International Journal of Innovative Computing, Information and Control
Volume **8**, Number **10(B)**, October **2012**

ICIC International ©2012 ISSN 1349-4198
pp. 7363–7379

A NEW MUTATION OPERATION FOR FASTER CONVERGENCE IN GENETIC ALGORITHM FEATURE SELECTION

Rubiyah Yusof*, Uswah Khairuddin and Marzuki Khalid

Centre for Artificial Intelligence and Robotics
Universiti Teknologi Malaysia
Jalan Semarak, 54100, Kuala Lumpur, Malaysia
*Corresponding author: rubiyah@ic.utm.my; uswah@ic.utm.my; marzuki@utm.my

Received April 2011; revised October 2011

Abstract. *Feature selection is an important step in data classification because it has a high impact on classification accuracy. Feature selection using Genetic Algorithm (GA) is usually done in a wrapper method. The process is time consuming especially for large dimensional database. We propose a new mutation operation for faster feature selection by GA based on elitism of the allele. Normal elitism in GA preserves the most fit chromosomes which are evaluated using the fitness function. In the same way, the highest fit allele will be preserved and the fitness of the allele is evaluated based on the frequency of occurrences. The chromosome undergoing this mutation process will have a high if not the highest fitness because it is created based on a high fit allele. It will be the catalyst to increase the rate of convergence towards achieving an optimal features combination. Experiments for feature selection using this method are conducted using a database of tropical wood species which has a large variation of features. Results of the experiments show that a high accuracy is obtained for the recognition of the tropical wood species using the feature selection method. In addition, it has also been shown that the chromosomes created by the new mutation operation have high fitness and the rate of optimal convergence is improved substantially. The new mutation operation is not only useful for large database, but also can be used for small or medium sized database.*
Keywords: Feature selection, GA, Wood recognition, Classification

1. **Introduction.** Feature selection is a process of selecting only important features from a large database to create subset of the original datasets and it retains the original features without changing the features' domain [1]. Other objectives of feature selection are to reduce database dimensionality, remove irrelevant features, reduce learning computational cost, increase variables prediction accuracy and keep only important features that give comprehensive understanding for all variables [2]. Demand in feature selection is currently rising because of the expanding sizes of databases in various applications. Feature selection has been used for many data mining operation, for example, gene microarray data analysis, medical data analysis [3], biometrics study, industrial operating system analysis and pattern recognition [4]. A research on feature selection done by C.-C. Lai et al. [5] for spam filtering also proves that a proper discriminative features selection increases classification performance.

There are many ways to do feature selection using search algorithms such as sequential forward selection (SFS) [6], sequential backward selection (SBS) [7], exhaustive search, and genetic algorithm (GA). Unfortunately, exhaustive search is not suitable for large dimensional data because it is very time consuming. SFS and SBS operation are good because all features in a database are evaluated at the same time. However, once a feature is removed, it will have no chance of being selected again. The selected features will remain

in the selection even after new features are selected, resulting in some redundancies, in some cases. GA differs from these types of feature selection method due to the capability to evolve new features from the selected features and a vast exploration of search space for new fitter solutions. The evolving process is made possible using the GA operators of selection, cross-over and mutation. The process will be continued until the best solution is found or the maximum numbers of generations are met. The probability of finding the global optimal solution is high in GA due to its enormous computation effort and extensive exploration. In applying GA, it is important to choose suitable fitness function and consider all possible constraints. Other researches [3,8-11] have shown that GA is a powerful tool for feature selection due to its robustness and ability to work in large databases. A review paper [12] also gives good credits to GA as a feature selector. The reported works using GA for feature selectors are mostly based on two class problem [8-11], and the output of the GA are the classified patterns which are obtained from the most fit chromosomes which represent the optimized features. In these cases, the fitness function of the GA calculates the fitness of the features of the chromosomes and classifies them. GA is used as optimized feature selector as well as the classifier.

The key to good feature selection is to obtain the best features combination. One of the main problems in the application of GA for feature selection is to determine the optimal convergence. In most cases, the GA has to be run through many generations in order to ensure that the optimal convergence is achieved and could result in a time consuming process. Yang et al. [13] proposed a segmented chromosome management scheme in GA in a way of overcoming the problem. All operations in GA such as the crossover and mutation are performed in separate segments although the accuracy is the same as that obtained by the conventional GA, and the convergence rate is faster. However, the method is only useful for multi-character feature set in which the features are segmented according to various feature sets such as shape, texture and color. Work by Ingu and Takagi [14] to increase the rate of convergence in GA by fitting single-peak function on searching surface using vertex of fitted function as new elite. Although the method is simple, the convergence rate is improved. However, these methods are suitable for singled valued solution problems only and are not suitable for feature selection.

Khairuddin et al. [15] developed a feature selection method using GA for tropical wood recognition. Although the method has shown a marked improvement in terms of the accuracy as compared with methods without feature selection, the problem of slow convergence rate remained. The convergence rate depends on the size of the database. The larger the database is, the longer the time needed to obtain the optimal solution. In order to increase the convergence rate, we propose a new mutation operation for feature selection process by GA. The proposed mutation operation for faster feature selection by GA is based on the elitism of the allele. Normally, elitism in GA is applied to preserving the most fit chromosomes which are evaluated using the fitness function. In the same way, in the new mutation method, the highest fit allele will be preserved and the fitness of the allele is evaluated based on the frequency of occurrences or the frequency of the features being selected. The chromosome undergoing this mutation process will have a high if not the highest fitness because it is created based on the most fit alleles.

As the mutation rate is usually very small, the method still preserves the randomness of the GA. Moreover, the method ensures the preservations of the fit or most occurring alleles which actually are representing the features of the problem at hand. In this way, the exploration of the GA is more focused towards achieving the optimal convergence value, and thus reduces the occurrence of suboptimal convergence.

The paper is organized as follows. Section 2 discusses the feature selection method using GA. Section 3 describes the improved mutation operation. Section 4 is the application

in tropical wood recognition system. Section 5 will have details on the results of the experiments and discussions while Section 6 will be the conclusion.

2. **Feature Selection Using GA.** Consider a number of combinations of n distinct features taken m at a time which can be defined by the following equation:

$$^{n}c_{m} = \frac{^{n}P_{m}}{m!} \tag{1}$$

where P is the number of permutation (possible arrangement) of n distinct features taken m at a time and $m!$ is the product of all positive integer less than or equal to m.

The equation can be simplified as:

$$^{n}c_{m} = \frac{n!}{(n-m)!m!} \tag{2}$$

Therefore, to get all combination for n distinct features taken from $m = 1$ to $m = n$ at a time can be defined as:

$$\sum_{m=1}^{n} {}^{n}c_{m} = \sum_{m=1}^{n} \frac{n!}{(n-m)!m!} \tag{3}$$

Or in general, the sum of all combination for n distinct features is 2^n

$$^{n}c_{0} + {}^{n}c_{1} + {}^{n}c_{2} + \cdots + {}^{n}c_{n} = 2^{n} \tag{4}$$

If a database has 20 features, there are 2^{20} or 1048575 possible combinations. For a large database containing 100 features, there will be 2^{100} or 1.2677×10^{30} possible combinations. It will be time consuming to test on all the combinations. It is expected that in using GA, some of the redundant features and weak features can be eliminated. For a normal GA process, 1000 generations with population size 20 will give 20000 possible combinations. Although the number of possible combination may result in a high computation time, it is much smaller as compared with the search through the total number of possible combinations.

In [15], the feature selection process by GA is done in wrapper method where a classifier is wrapped inside the GA and the result obtained from the classifier is the fitness function for the GA. Each chromosome generated will represent the feature vector of the original database. The alleles of the chromosome will act as masks for the feature vector where the alleles numbered as '1' are the selected features and alleles numbered as '0' are the eliminated features. Each chromosome will create a subset database based on its alleles. The subset database will be used as the training database and the classification accuracy is determined by the GA process using the classifier as the fitness function. Figure 1 shows the feature selection process by wrapper GA where the classifier used is the linear discriminant analysis (LDA).

3. **Improved Mutation Operation.** In the normal GA, a process starts with generating an initial population randomly. A population consisting of a set of chromosomes is evaluated using a fitness function and ranked accordingly. A selection process is then performed and the selected chromosomes will undergo three basic GA processes which are elitism (preservation of the best fitted chromosome), crossover (cross combination of two high fitted chromosomes to produce new child) and mutation (change of bit number randomly). These processes will produce new population for the next generation. The whole processes are repeated until final generation number is reached.

In most cases, mutation is implemented to expand the exploration capability of the GA, and in finding new fitter solutions which are not present in the population of the

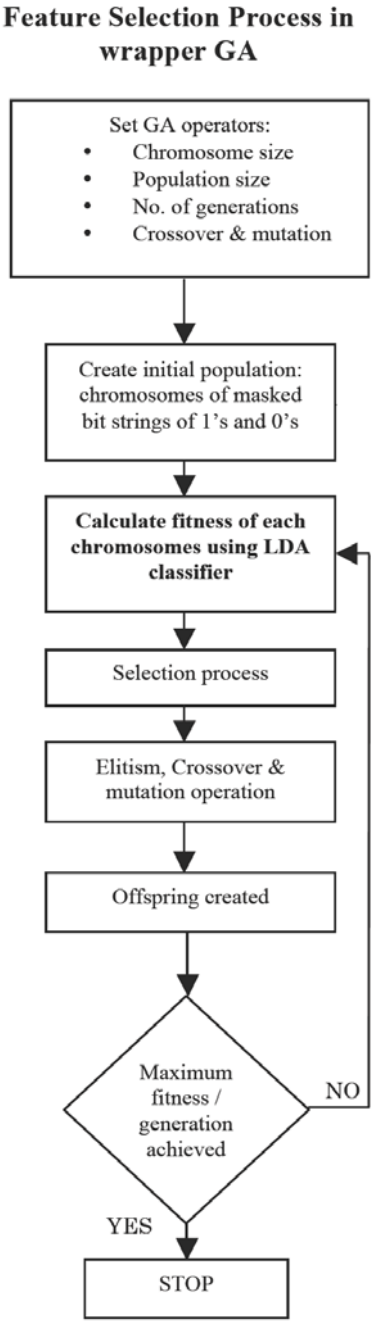


FIGURE 1. Flowchart showing feature selection process in wrapper GA

generation. On the other hand elitism is used as a means of preserving the good chromosomes for the next generation. In our proposed method, we use the principles of elitism for the mutation process in which the best alleles are preserved in the chromosome set.

In order to ensure that the randomness of the GA is preserved, the mutation process outputs a small percentage of the population. The mutation process focuses on recombination of elitist allele. The operation is very suitable for feature selection process due to the fact that feature selection focuses on the best combination of features and alleles are actually masks of the features. Therefore, the combination of elitist allele will produce the best feature combination. The final population will consist of the population of chromosomes from the normal GA process and additional chromosomes which have been

mutated by the new mutation process. The flowchart in Figure 2 below shows a feature selection process in GA with the addition of the improved mutation operation.

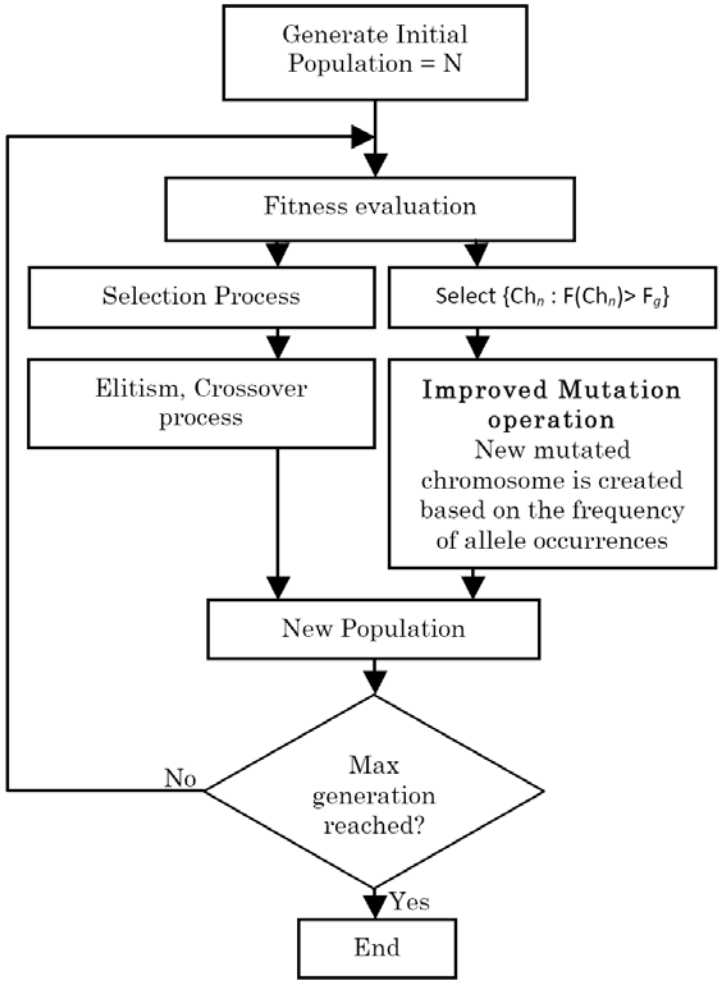


FIGURE 2. Flowchart for the new improved mutation operation in GA

3.1. **Formulation of the new mutation process.** In a population size of N chromosomes, a chromosome Ch_n where $n = 1$ to $n = N$, is an array of allele a_m where $m = 1$ to $m = M$. M is number of features in the database. A population of N chromosomes with M features is a matrix sized $N \times M$. Fitness of each chromosome Ch_n is defined as $\mathrm{F}(\mathrm{Ch}_n)$ and is based on the fitness calculated in the fitness evaluation process explained in 2.0. The average fitness in a population is F_{ave} and minimum fitness for a chromosome to be accepted as good chromosome is denoted as F_g.

When the fitness evaluation process is completed, the chromosomes will be chosen based on the selection process. These chromosomes will undergo the normal GA process of selection, crossover and mutation to generate a new set of population. Parallelly, out of these sets of N chromosomes, N' good chromosomes are selected for the new mutation process. In this case, we consider selecting the chromosome as good chromosome if it has fitness more than the desired minimum fitness (F_g) in the population. The desired minimum fitness F_g can be obtained using the following equation:

$$F_g = g \times F_{ave} \tag{5}$$

where g is a constant with a value between 0 and 1 and can be chosen judiciously by trial and error. F_{ave} is the average fitness of all the chromosomes in the population and given as follows:

$$F_{ave} = \frac{\sum_{n=N}^{n=1} F(Ch_n)}{N} \tag{6}$$

If the constant g is set to its highest value, which is $g = 1$, F_g will have the same value as F_{ave}. The idea behind choosing g to be less or equal to 1 is to ensure that the search space is large enough for the exploration of the most fit chromosome.

To create a chromosome for each generation using the new mutation operation, the summations of occurrences of "1s" alleles in the horizontal and vertical directions of the good chromosomes are considered. These alleles represent the occurrences of the features selected in the chromosomes. The new chromosomes formed will be based on the number of occurrences of the alleles in both directions. The sum of the alleles in the horizontal direction is used as an indicator of how many of the "1s" alleles should be present in the new mutated chromosome. In this case we set:

$$l = h \times L_{ave} \tag{7}$$

where h is a constant with a value between 0 and 1 and l is the number of "1s" alleles in the new mutated chromosomes, L_{ave} is the average number of horizontal "1s" alleles for all chromosomes and is given as:

$$L_{ave} = \frac{\sum_{n=1}^{N'} L_n}{N} \tag{8}$$

where L_n is the summation of the occurrence of "1s" allele in the horizontal direction or the total count of the number of features selected for each chromosome and is given as:

$$L_n = \sum_{m=1}^{M} a_m \tag{9}$$

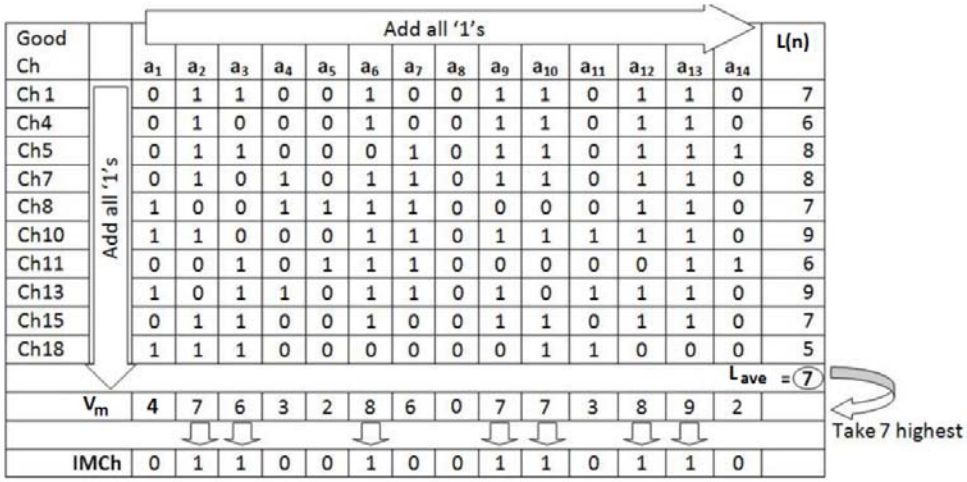
The summation of the "1s" in the vertical alleles in the set of chromosomes determines which of the alleles should be a "1" or the features to be selected. The number of occurrences of "1s" alleles in the vertical direction in the set of good chromosomes are denoted by V_m and is given by:

$$V_m = \sum_{n=1}^{N'} a_m \tag{10}$$

The mutated chromosome will have l alleles which are "1s" and is chosen among the highest count of V_m for $m = 1$ to M. In order to prevent the decrease in the exploration capability of the search space for the GA, we consider only mutating one chromosome for each generation. Therefore, the new set of population will consist of $N - 1$ chromosomes which undergo the normal GA process.

Figure 3 shows an example of how the mutated chromosome is created from a pool of 10 good chromosomes.

In the example, L_{ave} is given as 7, and l is set to be equal to L_{ave}. Therefore, the new chromosome will have 7 "1s" alleles. In order to choose the 7 alleles, the highest vertical occurrence of the alleles are considered; in this case it is found that the most occurring alleles are at $m = 13, 12, 6, 10, 9, 2, 3$ which have the count of 9, 8, 8, 7, 7, 7, 6 respectively. The alleles will be ranked based on the high V_m and the top l ranked alleles will be chosen. In this case, since only 7 alleles can be chosen, the alleles in the lowest rank have to compete with each other in a random competition.



Good Ch		a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	a_{11}	a_{12}	a_{13}	a_{14}	L(n)
Ch 1		0	1	1	0	0	1	0	0	1	1	0	1	1	0	7
Ch4		0	1	0	0	0	1	0	0	1	1	0	1	1	0	6
Ch5		0	1	1	0	0	0	1	0	1	1	0	1	1	1	8
Ch7		0	1	0	1	0	1	1	0	1	1	0	1	1	0	8
Ch8		1	0	0	1	1	1	1	0	0	0	0	1	1	0	7
Ch10		1	1	0	0	0	1	1	0	1	1	1	1	1	0	9
Ch11		0	0	1	0	1	1	1	0	0	0	0	0	1	1	6
Ch13		1	0	1	1	0	1	1	0	1	0	1	1	1	0	9
Ch15		0	1	1	0	0	1	0	0	1	1	0	1	1	0	7
Ch18		1	1	1	0	0	0	0	0	0	1	1	0	0	0	5
															L_{ave} = 7	
	V_m	4	7	6	3	2	8	6	0	7	7	3	8	9	2	
	IMCh	0	1	1	0	0	1	0	0	1	1	0	1	1	0	

FIGURE 3. Example of how the mutated chromosome (IMCh) is created from a pool of good chromosome in the improved mutation operation

3.2. **Advantages of the improved mutation operation in GA.** The mutated chromosome will have a high if not the highest fitness because it is created using the high fitted alleles from the good chromosomes. The concept is similar to the selection process where only the most frequent or most fit alleles are selected to be in the mutated chromosomes. This is done in consideration that if the features are selected more frequently, then the feature can be considered as the strong features. The mutation process preserves the existence of these strong features.

The new mutation operation is a valuable addition to the already powerful GA especially when it is used in a feature selection process because it focuses on individual allele's fitness. In a normal GA, the focus is not on the individual allele, and therefore the best feature combination or the optimal convergence process may take a long time. The proposed mutation process escalates the search for the most fit chromosomes, thereby reducing the time taken for optimal convergence.

In order to ensure that the GA process retains its randomness, the mutated chromosomes have no guarantee to be selected for the next generation of population. It will undergo the same process as the other chromosomes. However, due to the high fitness value, chances of these chromosomes to be selected as the elite chromosome are very high. This is done to ensure that the process will not result in local minimum.

In a normal mutation process, the rate is usually very small. More so, in the feature selection, mutation process hardly affects the fitness value of the chromosomes as only one or two allele(s) are changed if the mutation is applied. In this proposed method, the mutation process plays a more significant role, in that the mutated chromosome has a high fitness value and able to make significant changes to the population that are being selected.

4. **Application in Tropical Wood Recognition System.** In this subtopic, some background of the tropical wood species recognition system is explained. One of the most important part of the system is the feature selection process where GA is used. The improved mutation operation has been tested and applied in GA for this system.

4.1. **Background of automatic wood recognition system.** Conventional tropical wood species recognition system is done either by naked eye or magnifier by highly trained, certified personnel. However, this system is subjected to human error and biasness. The

fact that there are more than 3000 tropical wood species makes it harder to rely totally on this system. Therefore, an automatic tropical wood recognition system was built by Khalid et al. [16]. However, in this work, only 20 wood species are being tested. With the increase in the number of species to 52, it has been shown that the accuracy dropped tremendously [15]. One of the main reasons is that the wood species have a lot of noises as well as too many irregularities among the same species which made it difficult to discriminate the species. The use of feature selection improved the system accuracy by more than 30% [15]. The complete process of the wood recognition system is given in Figure 4.

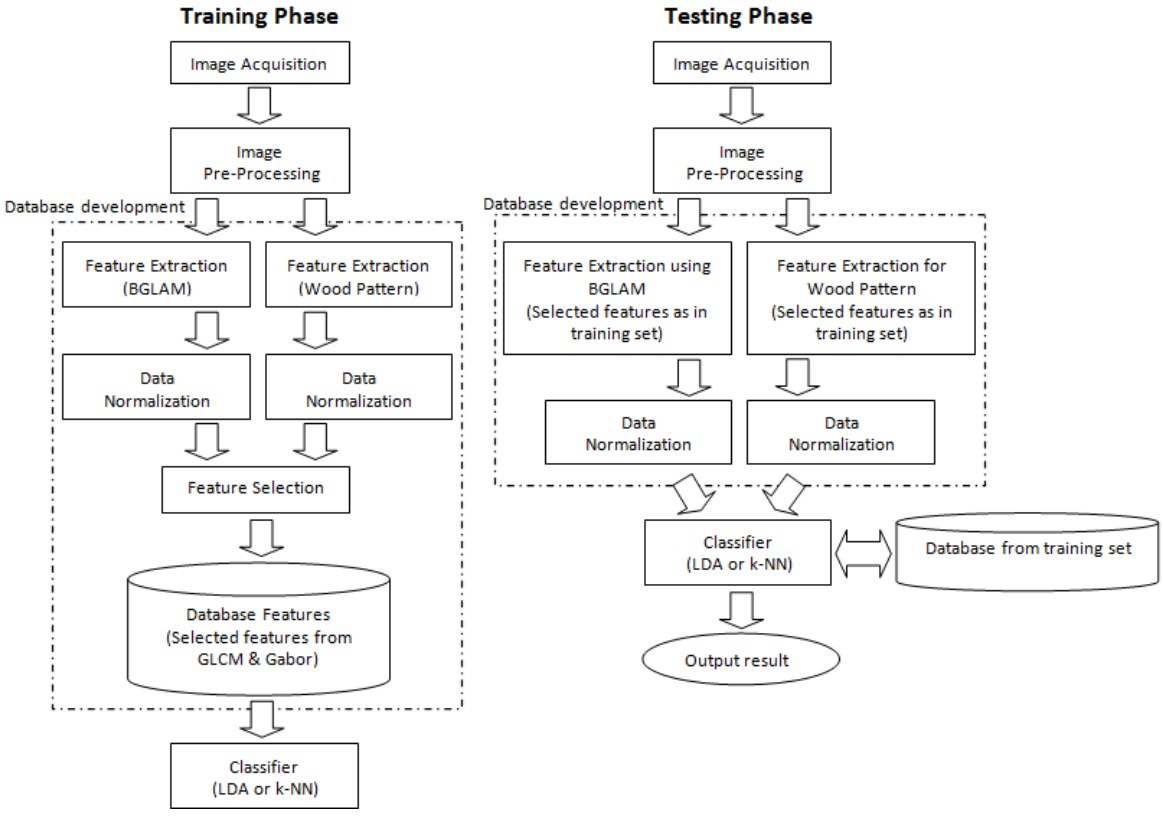


FIGURE 4. Flowchart of automatic wood species recognition system

The wood database that is being used has over 52 wood species, each with 5 different wood samples taken from different trees. A portable camera with 10 times magnification is used to capture the images of the wood surfaces. The resolution of each image is 576×768 pixels. There are 100 images taken from one species, out of which 90 images are used for training and 10 images are used for testing.

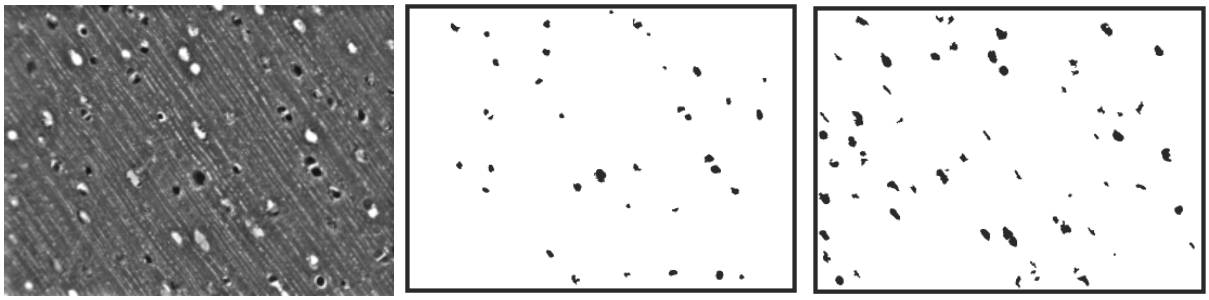
Several preprocessing techniques have been performed on the images in order to improve the image for feature extraction. One of the processes performed is homomorphic filtering to enhance image presentation. The filtering process uses a linear filter to perform non-linear mapping to a different domain and later it was mapped back to the original domain. The process removes illumination and reflectance of the images.

For feature extraction, Khairuddin [15] uses Basic Grey Level Aura Matrix (BGLAM) and Statistical Properties of Pores Distribution (SPPD) to extract the features of the wood images. X. Qin and Y.-H. Yang [17] have shown that BGLAM can give good performance as compared with many commonly used operations in this area such as Gabor texture features, Grey level Co-occurrence Matrix (GLCM), wavelet and Markov/Gibbs random field (MRF/GRF) texture models. The main advantage of BGLAM is that the features

are directly calculated from the images, and therefore the image can be reconstructed from the BGLAM features. Moreover, the method is easy to compute and no further filtering is needed.

Another sets of features used are those from SPPD. In this method, some statistical features of the images are calculated. The images are split into two main images which are

(1) Black and white images showing the black pores (Figure 5(b))
(2) Black and white images showing the "white pores" (Figure 5(c))



(a) (b) (c)

FIGURE 5. (a) Homomorphic image, (b) black and white images showing the black pores and (c) black and white images showing the "white pores"

From the images, these features are extracted:

(1) Mean size of the pores and the corresponding standard deviation (2 features)
(2) Mean distance between pores and the corresponding standard deviation (2 features)
(3) Number of small, medium and large pores (3 features)
(4) Number of pairs and solitary pores (2 features)
(5) Number of pores per square millimeters (1 feature)

In total, 10 features are extracted from one image. Additionally, the mean grey level of the original image is considered as one of the features. So, there are 21 features extracted from one image. The SPPD feature extraction method focuses more on the anatomical feature of the wood texture emulating the method used by experts to identify the wood species. Variations in the size of the pores and vessels as well as the pores arrangements are taken into considerations. In a way, the method allows the extraction of the distinct features of the pores.

4.2. **Feature selection and classification.** In this work, we employed the BGLAM with 16 grey levels and the total numbers of features extracted for each image is 136. The total features extracted from the SPPD method is 21 giving 157 total numbers of features extracted per image. Since there are 90 training images and 10 testing images per species, the training database is a matrix with a size of 4680×157. The testing database is matrix with a size of 520×157.

The database size is very large thus making the classification process very slow. Due to the irregularities of some of the wood species samples obtained such as the age of the trees, the weather and some contamination, the images are laced with various noises. Moreover, some of the features extracted maybe irrelevant and redundant which has been shown to decrease the accuracy rate. Therefore, feature selection is done in order to reduce the original database size and keep only the important and discriminative features.

In the GA process, the fitness of each chromosome will be calculated based on the fitness function. To calculate the fitness, a new reduced sized database will be formed.

This database contains only the selected number of features according to the chromosome bit. Linear Discriminant Analysis (LDA) is used in the training set in the new database. Testing samples in this new database are being classified using the discriminant function. The LDA algorithm [18] as used in the wrapper GA for feature selection is summarized as follows.

(a) Find mean μ_i for each features of a species i (μ_i is a matrix with number of column equal to number of features)

(These are the features that have been selected by the chromosomes which act as masked vectors for the feature vectors)

(b) Construct covariance matrix c_i for each species

(c) Construct pooled within group covariance matrix,

$$C(r,s) = \frac{1}{n}\sum_{i=1}^{g} n_i \cdot c_i(r,s) \tag{11}$$

i. where g is the number of species;

ii. it is calculated for each (r,s) entry in the matrix.

(d) Construct prior probabilities vector, p.

Prior probabilities for species i is equal number of samples in species i divided by number of total samples in database.

$$p_i = \frac{n_i}{N} \tag{12}$$

After discriminant analysis is done for the training set, a linear coefficient for each class is obtained. To make the classification for the testing set, the following discriminant function is used:

$$f_i = \mu_i C^{-1} x_k^T - \frac{1}{2}\mu_i C^{-1}\mu_i^T + \ln(p_i) \tag{13}$$

i. where x_k is the feature array of test sample k;

ii. sample k will be assigned to species i that has maximum f_i.

Classification done by LDA for the testing set is then compared with the true sample class. Classification accuracy is calculated based on the percentage of correctly classified samples over the total number of testing samples. The fitness value of the chromosomes in the GA is determined by the classification accuracy of the testing set.

5. **Experimental Results and Analysis.** The experiments conducted are based on the dataset used by [15]. The accuracy obtained in [15] is good at 95% with only 79 out of the 157 features are selected for the recognition. Comparatively, the accuracy for the recognition using 157 features without the feature selection is at 88%. However, one of the main problems in the feature selection method in [15] is the high cost of computational time in order to obtain the optimal convergence. Basic calculation of the computational time used is about 1.5 days for optimal convergence during the training phase.

In this paper, the same feature selection using wrapper GA as in [15] is used. In this case, the crossover rate and mutation rate is set to 1.0 and 0.01 respectively. The number of maximum generation is set to 2000 and the population is set to 20. Several experiments are conducted to show the applicability and effectiveness of the new improved mutation process.

5.1. **Setting minimum fitness F_g for good chromosomes.** The minimum fitness F_g will determine the number of chromosomes chosen to undergo the new mutation process. The lower the value of F_g is, the larger the number of chromosomes from the set of population would be chosen. In order to determine F_g , the constant g has to be determined.

Several sets of experiments are conducted for different values of g starting with $g = 0.1$ until $g = 1.0$, where $F_g = g \times F_{ave}$. The length of the mutated chromosome is set as L_{ave}. Table 1 shows the result of the different values of g and the number of generations needed for optimal convergence or solutions.

TABLE 1. Result of feature selection using new mutation GA with different F_g

g	Number of generation where optimal solution is found
0.1	1847
0.3	1956
0.5	1328
0.7	978
0.8	1249
0.9	885
1.0	467

For g less than 0.9, the number of generations needed to reach the optimal solution is very high and quite similar to the normal GA, which is the GA without the new mutation method. We may conclude that if F_g is less than the average fitness in population, the good chromosomes may not be good enough to produce a high fit mutated chromosome. Therefore, the GA with the new mutation method will operate as normal GA where the mutated chromosome did not play any role to improve the GA.

For g equal to 0.9 and 1.0, the optimal solutions are found after a smaller generation runs. This translates to less computational time needed for obtaining the optimal solutions. The following experiments are conducted based on $g = 1.0$. As seen from Table 1, the reduction in the computational time is more than 80% from the value of $g = 0.1$.

5.2. **Setting the parameter l of mutated chromosome.** The number of "1s" alelle in the mutated chromosome is determined by the parameter l. When creating a mutated chromosome from a pool of good chromosomes, it is important to set l judiciously, where $l = h \times L_{ave}$. l is determined by varying the constant h. Results of the experiments are recorded in Table 2.

TABLE 2. Results of the feature selection using new mutation GA with different l

h	Number of generation where optimal solution is found
0.1	1542
0.3	1054
0.5	1692
0.7	1272
0.9	974
1.0	467

When h is set to 0.1 until 0.7, the numbers of generations needed to obtain the optimal solutions are very high, similar to the normal GA. We may conclude that the mutated chromosomes created have lower fitness due to insufficient number of features selected and is not able to affect the normal GA operation.

When h is set higher at 0.9 and 1.0, the number of generations needed to obtain the optimal solution reduced tremendously. We maintain to keep the highest value of h at 1.0 as this is the value where $l = L_{ave}$. Increasing the value of h further may restrict the exploration capability of the GA. We may conclude that $l = L_{ave}$ gave the best performance for this dataset and this value is used for the following experiments.

5.3. **Comparison between normal GA and GA with the new mutation method.** We run feature selection process using both normal GA and GA with the new mutation operation. Both methods have the same GA parameters (Number of generations = 2000, Population size = 20, Crossover rate = 1.0, Mutation rate = 0.01). The new mutation components are set as $F_g = F_{ave}$ and $l = L_{ave}$. The objective of this experiment is to compare the optimal convergence time for both GA runs. The initial population is randomly generated and is the same for both GAs.

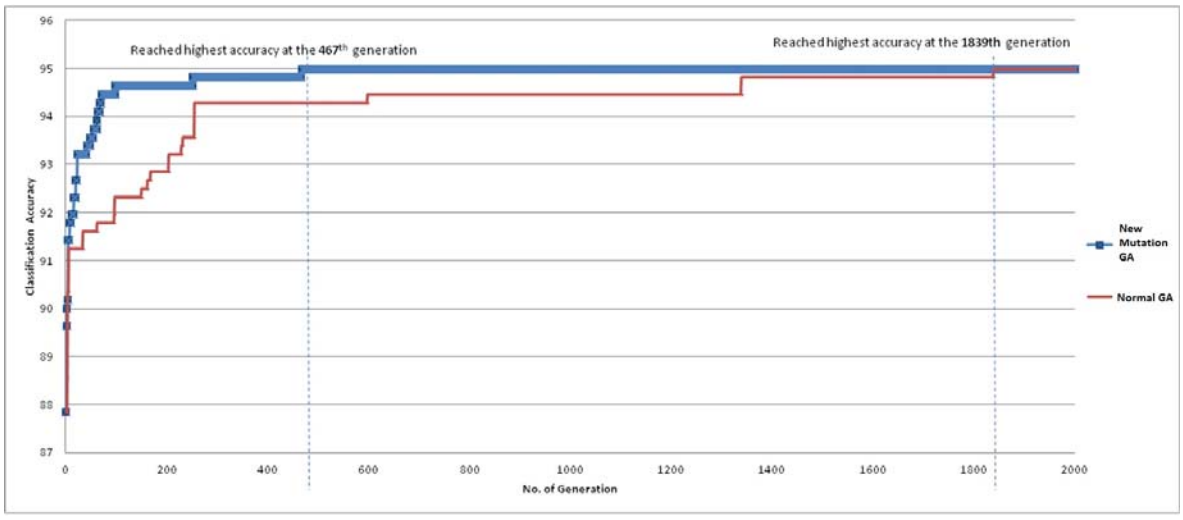


FIGURE 6. New mutation GA converges to optimal solution faster than normal GA

Figure 6 shows the result of the experiment. The thick line in the graph represents the fitness (classification accuracy) obtained per generation by using the improved mutation GA. The thin line represents the fitness obtained per generation by using the normal GA. Both feature selection processes converge on a single fitness value which is 95%. However, for the new mutation GA, the escalation towards the optimal value is very steep at the beginning of the generation runs, resulting in a faster convergence rate. The highest accuracy (95%) obtained by the new mutation GA is at the 467th generation. However, for the normal GA to reach the same value, it needs another 1372 generations. The computation cost is reduced by more than 70% when the new mutation process is performed in the GA.

Another criterion which can be used to evaluate the performance of the new mutation GA is to check on how many of the mutated chromosomes are chosen as elite chromosomes. Recalling that elite chromosomes are preserved in the next generation, any chromosomes which have been selected as the elite chromosomes are considered as very high fit chromosomes. In this experiments, test are conducted to check the elite chromosomes at every occurrence of the classification accuracy increment. In GA, an increment occurs when a chromosome in the new population has higher fitness than the previous generation. The

increment is defined by the classification accuracy rate (AR) and is given as:

$$AR(t) - AR(t-1) = \Delta AR(t) \tag{14}$$

$$\text{Increment} = \Delta AR(t) > 0 \tag{15}$$

In Figure 6, there are a total of 19 increments of the new mutation GA performance. Out of the 19 increments, the mutated chromosomes created by the new mutation operation are the elite chromosomes in 10 of them. Experimental results also show that more than 99% of the mutated chromosomes created have more than the average fitness of the generation, meaning that the mutated chromosome created are also a good chromosome.

TABLE 3. Results of the feature selection process using both normal GA and new mutation GA

No.	**Normal GA**		**New mutation GA**	
	No. of Generations when optimal convergence reached	% Accuracy	No. of Generations when optimal convergence reached	% Accuracy
1	1981	95	439	95
2	141	94.1071	255	94.2857
3	987	95	718	95
4	970	93.75	212	94.6429
5	1839	95	467	95
6	856	93.75	952	93.0357
7	203	93.2143	394	94.6429
8	1346	95	483	95
9	394	91.4286	664	94.6429
10	171	90.8929	544	94.6429
11	928	92.6786	500	94.6429
12	962	93.75	568	94.6429
13	962	95	568	95
14	928	95	501	95
15	1025	94.6429	145	93.0357
16	1547	93.2143	257	94.2857
17	1438	95	544	95
18	1358	94.6429	354	94.2857
19	1000	95	659	95
20	1844	93.75	852	94.6429
21	1000	95	394	95
22	847	95	352	95
Average no. of generations & % Accuracy	**1033.045**	**94.0828**	**491.9091**	**94.6104**
Standard Deviation	**515.4853**	**1.198892**	**199.8123**	**0.569081**

However, to prove the advantage of the proposed method by just one experiment is not enough as GA is a random stochastic search method where different runs might give

different results. Therefore, several experiments are done using the normal GA and the new mutation GA. Results of the experiments are recorded in Table 3.

The new mutation GA gave a more consistent result over the 22 runs of experiments with a standard deviation of only 199.8123 generations and 0.569081% accuracy as compared with the normal GA with standard deviation of 515.4853 generations and 1.198892% accuracy. In addition, all the optimal solutions are obtained at a much lower generation runs with the same or better accuracy.

Out of the 22 runs of the experiments, only 10 runs successfully reached the desired optimal convergence of 95%. Figure 7 shows the number of generations where the optimal value is obtained by both the new mutation GA and the normal GA for all the 10 experiments.

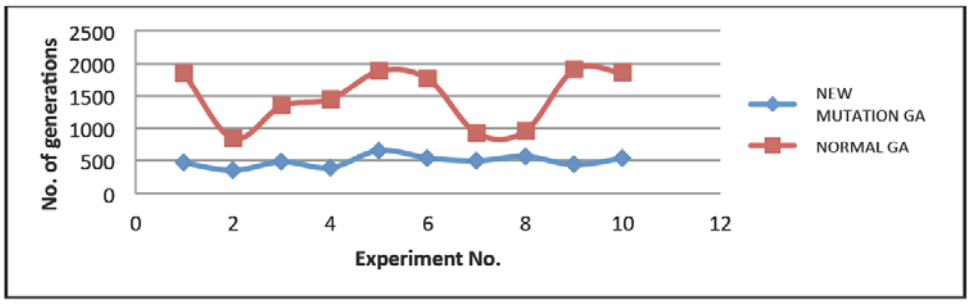


FIGURE 7. Number of generations required by each experiment to reach optimal value

The feature selection process performed by the new mutation GA reached the optimal convergence between 350th to 600th generations. Using the normal GA, feature selection process took more time to reach the optimal value, ranging between 800 to 1900 generations. The graph also indicates that the new mutation GA is able to give a more consistent result with respect to the optimal convergence as compared with the normal GA. Obviously, one of the reasons is due to the high fit mutated chromosomes, where the search is more directed towards getting the optimal convergence, resulting in a more consistent convergence rate for the experiments.

In order to show the high fitness of the mutated chromosomes, we performed several experiments to determine the frequency of the mutated chromosomes being selected as first ranked chromosome. The first rank chromosome in a population is usually the chromosome chosen as the final classification accuracy in any generation.

The analysis is conducted based on the 10 experiments that gave 95% classification accuracy. Table 4 shows the frequency of the mutated chromosomes being selected as the first ranked chromosome. The result shows that the percentage of the mutated chromosomes being ranked first is very high at almost more than 50% for all the experimental runs. This concludes that the new mutation method is able to generate a high fit chromosome which increases the fitness of the population, thereby increasing the convergence rate.

Table 5 shows the average rank of the mutated chromosome among normal chromosomes in a single population. Ranking is made based on the fitness of each chromosome. During the earlier generation in the new mutation GA, the mutated chromosomes have lower ranking; however, as number of generations increased, the ranking of the mutated chromosomes also rises and stabilizes. Average ranking of the mutated chromosome is 3 to 5 with standard deviation 0.9 to 4. It shows that the mutated chromosomes have high chance to be selected as parents by the roulette wheel selection in cross over operation

TABLE 4. Number of occurrence of 1st ranked mutated chromosome at the time of increment

Exp. No.	Number of increments	Number of mutated chromosome chosen as elite chromosome (resulting an increment)	Percentage of mutated chromosomes chosen as elite chromosome over number of increment
1	19	10	0.53
2	6	3	0.50
3	13	8	0.62
4	13	8	0.62
5	15	9	0.60
6	23	15	0.65
7	15	7	0.47
8	22	16	0.73
9	13	7	0.54
10	25	13	0.52

TABLE 5. Average rank of mutated chromosome in a population and its standard deviation

Exp. No.	Average rank of mutated chromosome in a population	Standard Deviation of rank of mutated chromosome in a population
1	3.09	0.96
2	4.39	2.63
3	4.97	4.14
4	3.61	1.96
5	3.53	1.81
6	4.26	2.36
7	4.73	0.94
8	3.20	2.56
9	3.67	2.10
10	3.54	0.98

of the GA. Figure 8 below shows the distribution of the mutated chromosome rank in 20 populations for 2000 generations in experiment 1.

6. **Conclusions.** A new mutation operation for faster feature selection by GA is proposed. The proposed method can increase the optimal convergence rate for feature selection while maintaining classification accuracy. This in turn reduces the computational time considerably. The proposed method is designed specifically for feature selection process which focuses on good combination of features since the method centered on preserving the high fit alleles of the chromosomes. The introduction of variables such as the constant h and g ensures the non conservative of the method and can be applied to various types of database. Obviously, the exploration capability of the GA is maintained and even better with the method.

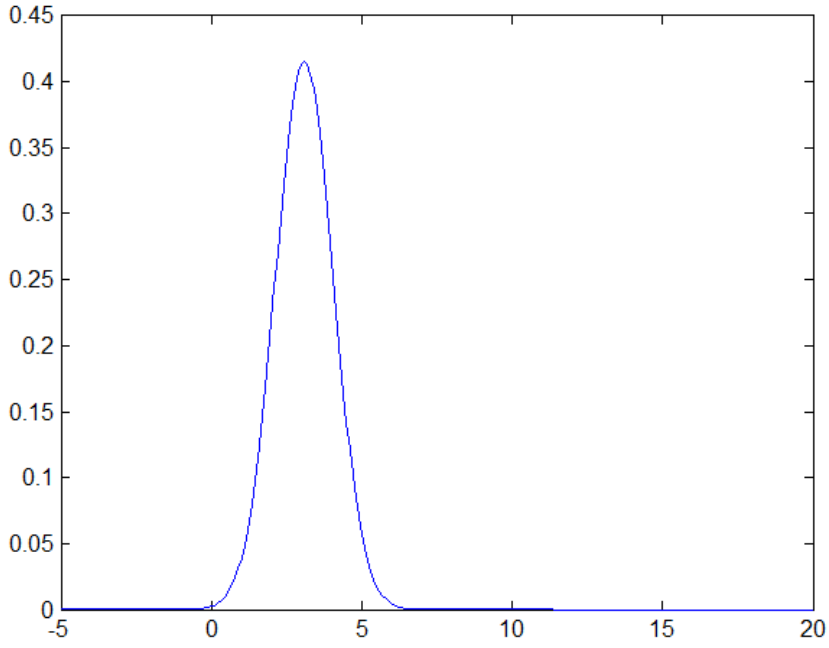


FIGURE 8. Distribution plot of the mutated chromosome's rank in experiment no. 1

Extensive analyses have been performed to show the mutated chromosomes are among the most fit chromosomes in a population, if not the highest fit. It has also been shown that the convergence rate and the classification accuracy is more consistent using the improved mutation in the GA process. This proves that the global optimal solution can be found at a faster rate as compared with the normal GA.

Acknowledgment. This research is partly funded by Malaysia Japan International Institute of Technology through the research grant vote number 4J006 and the Technofund grant from the mInistry of Science and Innovation. The authors gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation of this paper.

REFERENCES

[1] H. Liu and L. Yu, Toward integrating feature selection algorithms for classification and clustering, *IEEE Transactions on Knowledge and Data Engineering*, vol.17, no.4, pp.491-502, 2005.

[2] H. Liu, E. R. Dougherty, J. G. Dy et al., Evolving feature selection, *Intelligent Systems, IEEE*, vol.20, no.6, pp.64-76, 2005.

[3] Y. Sun, C. Babbs and E. Delp, A comparison of feature selection methods for the detection of breast cancers in mammograms: Adaptive sequential floating search vs. genetic algorithm, *Proc. of the IEEE Engineering in Medicine and Biology the 27th Annual Conference*, Shanghai, China, 2005.

[4] J. Lu, T. Zhao et al., Feature selection based-on genetic algorithm for image annotation, *Knowledge-Based Systems*, vol.21, no.8, pp.887-891, 2008.

[5] C.-C. Lai, C.-H. Wu and M.-C. Tsai, Feature selection using particle swarm optimization with application in spam filtering, *International Journal of Innovative Computing, Information and Control*, vol.5, no.2, pp.423-432, 2009.

[6] D. Q. Miao et al., A heuristic algorithm of reduction for knowledge, *Journal of Computer Research and Development*, vol.36, no.6, pp.681-684, 1999.

[7] S. K. Choubey et al., On feature selection and effective classifiers, *Journal of ASIS*, vol.49, no.5, pp.423-434, 1998.

[8] L. H. Chiang and R. J. Pell, Genetic algorithms combined with discriminant analysis for key variable identification, *Journal of Process Control*, vol.14, pp.143-155, 2004.
[9] M. Raymer, W. Punch, E. Goodman, L. Kuhn and A. Jain, Dimensionality reduction using genetic algorithms, *IEEE Transactions on Evolutionary Computations*, vol.4, no.2, 2000.
[10] J. D. Deng, C. Simmermacher and S. Cranefield, A study on feature analysis for musical instrument classification, *IEEE Transactions on Systems, Man and Cybernetics-Part B: Cybernetics*, vol.38, no.2, 2008.
[11] Y. Marinakis, G. Dounias and J. Jantzen, Pap smear diagnosis using a hybrid intelligent scheme focusing on genetic algorithm based feature selection and nearest neighbor classification, *IEEE Computers in Biology and Medicine*, vol.39, pp.69-78, 2009.
[12] F. Hussein, N. Kharma and R. Ward, Genetic algorithms for feature selection and weighting, a review and study, *The 6th International Conference on Document Analysis and Recognition*, pp.1240-1244, 2001.
[13] W. Yang, D. Li and L. Zhu, An improved genetic algorithm for optimal feature subset selection from multi-character feature set, *Expert Systems with Applications*, vol.38, pp.2733-2740, 2011.
[14] T. Ingu and H. Takagi, Accelerating a GA convergence by fitting a single-peak function, *Proc. of IEEE International Fuzzy Systems Conference*, Seoul, Korea, 1999.
[15] U. Khairuddin, R. Yusof, M. Khalid and F. Cordova, Optimized feature selection for improved tropical wood species recognition system, *ICIC Express Letters, Part B: Applications*, vol.2, no.2, pp.441-446, 2011.
[16] M. Khalid, E. Y. L. Lew, R. Yusof and M. Nadaraj, Design of an intelligent wood species recognition system, *IJSSST*, vol.9, no.3, 2008.
[17] X. Qin and Y.-H. Yang, Aura 3D textures, *IEEE Transactions on Visualization and Computer Graphics*, vol.13, no.2, 2007.
[18] K. Teknomo, *Discriminant Analysis Tutorial*, http://people.revoledu.com/kardi/ tutorial/LDA/.