International Journal of Engineering Research and Science & Technology

IJERST

ISSN : 2319-5991

www.ijerst.com

Email: editor@ijerst.com or editor.ijerst@gmail.com

International Journal of Engineering Research and Science & Technology

https://doi.org/10.62643/ijitce.2024.v20.i3.pp179-182

ISSN 2319-5991 www.ijerst.com

Vol. 20, Issue 3, 2024

Heart Disease Prediction Using Machine Learning Techniques

[1] D. Ratna Kumari, [2]G. Shanmukha Sai Santosh, [3] P. Guna Sai Yaswanth, [4] I. Sujeeva Raju, [5] B. Lohith Datta Sai

[1] Professor and Head of the department, Dept. of CSE, Ramachandra College of Engineering (A), Affiliated to JNTUK Kakinada, Eluru Andhra Pradesh, India.

[2,3,4,] UG Students, Dept. of CSE, Ramachandra College of Engineering (A), Affiliated to JNTUK Kakinada, Eluru Andhra Pradesh, India.

ABSTRACT

Heart disease is now days one of the primary basis of death in world. Prediction of heart diseases requires more precision, perfection and correctness because a little mistake can cause death of the person and also it associates with many risky factors. To deal with the problem there is essential need of prediction system to get accurate and reliable about diseases. Machine learning provides the way to predict any kind of event which take training from natural events. In this paper we have implemented supervise machine learning classification algorithms like K- Nearest Neighbour and Random Forest and calculate accuracy by using existing dataset from the Cleveland database, Kaggle database of UCI repository of heart disease patients.

KEYWORDS- Machine Learning, Flask, Prediction, Interface, Algorithms

1. INTRODUCTION

According to the World Health Organization, every year 12 million deaths occur worldwide due to heart disease. The load of cardiovascular disease is rapidly increasing all over the world from the past few years. Many researchers have been conducted in attempt to pinpoint the most influential factors of heart disease as well as accurately predict the overall risk. Heart Disease is even highlighted as a silent killer which leads to the death of the person without obvious symptoms. The early diagnosis of heart disease plays a vital role in making decisions on lifestyle changes in high risk patients and in turn reduce the complications. This project aims to predict future heart disease by analyzing data of patients which classifies whether they have heart disease or not using machine learning algorithms. Over the last decade, heart disease or cardiovascular remains the primary basis of death worldwide. An estimate by the World Health Organization, that over 17.9 million deaths occur every year worldwide because of cardiovascular disease, and of these deaths, 80% are because of coronary artery disease and cerebral stroke. Various habitual risk factors such as smoking, overuse of alcohol and caffeine, stress, and physical inactivity along with other physiological factors like obesity, hypertension, high blood cholesterol, and pre-existing heart conditions are predisposing factors for heart disease iterative refinement, and ensuring the system's usability and scalability.

2. DISCUSSION

Research on predicting heart disease in its early stages through machine learning has gained significant attention in recent years. Studies have extensively explored various aspects of this domain, focusing on feature selection and extraction techniques that identify critical variables such as age, BMI, family history, blood pressure, and habits. Machine learning algorithms, including Decision Trees, Support Vector Machines, and Ensemble methods like Random Forests, have been Implemented.

International Journal of Engineering Research and Science & Technology
https://doi.org/10.62643/ijitce.2024.v20.i3.pp179-182
ISSN 2319-5991 www.ijerst.com
Vol. 20, Issue 3, 2024

Data preprocessing techniques, such as imputation and standardization, are commonly applied to ensure robust performance.

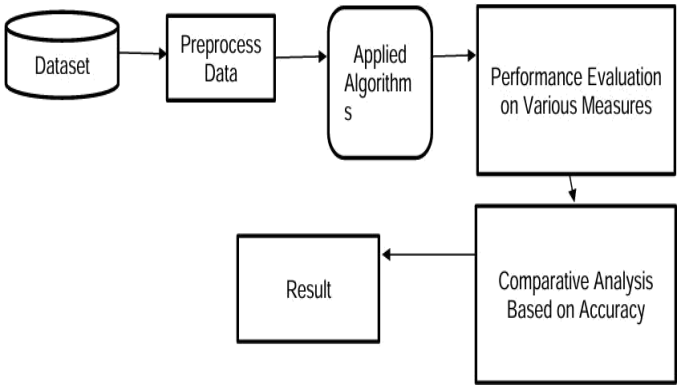


Fig 1: Architecture

IMPLEMENTATION

Data collection:

- The dataset for Heart disease prediction discovery employed in this design was sourced from Kaggle. The set data generally consists of rows and columns, where each row represents an individual and each column represents a specific point.

Data Preprocessing:

- Gathered a large, well- structured dataset containing applicable features for Heart disease prediction. Consider intimately available datasets.
- Preprocess the data to handle missing values, outliers and label encoding.

Splitting the data:

- The pre-processed data is splitting into training and testing data. Training Set contains 80% data. This larger portion of the data is used to train the machine learning models
- The Testing Set contains 20% It is used latterly to evaluate the model's performance on new data it has not encountered ahead.

Machine Learning Model Training:

- In these ML algorithms for classification, such as Support Vector Machines (SVMs), Random Forests are used.

Train the ML model on the training set and estimate its performance on the testing set using criteria like accuracy score, precision, recall, and F1-score.

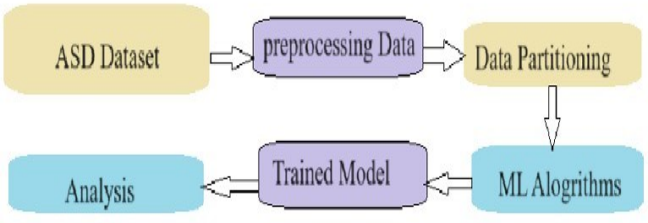
User Interface:

- Developed a user-friendly web interface using Flask to collect patient information from users.
- Integrated the trained ML model into the web operation to assay the entered data and provide a prediction (HD Present or Not).
- Emphasize that the prediction is for original webbing purposes and should not be considered a definitive opinion. Advise users to consult a healthcare professional for comprehensive evaluation.

METHODOLOGY

The methodology includes the choosing the machine learning algorithms suitable for classification tasks, considering the need for speed, delicacy, and real-time processing. consider using pre-trained models known for their efficiency and delicacy, similar as Random Forest Classifier, and AdaBoost Classifier. Use applicable features from the dataset, which may include Cholesterol Levels, medical history related to HD prediction.

Evaluating of the trained models using applicable performance criteria, such as accuracy, precision, recall, and F1-score.



Proposed Methodology

By following this methodology, experimenters and interpreters can develop and emplace machine learning models for prognosticating Heart Disease using algorithms like Random Forest Classifier while incorporating ways for data addition, pre-trained models, and optimization.

Random forest classifier:

- A random forest is a meta estimator that fits a number of decision tree classifiers on variety of sub-samples of the dataset and uses comprising to ameliorate the prophetic accuracy and control over-fitting. n_estimators

International Journal of Engineering Research and Science & Technology
https://doi.org/10.62643/ijitce.2024.v20.i3.pp179-182
ISSN 2319-5991 www.ijerst.com
Vol. 20, Issue 3, 2024

int, default=*100* indicates the number of trees in the forest.

- The evaluation of a machine learning Random Forest classifier performance is done by using the accuracy_score function from the sklearn_metrics module.
- One of the most important features of the Random Forest Algorithm is that it can handle the data set containing nonstop variables, as in the case of regression, and categorical variables, as in the case of classification.
- It performs better for classification and regression tasks.
- In prediction, the algorithm summations the results of all trees, either by voting.
- Random forests are extensively used for classification and regression functions, which are known for their capability to handle complex data, reduce overfitting.
- By incorporating advanced machine learning algorithms, such as ensemble methods, the proposed system aims to achieve higher predictive accuracy compared to existing approaches.
- This means more reliable predictions for identifying individuals at risk of heart disease.
- The system is designed to handle intricate relationships within the data. This is crucial because heart disease risk factors often interact in complex ways.

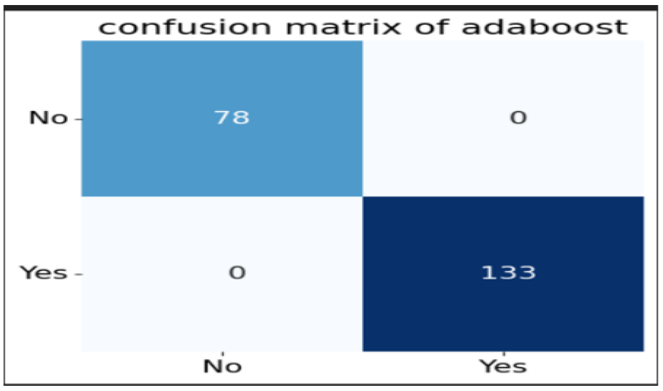
3. RESULTS

The system shows better performance comparing to the other being approach of Heart disease screening. The user answers all the mentioned questions given in the form page and other required details completely without missing any. Hybrid model which includes all the mentioned algorithm has been used for analysing and Heart disease. Finally, detects if a user has HD or not.

The evaluation of performance is done by using the accuracy_score function from the sklearn metrics module and it is used to cipher the delicacy of a classification model. The accuracy score of the model is 100.000000.

The evaluation of a machine learning Random Forest classifier performance is done by using the accuracy_score function from the sklearn_metrics module.

It is used to compute the accuracy of a classification model. The accuracy score of the model is 99.7959.



The classification report of adaboost classifier

	precision	recall	f1-score	support
No	1.00	1.00	1.00	78
Yes	1.00	1.00	1.00	133
accuracy			1.00	211
macro avg	1.00	1.00	1.00	211
weighted avg	1.00	1.00	1.00	211

User Interface:

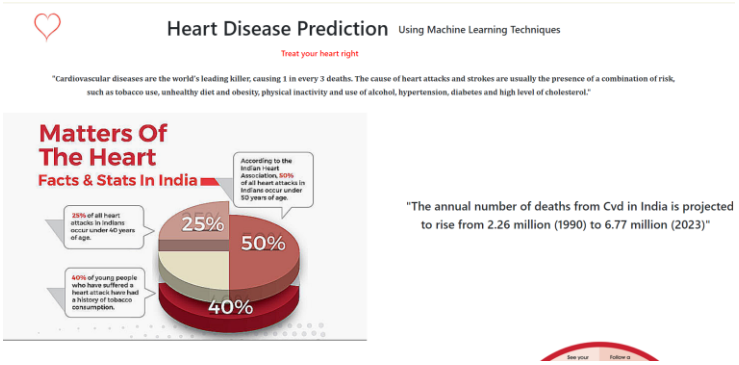


Fig 2: Home page

International Journal of Engineering Research and Science & Technology

https://doi.org/10.62643/ijitce.2024.v20.i3.pp179-182

ISSN 2319-5991 www.ijerst.com

Vol. 20, Issue 3, 2024

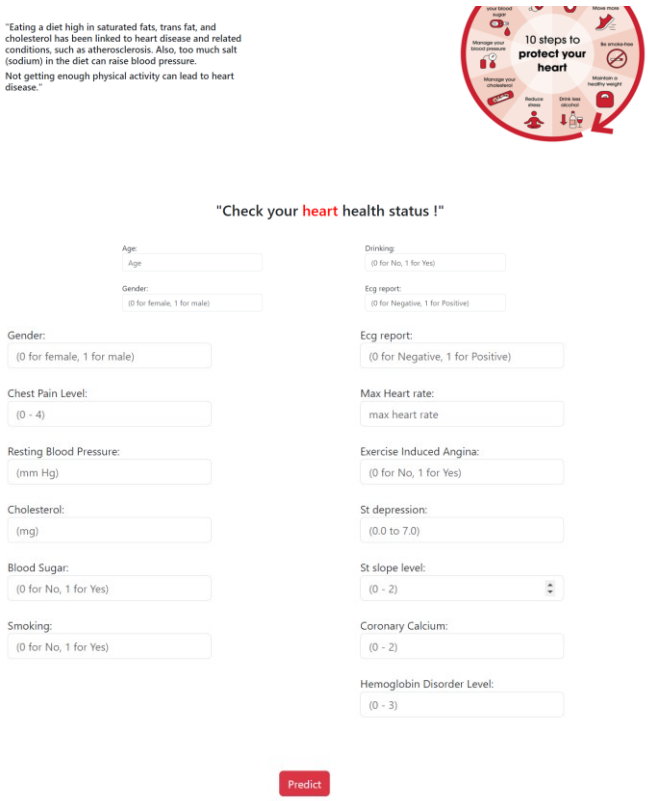


Fig 3:Form page

Fig 4:Precautions page

4. CONCLUSION

In this design, we embarked on a trip to influence machine learning ways for the early discovery and diagnosis of Heart Disease Prediction. The project encompassed several crucial stages, including data preprocessing, model training, and the development of a user-friendly interface using Flask.

We began by importing necessary libraries and performing data preprocessing to prepare our dataset for model training. This involved encoding categorical variables, handling missing values, and splitting the data into training and testing sets. We used the RF classifier as our machine learning model due to its capability to combine multiple weak learners to make a robust classifier.

After training the model on the dataset, we estimated its performance using criteria such as accuracy score, confusion matrix, and classification report. The results indicated promising delicacy in prognosticating the presence or absence of HD, showcasing the eventuality of machine learning in aiding clinical diagnosis.

To make the model accessible to end-users, we developed a user interface using Flask. This interface allowed users to input applicable information about a case, similar cholesterol, bp levels and admit instant feedback on the liability of HD presence. The interface also handed precious information about Heart Disease, including symptoms, precautions, and resources for further exploration.

REFERENCES

[1] Seckeler MD, Hoke TR. The worldwide epidemiology of acute rheumatic fever and rheumatic heart disease. Clin Epidemiol. 2011;3:67.

[2] Gaziano TA, Bitton A, Anand S, Abrahams-Gessel S, Murphy A. Growing epidemic ofcoronary heart disease in low-and middle-income countries. Curr Probl Cardiol. 2010;35(2):72–115.

[3] Weng SF, Reps J, Kai J, Garibaldi JM, Qureshi N. Can machine-learning improve cardiovascular risk prediction using routine clinical data? PLoS ONE. 2017;12(4):e0174944.

[4] Ramalingam VV, Dandapath A, Raja MK. Heart disease prediction using machine learning techniques: a survey. Int J Eng Technol. 2018;7(2.8):684–7. [5] Patel J, Tejal Upadhyay D, Patel S. Heart disease prediction using machine learning and data mining technique. Heart Dis. 2015;7(1):129–37.

[5] Fatima M, Pasha M. Survey of machine learning algorithms for disease diagnostic. J Intel Learn SystAppl.2017;9:1–16.

[1] https://doi.org/10.4236/jilsa.2017.91001