

Research Article

Probabilistic Adaptive Crossover Applied to Chilean Wine Classification

M. A. Duarte-Mermoud, N. H. Beltrán, and S. A. Salah

Electrical Engineering Department, University of Chile, Tupper Avenue 2007, Casilla 412-3, Santiago 8370451, Chile

Correspondence should be addressed to M. A. Duarte-Mermoud; mduartem@ing.uchile.cl

Received 17 September 2013; Revised 18 November 2013; Accepted 2 December 2013

Academic Editor: Jonathan N. Blakely

Copyright © 2013 M. A. Duarte-Mermoud et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Recently, a new crossover technique for genetic algorithms has been proposed. The technique, called probabilistic adaptive crossover (PAX), includes the estimation of the probability distribution of the population, storing the information regarding the best and the worst solutions of the problem being solved in a probability vector. The use of the proposed technique to face Chilean wine classification based on chromatograms obtained from an HPLC is reported in this paper. PAX is used in the first stage as the feature selection method and then support vector machines (SVM) and linear discriminant analysis (LDA) are used as classifiers. The results are compared with those obtained using the uniform (discrete) crossover standard technique and a variant of PAX called mixed crossover.

1. Introduction

Genetic algorithms (GAs) have been used in several fields such as adaptive filtering [1], classifiers [2], path planning [3], system identification [4], inventory modeling [5], fault diagnosis [6], portfolio investment [7–10], defining neural networks topology [11–13], function optimization [14–17], feature selection for classifiers [18–22], and wine classification [23–25], among several others.

GAs are one of the techniques known under the name of evolutionary algorithms (EAs) and are inspired by the concept of biological evolution [26]. The main idea is that each individual of the population is represented by a binary strip called chromosome, where each element in the strip is called gene or locus and the value that can be acquired is named allele. Each individual in the population corresponds to a possible solution of the problem being solved. An objective function is defined so that each individual has its own fitness. The probability that each individual has offspring is defined accordingly to its fitness. The evolution of the population will take place by generating randomly new possible solutions corresponding to the result of a crossover, the way in which the selected parents are mixed to generate the new solutions of the next generation. Then a new fitness

scoring is made and the possible solutions evolve along generations [26].

One of the most interesting applications of the GAs is certainly feature selection in the pattern recognition area [18–24, 27–29]. The methodology introduced by Holland in his 1975 book [30] has been referred to as GAs and was generalized later by De Jong [31] for the N crossover case. In 1987, Ackley [32] proposed the uniform crossover technique; however, this method is usually attributed to Syswerda who presented in 1989 a theoretical study of the uniform crossover [33, 34].

According to Yang [35] the way in which the crossover is adapted in a GA can be classified in three categories: adaptation of the crossover type [36–38], adaptation of the crossover rate [39, 40], and adaptation of the crossover position or the probability of change of each bit [41–44]. The new technique called probabilistic adaptive crossover (PAX) proposed in Salah et al. [26] belongs to the third category.

The purpose of this paper is to show an application of this new adaptive crossover technique PAX, which was recently introduced in [26], to an important problem of the Chilean economy related to the production and quality of its wine. In this technique the estimation of the population distribution is done in a new way rewarding the best individuals and

penalizing the worst individuals of the population. Once the estimation of the distribution is obtained, two parents are selected to generate new individuals, transferring the allele of each locus to the offspring in those places where the parents are equal and applying the estimated distribution to determine the alleles in those places where the parents have different values. Other similar techniques have been proposed by Varnamkhasti et al. [45] and the one reported by Talaslioglu [46]. In the former, a new crossover operator and probability selection technique are proposed based on the population diversity using a fuzzy logic controller, whereas in the latter a bipopulation-based genetic algorithm with enhanced interval search is introduced. In both cases the proposed methodologies are effective in finding better performance and quality solutions.

This paper presents the application of this new crossover technique used as the feature selection methodology in the problem of classifying the grape variety of Chilean red wines. To this extent a wine database was first built using different commercial Chilean wines of three type: cabernet sauvignon, carménère, and merlot. The classification is done based on the information contained in the liquid chromatograms of phenolic compounds provided by an HPLC. A brief description of the new PAX methodology is presented in Section 2 and its application to a wine database of 172 wine samples is developed in Section 3, where it is compared with uniform (discrete) crossover and a technique which is a variant of PAX called mixed crossover. Finally, some conclusions are drawn in Section 4.

2. Probabilistic Adaptive Crossover (PAX)

As mentioned above, in [26], a new crossover methodology called PAX was proposed for genetic algorithms, which uses a probability vector (probability that the bit in the i th position in one individual would be equal to 1) which is updated at each generation. The updating procedure is such that the evolution is done moving away from the worst genotypes and getting closer to the best genotypes. When the crossover is performed, this new method retains those bits that are equals in both parents in their offspring and randomly generates those bits that are different in both parents, using the previously updated probability vector. This probability vector presents characteristics of boundedness, forgetting factor, and learning rate. Thus, a methodology that feeds back the contribution of the different alleles to the fitness function is generated. This provides crossover with the kind of memory needed to use the information from previous evaluations. A new way of capturing population statistics is formulated favoring the best individuals and punishing the worst. The above allows stating a new feature selection methodology mixing two currently existing groups of methodologies: filtering and wrapped. In the following a brief description of PAX is presented although the reader is referred to [26] for a more detailed explanation of the method.

In the proposed method, those genes of the parents having the same values are directly transferred to the children. Since the fact that both parents have the same genetic

content in certain positions is expected by natural selection, these values are inherited to the children in a direct way, since they have produced good adaptation in their parents. For those genes where the parents have different values, what is analyzed is which of these values produces a greater adaptation to the environment, using the information from the experience of the entire population.

The method uses an adaptive vector of probabilities, P , which is constantly updated, computed for the entire population. This vector P represents the probability that a bit (loci) takes the value 1 in the children provided that both parents have different values. This vector also considers the information of the previous generations and is updated using the fitness of the individuals of the population. The updating procedure is done considering those individuals that present fitness over the average of the population as positive examples and those individuals presenting fitness below the average as negative examples.

In the updating process the information of previous generations is introduced, weighted by a forgetting factor α . This forgetting factor is introduced because the average of the fitness of the individuals theoretically increases, which is the reason why a solution that was over the average in generation t can be below the average in generation $t+1$. Thus, the probability vector is updated considering the population history as well as the current population. If the evolution has led to values near to extreme, almost all the individuals in the next generation will have the same value. In order to prevent this overfitting, this forgetting factor brings the probabilities values near to 0.5.

This updating methodology has another characteristic since the contribution to the adaptive vector of probabilities of those individuals which are over the average is weighted differently. Thus, each individual, that is, over the average does not contribute equally to updating the probability vector but rather the contribution, is made proportional to the fitness of each individual with respect to the others. The same concept is used for those individuals that are below the average, where those individuals of worst performance will be severely punished. In case that no allele stands out over the rest in some loci, the probability vector should converge to 0.5, which would transform the method into a discrete crossover [47].

The forgetting factor α is considered as follows:

$$\tilde{P}(t+1) = (P(t) - 0.5 \cdot \vec{1}) \cdot (1 - \alpha) + 0.5 \cdot \vec{1}, \quad (1)$$

where $\vec{1}$ denotes the unity vector of length L and \tilde{P} is an intermediate probability vector. After the application of the forgetting factor, the individuals x_t of the population Pobl_t are separated into the individuals whose fitness $f(x_t)$ is over the mean $\overline{f(t)}$, and those who are under the mean. These two groups are denoted as two subsets pobl_+ and pobl_- , respectively, and they are defined as

$$\begin{aligned} \text{pobl}_+ &= \{x_t \in \text{Pobl}_t \mid f(x_t) \geq \overline{f(t)}\}, \\ \text{pobl}_- &= \{x_t \in \text{Pobl}_t \mid f(x_t) < \overline{f(t)}\}. \end{aligned} \quad (2)$$

Then, the probability vector is updated using the following relationship:

$$\begin{aligned}
 P(t+1) &= \bar{P}(t+1) + \frac{\gamma}{2} \\
 &\times \left(\sum_{x_t \in \text{pobl}_+} \left(((f(x_t) - \overline{f(t)}) \cdot \vec{x}_t \right. \right. \\
 &\quad \left. \left. + (f(x_t) - \overline{f(t)}) \cdot (\vec{x}_t - \vec{1})) \right) \right. \\
 &\quad \left. \times \left(\sum_{x_t \in \text{pobl}_+} (f(x_t) - \overline{f(t)}) \right)^{-1} \right) \\
 &- \sum_{x_t \in \text{pobl}_-} \left(((f(x_t) - \overline{f(t)}) \cdot \vec{x}_t \right. \\
 &\quad \left. + (f(x_t) - \overline{f(t)}) \cdot (\vec{x}_t - \vec{1})) \right) \\
 &\quad \left. \times \left(\sum_{x_t \in \text{pobl}_-} (f(x_t) - \overline{f(t)}) \right)^{-1} \right), \quad (3)
 \end{aligned}$$

where γ is the learning rate that determinates the maximum value that can be used to update one position of the vector P . This maximum value is reached in the case where all the individuals in the set pobl_+ have the same value in one position, and the individuals in the set pobl_- have the opposite value in the same position (in the case of a binary alphabet).

Once the probability vector P is updated, the individuals that will be the parents of the next generation are selected according to the desired selection scheme (selection by tournament [48], ranking selection [49], stochastic universal sampling (SUS) [50], deterministic crowding [51], etc.). Next, the chosen parents are grouped in pairs, giving rise to two new children, which will have the same value of the parents in those positions where the values of the parents coincide. The rest of the positions will be filled with 1, with the probability indicated by vector P . After this it is possible to apply a tournament between children and parents, elitism operators or mutation, according to the desired scheme.

Tournament selection is a method for selecting the parents for the next generation making several tournaments between random selected individuals. The winner of each tournament is selected as a parent [48]. In the ranking selection method the probability of being selected as a parent for the next generation is assigned proportionally to the fitness of the individual [49]. SUS is an improvement of ranking selection using a single random value for selecting the whole population, reducing the bias of the ranking selection [50]. In this study the deterministic crowding technique [51] was chosen as selection method. In this method all the individuals are used as parents for the next generation.

In order to analyze empirically the performance of the proposed method PAX, in [26], the method is compared with six different crossover methodologies: one and two-point crossover, uniform crossover, discrete crossover, statistics-based adaptive nonuniform crossover (SANUX) [35], and selective crossover. From the studies performed in [26] over several databases it is concluded that PAX provides better solutions using up to 15% fewer computations compared with the other crossover methodologies, for the case of problems with one global optimum and when each bit contributes individually to the fitness function. On the other hand, in case of multiple optimum problems or when each bit does not make an explicit contribution to the fitness function, PAX performs equally well compared with the other methodologies studied.

3. Application of PAX as Feature Selection for Chilean Wine Classification

Recently, there has been a strong interest in applying techniques for wine classification. Numerous techniques and algorithms have been used in order to classify not only variety but also the geographical origin of wines. This classification has been done by using physical characteristics (color, density, conductivity, etc.), chemical features (anthocians, phenols, amino acids, etc.), and organoleptic characteristics (flavor, aromas, etc.) [23, 24, 26, 52–55].

In what follows, a methodology based on PAX [26] is presented for selecting the main variables from the viewpoint of wine variety classification, using information coming from liquid chromatograms supplied by an HPLC-DAD [53].

3.1. Problem Description. The information used for wine classification in this study corresponds to polyphenolic compounds of low molecular weights contained in liquid chromatograms obtained from an HPLC-DAD [53]. HPLC is an analytical chemistry technique used to separate the chemical compounds present in a liquid sample, in order to identify each component and to determine its concentration. Generally speaking, the method involves passing the liquid sample over a solid adsorbent material located inside a column, using a flow of liquid solvent. Each compound in the sample interacts differently with the adsorbent material so that light compounds flow off the column first, whereas heavier compounds flow off much later.

The equipment use in the study corresponds to a Merck-Hitachi model L-4200 UV-Vis Detector with a pump and column holder thermostat. The column is a Novapack C18, with a length of 300 mm and an inner diameter of 3.9 mm. For separation of the different phenolic compounds the following solvents were used:

- (A) 98% H₂O and 2% acetic acid;
- (B) 78% H₂O, 20% acetonitrilo, and 2% acetic acid;
- (C) 100% acetonitrilo.

After a series of trial experiments, it was determined that the best gradient to be used in the HPLC for the separation process was that shown in Table 1.

TABLE 1: Gradient used by the HPLC.

Time	Flux	Composition
0–55 min	1 mL/min	100% of A
55–57 min	1 mL/min	20% of A and 80% of B
57–90 min	1.2 mL/min	10% of A and 90% of B

Each chromatogram provided by the HPLC is a 90 minutes signal sampled every 800 ms and contains 6751 points. Each of these peaks has been identified by researchers in the area of chemistry and agronomy and associated to a particular chemical compound [53, 56, 57].

In general, the approach followed to identify the wine sample has been to identify each peak of the chromatogram (which is associated with a particular chemical compound) and then to use the concentration of these compounds as features for classification. The hypothesis is that each wine variety has a different combination of these compounds present in the chromatogram with a specific concentration. The approach we are proposing in this paper is to use all of the information contained in the chromatogram (electric signal), without identifying the chemical compounds associated with each peak, but using the complete chromatogram as a distinctive mark (a kind of “fingerprint”) associated to the wine sample. Here the hypothesis is that wines belonging to the same class have similar distinctive marks and wines belonging to different classes have distinctive marks that differ in some sense.

As an example, Figure 1 shows a typical normalized profile corresponding to a carménère wine sample supplied by the HPLC-DAD. The chromatograms last 90 minutes and contain 6751 points since the HPLC uses a sampling period of 800 milliseconds.

A spectral analysis of the chromatograms (DFT) revealed that, on average, 99.95% of the power spectrum is located at frequencies less than 0.125 Hz. Then using as Nyquist frequency for the data $f_n = 0.125$ Hz and applying the Shannon sampling theorem [58], it is possible to resample the chromatograms with a sampling period of $T_n = 1/(2f_n) = 1/(2 \cdot 0.125) = 4$ [s], losing about 0.05% of its information. Thus, after resampling the chromatograms at this new sampling period, the amount of points to be considered for each chromatogram are reduced from the original 6751 points to only 1350.

On the other hand, the first 5 minutes of the chromatograms were discarded since this part of the chromatogram contains basically information related to the effluents used to perform the liquid chromatogram but do not contain any useful information associated with the wine sample itself. Thus, we finally worked with chromatograms of 1276 points.

Since the size of the maxima depends upon the wine volume injected into the HPLC to obtain the chromatogram, a normalization procedure for the amplitudes was used, in order to avoid distortions. In some cases 20 mL of a prepared wine sample was injected in the HPLC but in others up to

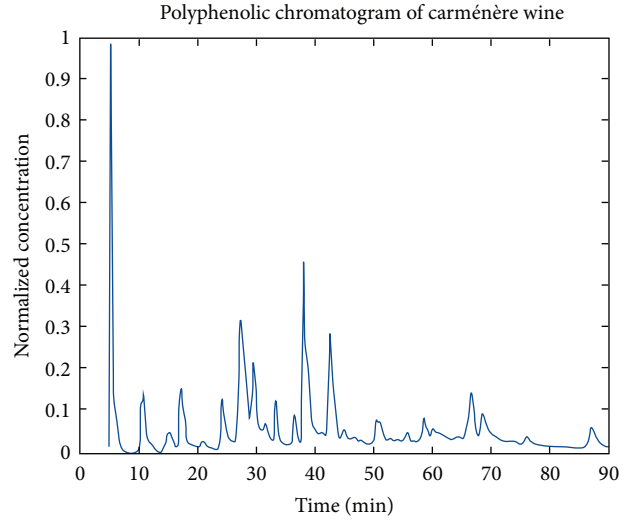


FIGURE 1: Normalized phenolic chromatogram of a typical Chilean carménère.

TABLE 2: Wine sample distribution according to the three grape varieties.

Variety	Training-validation samples	Test samples	Number of samples
Carménère	53	27	80
Cabernet Sauvignon	38	19	57
Merlot	23	12	35
Total	114	58	172

100 mL was used, and, as a consequence, peaks with different magnitudes were obtained in the chromatograms.

The normalization procedure was done using the following formula:

$$\tilde{X}_i = \frac{X_i - \min(X_i)}{\max(X_i) - \min(X_i)}. \quad (4)$$

Thus the values of all chromatograms will lie on the intervals 0 and 1, which allow comparing the chromatograms using a common base.

The 172 wine samples available for this study are distributed as shown in Table 2. These are commercial wines from different valleys of the central part of Chile (Maipo, Rapel, Curicó, Maule and Itata) and are from vintages from 2000 and 2001.

A standard practice in data mining is to separate data in two sets: one for training-validation purposes and the other for testing. With the first set, all the parameters of the classifier are suitably tuned and then the test set is presented to the classifier to evaluate its performance. The choice of the training-validation and the test sets was done at random, choosing the number of samples proportional to the total number of samples available for each class. Based on previous works [23, 24, 54], 2/3 of the database was used for training-validation and the rest for testing.

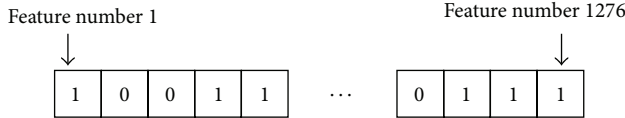


FIGURE 2: Strip representing a sample (chromatogram).

3.2. Methodology Description

3.2.1. Individual Coding. For individual coding purposes binary strips of length N were used. N is the number of features available for the problem, which in this case is $N = 1276$. (Each feature is associated with a point of the chromatogram). The existence of a 1 at the i th position indicates that the GA has chosen feature i of the sample to be considered in the classification stage and the existence of a zero means that this feature should not be considered in the classification process, as is shown in Figure 2.

The choice of the number of individuals for each new generation will be denoted as N_i . This number has a direct relationship to the amount of computation necessary to find the optimal solution. As this number increases the exploration is wider, but the number of computations is also greater.

In this study the value $N_i = 150$ was chosen, which is equal to that used in the study reported in [23] on a similar problem. A simple mutation probability of $P_m = 0.001$ was considered together with 150 generations per each population.

The individuals of the initial population were chosen in the following manner: a random number of characteristics r between 1 and 1276 was first chosen. Then these r features (numbers one) were randomly located in the binary strip of length 1271, obtaining an individual of the initial population.

3.2.2. Fitness Function. In order to define the fitness of each individual, classification using SVM was used considering as kernel the radial basis function

$$K(Y_i, Y_j) = \exp\left(\frac{\|Y_i - Y_j\|}{2\sigma^2}\right), \quad (5)$$

where Y_i, Y_j are the two patterns being compared and σ is the spread of the RBF.

The fitness considers the number of correct classifications normalized by the total number of individuals employed in the training. This number is penalized subtracting a term proportional to the number of characteristics, normalized by the total number of characteristics available (L), weighted by the factor C . The fitness is then determined by

$$f(x_i) = \frac{N_{\text{correct_classifications}}(x_i)}{N_{\text{training_samples}}} - C \cdot \frac{N_{\text{char}}(x_i)}{L}. \quad (6)$$

For this study a value of $C = 0.1$ was considered, since with this value an individual (solution) classifying correctly one sample less than another individual but uses 11 fewer features

TABLE 3: Parameters used for the SVM.

SVM	Q	σ
Cabernet-rest	4	0.048
Carménère-rest	4	0.28
Merlot-rest	2	0.022
Cabernet-carménère	28	0.004
Merlot-carménère	4	0.074
Cabernet-merlot	28	0.002

obtains a similar fitness to that of an individual classifying correctly one more sample but using more features. Therefore each sample wrongly classified is compensated by 11 features.

The percentage of correct classification was obtained by using the *K-fold-cross-validation* [36] methodology over the training-validation set with $K = 6$. *K-fold-cross-validation* consists of separating the data set into K groups and training the classifier with $K - 1$ of them and then validating with the remaining group. This is repeated for each of the K sets.

Since SVMs are defined for two classes, a scheme was developed in order to classify one class first with respect to the other two (“other classes”) and then, using another SVM, samples classified as “other classes” are separated. Since there are three possible schemes (depending on what class is left out) all of the three schemes were implemented. The results obtained from these three schemes were voted on, assigning the sample X to class (C_i), when this is the solution in two out of the three possible classification schemes.

To determine the SVM parameters, the methodology *K-fold-cross-validation* with $K = 6$ was used. Different combinations of σ values (spread of the radial basis function) between 0.001 and 1 and Q values (degree of penalization for wrongly classified samples in SVM) between 1 and 100 were chosen and evaluated. The best parameters in each case are shown in Table 3.

3.2.3. Crossover and Selection Methodology. In this study the niching method called deterministic crowding was used [51] for selecting the individuals to produce the next generation. Three different crossover techniques were used for feature selection with GA. These techniques were discrete crossover, PAX (with learning rate of 0.1, forgetting factor 0.1 and bounds between 0.3 and 0.7), and a modification of PAX called mixed crossover (with learning rate of 0.1, forgetting factor 0.1, and bounds between 0.3 and 0.7). In the mixed crossover methodology the probability vector is not directly updated using each individual *fitness* but rather using other measures applied to the data (Fisher criteria in this study). This combination would correspond to an approximation of the hybrid methodology between the filter feature selection method and the *filter-wrapped* selection method. This corresponds to an open loop-close loop approach that can be applied since in PAX individuals are compared according to the results of the SVM classifier, but they are led towards the feature sets presenting the best characteristics according to the separation measure being used (Fisher criteria in our case). Figure 3 presents the block diagram of the classification method used here.

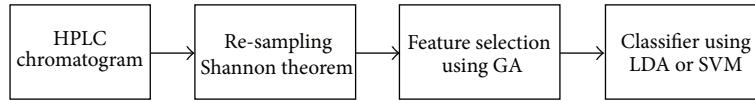


FIGURE 3: Block diagram of the classification system.

TABLE 4: Results obtained for feature selection of Chilean wines using SVM.

Run	Mixed			PAX			Discrete		
	Train./valid. (%)	Test. (%)	N° Char.	Train./valid. (%)	Test. (%)	N° Char.	Train./valid. (%)	Test. (%)	N° Char.
1	82.46	68.97	150	86.84	68.97	211	83.33	65.52	194
2	84.21	65.52	143	87.72	68.97	219	86.84	70.69	182
3	86.84	67.24	141	86.84	65.52	159	86.84	65.52	152
4	87.72	70.69	140	86.84	67.24	212	86.84	65.52	189
5	81.58	68.97	160	86.84	67.24	236	85.09	62.07	193
6	86.84	56.90	171	88.60	70.69	263	87.72	67.24	203
7	83.33	63.79	154	88.60	67.24	201	85.96	70.69	189
8	85.96	65.52	210	86.84	75.86	244	86.84	72.41	206
9	87.72	68.97	172	87.72	72.41	234	88.60	70.69	181
10	87.72	67.24	220	87.72	68.97	200	87.72	70.69	211
11	81.58	67.24	127	82.46	67.24	209	84.21	65.52	189
12	83.33	65.52	159	87.72	60.34	201	87.72	68.97	195
13	85.09	67.24	226	87.72	68.97	194	87.72	65.52	189
14	88.60	72.41	213	85.96	70.69	183	87.72	70.69	191
15	84.21	60.34	167	87.72	70.69	213	86.84	72.41	188
Average	85.15	66.44	170.20	87.08	68.74	211.93	86.67	68.28	190.13
Stand. dev.	2.38	3.90	31.95	1.46	3.44	25.50	1.45	3.18	13.40

3.3. *Results.* 15 runs using different randomly chosen initial populations of 150 individuals were analyzed. At each run the sets used to perform the training and the set for cross-validation were changed each time, choosing it randomly. The maximum percentage of correct classification obtained in training-validation and the number of features selected by the GA (N° Char.) are shown in Table 4 for the three crossover techniques employed using SVM as the classifier. Also, the percentage of correct classification obtained in testing is presented using the features proposed by the best individual of the population in the generation 150.

Figure 4 shows the average fitness of the best individuals at each generation when using the GA with mixed, PAX, and discrete crossover methodologies, using SVM as the classifier. Figure 5 shows the 213 characteristics selected by the mixed crossover technique plotted over the chromatogram of a cabernet sauvignon sample (Sample 1). These characteristics correspond to those obtained for the best individual in run 14.

The maximum percentage of correct classification obtained in training-validation and the number of features (N° Char.) selected by the three crossover techniques studied using LDA as the classifier are shown in Table 5. The classification rate obtained in testing is also presented, using the characteristics obtained for the best individual of the population after 150 generations.

The average fitness of the best individuals of each generation for the three crossover techniques (mixed, PAX, and

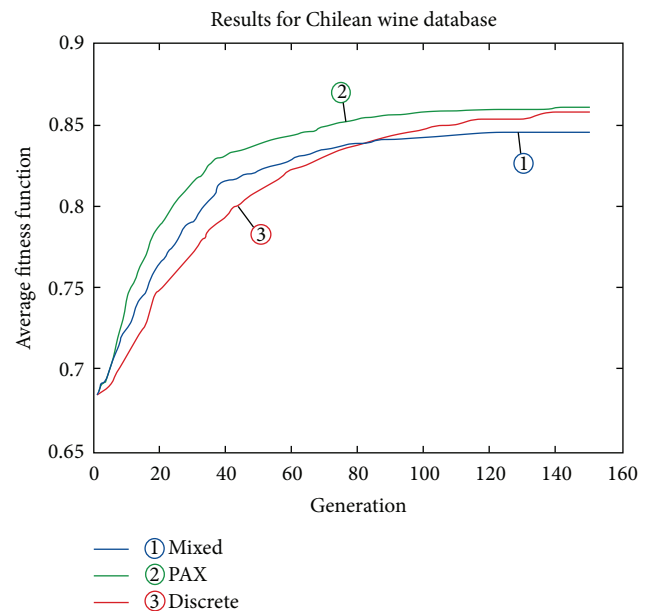


FIGURE 4: Average evolution of the best individual using SVM.

discrete) when LDA is used as the classifier are presented in Figure 6. In Figure 7 the 22 features selected by the mixed crossover methodology, plotted over the chromatogram of a cabernet sauvignon wine sample (Sample 1) are indicated.

TABLE 5: Results obtained for feature selection of Chilean wines using LDA.

Run	Mixed			PAX			Discrete		
	Train./valid. (%)	Test. (%)	N° Char.	Train./valid. (%)	Test. (%)	N° Char.	Train./valid. (%)	Test. (%)	N° Char.
1	96.49	74.14	43	96.84	70.69	48	96.49	67.24	27
2	91.23	70.69	45	99.12	72.41	43	97.37	70.69	26
3	94.74	68.97	39	99.12	75.86	43	92.98	62.07	23
4	96.49	65.52	73	94.74	70.69	42	95.61	74.14	28
5	94.74	75.86	41	100.00	72.41	54	96.49	75.86	35
6	97.37	75.86	54	98.25	72.41	48	94.74	67.24	30
7	96.49	70.69	36	98.25	75.86	50	92.98	70.69	30
8	93.86	70.69	53	96.49	67.24	47	93.86	75.86	35
9	95.61	70.69	36	99.12	70.69	53	95.61	70.69	31
10	89.47	67.24	39	95.61	67.24	46	97.37	72.41	28
11	90.35	75.86	22	100.00	74.14	48	98.25	65.52	26
12	89.47	72.41	31	97.37	67.24	53	96.49	68.97	27
13	92.11	68.97	29	100.00	74.14	42	94.74	70.69	38
14	90.35	75.86	22	95.61	72.41	42	91.23	68.97	22
15	92.11	72.41	30	96.49	74.14	46	93.86	75.86	52
Average	93.39	71.72	39.53	97.80	71.84	47.00	95.20	70.46	30.53
Stand. dev.	2.81	3.31	13.28	1.77	2.89	4.16	1.96	4.01	7.40

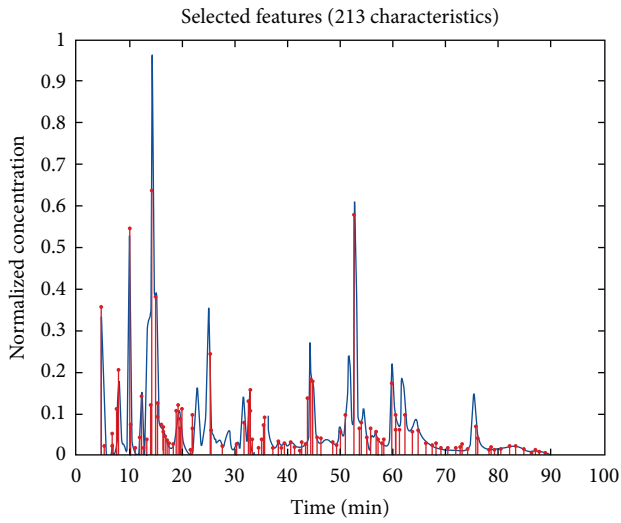


FIGURE 5: Selected features by the mixed crossover technique using SVM, plotted over the chromatogram of the wine Sample 1 (cabernet sauvignon).

These features correspond to those obtained for the best individual in run 14.

3.4. Analysis and Discussion of the Results. Simulations were performed on 12 Pentium IV computers of 1.5 GHz and 256 MB RAM running independently. The software used was MATLAB 6.5 and the GENETIC ALGORITHM TOOLBOX [59]. LDA classification was done using the toolbox DISCRIM of MATLAB [60] and the SVM classification was done using LIBSVM [61], an open source machine learning library.

From Table 4 and Figure 4 it can be seen that the three crossover methodologies achieve similar results, as far as

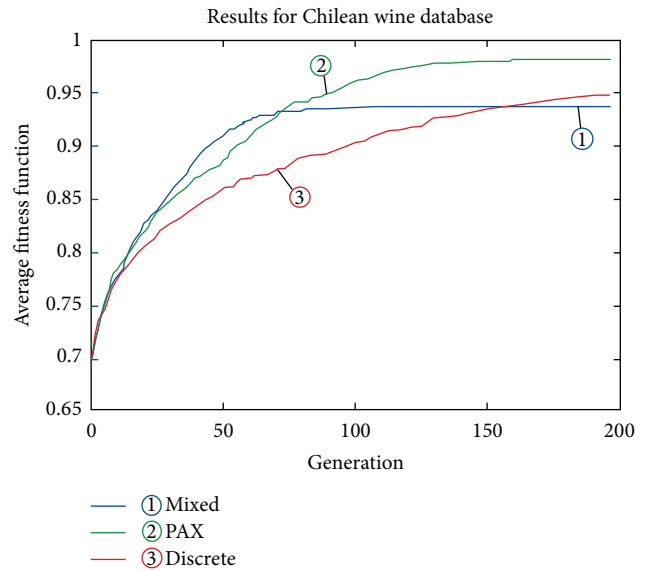


FIGURE 6: Average evolution of the best individual using LDA.

fitness is concerned, when using SVM. The number of selected features differs from one method to another, with the mixed crossover being the one with the smaller number of selected features and with a lower value of the objective function (6).

As far as the convergence speed is concerned it can be appreciated from Figure 4 that PAX increases during the first generations with a larger rate than the other techniques. It is important to note that the velocity at which PAX reaches values near 0.85 occurs around generation 70, whereas the same level is reached by the discrete crossover at around generation 120. This shows a significant computational saving

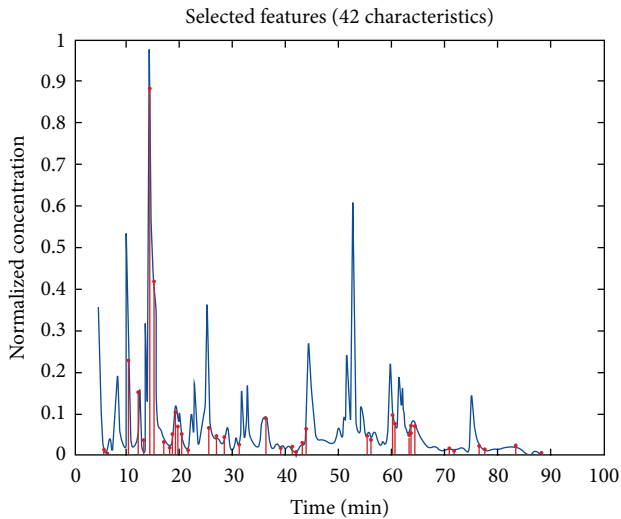


FIGURE 7: Selected features by the mixed crossover technique using LDA, plotted over the chromatogram of the wine Sample 1 (cabernet sauvignon).

when using PAX as compared with use of the discrete crossover.

Regarding the speed of convergence, similar results to those obtained with SVM are obtained when LDA is used as the classifier (see Table 5). However, the number of features selected with LDA is 75% or 80% less than the number obtained with SVM.

For both classifiers values around 70% are obtained in testing, attributable to the small number of wine samples in the database.

Other results on wine classification, using the same wine database, were obtained by the authors and reported in [62]. There, a feature extraction stage was used instead of feature selection, as is reported here. Several combinations of feature extraction methods followed by the quadratic discriminant analysis (QDA) classification technique were studied. Average correct classification rates as high as 99% in the test set were reached when using the quadratic fisher transformation (QFT) for feature extraction [63]. These results are superior to those reported here and a future investigation will be to analyze why this large difference is produced in both methodologies and how the results reported here could be improved.

4. Conclusions

The application of a genetic algorithm with a new crossover technique has been applied to the problem of classifying Chilean red wines. The GA was used as the feature selection method, and the classifying technique is based on the information contained in the liquid chromatograms of phenolic compounds present in the wine samples obtained from an HPLC. The results obtained using SVM and LDA with a database of 172 wine samples are quite acceptable and the PAX methodology (standard and modified) presents some advantages over the uniform (discrete) crossover technique.

Much better results are obtained when LDA is used as the classifier compared with those obtained when SVM is used as the classifier. In the first case, classification rates of 71% are achieved on average against only 68% in the SVM case. More important is the fact that the number of selected features decreases on average from 200 (in the case of SVM) to only 40 for LDA. Nevertheless, this last aspect has to be carefully examined. In fact, since the objective function (6) considers the number of features, suitably weighted through parameter C , as well as the percentage of correct classification, lower number of features does not necessarily mean better classification rates.

In general, the application of PAX to this wine database reveals that PAX maintains the good characteristics exhibited when applied to standard databases [26], regarding the speed of convergence and the quality of the obtained solutions. Unfortunately, the good results obtained in training-validation are not reflected in testing. This can be attributable to the fact that the low number of elements (samples) in the database is not representative enough for the problem under analysis, which could be causing an overtraining.

Finally, it is important to point out that PAX allows developing open loop-close loop optimization methodologies, as the one presented in the mixed crossover approach, leading to better solutions, as long as a suitable fitting function is defined. In this sense, other combinations of open loop-close loop measures applied to the case of SVM, as it was in the case of LDA where the Fisher criteria was used, could improve the results reported here using SVM. This statement can be viewed as a future work in this area.

Conflict of Interests

The authors declare that they do not have conflict of interests of any kind regarding the research contained in this paper.

Acknowledgment

The results presented in this work were supported by CONY-CIT-Chile, under Grant FONDEF D01-1016, "Chilean Red Wine Classification by means of Intelligent Instrumentation."

References

- [1] K. Hu, A. Song, M. Xia, X. Ye, and Y. Dou, "An adaptive filtering algorithm based on genetic algorithm-backpropagation network," *Mathematical Problems in Engineering*, vol. 2013, Article ID 573941, 8 pages, 2013.
- [2] J. Liu, Z. Yu, and D. Ma, "An adaptive fuzzy min-max neural network classifier based on principle component analysis and adaptive genetic algorithm," *Mathematical Problems in Engineering*, vol. 2012, Article ID 483535, 21 pages, 2012.
- [3] G. Giardini and T. Kalmár-Nagy, "Genetic algorithm for combinatorial path planning: the subtour problem," *Mathematical Problems in Engineering*, vol. 2011, Article ID 483643, 31 pages, 2011.
- [4] R. V. Cruz and L. C. Sandoval Ges, "Results of short-period helicopter system identification using output-error and hybrid

- search-gradient optimization algorithm,” *Mathematical Problems in Engineering*, vol. 2010, Article ID 231594, 17 pages, 2010.
- [5] S. Kar, D. Das, and A. Roy, “A production-inventory model for a deteriorating item incorporating learning effect using genetic algorithm,” *Advances in Operations Research*, vol. 2010, Article ID 146042, p. 26, 2010.
 - [6] T. Han, B. S. Yang, W. H. Choi, and J. S. Kim, “Fault diagnosis system of induction motors based on neural network and genetic algorithm using stator current signals,” *International Journal of Rotating Machinery*, vol. 2006, Article ID 61690, 13 pages, 2006.
 - [7] J. Shapcott, “Index Tracking: Genetic Algorithms for Investment Portfolio Selection,” Report EPCC-SS92-24, Edinburgh Parallel Computing Centre, The University of Edinburgh, JCMB, The King’s Buildings, Mayfield Road, Edinburgh, EH9 3JZ, UK, 1992.
 - [8] J.-S. Chen, J.-L. Hou, S.-M. Wu, and Y.-W. Chang-Chien, “Constructing investment strategy portfolios by combination genetic algorithms,” *Expert Systems with Applications*, vol. 36, no. 2, pp. 3824–3828, 2009.
 - [9] P.-C. Lin and P.-C. Ko, “Portfolio value-at-risk forecasting with GA-based extreme value theory,” *Expert Systems with Applications*, vol. 36, no. 2, pp. 2503–2512, 2009.
 - [10] S. M. Wang, J. C. Chen, and K.-J. Wang, “Resource portfolio planning of make-to-stock products using a constraint programming-based genetic algorithm,” *Omega*, vol. 35, no. 2, pp. 237–246, 2006.
 - [11] W. Schiffmann, M. Joost, and R. Werner, “Application of genetic algorithms to the construction of topologies for multilayer perceptrons,” in *Proceedings of the International Conference on Artificial Neural Networks and Genetic Algorithms (ICANNGA '93)*, pp. 675–682, 1993.
 - [12] T. Nazghelichi, M. Aghbashlo, and M. H. Kianmehr, “Optimization of an artificial neural network topology using coupled response surface methodology and genetic algorithm for fluidized bed drying,” *Computers and Electronics in Agriculture*, vol. 75, no. 1, pp. 84–91, 2011.
 - [13] M. Saemi, M. Ahmadi, and A. Y. Varjani, “Design of neural networks using genetic algorithm for the permeability estimation of the reservoir,” *Journal of Petroleum Science and Engineering*, vol. 59, no. 1-2, pp. 97–105, 2007.
 - [14] A. Eiben, “Multi-parent recombination,” in *Handbook of Evolutionary Algorithms*, pp. 25–33, IOP Publishing Ltd. and Oxford University Press, 1997.
 - [15] L.-N. Xing, Y.-W. Chen, and H.-P. Cai, “An intelligent genetic algorithm designed for global optimization of multi-minima functions,” *Applied Mathematics and Computation*, vol. 178, no. 2, pp. 355–371, 2006.
 - [16] N. Tutkun, “Optimization of multimodal continuous functions using a new crossover for the real-coded genetic algorithms,” *Expert Systems with Applications*, vol. 36, no. 4, pp. 8172–8177, 2009.
 - [17] K. Nath Das and R. Mishra, “Chemo-inspired genetic algorithm for function optimization,” *Applied Mathematics and Computation*, vol. 220, pp. 394–404, 2013.
 - [18] A. K. Jain, R. P. W. Duin, and J. Mao, “Statistical pattern recognition: a review,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 1, pp. 4–37, 2000.
 - [19] J. Yang and V. Honavar, “Feature subset selection using genetic algorithm,” *IEEE Intelligent Systems and Their Applications*, vol. 13, no. 2, pp. 44–48, 1998.
 - [20] S. Venkadesh, G. Hoogenboom, W. Potter, and R. McClendon, “A genetic algorithm to refine input data selection for air temperature prediction using artificial neural networks,” *Applied Soft Computing*, vol. 13, no. 5, pp. 2253–2260, 2013.
 - [21] M. E. ElAlami, “A filter model for feature subset selection based on genetic algorithm,” *Knowledge-Based Systems*, vol. 22, no. 5, pp. 356–362, 2009.
 - [22] M. Jalali-Heravi and A. Kyani, “Application of genetic algorithm-kernel partial least square as a novel nonlinear feature selection method: activity of carbonic anhydrase II inhibitors,” *European Journal of Medicinal Chemistry*, vol. 42, no. 5, pp. 649–659, 2007.
 - [23] S. A. Salah, M. A. Duarte-Mermoud, N. H. Beltrán et al., “Selección de Características usando Algoritmos Genéticos para Clasificación de Vinos Chilenos,” *Anales Instituto De Ingenieros De Chile*, vol. 115, no. 1, pp. 3–10, 2003.
 - [24] N. H. Beltrán, M. A. Duarte-Mermoud, S. A. Salah et al., “Feature selection algorithms using Chilean wine chromatograms as examples,” *Journal of Food Engineering*, vol. 67, no. 4, pp. 483–490, 2005.
 - [25] J. A. Roubos, M. Setnes, and J. Abonyi, “Learning fuzzy classification rules from labeled data,” *Information Sciences*, vol. 150, no. 1-2, pp. 77–93, 2003, Recent advances in soft computing (Leicester, 2000).
 - [26] S. A. Salah, M. A. Duarte-Mermoud, and N. H. Beltrán, “Probabilistic adapting crossover (PAX): a novel genetic algorithm crossover methodology,” *International Journal on Artificial Intelligence Tools*, vol. 17, no. 6, pp. 1131–1160, 2008.
 - [27] A. L. Blum and P. Langley, “Selection of relevant features and examples in machine learning,” *Artificial Intelligence*, vol. 97, no. 1-2, pp. 245–271, 1997.
 - [28] M. Dash and H. Liu, “Feature selection for classification,” *Intelligent Data Analysis*, vol. 1, pp. 131–156, 1997.
 - [29] G. H. John, R. Kohavi, and K. Pfleger, “Irrelevant features and the subset selection problem,” in *Proceedings of the 11th International Conference on Machine Learning*, pp. 121–129, Morgan Kaufmann, New Brunswick, NJ, USA, 1994.
 - [30] J. H. Holland, *Adaptation in Natural and Artificial Systems*, MIT Press, Cambridge, Mass, USA, 2nd edition, 1975.
 - [31] K. De Jong, *An Analysis of the Behavior of a Class of Genetic Adaptive Systems [Ph.D. thesis]*, University of Michigan, 1975.
 - [32] D. Ackley, *A Connectionist Machine for Genetic Hillclimbing*, vol. 28 of *The Kluwer International Series in Engineering and Computer Science*, Kluwer Academic, Boston, Mass, USA, 1987.
 - [33] G. Syswerda, “Uniform crossover in genetic algorithms,” in *Proceedings of the 3rd International Conference on Genetic Algorithm*, pp. 2–9, Morgan Kaufmann, San Mateo, Calif, USA, 1989.
 - [34] G. Syswerda, *Simulated Crossover in Genetic Algorithms*, Foundations of Genetic Algorithms 2, Morgan Kaufmann, 1993.
 - [35] S. Yang, “Adaptive non-uniform crossover based on statistics for genetic algorithms,” in *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO '02)*, pp. 650–657, Morgan Kaufmann, New York, NY, USA, 2002.
 - [36] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, Academic Press, Boston, Mass, USA, 2nd edition, 1990.
 - [37] W. Spears, “Adapting crossover in a genetic algorithm,” Technical Report AIC-92-025, Naval Research Laboratory, Navy Center for Applied Research on Artificial Intelligence, Washington, DC, USA, 1992.

- [38] P. M. Narendra and K. Fukunaga, "A branch and bound algorithm for feature selection," *IEEE Transactions on Computers*, vol. 26, no. 9, pp. 917–922, 1977.
- [39] D. Corne, P. Ross, and H. Fang, "Fast practical evolutionary timetabling," in *Proceedings of the AISB Workshop on Evolutionary Computing*, Springer, 1994.
- [40] D. Koller and M. Sahami, "Toward optimal feature selection," in *Proceedings of 13th International Conference on Machine Learning*, pp. 284–292, 1996.
- [41] J. Schaffer and A. Morishima, "An adaptive crossover distribution mechanism for genetic algorithms," in *Proceedings of the 2nd International Conference on Genetic Algorithms*, pp. 36–40, Lawrence Erlbaum Associates, 1987.
- [42] S. Louis and G. Rawlins, "Designer genetic algorithms: genetic algorithms in structure design," in *Proceedings of the 4th International Conference on Genetic Algorithms*, pp. 53–60, Morgan Kaufmann, 1991.
- [43] T. White and F. Oppacher, "Adaptive Crossover Using Automata," in *Proceedings of the 3rd Conference Parallel Problem Solving from Nature Conference*, pp. 229–238, Springer, 1994.
- [44] K. Vekaria and C. Clack, "Selective crossover in genetic algorithms," in *Proceedings of the 3rd Annual Conference on Genetic Programming*, pp. 609–611, Morgan Kaufmann, 1998.
- [45] M. J. Varnamkhasti, L. S. Lee, M. R. Abu Bakar, and W. J. Leong, "A genetic algorithm with fuzzy crossover operator and probability," *Advances in Operations Research*, vol. 2012, Article ID 956498, 16 pages, 2012.
- [46] T. Talaslioglu, "A new genetic algorithm methodology for design optimization of truss structures: bipopulation-based genetic algorithm with enhanced interval search," *Modelling and Simulation in Engineering*, vol. 2009, Article ID 615162, 28 pages, 2009.
- [47] H. Mühlenbein and D. Schlierkamp-Voosen, "Predictive models for the Breeder genetic algorithm: I. Continuous parameter optimization," *Evolutionary Computation*, vol. 1, no. 1, pp. 25–49, 1993.
- [48] T. Blicke and L. Thiele, "A mathematical analysis of tournament selection," in *Proceedings of the 6th International Conference on Genetic Algorithms (ICGA '95)*, pp. 9–16, Morgan Kaufmann, San Francisco, Calif, USA, 1995.
- [49] J. E. Baker, "Adaptive selection methods for genetic algorithms," in *Proceedings of the 1st International Conference on Genetic Algorithms*, pp. 101–111, Lawrence Erlbaum Associates, Hillsdale, NJ, USA, July 1985.
- [50] J. E. Baker, "Reducing Bias and inefficiency in the selection algorithm," in *Proceedings of the 2nd International Conference on Genetic Algorithms and Their Applications*, pp. 14–22, Lawrence Erlbaum Associates, Hillsdale, NJ, USA, July 1987.
- [51] S. W. Mahfoud, *Niching methods for genetic algorithms [Ph.D. thesis]*, University of Illinois at Urbana-Champaign, Illinois Genetic Algorithms Laboratory (IlligAL), Report No. 95001, 1995.
- [52] R. Marx, B. Holbach, and H. Otteneder, "Determination of nine characteristics anthocyanins in wine by HPLC," *Off. Int. Vigne Vin. Bulletin*. Paris, France. August 2000.
- [53] A. Peña-Neira, T. Hernández, C. García-Vallejo, I. Estrella, and J. A. Suarez, "A survey of phenolic compounds in Spanish wines of different geographical origin," *European Food Research and Technology*, vol. 210, no. 6, pp. 445–448, 2000.
- [54] N. H. Beltrán, M. A. Duarte-Mermoud, M. A. Bustos et al., "Feature extraction and classification of Chilean wines," *Journal of Food Engineering*, vol. 75, no. 1, pp. 1–10, 2006.
- [55] C. Flazy, *Enología: Fundamentos científicos y tecnológicos*, AMV-Mundi Prensa, Madrid, Spain, 2000.
- [56] V. S. Alamo, *Caracterización de la composición fenólica de vinos comerciales Merlot y Sauvignon Blanc de la vendimia 2002, provenientes de cinco valles de Chile*, Memoria de Ingeniero Agrónomo, Facultad de Ciencias Agronómicas, Universidad de Chile, 2002.
- [57] L. P. Muñoz, *Caracterización de la composición fenólica de vinos comerciales Cabernet Sauvignon*, Memoria de Ingeniero Agrónomo, Facultad de Ciencias Agronómicas, Universidad de Chile, 2002.
- [58] R. H. Middleton and G. C. Goodwin, *Digital Control and Estimation. A Unified Approach*, Prentice Hall, 1990.
- [59] A. Chipperfield and P. Fleming, "The MATLAB Genetic Algorithm Toolbox," Digest 1995/014, IEEE Colloquium on Applied Control Techniques Using MATLAB, 1995.
- [60] M. Kieft, "Discriminant Analysis Toolbox ver. 0.3 for MATLAB," 1999, <http://www.mathworks.com/matlabcentral/fileexchange/loadFile.do?objectType=file&objectId=189>.
- [61] C. C. Chang and C. J. Lin, "LIBSVM: A Library for Support Vector Machines," 2001, <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [62] M. A. Duarte-Mermoud, N. H. Beltrán, and M. A. Bustos, "Chilean wine varietal classification using quadratic Fisher transformation," *Pattern Analysis and Applications*, vol. 13, no. 2, pp. 181–188, 2010.
- [63] M. A. Bustos, M. A. Duarte-Mermoud, and N. H. Beltrán, "Nonlinear feature extraction using fisher criterion," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 22, no. 6, pp. 1089–1119, 2008.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

