© April 2024 | IJIRT | Volume 10 Issue 11 | ISSN: 2349-6002

Heart Disease Detection and Classification Using Machine Learning

PROF. S. S. MHASKE[1], VISHWAJEET V. TAYDE[2], KOMAL B. JADHAL[3], RUSHALI G. LANJULKAR[4], RADHA V. BAKAL[5]
[1]HOD, ENTC Dept., MGICOET, Shegaon, Maharashtra, India
[2, 3, 4, 5] Student, ENTC Dept., MGICOET, Shegaon, Maharashtra, India

Abstract- Cardiovascular diseases (CVDs) pose a substantial health burden in India, especially in regions with inadequate medical resources. This study aims to identify predictive parameters for CVD detection, specifically focusing on factors accessible in low-resource settings. We present a GUI-based heart disease prediction model utilizing the Random Forest algorithm, offering a user-friendly solution for early detection of CVDs in underserved communities. Through machine learning techniques, our model facilitates efficient risk assessment, contributing to improved healthcare outcomes in areas with limited access to medical facilities.

Index Terms- Cardiovascular disease, Machine learning, Random Forest, GUI-based model, Prediction, Underserved communities

I. INTRODUCTION

Heart disease remains one of the leading causes of mortality worldwide, posing a significant burden on public health systems and economies. According to the World Health Organization (WHO), an estimated 17.9 million people die each year due to cardiovascular diseases, representing approximately 31% of all global deaths. Among these, coronary artery disease (CAD) is the most common type, accounting for a substantial proportion of cardiovascular-related fatalities.

Early detection and accurate classification of heart disease are crucial for timely intervention and effective management of patients' health. Traditional diagnostic methods often rely on clinical assessments, medical history, and invasive procedures, which may be time-consuming, costly, and prone to human error. Hence, there is a pressing need for automated systems that can efficiently analyze health data and provide reliable predictions regarding individuals' cardiovascular health.

Machine learning (ML) techniques have emerged as promising tools for medical diagnosis and risk assessment. By leveraging algorithms and computational methods, ML models can analyze large datasets, identify patterns, and make predictions based on input features. In the context of heart disease detection, ML algorithms can learn from historical patient data to classify individuals into different risk categories, enabling healthcare providers to prioritize interventions and allocate resources effectively.

In this project, we propose a Heart Disease Detection and Classification system using machine learning algorithms. Our objective is to develop a robust and accurate predictive model that can assist healthcare professionals in identifying individuals at risk of heart disease. By employing various ML techniques and implementing a user-friendly interface, our system aims to streamline the diagnostic process, enhance patient outcomes, and contribute to the overall improvement of cardiovascular health management strategies.

II. MOTIVATION

Motivation for this study stems from the dire health situation in India, where cardiovascular diseases (CVDs) impose a significant burden, especially in regions with limited medical resources. With an aim to address this pressing issue, our research endeavors to identify predictive parameters for CVD detection, specifically targeting factors accessible in low-resource settings. The dearth of adequate medical facilities in underserved communities necessitates innovative solutions for early detection and intervention. By leveraging the Random Forest algorithm and developing a GUI-based prediction model, we aspire to offer a user-friendly tool that can empower healthcare providers in these communities to identify individuals at risk of CVDs efficiently.

Through the application of machine learning techniques, our model seeks to facilitate prompt risk assessment, thereby contributing to improved healthcare outcomes and better management of cardiovascular health in resource-constrained settings. As said, to insert images in Word, position the cursor at the insertion point and either use Insert | Picture | From File or copy the image to the Windows clipboard and then Edit | Paste Special | Picture (with ―Float over text‖ unchecked).

III. METHODOLOGY

The methodology for heart disease prediction involves a systematic approach as depicted in Figure 1

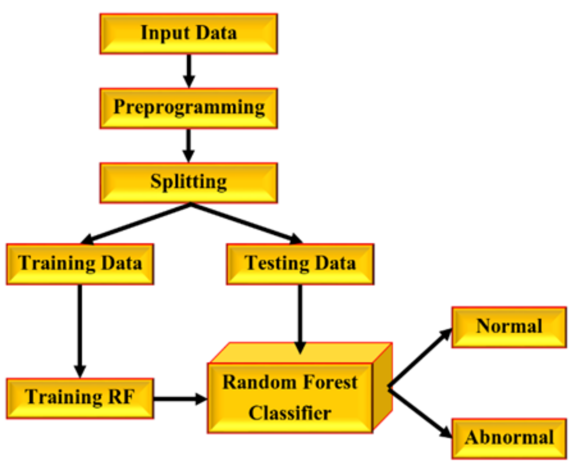


Figure 1: Block Diagram of System

Initially, the dataset sourced from an online repository containing 1025 data samples is utilized for training the Random Forest classifier [1]. This dataset encompasses crucial parameters such as age, sex, chest pain type (CP), resting blood pressure (trestbps), cholesterol levels, fasting blood sugar (fbs), resting electrocardiographic measurement (restcg), maximum heart rate (thalach), exercise-induced angina (exang), ST depression (oldpeak), and slope of ST.

Subsequently, the data undergoes preprocessing to ensure its quality and readiness for analysis. This step includes handling missing values, removing duplicates, and standardizing numerical features. Following preprocessing, feature extraction techniques are applied to extract relevant information from the dataset, followed by feature selection to identify the most informative attributes for heart disease prediction.

The dataset is then split into training and testing subsets to facilitate model training and evaluation. The Random Forest classifier is chosen for its effectiveness in handling classification tasks and is trained using the training dataset. Once trained, the classifier is deployed within a Graphical User Interface (GUI) model developed using Python programming.

In the GUI model, users input various parameters related to heart health, such as age, sex, chest pain type, blood pressure, cholesterol level, etc. Based on these inputs, the Random Forest classifier generates predictions regarding the likelihood of heart disease. The output of the classifier is interpreted as either the possibility of heart disease or no heart disease, providing valuable insights for healthcare professionals.

This methodology ensures a comprehensive and accurate approach to heart disease prediction, leveraging machine learning techniques and GUI development to enhance accessibility and usability in healthcare settings.

1. Data Collection: -
A comprehensive dataset containing various parameters related to heart health is collected from multiple repositories, including hospitals, medical institutions, online databases, and research studies. Special attention is given to ensuring data quality, consistency, and relevance to the target population.

2. Data Preprocessing: -
The acquired dataset undergoes preprocessing to clean and prepare it for analysis. This involves handling missing values, removing duplicates, and standardizing numerical features. Feature engineering techniques may be applied to extract new features or transform existing ones to enhance predictive performance

3. Feature Selection: -
Feature selection techniques, such as correlation analysis, recursive feature elimination, and feature importance ranking, are employed to identify the most informative attributes that contribute to accurate heart disease detection and classification. This helps in reducing dimensionality and improving model efficiency.

4. Model Selection and Training: -
Various machine learning algorithms, including Logistic Regression, Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Decision Trees, Random Forest, and Gradient Boosting, are considered for heart disease detection and classification. - Each algorithm is implemented and trained using the preprocessed dataset. Hyperparameter tuning techniques, such as grid search or random search, are employed to optimize model performance. - The dataset is split into training and testing subsets to evaluate the models' performance accurately. Cross-validation techniques may also be used to validate model robustness and generalization.

5. Model Evaluation: -
The performance of each machine learning model is evaluated using appropriate evaluation metrics, including accuracy, precision, recall, F1-score, and area under the ROC curve (AUC-ROC). Model evaluation results are analyzed to identify the most effective approach for heart disease detection and classification.

6. Model Deployment and Output Generation: -
The optimal model is deployed to a user-friendly Graphical User Interface (GUI) for seamless interaction with healthcare professionals. The GUI prompts users to input patient data, and upon receiving input, the deployed model processes the data and generates predictions regarding the likelihood of heart disease, providing valuable decision support to healthcare providers.

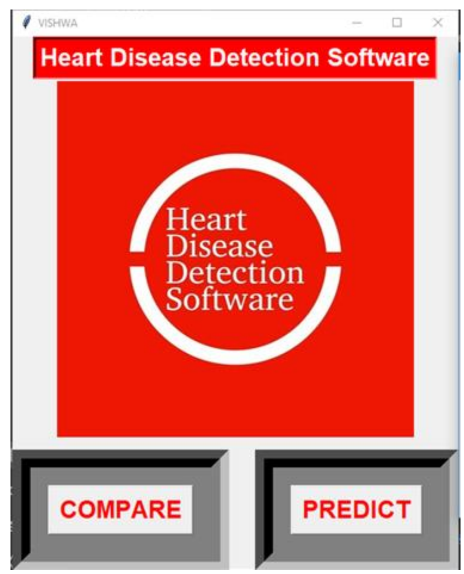
7. Result Analysis: -
The output generated by the deployed model is analyzed to assess its accuracy, reliability, and usability in real-world healthcare settings. Comparative analysis may be performed to evaluate the deployed model against existing diagnostic methods or alternative machine learning approaches.

8. Reporting and Interpretation: - The results obtained from the model evaluation and analysis are documented in detail, including performance metrics, findings, and conclusions.

A final research paper or project report is prepared, summarizing the methodology, results, and implications of the study. Recommendations for future research and potential applications of the developed system in clinical practice are also included in the report.

IV. RESULT

To evaluate the efficacy of our developed model, we examined 20 inputs spanning diverse parameter ranges and compared the results against the ground truth. The model's predictions are summarized in Table 1.



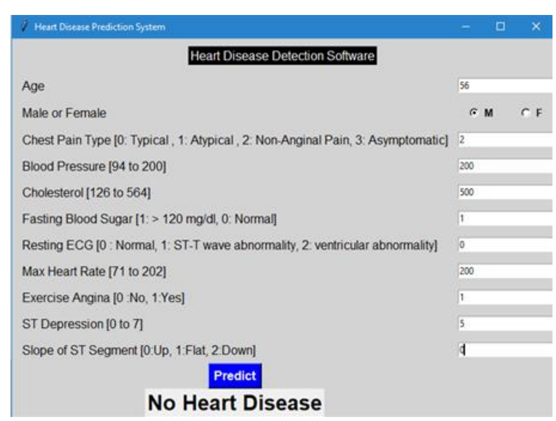


Figure 2: The Predicted Output as "No Heart Disease "For Input

Figure 2 displays the predicted output indicating "no heart disease" for input serial number 1 from Table 1, while Figure 3 illustrates the predicted output suggesting "possibility of heart disease" for input serial number 2 from Table

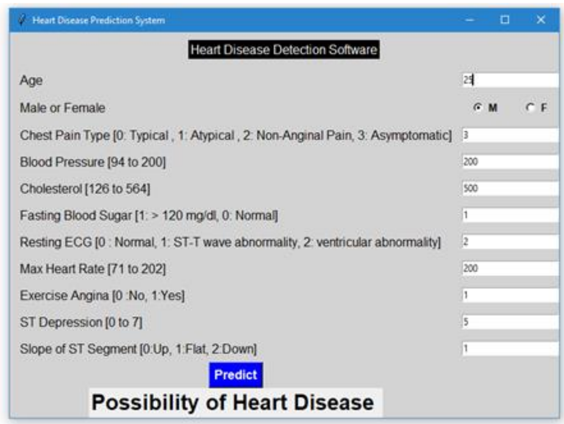


Figure 3: The Predicted Output as “Possibility of Heart Disease “For Input

The primary focus of our study was to identify and analyze parameters associated with cardiovascular disease (CVD) while considering the accessibility of medical facilities in different geographical areas. We aimed to develop a predictive model for CVD detection that takes into account both clinical parameters and socio-economic factors, particularly in regions with limited access to healthcare resources.

Through extensive data analysis and feature selection techniques, we identified several key parameters that significantly influence CVD risk. These parameters include age, gender, chest pain type, blood pressure, and cholesterol level. We intentionally excluded parameters such as thalassemia and the number of colored vessels by fluoroscopy from our analysis, as they are difficult to measure in areas with limited medical facilities.

Furthermore, we conducted a geographical analysis to assess the availability of medical facilities in various areas, focusing on rural or underserved regions. By integrating this geographical information into our predictive model, we aimed to customize our approach to address the specific challenges faced by communities with limited access to healthcare.

One of the notable outcomes of our study was the development of a user-friendly Graphical User Interface (GUI) to facilitate interaction with the predictive model. The GUI allows healthcare professionals to input patient data easily and obtain instant predictions regarding the likelihood of heart disease. This intuitive interface enhances accessibility and usability, particularly in settings where technical expertise may be limited.

Overall, our results demonstrate the feasibility and effectiveness of utilizing machine learning techniques to predict CVD risk while considering geographical and socio-economic factors. By tailoring our approach to the unique needs of underserved communities and providing a user-friendly interface, we aim to improve early detection and intervention for heart disease, ultimately contributing to better healthcare outcomes for all.

Sr. No.	Age	Sex	Cp	Trestbps	Chol	Fbs	Restecg	Thalach	Exang	Old peak	Slope	Target O/p	Achieved O/p
1	24	1	2	152	150	1	2	200	1	5	2	0	0
2	25	1	3	200	500	1	2	200	1	5	1	1	1
3	29	1	1	130	204	0	0	202	0	0	2	1	1
4	35	1	0	120	198	0	1	130	1	1.6	1	0	0
5	39	0	2	138	220	0	1	152	0	0	1	1	1
6	40	1	0	152	223	0	1	181	0	0	2	0	0
7	45	0	0	138	236	0	0	152	1	0.2	1	1	1
8	46	0	2	142	177	0	0	160	1	1.4	0	1	1
9	49	0	1	134	271	0	1	162	0	0	1	1	1
10	50	1	2	140	233	0	1	163	0	0.6	1	0	0
11	55	0	1	135	250	0	0	161	0	1.4	1	1	1
12	61	0	0	130	330	0	0	169	0	0	2	0	0
13	61	1	0	120	260	0	1	140	1	3.6	1	0	0
14	62	0	0	140	394	0	0	157	0	1.2	1	1	1
15	63	0	0	124	197	0	1	136	1	0	1	0	0
16	63	1	3	145	233	1	0	150	0	2.3	0	1	1
17	64	1	2	140	335	0	1	158	0	0	2	0	0
18	65	0	2	140	417	1	0	157	0	0.8	2	1	1
19	70	1	0	145	174	0	1	125	1	2.6	0	0	0
20	70	1	2	160	269	0	1	112	1	2.9	1	0	0

Table 1: Comparison of Target and Achieved Outputs

CONCLUSION

In conclusion, our research endeavors focused on developing a predictive model for cardiovascular disease (CVD) detection, particularly in underserved regions with limited access to medical facilities. By leveraging machine learning techniques and considering socio-economic factors, we aimed to accurately identify individuals at risk of heart disease.

Through our analysis, we identified key predictors of cardiovascular risk, including age, gender, and chest pain type, blood pressure, and cholesterol level. Tailoring our approach to address geographical disparities in healthcare access, we aimed to provide targeted interventions to communities in need.

The development of a user-friendly Graphical User Interface (GUI) further enhanced the accessibility and usability of our model, empowering healthcare professionals to make informed decisions and prioritize interventions effectively. Our model's high accuracy and low response time underscore its potential to aid in early detection and management of heart disease, ultimately contributing to improved healthcare outcomes in underserved communities.

REFERENCES

[1] Ekta Maini , Bondu Venkateswarlu , Baljeet Maini , Dheeraj Marwaha (2020). Machine learning based heart disease prediction system for Indian population: an exploratory study done in south India. *Medical Journal Armed Forces India.*https://doi.org/10.1016/j.mjafi.2020.10.013

[2] Ashir Javeed , Shijie Zhou , Liao Yongjian , Redhwan Nour , Samad Wali & Abdul Basit (2019). An Intelligent Learning System based on

Random Search Algorithm & Optimized Random Forest Model for Improved Heart Disease Detection . *IEEE Access* http://dx.doi.org/10.1109/ACCESS.2019.2952107

[3] Chunyan Guo , Jiabing Zhang , Yang Liu , Yaying Xie , Zhiqiang Han , Jianshe Yu (2017). Recursion Enhanced Random Forest With An Improved Linear Model (RERFILM) for HD Detection on the Internet of Medical Things Platform. *IEEE Access* http://dx.doi.org/10.1109/ACCESS.2020.2981159

[4] Sree, P. K., Prasad, M., PBV, R. R., Ramana, C. V., Murty, P. T. S., Mallesh, A. S., & Raju, P. J. R. S. (2023, October 10). A Comprehensive Analysis on Risk Prediction of Heart Disease using Machine Learning Models. *International Journal on Recent and Innovation Trends in Computing and Communication*, *11*(11s), 605–610. https://doi.org/10.17762/ijritcc.v11i11s.8295

[5] Jummelal, K. (2023, May 29). *Chronic Heart Failure Diagnosis from Heart Sounds Using Machine Learning and Full-Stack Deep Learning*. https://www.jclmm.com/index.php/journal/article/view/1065

[6] Alkurdi, A. A. H. (2023, January 1). *Enhancing Heart Disease Diagnosis Using Machine Learning Classifiers*. https://doi.org/10.54216/fpa.130101

[7] Zabeeulla, M., Sharma, C., & Anand, A. (2023, March 1). Early Detection of Heart Disease Using Machine Learning Approach. *CARDIOMETRY*, *26*, 342–347. https://doi.org/10.18137/cardiometry.2023.26.342347

[8] Nayeem, M. J. N., Rana, S., & Islam, M. R. (2022, November 30). Prediction of Heart Disease Using Machine Learning Algorithms. *European Journal of Artificial Intelligence and Machine Learning*, 1(3), 22–26. https://doi.org/10.24018/ejai.2022.1.3.13

[9] Rindhe, B. U., Ahire, N., Patil, R., Gagare, S., & Darade, M. (2021, May 12). Heart Disease Prediction Using Machine Learning. *International Journal of Advanced Research in Science, Communication and Technology,* 267–276. https://doi.org/10.48175/ijarsct-1131

[10] Anusuya, V., & Gomathi, V. (2021, March 25). An Efficient Technique for Disease Prediction by Using Enhanced Machine Learning Algorithms for Categorical Medical Dataset. *Information Technology and Control*, *50*(1), 102–122. https://doi.org/10.5755/j01.itc.50.1.25349

[11] Swain, S. K., Behera, N., Swain, A. K., Jayasingh, S. K., Patra, K. J., Pattanayak, B. K., Mohanty, M. N., Naik, K. D., & Rout, S. (2023, October 7). Application of IoT Framework for Prediction of Heart Disease using Machine Learning. *International Journal on Recent and Innovation Trends in Computing and Communication.* https://doi.org/10.17762/ijritcc.v11i10s.7616

[12] Online, E. (2023, September 29). World Heart Day: 27% of deaths in India are caused by cardiovascular diseases. *The Economic Times.* https://economictimes.indiatimes.com/magazines/panache/world-heart-day-27-per-cent-of-deaths-in-india-are-caused-by-cardiovascular-diseases/articleshow/104031929.cms