iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** V **Month of publication:** May 2023

DOI: https://doi.org/10.22214/ijraset.2023.51976

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

Study of Optimum Medical Diagnosis of Human Heart Disease using Machine Learning Technique

Aravind Gudapati[1], Arthik Shetty[2], Chandrashekar[3], Kushal S K[4], K. P. Mayuri[5]
[1, 2, 3, 4]*Student,* [5]*Assistant Professor, Department of Computer Science and Engineering, Sir M Visvesvaraya Institute of Technology Bengaluru and affiliated to Visvesvaraya Technological University, Belagavi, Karnataka, India*

Abstract: ***We want to set up a system to be able to identify the indication of a heart diseases at an initial stage in order to avoid the heart failures from occurring in the first phase.***
A system that can precisely forecast the possibility of getting heart disease must be in place because it is unfeasible for the common individual to frequently go through pricey tests like the ECG. Researchers used a range of methods, including the feature selection algorithm sequential feature selection, LDA, SVM, RF, DT, KNN, and GBC, to predict heart illness. As a result, we recommend developing a programme that can estimate the susceptibility of a heart ailment given basic inputs like age, sex, pulse rate and pO2.
Keywords: Heart disease, LDA, KNN, SVM, RF, GBC, DT, LDA, Sequential feature selection.

I. INTRODUCTION

In the field of healthcare, the study of disease diagnosis is essential [1]. A disease is defined as any factor or group of factors that results in pain, ill health, or dysfunction and lastly, a person's demise. Modern life's hectic pace results in an unhealthy lifestyle that breeds anxiety and sadness. There is a propensity to turn to excessive smoking, drinking, and drug use to combat these symptoms. All of these factors are the underlying causes of a number of deadly diseases, such as cancer and cardiovascular conditions. A number of disorders can affect your heart and are referred to as heart disease. In regarding to World Health Organization (WHO) report cardiovascular illnesses are now the major source of death worldwide, resulting for 17.9 million deaths each passing year. Various high obesity, cholesterol, an increase in triglyceride levels, hypertension, etc. are all risk factors for heart disease. According to WHO guidelines, everyone has a fundamental right to health. It is believed that suitable medical services should be made available for regular health examinations. With heart disease accounting for roughly 31% of all fatalities worldwide, it is the most common cause of death. As a result of lack of diagnostic facilities, certified medical professionals, and other resources that affect the accurate prediction of the heart illness, prior identification and treatment of many heart problems are very difficult, especially in developing nations. Utilizing the Random Forest method, the objective is to find out the age group and heart rate. In our project, we describe how the system estimates a user's pulse rate and condition by taking the inputs from them, such as blood pressure (BP) and many others. RFA execution provides improvement of the user experience and more exact results by comparing to other algorithms, which is a lot better method. This assists in the initial diagnosis of the illness and is employed in a variety of ways, including when the input is given to find out the pulse rate based on the obtained input data.

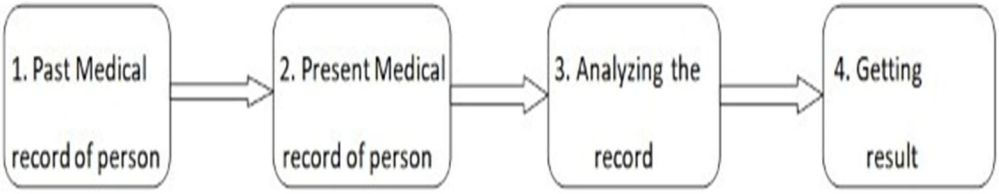


Fig 1 :- Process of diagnosis.

As shown in figure1, diagnosis can be defined as the process of find out which disease a person has based on their symptoms and examination results. Medical records provided the information. An individual with a medical pathology is physically examined to gain the knowledge required for diagnosis. At the minimum one diagnostic procedure, such as a medical test, is often carried out during this treatment. A medical professional will follow a procedure with several phases in order to establish an accurate diagnosis and obtain the greatest amount of data possible. The patient's required therapy may be put off or disregarded as a result of a diagnostic error, leading to serious health issues and it is very dangerous to patient.

II. LITERATURE REVIEW

In page [1] Prediction of the Heart Disease Using Machine Learning Methods Galla Munaga Meghana, Siva Sai Bindhika. Several record where collected to pre-process data on heart disease. The dataset consists of 303 patient records, 6 of which contain partial missing data. After those 6 records were removed from the dataset, pre-processing was performed on the remaining 297 patient records.

Selection and Reduction of Features: Age, testbps, chol, fbs, restecg, thalach, cp, exhang, sex, olpeak, slope, ca, thal, and target are the 13 qualities. Other than age and sex, two of the data set's 13 other qualities are used to determine the patient's personal information. The remaining traits are important since they comprise essential clinical records. Clinical data are crucial for cardiac illness diagnosis and severity assessment. The best-performing models among the many classification models utilised, including decision trees, language models, random forests, and support vector machines, are determined from the results based on their low error rates.

In page [2] Heart Disease can be predictied Using technique Machine Learning by Aditi Gavhane, Prof, Kailas Devadkar (PhD). Recent research have shown that the machine learning field's most well-liked and rapidly developing subfield is neural networks. Now, to test and train the dataset for the suggested system.A superintend neural network approach known as a multi-layer perceptron has one layer for input, one for output, and one or more invisible layers in between these two layers. The invisible layers connect each input layer node to the output layer nodes. Any two nodes' connections have weights assigned to them.

The formula for the calucation of the resultant input. Yin=$\sum wixi$ (i = 0 - n) where xi is ith input, and wi for matching weight. To balance the perceptron, a second identity input of weight b where is to introduced node and is referred to as Bias. Depending on the situation, the nodes' connection can either be feed forward or feedback.

In page [3] Muhammad Waseem Anwar The Framingham dataset is the dataset we are using in our research because it is one of the most popular datasets and contains all the necessary attributes that can aid us in the prognosis of these illnesses. The Framingham dataset was gathered in three stages. Data were gathered from 5209 men and women aged 30 to 62 during the first phase, which was done back in 1948. In the second phase, which took place in 1971, 5124 participants were requested to undergo the identical testing procedure. These individuals represented the second generation of first-phase participants. The final phase, in which the data were processed, was started in April 2002.

In page [4] The innovative strategy he proposed was the selection of important features from the data for machine learning classifier training and testing. They were classified with an accuracy rate of 88.07%. By utilising an improved SVM-based duality optimization technique, Geweid et al. created HD detection tools. To further understand the significance of our suggested methodology, Table 1 summarises the limitations and benefits of the HD diagnosis approaches that have been suggested in the aforementioned literature. To detect the HD, all of these current strategies utilised a variety of ways.

In page [5] Many of them conducted the studies in the medical area on disease detection systems using different machine learning approaches. Another was presented by Vikas Chourasia et al. Understanding of the Data Mining Techniques for Early detection of the Heart illness. They also employed decision tables, iterative dichotomization, classification and regression trees, and other techniques. The method involves 10-fold cross processing, and CART has one of the highest accuracy (83%) of the methods. To diagnose coronary artery disease (CAD), Lakshmana Rao et al. Alizadeh Sani et al. It is used as rule-based, cost-sensitive, and sequential minimum optimization. Techniques for Heart Disease Using Machine Learning figuring out the heart disease have many factors which may affects some of them are(stress, circulatory strain, diabetes, chain smokers, high cholesterol, etc.).

In page [6] Diagnosis And Prognosis Of Cardiac Disease Using various Machine Learning Techniques: J.Jeyaganesan1 ; A.Sathiya2 ; This essay's research involves looking through publications, magzines that have been released, and current heart disease statistics. The technique is a group of simulations that changes the incoming data into patterns of data that are familiar to the users. The introduced approach consists of three stages, the first is the data collection, the second is the extraction of relevant values, and the third is the pre-processing stages where we can study the data. In page [7] P. S. Kohli and S. Arora. On three distinct illness databases (Heart, Breast cancer, and Diabetes) available in the UCI repository for disease prediction, we employ various classification methods in this study, each with its own advantages. The p-value test and backward modelling were used to choose the features for each dataset. The study's conclusions lend credence to the idea that early disease detection may be possible using machine learning. In page [8] Lakshmanarao and T.S. R.Kiran.. Two different datasets from Kaggle and UCI were used to evaluate the suggested methodology. On the unbalanced dataset, we employed sampling techniques, and to discover the best features, we used feature selection techniques. Later, a number of classifier models were used, and the ensemble classifier produced good precision. On evaluation of two datasets have demonstrated that the proposed model is useful for forecastting heart disease. Each and every implementation made use of Python.

In page [9] Erdoğan and S. Güney .Cardiac disease, which kills the majority of patients, is one among significant illnesses in current times. Heart disease diagnosis is highly challenging in medicine. While cardiac conditions are medically diagnosed, they can be mistaken for other conditions that present with the same symptoms, including chest pain, breathing problem, palpitations, and nausea. Medical diagnosis of heart irregularities is very challenging. In this work, machine learning techniques were used to identify the presence of heart disorders. In this study, the patient data were weighted based on how they affected the success rate. In this study, a method for calculating the weight coefficient is proposed. The results of the proposed method show that 13 different patient features led to 86.90% success.

In page [10] The medical industry generates a large scale of data, which must be processed using any cutting-edge methods that will enable us to deliver accurate results, make accurate decisions based on the data, and obtain the intended outcomes. The main issue and one of the main contributors to the number of fatalities worldwide is heart disease. In this survey, a model for predicting heart disease is implemented using machine learning technique such as Gaussian Naive Bayes, Random Forest, and K-Nearest Neighbor supports vector machines 13 elements, including age, BMI, gender, sugar level, blood pressure, oxygen concentration, cholesterol, obesity, and cp, are used in the framework. We are using a user-friendly system that comprises several phases. We choose the algorithm to run on the chosen dataset and upload the dataset file in the first stage.

III. METHODOLOGY

A. *Description of the Dataset*

There are 14 qualities in use here. The patient's presence of heart disease is the target field. The following characteristics were employed in this study report.

1) Age—age patient's in years; sex—(1 for male; 0 for female).
2) Cp—type of chest pain.
3) Trestbps—resting blood pressure, measured at the time of hospital admission (in mm Hg).
4) Chol—serum cholesterol reveals the concentration of triglycerides.
5) Fbs—fasting blood sugar (FBS)
6) Restecg—results from a resting electrocardiogram.
7) Exang—Exercise-induced angina (1 yes)
8) Oldpeak—ST depression induced by exercise relative to rest.
9) Slope—the slope of the peak exercise ST segment.
10) Ca—number of major vessels (0-3) colored by fluoroscopy.
11) Thal—no explanation given, but it is likely thalassemia (3 normal; 6 fixed faults; 7 reversible defects).
12) Target (T): Geographic disease status, no disease=0 and disease=1.

B. *Pre-processing of the Dataset*

There are no null values in the dataset we have to handle the outliers and the dataset's distribution is also off. Two strategies were applied. Without using outliers, a feature selection procedure, or putting the data straight to machine learning algorithms, the results were not encouraging.

1) Checking the Distribution of the Data.

The distribution of the data is essential for problem prediction or categorization. According to the dataset, heart disease happened 54.46% of the time while it went undiagnosed 45.54% of the time.

2) Feature Selection.

The Lasso algorithm, which is a part of embedded approaches, is used to pick features and only select the crucial characteristics when doing feature selection. It has a greater prediction accuracy when compared to filter techniques. It generates useful feature subsets for the employed algorithm.

3) Checking Duplicate Values in the Data.

We need to remove the duplicate carefully if not that they will appear in the training dataset and testing dataset .

C. Evaluation Process Used.

Confusion matrix, accuracy score, precision, recall, sensitivity, and F1 score are all employed in the evaluation procedure.
Accuracy= (TP+TN) / (TP+TN+FP+FN).

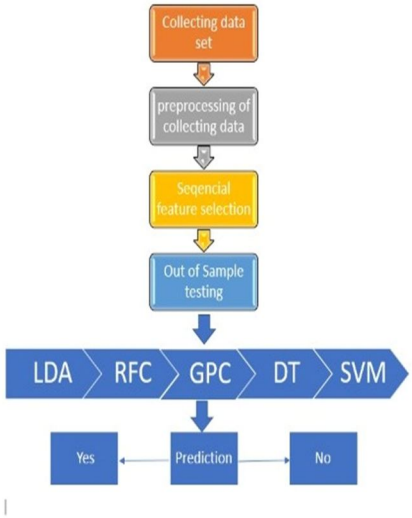


Fig:2. The framework for predicting heart disease.

IV. IMPLEMENTATION

The tech stack used in the project contains python libraries like NumPy, Pandas and MatPlotLib for dataset manipulation and visualization. PyTorch, Scikit-learn, Pickles are used for mode implementation and serialization. Flask framework is used as backend of the project. The HTML templates with CSS stylings are used to respond to the requests.

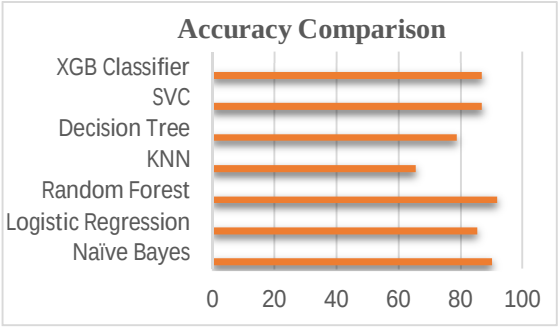


Fig:3. Accuracy comparison of different ML Algorithm

Algorithm for implementing of Random Forest: -

```
1  import RandomForestClassifier
2  n_estimators=50.
3  Fit the model (X_train, y_train)
4  Predict model.predict(X_test)
5  Get model Score
6  Rate people =0
7  Check if len(people) >0:
8  Rate people = len(people)/len(output)
9  Get the prediction of heart disease
10 import confusion_matrix
11 Compute confusion_matrix (y_pred, y_test)
12 Print Accuracy
13
```

V. RESULT AND DISCUSSIONS

Moreover, the maximum accuracy obtained by Random Forest are equal to 91.80% which is greater or almost equal to accuracies obtained from previous researches.

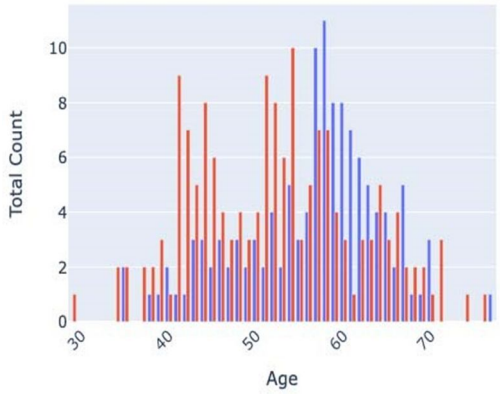


Fig:4. Shows the Risk of Heart disease based on their age.

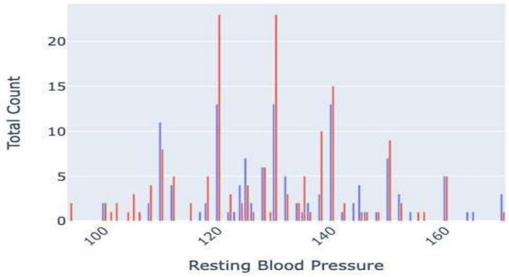


Fig:5. Shows the Risk of Heart disease based on their Resting Blood Pressure.

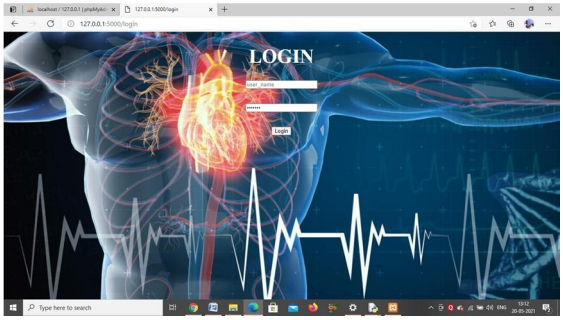
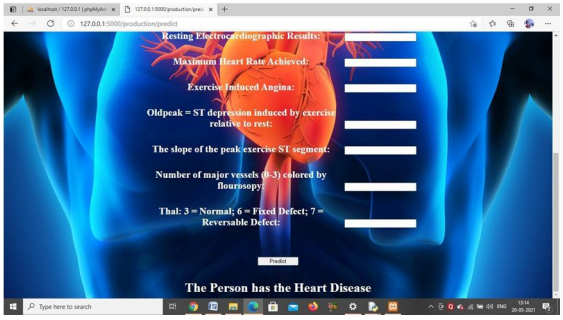


Fig:6. Login page



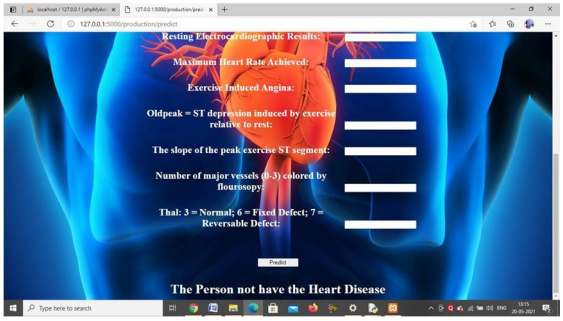


Fig :7. Detection of Heart Disease

VI. CONCLUSION

The Heart Abnormalities Prediction System, which employs the machine learning technique MLP, delivers its customers a calculated result that indicates the user's condition as it relates to developing CAD. Given the recent. Due to technological developments, machine learning algorithms have undergone significant evolution; as a result, we adopt Multi Layered Perceptron (MLP) in the suggested system due to its effectiveness and precision. Adding to this, based on the data entered by the users, this algorithm gives a trustworthy result that is nearby. If people start to use this system frequently, they will be much aware of their present heart status which will definitely decrease in the number of deaths from heart disease. The Heart Disease Forecast System, which employs the MLP machine learning algorithm, makes a prediction for its consumers.

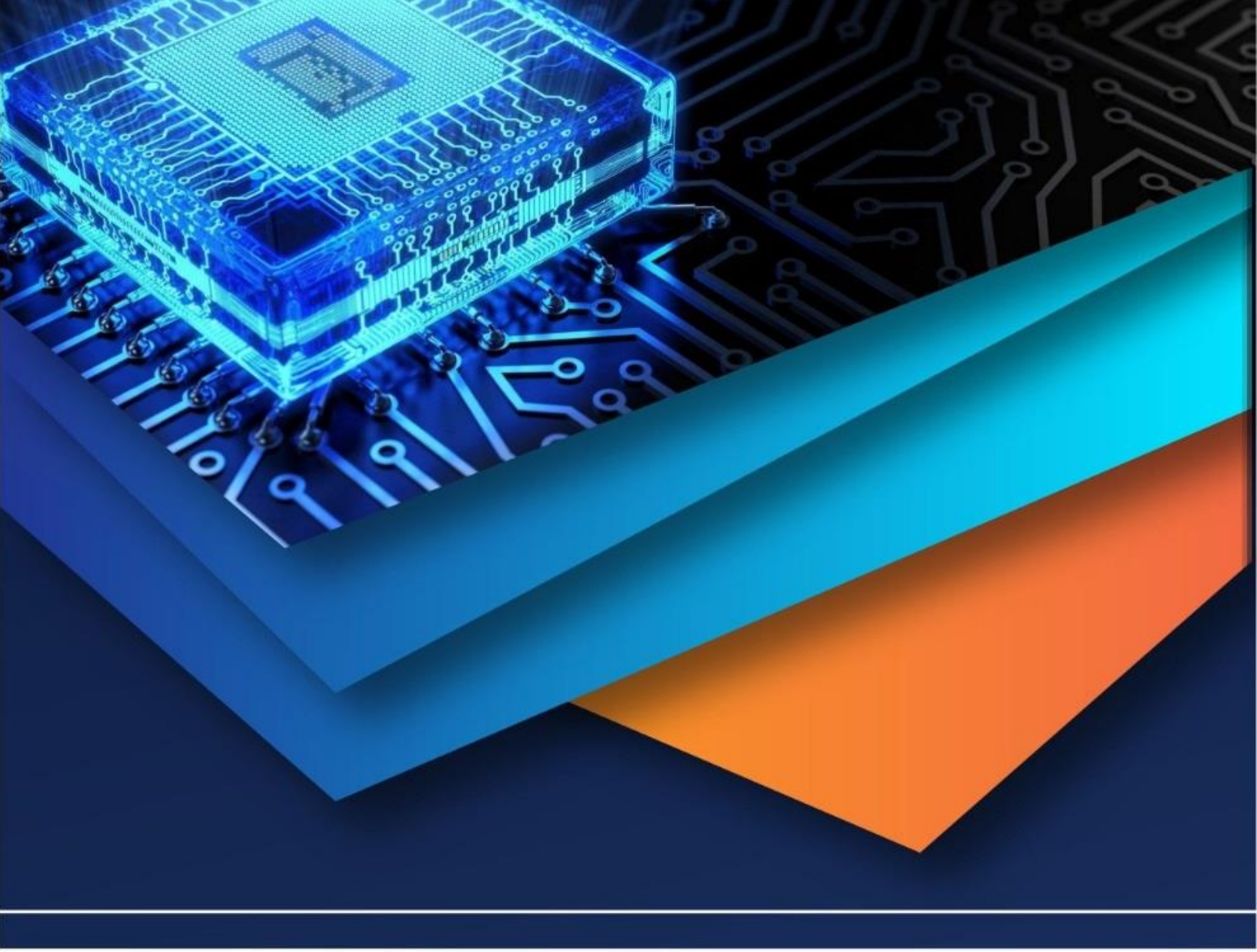
VII. FUTURE SCOPE

The proposed system can be enhanced in order to find various other methods of disease prediction, including those for diabetes, respiratory problems, lung infections, and a wide range of other conditions. We may also develop an Web application that the user can access whenever and wherever he wants The device will then compute all 13 elements, run the algorithms, and provide information about your heart's condition as a result, the patient need not visit the hospital to learn about his or her specific condition.

REFERENCES

[1] “Heart Disease Prediction Using Machine Learning Techniques” Galla Siva Sai Bindhika1, Munaga Meghana 2 International Research Journal. of Engineering and Technology (IRJET).
[2] **“**Prediction of Heart Disease Using Machine”, LearningAditi Gavhane, Prof.Kailas Devadkar (PhD) Proceedings of the 2nd International conference on Electronics, Communication and Aerospace Technology (ICECA 2018).
[3] Muhammad Waseem Anwar "An integrated Machine Learning Framework for effective Prediction on cardiovascular diseases", College of Computing and Informatics, Saudi Electronic University, Riyadh 11673, Saudi Arabia July 20 2021.
[4] Jian Ping Li and Amin Ul Haq "Heart disease identification method using ML classification in E-Healthcare" School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China June 9 2020.
[5] N. L. Fitriyani, M. Syafrudin, G. Alfian and J. Rhee, "Development of disease prediction model based on ensemble learning approach for diabetes and hypertension", IEEE Access, vol. 7, pp. 144777-144789, 2019.
[6] D. Singh and J. S. Samagh, ‘‘A comprehensive review of heart disease prediction using machine learning,’’ J. Crit. Rev., vol. 7, no. 12, p. 2020,2020.
[7] “Diagnosis And Prediction Of Heart Disease Using Machine Learning Techniques” J. Jeyaganesan1 and A. Sathiya2 Ilkogretim Online - Elementary Education Online, 2020; Vol 19 (Issue 2): pp. 1817-1827.
[8] P. Sujatha and K. Mahalakshmi, "Performance Evaluation of Supervised Machine Learning Algorithms in Prediction of Heart Disease," 2020 IEEE International Conference for Innovation in Technology (INOCON), 2020, pp. 1-7,doi: 10.1109/INOCON50539.2020.9298354.
[9] P. S. Kohli and S. Arora, "Application of Machine Learning in Disease Prediction," 2018 4th International Conference on Computing Communication and Automation (ICCCA), 2018.
[10] A. Ed-Daoudy and K. Maalmi, "Real-time machine learning for early detection of heart disease using big data approach," 2019 International Conference on Wireless Technologies, Embedded and Intelligent Systems (WITS), 2019.
[11] A. Lakshmanarao, A. Srisaila and T. S.R.Kiran,"HeartDisease Prediction using Feature Selection and Ensemble Learning Techniques," 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), 2021.
[12] A. Erdoğan and S. Güney, "Heart Disease Prediction by Using Machine Learning Algorithms," 2020 28th Signal Processing and Communications Applications Conference (SIU), 2020, pp. 1-4, doi: 10.1109/SIU49456.2020.9302468.
[13] S. Farzana and D. Veeraiah, "Dynamic Heart Disease Prediction using Multi- Machine Learning Techniques," 2020 5th International Conference on Computing, Communication and Security (ICCCS), 2020.
[14] “Comparative Study of Optimum Medical Diagnosis of Human Heart Disease Using Machine Learning Technique With and Without Sequential Feature Selection. Gulab Nabi and Shafiualla. Received January 26, 2022, accepted February 12, 2022, date of publication February 21, 2022, date of current version March 7, 2022.

[15] S. K. J. and G. S., "Prediction of Heart Disease Using Machine Learning Algorithms.," 2019 1st International Conference on Innovations in Information and Communication Technology (ICIICT), 2019.
[16] S. Doki, S. Devella, S. Tallam, S. S. Reddy Gangannagari, P. Sampathkrishna Reddy and G. P. Reddy, "Heart Disease Prediction Using XGBoost," 2022 Third International Conference on Intelligent Computing Instrumentation and Control Technologies (ICICICT), 2022.
[17] F. Deeba and S. R. Patil, "Implementation of Artificial Intelligence in Disease Prediction and Healthcare System- A Survey," 2021 Innovations in Power and Advanced Computing Technologies (i-PACT), 2021.
[18] A. B. Nassif, O. Mahdi, Q. Nasir, M. A. Talib and M. Azzeh, "Machine Learning Classifications of Coronary Artery Disease," 2018 International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP), 2018.
[19] S. Farzana and D. Veeraiah, "Dynamic Heart Disease Prediction using Multi- Machine Learning Techniques," 2020 5th International Conference on Computing, Communication and Security (ICCCS), 2020.
[20] K. Pahwa and R. Kumar, "Prediction of heart disease using hybrid technique for selecting features," 2017 4th IEEE Uttar Pradesh Section International Conference on Electrical, Computer and Electronics (UPCON), 2017.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY