



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 12, Special Issue 2, April 2023

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

9940 572 462

6381 907 438

ijirset@gmail.com

www.ijirset.com



Effective Prediction of Cardiovascular Disease using K-Nearest Neighbors Algorithm

Mrs. A. Jayanthi¹, Adlin Delma Y², Harini M², Nandhini S²

Assistant Professor, Department of Computer Science Engineering, Velammal Engineering College, Velammal Engineering College, Chennai, Tamil Nadu, India¹

U.G. Students, Department of Computer Science and Engineering, Velammal Engineering College, Velammal Engineering College, Chennai, Tamil Nadu, India²

ABSTRACT:One of the leading causes of death in the modern world is heart disease. There are many medical tools available in the industry for forecasting cardiac illnesses, however there are factors to be taken into account including accessibility and precision. With the use of machine learning (ML), it has been demonstrated that it is possible to make predictions and judgements from the vast amount of data generated by the healthcare sector. Additionally, we have observed the employment of ML approaches in recent advancements across several IoT domains (IoT). Only a few research have looked into using ML to predict cardiac disease. In this study, we propose to develop a machine learning model that can determine a person's potential for developing heart disease based on a combination of risk factors (disease-specific parameters). Several feature combinations and many well-known categorization strategies are used to introduce the prediction model. By combining a hybrid method of data insertion and processing to create a prediction model for heart disease, we achieve an improved performance level with an accuracy level of 88.7%. (HRFLM).

KEYWORDS:Machine learning, heart disease prediction, feature selection, prediction model, classification algorithms, cardiovascular disease (CVD).

I. INTRODUCTION

The heart plays a critical role in maintaining proper bodily function. Research conducted by the World Health Organization (WHO) reveals that each year, cardiovascular disease (CVD) accounts for 31% of global deaths, with more than three-quarters of these deaths occurring in low- and middle-income countries such as India. Accurately predicting the onset of cardiovascular disease remains a significant challenge. While traditional techniques have proven ineffective in this regard, there are now several medical devices available on the market. However, these devices tend to be prohibitively expensive and not entirely reliable. Key risk factors for heart disease include age, sex, blood pressure, cholesterol, blood sugar, diabetes, obesity, unhealthy diet, physical inactivity, smoking, and alcohol consumption, most of which are controllable through lifestyle changes. In recent years, medical professionals have capitalized on technological advancements in machine learning, artificial neural networks (ANN), deep learning, fuzzy logic, data mining, genetic algorithms, and other soft computing techniques to achieve more accurate disease diagnosis and prognosis. Machine learning, in particular, has emerged as an ideal method for predicting heart disease due to its ability to analyze large data sets and identify patterns and trends more rapidly and with greater reliability.

II. RELATED WORK

Due to the complexity of diagnosis of cardiovascular disease, an automatic medical diagnosis system would be beneficial. Various studies have been conducted to diagnose heart disease using data mining techniques, including multiple regression models, Naïve Bayes, decision tree, support vector machine, and artificial neural network. These studies achieved different probabilities for different methods.

For instance, Polaraju, Durga Prasad, and Tech Scholar (2017) proposed a multiple regression model that was appropriate for predicting the chance of heart disease. The model was trained on a data set of 3000 instances with 13 different attributes.



Organized by

Department of CSE, Velammal Engineering College, Chennai, India

Deepika and Seema (2017) focused on predicting chronic disease by mining historical health records using different classifiers. They found that SVM gave the highest accuracy rate for heart disease, while Naïve Bayes was more accurate for diabetes.

In their 2018 study, Beyene and Kamat suggested several algorithms, including Naive Bayes, Classification Tree, KNN, Logistic Regression, SVM, and ANN, for predicting heart disease using data mining techniques. Among these algorithms, Logistic Regression was found to yield the highest accuracy. The authors also utilized the WEKA software to automate disease diagnosis and improve healthcare center services. The study further tested various algorithms, such as SVM, Naïve Bayes, Association rule, KNN, ANN, and Decision Tree, and concluded that SVM was the most effective in terms of accuracy compared to other data mining algorithms.

III. METHODOLOGY

1) Data Pre-Processing

Heart disease data is pre-processed after collection of various records. The dataset contains a total of 903 patient records, where 6 records are with some missing values. Those 6 records have been removed from the dataset and the remaining 897 patient records are used in pre-processing. Figure 1(a) shows the dataset collected from four different countries like Cleveland, Hungarian, Switzerland, Virginia. Figure 1(b) displays the attributes and data in a combined dataset of four different countries.



1(a)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
2	63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
3	37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
4	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
5	56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
6	57	0	0	120	354	0	1	163	1	0.6	2	0	2	1
7	57	1	0	140	192	0	1	148	0	0.4	1	0	1	1
8	56	0	1	140	294	0	0	153	0	1.3	1	0	2	1
9	44	1	1	120	263	0	1	173	0	0	2	0	3	1
10	52	1	2	172	199	1	1	162	0	0.5	2	0	3	1
11	57	1	2	150	168	0	1	174	0	1.6	2	0	2	1
12	54	1	0	140	239	0	1	160	0	1.2	2	0	2	1
13	48	0	2	130	275	0	1	139	0	0.2	2	0	2	1
14	49	1	1	130	266	0	1	171	0	0.6	2	0	2	1
15	64	1	3	110	211	0	0	144	1	1.8	1	0	2	1
16	58	0	3	150	283	1	0	162	0	1	2	0	2	1
17	50	0	2	120	219	0	1	158	0	1.6	1	0	2	1
18	58	0	2	120	340	0	1	172	0	0	2	0	2	1
19	66	0	3	150	226	0	1	114	0	2.6	0	0	2	1
20	43	1	0	150	247	0	1	171	0	1.5	2	0	2	1

1(b)

2) Feature Selection and Reduction

From among the 13 attributes of the data set, two attributes pertaining to age and sex are used to identify the personal information of the patient. The remaining 11 attributes are considered important as they contain vital clinical records. Clinical records are vital to diagnosis and learning the severity of heart disease.

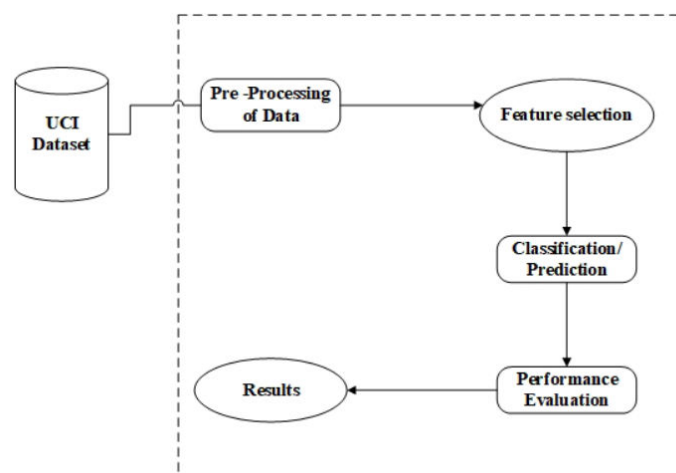
3) Classification Modelling

The clustering of datasets is done on the basis of the variables and criteria of Decision Tree (DT) features. Then, the classifiers are applied to each clustered dataset in order to estimate its performance. The best performing models are identified from the above results based on their low rate of error.

- Decision Trees Classifier - A decision tree classifier is a widely used supervised learning algorithm in machine learning that is designed to tackle classification problems. The classifier constructs a decision tree by recursively dividing the training data into smaller subsets based on the most informative features that predict the class labels of the target variable. Each decision in the tree is made based on a feature value, and each terminal node represents a class label. Decision trees are highly interpretable and visually appealing, making them a popular choice for classification tasks. They can accommodate both categorical and numerical data and are capable of capturing nonlinear relationships between the input features and the target variable. However, decision trees have a tendency to overfit and may not perform well on unseen data.
- Support Vector Classifier - The SVM algorithm aims to identify a hyperplane in an N-dimensional space that effectively separates the data points into distinct classes. The dimension of the hyperplane is determined by the number of features present in the data. In cases where there are only two input features, the hyperplane reduces to a line. If there are three input features, then the hyperplane becomes a 2-D plane. However, it becomes increasingly challenging to visualize the hyperplane as the number of features goes beyond three.
- Random Forest Classifier - The random forest algorithm is an ensemble method that combines multiple base-level models to achieve better results. To comprehend how this algorithm operates, it is necessary to first examine its building blocks, which are decision trees, and how they function.
- K- Nearest Neighbor - K-nearest neighbors (KNN) algorithm utilizes the concept of "feature similarity" to forecast the values of new data points. This implies that the new data point will be assigned a value based on its resemblance to the points in the training set.

4) Performance Measures:

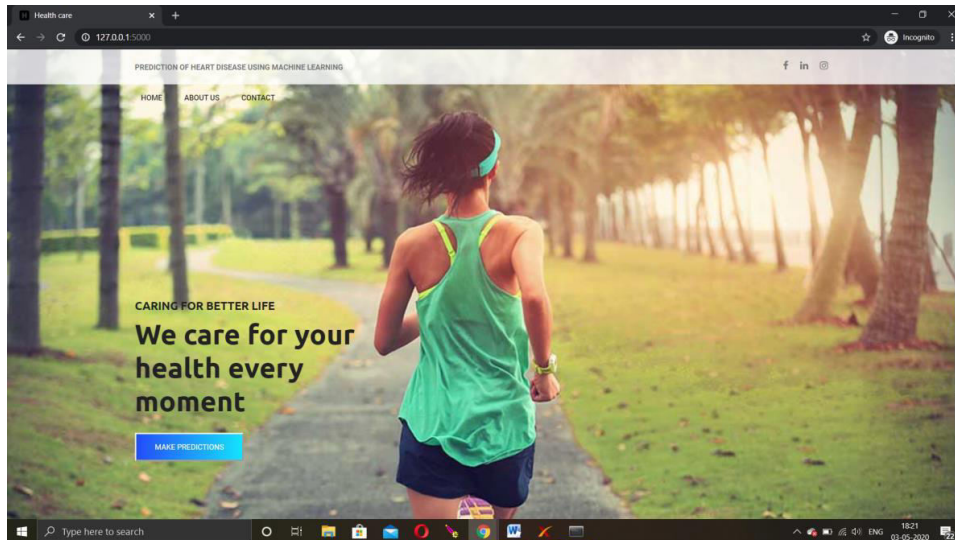
Several standard performance metrics such as accuracy, confusion matrix, precision and recall, error in classification have been considered for the computation of performance efficacy of this model.



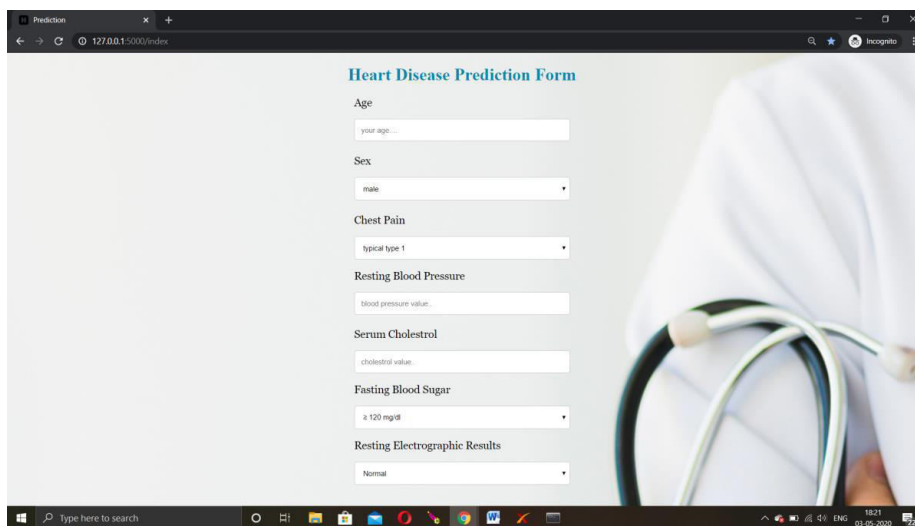


IV. EXPERIMENTAL RESULTS

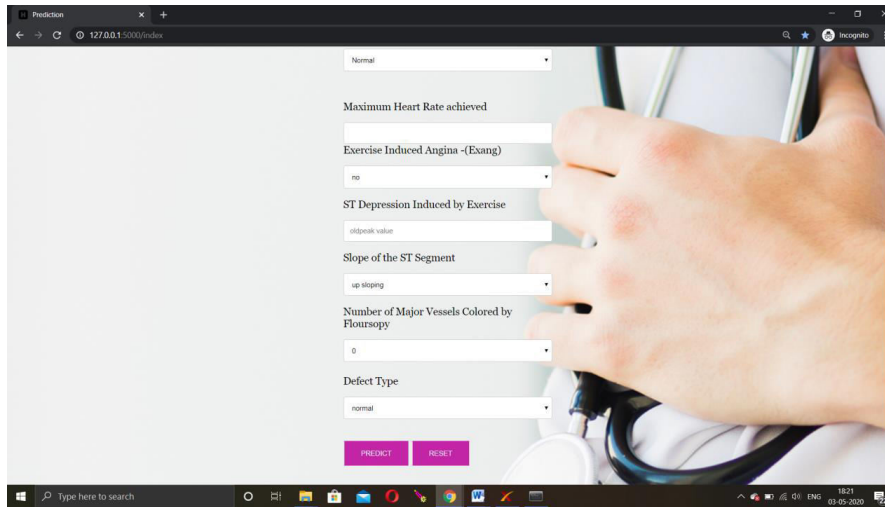
A user-friendly flask application enhances the entire user experience and serves the purpose of prediction of cardiovascular disease. The below shown figure displays the home screen with user login and register which can speed up the process.



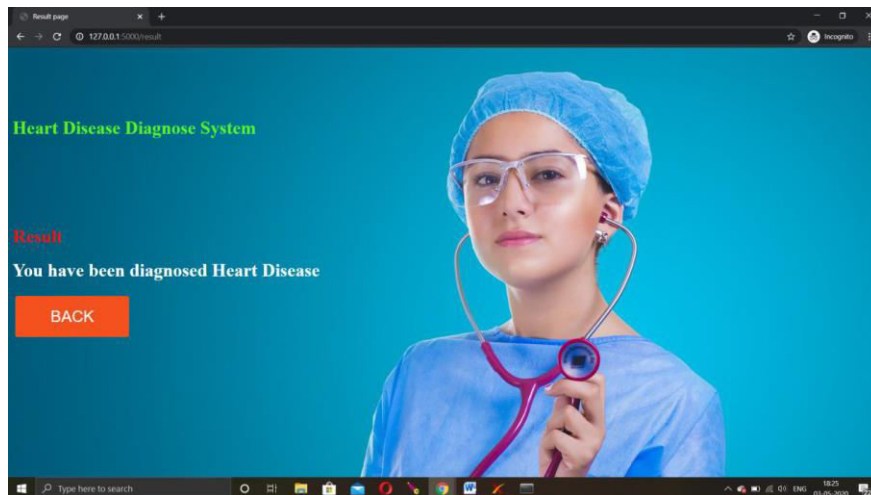
The below shown figure is the prediction page with comprises of text boxes and drop down boxes wherein the major factors of cardiovascular disease of the person is to be entered.



After entering the essential attributes for the prediction of cardiovascular disease, the prediction can be made by clicking the button. Here, the attributes are put through the machine learning process which comprises of the k-nn algorithm.



The below figure displays if the person is diagnosed with cardio vascular disease or not based on the factors or attributes entered using machine learning.



V. CONCLUSION

To begin with, the four algorithms were implemented and trained on separate datasets. Afterward, they were tested, and the most effective algorithm was chosen based on several factors. Our findings indicated that the KNN algorithm was the most efficient, with an accuracy rate of 87.0%, whereas the decision tree, SVM, and Random Forest Classifier had accuracy rates of 79.0%, 83.0%, and 84.0%, respectively. Consequently, the KNN algorithm was further developed into a web application with a more user-friendly interface, utilizing HTML5, CSS, JavaScript, and Flask (Python's microweb framework). This project aims to provide end-users with an idea of their heart's status, and the implementation of machine learning technology in the first prediction of heart problems will have a significant social impact since heart diseases are a major cause of death in India and worldwide. The user will be notified if they are at risk and should seek medical attention, ultimately reducing the number of fatal heart attacks. Therefore, the aforementioned approach successfully studied an individual's heart diseases, and the results projected the risk of heart disease based on the provided user factors.

REFERENCES

- [1] M.Akhiljabbar, Dr.PritiChandrab, Dr.B.LDeekshatuluc.: "Heart Disease PredictionSystem using Associative Classification and Genetic Algorithm", ICECIT, (2012)
- [2] Syed Umar Amin, Kavita Agarwal, Dr. Rizwan Beg.: "Genetic Neural Network BasedData Mining in Prediction of Heart Disease Using Risk Factors", ICT, (2013)
- [3] Bhuvanewari Amma N.G., .: "Cardiovascular Disease Prediction System usingGenetic Algorithm and Neural Network", IEEE, 2012
- [4] Oluwarotimi Williams Samuel, Grace MojisolaAsogbona, Arun Kumar Sangaiahc,PengFanga, and Guanglin Lia.: "An Integrated Decision Support System Based on ANN andFuzzy_AHP for Heart Failure Risk Prediction", SIAT, CAS, (2016)
- [5] Aditi Gavhane ;GouthamiKokkula ; Isha Pandya ; Prof. KailasDevadkar.: "Prediction of Heart Disease using Machine Learning", ICECA(2019)
- [6] K. Mathan, PriyanMalarvizhi Kumar, Parthasarathy Panchatcharam, GunasekaranManogaran, R. Varadharajan.: "A novel Gini index decision tree data mining method withneural network classifiers for prediction of heartdisease", Kluwer Academic PublishersNorwell, vol. 22, (2018)
- [7] Nilakshi P. Waghulde, Nilima P. Patil.: "Genetic Neural Approach for Heart DiseasePrediction", ISSN, vol. 4, no. 3, sep-2014
- [8] Ismail Babaoglu, Omer KaanBaykan, NazifAygul, KurtulusOzdemir, MehmetBayrak.: "Assessment of exercise stress testing with artificial neural networkin determiningcoronary artery disease and predicting lesion localization", Science Direct, 2009
- [9] William G. Bat.: "Use of an Artificial Neural Network for Data Analysis in ClinicalDecision-Making: The Diagnosis of Acute Coronary Occlusio", MITP, vol. 2, issue. 4, 1990
- [10] Himanshu Sharma, M A Rizvi. : "Prediction of Heart Disease using MachineLearning Algorithms: A Survey ",International Journal on Recent and Innovation Trends inComputing and Communication, Volume: 5 Issue: 8 , 2017.
- [11] Jae-Hong Eom, Sung-Chun Kim, Byoung-Tak Zhang.: "AptaCDSS-E: A classifierensemble-based clinical decision support system for cardiovascular disease level prediction,ScienceDirect", 2008
- [12] Indira S. FalDessai.: "Intelligent Heart Disease Prediction System Using ProbabilisticNeural Network", IJACTE, vol. 2, issue. 3, 2013
- [13] H.Jayasree, D. S. S. K. R. T. Naren, K. Sai Sankeerth, T. Kumidini.: "HEARTDISEASE PREDICTION SYSTEM", www.researchgate.net/publication/334612815, July2019
- [14] AH Chen, SY Huang, PS Hong, CH Cheng, EJ Lin.: "HDPS: Heart DiseasePrediction System", IEEE, 2011
- [15] Latha Parthiban and R.Subramanian.: "Intelligent Heart Disease Prediction Systemusing CANFIS and Genetic Algorithm", International Journal of Medical and Health Sciences, Vol. 1, No. 5, 2007



SJIF: 8.423



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INTERNATIONAL CENTRE



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 **9940 572 462**  **6381 907 438**  **ijirset@gmail.com**



www.ijirset.com

Scan to save the contact details