WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Heart Disease Prediction Using Machine Learning.

Chethan D

Rani Channamma University, Belagavi

ABSTRACT

Coronary heart disease and cardiovascular illnesses (CVDs) have been the leading cause of numerous deaths worldwide over the past few decades, emerging as the most life-threatening disease not only in India but globally. Therefore, there is an urgent requirement for a dependable, precise, and feasible system to diagnose such illnesses in a timely manner for effective treatment. Machine learning algorithms and methods have been utilized to analyze extensive and intricate medical datasets, automating the assessment of large and complex data. The prognosis of heart ailment in maximum cases depends on a complicated aggregate of medical and pathological records. due to this complexity, there exists a sizable quantity of hobby amongst clinical experts and researchers regarding the green and correct prediction of heart disease. a number of the information mining and machine studying strategies are used to are expecting coronary heart illnesses, along with Random wooded area, and guide Support vector machine (SVM).Prediction and diagnosing of heart ailment become a tough issue faced with the aid of doctors and hospitals both in India and abroad. records mining techniques and machine mastering algorithms play a totally vital function in this location. The researchers accelerating their studies works to increase software with the assist of gadget learning algorithms which can assist medical doctors to determine both prediction and diagnosing of heart sickness.

Keywords: Neural network, Machine learning, Supervised learning, Support vector meachine, Random forest, Decision tree Cardiovascular disease.

1. INTRODUCTION:

The human heart is one of the most important organs in the body. It functions as a pump and is a muscular organ comprising four chambers separated by valves and also divide into a halves. Each consists of atrium and ventricle. The atria are responsible for receiving blood, while the ventricles work to push blood out of the heart. The right side of the heart pumps blood with low oxygen levels to the lungs, where the blood cells can acquire more oxygen.

The lungs send freshly oxygenated blood to the left atrium and then to the left ventricle. From the body's organs and tissues receive the newly oxygen-rich blood pumped by the left ventricle.providing energy and maintaining overall health. Heart disease, which encompasses various conditions affecting the heart's structure and function, poses a significant global health concern, extending beyond heart attack and coronary issues.

Failure, different forms of coronary heart disease include arrhythmias, coronary heart valve problems, and congenital coronary heart defects. knowledge the diverse manifestations of coronary heart disease is vital for comprehensive healthcare. heart assaults, medically termed myocardial infarctions, arise while blood drift to part of the heart muscle is blocked, often by way of a blood clot. This obstruction can result in harm or loss of life of the heart muscle cells, necessitating prompt scientific attention to decrease complications. coronary heart failure, then again, refers back to the coronary heart's inability to pump blood efficaciously, ensuing in insufficient oxygen supply to meet the body's wishes. This circumstance can also expand step by step over the years because of underlying coronary heart sickness, leading to signs and symptoms together with fatigue, shortness of breath, and fluid retention. Preventative measures play a pivotal role in mitigating the hazard of heart disease. lifestyle modifications, consisting of a heart-healthy weight-reduction plan, ordinary exercising, and smoking cessation, can considerably contribute to cardiovascular health. additionally, managing conditions which include diabetes, hypertension, and obesity is indispensable to lowering the probability of heart-related complications. As we navigate these complicated interconnections among infectious illnesses, heart health, and healing, continued studies and attention to person cases emerge as paramount. Heart diseases have become a prominent cause of death globally. According to the World Health Organization, heart-related illnesses claim 17.7 million lives annually, accounting for 31% of all global deaths. In India as well, these diseases pose a significant threat.coronary heart-associated illnesses have emerge as the main purpose of mortality . by using fostering cognizance, enforcing preventative techniques, and advancing medical studies, we will collectively work in the direction of selling basic cardiovascular well-being and mitigating the impact of coronary heart ailment in numerous populations.

Heart disease prediction involves using medical data and advanced algorithms to assess an individual's risk of developing heart-related conditions, such as coronary artery disease or heart failure. By analyzing key factors like age, cholesterol levels, blood pressure, lifestyle habits, and family history, machine learning models can predict the likelihood of heart disease. This early prediction allows for timely interventions, potentially preventing serious health issues and improving patient outcomes. The goal is to leverage data to provide personalized healthcare and guide preventive measures.

2. LITERATURE REVIEW:

Machine learning algorithms like Random Forest, SVM (support vector machine), and KNN were utilized to predict Cardio Vascular Disease, with the Random Forest algorithm achieving the highest accuracy of 85% [1]. The concealed naïve Bayes algorithm has the capability to forecast heart disease and it attained a 100% accuracy rate, surpassing naïve Bayes. [2]. Sumit S and colleagues [3] conducted a study aiming to develop a heart disease prediction model using a deep learning neural network and the novel optimization method known as Talos Hyperparameter. The UCI Heart Disease Dataset was utilized for this purpose. Various learning algorithms such as Logistic regression, KNN, SVM, Naïve Bayes, and Hyperparameter optimization (Talos) were employed. This study was also referenced by Bharati R and others [4]. In their data pre-processing, they utilized a public dataset, examined the data distribution, selected features, and checked for duplicates. Following that, they suggested machine learning classifiers and deep learning classifiers. This was mentioned in the study by Rajni Bhalla et al [5]. The three primary elements of the suggested method are dataset alignment, model training, and prediction aggregation. During the dataset alignment phase, the main emphasis is on aligning different datasets to guarantee consistency and compatibility. This was discussed by Rajni Bhalla et al [6]. The effectiveness and performance of proposed and existing methodologies for classification tasks are thoroughly examined, with a focus on collecting and analyzing data for structured data. This analysis is presented by R Silvia et al [7]. This review article aims to outline the current status of artificial intelligence application in clinical practice. The paper effectively discusses Artificial Intelligence and various subfields including Machine Learning and Deep Learning along with their respective subfields [8]. Regression, and Random Forest Classifiers algorithms, which can aid practitioners and medical analysts in accurately diagnosing heart disease. The content of this document involves an analysis of recent journals, published papers, and cardiovascular disease data. The methodology section provides a framework for the proposed model [9]. The construction of a deep Neural Network included the relevant disease-related attributes, which is the fundamental and most crucial method to ensure accurate results in detecting heart disease when applying the model to the Test Dataset. It has been recommended to use the supervised network for diagnosing heart diseases [10]. When using the 10-cross validation technique, KNN achieves an accuracy of 83.16% with a k value of 9 [11]. This information is also presented in [12]. The accuracy of KNN with Ant Colony Optimization is 70.26%, and it has an error rate of 0.526, making it outperform other techniques. However, the decision tree performs inadequately, with a correctly classified instance percentage of 42.8954%, as mentioned in [13]. The same dataset was utilized, but the J48 algorithm was employed to implement Decision Trees according to Tahira Mahboob et al [14].

3. SYSTEM DESIGN AND IMPLEMENTATION:

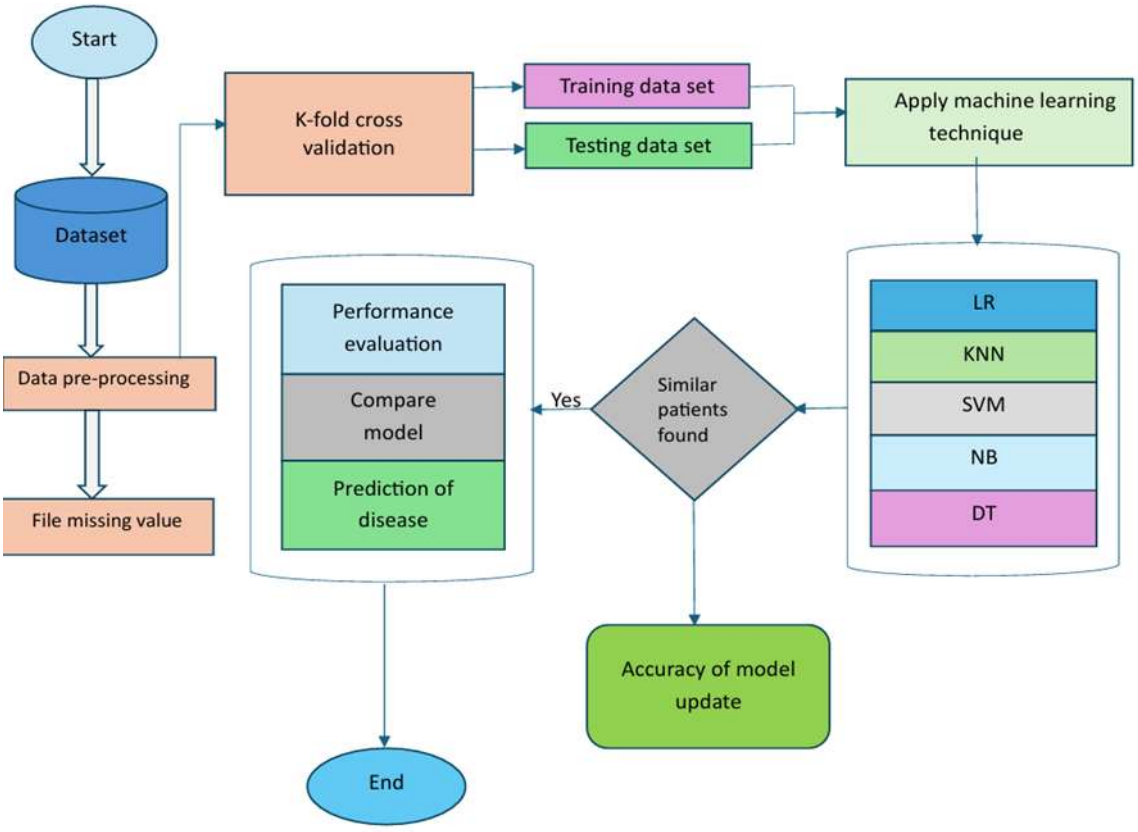

Figure 1. Flow Diagram of the Proposed Methodology

The conceptual version that outlines the structure, behavior, and additional perspectives of a device is known as a system architecture. [Figure 1] A formal description and illustration of a system, created in a way that facilitates understanding of the structures and behaviors of the system, is referred to as an architecture description.

Begin with the first step of the process, which involves acquiring and compiling dataset contain medical information for predicting heart conditions. This dataset commonly includes features consisting of age, sex, blood strain, cholesterol levels, and many others. information Pre-processing clean the facts through coping with missing values, disposing of duplicates, numerical features, encoding express capabilities, document missing price mainly cope with lacking values within the dataset. strategies like imply/mode/median imputation can cope with missing records immediately, may be used, ok-fold move Validation cut up dataset into a 'k' subsets. This process will repeated 'okay' times ensuring every subset gets the risk to be the checking out set, education and checking out facts Set break up the dataset right into a schooling set and a checking out set after preprocessing and go-validation. The training set will be used to teach the machine learning model, whereas the testing set is used to assess its performance.

Practice gadget getting to know method practice diverse machine getting to know algorithms to the training records. not unusual algorithms encompass LR (Logistic Regression) KNN (K-Nearest Neighbors) SVM (Support Vector Machine) NB (Naive Bayes) DT (Decision Tree), overall performance evaluation will be compare the performance for each machine getting to know version that use of suitable metrics inclusive of accuracy, precision, take into account, F1-rating, ROC-AUC, and so on, examine model examine the performance metrics of the extraordinary system learning fashions to determine which one plays nice to your dataset, Prediction of disorder use the satisfactory-performing version to make predictions on new or unseen statistics regarding the chance of heart disorder, comparable patients discovered discover sufferers inside the dataset who've similar characteristics to the ones anticipated to have coronary heart disorder. this could help in know-how commonplace patterns and doubtlessly tailoring interventions, stop step indicates the quit of the present day new release of the process. If the model performance is great, the assignment can be concluded, otherwise, steps might be revisited for improvement.

3.1 Methodology:

The main reason of the proposed technique is to expect the occurrence of heart ailment for early detection to the sickness in a short time to expect the coronary heart sickness primarily based on a few fitness parameters.

1. Data Collection

Collect a comprehensive dataset that includes historical health facts of patients with and with out heart sickness. This dataset will function the inspiration for training and evaluating the device mastering version. Publicly available datasets like the Cleveland heart sickness dataset or Framingham coronary heart observe dataset can be used.

2. Data Preprocessing

Easy and preprocess the information to handle lacking values, normalize capabilities, and encode express variables. This step guarantees that the information is in a appropriate format for schooling gadget getting to know algorithms.

3. Feature Selection

Become aware of the most applicable functions that contribute to coronary heart disease prediction. function selection techniques along with correlation evaluation or recursive function elimination can be employed to enhance version performance.

4. Model Development

Put into effect numerous machine mastering algorithms to are expecting heart disorder hazard. commonly used algorithms for this sort of problem encompass logistic regression, decision trees, random forests, aid vector machines, neural network. The version might be educated and tested using the preprocessed information.

5. Model Evaluation

Examine overall performance of evolved models using metrics inclusive of accuracy, precision, don't forget, F1-rating, and the location underneath the ROC curve (AUC). This evaluation facilitates in choosing the best model for heart disorder prediction.

6. Implementation

Broaden a user-friendly interface or application wherein healthcare experts can enter affected person facts and get hold of chance predictions. This interface might be an internet utility or a mobile app, relying at the intended use.

7. Deployment

Install the gadget getting to know model and the related software in a actual-global placing. make certain that the deployment procedure consists of right integration with existing healthcare systems and adheres to records privateness rules.

8. Continuous Improvement

Reveal the model's performance over the years and replace it with new data to hold its accuracy and relevance. contain feedback from healthcare experts to refine the model and enhance its predictive capabilities.

3.2 Algorithms and Defination:

a. Random Forest:

Random forest is a supervised machine learning algorithm it can be use for both regression and classification tasks, but it typically performs better in classification tasks. This technique involves considering multiple decision tree before producing an output, essentially creating a decision trees. The approach it can based on the belief that a greater number of trees will lead to the correct decision. In classification, it uses a voting system to determine the class, while in regression, it calculates the mean of all the outputs from each decision tree. Random forest works effectively with large datasets that have high dimensionality**.**

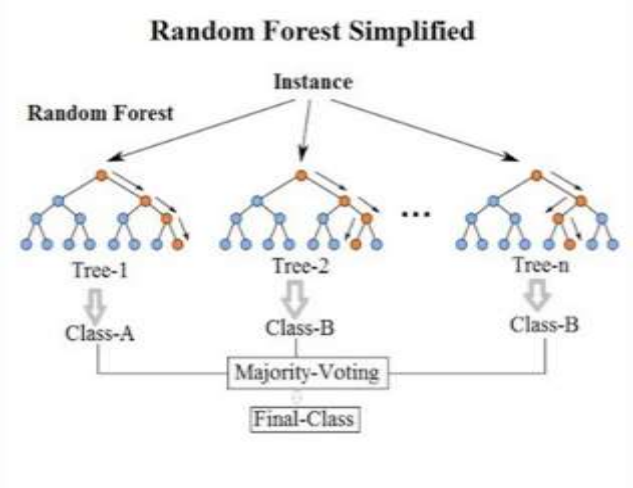

Figure 2. Random Forest

b. Decision Tree

The decision tree is a supervised learning algorithm used primarily for solving classification problems. It can handle both continuous and categorical attributes, dividing the population into two or more similar groups based on the most important predictors. To begin, the algorithm calculates the entropy of each feature and then splits the dataset using the variables or predictors with the highest information gain or lowest entropy. These steps are performed recursively with the remaining attributes.

$$Entropy(S) = \sum_{i=1}^{c} -p_i \log_2 p_i$$

$$Gain(S, A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

Figure 2: CNN Flow chart

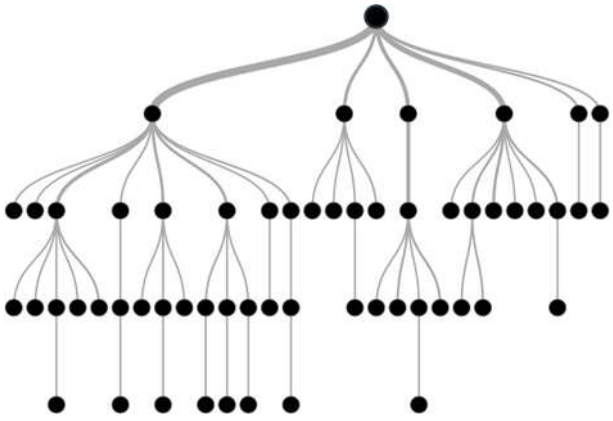

Figure 3. Decision Tree

The decision tree had the lowest accuracy at 77.55% in [10], but its performance improved to 82.17% when used with boosting technique. In the decision tree's performance was very poor with a correctly classified instance percentage of 42.8954%. However, in [26], using the J48 algorithm for decision trees on the same dataset resulted in an accuracy of 67.7%, which is an improvement. Renu Chauhan et al. achieved an accuracy of 71.43% M.A. Jabbar et al. employed alternating decision trees with principal component analysis to achieve an accuracy of 92.2%. Kamran Farooq et al. obtained the best results by using a decision tree-based classifier combined with forward selection, achieving a weighted accuracy of 78.4604% .

4. RESULTS:

In heart disease prediction, results usually tell you the likelihood that a person will develop heart disease based on various factors. These results are generated from models that look at patterns in health data, like age, cholesterol levels, blood pressure, and other lifestyle factors. Here's a simple breakdown of how to interpret the results:

4.1 Attribute information:

Attribute information refers to the specific features or variables in a dataset that are used for analysis, particularly in machine learning and data science. Each attribute represents a characteristic or property of the data being studied. For example, in heart disease prediction, attributes might include age, blood pressure, cholesterol level, and smoking status. These attributes are essential for building predictive models, as they provide the input data that algorithms use to identify patterns, correlations, and make predictions. Understanding attribute information helps in selecting relevant features for accurate analysis and model training.

Age: Patient's age.

Sex: Gender of the patient.

Chest Pain Type: Type of chest pain experienced.

Blood Pressure: Resting blood pressure.

Cholesterol Levels: Serum cholesterol levels.

Fasting Blood Sugar: Whether fasting blood sugar is greater than 120 mg/dl.

Resting ECG: Results of resting electrocardiographic measurements.

Max Heart Rate: Maximum heart rate achieved during exercise.

Exercise Induced Angina: Whether angina was induced by exercise.

Oldpeak: Depression caused by physical activity compared to resting state.

Slope of ST Segment: The incline of the highest exercise ST segment.

Number of Vessels: Quantity of significant vessels stained using fluoroscopy.

Thalassemia: A blood disorder related to the heart.

4.2 User Login Page:

User Login Form

Purpose: To allow users to securely log into the application.

Components:

Type: Text input

Placeholder: "Enter your username"

Validation: Required, should be in email format (if using email for login).

Type: Password input

Placeholder: "Enter your password"

Validation: Required, minimum length (e.g., 6 characters).

Login Button:

Type: Button

Label: "Login"

Action: Submits the form to authenticate the user.

Type: Link

Label: "Forgot Password?"

Action: Redirects to a password recovery page.

Sign Up Link:

Type: Link

Label: "Create an Account"

Action: Redirects to the registration page for new users.

Error Messages:

Display: Inline, near the relevant field

Content: Descriptive error messages for invalid inputs or authentication failures.

Success Message:

Figure 4. User Login Page

SCREEN SHOTS:

Figure 5. Home Page

Figure 6. Result No Risk of Heart Attack

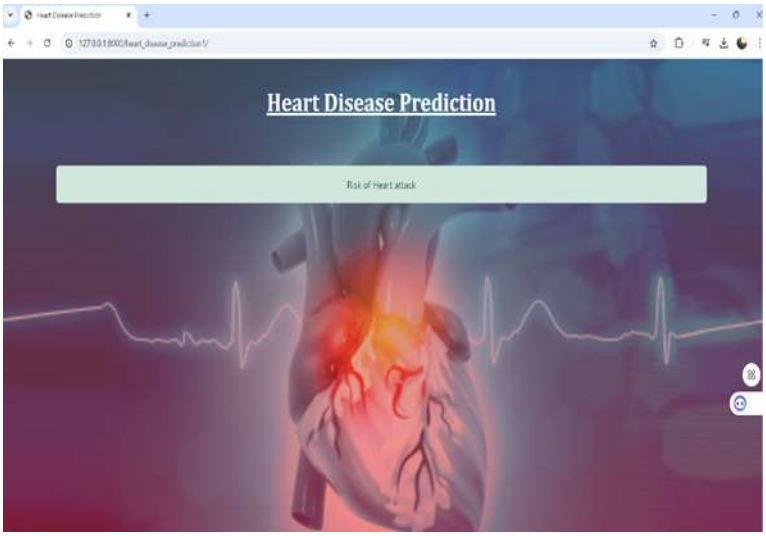

Figure 7. Risk of Heart Attack

Explanation:

In the above depiction (figure 5), the contents depict the User Interface (UI) of my project. The page includes attributes such as age, gender, chest pain type, blood pressure, and more. Users can select the appropriate values and then click on the predict button in the UI. This action triggers processing in the backend. In the subsequent step (figure 6), the actual process commences. The UI validates the entered values and if they are within the normal range, it displays the result as "No Risk of Heart Attack." Similarly, in the following step (Figure 7), if the entered values are high, it indicates a result of "Risk of Heart Attack." The heart disease prediction using ML involves creating a system that leverages machine learning algorithms to predict the likelihood of heart disease based on various patient data.

5. CONCLUSION:

This study managed to provide a significant contribution in computing the strength scores with significant predictors in heart disease prediction. From the evaluation results, we obtained important rules and achieved highest confidence score by utilizing the computed strength scores of significant predictors on Weighted Associative Rule Mining in predicting heart disease..The identification of important rules carries direct clinical relevance, offering actionable insights for healthcare professionals to streamline early detection and tailored management of heart disease. Moreover, the study's findings empower the development of an enhanced decision support system, facilitating informed decision-making in a clinical setting.

Our approach, which prioritizes transparency and explainability, lays the groundwork for a practical implementation of the predictive model within healthcare systems. The computed strength scores not only contribute to the reliability of the model but also open avenues for personalized medicine approaches, allowing for the optimization of preventive strategies and treatment plans according to individual risk profiles.

Future Work:

Future work on heart disease is focused on improving prevention, diagnosis, and treatment through advancements in technology and medicine. This includes the development of personalized treatments using genetic data, the use of artificial intelligence for early detection, and innovations in wearable devices to monitor heart health. Researchers are also exploring regenerative therapies, such as stem cell treatments, to repair damaged heart tissue. Additionally, efforts are ongoing to reduce risk factors through lifestyle interventions and public health initiatives aimed at addressing obesity, smoking, and hypertension.

Acknowledgement:

I am grateful to Assistant Prof. Mallamma V. Reddy Department of Computer Science, Rani Channamma University, Belagavi for his valuable guidance for completion of this work.

REFERENCES:

*[1] Komal Kumar Napa, G.Sarika Sindhu, D.Krishna Prashanthi, A.Shaeen Sulthana, "Analysis and Predictionof Cardiovasculsr Disease using Machine Learning Classifiers",*https://www.researchgate.net/publication/340885231_*Analysis Prediction _of_Cardio_Vascular_Disease_using_Machine_ learning_Classifiers, April 2020.*

[2] Hossam Meshref, "Cardiovascular Disease Diagnosis: A Machine Learning Interpretation Approach",https://www.researchgate.net /publication/338428682_Cardiovascular _Disease_Diagnosis_A_Machine_Learni ng_Interpretation_Approach, January 2019.

[3] Sumit Sharma, Mahesh Parmar (2020), Heart Diseases Prediction using Deep Learning Neural Network Model, International Journal of Innovative Technology and Exploring Engineering (IJITEE), Volume-9, Issue-3. https://doi.org/10.35940/ijitee.C9009.019320.

[4] Bharti R, Khamparia A, Dhiman G, Pande S, Singh P. Prediction of Heart Disease Using a Combination of Machine Learning and Deep Learning. Comput Intell Neurosci. 2021 Jul 1; 2021:8387680. doi- 10.1155/2021/8387680. PMID: 34306056; PMCID: PMC8266441. https://pubmed.ncbi.nlm.nih.gov/34306056/

[5] R. Bhalla, A. Sharma, and G. Ganesan, "A Bibliographic Survey of Sentiment Classification using Hybrid Ensemble-based Machine Learning Approaches," in ceur workshop proceedings, 2022, vol. 3146, pp. 33–38, [Online]. Available: https://ceur-ws.org/Vol-3146/PAPER_04.pdf. https://ceur-ws.org/Vol-3146/PAPER_04.pdf

[6] Bhalla R, Amandeep. A comparative analysis of application of proposed and the existing methodologies on a mobile phone survey. In Futuristic Trends in Networks and Computing Technologies: Second International Conference, FTNCT 2019, Chandigarh, India, November 22 23, 2019, Revised Selected Papers 2 2020 (pp. 107-115). https://link.springer.com/chapter/10.1007/978-981-15-4451-4_10

[7] Romiti S, Vinciguerra M, Saade W et al.,2020. "Artificial Intelligence (AI) and Cardiovascular Diseases: An Unexpected Alliance", Cardiology Research and Practice, vol. 2020, Article ID 4972346, 8 pages. https://doi.org/10.1155/2020/4972346

[8] Virani S. S., Alonso A., Benjamin E. J. et al., Heart disease and stroke statistics-2020 update: a report from the American heart association, Circulation. (2020) **141**, no. 9, e139–e596, https://www.ahajournals.org/doi/10.1161/CIR.0000000000000757.

[9] Wolgast G, Ehrenborg C, Israelsson A, Helander J, Johansson E & Manefjord H (2016). Wireless body area network for heart attack detection [Education Corner]. IEEE antennas and propagation magazine, 58(5), 84-92.

[10] Raihan M, Mondal S, More A, Sagor M O F, Sikder G, Majumder M A & Ghosh K (2016, December). Smartphone based ischemic heart disease (heart attack) risk prediction using clinical data and data mining approaches, a prototype design. In 2016 19th International Conference on Computer and Information Technology (ICCIT) (pp. 299-303).

[11] Seyedamin Pouriyeh, Sara Vahid, Giovanna Sannino, Giuseppe De Pietro, HamiArabnia, Juan Gutierrez et al. " A Comprehensive Investigation and Comparison of Machine Learning Techniques in the Domain of Heart Disease", 22nd IEEE Symposium on Computers and Communication (ISCC 2017): Workshops - ICTS4eHealth 2017.

[12] S.Rajathi and Dr.G.Radhamani et al. "Prediction and Analysis of Rheumatic Heart Disease using kNN Classification with ACO ", 2016.

[13] Simge EKIZ and Pakize Erdogmus et al. "Comparitive Study of heart Disease Classification", 978-1-5386-0440-3/17/$31.00 ©2017 IEEE.

[14] Tahira Mahboob, Rida Irfan and Bazelah Ghaffar et al. "Evaluating Ensemble Prediction of Coronary Heart Disease using Receiver Operating Characteristics", 978 - 1 -5090 -4815 - 1/17/$31.00 ©2017 IEEE