

A Novel Swarm based Feature Selection Algorithm in Multifunction Myoelectric Control

Rami N. Khushaba¹, Akram AlSukker¹, Ahmed Al-Ani¹, Adel Al-Jumaily¹
Albert Y. Zomaya²

¹ Mechatronics and Intelligent Systems Group, Faculty of Engineering,
University of Technology, Sydney
P.O. Box 123, Broadway 2007, Sydney - NSW, Australia.

² School of Information Technologies, Building J12
University of Sydney, NSW 2006, Australia

Corresponding author

Rami N. Khushaba (*PhD Candidate*)

Faculty of Engineering, B2 – L6 – R32 – Desk02

University of Technology, Sydney

P. O. Box: 123, Broadway 2007.

NSW – Australia.

Tel: (612)-9514-3140

E-mail: rkhushab@eng.uts.edu.au

ABSTRACT

Accurate and computationally efficient myoelectric control strategies have been the focus of a great deal of research in recent years. Although many attempts exist in literature to develop such strategies, deficiencies still exist. One of the major challenges in myoelectric control is finding an optimal feature set that can best discriminate between classes. However, since the myoelectric signal is recorded using multi channels, the feature vector size can become very large. Hence a dimensionality reduction method is needed to identify an informative, yet small size feature set. This paper presents a new feature selection method based on modifying the Particle Swarm Optimization (PSO) algorithm with the inclusion of Mutual Information (MI) measure. The new method, called BPSOMI, is a mixture of filter and wrapper approaches of feature selection. In order to prove its efficiency, the proposed method is tested against other dimensionality reduction techniques proving powerful classification accuracy.

Index Terms— Mutual information, Myoelectric Control, Particle Swarm Optimization, Pattern Recognition

I. INTRODUCTION

The field of biosignals pattern recognition and interpretation has enjoyed a rapid increase in popularity in the past few years. This kind of research offers a natural way to interface with the external world for the amputees and people with neuromuscular handicaps. The term myoelectric control (MEC) refers to the process of utilizing the Myoelectric Signals (MES) from human muscles in order to control a powered external device. Specifically, the MEC is usually utilized in controlling prosthetic devices that function as artificial alternatives to missing limbs. The MES, also known as the Electromyogram (EMG) signal, is one of the biosignals generated by the human body [20]. It represents the muscles activity or the summation of the action potentials from many motor units [10, 15]. The MES signal is a one dimensional non-stationary signal that carries the distinct signature of the voluntary intent of the central nervous system [7].

The MES exhibit distinct differences in the temporal structure for different kinds of arm movements. An example of the MES collected using two electrodes placed on the Biceps and triceps muscles are shown in Fig.1 for different forearm movements [5, 4]. This in turn facilitates the use of a pattern recognition approach to identify the user's intention based on the rich information presented in the form of a muscular contraction. The most significant advantage of pattern recognition based myoelectric control strategies is that it represents an intention based control, i.e., a hand-free control scheme [14]. In most of the amputees, the nerves underlying the skin can still be innervated and are thus subject to voluntary control based on the user's intention.

In spite of being an informative signal, the MES detected at different locations over the same muscle may have significantly different amplitudes [19]. This in turn would increase the complexity in recording such a signal. In order to capture the complete amputee's muscle activities, the MES is usually recorded using a number of channels, e.g. 4, 8, or 16. In myoelectric control, various techniques and methods were developed and utilized to extract features from the various MES channels, which usually leads to large feature vectors. The high dimensionality of

the feature vector causes an increase in the learning parameters of the pattern classifier, and the convergence of learning deteriorates [8]. Thus dimensionality reduction becomes an indispensable step in such biosignal pattern recognition system.

Dimensionality reduction is the process of reducing the number of variables under consideration. It can be implemented using Feature Selection (FS) methods that are semantics-preserving, or Feature Projection (FP) methods that transform the underlying meaning of the original features [17]. In a study conducted by Englehart [11], a comparison was made between feature selection and projection on both transient and steady-state MES datasets. Different feature selection techniques were tested against principal component analysis (PCA), which is a well-known feature projection method. Englehart's comparison revealed that the PCA technique gave better performance than all other feature selection techniques [12]. This formed a motivation to most of the researches working in this field to focus on feature projection methods in myoelectric control. Some of the proposed methods include the combination of PCA and a self organizing feature map (SOFM) [8], linear discriminant analysis (LDA) based feature projection [9], and the uncorrelated linear discriminant analysis (ULDA) that requires the reduced features to be statistically uncorrelated with one another [2].

This paper presents a new approach to re-evaluate the significance of feature selection on MES classification problems. This study is motivated by the fact that the feature selection methods used by Englehart were not optimal. Hence, we are proposing here a new hybrid swarm based feature selection method to be used as a dimensionality reduction tool in myoelectric control. The block diagram of the proposed system is shown in Fig.2, given with Hudgins time domain features as an example [4]. The new method, abbreviated as BPSOMI, is applied on the features extracted from the MES data collected from multiple channels in order to reduce the dimensionality of the extracted features set. The BPSOMI is based on modifying the binary PSO method with the inclusion of a mutual information (MI) measure. The MI concept is employed to estimate the

relevance and redundancy of feature subsets. The justification behind using PSO is due to its parallel computational nature, which makes it attractive for such a problem. On the other hand, it has been found that maximizing the MI between transformed data and the desired target achieves the lowest probability of error [21]. The proposed method will be compared with other dimensionality reduction methods utilized in myoelectric control.

This paper is organized as follows: section II explains the PSO algorithm and the proposed feature selection algorithm. In section III, the experimental results are presented and analyzed, and finally a conclusion is given in section IV.

II. PARTICLE SWARM OPTIMIZATION AND FEATURE SELECTION

A. Particle Swarm Optimization

Particle swarm optimization, also commonly known as PSO, is a population based stochastic optimization technique developed by Eberhart and Kennedy in 1995 [16]. It mimics the behavior of a swarm of birds or a school of fish. If one of the particles discovers a good path to food then the rest of the swarm will be able to follow instantly even if they are far away in the swarm. Swarm behavior is modeled by particles in multidimensional space that have two characteristics: position and velocity. These particles wander around the hyperspace and remember the best position that they have discovered. A particle's position in the multi-dimensional problem space represents one solution for the problem. They exchange information about good positions to each other and adjust their own position and velocity with certain probabilities based on these good positions.

There are two versions of the PSO algorithm, these are: the canonical PSO and the binary PSO [6]. The canonical PSO was developed for continuous optimization problems. However, lots of

practical engineering problems are formulated as combinatorial optimization problems, which require a binary version of the PSO algorithm. The binary PSO model is based on a very simple modification of the real valued PSO, where a mapping from a given problem-domain to a set of bit strings is implemented. In such a problem every bit represents an attribute. A value of “1” means that the attribute is selected while a value of “0” means it not selected.

During iterations each particle adjusts its own trajectory in the space in order to move towards its best position and the global best according to the following equations:

$$v_{ij}(t+1) = w * v_{ij}(t) + c_1 * r_1 * (lbest_{ij} - x_{ij}) + c_2 * r_2 * (gbest_{ij} - x_{ij}) \quad \dots (1)$$

$$s(v_{ij}) = \frac{1}{1 + \exp(-v_{ij})} \quad \dots (2)$$

$$\text{if } p_{ij} < s(v_{ij}(t)) \text{ then } x_{ij}(t) = 1; \text{ else } x_{ij}(t) = 0 \quad \dots (3)$$

where

i is the particle index,

j is the current dimension under consideration.

x_{ij} is the current state of the bit on the j 'th *dimension* of individual i (position in canonical PSO).

v_{ij} is the velocity of the particle, or current probability of deciding 1.

t is the current time step, and $t-1$ is the previous time step.

w is the inertia weight,

r_1 and r_2 are two random numbers uniformly distributed in the range (0,1), c_1 and c_2 are cognitive and social parameters respectively, $lbest$ is the local best position, the one associated with the best fitness value the particle has achieved so far, and $gbest$ is the global best position, the one associated with the best fitness value found among all of the particles. p_{ij} is a vector of random numbers, drawn from a uniform distribution between 0.0 and 1.0. These formulas are iterated repeatedly over each dimension of each individual, testing every time to see if the current value of

x_{ij} results in a better evaluation than $lbest_{ij}$, which will be updated if it does. For more information about PSO, the reader can refer to [6].

B. The Proposed Feature Selection Algorithm

The BPSOMI algorithm combines the power of binary PSO and the concept of mutual information. To guide the search; a filter method implemented using mutual information is adopted. The binary PSO is used to implement a wrapper feature selection using a suitable classifier. When the binary PSO is applied to the feature selection problems, the number of the selected features must be equal to m , i.e., the desired number of features, while each particle is represented with an initial dimensionality equal to the total number of features. In order to explain how this task is accomplished, consider the example shown in Fig. 3, that is explained below:

Assuming the total number of features to be eight, the generated probabilities from Eq. 2 above are sorted in a descending manner. A random vector with the same dimensionality is generated. Then the sorted probabilities are compared, point by point comparison, with the randomly generated vector according to Eq. 3 above. The results from this task are then mapped to the initial probability vector to extract the indices of the selected features. The first member in the comparison result refers to the feature with the highest probability, given as 0.987 in the sorted probability vector. This in turn maps to the second location in the original vector and so on. In order to extract only three features from the resultant subset $\{2, 7, 8, 4\}$, then we only pick the first three indices $\{2, 7, 8\}$ and so on. If the number of selected features is not equal to m , then we simply multiply the randomly generated numbers by a factor ϵ , where $\epsilon < 1$, in an attempt to reduce the threshold compared with, thus generating more features. Hence, this approach forms a kind of constraint optimization and will guarantee that the PSO will produce the desired number of features.

The following parameters are used in the algorithm:

- n : number of features that constitute the original set.
- m : is the desired number of the features to be selected.
- np : number of particles through the feature space.
- SM_i : importance of the generated subset (explained later).
- $S_j = \{s_1, \dots, s_m\}$: a list that contains the selected feature subset for particle j . For example if $x_j = 1001011$ then the indices selected are $S_j = \{1, 4, 6, 7\}$ which basically represent the features selected by particle j .
- PL : list of the previously tested subsets

A flowchart of the general steps of the algorithm is given in Fig.4. The detailed description of the algorithm steps is listed below:

1. Initialization:

- Set the values of c_1 , c_2 , and w (w usually decrease from 0.9 to 0.4).
- Define the maximum number of iterations (MAXiter).
- Define the maximum velocity V_{\max} .

2. In the first iteration,

- Randomly assign subsets of m features to x_i .
- Randomly initialize the velocity matrix to values in the range $[0, 1]$.
- Estimate the Mean Square Error (MSE) of subset S_j .
- Assign the obtained result by particle j as its local best ($lbest$) and determine the global best particle ($gbest$).

3. For $j = 1$ to np ,

- Apply Eq. (1) mentioned earlier to compute new velocities.
- Clip the value of the velocities between $[-V_{\max}, V_{\max}]$.

- Determine the new values of the current solution of each particle by using Eq. 2 and 3.
 - Estimate the Mean Square Error (*MSE*) of the classification results of S_j .
 - Compare the current fitness (error) achieved by each particle at iteration t with the fitness of the solution stored in its local best memory $lbest$ and replace old one if the new solution achieved better fitness.
4. Working on the solutions stored in the local memory $lbest$, keep $m-p$ features of $lbest$, where $p < m$. The remaining p features are to be found using the following information theoretical approach.

- Select the remaining p features for each particle:

- For $mm = m - p + 1$ to m ,

- For $j = 1$ to np ,

- Given subset S_j which contains $(m-p)$ features, choose feature f_i that maximizes the **Selection Measure**, $SM_i^{S_j}$, as given below.

$$SM_i^{S_j} = a_1 * ICSx_i + a_2 * \frac{PD}{PD+ND} + a_3 * \left(1 - \frac{(PD+ND)}{\max(PD+ND)}\right) \quad \dots (4)$$

where a_1 , a_2 , and a_3 are constants and $ICSx_i$ is the local importance of feature f_i given the subset S_j , this is measured by the mutual information evaluation function (MIEF) proposed by Al-Ani et al [1]. The MIEF measure is defined as:

$$ICSx_i = I(C; f_i) \times \left[\frac{2}{1 + \exp(-\alpha D)} - 1 \right] \quad \dots (5)$$

where

$$D = \min_{f_j \in K} \left[\frac{H(f_i) - I(f_i; f_j)}{H(f_i)} \right] * \frac{1}{|S|} \sum_{f_j \in K} \exp \left[\beta \left(\frac{I(C; \{f_i, f_j\})}{I(C; f_i) + I(C; f_j)} \right)^\gamma \right] \quad \dots (6)$$

The parameters α, β , and γ are constants. $|S|$ is the cardinal of S that represents the selected feature subset. $I(C; f_i)$ is the mutual information between feature f_i and the class labels C . $I(C; \{f_i, f_j\})$ is mutual information between the two features f_i and f_j and the class labels. $H(f_i)$ is the entropy of f_i , $I(f_i, f_j)$ is the mutual information between the two features f_i and f_j . f_i represents a candidate feature, while f_j is a feature that is already in S_j . For more information the reader can refer to [1]. In addition we suggest the use of the positive and negative distribution factors (PD and ND respectively). PD is the positive distribution factor that is computed from the subsets that achieved an accuracy that is higher than the average accuracy of the whole subsets. ND is the negative distribution factor that is computed from the subsets that achieved an accuracy that is lower than the average accuracy of the whole subsets. This is shown in Fig.5 schematically with the light gray region being the region of elements achieving less error than the average error values and the dark gray being the region with elements achieving higher error rates than the average.

- Augment the selected feature f_i , to S_j , $S_j = S_j \cup \{f_i\}$.

5. Evaluate the subsets generated by modifying the particle's solution in the *lbest* memory using a chosen classification algorithm:

- For $j = 1$ to np ,
 - Estimate the Mean Square Error (MSE_j) of the classification results obtained by classifying the features of S_j .
 - Compare the current fitness achieved by the solution generated with the aid of equation (4) with the fitness of the solution stored in the *lbest* memory in its local best memory and replace old one if the new solution achieves better fitness.

- Sort the subsets according to their MSE . Update the minimum MSE (global fitness) if achieved by any particle, and store the corresponding subset of features.
- Update the list of the previously tested subsets. $PL = [PL ; S_j]$, where $(j=1:np)$.

6. If the number of iterations is less than the maximum number of iterations (MAXiter), then go to step 3.

The rationale behind Eq. 4 is to replace part of the solutions stored in the *lbest* memory according to two measures: the MIEF and the Estimation of features distribution. The MIEF measure will build on the achieved solutions through considering features that interact well with the already selected ones. Hence, it will enhance the chance of finding “good” solutions. The $PD/(PD+ND)$ factor indicates the degree to which the features contribute in forming good subsets, where PD is computed from the subsets with less fitness than the average fitness of PL . On the other hand the last term in Eq. (4) aims at favoring exploration, where this term will be close to 1 if the overall usage of f_i is very low.

III. EXPERIMENTAL RESULTS

Two sets of experiments are implemented. In the first experiment the performance of the proposed method is tested on different MES datasets and compared with the performance of feature projection techniques. In the second experiment, another set of MES datasets is used upon which the performance of the BPSOMI is compared with other evolutionary algorithms like Genetic Algorithm (GA). In order to make a good judgment on the performance of the proposed technique, we have also included the performance of the original binary PSO algorithm in the comparisons.

The goal behind both experiments is to first prove that an effective feature selection technique can be as powerful as the state of the art feature projection techniques, if not better. Secondly, we aim to prove that the proposed BPSOMI method can outperform other evolutionary techniques by means of convergence speed and solution optimality.

The testing is performed by adopting a three way data split scheme in which the datasets were divided into a training, validation, and testing sets. The features that minimize the difference between the training error and validation error are chosen as the members of the best solution. Then the generalization capability of the classifier is tested based upon the completely unseen testing data (unseen during training and validation). In this way one can ensure that the performance of the classifier will not be biased toward any of the training, or validation sets.

A. A Comparison with Feature Projection Techniques

The MES dataset utilized in this experiment was collected and used by Adrian et al [2]. Eight channels of surface MES records were collected from the right arm of thirty normally limbed subjects (twelve males and eighteen females). Each subject underwent four sessions, with one to two days separation between sessions. Each session consisted of six trials. Seven distinct limb motions were used, hand open (HO), hand close (HC), supination (S), pronation (P), wrist flexion (WF), wrist extension (WE), and rest state (R). The positions of the electrodes are shown in Fig.6. As in the original research data from only session four were used. Data from the first two trials were used as training set and data from the remaining four trials were divided equally into two trials for validation (trials 1 and 2) and two trials for testing (trials 3 and 4). The testing scheme is hence a bit different from the one used in [2] which utilized the first two trials for training and the other four trials for testing. The MES dataset utilized in this paper was already down-sampled and processed off-line in the original research [2], thus no further preprocessing was required.

Two sets of features were extracted from the dataset in order to test the performance of the proposed method with different feature extraction techniques. The first set of extracted features included a combination of the first four autoregressive (AR) coefficients and the root mean square value as the feature vector (total of 40 features = 8 channels \times 5 features/channel). In this paper, the combination of the four AR features with the root mean square feature, being a simple time domain (TD) feature, will be referred to as TDAR, which stands for the combination of **T**ime **D**omain **A**uto**R**egressive feature set. The analysis window size was 256 ms which were spaced 128 ms apart for training data and 32 ms apart for testing data. Data that were 256 ms before or after a change in limb motion were removed from the training set to avoid transitional data. The second feature set extracted included the mean of the square values of the wavelet transform coefficients using a Symmlet wavelet family with five levels of decomposition (total of 48 features = 8 channels \times 6 features/channel). This feature set was referred to as the **W**T, which stands for the **W**avelet **T**ransform based feature set.

It is worth mentioning here that the TDAR feature extraction code is the same as the one used in [2]. The desired number of features was set to be equal to only 10 features. Classification is performed using a linear discriminant analysis classifier (LDA). The advantage of this classifier is that it does not require iterative training, avoiding the potential for under- or over-training. The classification results averaged across thirty subjects using both the TDAR and the WT feature sets are shown in Fig.7. It should be mentioned here that the output of the MES pattern recognition system is usually smoothed using a majority vote post processing technique. It has been found that applying majority vote in MES classification problems represents a necessary step as it can achieve an enhancement in the MES classification accuracy of about 2% [3]. Another step usually utilized in MES recognition problems is to remove the transitional data between classes. This is due to the fact that the system is in an undetermined state between contractions. The results shown for both the validation and the testing sets were given first without a majority vote (referred to as

Initial), then with a majority vote (MV), followed by results with the transitional data between classes removed (NT), and finally with both majority vote and removal of the transitional data (MV+NT).

When analyzing the results, it was obvious that the hit rates obtained by both ULDA and the proposed BPSOMI algorithms highly outperform PCA. This is expected as the later does not take the relation between the features and the class label into consideration. On the other hand, ULDA projects the data into the direct that maximizes the ratio of the between scatter matrix to the within scatter matrix. One issue to be mentioned regarding ULDA is that the resulting dimensionality is limited to $C-1$ where C represents the number of problem classes (seven for this problem),. Although this might be an advantage since it highly reduces the number of projected features; but it could also serve as a limitation to this technique. In simple words, projecting the data into $C-1$ new feature may not be an optimal solution as we might lose important information for this particular dimension.

The BSPOMI, one the other hand, selects feature subsets by means of a hybrid technique that employs a mutual information based measure. This will enhance the chance of selecting subsets of maximum relevance to the problem, whose features carries the minimum amount of redundancies in the information content. As a result, the BPSOMI achieved the highest classification accuracies across different feature sets. This in turn proves that if the interaction property between the features in the selected subset is high, then this could result in a very powerful combination of features. In other words, it is important to search for features that complement each other.

In order to provide a rigorous validation or comparison with existing techniques for dimensionality reduction, the confusion matrix for all the subjects was also computed for the different feature sets. A plot of the diagonal values of the class-wise classification accuracy matrices with both the TDAR and the WT feature is presented in Fig. 8 including the validation

and testing sets results. The figure indicates that for almost all classes BPSOMI achieved better performance than ULDA and that both BPSOMI and ULDA clearly outperform PCA.

B. A Comparison with Other Evolutionary Techniques

In this experiment, the MES dataset used to test the proposed method was acquired from the University of New Brunswick in Canada [13]. The dataset consisted of ten motions associated with three degrees of freedom (DOF's) of the wrist, two different hand grips, and a rest state. In particular they were: forearm pronation, forearm supination, wrist flexion, wrist extension, radial deviation, ulnar deviation, key grip, chuck grip, hand open, and a rest state. Each session of the database consisted of two trials for each motion. Six subjects (abbreviated as AW, KS, LH, MW, SM, and WM) were prompted to complete medium force isometric contractions of 5 seconds duration followed by a brief rest period. Each record was 256 ms in duration (256 points sampled at 1024 Hz, pre-filtered between 10-500 Hz).

Features were extracted by means of the wavelet packet transform (WPT). The MES records were decomposed using a Symmlet family of wavelets with four levels of decomposition. This in turn resulted in generating 16 subspaces and those were chosen as features along each channel (total of 256 features = 16 channels \times 16 features/channel). To this end, the BPSOMI was tested against GA and the Binary PSO (abbreviated as BPSO). The parameters for BPSO were: population size =50, number of iterations =100, $V_{max}=4$, $c_1=c_2=2$, the same applies to BPSOMI in addition to $\alpha=1$, $\beta=1.65$, and $\gamma=3$. The parameters for the GA based feature selection were: population size =50, number of iterations =100, probability of crossover =0.8, and the probability of mutation =0.05. The reader is referred to [18] for more information about GA and the selection of its parameters. All of the methods were made to start from the same initial population. A three way data split was used here as well. The first two trails in the datasets were used for training; the third for validation; and the fourth for testing.

The results in Fig.9 are the average error rates achieved across the six subjects when selecting only nine features from the total of 256 features. The number of features was selected as nine as this proved to be enough to produce powerful results, while at the same time reducing the computational cost. These results are plotted without MV post-processing to indicate the real power of each method. It is clearly shown that the proposed BPSOMI outperformed both the GA and BPSO methods in terms of convergence speed and optimality. This can be realized from the fact that BPSOMI approached the optimum solution within less iteration than both GA and BPSO. The justification behind this case is that GA and BPSO may require large number of iterations to converge when dealing with a complex search space. The BPSOMI on the other hand, converged faster and achieved a lower error rate (3.37% compared to 4.55% achieved by GA and BPSO).

The rationale behind the enhanced performance of BPSOMI method is driven by the fact that the mutual information measure directs the swarm into the area of promising solutions. i.e., features that best interact together. The swarm would then employ the distribution factors along with its internal search capabilities to quickly reach the optimal solution.

IV. CONCLUSION

In this paper, a novel feature selection algorithm was presented based on a mixture of swarm intelligence and information theory. The concept of mutual information was utilized within the search procedure of the swarm to guide the particles to the vicinity of the promising solutions. The search procedure was also coupled with an estimation of distribution factors to enhance the search procedure. The method was applied to the myoelectric control problem as a dimensionality reduction tool. The BPSOMI method was tested against the state of the art feature projection techniques and it was able to outperform their performance. The proposed method was also tested against other evolutionary techniques and also proved its superiority.

ACKNOWLEDGMENT

The authors would like to acknowledge the support of the following researchers for providing us with the MES datasets:

- 1- **Dr. Adrian D. C. Chan** from Carleton University
- 2- **Levi Hargrove** and **Dr. Bernard Hudgins** from the University of New Brunswick.

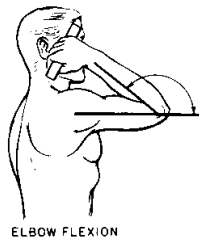
REFERENCES

- [1] A. Al-Ani, M. Deriche, and J. Chebil, "A new mutual information based measure for feature selection," *Intelligent Data Analysis*, 7 (2003), 43-57.
- [2] A. D. C. Chan and G. C. Green, "Myoelectric control development toolbox," in accepted to the 30th Conference of the Canadian Medical & Biological Engineering Society, Toronto, ON, 2007.
- [3] A. D. C. Chan and K. B. Englehart, "Continuous myoelectric control for powered prostheses using hidden Markov models," *IEEE Transactions on Biomedical Engineering*, 52 (2005), 121-124.
- [4] B. Hudgins, P. Parker, and R. N. Scott, "A new strategy for multifunction myoelectric control," *IEEE Transactions on Biomedical Engineering*, 40 (1993), 82-94.
- [5] C. M. Lighty, P. H. Chappelly, B. Hudgins, and K. Englehart, "Intelligent multifunction myoelectric control of hand prostheses," *Journal of Medical Engineering & Technology*, 26 (2002), 139- 146.
- [6] J. Kennedy, R. C. Eberhart, and Y. Shi, *Swarm intelligence*. London: Morgan Kaufmann Publishers, 2001.
- [7] J. Kimura, *Electrodiagnosis in diseases of nerve and muscle: principles and practice*, 3rd ed.: Oxford University Press, 2001.
- [8] J. U. Chu, I. Moon, and M. S. Mun, "A real-time EMG pattern recognition system based on linear-nonlinear feature projection for a multifunction myoelectric hand," *IEEE Transactions on Biomedical Engineering*, 53 (2006), 2232-2239.
- [9] J. U. Chu, I. Moon, and M. S. Mun, "A supervised feature projection for real-time multifunction myoelectric hand control," in *The 28th IEEE EMBS Annual International Conference*, New York City, USA, (2006), 2417-2420.
- [10] J. R. Cameron, *Medical physics*: John Wiley & Sons Inc, 1978.
- [11] K. Englehart, "Signal representation for classification of the transient myoelectric signal" Ph.D. Dissertation, University of New Brunswick, 1998.
- [12] K. Englehart, B. Hudgins, P. Parker, and M. Stevenson, "Time-frequency representation for classification of the transient myoelectric signal," *Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 1998, 2627-2630.
- [13] L. Hargrove, K. Englehart, and B. Hudgins, "A comparison of surface and intramuscular myoelectric signal classification," *IEEE Transactions on Biomedical Engineering*, 54 (2007), 847-853, 2007.

- [14] M. Asghari Oskoei and H. Hu, "Myoelectric control systems--A survey," *Biomedical Signal Processing and Control*, 2(2007), 275-294, 2007.
- [15] R.A. Rhoades and G. A. Tanner, *Medical physiology*, 2nd ed.: Lippincott Williams & Wilkins, 2003.
- [16] R. C. Eberhart and J. Kennedy, "A new optimizer using particle swarm theory," in *IEEE 6th Symposium on Micro Machine and Human Centre*, (1995), 39-43.
- [17] R. Jensen, "Combining rough and fuzzy sets for feature selection." Ph.D. Dissertation, University of Edinburgh, 2005.
- [18] R. L. Haupt and S. E. Haupt, *Practical genetic algorithms*, 2nd ed.: John Wiley & Sons, Inc, 2004.
- [19] R. Merletti, A. Bottin, C. Cescon, D. Farina, M. Gazzoni, S. Martina, L. Mesin, M. Pozzo, A. Rainoldi, and P. Enck, "Multichannel surface EMG for the non-invasive assessment of the anal sphincter muscle", *Digestion*, 69 (2004), 112-122.
- [20] R. Merletti and P. Parker, "Electromyography physiology, Engineering, and noninvasive applications," *IEEE Press Engineering in Medicine and Biology Society*, 2004.
- [21] R. M. Fano, *Transmission of information; a statistical theory of communications*: MIT Press, Cambridge, MA, 1961.

FIGURE CAPTIONS

- Fig.1** Temporal structures associated with different contractions using two sets of electrodes placed on the Biceps and Triceps muscles.
- Fig.2** Block diagram of the proposed system employing feature selection on the MES features extracted from multiple channels.
- Fig.3** BPSOMI formation of the reduced feature space.
- Fig.4** Flowchart of the proposed feature selection algorithm reflecting the general steps of the algorithm
- Fig.5** The feature distribution factors
- Fig.6** Electrodes placement on the right arm.
- Fig.7** Validation and testing classification results achieved with **BPSOMI**, **ULDA**, and **PCA** across different feature sets.
- Fig.8** Diagonal values of the confusion matrices achieved with **BPSOMI**, **ULDA**, and **PCA** across different feature sets.
- Fig.9** Average error rates for different optimization techniques.



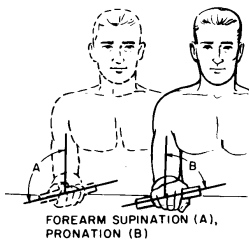
Elbow flexion



Elbow extension



(a) Elbow Flexion/Extension



Forearm pronation



Forearm supination



(b) Forearm Pronation/ Supination

Fig.1 Temporal structures associated with different contractions using two sets of electrodes placed on the Biceps and Triceps muscles (A modified version of the one presented in [5]).

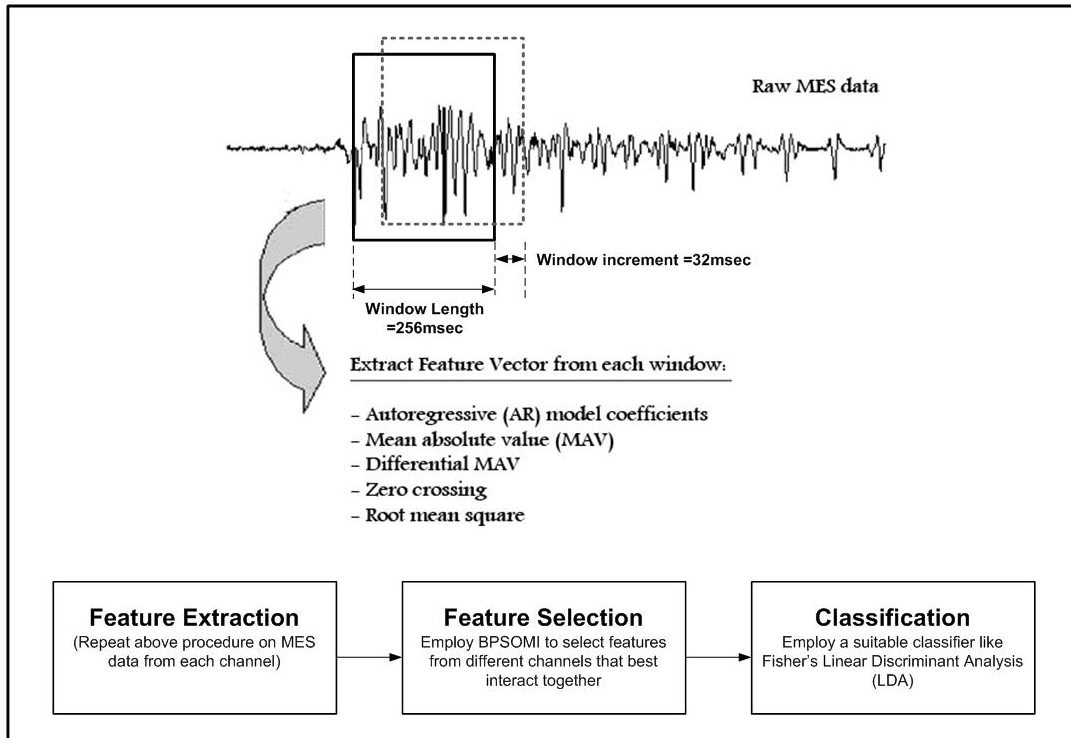


Fig.2 Block diagram of the proposed system employing feature selection on the MES features extracted from multiple channels.

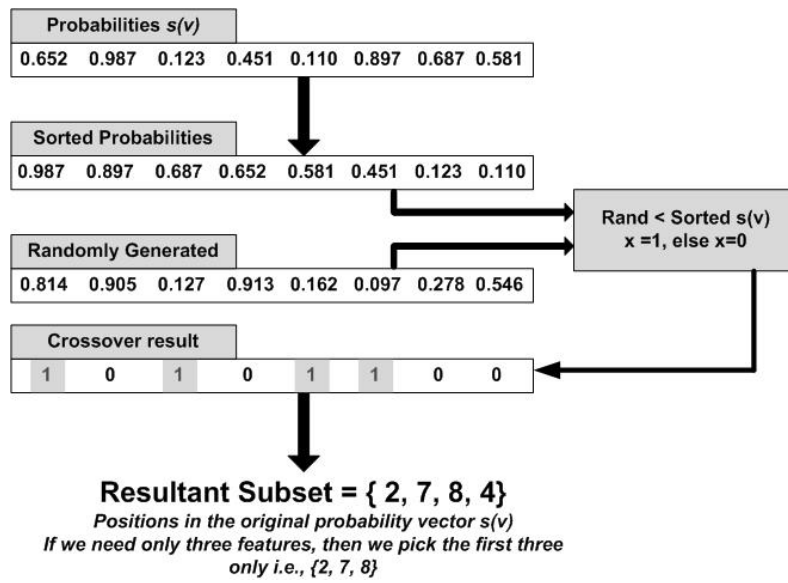


Fig.3 BPSOMI formation of the reduced feature space.

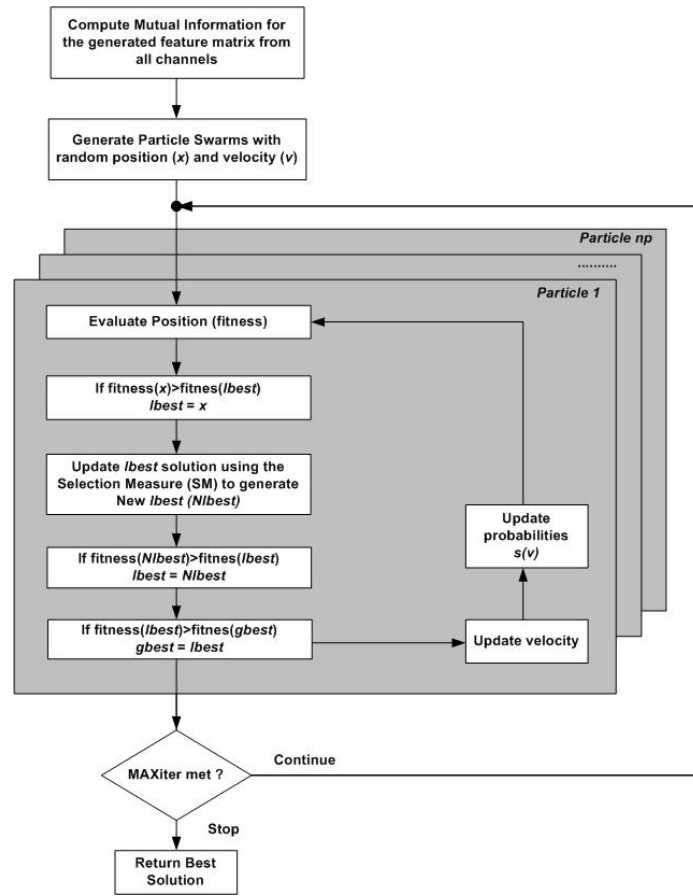


Fig.4 Flowchart of the proposed feature selection algorithm reflecting the general steps of the algorithm

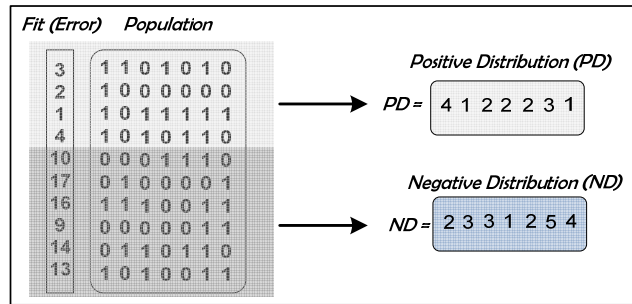


Fig.5 The feature distribution factors PD and ND.

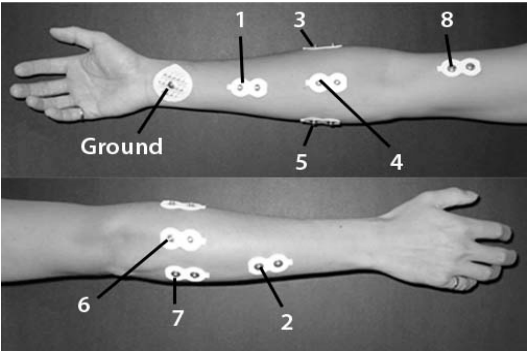


Fig.6 Electrodes placement on the right arm.

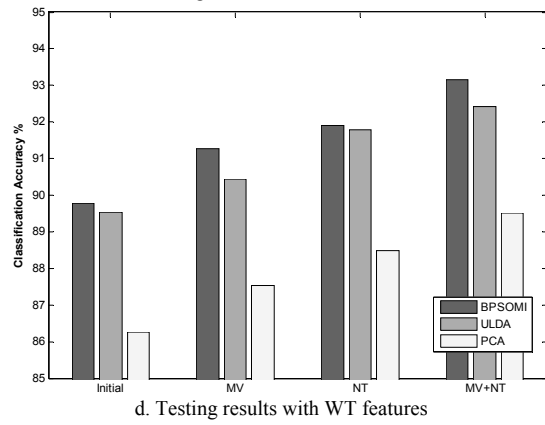
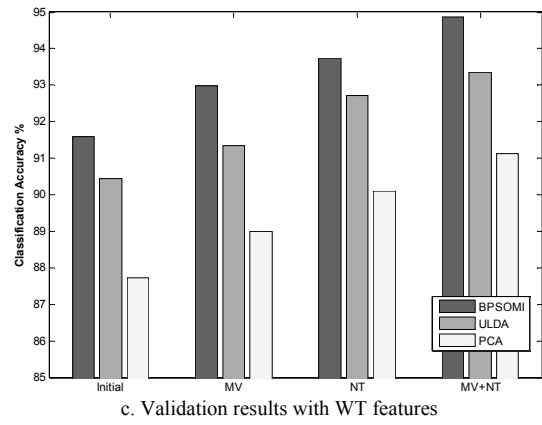
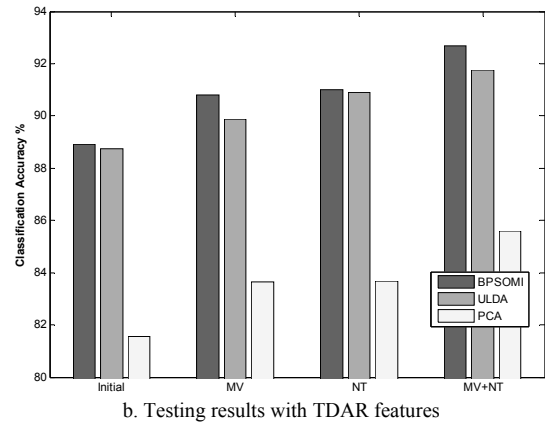
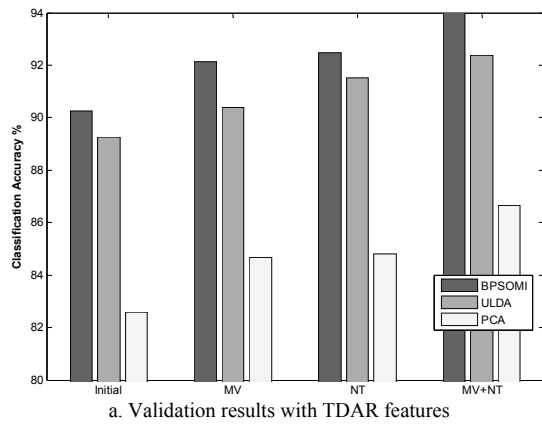
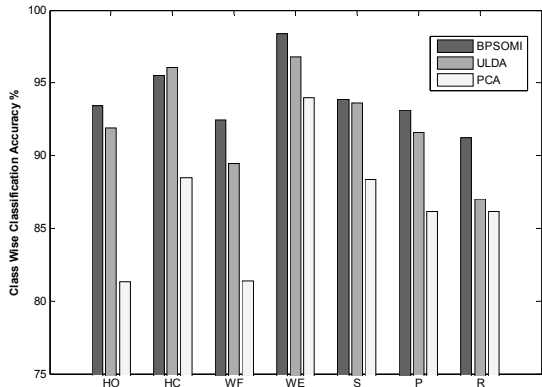
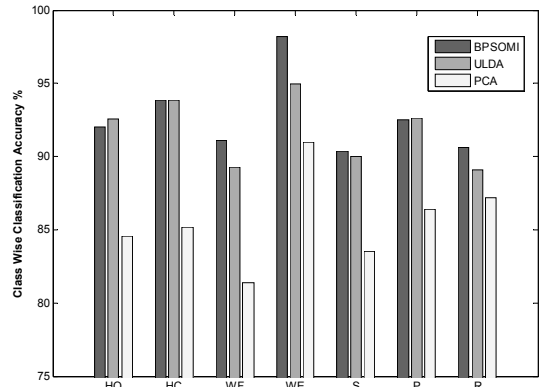


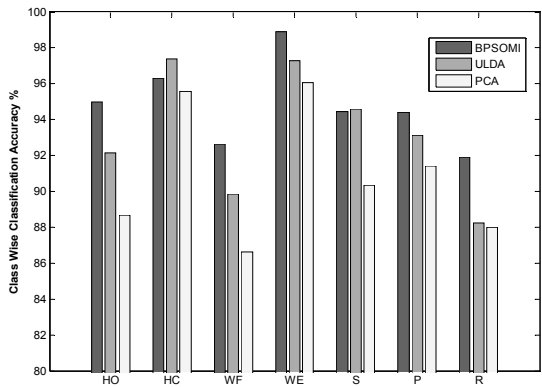
Fig.7 Validation and testing classification results achieved with **BPSOMI**, **ULDA**, and **PCA** across different feature sets.



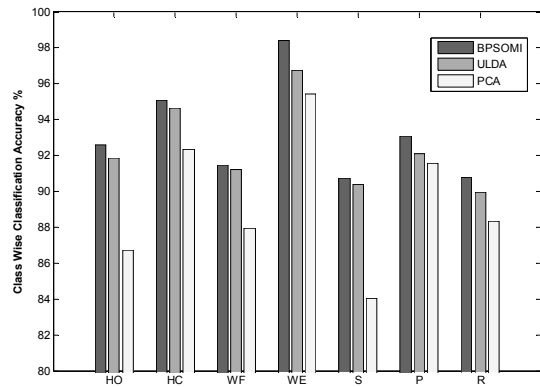
a. Validation results with TDAR features



b. Testing results with TDAR features



c. Validation results with WT features



d. Testing results with WT features

Fig.8 Diagonal values of the confusion matrices achieved with **BPSOMI**, **ULDA**, and **PCA** across different feature sets.

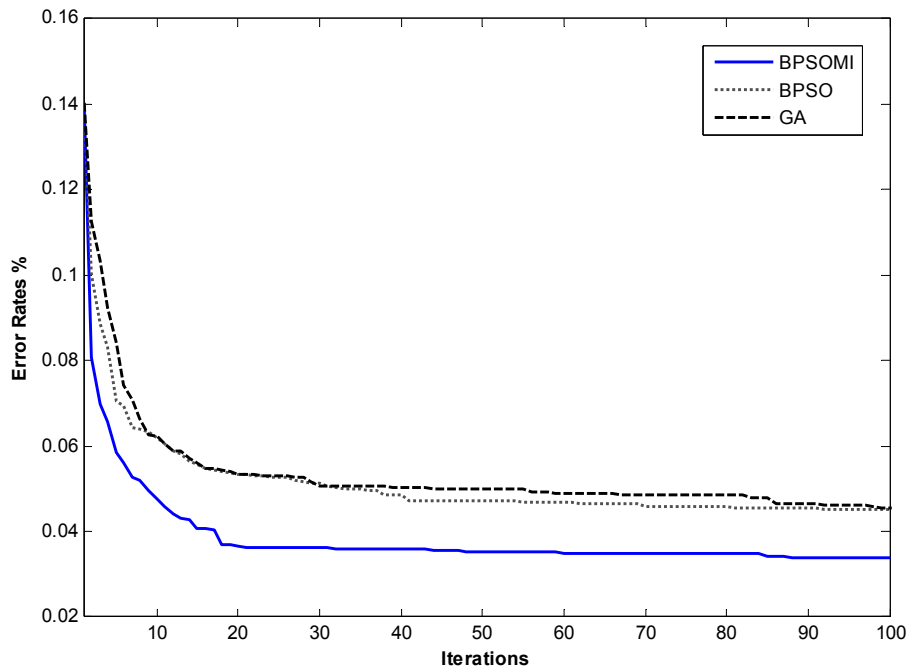


Fig.9 Average error rates for different optimization techniques