Predictive Analysis of Cardiovascular Health Through Machine Learning

Prof. Takbhate T.K[1], Siddhant Amar Singh[2], Rohit Gore[3], Aniket Kshirsagar[4], Sumit Bhosle[5]

[1] Prof. Takbhate T.K, Computer Science Engineering Department, MIT College Of Railway Engineering And Research, Barshi, Maharashtra, India,
Email: tushar.takbhate@gmail.com
[2] Siddhant Amar Singh, Computer Science Engineering Department, MIT College Of Railway Engineering And Research, Barshi, Maharashtra, India,
Email: sidavsingh2902@gmail.com
[3]Rohit Prakash Gore, Computer Science Engineering Department, MIT College Of Railway Engineering And Research, Barshi, Maharashtra, India,
Email: rohitgore003@gmail.com
[4]Aniket Anil Kshirsagar, Computer Science Engineering Department, MIT College Of Railway Engineering And Research, Barshi, Maharashtra, India,
Email: aniketkshirsagar8009@gmail.com
[5]Sumit Sanjay Bhosle, Computer Science Engineering Department, MIT College Of Railway Engineering And Research, Barshi, Maharashtra, India,
Email: sumitbhosle2002@gmail.com

ABSTRACT

The "Predictive Analysis of Cardiovascular Health Through Machine Learning" project ventures into the critical realm of Healthcare. Just as financial markets endeavor to predict the future value of stocks, this project aspires to preict something even more precious – the state of one's cardiovascular health. Leveraging the power of machine learning, this undertaking explores the intricate web of factors influencing heart disease, aiming to offer early, accurate prediction and, consequently, lifesaving insights.

The abstract heartbeats with the promise of a future where healthcare becomes increasingly proactive, where personalized assessments can identify before they manifest. Much like the stock market's complexity, understanding the heart's intricate dance with various variables presents formidable challenges. Nevertheless, the otential rewards are equally great – the possibility of intervening in heart disease, a leading cause of mortality worldwide.

This projects's significance transcends the confines of algorithms and datasets; it embodies the intersection of technology and human well-being, providing a compelling narrative of using advanced data analytics and predictive modeling to save lives. Just as stock proce predictions hold the key to finacials gains, here, we unlock the potential to predict and prevent cardiovascular ailments. The project exemplifies the potency of technology to illuminate the path towards better health and longevity.

We also analyze the advantages and disadvantages of using machine learning in this context. By the end of this presentation, you will gain insights into the potential of machine learning to revolutionize heart disease diagnosis and contribute to better healthcare outcomes.

Keyword: -

Heart Disease Prediction, Machine Learning, Python, Data Preprocessing, Feature Engineering, Model Selection, Data Collection, Evaluation Metrics, Logistic Regression, Decision Trees, Random Forest, Support Vector Machines (SVM), Naive Bayes, K-Nearest Neighbors (KNN), Ensemble Learning.

1. INTRODUCTION

Heart disease refers to a range of conditions affecting the heart, often linked to issues with blood vessels, heart rhythm, or structural problems. It's broad term encompassing various conditions that impact the heart's ability to function efficiently. From coronary artery disease and heart attacks to heart failure and arrhythmias, heart disease poses a significant global health concern. Coronary artery disease, the most common type, occurs when plaque buildup narrows the arteries supplying blood to heart, leading to reduced blood flow and potential blockages. Heart attacks, a severe manifestation of this condition, result from the sudden blockage of blood flow to a part of the heart.

1.1 Traditional Risk Factors:

Numerous studies have highlighted the significance of traditional risk factors such as hypertension, diabetes, hyperlipidemia, smoking, and family history in predicting heart disease. Most prediction models incorporate these factors as primary predictors due to their well-established association with cardiovascular outcomes. However, the reliance solely on traditional risk factors may overlook other important predictors and limit the accuracy of the models.

1.2 Literature Review:

Heart disease would typically involve an extensive review of existing research, studies, and scholarly articles related to various aspects of heart disease. Due to the breadth and depth of the topic, a comprehensive literature review might cover several areas.

1.4 Pathophysiology and Mechanisms:

Delve into the physiological and molecular mechanisms underlying heart diseases like coronary artery disease, myocardial infarction, heart failure, arrhythmias, etc.
Understand the role of inflammation, lipid metabolism, genetics, and other factors contributing to the development and progression of heart disease.

1.5 Diagnostic Methods and Technologies:

Review studies on diagnostic tools and technologies used in assessing heart disease, including imaging techniques (echocardiography, MRI, CT scans), biomarkers, and genetic testing.
Evaluate the accuracy, efficiency, and advancements in diagnostic methods for different heart conditions.

1.6 Machine Learning Approaches:

In recent years, machine learning techniques have gained prominence in heart disease prediction due to their ability to handle large and complex datasets and capture nonlinear relationships. Models such as decision trees, random forests, support vector machines, and neural networks have been applied to various healthcare datasets to predict heart disease risk. These models demonstrate promising results in terms of accuracy and sensitivity, outperforming traditional risk assessment tools in some cases.

1.7 Data Source:

An Organized Dataset of individuals had been selected keeping in mind their history of heart problems and in accordance with other medical conditions. Heart disease are the diverse condition by which the heart is affecred. According to World Health Organization (WHO), the greatest number of death in middle aged people are due to Cardiovascular diseases. We take a data source which is comprised of medical history of 304 different patient of different age group. This dataset gives us the much-needed information i.e. the medical attributes such as age, resting blood pressure, fasting sugar level etc. of the patient that helps us in detecting the patient that is diagnosed with any heart disease or not. This dataset contains 13 medical attributes of 304 patients that helps us detecting if the patient is at risk of getting a heart disease or not and it helps us classify patient that are at risk of having a heart disease and that who are not at risk. This heart Disease dataset is taken from the UCI repository. According to ths dataset the pattern which leads to the detection of patient prone to getting a heart diseaseis extracted. All attributes are listed in 'Table1'

Table 1. Various Attributes used are listed

S. No	Observation	Description	Values
1.	Age	Age in Years	Continuous
2.	Sex	Sex of Subject	Male/Female
3.	CP	Chest Pain	Four Types
4.	Trestbps	Resting Blood Pressure	Continuous
5.	Chol	Serum Cholesterol	Continuous
6.	FBS	Fasting Blood Sugar	< ,or > 120 mg/dl
7.	Restecg	Resting Electrocardiograph	Five Values
8.	Thalach	Maximum Heart Rate Achieved	Continuous
9.	Exang	Exercise Induced Angina	Yes/No
10.	Oldpeak	ST Depression when Workout compared to the Amount of Rest Taken	Continuous
11.	Slope	Slope of Peak Exercise ST segment	up/ Flat /Down
12.	Ca	Gives the number of Major Vessels Coloured by Fluoroscopy	0-3
13.	Thal	Defect Type	Reversible/Fixed/Normal
14.	Num(Disorder)	Heart Disease	Not Present /Present in the Four Major types.

2. Methodology

Utilize publicly available datasets such as the Framingham Heart Study dataset or the Cleveland Heart Disease dataset, handle missing values through imputation techniques such as mean imputation or predictive modeling.
Create new features from existing ones, such as BMI (Body Mass Index) from height and weight, extract temporal features from time-series data, such as average heart rate over time. Experiment with a variety of machine learning algorithms suitable for classification tasks, such as logistic regression, decision trees, random forests, support vector machines, and neural networks.

2.1 Project Planning:

Define Project Goals and Objectives: Clearly articulate the goals and objectives of the project, such as developing a machine learning model to predict heart disease risk accurately and early.

Create a Project Timeline: Develop a detailed timeline that outlines key project milestones, tasks, and deadlines. Break down the project into manageable phases, such as data collection, preprocessing, model development, evaluation, and deployment.

Resource Allocation: Identify the resources needed to execute the project successfully, including human resources, computational resources, and data sources. Allocate resources effectively to ensure that project tasks are completed on time and within budget.

Communication and Collaboration: Establish clear communication channels and protocols for team members to collaborate effectively. Schedule regular meetings and status updates to review progress, discuss challenges, and make decisions.

2.2 System Design:

Architecture Design: Define the overall architecture of the system, including data flow, processing steps, and model deployment. Determine how data will be collected, preprocessed, and fed into the machine learning model.

Model Selection and Integration: Select appropriate machine learning algorithms for heart disease prediction based on the characteristics of the data and the project goals. Integrate the selected model(s) into the system architecture, ensuring compatibility with data sources and processing pipelines.

Scalability and Performance: Consider scalability requirements to handle large volumes of data and accommodate future growth. Optimize system performance to minimize latency and maximize throughput, especially for real-time prediction scenarios.

2.3 Development:

Data Collection and Preprocessing: Collect relevant datasets containing patient demographics, medical history, diagnostic tests, and other relevant features. Clean the data by handling missing values, removing duplicates, and addressing outliers.

Feature Engineering: Engineer new features or transform existing ones to capture relevant information for heart disease prediction. Utilize domain knowledge and medical expertise to identify informative features that may improve model performance.

Model Selection and Training: Experiment with different machine learning algorithms suitable for classification tasks, such as logistic regression, decision trees, random forests, support vector machines, and neural networks.

2.4 Testing:

Validation on Hold-Out Test Set: Reserve a portion of the collected data as a hold-out test set, separate from the data used for training and validation. Use this test set to evaluate the final trained model's performance on unseen data and assess its generalization ability.

Cross-Validation Techniques: Employ cross-validation techniques such as k-fold cross-validation to assess the model's robustness and variability. Divide the dataset into k subsets (folds), train the model on k-1 folds, and evaluate it on the remaining fold. Repeat this process k times, rotating the validation fold each time.

External Validation: Validate the trained model externally with independent datasets from different sources or healthcare institutions. Collaborate with healthcare professionals or domain experts to validate the model's clinical relevance and applicability.

2.4 Testing:

Integration into Clinical Workflow: Deploy the trained machine learning model into the clinical workflow of healthcare institutions, ensuring seamless integration with existing systems and processes. Collaborate with healthcare professionals to identify the most suitable points in the workflow for model deployment, such as during patient consultations or diagnostic screenings.Scalability and Performance Optimization: Optimize the deployed model for scalability and performance to handle real-time predictions and accommodate increased demand as the system scales. Utilize cloud-based infrastructure or distributed computing resources to support concurrent requests and minimize latency.

Monitoring and Maintenance: Establish monitoring mechanisms to track the performance of the deployed model over time and detect any anomalies or deviations from expected behavior. Monitor key metrics such as prediction accuracy, model drift, and system latency to ensure ongoing reliability and effectiveness.

References	Findings	Method used	Results obtained
Ramprakash et al. (2)	Prediction of heart disease on Cleveland dataset	Machine learning algorithms—Deep Neural Network and χ^2 —a statistical method	Accuracy—94%
Mohan et al. (6)	Prediction of heart disease on Cleveland dataset	Proposed HRFLM model using Random Forest, and Linear model	Proposed model accuracy—88%
Sowmiya and Sumitra (19)	Used Cleveland dataset to implement heart disease prediction	Used ant colony optimization technique for hybrid KNN classifier and compared with other ML models	Proposed approach HKNN proves efficient and effective technique
Made et al. (34)	Estimation of 10-year coronary heart risk	Using statistical approach—regression analysis in age group of 18 and above	Study estimates heart risk for the next 10-years
Proposed work	Detection and classification of cardiovascular diseases in different age-based group and gender-based people, collected vital signs of an individual using wearables and medical equipments	Used 6 Machine Learning algorithm—Decision Tree, SVM, ANN, RFC, XGBoost, and NB. Further, comparison is made among 6 classifiers. Also, we developed two models based on Kaplan-Meier method (For Gender-Based) and Cox-Regression Proportional Model (For Age-Based) using scoresheet	Among 6 Machine learning models-Decision Tree achieves maximum accuracy of 98%. Our proposed approach, for age-based and gender-based shows risk score (in percentage) of cardiac arrest for the next 10-years

Heart Disease Prediction System (EHDPS) has been developed using different classifiers. This model uses 13 medical parameter such as chest pain, fasting sugar, blood pressure, cholesterol, age, sex, etc.

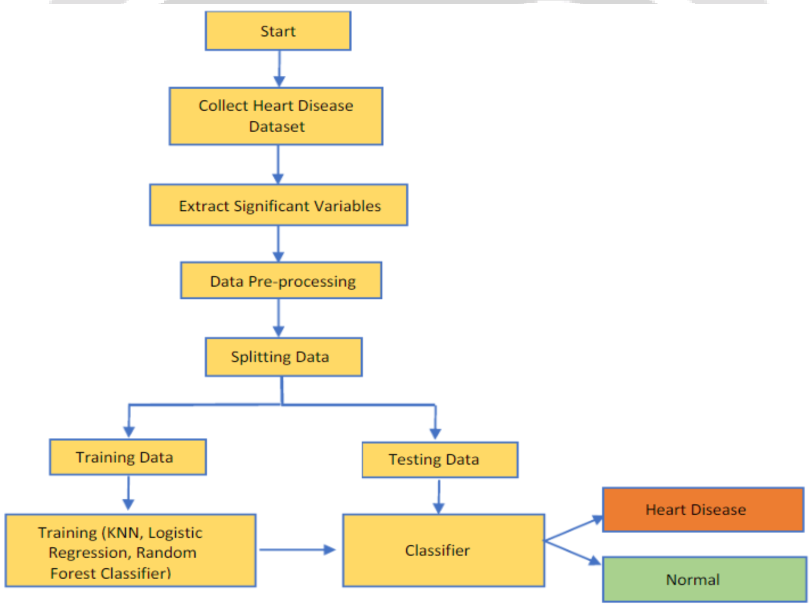


Figure 1. Proposed Model

3. Result & Discutions

From these result we can see that although most of the researchers are using different algorithms such as SVC, decision tree for the detection of patient diagnosed with Heart Disease, KNN, Random Forest Classifier and Logistic regression yield a better result to out rule them. The algorithms that we used are more accurate, saves a lot of money i.e. it is cost efficient and faster than the previous researches used. Moreover, the maximum accuracy obtained by KNN and Logistic Regression are equal to 88.5% which is greater or almost equal to accuracies obtained from previous researches. So, we summarize that our accuracy is improved due to the increased medical attributes that we used from the dataset we took. Our project also tells us that us that Logistic Regression and KNN outperforms Random Forest classifier in the prediction of the patient diagnosed with a heart disease. This proves that KNN and Logistic Regression are better in diagnosis pf a heart Disease. The following 'figure 2','figure 3','figure 4','figure 5' shows a plot of the number of patient that are been segregated and predicted by the classifier depending upon the age group, Resting Blood Pressure, Sex, Chest pain:

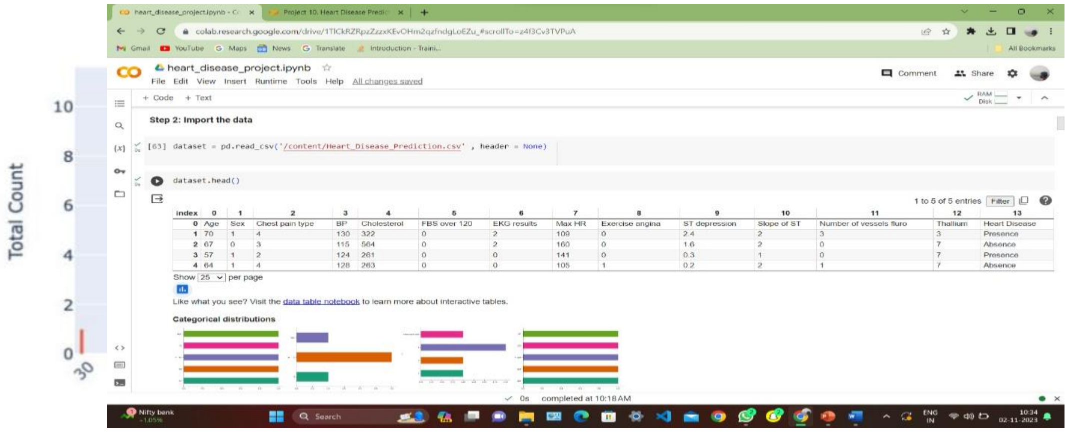
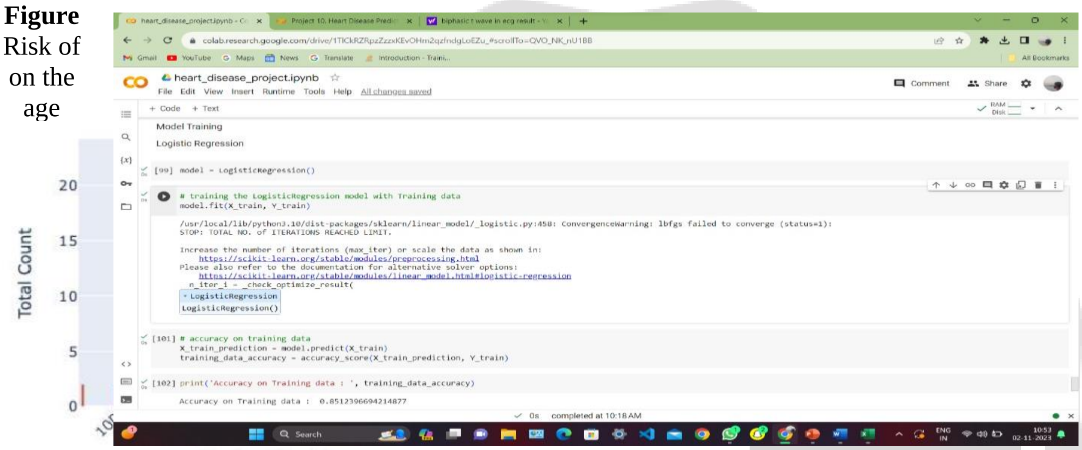


Figure 2. Shows the Risk of Heart Attack on the basis of their age



Resting Blood Pressure

Figure 3. Shows the Risk of Heart Attack on the basis of their Resting Blood Pressure

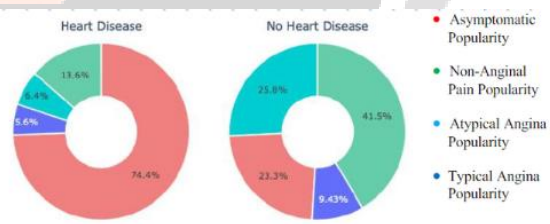
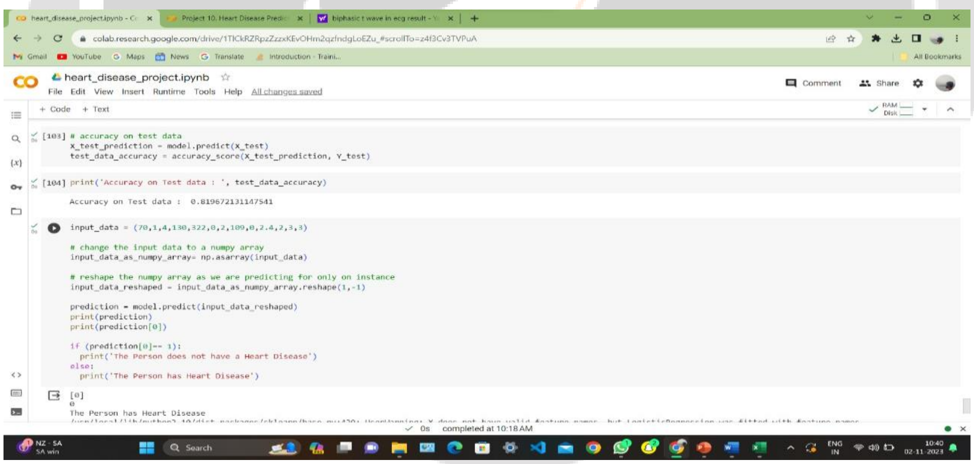
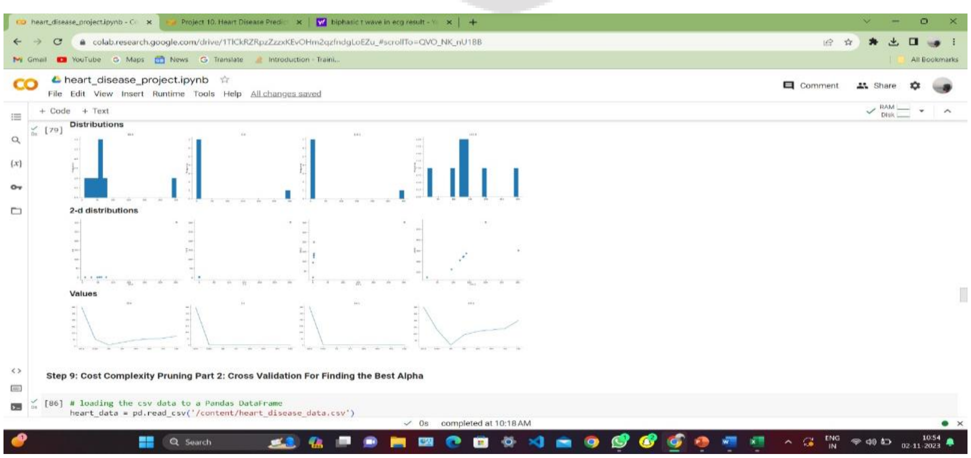


Figure 4. Shows patient. having or not having HD Figure 5. Shows Chest Pain basis of having or not having HD





4.Future Scope

The future of heart disease prediction using machine learning is exceptionally promising. With ongoing advancements in data collection, algorithm development, and computational power, we anticipate significant strides in accuracy, reliability, and accessibility of predictive models.

4.1 Precision Medicine:

Machine learning algorithms will increasingly tailor predictions to individual patients, considering a comprehensive range of factors including genetics, lifestyle, medical history, and environmental influences. This personalized approach will enhance risk assessment and enable targeted interventions, leading to better health outcomes.

4.2 Early Detection and Prevention:

Machine learning models will continue to improve in their ability to detect subtle patterns and early indicators of heart disease, enabling healthcare providers to intervene proactively. By identifying at-risk individuals before symptoms manifest, we can implement preventive strategies and lifestyle modifications to mitigate risk factors and prevent adverse cardiac events.

4.3 Integration with Healthcare Systems:

As machine learning algorithms become more robust and validated, we anticipate seamless integration into existing healthcare systems. Clinicians will have access to user-friendly tools that leverage patient data to generate actionable insights in real-time, facilitating informed decision-making and personalized patient care.

4.4 Empowering Patients:

Machine learning-powered predictive tools will empower patients to take an active role in managing their heart health. Through user-friendly interfaces and personalized risk assessments, individuals will gain valuable insights into their cardiovascular risk factors, enabling them to make informed decisions about their lifestyle, treatment options, and preventive measures.

4.5 Collaborative Research and Innovation:

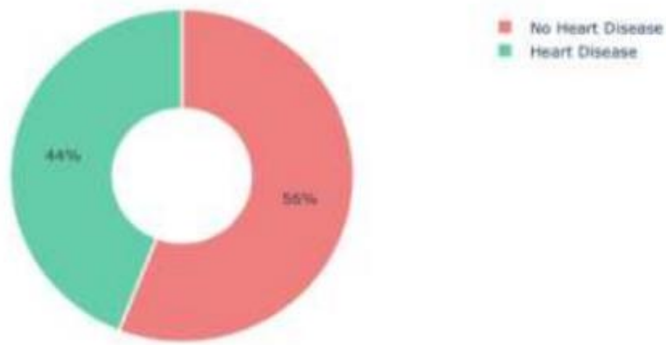
Collaboration between multidisciplinary teams of researchers, clinicians, data scientists, and industry partners will drive innovation in heart disease prediction. By sharing expertise, resources, and data, we can accelerate progress, validate predictive models, and ensure their scalability and generalizability across diverse populations.

4.6 Ethical and Regulatory Considerations:

As with any healthcare technology, it's crucial to prioritize ethical principles, patient privacy, and regulatory compliance. Transparency, accountability, and equity must guide the development and deployment of machine learning algorithms for heart disease prediction, ensuring that they benefit all individuals regardless of demographic or socioeconomic factors.

4. CONCLUSION

A cardiovascular disease detection model has been developed using three ML classifications modeling techniques. This project predicts people with cardiovascular disease by extracting the patient medical history that leads ti a fatal heart disease from a dataset that includes patient's medical history such as chest pain, sugar level, blood pressure, etc. This Heart Disease detection assists a patient based on his/her clinical information of them been diagnosed, Random Forest Classifier. The accuracy of our model is 87.5% Use of more training data ensures the higher chances of the model to accurately predict whether the given person has a heart disease or not. By using these, computer aided technique we can predict the patient fast and the cost can be reduced very much. There are a number of medical database that we can work on as these Machine learning techniques are better and they can predict better than a human being which helps the patient as well as the doctors. Therefore, in conclusion this project helps us predict the patient who are diagnosed with heart disease by cleaning the dataset and applying logistic regression to get an accuracy of an average of 85% on our model which is better than the previous models having an accuracy of 85% also, it is concluded that accuracy of KNN highest between the three algorithm that we have used i.e.88.52%. shows 44% of people that are listed in the dataset are suffering from Heart Disease.



5. REFERENCES

[1]. Soumya Sahoo, Sukant Kishoro Bisoy, Pradeep Kumar Mallick, "Cardiac Disease Detection Using Machine Learning Algorithms: A Review", 2024 International Conference on Advancements in Smart, Secure and Intelligent Computing (ASSIC), pp.1-5, 2024.

[2]. Amruta Amune, Shraddha Chavanke, Shushwet Joshi, Shubham Kuttarmare, Shubham Kathane, "Integrated Healthcare System for Saline Monitoring, Patient Communication, and Heart Disease Prediction Using IoT and Machine Learning Algorithms", 2024 International Conference on Emerging Smart Computing and Informatics (ESCI), pp.1-6, 2024.

[3]. Manika Manwal, Tanuj Maithani, Sristhi Mall, Kamlesh Chandra Purohit, Tanupriya Choudhury, "Cardiovascular Disease and Diabetes Disease Prediction Using Machine Learning and Streamlit API", 2023 12th International Conference on System Modeling & Advancement in Research Trends (SMART), pp.543-549, 2023.

[4]. Md. Tanziar Rahman, Md. Shake Farid Uddin, Mosaddek Adnan Sikder, "Heart Disease Prediction using Supervised Learning Classifiers", 2023 26th International Conference on Computer and Information Technology (ICCIT), pp.1-5, 2023.

[5]. Sai Sri Surya Yandamuri, Jodu Manjula, Vinay Varma R, Vetrivelan P, Ajeyprasaath KB, "Remote Vital-Sign Monitoring for Heart-Attack Patients in Telemedicine", 2023 Innovations in Power and Advanced Computing Technologies (i-PACT), pp.1-7, 2023.

[6]. Akash Rajak, Rabi Narayan Panda, "Prediction of Heart Disease from the Multiple Features Available in the Database by Applying Classification Models of Machine Learning", 2023 3rd International conference on Artificial Intelligence and Signal Processing (AISP), pp.1-4, 2023.

[7]. S. Hemalatha, T. Kavitha, D. Niruba, S. Nandhakumar, R. Venkatesh, "Comparative Performance Assessment Of Machine Learning Algorithms To Predict Cardiovascular Disease", 2023 International Conference on Computer Communication and Informatics (ICCCI), pp.1-9, 2023.

[8]. Islam Daoud Suliman, D. Vasumathi, "Prediction of Heart Disease Using Machine Learning Algorithms", 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), pp.1-12, 2023.

[9]. Anshul Pandey, Mrityunjay Prasad Gupta, Manoj Diwakar, Sarishma Dangi, Parul Madan Prabhishek Singh, "An Effective Machine Learning based Heart Disease Diagnosis Analysis", 2023 International Conference on Computational Intelligence, Communication Technology and Networking (CICTN), pp.536-540, 2023.

[10]. Akash Rajak, Rabi Narayan Panda, "Prediction of Heart Disease from the Multiple Features Available in the Database by Applying Classification Models of Machine Learning", 2023 3rd International conference on Artificial Intelligence and Signal Processing (AISP), pp.1-4, 2023.

[11]. Jay Raval, Jai Prakash Verma, Sardar N. M. Islam, Rachna Jain, Narina Thakur, "AI Based Prediction for Heart Disease: A Comparative Analysis and an Improved Machine Learning Approach", 2022 6th Asian Conference on Artificial Intelligence Technology (ACAIT), pp.1-9, 2022.

[12]. Senthil G. A, R. Prabha, M Razmah, T. Veeramakali, Sridevi S, Yashini R, "Machine Learning Heart Disease Prediction Using KNN and RTC Algorithm", 2022 International Conference on Power, Energy, Control and Transmission Systems (ICPECTS), pp.1-5, 2022.

[13]. E. M. H. I. Ekanayaka, P. P. N. V. Kumara, "Machine Learning Approach for New Crop Disease Predict and Alert System: A Review", 2022 International Conference on Emerging Techniques in Computational Intelligence (ICETCI), pp.45-49, 2022.

[14]. Syeda Urwa Warsi, Saba Mohsin, Muhammad Asif, Arfa Hassan, Rabia Khan, Tahir Alyas, "A Hybrid Approach for Heart Disease Prediction using Genetic Algorithm and SVM", 2024 5th International Conference on Advancements in Computational Sciences (ICACS), pp.1-6, 2024.

[15]. Deepali Yewale, S.P. Vijayaragavan, "Data-Driven Insights: A Genetic Algorithm Feature Optimization Approach to Heart Disease Prediction", 2024 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE), pp.1-6, 2024.

[16]. Kireet Joshi, Vishan Kumar Gupta, Paras Jain, Anurag Shukla, Monika Bharti, Himanshu Jindal, "Disease Prediction System in Human Beings using Machine Learning Approaches", 2024 Fourth International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT), pp.1-5, 2024.

[17]. Arti Badhoutiya, Rajesh Prasad Verma, Anurag Shrivastava, K Laxminarayanamma, A L N Rao, Akhilesh Kumar Khan, "Random Forest Classification in Healthcare Decision Support for Disease Diagnosis", 2023 International Conference on Artificial Intelligence for Innovations in Healthcare Industries (ICAIIHI), vol.1, pp.1-7, 2023.

[18]. Rushita Gandham, Keerthi Reddy Manambakam, Sai Venkat Naveen Madala, Navya Sri Nannapaneni, Srilatha Tokala, Murali Krishna Enduri, "Predictive Modeling for Heart Disease Detection with Machine Learning", 2023 IEEE 15th International Conference on Computational Intelligence and Communication Networks (CICN), pp.325-329, 2023.

[19]. Shawlin Shadat Aishee, Sayma Khorshed Chowdhury, Fariha Mahmud, Sayma Alam Suha, "Design and Development of an Ensemble Feature Selection and Machine Learning Based Cardiac Disease Prediction System", 2023 26th International Conference on Computer and Information Technology (ICCIT), pp.1-6, 2023.

[20]. Santosh Kumar Satapathy, Harsh Makadiya, Mohammed Hamza Malik, Manav Dhami, Dhrupad Chauhan, Kunj Manavadariya, "Heart Disease: Automatic Prediction from the Numerical and Categorical Features by Machine Learning Methods", 2023 IEEE 8th International Conference for Convergence in Technology (I2CT), pp.1-6, 2023.