Hybrid Machine Learning Classification Technique for Improve Accuracy of Heart Disease

[1]Ajay Maru
Research Scholar
Computer Science & Engineering
Geetanjali Institute of technical
studies, Udaipur,
ajaymaru100@gmail.com

[2]Ajay Kumar Sharma
Associate Professor
Computer Science & Engineering
Geetanjali Institute of technical
studies, Udaipur
profsharmaak@gmail.com

[3]Mayank Patel
Associate Professor
Computer science & Engineering
Geetanjali Institute of technical
studies, Udaipur
mayank999_udaipur@yahoo.com

Abstract— The area of medical science has attracted great attention from researchers. Several causes for human early mortality have been identified by a decent number of investigators. The related literature has confirmed that diseases are caused by different reasons and one such cause is heart-based sicknesses. Many researchers proposed idiosyncratic methods to preserve human life and help health care experts to recognize, prevent and manage heart disease. Some of the convenient methodologies facilitate the expert's decision but every successful scheme has its own restrictions. The proposed approach robustly analyze an act of Hidden Markov Model (HMM), Artificial Neural Network (ANN), Support Vector Machine (SVM), and Decision Tree J48 along with the two different feature selection methods such as Correlation Based Feature Selection (CFS) and Gain Ratio. The Gain Ratio accompanies the Ranker method over a different group of statistics. After analyzing the procedure the intended method smartly builds Naive Bayes processing that utilizes the operation of two most appropriate processes with suitable layered design. Initially, the intention is to select the most appropriate method and analyzing the act of available schemes executed with different features for examining the statistics.

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

I. INTRODUCTION

Some of the related works represent various convenient methods with the implication but none of the methods aid professionals under different characteristics [1-4]. 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 [5-8]. With a discussion of research objectives, motivation, and key findings this chapter describes the contribution towards the direction to improve the QoS of 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.

II. ISSUES WITH PREDICTION SYSTEM

1) A huge number of methodologies of available algorithms are not appropriate to aid professionals under the different area, highly depends on the employed statistics, therefore, struggle to maintain recitation with the variation of data properties.

2) The majority of the algorithms intended to accomplish only a specific task hence does not fit for the real-time scenario where the cost is one of the primary factors.

3) The exploitation of the random method is one of the key issues and the implementation of the method is based on the modified approach without consideration.

4) The formation of considerable false prediction is still a challenging issue of accessible algorithms.

5) Similar to the classical approaches most of the recently suggested algorithms have unique restraints and possess the urge of human efforts to destroy the obstacles.

III. PREVIOUS MODELS

1. Hidden Markov Model

In 1966, Baum & Petrie, the authors proposed this statistical model [9]. In simple terms, it can be considered

as the simplest Bayesian network that identifies the Markov procedure [10]. It is related to an optimal non-linear filtering dilemma called stochastic processes. Initially, this model was performed for speech recognition but with extreme mathematical formation, it can be exploited to aid in different fields like pattern, handwriting and gesture recognition, tagging of part-of-speech, a score of music, partial discharges, and bioinformatics.

2. Support Vector Machine (SVM)

Support Vector Machine (SVM) is a linear or supervised learning algorithm of machine learning that is capable of an act of classification and regression process. Additionally, this method is more efficient to resolve a variety of practical issues. Typically, it is a straightforward scheme that builds a line/hyperplane that divides data into classes. This scheme is more suitable for the various issues of real-world applications like offer significant aid in the categorization of hyper and simple text, image classifications and segmentation, recognition of handwritten characters, and for much of biological & supplementary sciences [11].

3. Artificial Neural Network (ANN)

Artificial Neural Networks (ANN) is a model of information processing, stimulated by a biological sensitive method like the process of information of a brain. [12] It is a prominent pattern of interrelated processing elements (neurons) that operate in unity to determine a precise task. This network is formed with an interrelated processing component that has direct links for carrying out a task. Every node corresponds to a processing unit and the links indicate the underlying connections   [13].
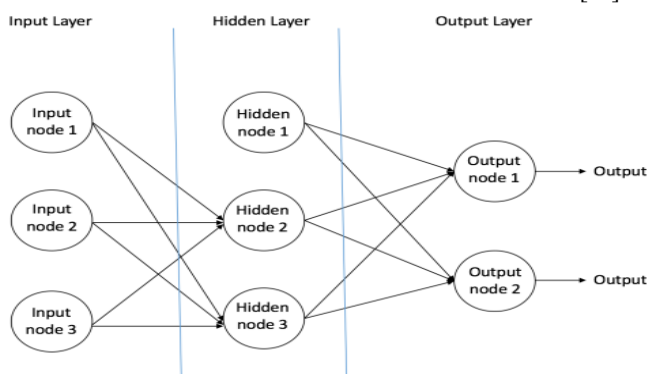

Figure 1 Neural Network

4. J48

This scheme is a form of a Decision Tree and represents the classification process of enormous data in acceptable mode. This model arranges the data into a tree form where every inner node signifies a determination and each leaf explains the classes.[14] Typically this method recursively split a branch into sub-branches, the process has been executed until the scheme has not resolved the difficulty.[15] The branch attributes become nodes of a constructed tree and their proportional assessment decide the pathway of the tree. Due to its simple working functionalities and robust outputs, this technique has lead over many techniques. Finally, a decision tree integrates two phases as Tree Construction, and Tree Pruning [16].

IV. PROPOSED METHODOLOGY

After reviewing all the existing techniques, some of the researchers signifying the various advantages of each suggested technique and elaborated several restraints that are still associated with obtainable methods and highly affect the working behavior of the techniques. Among several associated issues, some of the key restraints such as inflexibility time-consuming for building a model, alternative parameters, and inaccurate verdicts.

1. Proposed Scheme

In the proposed work, there are four different classification algorithms were selected along with the two feature compressing methods as CFS with best-first search and Gain ratio with ranker mechanism. As described in the literature survey each algorithm is designed with an obtainable process in an optimized form, such a selected process may not be utilized to build a more competent method. The proposed method investigate and analyze four chosen method such as Hidden Markov Model (HMM), Artificial Neural Network (ANN), Support Vector Machine (SVM) and Decision Tree (J48) along with two other feature compressing methods. [2, 17] After analyzing these feature compressing methods, combine them with the linear models. And if any data is mismatched then re-examine with the other employed technique to improve the QoS.
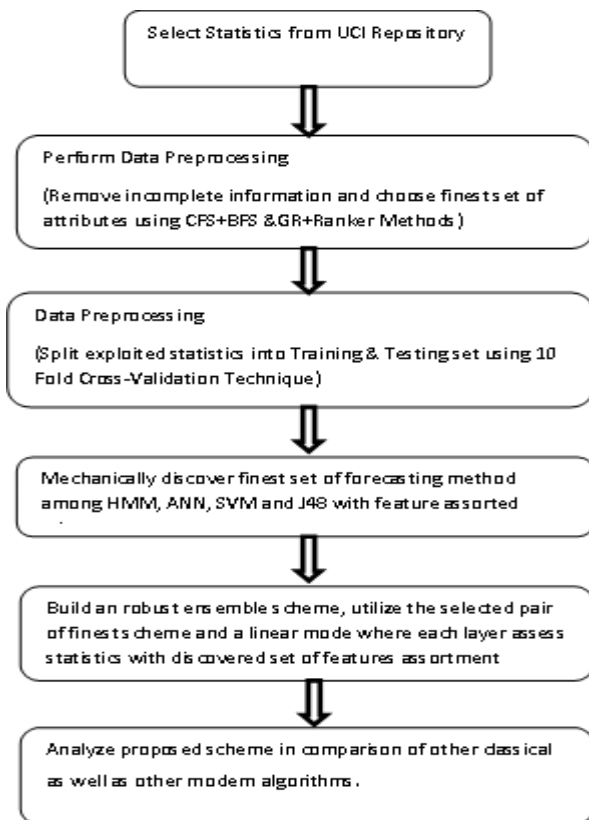
Figure 2 Process Flow Diagram of Proposed Work

Initially, several statistics have been obtained from the UCI data pool, it is an online repository of abundant datasets that can be freely acquired for the input investigation. Ultimately choose a different set of classification techniques such as HMM, ANN, SVM, and J48. These techniques were analyzed to overcome the shortcoming.

V. EXPERIMENTAL RESULTS

To demonstrate the effectiveness and suitability of the proposed approach over the other classical and modern algorithms, several investigations were initiated and discussed in this chapter.
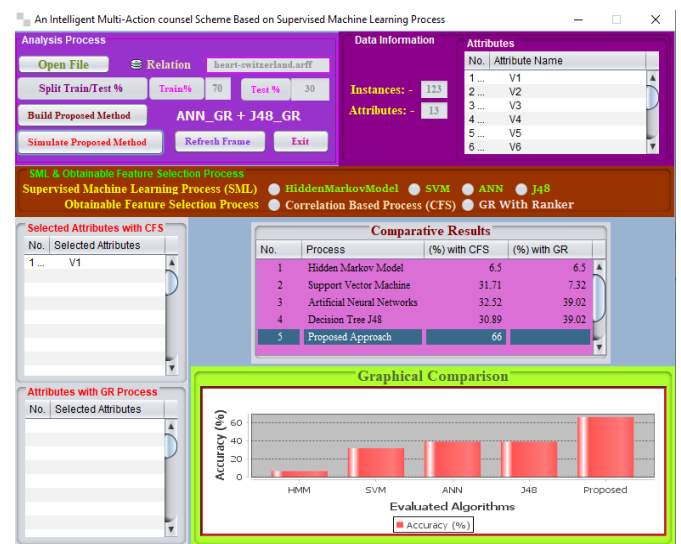

Figure 3 Designed model of Accuracy prediction

The values designated in the below-mentioned figure elaborates the efficiency of the proposed algorithm even the parameters of the modified statistics. The comparative values show that the proposed method obtains higher accuracy when compared with other existing methods.

Figure 4 Comparative results with previous methods

Table 1 Proposed Approach V/s N. Senthil kumar MohanAlgorithm over Heart Diseases Dataset [14]

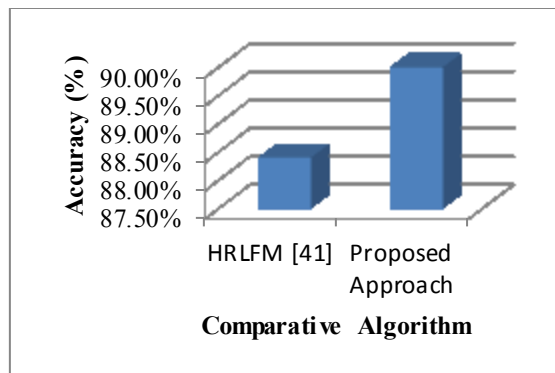| Algorithms	Attained Accuracy
HRLFM [41]	88.4%
Proposed Approach	**90%**

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

The above-mentioned values in the table and graph describe the significance of the proposed approach of this investigation. The efficiency and suitability of the proposed approach are compared with other suggested methods.

VI. CONCLUSION

The work that was carried out in this investigation endeavors 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 Hidden Markov Model (HMM), Support Vector Machine (SVM), Artificial Neural Network (ANN), and a form of Decision Tree (J48). The proposed method robustly analyzes 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.

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] AaayushiVerma, 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] [16] 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] 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, HeenaL., 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. https://doi.org/10.1007/978-981-15-7907-3_35.

[15] H. Gupta and M. Patel, "Study of Extractive Text Summarizer Using The Elmo Embedding," 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), Palladam, India, 2020, pp. 829-834, doi: 10.1109/I-SMAC49090.2020.9243610.

[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.