

Enhancing Heart Disease Prediction Accuracy Using Hybrid Machine Learning Techniques

¹B. Naresh, ²C. Jyothi

¹Assistant Professor, ²M.Tech Student, Department of CSE,
Megha Institute Of Engineering And Technology For Womens,
Eduleabad (Village), Ghatkesar (Mandal), Medchal District, Telangana

Abstract—Researchers have shown a lot of interest in the field of medical science. A fair amount of researchers have identified many reasons for early death in humans. According to the relevant literature, there are several causes of diseases, and heart-related illnesses are one of them. In an effort to save lives and aid medical professionals in the detection, prevention, and management of cardiovascular disease, several researchers have put forward unique approaches. Every effective plan has its limitations, but there are several easy approaches that help the expert make a conclusion. The two feature selection approaches, Correlation Based Feature Selection (CFS) and Gain Ratio, as well as the Hidden Markov Model (HMM), Artificial Neural Network (ANN), Support Vector Machine (SVM), and Decision Tree J48, are thoroughly examined using the suggested method. When applied to a separate set of data, the Ranker technique is complemented by the Gain Ratio. The proposed technique analyses the process and then intelligently constructs Naive Bayes processing by combining the two best processes using an appropriate layered architecture. Choosing the best approach and comparing the available schemes with various characteristics for statistical analysis is the primary goal at the outset.

Keywords— Machine learning, Classification Technique, Naïve Bayes, neural networks, supervised machine learning

I. INTRODUCTION

While a number of similar publications provide numerous implication-based, handy approaches, none of these ways really help professionals in any way, shape, or form [1-4]. Consequently, new lines of enquiry may be opened thanks to the development and use of these methodologies. Data mining is also shown to be more effective than other methods in the current research [5-8]. In this chapter, we will examine the study aims, motivation, and major

Results. We will also outline the contribution towards improving the system's QoS. Instead of using a whole set of features linked with the chosen dataset, selection and formation are the best features to use.

II. ISSUES WITH PREDICTION SYSTEM

1. It's difficult to keep up with the difference in data qualities, and a lot of the algorithms that are now accessible don't have the right methodology to help experts in diverse fields.

2. In a real-time setting, when cost is a crucial consideration, most algorithms were designed to complete a single job and are therefore unsuitable.
3. The technique's execution relies on the altered approach without thinking about it, and one of the main problems is that it exploits the random method.
4. A big problem with accessible algorithms is that they still make a lot of inaccurate predictions.
5. Most of the newly proposed algorithms, like the traditional techniques, have unique constraints and have the desire to eliminate the impediments.

III. PREVIOUS MODELS

This statistical model was introduced by Baum and Petrie in 1966 [9] and is known as the Hidden Markov Model. It is the simplest Bayesian network that can detect the Markov process, to put it simply [10]. Optimal nonlinear filtering problems involving stochastic processes are associated with this. Although

this model was originally developed for voice recognition, it has now found applications in several domains including as pattern recognition, handwriting and gesture identification, part-of-speech tagging, music scores, partial discharges, bioinformatics, and more. Secondly, there is Support Vector Machine (SVM), a supervised learning technique in machine learning that can do regression and classification.

SVM is a linear algorithm. More than that, this approach works better for fixing a wide range of real-world problems. Conventionally, it is a simple strategy to construct a line or hyper plane that partitions data into categories. Much of the biological and auxiliary scientific fields, as well as picture classifications and segmentation, handwriting recognition, and the categorization of hyper and basic language are better served by this system [11]. 3. A Neural Network (ANN) that is Artificial A model of information processing, Artificial Neural Networks (ANN) are stimulated by a biologically sensitive technique similar to the information processing in a brain. [12] It is a well-known arrangement of neurons that work together to accomplish a certain goal. An interconnected processing node with direct linkages to other nodes in the network allows it to complete a job. There is a processing unit for every node, and the connections between them are shown by the links [13].

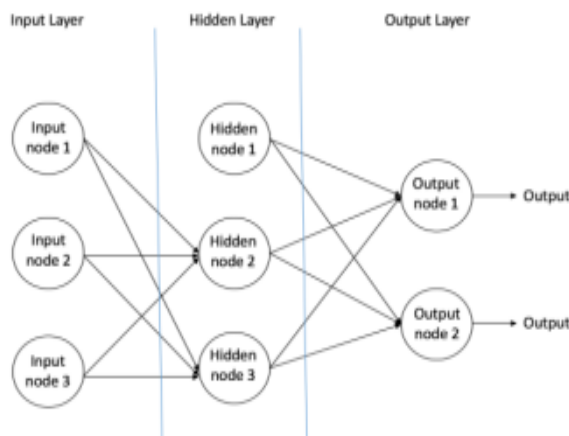


Figure 1 Neural Network

This plan, which depicts the accepted way of data categorization for massive datasets, is a Decision Tree in scheme form. Every inner node in this model's data tree represents a determination, and each leaf describes a class. According to [14], this approach usually involves recursively splitting a branch into sub-branches and continuing to perform the procedure until the problem is not addressed. [15] A built tree is determined by the proportionate evaluation of the branch qualities, which constitute its nodes. Its straightforward operational functions and strong results have placed it ahead of many others. At last, a decision tree combines the Tree Construction and Tree Pruning processes [16].

IV. PROPOSED METHODOLOGY

Following a comprehensive evaluation of all available methods, a number of researchers highlighted the benefits of each proposed approach and focused on a number of limitations that are still present with practical approaches and significantly impact the methods' operational behaviour. Inflexibility, lengthy model construction, lack of alternate parameters, and erroneous judgements are a few of the major constraints among many related concerns. 1. Suggested Plan Two feature compression techniques, CFS with best-first search and Gain ratio with ranker mechanism, and four distinct classification algorithms were chosen for the proposed study.

It is not possible to construct a more competent approach using a procedure that is optimised for each algorithm, as stated in the literature review. In addition to two additional feature compression approaches, the suggested approach investigates and analyses four selected methods: Decision Tree (J48), Artificial Neural Network (ANN), Hidden Markov Model (HMM), and Support Vector Machine (SVM). Integrate the feature compression techniques with the linear models once you've analysed them [2, 17]. Furthermore, compare the data with the other method used to enhance QoS if there are any discrepancies.

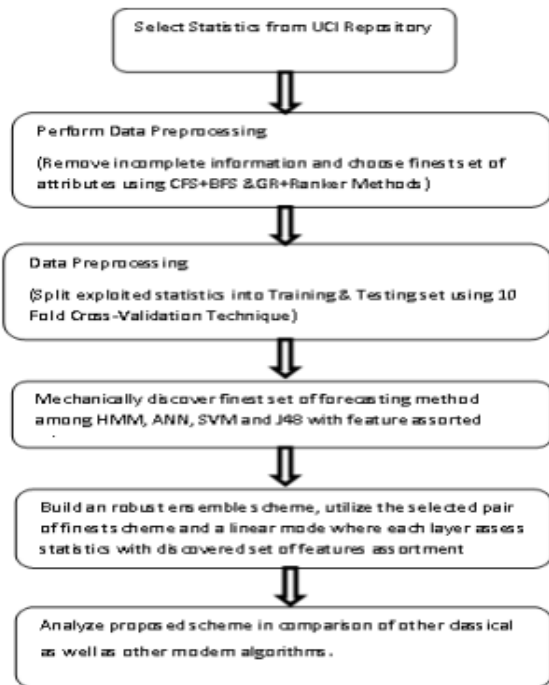


Figure 2: Process Flow Diagram of Proposed Work

V. EXPERIMENTAL RESULTS

Several researches were launched and addressed in this chapter to prove that the suggested method was better and more appropriate than the other classical and contemporary algorithms.

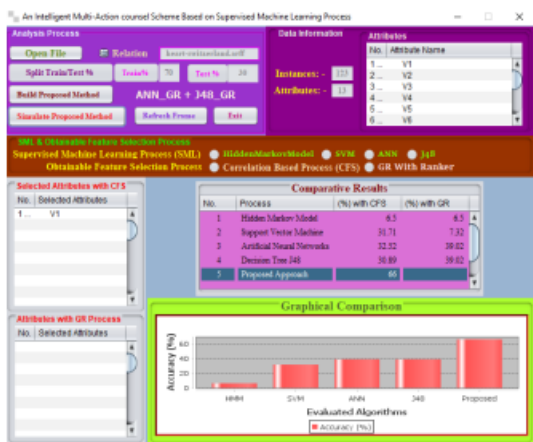


Figure 3: Designed model of Accuracy prediction

The parameters of the updated statistics, as well as the suggested algorithm's efficiency, are shown by the values indicated in the figure below. The results demonstrate that the suggested strategy outperforms the alternatives in terms of accuracy.

No.	Process	(%) with CFS	(%) with GR
1	Hidden Markov Model	55.56	55.56
2	Support Vector Machine	81.11	77.78
3	Artificial Neural Networks	78.52	79.63
4	Decision Tree J48	80	79.26
5	Proposed Approach	90	

Figure 4: Comparative results with previous methods

Table 1: Selected Method vs. N. Senthil kumar Mohan's Algorithm on the Heart Diseases Population [14]

Algorithms	Attained Accuracy
HRLFM [41]	88.4%
Proposed Approach	90%

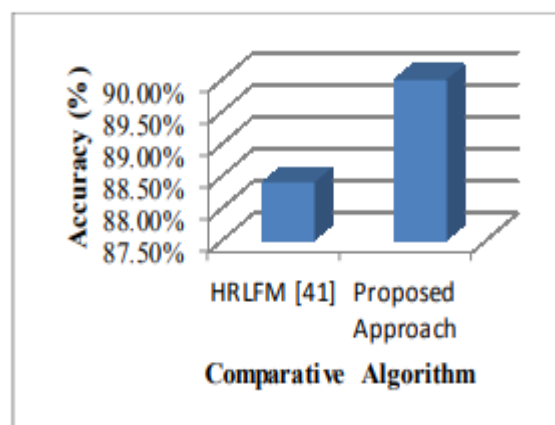


Figure 5: Proposed & Senthil kumar Mohan Algorithm over Heart Diseases Dataset [14]

Values in the table and graph that have been given above illustrate the importance of the recommended strategy to this enquiry. There is a comparison of the recommended method's efficacy and appropriateness to other approaches.

VI. CONCLUSION

Improving efficiency, appropriateness, and quality of service was the goal of the work done in this enquiry. The purpose of the literature review was to develop a more effective approach by discussing the features and shortcomings of current methods. A variety of methods, including ANN, a Hidden Markov Model (HMM), a Support Vector Machine (SVM), and a Decision Tree variant (J48), are explored in the proposed study. Following a thorough evaluation of all four approaches, the suggested technique selects the best two algorithms, one of which uses a linear model based on the feature selection process with best-first search and gain ratio, and the other of which employs the Ranker method. To prove that the suggested method works, several simulations have been run. The suggested method successfully addresses the shortcomings of both classic and cutting-edge algorithms, according to all comparisons.

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.
9. Sumalatha.V , Dr.Santhi.R "A Study on Hidden Markov Model (HMM)" International Journal of Advance Research in Computer Science and Management Studies, Volume 2, Issue 11, November 2014, pp.- 465-469.
10. Zhang Youzhi "Research and Application of Hidden Markov Model in Data Mining" Second IITA International Conference on Geoscience and Remote Sensing, IEEE, 2010, pp.-459-462.
11. PadmavathiJanardhanan, Heena L., and FathimaSabika "Effectiveness Of Support Vector Machines In Medical Data Mining" Journal Of

- Communications Software And Systems, Vol. 11, No. 1, March 2015, pp.- 25-30.
12. GaganjotKaur, AmitChhabra "Improved J48 Classification Algorithm for the Prediction of Diabetes" International Journal of Computer Applications, Volume 98 – No.22, July 2014, pp.- 13-17.
 13. Senthilkumar Mohan, ChandrasegarThirumalai, &GautamSrivastava "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques" Special Section On Smart Caching, Communications, Computing And Cybersecurity For Information-Centric Internet of Things, IEEE, 2019, pp.- 81542-81554.
 14. Shekhawat V.S., Tiwari M., Patel M. (2021) A Secured Steganography Algorithm for Hiding an Image and Data in an Image Using LSB Technique. In: Singh V., Asari V.K., Kumar S., Patel R.B. (eds) Computational Methods and Data Engineering. Advances in Intelligent Systems and Computing, vol 1257. Springer, Singapore. .
 15. H. Gupta and M. Patel, "Study of Extractive Text Summarizer Using The Elmo Embedding," 2020 Fourth International Conference on ISMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), Palladam, India, 2020, pp. 829-834, doi: 10.1109/ISMAC49090.2020.9243610.
 16. Neelam Badi, Mayank Patel and Amit Sinhal (2019). The Role of Fuzzy Logic in Improving Accuracy of Phishing Detection System. International Journal of Innovative Technology and Exploring Engineering, Volume-8 Issue-8, ISSN: 2278-3075, pp.3162-3164.
 17. Menaria H.K., Nagar P., Patel M. (2020) Tweet Sentiment Classification by Semantic and Frequency Base Features Using Hybrid Classifier. In: Luhach A., Kosa J., Poonia R., Gao XZ., Singh D. (eds) First International Conference on Sustainable Technologies for Computational Intelligence. Advances in Intelligent Systems and Computing, vol 1045. Springer, Singapore. https://doi.org/10.1007/978-981-15-0029-9_9.