Anticipation of Heart Disease Using Machine Learning Algorithms Optimised By Feature Optimisation

Jishnu Bhardwaj[1, #], Preeti Nagrath[1]

[1]*Computer Science and Engineering Department,*

Bharati Vidyapeeth's College of Engineering

[#]Corresponding Author, Email: jishnubhardwaj400@gmail.com

Abstract—Machine learning has been the most trusted source in the anticipation of the heart disease. Complicated nonlinear problems with wonderful resistance and versatility can be easily handled with algorithms for optimization. Actually, there have been tests on various methods to examine and record different features hence forth resulting in improvement in categorization of heart diseases. Later, categorization on the basis of various machine learning algorithms like KNN, SVM, DT and RF is optimized by PSO and ACO method. This blended path is activated to the heart disease and the amazing outcome gives the capability and durability of amalgam methods in giving out various types of statistics for categories of heart disease. Hence to conclude, this paper evaluates various machine learning designs thus analysing the outcome on various domains e.g. exactness, recall, F1 score etc and the topmost accuracy comes out to be 87.0% using PSO and ACO.

Keywords: **Optimization, PSO, ACO, Heart disease, Accuracy**

1. INTRODUCTION

Cardio vascular Disease (CVD) [1] is going on hype now a days globally. WHO study says that 17 million people lose their lives every year from cardio vascular diseases, like strokes and heart attacks. Hence, causes and symptoms for the same are necessary to be recorded including ECG, Echo BP blood sugar, cholesterol [2] etc. The tests should be prioritized to let the patent start with the medication at the earliest. Machine learning is an upcoming area and it makes it possible to attain information and knowledge from a huge data which is just next to impossible for humans. The paper aims diagnose and investigate on various features contributing to CVD. Many machine learning algorithms [12] like SVM (Support vector machines), KNN(K-Nearest neighbor), DT(Decision tree) and RF(Random Forest) are compared by applying PSO(Particle swarm optimization) and ACO(Ant colony optimization) like smart algorithms on them [8, 9]. Initially manual classification is done. This categorization is between healthy or unhealthy. Based on this classification, 67% of data is cross checked and 33% is tested on the basis of this article.

The research paper is organized into six sections: Section 1 introduces the topic with its importance. Section 2 explains problem definition the methodology of the paper. Section 3 explores work related to our model. Section 4 explains the methodology and the architecture the specific models presented in this paper. The section also explains the dataset used to train and test the four models. Section 5 showcases the results achieved by each model, and Section 6 concludes the research paper.

2. PROBLEM OF STUDY

Some earlier researches applied machine learning techniques for the diagnosis and detection of heart diseases. Researchers analysed these machine learning techniques and worked less on optimization of these techniques. The optimization study of PSO and ACO are exploited with any machine learning technique. Here the FCBF method is

used as an initial step. When all the rigorous attributes are discrete, the attribute selection used for mining is selected. A pre requisite to machine learning is the feature selection that effectively reduces dimensions, eradicating unimportant data, making learning accuracy higher, thus improving the result learning and understanding. In the next step, PSO and ACO are implemented for optimizing the features. This selection of features helps in getting the better results. Hence, the main step uses classification for diagnosing heart problems. The research work emphasizes on the classification by enhancing accuracy to analyse the performance of characteristic selection method.

3. Related Work

Various demonstrations take place on medical data sets using several classifiers and portrays selection criteria. There is a very minute research done on categorization of heart disease dataset [6] which is used here. A lot of them have shown good classification in precision. There was a hybrid method where the two machine learning algorithms SVM and GA [3] are combined with the wrapper approach. Different tools like WEKA and data mining are used to realize the outcome of this methodology on different datasets [7]. A fully automated Clinical Decision Support System (CDSS) [15] is presented for the treatment of heart disease, automatically which scans the patient's data from his medical history. The proposed CDSS for the risk detection of heart patients contains two parts: (1) A fully optimized automated methodology for the generation of weighted rules and creating a rule-bound decision system. In the first part, mining technique is used, attribute selection and weightage technique to obtain the fuzzy rules. Then, the fuzzy system is made according to the fuzzy rules and the chosen attributes. Another proposed system has been developed with the aim to classify people with heart disease and healthy people. The performances of different ML models for heart disease diagnosis on full and selected features were tested. Feature selection algorithms such as Relief, mRMR, and LASSO were used to select important features, and on these selected features, the performance of the classifiers was tested. The Cleveland heart disease dataset was implemented in several studies and is used.

4. Methodology

This segment has PSO and ACO based feature selection and system to categorize. Following is the structure in the proposed system. Binary classification is done on the training dataset. The features are picked up and PSO and ACO are applied in combination of the subset of features for optimizing the data [14]. The new features are categorized using python SK Learn [4]. The features of the system are hereby given in detail in the following sections. The proposed architecture is shown in Figure 1. The data is obtained from UCI Machine Learning Repository [5].

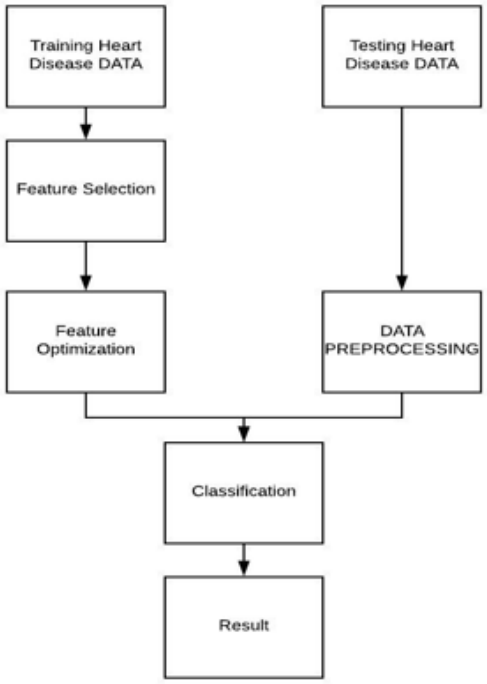

Fig 1. Proposed Architecture

4.1. *Aspect Selection*

Number of features can rise up to 14 attributes in heart disease. Because a huge number of irrelevant and duplicate attributes are involved, the categorization of heart disease becomes more intrinsic. Accuracy hits with costing going high when whole data is used for heart disease classification. The feature

selection leads to reduction of size, elimination of unresolved data, and increase of accuracy with the improvement in outcomes. A serious threat to selection of efficiency and effectiveness is increase in dimensionality of data. For choosing the effective features, the FCBF's method is adopted to categorize. The method helps in reduction of attributes and helps in providing outstanding results considering the feature co-relation and duplication. The reduction in the rate of heart disease is from 14-7.

4.2 *Feature Optimization*

4.2.1. Particle Swarm Optimization (PSO) [15]
PSO has been used to solve complex problems by answering the complexities by interaction between simple agents and their works. Russel Eberhart, and James Kennedy in 1995[16] gave optimization method for real world problems. It talks about individual collaborations, each particle moving at each iteration, resulting one in closest position to the other to move/ change their path. Due to this nature, a tremendous research is going on PSO by extending it with optimization framework

4.2.2. Ant Colony Optimization (ACO)
It is a simple procedure to find out the optimal feature subset using very less iterations. The aim of ACO [17] method is to reduce duplication amongst them by choosing a subset of features. This process relates each Ant to earlier selected features thus selecting the lowest similar feature. Hence if majority of ants selects a feature, it implies that the feature has uniqueness as compared to others. The feature gets the maximum portion of Pheromone and other ants' selections will be more in further iterations. The features with high similarity gets high pheromone values. So ACO method decide on the best features having minimum duplication.

5. RESULTS

The objective of the experiment was to test the best algorithm for heart disease with the above-mentioned optimization methods. The experiment and the paper were done under Jupyter environment. Further, due to less selected features, tenfold cross validation was applied here. We examined the accuracy and effectiveness of all classifiers with respect to the time taken to build the model, correct examples, incorrectly classified examples and accuracy without optimization and with optimization (after applying PSO and ACO methods)

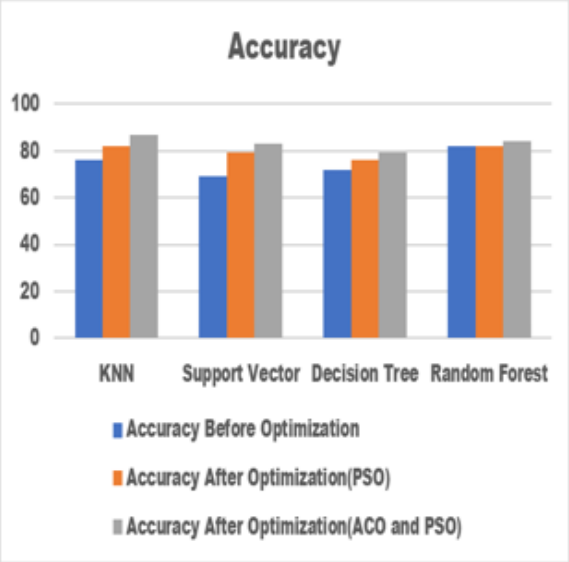

Fig 2. Accuracy score of KNN, SVM, DT(Decision Tree) and RT(Random Forest)

5.1 Accuracy Results

The accuracy result can be analysed by the following table 1:

Table 1: Accuracy score in (%) on various algorithms

Algorithm	Accuracy Before Optimization (in %)	Accuracy After Optimization (PSO)	Accuracy After Optimization (ACO and PSO)
KNN	76	82	87
SVM	69	79	83
DT	72	76	79
RF	82	82	84

6. CONCLUSION

This work aimed at comparison of algorithms with different performance measures using machine learning. In this process the data work preprocessed and used for test prediction. Each algorithm showed their workability factors. The best was KNN, RF, SVM and DT. This experiment also showed that the predictive accuracy increases with optimization hybrid approach. Hence the analysis section clearly indicates the impact of hybrid PSO and ACO approach to disease diagnosis as compared to any other approach. This model by PSO and ACO gets an accuracy score of 87% with KNN and 84% with RF. This kind of study requires best resources and expert advice in their respective fields.

REFERENCES

[1] Rumsfeld J. S., Joynt K. E., & Maddox T. M., "Big data analytics to improve cardiovascular care: promise and challenges", Nature Reviews Cardiology, 13(6), pp.350, 2016.

[2] Fatima M. & Pasha M., "Survey of machine learning algorithms for disease diagnostic. Journal of Intelligent Learning Systems and Applications", Journal of Intelligent Learning Systems and Applications, 9(1), pp.1, 2017.

[3] Tan K. C., Teoh E. J., Yu Q., & Goh K. C., "A hybrid evolutionary algorithm for attribute selection in data mining", Expert Systems with Applications, 36(4), 8616-8630, 2009.

[4] Sklearn repository, [online] https://scikitlearn.org/stable/index.html

[5] Dataset U. H., "UCI Machine Learning Repository",[online].https//archive.ics.uci.edu/ml/machine-learning databases/heartdisease/ Heart

[6] Otoom A. F., Abdallah E. E., Kilani. Y, Kefaye A., & Ashour M., "Effective diagnosis and monitoring of heart disease", International Journal of Software Engineering and Its Applications, 9(1),143-156, 2015.

[7] Parthiban G. & Srivatsa S. K., "Applying machine learning methods in diagnosing heart disease for diabetic patients", International Journal of Applied Information Systems, 3(7), 2249-0868, 2012.

[8] Chaurasia V. & Pal S., "Data mining approach to detect heart diseases", International Journal of Advanced Computer Science and Information Technology, 2(4),56-66, 2014.

[9] Vembandasamy K., Sasipriya R., & Deepa E., "Heart Diseases Detection Using Naive Bayes Algorithm", IJISET-International Journal of Innovative Science, Engineering & Technology, Vol.2, 441-444, 2015.

[10] Liu X., Wang X., Su Q., Zhang M., Zhu Y., Wang Q., & Wang, Q.,"A hybrid classification system for heart disease diagnosis based on the rfrs method", Computational and Mathematical Methods in Medicine, Vol.2017, Article ID 8272091, 2017.

[11] Malav A., Kadam K., & Kamat P., "Prediction of heart disease using k-means and artificial neural network as a hybrid approach to improve accuracy", International Journal of Engineering and Technology, Vol.9, No.4, 2017.

[12] Cauwenberge L. Van, "Top 10 Machine Learning Algorithms", Data Sci. Cent., 2015.

[13] Thirumuruganathan S., "A Detailed Introduction to K-Nearest Neighbor (KNN) Algorithm", [Online].

[14] Khourdifi Y. & Bahaj M. "Heart Disease Prediction and Classification Using Machine Learning Algorithms Optimized by Particle Swarm Optimization and Ant Colony Optimization" International Journal of Intelligent Engineering and Systems, 12(1), 2019

[15] Anooj P.K., "Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules", Journal of King Saud University - Computer and Information Sciences,24(1),27-40,2012 https://doi.org/10.1016/j.jksuci.2011.09.002

[16] Song M. P. & Gu G. C, "Research on particle swarm optimization: a review," Proceedings of 2004 International Conference on Machine Learning and Cybernetics (IEEE Cat. No.04EX826), Shanghai, China, vol.4, 2236-2241,2004. doi: 10.1109/ICMLC.2004.1382171

[17] Pei Y., Wang W. & Zhang S., "Basic Ant Colony Optimization," 2012 International Conference on Computer Science and Electronics Engineering, Hangzhou, 665-667, 2012.doi: 10.1109/ICCSEE.2012.178

[18] Li J. P., Memon M. H., Nazir S., & Sun R. "A hybrid Intelligent System Framework for the prediction of heart disease using Machine Learning Algorithms.", Wearable Technology and Mobile Applications for Healthcare, 2018.