cardiovascular diseases, thereby reducing the fatality associated with such conditions. A method of k-modes clustering with Huang starting is proposed in this paper, aiming to enhance classification accuracy. Models including random forest (RF), decision tree classifier (DT), multilayer perceptron (MP), and XGBoost (XGB) are employed. GridSearchCV used to hypertune the parameters of the applied model, optimizing the results.

Sivakannan et.al., proposed cardiovascular diseases prediction by machine learning incorporation with deep learning. This research presents a collection of machine learning models that are to address this problem. These models take into account the data observation mechanisms and training procedures of a number of different algorithms. In order to verify the efficacy of our strategy, we combined the Heart Dataset with other classification models.

Neha et.al., proposed Machine learning-based heart attack prediction: A symptomatic heart attack prediction method and exploratory analysis. In this paper, a machine learning-based heart attack prediction (ML-HAP) method is analysed, wherein the analysis of various risk factors and prediction for heart attacks is predicted using ML approaches, including Support Vector Machines, Logistic Regression, Naïve Bayes, and XGBoost.

Kathiravan Srinivasan et al., proposed a machine learning and IOT based heart disease detection system. It discusses the development of an IoT-Cloud Based Smart Healthcare Monitoring System for Heart Disease Prediction via Deep Learning. It covered related work, methodology, experimental setup, performance assessment, experimental results and discussion, comparative analysis, future directions, and conclusions. The proposed system's performance surpasses that of previous systems, and it utilizes edge computing and fog computing to handle the immense data traffic generated by smart devices. The article also presents various models and approaches for heart disease prediction, including ensemble classifiers, fuzzy logic systems, and machine learning hybrid models.

Victor Chang et al proposed a work on smart healthcare. The work describes the construction of an artificial intelligence-based heart disease detection system using artificial intelligence and machine learning algorithms. The system uses a random forest classifier algorithm for predictive analytics. However, there is no information provided about the accuracy of the system or whether it can be applied to other types of health monitoring applications beyond heart disease detection.

Yang Yan et.al., contributed over the cardio vascular disease detection. It discusses the potential role of AI in the diagnosis and treatment of cardiovascular diseases. It highlights the current shortcomings of AI in cardiovascular medicine and the need for guidelines to protect human rights in AI research. Overall, the article suggests that AI has the potential to revolutionize the clinical application of CVD in the future.

Paul A.Friedman et.al., adapted AI techniques for the heart disease classification. It provides the information about Mayo Clinic’s resources and services, including patient-centered care, clinical trials, support groups, and educational programs for medical professionals. Mayo Clinic is a leader in the movement to bring artificial intelligence (AI) tools and technology into clinical practice to benefit people who have or are at risk of heart disease. The clinic's AI cardiology team is applying these new approaches to early risk prediction and diagnosis of serious or complex heart problems. This paper also includes sections on diseases and conditions, symptoms, tests and procedures, drugs and supplements, healthy lifestyle, books and subscriptions, and more.

**Proposed Framework**

There are so many research work and studies are provided for the heart disease classification using the machine learning and deep learning approaches. All the existing systems are using the clinical features and the patient symptoms and the health unit measures for the heart disease classification. In this proposed system we have analysed the risk factors that are highly contributing for the heart disease by various data analysis and the feature selection methodology. Each classification algorithms have its own characteristics for classification. In order to identify the best learner for the heart disease classification we have trained with some of the classifiers with the considered features and the performance study conducted in order to decide the best classifier for heart disease detection task. Once the disease detected it is very important to let know the patient that which factor contributing high risk for their disease and how to overcome the situation. For that a health care chat bot is developed to have a AI bot-patient conversation regarding the health advisories. The architecture of the proposed system is given in the figure 1.

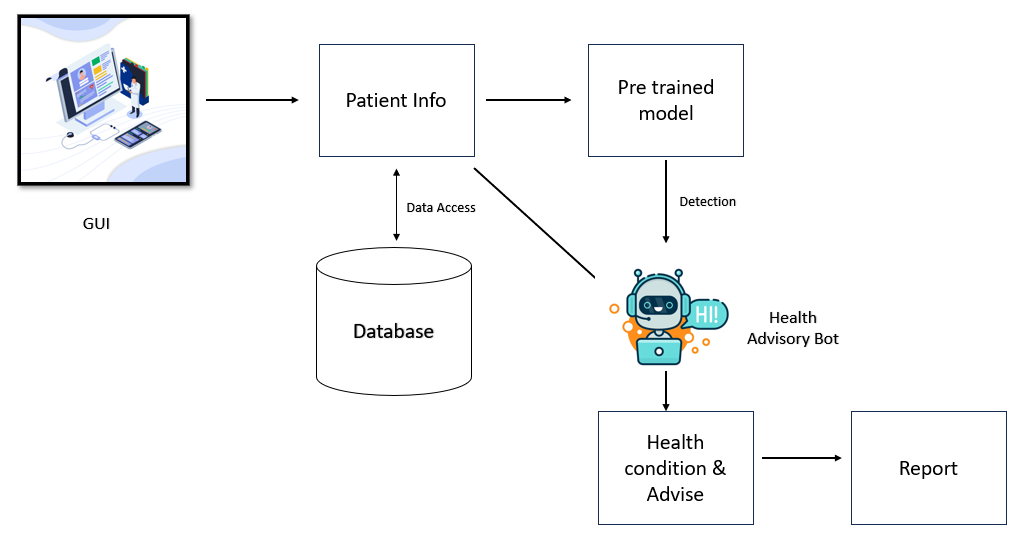


Fig 1. Heart Disease Classification System

1. **DATASET**

There are so many different datasets are available for the heart disease detection. However, the data should have the main risk factors such as blood pressure, cholesterol and smoking. The 2022 annual CDC survey data have all the important factors for the heart disease detection. The dataset contains 319795 different samples with 18 features. The table 1 describes about the number of samples in each category.

|  |  |  |
| --- | --- | --- |
| Labels | Yes | No |
| Number | 27373 | 292422 |
| Percentage | 9% | 91% |

Table 1. Dataset

1. **METHODOLOGY**

The proposed system classifies whether the patient have heart disease or not and once if heart disease is detected, the type of heart disease is identified. The project consists of 3 important modules such as

1. Data preprocessing
2. Model(s) generation and performance analysis
3. Deployment and chatbot

**Data Preprocessing**

Data preprocessing is the important phase. In order to feed into the machine learning algorithm, the dataset to be pre-processed. The dataset considered for this project did not have any null value. Some of the columns are contain the string data which is converted into numerical form by using the label encoding. There are 18078 duplicate data samples found in the dataset. The duplicate sample are dropped from the dataset. There are 18 features exists in this dataset in order to reduce the directionality the feature selection is performed using the selectKbest technique. SelectKBest employs statistical tests such as the chi-squared test, ANOVA F-test, or mutual information score to assess and rank features according to their association with the output variable. Subsequently, it chooses the top K features with the highest scores for inclusion in the ultimate feature subset.

**Model Generation**

This project utilizes the machine learning algorithms for the heart disease classification. The training dataset trained with Decision Tree (DT), Logistic Regression (LR), Random Forest (RF), K Nearest Neighbour (KNN), Gradient Boosting (GB) and the Artificial Neural Network (ANN). Once the data split into training and testing, the classifiers are fitted with the train data for train the model. In this project there are two different model to be developed, one is for heart disease detection and another one is for detecting the various heart disease means type of the heart disease. There is a lack of data in the context of detecting the different type of heart disease. Hence, we are detecting whether the patient have heart disease or not and if the heart disease is detected then the type of heart disease is identified using another model. Both models are trained with the above mentioned classifies and the best model chosen for the deployment in the heart disease detection application. The best model chosen based on the performance analysis, which is discussed in the next section.

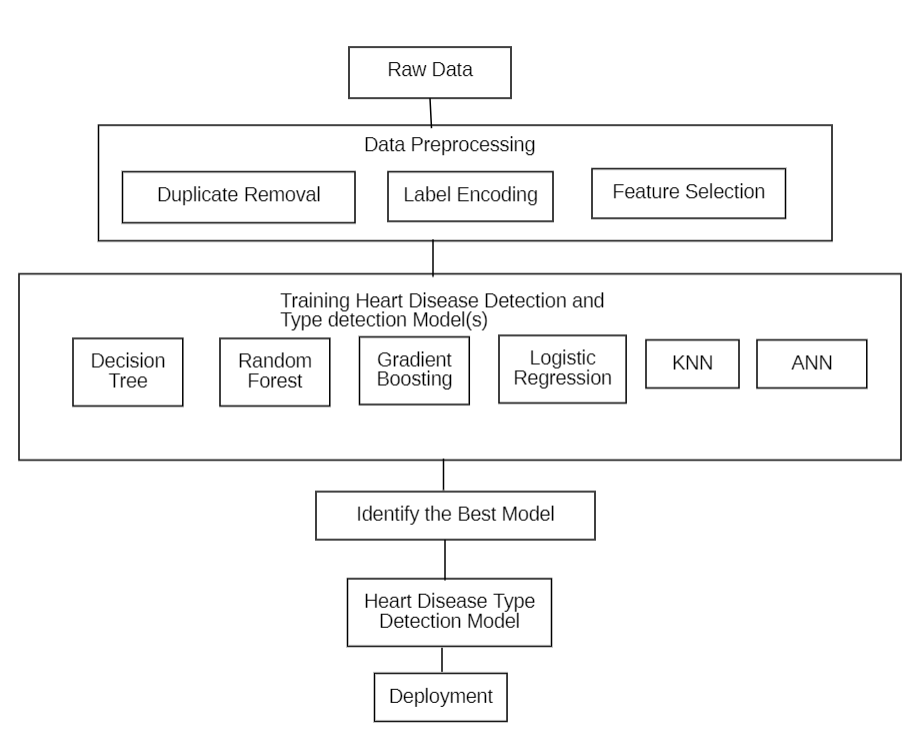


Fig 2. Flow diagram

1. **RESULT AND DISCUSSION**

**Result 01 – Tools and Libraries**

The experiments conducted in this project done in the windows 11 OS with 8GB RAM. The project developed using the Jupyter notebook. This project adapts the python libraries for the preprocessing and training tasks. The process and the used libraries are as tabulated below.

|  |  |
| --- | --- |
| Label encoding | Label encoding from preprocessing library |
| Duplicate Removal | drop\_duplicate from Pandas library |
| Feature Selection | SelectKBest, Chi2 |
| Training the classifier | Model from sklearn library |
| Performance | metrics from sklearn library |

Table2. Tools used

**Result 02 – Performance Metrics**

This project used the accuracy, precision, recall and ROC curve for the performance analysis. Once the model is trained the model evaluated with the test data and the accuracy and ROC curve is plotted for each algorithm. The accuracy gained by each algorithm is tabulated below. All the models are trained and evaluated with same set of training and test data.

|  |  |  |  |
| --- | --- | --- | --- |
| Algorithm | Accuracy | Precision | Recall |
| DT | 86 | 25 | 20 |
| LR | 91 | 50 | 74 |
| RF | 89 | 27 | 15 |
| KNN | 90 | 34 | 59 |
| GB | 91 | 50 | 76 |
| ANN | 90 | - | - |

From the above tabulated results, it is shows that the Gradient boosting algorithm outperforms well with good accuracy, precision and recall scores when compared with another models. The figure 3 describes about the Receiver operating characteristic curve (ROC) attained by each algorithm. The ROC curve is a graphical representation of the True positive and False positive rate. The ROC curve shows that the Gradient boosting algorithm attained the best true positive and false positive rates in the context of heart disease classification.

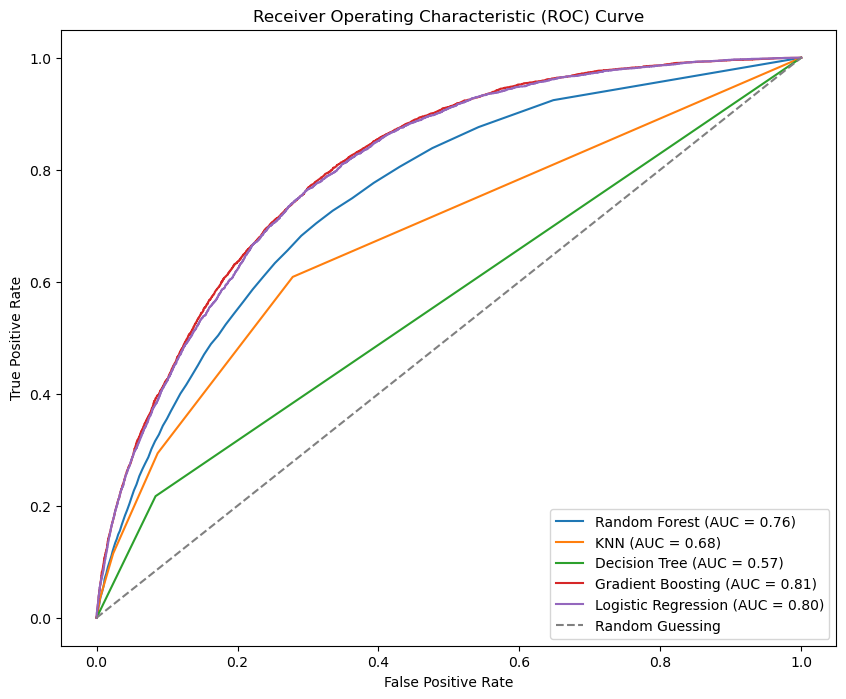


Fig 3. ROC curve

1. **CONCLUSION**

While considering the heart disease detection and the various disease type classification it is important to identify the high-risk factors that are led to heart disease. There are several factors available in the considered data however, the important factors are identified through the feature selection by verify against the class label. The machine learning models were trained and tested with same data and the results suggests that the gradient boosting algorithm attains good accuracy rate of 91% with high precision, recall scores and good ROC curve. Further the heart disease type classification model is developed in order to classify the possible type of heart disease type. The heart disease identification model deployed as the chatbot in order to provide a use friendly model utilization and the health suggestions also provided based on that. The model implemented in a web portal for the deployment.

**REFERENCES**

[1] Hossain, Md & Maruf, Mehadi & Khan, Dr. Md. Ashikur & Prity, Farida & Fatema, Sharmin & Ejaz, Md & Khan, Md. (2023). Heart disease prediction using distinct artificial intelligence techniques: performance analysis and comparison. Iran Journal of Computer Science. 6. 10.1007/s42044-023-00148-7.

[2] Bhatt, Chintan & Patel, Parth & Ghetia, Tarang & Mazzeo, Pier Luigi. (2023). Effective Heart Disease Prediction Using Machine Learning Techniques. Algorithms. 16. 88. 10.3390/a16020088.

[3] Subramani Sivakannan, Varshney Neeraj, Anand M. Vijay, Soudagar Manzoore Elahi M., Al-keridis Lamya Ahmed, Upadhyay Tarun Kumar, Alshammari Nawaf, Saeed Mohd, Subramanian Kumaran, Anbarasu Krishnan, Rohini Karunakaran

Cardiovascular diseases prediction by machine learning incorporation with deep learning, Frontiers in Medicine, Volume 10, 2023, DOI=10.3389/fmed.2023.1150933

[4] Nandal, Neha & Goel, Lipika & Tanwar, Rohit. (2022). Machine learning-based heart attack prediction: A symptomatic heart attack prediction method and exploratory analysis. F1000Research. 11. 1126. 10.12688/f1000research.123776.1.

[5] Nancy, Angel & Ravindran, Dakshanamoorthy & Vincent P M, Durai & Srinivasan, Kathiravan & Gutiérrez, Daniel. (2022). IoT-Cloud-Based Smart Healthcare Monitoring System for Heart Disease Prediction via Deep Learning. Electronics. 11. 2292. 10.3390/electronics11152292.

[6] Chang, Victor & Cao, Yi & Li, Taiyu & Shi, Yujie & Baudier, Patricia. (2019). Smart Healthcare and Ethical Issues. 53-59. 10.5220/0007737200530059.

[7] Yan, Yang & Zhang, Jia-Wen & Zang, Guang-Yao & Pu, Jun. (2019). The primary use of artificial intelligence in cardiovascular diseases: What kind of potential role does artificial intelligence play in future medicine?. Journal of geriatric cardiology : JGC. 16. 585-591. 10.11909/j.issn.1671-5411.2019.08.010.

[8] Friedman, H. S., & Booth-Kewley, S. (1987). Personality, Type A behavior, and coronary heart disease: The role of emotional expression. Journal of Personality and Social Psychology, 53(4), 783–792. https://doi.org/10.1037/0022-3514.53.4.783

[9] Akhtar and Nayab. (2021), Heart Disease Prediction, Reseasrch gate, 2021.

[10] Gupta, Neha & gandhi, kriti & Dhall, Shafali. (2020). Disease prediction using machine learning.

[11] Gandhi, Kriti. (2020). Disease Prediction using Machine Learning. International Journal for Research in Applied Science and Engineering Technology. 8. 500-507. 10.22214/ijraset.2020.6077.