Heart Disease Prediction Using ML with Increased Accuracy

Ms. B.Vanitha (M.C.A). Rajeev Gandhi Memorial college Of Engineering and Technology, Nandyal

Mrs.G. Naga pavani (MTech.,)Assistant professor. Rajeev Gandhi Memorial college Of Engineering and Technology, Nandyal

Abstract

Some of the connected works depict a variety of practical techniques with the impression that none of the techniques help experts in different ways. The development and application of these techniques thus pave the path for additional study. The work that has been given also shows that using the data mining method is more effective than using other strategies. This chapter discusses the contribution to the direction to improve the QOS of the system with a discussion of research aims, motivation, and significant findings. Instead of using a whole list of features connected to the chosen dataset, selection and formation are the most applicable characteristics. The practice of removing unknown and prognostic information from a vast volume of data is known as data mining. It is a cutting-edge solution with huge potential to assist businesses and principally concentrate on the most crucial data in their data warehouses. Data mining is most frequently referred to as Knowledge Discovery in Databases (KDD). KDD is a crucial procedure for finding true, fresh, possibly beneficial, and ultimately understood patterns in data. Data mining is one of the iterative sequential elements that make up the knowledge discovery process (KDD). Over the past ten years, undertaking research utilizing ensemble learning approaches has piqued the curiosity of numerous scholars. Several authors have found a significant improvement in performances when using ensemble techniques. These ensemble methodologies have broadened their application in a number of industries, including aerospace, automotive, financial services, health care, and manufacturing. An ensemble of classifiers is a collection of independent classifiers that, when combined and evaluated separately, aid in the classification of fresh test samples. Ensemble learning has developed into one of the most in-depth fields of study for machine learning researchers in supervised learning.

1. INTRODUCTION

1.1 Introduction

According to research, this composite model's performance has largely surpassed that of a single model when employing ensemble classification. Multiple experimental studies and analyses have shown that integrating the outputs of different classifiers reduces the generalized error. The main goal of employing ensemble approaches is to create an ensemble model that accurately combines a variety of different classifier algorithms. Due to the enormous popularity of the topic of bioinformatics in the scientific community, we have concentrated our attention on it in this work. The following is a description of the several ensemble learning techniques employed in this paper. Bagging is a popular ensemble learning methodology that was developed by Bremen and is based on the Bootstrap sampling method. To build the individual classifier of

the same algorithm or to change the choice of training data, samples known as bags are made in this process. In this case, the data point is chosen at random in conjunction with the replacement approach, meaning that some data is obtained from a random sample and some is missing from the original dataset. From each bag, a different classifier is learned. By giving each classifier a vote individually, the test sample is predicted by the combination of all the classifiers built in the earlier phase. Another effective ensemble mechanism based on learning in succession is called "boosting," which was put forth by Freund and Shapira. The learning process begins with the overall, comprehensive data set, and is followed by the outcome data.

2. Literature Survey

S. NO	Journal Type with year	Authors	Title	Outcomes
1	Journal, 2015	Ritu. Sharma, Mr Shiv Kumar, Mr. RohitMaheshwari	Comparative Analysis of Classification Techniques in Data Mining Using Different Datasets"	comparative analysis of several data mining classification techniques on the basis of parameters accuracy, execution time, types of datasets and applications
2	IEEE 2016	SobhanSarkar, Atul Patel, SarthakMadaan, JhareswarMaiti	Prediction of Occupational Accidents Using Decision Tree Approach	predictive model which not only could predict the occupational incidents but also provide rules for explaining accident scenarios like near-miss, property damage, or injury cases. Classification and regression tree (CART) is used for prediction purpose.

S. NO	Journal Type with year	Authors	Title	Outcomes
3	IEEE 2017	AayushiV erma, Shikha Mehta	A Comparative Study of Ensemble Learning Methods for Classification in Bioinformatics	novel ensemble learning approach "BBS method" which stands for Bagging, Boosting and Stacking with appropriate base classifiers for the classification of the five UCI datasets taken from the field of Bioinformatics
4	Conference , 2019	K. C. Giri, M. Patel, A. Sinhal and D. Gautam,	A Novel Paradigm of Melanoma Diagnosis Using Machine Learning and Information Theory	demonstrate that Havrda Entropy and Harris Corner Detector based melanoma analysis approach accomplish greater affectability

S. NO	Journal Type with year	Authors	Title	Outcomes
5	Journal 2018	AyisheshimAlma w, KalyaniKadam	Survey Paper on Crime Prediction using EnsembleApproach	Crime forecasting is a way of trying to mining out and decreasing the upcoming crimes by forecasting the future crime that will occur
6	Journal 2018	ShakuntalaJatav and Vivek Sharma	An Algorithm For Predictive DataMining Approach In Medical Diagnosis	The data mining classification techniques, namely Support Vector Machine(SVM) and Random Forest (RF) are analyzed on Diabetes, Kidney and Liver disease database. The performance of these techniques is compared, based on precision, recall, accuracy, f_measure as well as time

3. OVERVIEW OF THESYSTEM

3.1 Existing System

Machine learning has led to the division of computer approaches into two groups: traditional methods and machine learning methods. The identification of sentimental analysis and how machine learning techniques outperform conventional methods are covered in this section. The current methodology in this project has a specific flow, and traditional sentimental analysis is also used for development. It requires a lot of memory, though, and the results are not exact.

3.1.1 Disadvantages of Existing System

- Less feature compatibility
- Low accuracy.

3.2 Proposed System

After analyzing every strategy already in use, several researchers outlined the many benefits of each suggested methodology and commented on a number of limitations that are still connected to practical approaches and have a significant impact on how well the techniques' function. Some of the main obstacles include rigidity, which makes developing a model time-consuming, alternate parameter, and erroneous judgments, among other related problems.

3.3 Methodology
Upload File: uploads CSV (pre-processed dataset) file

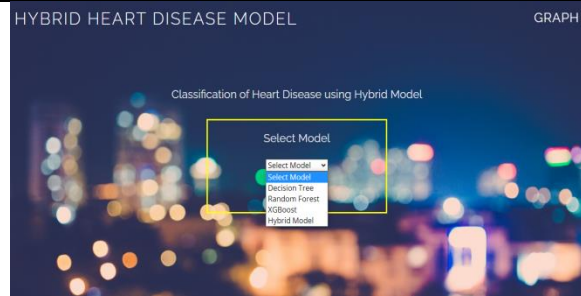**View Dataset:** View all the content in the dataset.

Pre-processing: To clear all the null values and
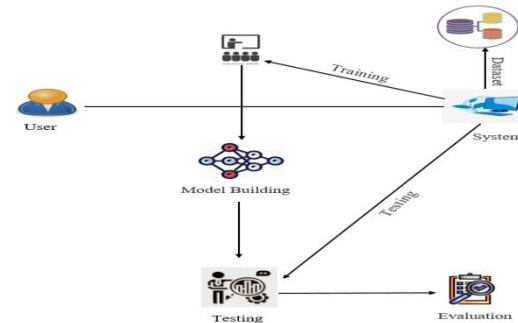
outliers and performing Label Encoding

Model Training: The dataset divides into train and test training will be done on train set

Graph: view graph accuracy

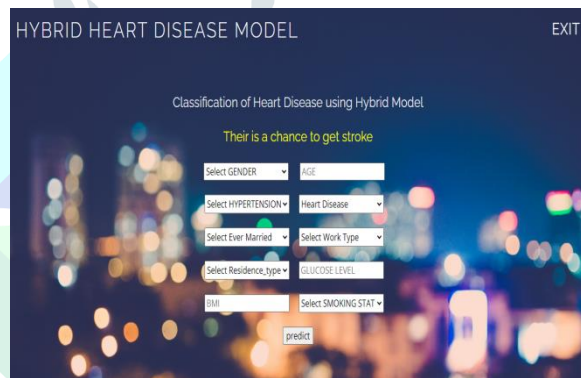**Predict:** we have to pass values to predict that person commit to suicide or not
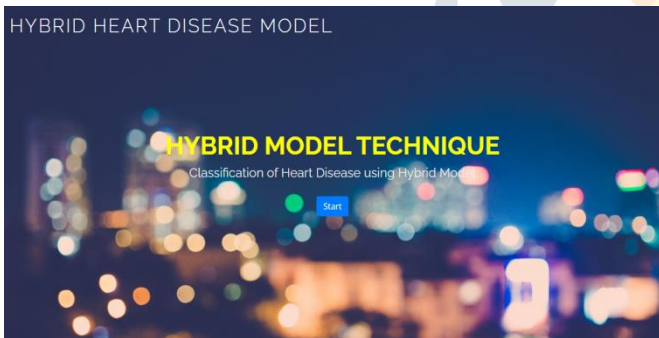
.

4   Architecture

Fig 1: Frame work of proposed method

Above architecture diagram shows three stages of data flow form one module to another module. Data collection, preprocessing, and algorithm training.
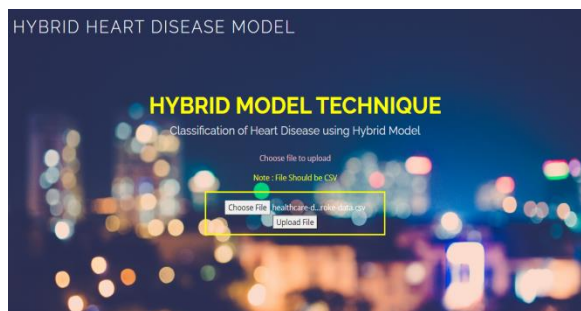
5   RESULTS SCREEN SHOTS

Home Page:

Upload Data:

Choose options:

Predict Result:

6. CONCLUSION

✓   The work done for this inquiry is to increase effectiveness, suitability, and QoS. In order to develop a more effective approach, the characteristics and restrictions of existing methods were reviewed in the literature review. The proposed research examines four distinct algorithms, including the Random Forest, XGBoost, and a decision tree variant (J48). The suggested technique thoroughly evaluates these four methods for exploiting statistics and selects the two best algorithms, which combine the Ranker method and a linear model based on feature selection with best-first search and Gain ratio. Numerous simulations have been run to show how effective the suggested strategy is. Every comparison has shown that the suggested strategy significantly resolves the problems with traditional methods.

Future Enhancement

- ✓ • The future capability of this program to recognize heart attacks.
- ✓ • Using the updated data set, we plan to evaluate the prediction approach.
- ✓ • The most reliable and pertinent machine learning detection techniques.

7. 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 Decision Tree Approach" IEEE Annual India Conference (INDICON), 2016, pp. - 1-6.

[3] AayushiVerma, Shikha Mehta "A Comparative Study of Ensemble Learning Methods 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. Gautama, "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 Ensemble Approach" 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., Varma 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-