

© [2008] IEEE. Reprinted, with permission, from [Rami N. Khushaba, Akram AISukker, Ahmed Al-Ani *Intelligent Artificial Ants based Feature Extraction from Wavelet Packet Coefficients for Biomedical Signal Classification*, 2008 3rd International Symposium on Communications, Control and Signal Processing]. This material is posted here with permission of the IEEE. Such permission of the IEEE does not in any way imply IEEE endorsement of any of the University of Technology, Sydney's products or services. Internal or personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution must be obtained from the IEEE by writing to pubs-permissions@ieee.org. By choosing to view this document, you agree to all provisions of the copyright laws protecting it.

Intelligent Artificial Ants based Feature Extraction from Wavelet Packet Coefficients for Biomedical Signal Classification

Rami N. Khushaba, *Student Member, IEEE*, Akram AlSukker, Ahmed Al-Ani

Mechatronics and Intelligent Systems

University of Technology, Sydney

Broadway 2007, Sydney, Australia

Emails: [Rkhushab, AlSukker, Ahmed @eng.uts.edu.au]

Abstract— In this paper, a new feature extraction method utilizing ant colony optimization in the selection of wavelet packet transform (WPT) best basis is presented and adopted in classifying biomedical signals. The new algorithm, termed Intelligent Artificial Ants (IAA) searches the wavelet packet tree for subsets of features that best interact together thus producing high classification accuracies. While traversing the WPT tree, care is taken so that no redundancy in the information is selected by the Ants. The IAA method is a mixture of filter and wrapper approaches in feature subset selection. The pheromone that the ants lay down is updated by means of an estimation of the information contents of a single feature or feature subset. The significance of the subsets selected by the ants is measured using linear discriminant analysis (LDA) classifier. The IAA method is tested on one of the most important biosignal driven applications, which is the Brain Computer Interface (BCI) problem with 56 EEG channels. Practical results indicate the significance of the proposed method achieving a maximum accuracy of 83%.

Keywords- *Ant colony optimization, wavelet packet transform, features extraction, brain computer interface*

I. INTRODUCTION

It has long been recognized that important features of biomedical signals exist in both the time and frequency domains [1]. This fact made the time-frequency (TF) analysis very important when dealing with such signals. The TF analysis is a process of search and investigation of intermediate representations that combine the information from both the time and frequency domains. One of the best TF analysis tools is the wavelet transform (WT) [2], and its generalization, known as wavelet packet transform (WPT) [3]. The wavelet theory in general has unified and extended ideas from several domains, including subband filtering, approximation theory, signal and image processing, and non-parametric estimation. It provides a very general technique that can be applied to many tasks in signal processing and therefore it has numerous potential applications.

The most fascinating area of signal processing with practical applications is pattern recognition [4]. Pattern recognition aims to classify data (or patterns) based on either a priori knowledge or on statistical information extracted from the patterns. The patterns to be classified are usually groups of measurements or observations, defining points in an appropriate multidimensional space.

Fundamental to the success of any pattern recognition system is the extraction of a set of informative features that best describe the content of the input signals. Although the literature includes a lot of research on the extraction of features from biomedical signals, no much effort has been made towards establishing a qualitative comparison of their quality.

There are two categories of classification methods employing WPT in features extraction [5]. The first uses abstract aggregates of the original wavelet packet features such as: energy, distance, or clusters. The second category is based on using the decomposition coefficients to form a feature space by merging specific nodes of the WPT tree and splitting others in order to produce a WPT tree that represent the best reflection of the properties of the signal. The common methods known, in the second category, are the joint best basis (JBB) [6], the local discriminant basis (LDB) [7], and fuzzy wavelet packets based features extraction method (denoted here as FWP) that was developed in [5] and proved to outperform both the JBB and the LDB methods in classification of biomedical signals. The FWP uses Fuzzy c-means (FCM) clustering [8] to determine the optimal wavelet packet decompositions and ranks the features based on their memberships in the classes, thus forming a kind of feature selection method. Although the FWP proved very successful on certain kinds of biomedical signals, but using such method based on features ability to separate different classes alone does not perform well on other biomedical signals, like the Electroencephalogram (EEG) and also the Electromyogram (EMG), due to the large variance associated with those signals.

In order for any feature selection techniques to be successful, such a method should be capable of measuring the following quantities:

1. **Relevance**: The selected subset should be of maximum relevance to the problem.
2. **Redundancy**: The selected subset should contain features that are of minimum redundancy among themselves.
3. **Interaction**: This is the most important property as certain features might not function well when used alone, but when considered with others they might prove very successful.

The only point against the JBB, LDB, and FWP methods is that these methods rank the features according to the relevance of individual features only. In other words, these methods do not account for how the selected features complement each other. When considering the effect of interaction, the extracted feature subset can produce higher classification accuracies, as the optimal subset usually consists of features that best interact among themselves.

In this paper, a novel feature extraction method utilizing the WPT and Ant Colony Optimization (ACO) technique is presented. The ACO is adopted to search for the subset of features that can best interact together. The set of ants, we termed Intelligent Artificial Ants (IAA), estimates the importance of feature subsets by means of mutual information measure (MI). It also employs LDA in order to measure the classification accuracy of the selected subset. Thus, it forms a mixture of filter and wrapper approaches to feature selection.

The paper is organized as follows: Section II gives an introduction to the ACO technique and explains the modification applied to it. In section III experiments and results are presented. Finally the conclusion is given in section IV.

II. BACKGROUND

A. Ant Colony Optimization

Swarm intelligence is an innovative computational way to solving hard problems. This discipline is inspired by the behavior of social insects, such as fish schools, bird flocks and colonies of ants, termites, bees and wasps. In general, this is done by mimicking the behavior of the biological creatures within their swarms and colonies.

Ant colony optimization (ACO) is one of the swarm based techniques that takes inspiration from the foraging behavior of some ant species. Initially, real ants wander randomly to find food [9]. These ants deposit pheromone on the ground in order to mark some favorable path that should be followed by other members of the colony. If other ants find such a path, they are likely to follow the trail with some pheromone and deposit more pheromone if they eventually find food, thus forming a kind of indirect communication called stigmergy [10]. In ACO, a number of artificial ants build solutions to the considered optimization problem at hand and exchange information on the quality of these solutions via a communication scheme that is reminiscent of the one adopted by real ants.

The ACO is a probabilistic technique for solving computational hard problems which can be reduced to finding an optimal path problem [11]. The ACO was initially applied in the optimization of the traveling salesman problem (TSP) because this problem closely resembles finding the shortest path to a food source [12]. In such a problem, a set of cities is given and the distance between each of them is known. The goal is to find the shortest tour that allows each city to be visited once and only once. A number of paths are generated in a Monte Carlo fashion on the basis of a probabilistic model whose parameters are called artificial pheromone—or more

simply pheromone. In the ACO metaphor, these paths are said to be constructed by artificial ants walking on the graph that encodes the problem in which each vertex represents a city and each edge represents a connection between two cities.

Initial attempts for building an ACO algorithm were not very satisfying until the algorithm was coupled with a local optimizer [13]. One problem is premature convergence to a less than optimal solution because too much virtual pheromone was laid quickly. To avoid this *stagnation*, pheromone evaporation is implemented. In other words, the pheromone associated with a solution disappears after a period of time.

In the construction of a solution, ants select the following city to be visited through a stochastic mechanism. When ant k is in city i and has so far constructed the partial solution s_p , the probability of going to city j is given by:

$$p_{ij}^k = \begin{cases} \frac{\tau_{ij}^\sigma \cdot \eta_{ij}^\nu}{\sum_{c_{ij} \in N(s^p)} \tau_{il}^\sigma \cdot \eta_{il}^\nu} & \text{if } c_{ij} \in N(s^p) \\ 0 & \text{otherwise} \end{cases} \quad \dots (1)$$

where $N(s^p)$ represent the set of feasible nodes. σ and ν are constants to control the relative importance of the pheromone versus the heuristic information η_{ij} , which is given by:

$$\eta_{ij} = \frac{1}{d_{ij}} \quad \dots (2)$$

Where d_{ij} is the distance between city i and city j .

During each of the iterations the pheromone values are updated by all the m ants that have built solutions in the iteration itself. The pheromone τ_{ij} , associated with the edge joining cities i and j is updated as follows:

$$\tau_{ij} \leftarrow (1 - \rho) \cdot \tau_{ij} + \sum_{k=1}^m \Delta \tau_{ij}^k \quad \dots (3)$$

Where ρ is the evaporation rate, m is the number of ants, and $\Delta \tau_{ij}^k$ is the quantity of pheromone laid on edge (i, j) by ant k , where:

$$\Delta \tau_{ij}^k = \begin{cases} Q/L_k & \text{if ant } k \text{ used edge } (i, j) \text{ in its tour,} \\ 0 & \text{otherwise} \end{cases} \quad \dots (4)$$

Where Q is a constant and L_k is the length of the tour constructed by ant k .

B. Wavelet Packet Based Features Extraction

The uneven distribution of signal energy in the frequency domain has made signal decomposition an important practical problem. The basic concept here is to divide the signal spectrum into its sub-spectra or subbands, and then to treat those sub-spectra individually for the purpose at hand.

With multiresolution analysis, the fast wavelet transform (FWT) leads to a dyadic pyramidal implementation using filter banks and the corresponding Mallat algorithm [14]. FWT develops the two channel filter banks through which the signal is split into two subspaces, V_j and W_j , which are orthonormally

complementary to each other. With V_j being the space that includes the low frequency information about the original signal and W_j includes the high frequency information (details may be found in [3]). One keeps repeating the decomposition of the low frequency to a specified number of levels. The Wavelet packets (WP) Compared to DWT, WP not only decomposes the approximation coefficients, but also the detail coefficients as shown in Fig.1. It was introduced by Coifmann, Meyer and Wicker Hauser [6].

It works by generalizing the link between multiresolution approximation and wavelet bases. A signal space V_j of a multiresolution approximation is decomposed in a lower resolution space V_{j+1} plus a detail space W_{j+1} . The decomposition is achieved by dividing the orthogonal basis $\{\phi_j(t - 2^j n)\}_{n \in \mathbb{Z}}$ of V_j into two new orthogonal bases $\{\phi_{j+1}(t - 2^{j+1} n)\}_{n \in \mathbb{Z}}$ of V_{j+1} and $\{\psi_{j+1}(t - 2^{j+1} n)\}_{n \in \mathbb{Z}}$ of W_{j+1} , where $\phi(t)$ and $\psi(t)$ are scaling and wavelet functions respectively.

The decomposition for WP can be implemented by using a pair of Quadrature Mirror Filter (QMF) bank that divides the frequency band into equal halves. Due to the decomposition of the approximation space (low frequency band) as well as the detail space (high frequency band), the frequency division of the signal take place on both the lower and higher sides. This recursive splitting of vector space is represented by admissible WP tree. Features are usually computed by accumulating the square of the expansion coefficients of the signal at each position in the binary table followed by the normalization of the total energy of the signal belonging to class I according to:

$$\Gamma_i[X] \triangleq E[(X_i)^2] / \sum_{i=1}^L E[(X_i)^2] \text{ or} \quad \dots (5)$$

$$\Gamma_i[X] \triangleq Var[(X_i)^2] / \sum_{i=1}^L Var[(X_i)^2]$$

where L is the dimension of X .

The JBB and LDB rank the features according to an information measure for which an entropy measure is adopted in JBB and one of the following measures in LDB: Relative Entropy, Symmetric Relative Entropy, Euclidean Distance, and l^2 norm. In contrast, the FWP employs the concept of fuzzy memberships as an information measure. All of the JBB, LDB, and FWP rank the features according to their power or relevance to the problem only. Those methods do not account for extracting pieces of information that complement each other.

To this end, we propose to employ an ant based technique for features extraction from the WPT decomposition tree. If we simply consider the set of features to be extracted as a set of nodes through which the ants will travel, then an important issue will be how the ant will choose the next node. This in turn tends to be a process of feature selection that is more complicated than that of the TSP. This is due to the fact that choosing the next feature to be added to the already selected subset cannot be made based on the distance between features only as it is the case in the TSP problem.

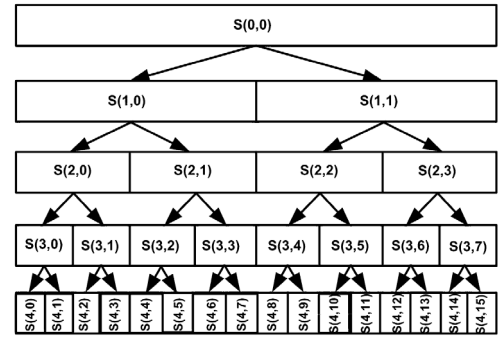


Fig.1 Decomposition trees for the FWT and WPT.

In TSP problem, when adding one more city, the change in the objective function is affected only by the distance between last two cities. In contrast to TSP, adding a feature to an existing subset of features can have an impact on the overall performance. A relevant feature will produce a better subset, and hence improve the performance, while an irrelevant feature may degrade the performance of the original subset. When adding a feature to the current feature subset the local performance measure should take into account the relationship with all previously selected features and not only the last one.

The most applicable local measure that was implemented in ACO in the literature employs the concept of Mutual Information (MI) [15]. For this purpose we adopted Al-Ani et al [16] approach known as the enhanced mutual information feature selection method (MIEF) that was coupled with ant colony optimization in [15] and modified that approach in such a way that when selecting certain father nodes this will prevent the ants from selecting the children nodes associated with that father node. In the proposed IAA method, the selection of next node (feature) is based on the following selection equation:

$$SM_i^{S_j} = \begin{cases} \frac{(\tau_i)^\eta (LI_i^{S_j})^K}{\sum_{g \in S_j} (\tau_g)^\eta (LI_g^{S_j})^K} & \text{if } i \in S_j \\ 0 & \text{otherwise} \end{cases} \quad \dots (6)$$

where $LI_i^{S_j}$ is the local importance of feature f_i given the subset S_j . The parameters η and K control the effect of trail intensity and local feature importance respectively.

$$LI_i^{S_j} = I(C; f_i) \times \left[\frac{2}{1 + \exp(-\alpha D_i^{S_j})} - 1 \right] \quad \dots (7)$$

where

$$D_i^{S_j} = \min_{f_s \in S_j} \left[\frac{H(f_i) - I(f_i, f_s)}{H(f_i)} \right] \times \frac{1}{|S_j|} \sum_{f_s \in S_j} \left[\beta \left(\frac{I(C; \{f_i, f_s\})}{I(C; f_i) + I(C; f_s)} \right)^\gamma \right] \quad \dots (8)$$

The parameters α , β , and γ are constants, $H(f_i)$ is the entropy of f_i , $I(f_i; f_s)$ is the mutual information between f_i and f_s , $I(C; f_i)$ is the mutual information between the class labels and f_i , and $|S_j|$

is the cardinal of S_j . The MIEF measure is employed in selecting the next feature according to (7) and (8) above. The first term in (8) reflects the degree of dependency between the next node to visit and the already visited nodes (the already selected features). When the next node to be selected is highly dependent on any of the already selected features then $D \rightarrow 0$ and $LI \rightarrow 0$. This will prevent the ants from selecting such a node as the next one to be visited. If the new node is independent or partially dependant on any of the features in the already chosen subset (chosen cities) then the importance of such a node will be calculated by the second term. For a detailed explanation of the method, the reader is referred to [16].

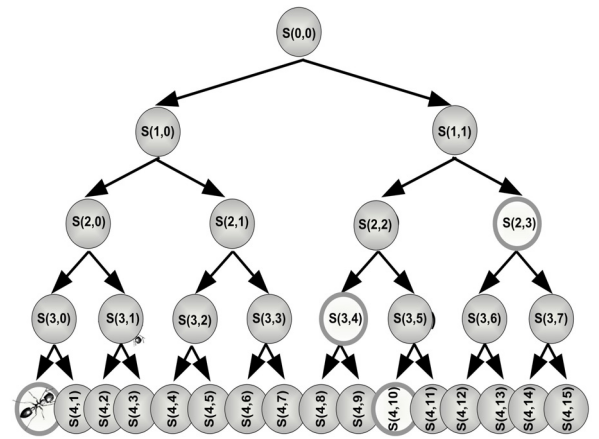
As an example, consider the graph in Fig.2, in which decomposition to level 4 is presented, assuming the first level to be 0. The example assumes that the ant at node $S(4,0)$ is faced with the fact that it has three nodes ($S(2,3)$, $S(3,4)$, and $S(4,10)$) to choose one from. This is implemented using the MIEF method according to (7) and (8). In the real process, each of the ants will be faced with the decision to choose the next nodes available given that these nodes are not the ascendant or descendant of any of the already chosen nodes by the ant. As an example, if the ant at node $S(4,0)$ chose the next node to be $S(3,4)$ then the following nodes given by [$S(1,1)$, $S(2,2)$, $S(4,8)$, and $S(4,9)$] will be excluded from the list of remaining nodes. The approach of removing the ascendant or descendant of the chosen nodes was implemented in order to present a fair comparison with the FWP method in literature in which the same approach is applied, but the features are selected only according to their memberships.

As a swarm technique, a population of ants is utilized on the tree each wandering randomly on the tree to make its final decision. A summary of the full algorithm that we modified is given below, refer to [15] for more info:

- 1- Decompose each of the original training signals to the maximum level J . Extract features from each of the nodes in the WPT tree according to (5). This step will produce N features (depending on the decomposition levels used).
- 2- Specify the desired number of extracted features (DNF). Usually $DNF \ll N$ thus selecting only certain features that best interact together.
- 3- Randomly choose a certain set of ants to wander along the tree to choose the DNF features according to the modified equation (6-8).
- 4- Measure the selected subsets performance employing linear discriminant analysis (LDA) classifier.
- 5- Update the pheromone trails.
- 6- Iterate the algorithm searching for features.

III. APPLICATION AND EXPERIMENTAL RESULTS

The new method was applied in one of the most important biosignals driven applications that is Brain Computer Interface (BCI), utilizing the Electroencephalogram (EEG) signal.



An ant at node $S(4,0)$ employs the MIEF measure to choose on the of the nodes $S(2,3)$, $S(3,4)$, and $S(4,10)$.

Fig.2 A set of ants traversing the wavelet packet tree.

The EEG is a recording of electrical activity originating from the brain. It is recorded on the surface of the scalp using electrodes, thus the signal is collected non-invasively. The process of communicating with a computer through EEG signals is referred to as BCI. The BCI is one of the most challenging research areas due to the complexity of the problem, specifically the complexity of information extraction task. Another challenge with such a problem is how to classify the time-varying EEG signals as accurately as possible. This is due to the high variability of EEG recordings that makes it a difficult task to classify different EEG signals accurately and necessitates adaptive learning to boost up the performance of existing BCIs.

There are four main EEG frequency rhythms. Each rhythm is defined by its frequency, although there is no exact agreement on these frequencies. These rhythms can be described as follows:

- **Delta (δ):** the frequency between 0-4 Hz.
- **Theta (θ):** the frequency between 4-8 Hz, with amplitude of less than $30\mu V$.
- **Alpha (α):** the posterior dominant rhythm (PDR), that first discovered by Berger in 1929. Alpha has a frequency between 8-12 Hz, and has its maximum amplitude in occipital regions.
- **Beta (β):** the frequency greater than 12 Hz, with amplitude of less than $20\mu V$ and commonly found in frontal and central lobe [17].

Through the literature, different features were extracted from EEG signals; some of these methods are: Power estimate [18, 19], Auto Regressive model (AR) [20, 21], Principal Component Analysis (PCA) [22], wavelet [19, 23], and Adaptive Gaussian representation. Since EEG is a time-varying and space-varying non-stationary signal, this made both the WT and WPT excellent candidates for features extraction from such data. The WT was found to provide a good way to visualize and decompose EEG signals into measurable components events [24]. Frequency component

selection has been considered in different studies. Pregonzer et al [25] analyzed the relevance of different spectral components for only two channels. In [26], Al-Ani et al used the genetic algorithm (GA) to select the best frequency bands from two different wavelet tree that would represent two channels (same channels as those presented in [25]). The best frequency bands from which the most discriminant features were extracted was found to be 4–8, 8–16 and 16–24 Hz [26].

The EEG dataset used in this paper was taken from the Department of Medical Informatics in the University of Technology, Graz in Austria. The EEG signals were recorded for three right handed females with 56 Ag/AgCl electrodes, with reference electrode on the right ear. The subjects were individually placed in an armchair and asked to imagine right or left finger movements according to stimuli on screen. A total of 8 seconds of data were recorded at 128 Hz sampling rate, 2 seconds before the stimuli and 6 after it. A total of 406 trials were used, 208 for the left movement and 198 for right movement. More details on experiment set-up can be found in [27]. This dataset was also used in [26], and the obtained results indicated that there is a small variation between the classification accuracy of feature subsets with the desired number of features ranging between 3 and 9. As using a small number of features is more desirable, a subset of 3 features that represent the energy values of the frequency bands 4 – 8, 8 – 16 and 16 – 24 Hz was chosen. The same approach of extracting only 3 features was followed in this paper, but the difference is that we employ IAA to search for features from each channel, while in [26], features were extracted from a mixture of two channels.

A number of experiments were carried out. During the first experiment the IAA method was utilized along each channel to extract the best subset of 3 features, for a total of 56 channels this in turn produced 168 features. For the same channels, 3 features were also extracted using the FWP method mentioned earlier. Both the JBB and LDB methods were not included in the comparison since the FWP proved to outperform those methods in classification of biomedical signals. A bar plot of the classification accuracy of the features extracted from each of the 56 channels is shown in Fig.3 employing the IAA and the FWP methods. It is obvious from these results that the IAA method extracted the features that can best discriminate between the given classes while for the FWP it extracted the features according only to their memberships in the classes (an approach which does not guarantee to provide optimum results).

In the second experiment, it was decided to reduce the size of the extracted feature sets that consist of 168 features (obtained from all the 56 channels), as it is not practical to consider using all of those features for classification. In order to produce a fair comparison that is unbiased toward the estimation of any of the methods; it was decided to employ a GA-based features selection approach on both methods. The adoption of GA is supported by the good results that it managed to achieve in the BCI problem [26]. To further illustrate this issue, the 168 features extracted by the FWP method were ranked according to their memberships in the

classes and subsets of varying numbers of features were formed. Fig.4 represents a comparison between those subsets and subsets of same number of features selected using the GA approach. The result shows a significant enhancement of the classification accuracy of the FWP results than the original approach by at least a factor of 10%.

GA as a dimensionality reduction method was applied to both the IAA and FWP and the performance was compared, as shown in Fig.5. The number of selected features ranged from 3 to 99 features to measure the accuracy along subsets of various sizes. It is clear that even with the enhancement gained by the using GA, IAA continued to outperform FWP. This proves that the performance of the classification task is mainly dependant on the quality of the extracted features. The experiments also prove the functionality of the IAA method in extracting informative subset of features.

As a final step, we compare the results obtained with the approach presented by Yang et al [28], in which they utilized the FWP method to design a subject dependant feature extraction for a BCI system. A maximum accuracy of 76% was achieved using Yang method, while in our experiment the FWP achieved a maximum of 81.73%, and the IAA achieved a maximum of 83.4%.

IV. CONCLUSION

In this paper, a new features extraction method was presented by utilizing a mixture of wavelet packet transform, ant colony optimization, and mutual information. The new method was compared with a recently presented fuzzy wavelet packet features extraction method in the problem of brain computer interface. Various experiments were made while applying GA as a dimensionality reduction to the features extracted from both methods. The results proved that the IAA method outperformed the FWP. The maximum accuracy achieved by the IAA method was 83.4%, compared to 81.73% achieved by the FWP method.. It is worth mentioning that the FWP needed 60 features to achieve 81.73%, while the same accuracy was achieved by the IAA method with only 39 features, i.e., a reduction of 35%. Further experiments are required to support our findings in different and larger datasets and applications.

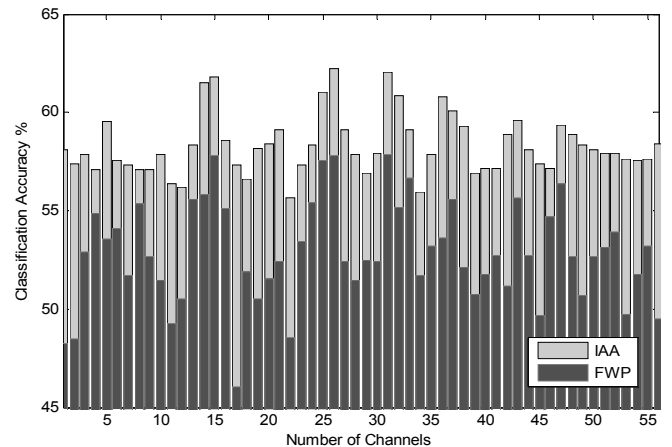


Fig.3 A comparison of the channel accuracy of the features extracted with both the IAA and the FWP methods.

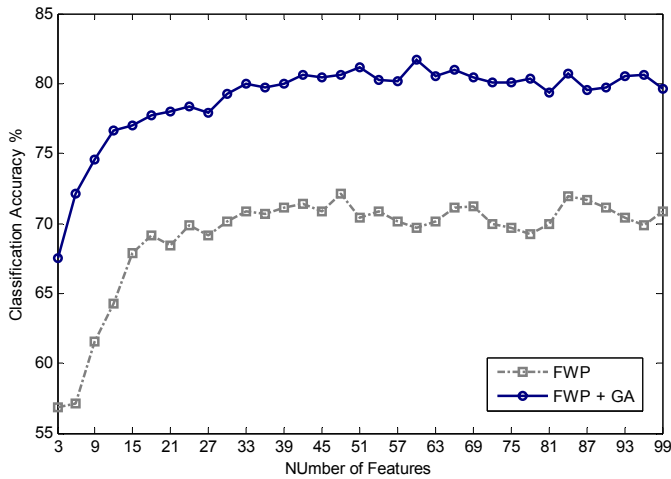


Fig.4 A comparison of the classification results of the 168 features extracted using FWP with and without GA.

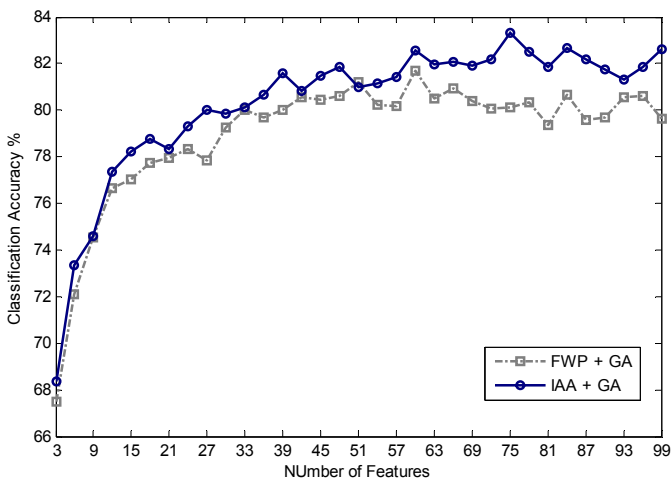


Fig.5 A comparison of the classification results of the 168 features extracted with the IAA and FWP with GA.

REFERENCES

- [1] K. Englehart, P. Parker, and B. Hudgins, "Wavelet methods in biomedical signal processing," in *Handbook of neuroprosthetic methods, The Biomedical Engineering Series*, M. Neuman, Ed.: CRC Press, 2003.
- [2] J. C. Goswami and A. K. Chan, *Fundamentals of wavelets : theory, algorithms, and applications*. New York: Wiley, 1999.
- [3] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 11, pp. 674-693, 1989.
- [4] Y. Y. Tang, L. H. Yang, J. Liu, and H. Ma, "Wavelet theory and its application to pattern recognition," in *Series in Machine Perception and Artificial Intelligence*, H. Bunke and P. S. P. Wang, Eds.: World Scientific Publishing, 2000.
- [5] L. Deqiang, W. Pedrycz, and N. J. Pizzi, "Fuzzy wavelet packet based feature extraction method and its application to biomedical signal classification," *IEEE Transactions on Biomedical Engineering*, vol. 52, pp. 1132-1139, 2005.
- [6] R. R. Coifman, Y. Meyer, and M. V. Wickerhauser, "Wavelet analysis and signal processing," in *Wavelets and Their Applications*, M. B. Ruskai, Ed. Boston: Jones and Bartlett, 1992.

- [7] N. Saito, "Local feature extraction and its applications using a library of bases," in *Department of Mathematics*, vol. PhD Thesis: Yale University, 1994.
- [8] T. J. Ross, *Fuzzy logic with engineering applications*, 2nd ed. Hoboken, NJ: Wiley, 2004.
- [9] M. Dorigo, M. Birattari, and T. Stutzle, "Ant colony optimization," *IEEE Computational Intelligence Magazine*, vol. 1, pp. 28-39, 2006.
- [10] D. Merkle, M. Middendorf, and H. Schmeck, "Pheromone evaluation in Ant Colony Optimization," presented at 26th Annual Conference of the IEEE Industrial Electronics Society, IECON, 2000.
- [11] A. Abraham, C. Grosan, and V. Ramos, "Swarm intelligence in data mining," in *Studies in Computational Intelligence*, vol. 34, J. Kacprzyk, Ed.: Springer, 2006.
- [12] M. Dorigo and L. M. Gambardella, "Ant colony system: a cooperative learning approach to the traveling salesman problem," *IEEE Transactions on Evolutionary Computation*, vol. 1, pp. 53-66, 1997.
- [13] R. L. Haupt and S. E. Haupt, *Practical genetic algorithms*, 2nd ed: John Wiley & Sons, Inc., 2004.
- [14] S. G. Mallat, *A wavelet tour of signal processing*, 2nd ed. San Diego, Calif.; London: Academic, 1999.
- [15] A. Al-Ani, "Feature subset selection using ant colony optimization," *International Journal of Computational Intelligence*, vol. 2, pp. 53 - 58, 2005.
- [16] A. Al-Ani, M. Deriche, and J. Chebil, "A new mutual information based measure for feature selection," *Intelligent Data Analysis*, vol. 7, pp. 43-57, 2003.
- [17] W. T. Blume, M. Kaibara, and G. B. Young, *Atlas of adult electroencephalography*, 2nd ed. Philadelphia: Lippincott Williams & Wilkins, 2002.
- [18] G. Pfurtscheller, C. Neuper, and D. Flotzinger, "EEG-based discrimination between imagination of right and left hand movement," *Electroencephalography and Clinical Neurophysiology*, vol. 103, pp. 642-651, 1997.
- [19] D. Garrett, D. A. Peterson, C. W. Anderson, and M. H. Thaut, "Comparison of linear, nonlinear, and feature selection methods for EEG signal classification," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 11, pp. 141-144, 2003.
- [20] T. N. Lal, M. Schroder, T. Hinterberger, J. Weston, M. Bogdan, N. Birbaumer, and B. Scholkopf, "Support vector channel selection in BCI," *IEEE Transactions on Biomedical Engineering*, vol. 51, pp. 1003-1010, 2004.
- [21] G. S. Dharwarkar and O. Basir, "Enhancing Temporal Classification of AAR Parameters in EEG single-trial analysis for Brain-Computer Interfacing," presented at 27th Annual International Conference of the Engineering in Medicine and Biology Society, pp - 5358-5361, 2005.
- [22] L. Yong, G. Xiaorong, L. Hesheng, and G. Shangkai, "Classification of single-trial electroencephalogram during finger movement," *IEEE Transactions on Biomedical Engineering*, vol. 51, pp. 1019-1025, 2004.
- [23] V. Bostanov, "BCI competition 2003-data sets Ib and IIb: feature extraction from event-related brain potentials with the continuous wavelet transform and the t-value scalogram," *IEEE Transactions on Biomedical Engineering*, vol. 51, pp. 1057-1061, 2004.
- [24] V. J. Samar, A. Bopardikar, R. Rao, and K. Swartz, "Wavelet analysis of neuroelectric waveforms," *Brain and Language*, vol. 66, pp. 7-60(54), 1999.
- [25] M. Pregenzer and G. Pfurtscheller, "Frequency component selection for an EEG-based brain to computer interface," *IEEE Transactions on Rehabilitation Engineering*, vol. 7, pp. 413-419, 1999.
- [26] A. Al-Ani and A. Al-Sukker, "Effect of Feature and Channel Selection on EEG Classification," presented at 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS'06, 2006.
- [27] H. Ramoser, J. Muller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement," *IEEE Transactions on Rehabilitation Engineering*, vol. 8, pp. 441-446, 2000.
- [28] B.-h. Yang, G.-z. Yan, T. Wu, and R.-g. Yan, "Subject-based feature extraction using fuzzy wavelet packet in brain-computer interfaces," *Signal Processing*, vol. 87, pp. 1569-1574, 2007.