International Journal for Modern Trends in Science and Technology, 9(05): 545-548, 2023

ISSN: 2455-3778 online
DOI: https://doi.org/10.46501/IJMTST0905092
Available online at: http://www.ijmtst.com/vol9issue05.html

Hybrid Machine Learning Classification Technique to Improve the Accuracy of Heart Disease

U. Mounika | V. Radhika | P. Jaya Sree | Sk. ArshiyaT

Department of CSE, Narayana Engineering College , Gudur , India.

To Cite this Article
U. Mounika, V. Radhika, P. Jaya Sree and Sk. ArshiyaT. Hybrid Machine Learning Classification Technique to Improve the Accuracy of Heart Disease. International Journal for Modern Trends in Science and Technology 2023, 9(05), pp. 545-548. https://doi.org/10.46501/IJMTST0905092

Article Info
Received: 16 April 2023; Accepted: 10 May 2023; Published: 19 May 2023.

ABSTRACT

Heart disease causes a significant mortality rate around the world, and it has become a health threat for many people. Early prediction of heart disease may save many lives; detecting cardiovascular diseases like heart attacks, coronary artery diseases etc., is a critical challenge by the regular clinical data analysis. Machine learning (ML) can bring an effective solution for decision making and accurate predictions. The medical industry is showing enormous development in using machine learning techniques. In the proposed work, a novel machine learning approach is proposed to predict heart disease. The proposed study used the Cleveland heart disease dataset, and data mining techniques such as regression and classification are used. Machine learning techniques Random Forest and Decision Tree are applied. The novel technique of the machine learning model is designed.

KEYWORDS: Machine Learning, Hybrid model, RF, Decision tree, XGBoost.

1. INTRODUCTION

Some of the related works represent various convenient methods with the implication but none of the methods aid professionals under different characteristics. Therefore the design and implementation of these methods pave the way for further research. Additionally, the presented work indicates that the utilization of the data mining method works better than other approaches. With a discussion of research objectives, motivation, and key findings this chapter describes the contribution towards the direction to improve the QOS of employing a complete list of features that are associated with the selected dataset.

The data mining is the process of extracting unknown and predictive information from huge amount of data. It is an innovative tool with great potential to help companies and mainly focus on the most essential information in their data warehouses. Most commonly data mining is also known as Knowledge Discovery in Databases (KDD).the system. Selection and formation are the most appropriate features instead of employing a complete list of features that are associated with the selected dataset.The data mining is the process of extracting unknown and predictive information from huge amount of data.

It is an innovative tool with great potential to help companies and mainly focus on the most essential information in their data warehouses. Most commonly data mining is also known as Knowledge Discovery in Databases (KDD). KDD is the important process of identifying valid, new, potentiallyuseful, and finally understandable patterns in data. Knowledge discovery process has iterative sequential steps of processes and data mining is one of the KDD processes.

2. LITERATURE SURVEY

There are many current works studied by the researchers about heart disease prediction and analysis. Some of such works are addressed below.

The author studies heart disease using the random forest in [1] with the Cleveland dataset. The author used the Chi Square feature selection model and genetic algorithm (GA) based feature selection model for the study. They proved in the experimental results that their proposed model with Genetic algorithm feature selection has given high accuracy than the existing models. However, the results are evaluated with existing machine learning models.

In [2], the author has generated specific rules based on this PSO algorithm and evaluated different rules to get a more accurate rule for heart disease identification. After evaluating the rules, C 5.0 is used for the classification of disease based on binary classification. The author used UCI repository data for implementation and evaluated high accuracy using PSO and the Decision tree algorithm[8].

3. PROPOSED WORK

A hybrid model is a novel technique, which uses the probabilities arrived from one machine learning model is given as input to the other machine learning model. This hybrid model gives us the better-optimized results based on both machine learning algorithm, which is considered for the implementations. The architecture of the proposed system is shown below.

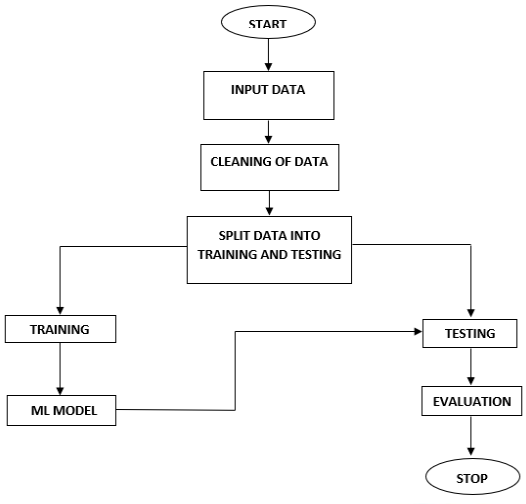

Fig. 1: System workflow.

Implementation Modules

1. System:

1.1 Store Dataset: The System stores the dataset given by the user.

1.2 Pre-processing

To treat null values, outliers and Label Encoding.

1.3 Model Training:

The system takes the data from the user and fed that data to the selected model.

2 1.4 Graphs Generation:

3 The system takes the dataset given by the user, selects the model and generates the accuracy corresponding to the selected model

4 2. User:

5 2.1 Upload Dataset:

6 The user can load the dataset he/she want to work on.

7 2.2 View Dataset:

8 The User can view the dataset.

9 2.3 Select model:

User can apply themodel to the dataset for accuracy.

2.4 Graphs:

User can evaluate the model performance using the graphs.

2.5Prediction:

Passing parameters to predict the output

Implementation Algorithms

DECISION TREE

Decision tree is a type of flowchart that shows a clear pathway to a decision. In terms of data analytics, it is a type of algorithm that includes conditional 'control' statements to classify data. A decision tree starts at a single point (or 'node') which then branches (or 'splits') in two or more directions. Each branch offers different possible outcomes, incorporating a variety of decisions and chance events until a final outcome is achievedDecision trees are extremely useful for data analytics and machine learning because they break down complex data into more manageable parts. They're often used in these fields for prediction analysis, data classification, and regression

- Decision nodes: Representing a decision (typically shown with a square)
- Chance nodes: Representing probability or uncertainty (typically denoted by a circle)
- End nodes: Representing an outcome (typically shown with a triangle)

Connecting these different nodes are what we call 'branches'. Nodes and branches can be used over and over again in any number of combinations to create trees of various complexity.

RANDOM FOREST

Random forest is a flexible, easy to use machine learning algorithm that produces, even without hyper-parameter tuning, a great result most of the time. It is also one of the most used algorithms, because of its simplicity and diversity (it can be used for both classification and regression tasks).Random forest is a supervised learning algorithm. The "forest" it builds, is an decision trees, usually trained with the "bagging" method. The general idea of the bagging method is that a combination of learning models increases the overall result.

XGBoost

XGBoost is the most popular machine learning algorithm. Extreme Gradient Boosting (XGBoost) is similar to gradient boosting framework but more efficient. It has both linear model solver and tree learning algorithms. So, what makes it fast is its capacity to do parallel computation on a single machine.

This makes XGBoost at least 10 times faster than existing gradient boosting implementations. It supports various objective functions, including regression, classification and ranking XGBoost is an efficient and straightforward to use algorithm.

4. RESULTS

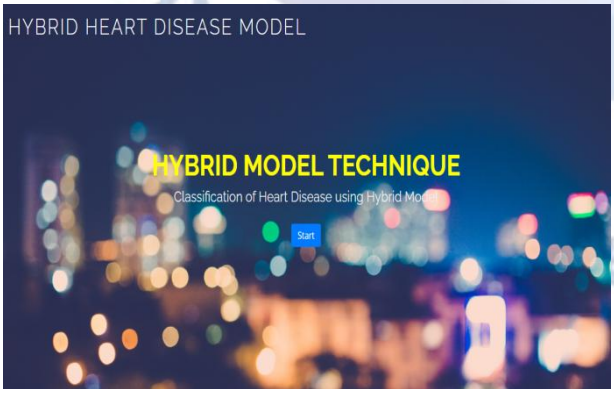

Fig. 2: Home Page

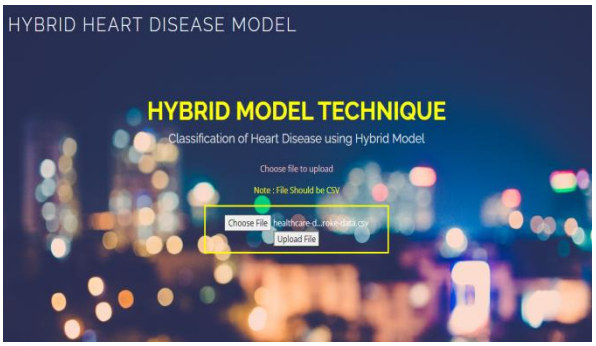

Fig. 3: Upload Dataset

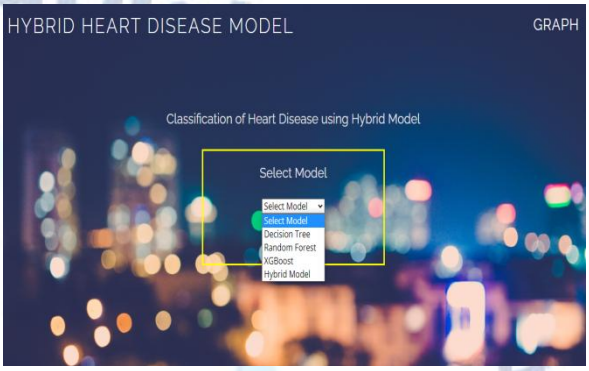

Fig. 4: Model Selection

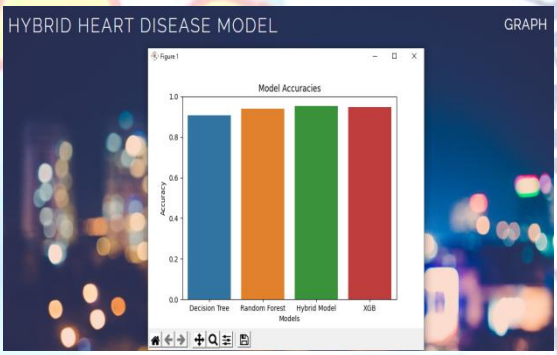

Fig. 5: Comparative Analysis of model Accuracies

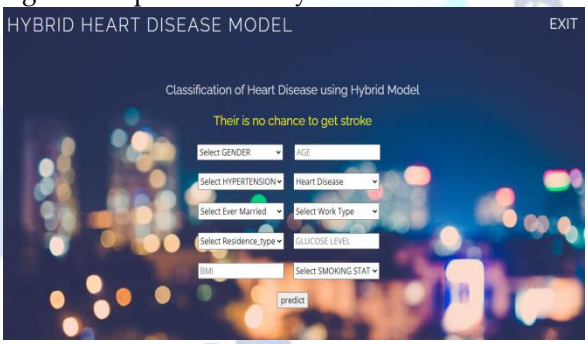

Fig. 6: Output of the Proposed Model

5. CONCLUSION

The work that was carried out in this investigation endeavours to improve efficiency, suitability, and QoS. The characteristics and limitations of existing methods were discussed in the literature survey to build a more efficient method. The proposed work investigates four different algorithms such as the Random Forest, XGBoost and a form of Decision Tree (10). The proposed

method robustly analyses these four methods to exploited statistics and opts for the pair of the finest algorithm that utilizes a linear model based on the feature selection process with best-first search and Gain ratio along with the Ranker method. Several simulations have been carried out to demonstrate the efficiency of the proposed approach. Each comparison has indicated that the proposed approach effectively improves the issues of traditional as well as modern algorithms.

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

REFERENCES

[1] Ritu. Sharma, Mr Shiv Kumar, Mr. RohitMaheshwari "Comparative Analysis of Classification Techniques in DataMining Using Different Datasets" International Journal of Computer Science and Mobile Computing, IJCSMC, Vol. 4, Issue. 12, December 2015, pp.-125 – 134.

[2] SobhanSarkar, Atul Patel, SarthakMadaan, JhareswarMaiti "Prediction of Occupational Accidents Using DecisionTree Approach" IEEE Annual India Conference (INDICON), 2016, pp.- 1-6.

[3] AayushiVerma, Shikha Mehta "A Comparative Study of Ensemble LearningMethods for Classification in Bioinformatics" IEEE 7th International Conference on Cloud Computing, Data Science & Engineering – Confluence, 2017, pp.- 155-158.

[4] K. C. Giri, M. Patel, A. Sinhal and D. Gautam, "A Novel Paradigm of Melanoma Diagnosis Using Machine Learning and Information Theory," 2019 International Conference on Advances in Computing and Communication Engineering (ICACCE), Sathyamangalam, Tamil Nadu, India, 2019, pp. 1-7, doi: 10.1109/ICACCE46606.2019.9079975.

[5] AyisheshimAlmaw, KalyaniKadam "Survey Paper on Crime Prediction using EnsembleApproach" International Journal of Pure and Applied Mathematics, Volume 118 No. 8 2018, pp.-133-139.

[6] ShakuntalaJatav and Vivek Sharma "An Algorithm For Predictive DataMining Approach In Medical Diagnosis" International Journal of Computer Science & Information Technology (IJCSIT) Vol 10, No 1, February 2018, pp.- 11-20.

[7] Han Wu, Shengqi Yang, Zhangqin Huang, Jian He, Xiaoyi Wang "Type 2 diabetes mellitus prediction model based on data mining" ELSEVIER Informatics in Medicine Unlocked, 2018, pp.- 100-107.

[8] Patel M., Choudhary N. (2017) Designing an Enhanced Simulation Module for Multimedia Transmission Over Wireless Standards. In: Modi N., Verma P., Trivedi B. (eds) Proceedings of International Conference on Communication and Networks. Advances in Intelligent Systems and Computing, vol 508. Springer, Singapore. https://doi.org/10.1007/978-981-10-2750-

[9] V.Sucharita, S.Jyothi, P.Venkateswara Rao " Comparison of Machine Learning Algorithms for the classification of Penaeid Prawn Species" in IEEEXplore. 2016

[10] V.Sucharita,P.Venkateswara Rao,A.Rammohan Reddy" Advances in Machine Learning Techniques for Penaeid Shrimp Disease Detection: A Survey" IJEAS, ISSN: 2394-3661, Volume-3, Issue-8, August 2016.