



e-ISSN: 2319-8753 | p-ISSN: 2320-6710

IJIRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 10, Issue 3, March 2021

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 7.512

9940 572 462

6381 907 438

ijirset@gmail.com

www.ijirset.com



A Systematic Cardiovascular Disorder Identification Using Machine Learning Methods

Jishnu D, Rahul J A, Santhosh A, Mr.S.Ezra Vethamani

U.G Student, Department of Computer Science and Engineering, Velammal Institute of Technology, Chennai,
Tamil Nadu, India

U.G Student, Department of Computer Science and Engineering, Velammal Institute of Technology, Chennai,
Tamil Nadu, India

U.G Student, Department of Computer Science and Engineering, Velammal Institute of Technology, Chennai,
Tamil Nadu, India

Associate Professor, Department of Computer Science and Engineering, Velammal Institute of Technology, Chennai,
Tamil Nadu, India

ABSTRACT: Heart disease is one of the most significant causes of mortality in the world today. Prediction of cardiovascular disease is a critical challenge in the area of clinical data analysis. Machine learning (ML) has been shown to be effective in assisting in making decisions and predictions from the large quantity of data produced by the healthcare industry. The ML techniques have been used in recent developments in different areas of the Internet of Things (IoT). Various studies give only a glimpse into predicting heart disease with ML techniques. In this project, a novel method that aims at ending significant features by applying machine learning techniques resulting in improving the accuracy in the prediction of cardiovascular disease is proposed. The prediction model is introduced with different combinations of features and several known classification techniques. An enhanced performance level with an accuracy level of 88.7% through the prediction model for Heart disease with the Hybrid Random Forest with a Linear Model (HRFLM) is proposed.

KEYWORDS : Heart disease, Cardiovascular Disease, Data mining techniques, Machine Learning, Decision Tree.

I. INTRODUCTION

To apply machine learning techniques resulting in improving the accuracy in the prediction of cardiovascular disease.

Heart disease is one of the complex diseases and globally many people suffered from this disease. On time and efficient identification of heart disease plays a key role in healthcare, particularly in the field of cardiology. The system proposed is an efficient and accurate system to diagnosis heart disease and the system is based on data mining and machine learning techniques. Data mining helps in finding out patterns in the data collected, by utilizing data mining we can assess numerous examples which will be used in the future to settle on wise frameworks and choices. Data mining alludes to different strategies for distinguishing data or the reception of arrangements dependent on learning and information extraction of this information with the goal that they can be utilized in different territories, for decision making and prediction calculation of probability.

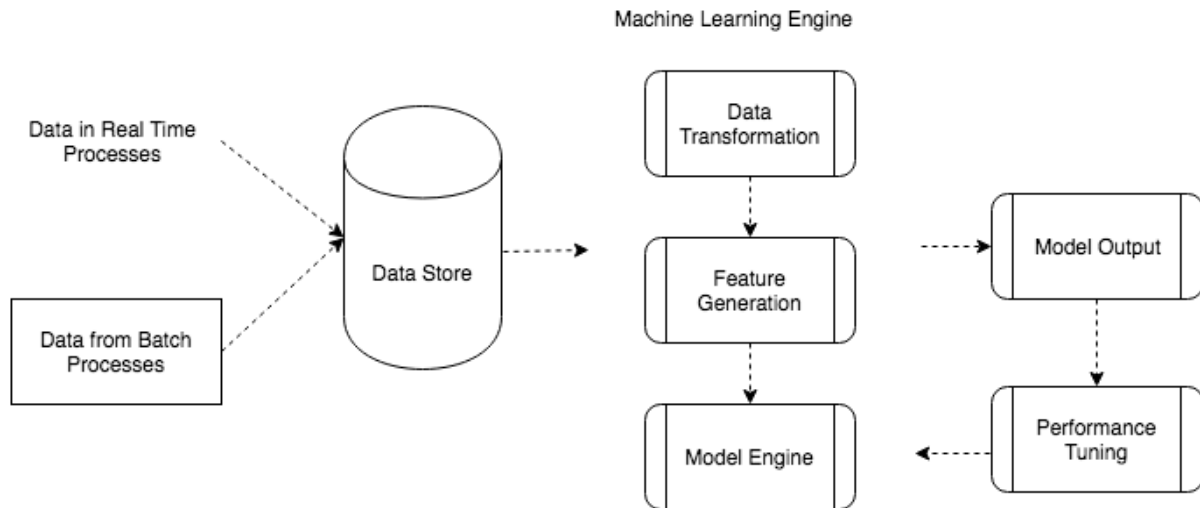


FIG 1.1 MACHINE LEARNING BLOCK DIAGRAM

II. LITARATURE SURVEY

Heart disease is a word that assigns to a huge number of clinical conditions associated to heart. These clinical conditions explain the strange health conditions that straight influence the heart and all its part. In this paper discussed about analyzing the various data mining techniques introduced in recent years for heart disease prediction. Some of papers mentioned as follows:

In this section the data mining systems used for the classification of heart disease is analyzed.

In year 2000, Shusaku Tsumoto [2] performed a work, "Problems with Mining Medical Data". In this paper well projected that data mining methods will find interesting patterns from databases as use again of stored data and be important for clinical research and perform, because human beings cannot arrangement with such a huge amount of data. In this paper alert on the characteristics of clinical data and talk about how data miners deal with clinical data.

In year 2004, Y. Alp Aslandogan, et. al. [3] performed a work, "Evidence Combination in Medical Data Mining". In this paper merged the viewpoint of three classifiers: KNearest Neighbor (K-NN), Naïve Bayesian and Decision Tree. Dempster's rule of combination combines three viewpoints to appear at one concluding decision. The experiments with k-fold cross validation illustrate that the nature of the data set has an improved impact on some classifiers than others and the classification based on joint idea shows improved largely accuracy than any individual classifier. Compared the performance of Dempster's combination (with differentiation-based uncertainty assignment) with those of performance-based linear and majority choose combination models and circumstances beneath which the proof combination gives to improves classification.

In year 2004, The inhibited problem to identify and to predict the rules of association for the heart disease according to Carlos Ordonez as presented in "Improving Heart Disease Prediction Using Constrained Association Rules," [4]. In this paper assessed set of data is encompassed medical records of the people of heart disease with attributes for the risk factors, measurements of heart perfusion and narrowed artery. The three constraints were introduced to reduce the number of patterns, they are as follows: 1) the attributes have to appear on one side of the rule only. 2) The rule separates the attributes into the uninteresting groups. 3) The number of attributes from the rule is controlled by the medical records of the people of heart disease finally. Further falling the running time as per the experiments illustrated the constraints of showing rules have been extremely decreased the number. The author, Carlos Ordonez, anticipated the presence or the absence of heart disease in four specific heart arteries into the two groups of rules. Data mining methods may aid the clinicians in the prediction of the survival of patients and in the adaption of the practices consequently.

III. PROPOSED METHODOLOGY DISCUSSION

3.1 EXISTING SYSTEMS :

The obtained results are compared with the results of existing models within the same domain and found to be improved. The data of heart disease patients collected from the UCI laboratory is used to discover patterns with NN, DT, Support Vector machines SVM, and Naive Bayes. The results are compared for performance and accuracy with

these algorithms. The proposed hybrid method returns results of 86:8% for *F*-measure, competing with the other existing methods.

3.1.1 DRAWBACKS :

- If the number of observations is lesser than the number of features, it may lead to overfitting resulting in reduction in accuracy.
- For very large data sets, the size of the trees can take up a lot of memory.
- In cases where the number of features for each data point exceeds the number of training data samples, the SVM will underperform.

3.2 PROPOSED SYSTEM :

We have used python and pandas operations to perform heart disease classification of the Cleveland UCI repository. It provides an easy-to-use visual representation of the dataset, working environment and building the predictive analytics. ML process starts from a pre-processing data phase followed by feature selection based on data cleaning, classification of modeling performance evaluation, and the results with improved accuracy.

3.2 ARCHITECTURE DIAGRAM :

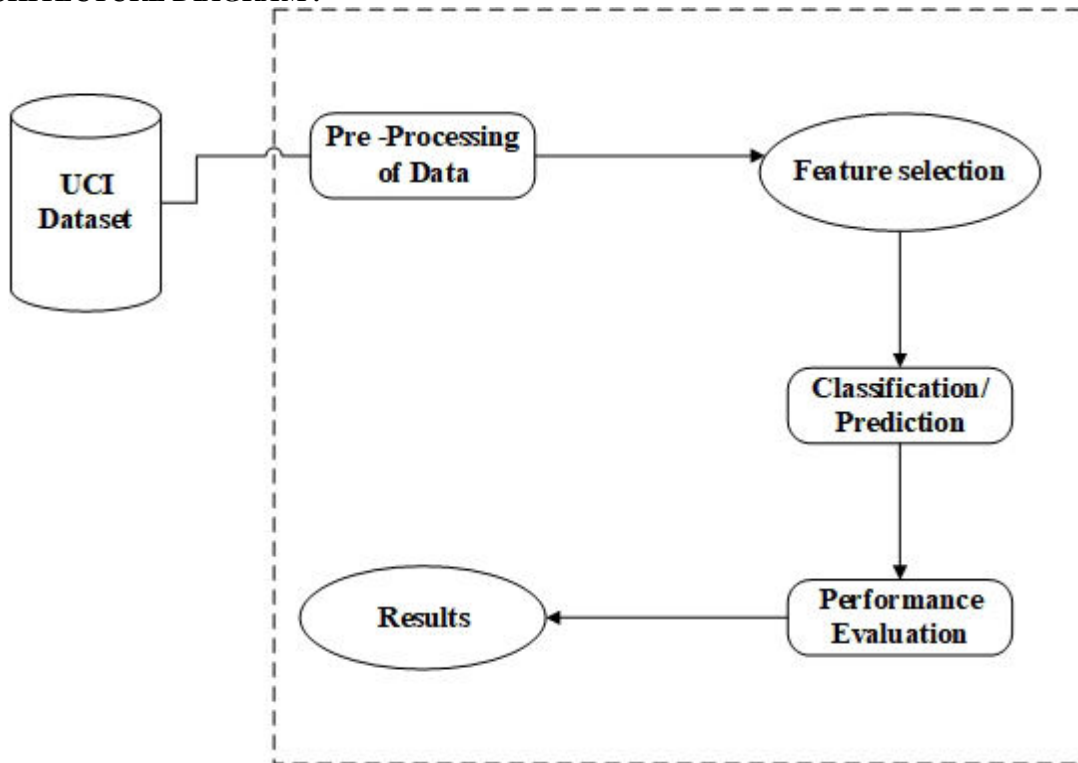


FIG 3.2 ARCHITECTURE DIAGRAM

3.3 Data Pre-Processing

Heart disease data is pre-processed after collection of various records. The dataset contains a total of 303 patient records, where 6 records are with some missing values. Those 6 records have been removed from the dataset and the remaining 297 patient records are used in pre-processing.

3.4 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.



3.5 Classification Modeling

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.

IV. EXPERIMENTAL RESULT

Figures shows the results of Identifying the Heart Disease by using Machine Learning Techniques based on K-NN and NAÏVE BAYES algorithm.

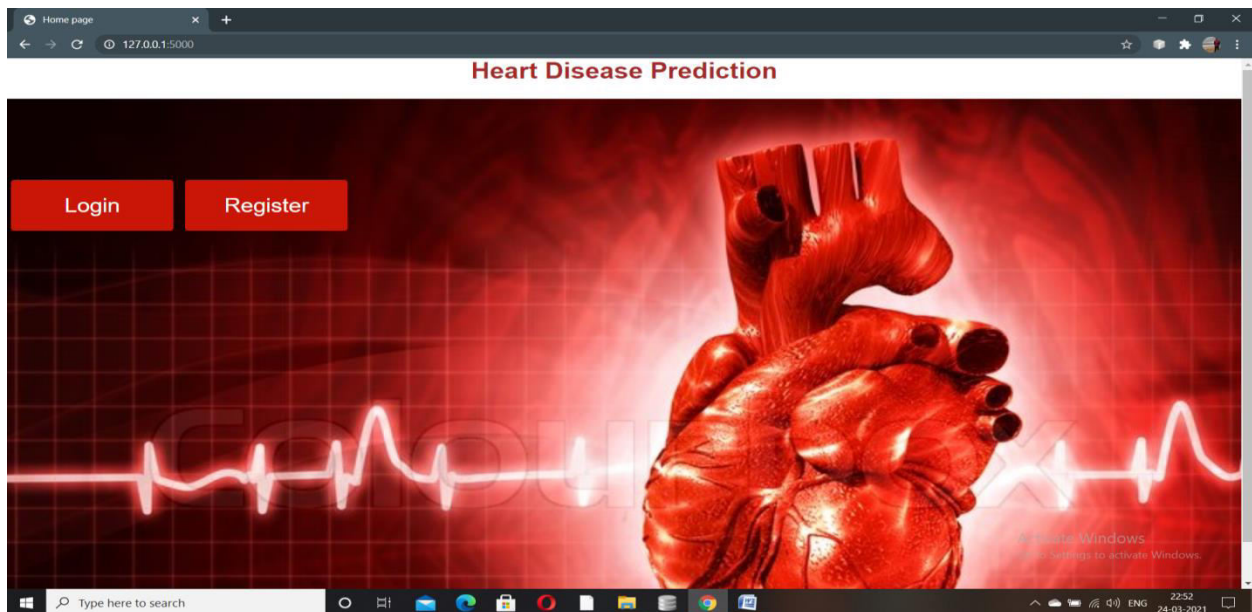


FIG 4.1. HOME PAGE

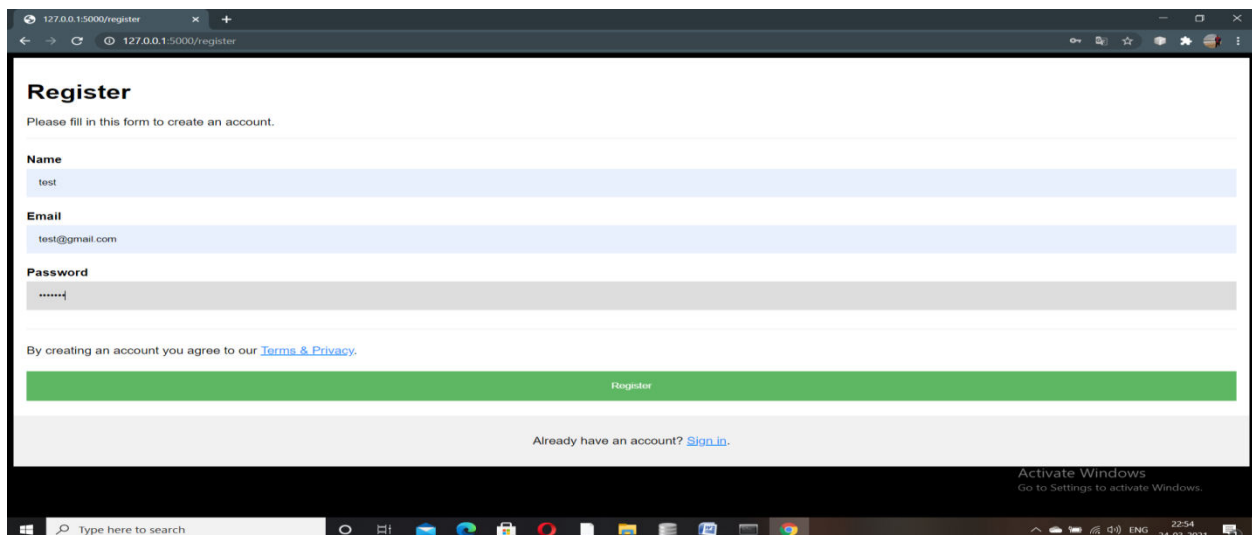


FIG 4.2. REGISTRATION NAME

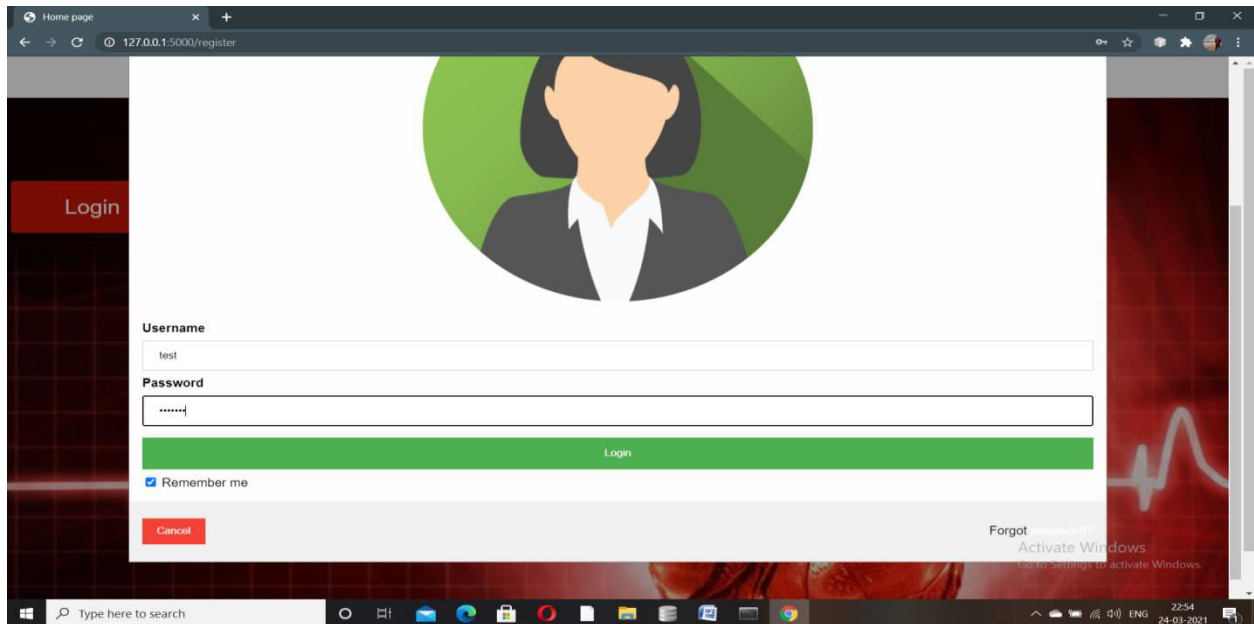


FIG 4.3.LOGIN PAGE

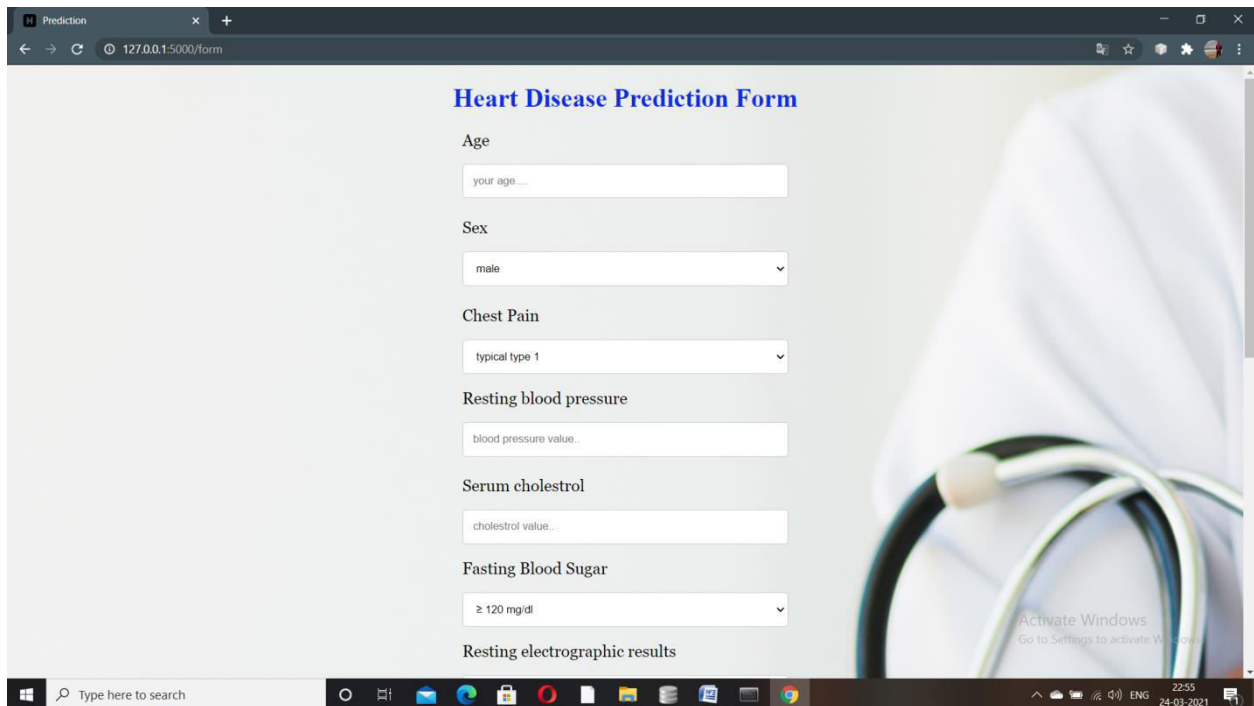


FIG 4.4.HEART DISEASE PREDICTION PAGE

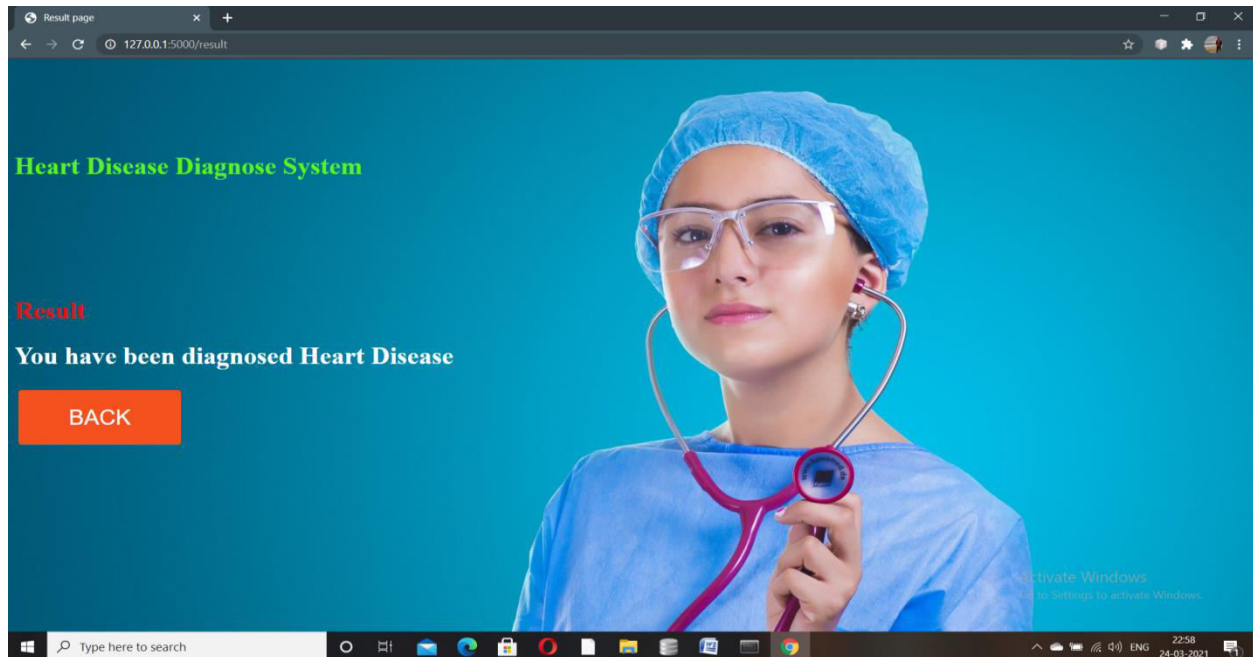


FIG 4.5.RESULT PAGE

V. CONCLUSION AND FUTURE WORK

The overall objective of the work is to predict more exactly the occurrence of heart disease using data mining techniques. In this research work, the UCI data repository is used for performing the comparative analysis of three algorithms such as Random Forest, Decision trees and Naive Bayes. From the research work, it has been experimentally proven that Random Forest provides perfect results as compare to Decision tree and Naive Bayes. The Future work of this project can be made to produce an impact in the accuracy of the Decision Tree and Bayesian Classification for additional improvement after applying genetic algorithm in order to decrease the actual data for acquiring the optimal subset of attribute that is enough for heart disease prediction. The automation of heart disease prediction using actual real time data from health care organizations and agencies which can be built using big data. They can be fed as a streaming data and by using the data, investigation of the patients in real time can be prepared.

REFERENCES

- [1] A. S. Abdullah and R. R. Rajalaxmi, "A data mining model for predicting the coronary heart disease using random forest classifier," in Proc. Int. Conf. Recent Trends Comput. Methods, Commun. Controls, Apr. 2012, pp. 22_25.
- [2] A. H. Alkeshuosh, M. Z. Moghadam, I. Al Mansoori, and M. Abdar, "Using PSO algorithm for producing best rules in diagnosis of heart disease," in Proc. Int. Conf. Comput. Appl. (ICCA), Sep. 2017, pp. 306_311.
- [3] N. Al-milli, "Backpropagation neural network for prediction of heart disease," J. Theor. Appl. Inf. Technol., vol. 56, no. 1, pp. 131_135, 2013.
- [4] C. A. Devi, S. P. Rajamhoana, K. Umamaheswari, R. Kiruba, K. Karunya, and R. Deepika, "Analysis of neural networks based heart disease prediction system," in Proc. 11th Int. Conf. Hum. Syst. Interact. (HSI), Gdansk, Poland, Jul. 2018, pp. 233_239.
- [5] P. K. Anooj, "Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules," J. King Saud Univ.-Comput. Inf. Sci., vol. 24, no. 1, pp. 27_40, Jan. 2012. doi: 10.1016/j.jksuci.2011.09.002.
- [6] L. Baccour, "Amended fused TOPSIS-VIKOR for classification (ATOVIC) applied to some UCI data sets," Expert Syst. Appl., vol. 99, pp. 115_125, Jun. 2018. doi: 10.1016/j.eswa.2018.01.025.
- [7] C.-A. Cheng and H.-W. Chiu, "An artificial neural network model for the evaluation of carotid artery stenting prognosis using a national-wide database," in Proc. 39th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC), Jul. 2017, pp. 2566_2569.
- [8] H. A. Esfahani and M. Ghazanfari, "Cardiovascular disease detection using a new ensemble classifier," in Proc. IEEE 4th Int. Conf. Knowl.- Based Eng. Innov. (KBEI), Dec. 2017, pp. 1011_1014.



- [9] F. Dammak, L. Baccour, and A. M. Alimi, "The impact of criterion weights techniques in TOPSIS method of multi-criteria decision making in crisp and intuitionistic fuzzy domains," in Proc. IEEE Int. Conf. Fuzzy Syst. (FUZZ-IEEE), vol. 9, Aug. 2015, pp. 1_8.
- [10] R. Das, I. Turkoglu, and A. Sengur, "Effective diagnosis of heart disease through neural networks ensembles," Expert Syst. Appl., vol. 36, no. 4, pp. 7675_7680, May 2009. doi: [10.1016/j.eswa.2008.09.013](https://doi.org/10.1016/j.eswa.2008.09.013).



INNO  **SPACE**
SJIF Scientific Journal Impact Factor

**Impact Factor:
7.512**

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



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