A Comparative Study of Hybrid Evolutionary Based Algorithms with Machine Learning Classifiers for the Prediction of Medical Database

S Poongothai¹*, C Dharuman² and P Venkatesan³

¹Asst. Professor, Department of Mathematics, SRM IST, Ramapuram Campus, Chennai, India.
²Professor, Department of Mathematics, SRM IST, Ramapuram Campus, Chennai, India.
³Faculty of Research, Sri Ramachandra University, Chennai, India.

*Corresponding author: poongothaikannan25@gmail.

Abstract. In the current world, day-by-day, huge amount of data being generated at a rate far higher than by which it can be examined by human comprehension alone, in this case, data mining became an important task for extracting useful information from large amount of data. Especially in these cases, the standard techniques of datamining plays in a better way such as Evolutionary approaches performs more efficient as well as more accurate. The objective of this paper is to compare Fuzzy based Evolutionary Algorithm with other machine learning classifiers to improve the classification of medical dataset of heart disease.

1. Introduction
In today medical world, Artificial Intelligence played a vital role in diagnosing the disease [1-3]. One of the major challenging diseases to the human society is heart problems. In this paper dataset of Heart disease is taken from UCI machine learning repository for classification. Genetic algorithm is used to find the relevant variables and combined with fuzzy to find the optimal solution [4] and also combined with ensemble methods like Bagging, Adaboost and Stacking. The performance of all the methods are compared and discussed.

1.1. Database
The dataset of heart disease is taken from UCI machine learning repository for the proposed model [4]. The dataset contains 294 instances with 14 variables including class variable. The variables are age, sex, chestpain, trestbps, chol, fbs, restecg, thalach, exang, oldpeak, slope, ca, thal and num.

2. Preliminaries
2.1. Fuzzy Logic
Despite of many forefathers in the mathematical field, Lofti A. Zadeh, a professor at University of California, fixed his place in the field of uncertainty or vagueness concepts. He proposed a theory of fuzzy sets and logic namely ‘Fuzzy Logic’ in the year 1965. He defined the fuzzy set as a set in which each element have membership values that lies between 0 and 1.
The first and foremost step of fuzzy logic is defining the linguistic variables. Next step is designing the membership function based on the problem. The other main components of Fuzzy logic controller are fuzzification system, inference system and defuzzification system [5,6]. The process of converting a crisp set into a fuzzy set i.e., changing crisp input values to linguistic variables is called Fuzzification. Support Fuzzification Method (SFM) and Grade Fuzzification Method (GFM) are the two methods attempted for fuzzification. In SFM method, using the following relation, the set can be fuzzified.

\[ \tilde{A} = \mu_1S(y_1) + \mu_2S(y_2) + \ldots + \mu_nS(y_n) \]  

Out of many fuzzy classifier systems, most famous classifier systems are Fuzzy Inference System, Fuzzy C means, and Fuzzy Grid Partition System. The fuzzy inference system (FIS) follows the technique as “If antecedent then consequent”. Fuzzy C means (FCM) is a data clustering technique in which a database is divided into n clusters such that every point in the database belongs to every cluster to some degree. In the simple fuzzy partition grid methods, each variable can be divided by various linguistic values [7]. The extraction of features from the set of attributes obtained by Genetic Algorithms acts as an input to the fuzzy system and simple fuzzy grid system yields the excellent output. Also Fuzzy Unordered Rule Induction Algorithm (FURIA) is used for classification. FURIA method is expertise in differentiate and separating every class from other ones, so no alternative rule is applied and the classes order are not relevant [8]. This algorithm is developed from RIPPER Algorithm which applies pruning technology while creating the replacement and the revision rule.

2.2. Ensembled systems
Ensembled systems is a combination of various machine learning algorithms to get good optimal solution i.e., ensemble learning is a machine learning technique in which many models are considered for training to solve one problem and combined to obtain a good output. The main idea behind the ensemble methods are when the weak models are correctly identified and then combined gives more better results. Many experimental studies proves that by combining the results of various classifiers gives accurate results and minimize the errors [9-12]. Most commonly used ensemble algorithms are Bagging, Adaboost and Stacking.

2.2.1. Bagging. The process of making many subsets of data from a sample (used for training) randomly chosen with replacement is termed as Bagging. It is also called Bootstrap aggregating which was developed by Leo Breiman in 1994 [13].

2.2.2. AdaBoost. It is the shortest form of Adaptive Boosting and was proposed by Freund and Schapire in the year 1996. The process of boosting the weaker classifiers to strong classifiers is termed as Adaboost [14].

2.2.3. Stacking. The concept of stacking was developed by Leo Breiman in the year 1996 [15]. The process of finding new algorithm by combining the prediction of several other learners is termed as stacking.
2.3. Evolutionary Algorithms

I. Rechenberg introduced the concept of evolutionary computation in his paper “Evolutionary strategies” in the year 1960. Based on Darwinian model, EAs are developed and it is a class of stochastic optimization algorithm. The basic concept of EAs are survival of the fittest. Selection, Crossover and Mutation are the major operators of EAs. Genetic Algorithms one of the branch of Evolutionary Algorithms was introduced and developed by Holland in 1975 [16] inspired by Darwin’s theory of evolution which states that the survival of an organism is affected by rule "the strongest species that survives". In Selection process, The individuals of the population should be selected at random in initial stage. Roulette Wheel Process is one of the common selection operator in which each member has ability to survive for the next generation. New offsprings are obtained by crossover method by crossing a couple of individuals from the current individuals. Mutation is a unary operator i.e., it will change individuals randomly in the current population. Then the original population are replaced by the newly created offsprings (variables). This replacement always gives the best set of variables deleting the worst ones. Every generation is better than its previous one so that the possibility of getting the optimal solution is better. The entire process should be repeated till the desired output or stopping criteria met.

3. Methods and Results

In this paper, by using GA, the variables in the dataset of heart disease are reduced to 6 namely Sex, Chestpain, FBS, Exang, Oldpeak and Slope. These variables are considered as relevant information to classify the database. Then the reduced variables are applied to fuzzy systems and ensembled methods such as Bagging, Adaboost and Stacking for classification. Suppose that fuzzy rules $r_1^j, \ldots, r_n^j$ have been trained to learn for class $\lambda_j$. For a new case $y$, the support of this class is described by

$$s_j(y) = \sum_{i=1}^{n} \mu_{r_i^j}(y) \, cf\left(r_i^j\right) \tag{2}$$

where $cf\left(r_i^j\right)$ is the certainty factor of the rule $r_i^j$. It is defined as follows

$$cf\left(r_i^j\right) = \frac{|D_T^{(j)}|}{|D_T|} \frac{\sum_{y \in D_T^{(j)}} \mu_{r_i^j}(y)}{\sum_{y \in D_T} \mu_{r_i^j}(y)} \left(1 + \frac{1}{r_i^j} \right)^{-2} \tag{3}$$

where $D_T^{(j)}$ denotes the subset of training instances with label $\lambda_j$. According to these equations the rules accepts to model more flexible boundaries which improves the classification rate.

In bagging, each collection of data is applied to get their respective decision trees. By this method, the resultant obtained is ensemble of various models. Finding the average of all the output from various decision trees gives the desired result of the problem taken. Consider the dataset $D$ of size $r$, then create $n$ new subsets of data $D_i$ of size $r_i$. Few observations may get repeated in each $D_i$. If $r_i^j = r$, then the expected value of the set is given by $1 - \frac{1}{e}$ being distinct subsets of $D$ and the remaining sets are duplicates [14]. These types of datasets are bootstrap sample. Afterwards, that $n$ dataset models are fitted by the $n$ samples of bootstrap. Then the results are combined by finding the mean of all the output.
In Adaboost, consider the dataset \( \{(x_1, y_1), \ldots, (x_r, y_r)\} \) where each \( x_j \) has an associated class \( y_j \in \{-1, 1\} \). Consider the set of weak classifiers \( \{w_1, w_2, \ldots, w_n\} \) which gives the output \( w_j(x_j) \in \{-1, 1\} \) for each object. After the \((m-1)\)th iteration, boosted classifier will be a linear combination of weak classifiers which is of the form

\[
C_{(m-1)}(x_i) = y_1 w_1(x_i) + \cdots + y_{m-1} w_{m-1}(x_i)
\]

At \( m \)th iteration, by extending this to a better classifier by adding another weak classifier \( w_m \) with weight \( y_m \),

\[
C_m(x_i) = C_{(m-1)}(x_i) + y_m w_m(x_i)
\]

The process continues to find which weak classifier is the good choice for \( w_m \) with weight \( y_m \). Total error \( TE \) can be defined as summation of its exponential loss on each object.

\[
TE = \sum_{i=1}^{N} e^{-y_i C_m(x_i)}
= \sum_{i=1}^{N} e^{-y_i C_{m-1}(x_i)} e^{-y_i y_m w_m(x_i)}
\]

By considering \( K_i(m) = \begin{cases} 1, & \text{if } m = 1 \\ e^{-y_i C_{m-1}(x_i)}, & \text{if } m > 1 \end{cases} \). So we have

\[
TE = \sum_{i=1}^{N} K_i(m) e^{-y_i y_m w_m(x_i)}
\]

Splitting the summation between correctly classified and misclassified object considering by \( y_i w_m(x_i) = 1 \) and \(-1\) respectively.

\[
TE = \sum_{y_i w_m(x_i) = 1} K_i(m) e^{-\alpha_m} + \sum_{y_i w_m(x_i) = -1} K_i(m) e^{\alpha_m}
\]

To find the desired weight \( \alpha_m \), which minimizes the total error by differentiating the total error w.r.t. \( \alpha_m \), we get

\[
\frac{d(TE)}{d\alpha_m} = \frac{d\left(\sum_{y_i w_m(x_i)} K_i(m) e^{-\alpha_m} + \sum_{y_i w_m(x_i)} K_i(m) e^{\alpha_m}\right)}{d\alpha_m}
\]

Equating the above equation to be zero and solving, we get

\[
\alpha_m = \frac{1}{2} \log \left(\frac{\sum_{y_i w_m(x_i)} K_i(m)}{\sum_{y_i w_m(x_i)} K_i(m)}\right)
\]

In stacking, the step by step procedure is given by

Step 1: The training set is separated into two disjoint sets.
Step 2: Train the set (base level) present in the first set.
Step 3: Using the training, test the remaining set (base level) present in the second set.
Step 4: Using the results obtained from step 3 as inputs and the perfect responses as the outputs, training should be updated on higher level learner.

Table 1 shows the classification rate of heart disease database with and without reduction of variables. Also the results are compared. Figure 1 shows the performance of classification rate of dataset before and after reduction of variables. It is obviously seen that hybrid techniques
performs better than the normal techniques. Also GA with Fuzzy is good in classification compared to other techniques.

Table 1. Comparison of hybrid classification of Fuzzy and Ensembled methods.

| Classification Techniques | Attributes used | Rate of accuracy (%) |
|---------------------------|----------------|----------------------|
| Adaboost                  | 14             | 77.89                |
| Bagging                   | 14             | 78.57                |
| Stacking                  | 14             | 63.94                |
| Fuzzy                     | 14             | 78.93                |
| GA-Adaboost               | 6              | 81.97                |
| GA-Bagging                | 6              | 80.03                |
| GA-Stacking               | 6              | 64.93                |
| GA-Fuzzy                  | 6              | 82.61                |

Figure 1. Performance of classification rate of dataset with and without GA.
4. Conclusion
By the proposed method the hybrid based models performs better in classifying the data even if it is noisy. GA used to reduce the variables then the classification is done by fuzzy and ensemble methods with and without GA. By reducing the variables it is cost effective and time effective to solve the particular case. The results are compared and discussed briefly which shows hybrid techniques are good in classifying the database. Also among all hybrid techniques GA with fuzzy is good. In future, some more social relevant problems are going to be discussed by applying fuzzy relation based algorithms with advanced version of evolutionary computations having hope that it would perform better in classification.

References
[1] Srimani P K and Koti M S 2014 Knowledge discovery in medical data by using rough set rule induction algorithms Indian Journal of Science and Technology vol 1 pp 905–15
[2] Overiu M and Simon D 2010 Biogeography-based optimization of neuro-fuzzy system parameters for diagnosis of Cardiac disease Proceedings of the 12th annual conference on Genetic and evolutionary computation pp 1235–42
[3] Rajeswari K, Vaithiyananthan V and Amirtharaj P 2011 Prediction of risk score for heart disease in India using Machine intelligence International Conference on Information and Network Technology pp19–22
[4] UCI : Machine Learning Repository, http://archive.ics.uci.edu/ml/.
[5] Bezdek J C, Keller J, Krasnapuram R and Pal N R 2005 Fuzzy Models and Algorithms for Pattern Recognition and Image Processing (The Handbooks of Fuzzy Sets). Secaucus, NJ, USA: Springer-Verlag New York, Inc.
[6] Verikas A, Guzaitis J, Gelzinis A and Bacauskiene M 2010 A general framework for designing a fuzzy rule-based classifier Knowledge and Information Systems pp 1-19
[7] Poongothai S, Dharuman C and Venkatesan P 2016 Fuzzy Evolutionary Computing in Biological Data Mining Global Journal of Pure and Applied Mathematics vol 12 pp 306-309
[8] del Jesus M, Hoffmann F, Navascues L and Sánchez L 2004 Induction of fuzzy rule-based classifiers with evolutionary boosting algorithms IEEE Transactions on Fuzzy Systems, vol 12 p 96
[9] Bauer E and Kohavi R 1999 An Empirical Comparison of Voting Classification Algorithms: Bagging, Boosting, and Variants Machine Learning vol 35 pp 1-38
[10] Domingos P 1996 Using Partitioning to Speed Up Specific-to-General Rule Induction. In Proceedings of the AAAI-96 Workshop on Integrating Multiple Learned Models pp 29-34
[11] Opitz D and Maclin R 1999. Popular Ensemble Methods: An Empirical Study Journal of Artificial Research vol 11 pp 169-198
[12] Quinlan J R 1996 Bagging, Boosting, and C4.5 In Proceedings of the Thirteenth National Conference on Artificial Intelligence pp 725-730
[13] Breiman, Leo 1994 Bagging Predictors (PDF) Department of Statistics University of California Berkeley. Technical Report No. 421. Retrieved 2019-07-28
[14] Aslam, Javed A, Popa, Raluca A and Rivest R L 2007 On Estimating the Size and Confidence of a Statistical Audit
[15] Breiman, Leo 1996 Stacked Regressions Machine Learning vol 24 pp 49–64
[16] Holland J 1975 Adaptation in Natural and Artificial Systems The University of Michigan Press Ann Arbor.