A Novel Genetic Algorithm Approach for Simultaneous Feature and Classifier Selection in Multi Classifier System

Tien Thanh Nguyen, Alan Wee-Chung Liew
School of Information and Communication Technology
Griffith University
Gold Coast, QLD, Australia
tienthanh.nguyen2@griffithuni.edu.au,
a.liew@griffith.edu.au

Minh Toan Tran
School of Applied Mathematics and Informatics
Hanoi University of Science and Technology
Hanoi, Vietnam

Xuan Cuong Pham
School of Information and Communication Technology
Hanoi University of Science and Technology
Hanoi, Vietnam
cuong.phamxuan.hust@gmail.com

Mai Phuong Nguyen
College of Business, Massey University
Albany, New Zealand

Abstract—In this paper we introduce a novel approach for classifier and feature selection in a multi-classifier system using Genetic Algorithm (GA). Specifically, we propose a 2-part structure for each chromosome in which the first part is encoding for classifier and the second part is encoding for feature. Our structure is simple in the implementation of the crossover as well as the mutation stage of GA. We also study 8 different fitness functions for our model. Experiments are conducted on both 14 UCI Machine Learning Repository and CLEF2009 medical image database to demonstrate the benefit of our model on reducing classification error rate.

Keywords—Genetic Algorithm, multi-classifier system, classifier fusion, classifier selection, feature selection, combining rules, combining classifiers algorithm.

I. INTRODUCTION

In recent years, ensemble learning is one of the most active research areas in supervised learning [1, 2]. Ensemble methods can be divided into two categories [2]:

- Mixture of experts: Using a fixed set of classifiers and making decision from outputs of these classifiers.
- Coverage: Generating generic classifiers, which are classifiers from the same family but have different parameters. The classifiers are combined to reach a final decision.

In this paper, we focus on the first type of ensemble methods where decision is formed by combining outputs of different base classifiers. There are several combining strategies and among them, Stacking-based approaches are one of the most popular ensemble methods. Stacking was first proposed by Wolpert [3] and was further developed by Ting and Witten [7]. In this model, the training set is divided into several equal disjoint parts. One part plays as the test set in turn and the rest play as the training set during training. The outputs of Stacking are posterior probabilities that observations belong to a class according to each base classifier. Posterior probabilities of all observations form meta-data or Level1 data. The original data is denoted as Level0 data.

We denote class set by \( \{W_j\} \), \( N \) as the number of observations, \( K \) as the number of base classifiers and \( M \) as the number of classes. For an observation \( X_i \), \( P_k(W_j | X_i) \) is the probability that \( X_i \) belongs to class \( W_j \) given by \( k^{th} \) classifier. Level1 data of all observations, a \( N \times MK \) posterior probability matrix \( \{P_k(W_j | X_i)\} \) where \( j = 1 \ldots M, k = 1 \ldots K, i = 1 \ldots N \) is in the form:

\[
\begin{align*}
&\begin{bmatrix}
P_1(W_1 | X_1) & \ldots & P_1(W_M | X_1) \\
\vdots & \ddots & \vdots \\
P_K(W_1 | X_1) & \ldots & P_K(W_M | X_1)
\end{bmatrix} \\
&\ldots \\
&\begin{bmatrix}
P_1(W_1 | X_N) & \ldots & P_1(W_M | X_N) \\
\vdots & \ddots & \vdots \\
P_K(W_1 | X_N) & \ldots & P_K(W_M | X_N)
\end{bmatrix}
\end{align*}
\]

Level1 data of an observation \( X \) is defined as:

\[
\text{Level1}(X) := \begin{bmatrix}
P_1(W_1 | X) & \ldots & P_1(W_M | X) \\
\vdots & \ddots & \vdots \\
P_K(W_1 | X) & \ldots & P_K(W_M | X)
\end{bmatrix}
\]

GA model for multi-classifier system to improve classification performance has been shown to be an effective strategy [6, 10, 11]. Our aim here is to solve both the classifier selection and feature selection problem. Feature selection is an important problem in pattern recognition, data analysis and data mining [12]. Feature selection tries to reduce the number of features while maintaining accuracy at an acceptable value. Generally speaking, methods that transform features to a new domain with a reduction in the dimension of feature can be treated as feature selection. Therefore, strategy to solve this problem is very diverse, for instance, linear transformations, search techniques and GA. Classifier selection is an interesting topic in classifier fusion. The presence of some classifiers may actually degrade the accuracy of the system because of their wrong predictions. So removing them would reduce predicted error rate. Here, we propose a GA approach to not only search for the best feature subset but also explore the best classifier subset in a multi classifier system. Another aspect of our work is that we also empirically evaluate 8 different fitness functions for combining classifiers. As there are many state-of-the-art combining algorithms for classifier fusion, it is important to assess their performance. We conduct extensive experiments.
on the UCI data files and CLEF 2009 medical image database to demonstrate the performance of our approach.

II. RECENT WORK

Based on Stacking model, a number of combining algorithms had been introduced with the purpose of reducing the error rate of classification task. Ting and Witten [7] proposed Multiple Response Linear Regression algorithm (MLR) to combine posterior probabilities of each observation based on the sum of weights calculated from K Linear Regression functions. Kuncheva et al. [4] applied Fuzzy Relation to find the relationship between posterior probability matrix of (2) and Decision Template for each class. These methods are trainable combining algorithms since Level1 data is used in training to form the prediction framework.

Besides, Kittler et al. [8] presented six fixed combining rules named Sum, Product, Vote, Min, Max and Average. These rules are simple in calculation and in several applications they give lower classifying error rate compared with that of base classifiers. The advantage of applying fixed rules for ensemble system is that no training based on Level1 data is needed.

On the other hand, GA approaches to improve accuracy of classifier fusion have also been proposed recently. Kuncheva and Jain [6] introduced two GA based algorithms in which features are selected by join and disjoin mechanism. In the former, features were encoded by \( \{0,1,\ldots,K\} \) where \( k \) means that feature is only used by \( k \)-th classifier and 0 means that feature is not used by any classifiers. In the latter, classifier encoding was added in the same chromosome with feature encoding and both of them work independently in crossover and mutation stage. An encoding method was developed based on Venn diagram for feature encoding and integer values for classifier encoding. The first algorithm did not perform well according to their experiment while the second algorithm is quite hard to implement since Venn diagram becomes more complicate with many classifiers. Nanni et al. [10] employed GA to improve the SCANN algorithm [9] by building representations where each includes encoding of M classes. Gabrys and Ruta [11] tried to put classifier, feature and fixed representations where each includes encoding of M classes.metis with a new GA model for multi classifier system which achieves both effectiveness and easy implementation.

III. PROPOSED GA APPROACH FOR MULTI CLASSIFIER SYSTEM

A. Chromosome Encoding and Crossover

To implement GA for classifier fusion, first, we propose the structure of a chromosome as illustrated in Fig. 1. Each chromosome has two parts. The first part includes \( K \) genes based on the number of classifiers. We use two elements \( \{0,1\} \) to encode for each gene in a chromosome in which:

\[
\text{Gene}(k) = \begin{cases} 
1 & \text{if } k\text{-th classifier is selected} \\
0 & \text{otherwise}
\end{cases} 
\]  

(3)

The second part is \( K \) feature encodings associated with the \( K \) classifiers. Each of them includes \( D \) genes according to the number of features. We again use \( \{0,1\} \) to encode for a gene of feature encoding. For \( k = 1,K \) and \( d = 1,D \) we have:

\[
\text{Gene}_d(k) = \begin{cases} 
1 & \text{if } d\text{-th feature is selected by } k\text{-th classifier} \\
0 & \text{otherwise}
\end{cases} 
\]  

(4)

Next, we discuss the 2-stage crossover process on two parents \( A \) and \( B \). In the first stage, crossover is conducted on classifier encoding part. Here we employ single point splitter. Each classifier encoding exchanges its head with the other while retains its tail, and their feature encodings are swapped accordingly based on the classifier encoding on the first part (Fig. 2). This ensures the consistency between first part and second part of each chromosome. In the second stage, single point splitter is again applied since crossover is continued to be applied on feature encoding of \( i\)-th classifier of both \( A \) and \( B \) (Fig. 3) (i = 1, K). After crossover is performed on all pair of feature encodings, we have 2 new offspring chromosomes.

As GA search does not always resulted in better accuracy than the model on the supersets, we add a special chromosome to initialize the population in which all genes of this element are equal to 1. This encoding will be evaluated based on fitness value and replaced if its accuracy is inferior or retained if its error rate is competitive. In doing so, we ensure that our GA approach is always better than or equal to the original superset model.

Algorithm 1: Crossover

Input: Chromosome A and Chromosome B
Output: Two new offspring chromosomes

Step 1: Crossover on classifier encoding part of \( A \) and \( B \) by single point splitter as in Fig. 2

Step 2: Crossover each pair of feature encoding on second part of \( A \) and \( B \) respectively by single point splitter (Fig. 3)

B. Fitness Function

We introduce 8 fitness functions for our GA approach. The accuracy of the combining classifiers algorithm is use as fitness in evaluations. Our purpose is to explore which algorithm is suitable for our model. First, six fixed combining rules [5, 7], namely Sum, Product, Vote, Min and Median, on Level1 data of unlabeled observation \( X_{\text{Test}} \) (2) are selected to predict class label. The details of the six fixed rules are given by:

• Sum rule:

\[
X_{\text{Test}} \in W_i \text{ if } t \arg \max_{n=1,M} \left\{ K \sum_{k=1}^{K} P_i(W_n | X_{\text{Test}}) \right\} 
\]  

(5)

• Product rule:

\[
X_{\text{Test}} \in W_i \text{ if } t \arg \max_{n=1,M} \left\{ K \prod_{k=1}^{K} P_i(W_n | X_{\text{Test}}) \right\} 
\]  

(6)

• Vote rule:

\[
X_{\text{Test}} \in W_i \text{ if } t \arg \max_{n=1,M} \left\{ K \Delta_{in} \right\} 
\]  

(7)
We also employ MLR algorithm [7] as fitness function. The idea of MLR is that each classifier put a different weight on each class and then combining algorithm is conducted based on posterior probability and its associated weight. Ting et al. [7] proposed solving M Linear Regression model corresponding with M class based on Level1 data (1) of training set to find these combining weights (Fig. 4). Here we denote the weight matrix by \( \Psi = \{ \omega_j \} \) in which \( \omega_j \) is the weight of \( i^{th} \) classifier on \( j^{th} \) class.

The regression model for \( m^{th} \) class is given by:

\[
LR_m(X) = \sum_{k=1}^{K} \alpha_{km} P_k(W_m | X)
\]  

(11)

and the predicted label of an observation \( X_{Test} \) is given by:

\[
X_{Test} \in W_i \text{ if } t = \text{arg max}_{n=1,M} \left\{ LR_n(X_{Test}) \right\}
\]  

(12)

To find \( \Psi = \{ \omega_j \} \), we solve M Linear Regression model independently by minimizing M objective functions:

\[
\sum_{j=1}^{N} (Y_m(X_j) - LR_m(X_j))^2 \rightarrow \min
\]  

(12)

where \( X_j \) is an observation in training set and \( Y_m \) (\( m = 1,M \)) is a crisp label vector of \( X_j \) given by

\[
Y_m(X_j) = \begin{cases} 
1 & \text{if } X_j \in \text{class } W_m \\
0 & \text{otherwise}
\end{cases}
\]  

(13)

Finally, we also evaluate the Decision Template combining algorithm. Decision Template of \( i^{th} \) class as a \( K \times M \) matrix defined in [4] is given by:

\[
DT_i(k, m) = \frac{\sum_{j=1}^{N} I[Y_j = W_i] \times P_k(W_m | X_j)}{\sum_{j=1}^{N} I[Y_j = W_i]}
\]  

(14)

where \( Y_j \) is the crisp label of observation \( X_j \) and

\[
I[Y_j = W_i] = \begin{cases} 
1 & \text{if } Y_j = W_i \\
0 & \text{otherwise}
\end{cases}
\]  

(15)

The Decision Profile of an observation \( X \) (denoted by \( DP(X) \)) is also defined as similar to Level1(X) (2). Actually, the Decision Template of \( i^{th} \) class is the average of Decision Profile of observations in training set where their labels are \( W_i \). Eleven measurements between \( DP(X) \) and \( DT_i \) was proposed in [4]. Here we employ a similarity measurement \( S \) defined by:

\[
S(DP(X), DT_i) = \frac{\|DP(X) \cap DT_i\|}{\|DP(X) \cup DT_i\|}
\]  

(16)

where \( \|e\| \) is the relative cardinality of the fuzzy set \( a \).

The class label for an observation \( X_{Test} \) is predicted by:

\[
X_{Test} \in W_i \text{ if } t = \text{arg max}_{j=1,M} \left\{ S(DP(X), DT_j(X)) \right\}
\]  

(17)

The pseudo code of our GA algorithm for classifier fusion is given by:

**Algorithm 2:** Compute fitness value based on a specific chromosome encoding

**Input:** Training set Level0 data, chromosome \( c \) and combining classifier algorithm.

**Output:** fitness value for \( c \)

Step 1: Use Stacking to generate Level1 data from Level0 data by selected classifiers and their associated features based on encoding of \( c \)

Step 2: Run combining classifiers algorithm on Level1 data, return accuracy as fitness value.

**Algorithm 3:** GA approach for multi-classifier system

**Training process:**

**Input:** Training set Level0, \( K \) base classifiers, PMul; mutation probability, TMax; maximum number of generations, L: population size, combining classifiers algorithm

**Output:** selected subsets of classifiers and features and classification model corresponding with optimal solution.

Step 1: Initialize population with L chromosomes;

Step 2: For each chromosome, call Algorithm 2 to obtain fitness value

Step 3: Do

\* Withdraw with replacement to generate L/2 pair of chromosome.
\* Call Algorithm 1 to conduct crossover for each pair
\* Perform mutation based on PMul
\* Add new L offspring chromosomes to population
\* Compute fitness function of all L offspring chromosomes (Algorithm 2), select L chromosomes with highest fitness

of GA on Sum Rule, Decision Template and MLR also show from all GA approaches (called Select Best). Our objective those of GA on the others (Table IV). Our objective is to discover which fitness function is best among the data. To assess statistical significance, we used paired t-test to compare two classification results (level of significance set to 0.05).

IV. EXPERIMENTAL RESULTS

We conducted experiments on data from the UCI Machine Learning Repository data files and CLEF 2009 medical image dataset. For evaluation, 10-fold cross validation was performed on each dataset and the test was run 10 times. So in total we had 100 test outcomes.

To compute fitness value, as mentioned earlier we used 6 fixed combining rules (denoted by GA Sum, GA Product, GA Max, GA Min, GA Median and GA Majority Vote) and 2 well-known combining classifiers methods (denoted by GA Decision Template and GA MLR). Error rate of these approaches were compared. To initialize the parameters of GA, we set the mutation probability PMut=0.015, population size L=20, maximum number of generations TMax=50. We selected Linear Discriminant Analysis, Naïve Bayes and K Nearest Neighbor (with K set to 5, denoted as 5-NN) as our base classifiers. As these classifiers are different to each other in their approach, the diversity of the ensemble system is ensured. To assess statistical significance, we used paired t-test to compare two classification results (level of significance set to 0.05).

A. UCI Data

We chose 14 UCI files as input data for combining classifier system (Table I). Experimental results of all files are reported in Table II and III. As can be seen, GA approach outperforms its counterparts with the same combining algorithm. For example, GAs on 5 fixed rules, namely Product, Max, Min, Median and Vote, all post more than 10 wins and 0 loss compared with the classical combining strategy. The results of GA on Sum Rule, Decision Template and MLR also show good performance although they are less remarkable than those of GA on the others (Table IV).

Next, we compare each GA approach with the best result from all GA approaches (called Select Best). Our objective is to discover which fitness function is best among the data. Table V illustrates result of statistical test in which GA Product achieves the best result, obtaining only 3 losses. GA Median is follows with 4 losses, and next is GA Sum, GA MLR and GA Max with 6 losses. GA Decision Template is the worst approach with up to 9 losses in comparison with Select Best. In reality, it is difficult to choose a best GA approach for all data files. One may perform well on some files but poor on others. For example, GA Decision Template is the worse in our experiment but it obtains good result on Balance (8.32%). Based on our experiments, we suggest using fixed rules as fitness function for our GA approach since fixed rules are very simple in implementation. We only need to compute posterior probabilities of unlabeled observation to form the prediction; as a result, computation cost will be reduced. However, when GA with fixed rules do not achieve acceptable results, other complex combining classifier algorithms like Decision Template and MRL can be employed.

Another advantage of our GA approach is that it helps to select optimal subsets of classifiers and features which produce better result than their supersets. This is due to the fact that the proposed GA approach solves both classifier selection and feature selection problems by using a 2-part structure for chromosomes, the first for classifier encoding and the second for feature encoding.

B. CLEF 2009

The other experiment was evaluated on CLEF2009, a medical image database collected by Archen University, Germany. It is a large database containing 15363 images allocated in 193 hierarchical categories. Here, we chose 10 classes with different number of observations in each (Table VI). Histogram of Local Binary Pattern (HLBP) [14] was selected as feature vector of each image.

Once again, the benefit of our approach is evident from the results in Table VII and VIII. The reduction of error rate is significant on GA with fixed rules, for example, GA Min helps to reduce error rate by 9.63%. Error rate of GA MLR decreases by 4% compared with MLR while the decrease for GA Decision Template is 1.5%.

V. CONCLUSION AND FUTURE WORK

In this paper, we have introduced a method to solve both the feature and classifier selection problems in a multi-classifier system based on GA. Our aim is to find subset of classifier and feature set which have more discrimination ability and to search the optimal subset from base classifiers to achieve lower error rate than the original sets. We have conducted extensive experiments and compare our GA approach among 8 different fitness functions. Experimental results on UCI files and CLEF2009 dataset has demonstrated that our model considerably reduces error rates of classification by exploring optimal subsets of the original feature and classifier set. We suggest using fixed rules as fitness function for our model because it helps to save computation cost while achieving acceptable accuracy.

In the future, we plan to explore new fitness functions for our GA model as well as discover new effective encoding to boost GA performance.

REFERENCE


### TABLE I. INFORMATION OF UCI FILES IN EXPERIMENTS

<table>
<thead>
<tr>
<th>Number</th>
<th>File name</th>
<th># of attributes</th>
<th>Attribute Type</th>
<th># of observation</th>
<th># of class</th>
<th># of attribute on Level 1 (3 classifiers)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Bupa</td>
<td>6</td>
<td>C,LR</td>
<td>345</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>Artificial</td>
<td>10</td>
<td>R</td>
<td>700</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>3</td>
<td>Pima</td>
<td>6</td>
<td>R, I</td>
<td>768</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>Sonar</td>
<td>60</td>
<td>R</td>
<td>208</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td>Heart</td>
<td>13</td>
<td>C, I, R</td>
<td>270</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>6</td>
<td>Haberman</td>
<td>3</td>
<td>I</td>
<td>306</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td>Balance</td>
<td>4</td>
<td>C</td>
<td>625</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>8</td>
<td>Fertility</td>
<td>9</td>
<td>R</td>
<td>100</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>9</td>
<td>Wdbc</td>
<td>30</td>
<td>R</td>
<td>569</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>10</td>
<td>Australian</td>
<td>14</td>
<td>C, I, R</td>
<td>690</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>11</td>
<td>Tae</td>
<td>20</td>
<td>C, I</td>
<td>151</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>12</td>
<td>Contraceptive</td>
<td>9</td>
<td>C</td>
<td>1473</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>13</td>
<td>Vehicle</td>
<td>18</td>
<td>I</td>
<td>946</td>
<td>4</td>
<td>12</td>
</tr>
<tr>
<td>14</td>
<td>Iris</td>
<td>14</td>
<td>R</td>
<td>150</td>
<td>3</td>
<td>9</td>
</tr>
</tbody>
</table>

### TABLE II. CLASSIFYING RESULTS ON UCI FILES BY 6 FIXED RULES, DECISION TEMPLATE AND MLR

<table>
<thead>
<tr>
<th>File name</th>
<th># of attributes</th>
<th>Attribute Type</th>
<th># of observation</th>
<th># of class</th>
<th>Mean</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bupa</td>
<td>6</td>
<td>C,LR</td>
<td>345</td>
<td>2</td>
<td>5.7E-03</td>
<td>3.9E-03</td>
</tr>
<tr>
<td>Artificial</td>
<td>10</td>
<td>R</td>
<td>700</td>
<td>2</td>
<td>5.8E-03</td>
<td>5.9E-03</td>
</tr>
<tr>
<td>Pima</td>
<td>6</td>
<td>R, I</td>
<td>768</td>
<td>2</td>
<td>6.9E-03</td>
<td>6.8E-03</td>
</tr>
<tr>
<td>Sonar</td>
<td>60</td>
<td>R</td>
<td>208</td>
<td>2</td>
<td>1.6E-02</td>
<td>1.5E-02</td>
</tr>
<tr>
<td>Heart</td>
<td>13</td>
<td>C, I, R</td>
<td>270</td>
<td>2</td>
<td>1.7E-03</td>
<td>1.6E-03</td>
</tr>
<tr>
<td>Haberman</td>
<td>3</td>
<td>I</td>
<td>306</td>
<td>2</td>
<td>2.4E-03</td>
<td>2.3E-03</td>
</tr>
<tr>
<td>Balance</td>
<td>4</td>
<td>C</td>
<td>625</td>
<td>3</td>
<td>2.8E-03</td>
<td>2.7E-03</td>
</tr>
<tr>
<td>Fertility</td>
<td>9</td>
<td>R</td>
<td>100</td>
<td>2</td>
<td>3.2E-03</td>
<td>3.1E-03</td>
</tr>
<tr>
<td>Wdbc</td>
<td>30</td>
<td>R</td>
<td>569</td>
<td>2</td>
<td>5.2E-03</td>
<td>5.1E-03</td>
</tr>
<tr>
<td>Australian</td>
<td>14</td>
<td>C, I, R</td>
<td>690</td>
<td>2</td>
<td>5.7E-03</td>
<td>5.6E-03</td>
</tr>
<tr>
<td>Tae</td>
<td>20</td>
<td>C, I</td>
<td>151</td>
<td>2</td>
<td>5.8E-03</td>
<td>5.7E-03</td>
</tr>
<tr>
<td>Contraceptive</td>
<td>9</td>
<td>C</td>
<td>1473</td>
<td>3</td>
<td>6.1E-03</td>
<td>6.0E-03</td>
</tr>
<tr>
<td>Vehicle</td>
<td>18</td>
<td>I</td>
<td>946</td>
<td>4</td>
<td>3.1E-03</td>
<td>3.0E-03</td>
</tr>
<tr>
<td>Iris</td>
<td>14</td>
<td>R</td>
<td>150</td>
<td>3</td>
<td>1.7E-03</td>
<td>1.6E-03</td>
</tr>
</tbody>
</table>

### TABLE III. CLASSIFYING RESULTS ON UCI FILES BY GA APPROACH WITH 6 FIXED RULES, DECISION TEMPLATE AND MLR AS DIFFERENT FITNESS FUNCTIONS

<table>
<thead>
<tr>
<th>File name</th>
<th>GA Sum</th>
<th>GA Product</th>
<th>GA Max</th>
<th>GA Min</th>
<th>GA Median</th>
<th>GA Majority Vote</th>
<th>GA Decision Template</th>
<th>MLR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bupa</td>
<td>0.5029</td>
<td>3.49E-03</td>
<td>5.49E-03</td>
<td>1.31E-03</td>
<td>5.39E-03</td>
<td>5.49E-03</td>
<td>5.49E-03</td>
<td>3.9E-03</td>
</tr>
<tr>
<td>Artificial</td>
<td>0.3917</td>
<td>2.08E-03</td>
<td>4.81E-03</td>
<td>2.80E-03</td>
<td>4.81E-03</td>
<td>4.81E-03</td>
<td>4.81E-03</td>
<td>4.81E-03</td>
</tr>
<tr>
<td>Pima</td>
<td>0.2241</td>
<td>1.27E-03</td>
<td>2.12E-03</td>
<td>0.2231</td>
<td>2.12E-03</td>
<td>2.12E-03</td>
<td>2.12E-03</td>
<td>2.12E-03</td>
</tr>
<tr>
<td>Sonar</td>
<td>0.1485</td>
<td>6.71E-04</td>
<td>6.71E-04</td>
<td>0.1429</td>
<td>6.71E-04</td>
<td>6.71E-04</td>
<td>6.71E-04</td>
<td>6.71E-04</td>
</tr>
<tr>
<td>Heart</td>
<td>0.1433</td>
<td>3.20E-04</td>
<td>3.20E-04</td>
<td>0.1385</td>
<td>3.20E-04</td>
<td>3.20E-04</td>
<td>3.20E-04</td>
<td>3.20E-04</td>
</tr>
<tr>
<td>Haberman</td>
<td>0.2343</td>
<td>2.04E-03</td>
<td>2.04E-03</td>
<td>0.2343</td>
<td>2.04E-03</td>
<td>2.04E-03</td>
<td>2.04E-03</td>
<td>2.04E-03</td>
</tr>
<tr>
<td>Balance</td>
<td>0.1126</td>
<td>4.12E-04</td>
<td>5.01E-04</td>
<td>0.1107</td>
<td>5.01E-04</td>
<td>5.01E-04</td>
<td>5.01E-04</td>
<td>5.01E-04</td>
</tr>
</tbody>
</table>

Fig. 4. MLR method on Level1 data
### TABLE IV. STATISTICAL TEST COMPARING GA APPROACH WITH COMBINING CLASSIFIERS APPROACHES

<table>
<thead>
<tr>
<th></th>
<th>GA Sum vs. Sum</th>
<th>GA Product vs. Product</th>
<th>GA Max vs. Max</th>
<th>GA Min vs. Min</th>
<th>GA Median vs. Median</th>
<th>GA Majority Vote vs. Majority Vote</th>
<th>GA Decision Template vs. Decision Template</th>
<th>GA MLR vs. MLR</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Better</strong></td>
<td>7</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>11</td>
<td>10</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td><strong>Competitive</strong></td>
<td>7</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td><strong>Worse</strong></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

### TABLE V. STATISTICAL TEST COMPARING EACH GA APPROACH WITH BEST RESULT SELECTED FROM ALL GA APPROACHES

<table>
<thead>
<tr>
<th></th>
<th>GA Sum vs. Select Best</th>
<th>GA Product vs. Select Best</th>
<th>GA Max vs. Select Best</th>
<th>GA Min vs. Select Best</th>
<th>GA Median vs. Select Best</th>
<th>GA Majority Vote vs. Select Best</th>
<th>GA Decision Template vs. Select Best</th>
<th>GA MLR vs. Select Best</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Competitive</strong></td>
<td>8</td>
<td>11</td>
<td>8</td>
<td>7</td>
<td>10</td>
<td>6</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td><strong>Worse</strong></td>
<td>6</td>
<td>3</td>
<td>6</td>
<td>7</td>
<td>4</td>
<td>8</td>
<td>9</td>
<td>6</td>
</tr>
</tbody>
</table>

### TABLE VI. INFORMATION OF 10 CLASSES CHOSEN FROM CLEF2009 MEDICAL IMAGE DATABASE

<table>
<thead>
<tr>
<th>Image</th>
<th>Description</th>
<th>Number of observation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Abdomen</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>Cervical</td>
<td>81</td>
</tr>
<tr>
<td></td>
<td>Chest</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>Facial cranium</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>Left Elbow</td>
<td>69</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Image</th>
<th>Description</th>
<th>Number of observation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Left Shoulder</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>Left Breast</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>Finger</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td>Left Ankle Joint</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>Left Carpal Joint</td>
<td>80</td>
</tr>
</tbody>
</table>

### TABLE VII. CLASSIFYING RESULTS ON CLEF2009 BY 6 FIXED RULES, DECISION TEMPLATE AND MLR

<table>
<thead>
<tr>
<th>HLBP 10 classes</th>
<th>Sum</th>
<th>Product</th>
<th>Max</th>
<th>Min</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Variance</td>
<td>Mean</td>
<td>Variance</td>
</tr>
<tr>
<td></td>
<td>0.2023</td>
<td>1.85E-03</td>
<td>0.2300</td>
<td>2.05E-03</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>Vote</td>
<td>Decision Template</td>
<td>MLR</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>Variance</td>
<td>Mean</td>
<td>Variance</td>
</tr>
<tr>
<td></td>
<td>0.2144</td>
<td>2.13E-03</td>
<td>0.2266</td>
<td>1.36E-03</td>
</tr>
</tbody>
</table>

### TABLE VIII. CLASSIFYING RESULTS ON CLEF2009 BY GA APPROACH WITH 6 FIXED RULES, DECISION TEMPLATE AND MLR AS DIFFERENT FITNESS FUNCTIONS

<table>
<thead>
<tr>
<th>HLBP 10 classes</th>
<th>GA Sum</th>
<th>GA Product</th>
<th>GA Max</th>
<th>GA Min</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Variance</td>
<td>Mean</td>
<td>Variance</td>
</tr>
<tr>
<td></td>
<td>0.1543</td>
<td>1.59E-03</td>
<td>0.1438</td>
<td>1.36E-03</td>
</tr>
<tr>
<td></td>
<td>GA Median</td>
<td>GA Majority Vote</td>
<td>GA Decision Template</td>
<td>GA MLR</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>Variance</td>
<td>Mean</td>
<td>Variance</td>
</tr>
<tr>
<td></td>
<td>0.1491</td>
<td>1.60E-03</td>
<td>0.1647</td>
<td>1.72E-03</td>
</tr>
</tbody>
</table>

### TABLE IX. STATISTICAL TEST COMPARING GA APPROACH WITH NORMAL COMBINING APPROACH

<table>
<thead>
<tr>
<th>Fitness Function</th>
<th>GA Sum vs. Sum</th>
<th>GA Product vs. Product</th>
<th>GA Max vs. Max</th>
<th>GA Min vs. Min</th>
<th>GA Median vs. Median</th>
<th>GA Majority Vote vs. Majority Vote</th>
<th>GA Decision Template vs. Decision Template</th>
<th>GA MLR vs. MLR</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Better</strong></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><strong>Worse</strong></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>-----</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td><strong>Competitive</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Worse</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>