Ant Colony Optimization Based Feature Selection Method for QEEG Data Classification

Turker Tekin Erguzel[1] ✉, Serhat Ozekes[1], Selahattin Gultekin[2], and Nevzat Tarhan[3,4]

[1]Department of Computer Engineering, Faculty of Engineering and Natural Sciences, Uskudar University, Istanbul, Turkey
[2]Department of Bioengineering, Faculty of Engineering and Natural Sciences, Uskudar University, Istanbul, Turkey
[3]Department of Psychology, Faculty of Humanities and Social Sciences, Uskudar University, Istanbul, Turkey
[4]Department of Psychiatry, NPIstanbul Hospital, Istanbul, Turkey

Objective   Many applications such as biomedical signals require selecting a subset of the input features in order to represent the whole set of features. A feature selection algorithm has recently been proposed as a new approach for feature subset selection.

Methods   Feature selection process using ant colony optimization (ACO) for 6 channel pre-treatment electroencephalogram (EEG) data from theta and delta frequency bands is combined with back propagation neural network (BPNN) classification method for 147 major depressive disorder (MDD) subjects.

Results   BPNN classified R subjects with 91.83% overall accuracy and 95.55% subjects detection sensitivity. Area under ROC curve (AUC) value after feature selection increased from 0.8531 to 0.911. The features selected by the optimization algorithm were Fp1, Fp2, F7, F8, F3 for theta frequency band and eliminated 7 features from 12 to 5 feature subset.

Conclusion   ACO feature selection algorithm improves the classification accuracy of BPNN. Using other feature selection algorithms or classifiers to compare the performance for each approach is important to underline the validity and versatility of the designed combination.   **Psychiatry Investig 2014;11(3):243-250**

Key Words   Neural networks, Feature selection, QEEG, Major depressive disorder.

INTRODUCTION

Reduction of pattern dimensionality using feature extraction is one of the most important steps for classification process. Feature selection has also considerable importance in areas such as bioinformatics,[1-3] signal processing,[4-8] image processing,[9-11] text categorization,[12] data mining,[13] pattern recognition[14-18] and medical diagnosis.[19,20] The aim of feature selection is to choose a subset of available features by eliminating less important or unnecessary features. To extract as much information as possible from a given set while using a smaller number of features, the features with little or no predictive information is to be

eliminated, and strongly correlated redundant features are to be ignored.[21] Thus, a large amount of computation time can be saved with a valuable subset. The selected subset of features used to represent such classification function influences several aspects of classification, including the time required to learn a classification function, the accuracy of the learned classification algorithm, the time-space cost associated with the features, and the number of samples required for training. MDD is considered to be a chronic, relapsing and remitting illness and early medical diagnosis is important for the consequent treatment process. Many of the patients (30−50%) fail to respond to initial antidepressant treatment process.[22] So there is a clear need for methods that select the right treatment for the right patient.[23] Repetitive transcranial magnetic stimulation (rTMS) has been proposed as an alternative[24] with its less invasive and less painful treatment process compared to electrical brain stimulation application.[25] In the light of "Personalized Medicine" perspective to depression, recently both ACO and neuroimaging biomarkers have been studied and point promising results in aiding treatment prediction using pre-treatment measures.[26-29]

Studies have been conducted mainly with neurophysiological EEG biomarkers[30,31] and functional neuroimaging biomarkers[32,33] which focused on the predictive effect of change of frontal quantitative EEG (QEEG) cordance in theta and delta frequency bands. In,[34] EEG data is analyzed to compare normal subjects versus subjects suffering from various mental disorders. It was found that a change in delta or theta band EEG power can be evaluated as a specific marker of brain dysfunction.[35] Considerable number of applications underline that the AD medication effects are physiologically detectable in the EEG and QEEG cordance is one of the auspicious biomarkers used to predict the treatment response which has generated research interest. In addition to its valuable contribution as biomarker, EEG patterns with optimized subset using ACO to minimize the number of features while maximizing classification performance.

In a study, ACO was compared with other well-known feature selection and projection techniques using two different biosignal-driven applications,[36] ACO was used as a feature selection method to classify hand motion surface electromyography signals in another study.[27] Another feature selection application using ACO was used for images from the mammography image analysis society database.[28] ACO method was also tested on one of the most important biosignal driven applications, which is the Brain Computer Interface (BCI) problem with 56 EEG channels.[29] In a pilot study the algorithm was introduced to select genes relevant to cancers first, then the multilayer perceptron neural network and support vector machine classifiers were used for cancer classification.[37] The main goal for the clinical research in the MDD is predicting the response of MDD subjects to rTMS therapy using their pre-treatment QEEG cordance and enhancement of the diagnostic accuracy. These are crucially important for the proper medical treatment and slowing down of the progress of the illness. An ANN based model combined with an optimization algorithm was designed as a tool in order to reduce the number of features while increasing the prediction accuracy.

In this paper, an artificial intelligence approach combining ACO and BPNN is proposed to classify MDD subjects as responder (R) or non-responder (NR) to rTMS treatment using most relevant features.

METHODS

Participants

The research was conducted in Neuropsychiatry Istanbul Hospital to understand from the value of QEEG to classify the MDD subjects before rTMS treatment as R or NR. The study has been formally approved by the local Medical Research Ethics Committee. This study was based on an open-label design. Patients who were willing to participate first visited a psychiatrist in order to assess if they met the inclusion criteria. All subjects were free of psychotropic medication for at least two weeks prior to enrollment. Subjects with nonpsychotic depressive disorder as defined by International Statistical Classification of Diseases and Related Health Problems (ICD-10) criteria and determined by 17-item Hamilton Depression Rating Scale (HAM-D) score higher than 14 were eligible.

A total of 147 major depressive disorder subjects, resistant to medication treatment, completed the protocols and were examined for the study. Responder and non-responder groups did not differ with respect to the psychopharmacological treatment process. In order to minimize potential confusing outcomes of pharmacological withdrawal symptoms, all subjects were on a monotherapy regimen and received concurrent selective serotonin reuptake inhibitor (SSRI) antidepressant medication during their 3 weeks, 20 sessions of rTMS therapy. No patients were receiving lithium or mood stabilizer or benzodiazepines. A baseline clinical assessment was conducted in the day prior to rTMS treatment by a psychiatrist using the 17 item Hamilton depression scale. Patients were assessed twice during the study using clinical, neuropsychological and QEEG assessments. Routine laboratory studies (complete blood count, chemistry, thyroid stimulating hormone), urine toxicology screen, and electrocardiogram were performed at study screening, and subjects were required to be medically stable before entry. Patients with organic brain disorders, with pacemakers, psychotic symptoms, dementia, delirium, substance-related disorders, cluster A or B axis II disorders, patients treated with electroconvulsive therapy (ECT) in the prior six months, patients having any past history of craniotomy, skull fracture, seizures, or significant neurological illness and the ones who had past history of suicidal intent, plan, or attempt were ineligible (exclusion criteria).

EEG recordings

During pre-treatment QEEG, subjects were instructed to rest in the eyes-closed, maximally alert state, in a quiet room with subdued lighting. The technicians monitored the QEEG data during the recording and re-alerted the subjects every minute as needed to avoid drowsiness. Electrodes were placed with an electrode using 19 recording electrodes distributed across the head according to the international 10−20 system arrangement. Three minutes of eye-closed EEG at rest were acquired using Scan LT EEG amplifier and electrode cap (Compumedics/Neuroscan, USA) with the sampling rate of 250 Hz. 19 sintered Ag/AgCl electrodes positioned according to the 10/20 International System with binaural reference. EEG signals received from 6 electrodes (Fp1, Fp2, F3, F4, F7, and F8)

in slow bands (delta and theta). Raw EEG signal was filtered through a band-pass filter (0.15−30 Hz) before artifact elimination. Manually selected (minimum 2 minutes) artifact-free EEG data which has minimum split-half reliability ratio of 0.95 and test-retest reliability ratio of 0.90 were used for cordance calculations. Fast Fourier transform was used to calculate absolute and relative power in each of two non-overlapping frequency bands: delta (1−4 Hz) and theta (4−8 Hz). Leuchter and colleagues stated the EEG cordance method first to provide a measure, which had face validity for the detection of cortical elimination or interruption of sensory nerve fibers. It was noticed that the EEG over a white-matter lesion caused absolute theta power decrease but relative theta power increase which is called as "discordant". So, the EEG cordance calculation process deals with both absolute and relative EEG power. Negative values of cordance (discordance) are used to underline low perfusion or metabolism, while positive values (concordance) are evaluated as high perfusion or metabolism indication.[38] A subsequent study corroborated the method comparing EEG cordance with simultaneously recorded PET scans reflecting perfusion.[39]

rTMS session procedures and ratings

rTMS was applied using the Magstim Super Rapid2 stimulator (Magstim Company, Whitland, UK) with figure-of-eight shaped Air Film Coil in all patients in an open-label manner. The rTMS intensity was set at 100% of the motor threshold which was determined by visual inspection. Stimulations were given to the left prefrontal cortex, deemed to be located anterior to the cortical motor area of the abductor pollicis brevis of which the motor threshold was determined. The treatment schedule was six days in a week, from Monday to Saturday for three weeks. 25Hz stimulation with the duration of 2 seconds was delivered 20 times with 30-second intervals. A full course comprised 1000 magnetic pulses.

Subjects were classified as "responders" if the HAM-D score at three weeks showed at least a 50% improvement over the pre-treatment HamD score. The HamD is a well-accepted means of quantifying the severity of depression. For our purposes, the HamD percentage change value is discretized into two values (or classes), corresponding to responder (R) when it is larger than or equal to 50%, and non-responder (NR) otherwise.[40] Table 1 gives the HAMD scores for each group before and after rTMS treatment.

BP neural network

Artificial neural networks are widely used solving problems in analyzing biomedical signals, because of their variety of applicability and their ability to learn complex and nonlinear relations. ANNs are trained by example data set instead

Table 1. Hamilton Depression Rating Scale (HAMD) Scores of R and NR Subjects before and after repetitive transcranial magnetic stimulation treatment

Number of subjects	Responder	Non-responder
	90	57
Pre-treatment HAMD (±SD)	23.84±3.57	22.93±3.66
Post-treatment HAMD (±SD)	9.13±1.93	14.87±3.11

of rules. When used in diagnosis of neuromuscular disorders, ANNs are not affected by factors such as human fatigue, emotional states, and habituation. They are also capable of rapid identification, analysis of conditions, and diagnosis in real time.[41] There are various types and architectures of neural networks varying fundamentally in the way they learn; the details of which are well documented in literature.[42,43]

BP neural network is a typical multilayer feed forward network trained according to back propagation algorithm. BP neural network uses parallel distributed processing approach to handle both qualitative and quantitative knowledge. It has strong robustness, fault tolerance and adaptability and can fully approximate any complex nonlinear relationship.[44] Because of these advantages, BP neural network is more appropriate for processing EEG data which is possible noisy, unstable and nonlinear. In this study, for modeling process, feedforward neural network trained by a backpropagation algorithm is used. The network is based on the supervised procedure, i.e. the network constructs a model based on examples of data with known outputs. The architecture of the network is a layered feed-forward neural network, in which the non-linear elements (neurons) are arranged in successive layers, and the information flows from input layer to output layer, through the hidden layer(s).[45] Input data is received from 6 electrodes as QEEG cordance, 10 neurons were used in hidden layer and sigmoid transfer function used in each neuron because of its nonlinear behavior. In order to minimize the error between the model output and a reference value MSE (mean square error) is used as the cost function, given in equation 1. The cost function is minimized by ACO.

$$J(w) = \frac{1}{N} \sum_{k=1}^{N} (y_k - z_k)^2 \qquad (1)$$

Where y_k is the output of the model and z_k is the reference output.

Feature selection with ACO algorithm

Feature selection and dimension reduction are important steps in a pattern recognition tasks. In this study, although the feature set was not excessive and giving satisfactory outcomes, using the most informative features increased the classification rate. Reducing the number of features also enabled the classifier to learn a more robust solution and achieved a bet-

ter generalization performance. In order to get an optimal sub-set of features ACO algorithm was employed and the flow chart of optimization process is given in Figure 1.

The feature selection optimization process steps are performed combining the optimization algorithm and BPNN classifier. Selected features by ACO are transferred to the classifier and generated model is tested using test set with the assigned features. The performance of each ant is then evaluated with MSE error to update the pheromone table finally. The process continues to satisfy the stopping ciriteria which is defined as an optimal error value.

ACO is are an iterative, probabilistic meta-heuristic for finding solutions to combinatorial optimization problems. It is based on the foraging mechanism employed by real biological ants attempting to find a short path from their nest to a food source. While foraging, the ants communicate indirectly using their pheromone, which they use to mark their respective paths and which attract other ants. In the ACO algorithm, artificial ants (agents) generate virtual pheromone to update their path through the decision graph, i.e. the path that reflects which alternative node an agent will chose. The amount and density of pheromone an agent uses to update its path depends on how good the solution is copared to those found by competing former agents of the same iteration. Following agents use the pheromone markings of previous good agents as a means of orientation when making their own selections to find the shortest path of all possible alternatives.[46]

Since this problem closely resembles finding the shortest path to a food source, the ACO was first applied to the optimization of traveling salesman problem (TSP).[47] In such a problem, a set of cities (nodes) is given and the distance between each is known. The aim is to find the shortest path that allows each city to be visited just once. Alternative paths are generated on the basis of a probabilistic model and 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 satisfying until the algorithm was coupled with a local optimizer.[48] One problem is early convergence to a less than optimal solution because too much virtual pheromone was laid quickly. To avoid this problem, 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 as:

$$p_{ij}^k = \begin{cases} \frac{\tau_{ij}^\sigma \cdot \eta_{il}^\upsilon}{\sum_{c_{il} \in N(s^p)} \tau_{ij}^\sigma \cdot \eta_{il}^\upsilon} & \text{if } c_{ij} \in N(s^p) \\ 0 & \text{otherwise} \end{cases} \qquad (2)$$

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

$$\eta_{ij} = \frac{1}{d_{ij}} \qquad (3)$$

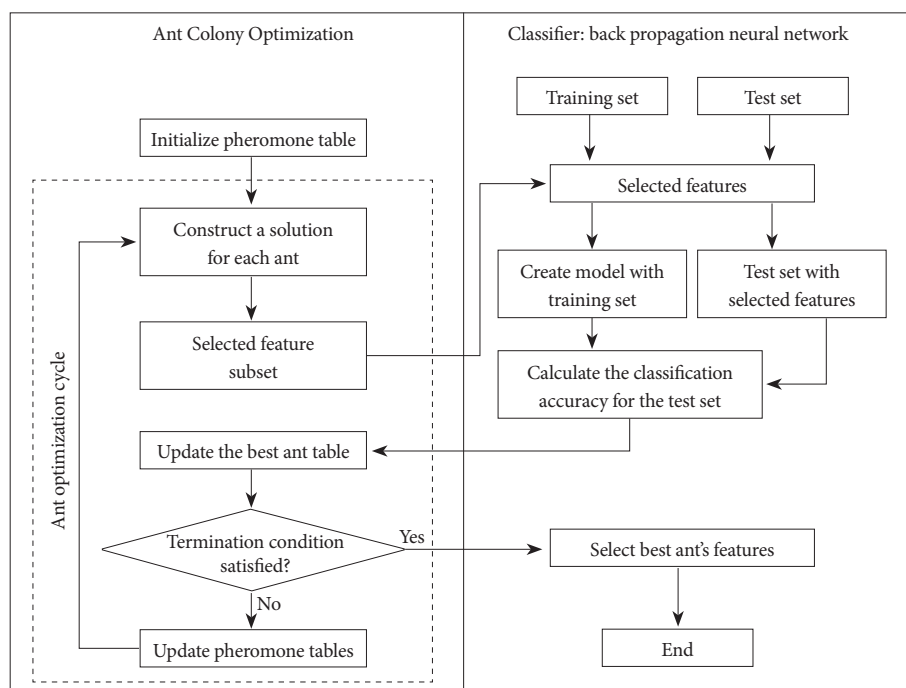Where d_{ij} is the distance between city i and city j.

Figure 1. Design of proposed ant colony optimization based feature selection of parameters.

During each of the iterations the pheromone values are updated by all the mants 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:

$$\tau_{ij} \leftarrow (1-\rho).\tau_{ij} + \sum_{k=1}^{m} \Delta\tau_{ij}^{k} \qquad (4)$$

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:

$$\Delta\tau_{ij}^{k} = \begin{cases} \frac{\delta}{L_k} & \text{if ant used edge (i, j) in its tour,} \\ 0 & \text{otherwise} \end{cases} \qquad (5)$$

Where δ is a constant and L_k is the length of the tour constructed by ant k.[29]

In this study, the value for each feature is represented by a node and the vectors between nodes can be considered as the paths between nodes. Before an ant starts from a randomly selected path, a hundred possible values were proposed for each node to enable variety. Each one of the connections between nodes from the 1st to the 12th node via various paths is a solution created by the visit of an ant to be evaluated. So a cost value is assigned to the travelling ant regarding to the path performance. Therefore better ants will track the trajectory closer and will generate lower costs ensuring higher pheromone density on the path. Same loop is repeated by other ants and the feature optimization process is not terminated unless the desired fitness function is met. The process is started by a simple individual ant from the randomly selected path, and the optimal path is found by colony evolution. Travelling ants also deposit pheromones to the paths they passed to enable others follow their trail. The pheromones are updated and evaporated regularly to let others find alternative paths and solutions. Local pheromone updates and global pheromone updates were used to update pheromone table after each iteration and tour respectively. After an ant completes iteration, it updates local pheromone table as given in equation 6;

$$\tau(n)_{ij} = \tau(n-1)_{ij} + \frac{(0.01 \times \theta)}{J} \qquad (6)$$

where $\tau(k)_{ij}$ is the pheromone value between nest (i) and (j) at the nth iteration,
θ is the general pheromone updating coefficient,
J is the cost function for the tour travelled by the ant.

After each tour, a hundred iterations are completed and global pheromone update process starts. Pheromones of the paths belonging to the best and worst of the tour are updated as given in the following equations 7 and 8 respectively:

$$\tau(n)_{ij}^{best} = \tau(n)_{ij}^{best} + \frac{\theta}{J_{best}} \qquad (7)$$

$$\tau(n)_{ij}^{worst} = \tau(n)_{ij}^{worst} - \frac{0.3 \times \theta}{J_{worst}} \qquad (8)$$

where $\tau(n)_{ij}^{best}$ and $\tau(n)_{ij}^{worst}$ are the pheromones of the paths followed by the ant in the tour with the lowest (J_{best}) and highest cost value (J_{worst}) in one iteration respectively. The pheromone evaporation, given in equation 8, decrease the pheromone density of the visited paths to let the ants visit low density paths assuring diversity.

$$\tau(n)_{ij} = \tau(n)_{ij}^{\lambda} + [\tau(n)_{ij}^{best} + \tau(n)_{ij}^{worst}] \qquad (9)$$

where λ is the evaporation constant.

RESULTS

In this study, an up-to-date swarm intelligence method, ACO, was employed as feature selection algorithm for 12 inputs and then BP neural network was used to classify 147 subjects as responder or non-responder. 6-fold CV was performed to train and test the classifier with stratified sampling. That hybrid approach, combining BPNN classifier with ACO algorithm, was significantly affected by the number of the selected features and contributed to the performance of classification. The combination of classification results before and after feature selection process are given with overall accuracy, sensitivity and area under Receiver Operating Characteristic (ROC) curve parameters in table 2 and the ROC curve for the compared approaches is plotted in figure 2.

Throughout the classification process, the intersection point of true positive rate (TPR) and false positive rate (FPR) at each threshold is plotted to form the ROC curve. Each point on the ROC curve represents a sensitivity/(1-specificity) pair corresponding to a particular decision threshold. Depending on the classification performance, the relative changes of TPR and FPR may differentiate causing sharp transitions between cut off points in ROC curve.

After the frequency band and channel selection phase, ACO algorithm was used to reduce the feature set by considering the classification error as cost function. The contribution of feature selection process to the accuracy is quite satisfacto-

Table 2. Repetitive transcranial magnetic stimulation treatment responder results using ant colony optimization (ACO)

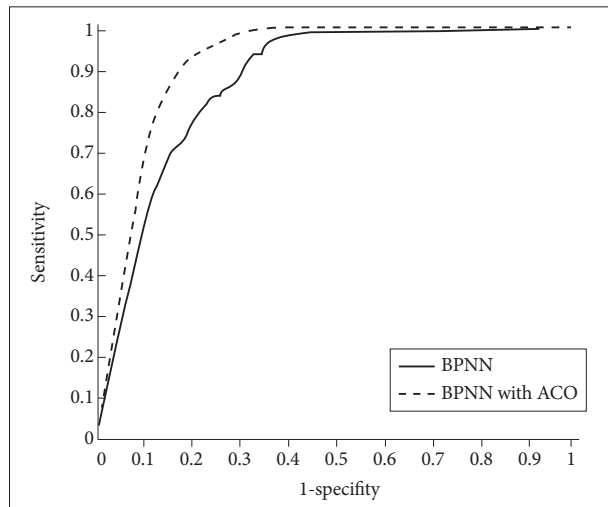| Feature selection method	Number of features	Accuracy %	Sensitivity	AUC
None	12	80.25	84.44%	0.8531
ACO Algorithm	5	91.83	95.55%	0.911

AUC: area under curve

Figure 2. Receiver operating characteristic curves for back propagation neural network and back propagation neural network with ant colony optimization (ACO). BPNN: back propagation neural network.

ry. NN classified R subjects with 91.83% overall accuracy, percentage of examples been classified correctly, and 95.55% subjects detection sensitivity. Area under ROC curve (AUC) was also used to evaluate the performance of ACO algorithm. AUC value after feature selection increased from 0.8531 to 0.911. The features selected by the optimization algorithm were Fp1, Fp2, F7, F8, F3 for theta frequency band and eliminated 7 features from 12 to 5 feature subset.

Since cordance values are correlated with regional cerebral blood flow, findings with this measure could be interpreted within the same conceptual framework as other functional neuroimaging studies demonstrating an abnormal pattern of metabolism or perfusion in the prefrontal cortex and the anterior cingulate in depressed patients. Moreover, frontal electrical activity in theta frequency band has been associated with the function of these structures and previous research has linked pretreatment theta activity of the anterior cingulate with clinical response.[23,26,49] The results of our study support the former clinical researches and focus on the prefrontal region and theta frequency band for MDD patients.

DISCUSSION

High dimensionality nature of QEEG data caused by the use of high number of electrodes and long periods of task time is one of the drawbacks in QEEG study. Evolutionary based approaches are alternative methodologies to conventional dimension reduction methods with the advantage of not requiring the entire recording sessions for operation.

ACO is an evolutionary method that achieves performance through evaluation of several generations of possible solutions. Optimizing the feature set enables the classifier to

work with a reduced sized data set. Treatment response prediction is crucially important for proper clinical treatment and medical research and various classification methods are proposed in literature.

This paper utilizes and combines a machine learning technique and a meta-heuristic approach to classify subjects using pre-rTMS treatment data. The ant colony optimization algorithm was first introduced to select features relevant to MDD subjects before classification then the ANN classifier was used for classification. Experimental results show that selecting features by using ACO can improve the accuracy of the classifier.

Similar studies for performance evaluation using the combination of modeling approaches and feature selection techniques generated quite satisfactory outcomes to underline the validity and reliability of the method used in this study. Feature selection of EEG signals for schizophrenic patients,[50] alzheimer patients,[51] depression patients[52] and patients suffering from epilepsy[53] were also studied and contributed to the combination of optimization algorithms and neural networks to increase the classification performance.

Using ACO as a feature selection method, various studies was also used biomedical data.[54,55] The machine learning paradigm has been applied in a study using ANN fed with EEG data to differentiate three classes of subjects: those with schizophrenia, those with depression, and healthy subjects.[56] Combining various biomarkers, statistical methods were also used to predict[23,40,57] treatment results. In order to increase the prediction performance, various feature selection methods were proposed for multi-channel EEG data.[58,59] Some other studies underlined the performance of ACO as feature selection method comparing to principal component analysis, genetic algorithm, random tree generation and differential evolution methods.[27,28,36,54,60]

Although the proposed ACO feature selection algorithm improves the classification accuracy of ANN, it still needs further investigation on other type of classifiers. Using other feature selection algorithms or classifiers to compare the performance for each approach is important to underline the validity and versatility of the designed combination. The results show that the approach is suitable for biological data classification and promising which is thus highly applicable to clinical studies requiring diagnostic results.

REFERENCES

1. Basiri ME, Ghasem-Aghaee N, Aghdam MH. Using ant colony optimization-based selected features for predicting post-synaptic activity in proteins. LNCS 2008;4973:12-23.

2. Saeys Y, Inza I, Larranaga P. A review of feature selection techniques in bioinformatics. Bioinformatics 2007;23:2507-2517.

3. Nemati S, Basiri ME, Ghasem-Aghaee N, Aghdam MH. A novel ACO-GA hybrid algorithm for feature selection in protein function prediction. Expert Syst Appl 2009;36:12086-12094.

4. Yeh YC, Wang WJ, Chiou CW. Feature selection algorithm for ECG signals using range-overlaps method. Expert Syst Appl 2010;37:3499-3512.

5. Liao TW. Feature extraction and selection from acoustic emission signals with an application in grinding wheel condition monitoring. Eng Appl Artif Intel 2010;23:74-84.

6. Nemati S, Boostani R, Jazi MD. A novel text-independent speaker verification system using ant colony optimization algorithm. LCNS 2008;5099:421-429.

7. Ververidis D, Kotropoulos C. Fast and accurate sequential floating forward feature selection with the Bayes classifier applied to speech emotion recognition. Signal Process 2008;88:2956-2970.

8. Roda A. Perceptual tests and feature extraction: Toward a novel methodology for the assessment of the digitization of old ethnic music records. Signal Process 2010;90:1000-1007.

9. Lu JJ, Zhao TZ, Zhang YF. Feature selection based-on genetic algorithm for image annotation. Knowl-Based Syst 2008;21:887-891.

10. Tsai JS, Huang WB, Kuo YH, Horng MF. Joint robustness and security enhancement for feature-based image watermarking using invariant feature regions. Signal Process 2012;92:1431-1445.

11. Choi SI, Oh J, Choi CH, Kim C. Input variable selection for feature extraction in classification problems. Signal Process 2012;92:636-648.

12. Aghdam MH, Ghasem-Aghaee N, Basiri ME. Application of ant colony optimization for feature selection in text categorization. Proceeding of the 5th IEEE Congress on Evolutionary Computation, IEEE Press: Hong Kong; 2008.

13. Patricia EN, Andries L, Engelbrecht P. A decision rule-based method for feature selection in predictive data mining. Expert Syst Appl 2010; 37:602-609.

14. Jensen R. Combining Rough and Fuzzy Sets for Feature Selection. Ph.D. Thesis. Edinburgh: University of Edinburgh; 2005.

15. Kanan HR, Faez K. An improved feature selection method based on ant colony optimization (ACO) evaluated on face recognition system. Appl Math Comput 2008;205:716-725.

16. Wang Y, Dahnoun N, Achim A. A novel system for robust lane detection and tracking. Signal Process 2012;92:319-334.

17. Huang K, Aviyente S. Information theoretic wavelet packet sub-band selection for texture classification. Signal Process 2006;86:1410-1420.

18. Awaidah SM, Mahmoud SA. A multiple feature/resolution scheme to Arabic (Indian) numerals recognition using hidden Markov models. Signal Process 2009;89:1176-1184.

19. Polat K, Gunes S. Medical decision support system based on artificial immune recognition immune system, fuzzy weighted pre-processing and feature selection. Expert Syst Appl 2007;33:484-490.

20. Zhang GP. Neural networks for classification: A Survey. IEEE Trans Syst Man Cybern Part C Appl Rev 2000;30:451-462.

21. Guyon I, Elisseeff A. An introduction to variable and feature selection. JMLR 2003;3:1157-1182.

22. Trivedi MH, Morris DW, Grannemann BD, Mahadi S. Symptom clusters as predictors of late response to antidepressant treatment. J Clin Psychiatry 2005;66:1064-1070.

23. Bares M, Brunovsky M, Novak T, Kopecek M, Stopkova P, Sos P, et al. The change of prefrontal QEEG theta cordance as a predictor of response to bupropion treatment in patients who had failed to respond to previous antidepressant treatments. Eur Neuropsychopharmacol 2010;20:459-466.

24. O'Reardon JP, Solvason HB, Janicak PG, Sampson S, Isenberg KE, Nahas Z, et al. Efficacy and safety of transcranial magnetic stimulation in the acute treatment of major depression: a multisite randomized controlled trial. Biol Psychiatry 2007;62:1208-1216.

25. Im CH, Lee C. Computer-aided performance evaluation of a multi-channel transcranial magnetic stimulation system. IEEE Trans Magn 2006;42:3803-3808.

26. Arns M, Drinkenburg WH, Fitzgerald PB, Kenemans JL. Neurophysiological predictors of non-response to rTMS in depression. Brain Stimul 2012;5:569-576.

27. Huang H, Xie HB, Guo JY, Chen HJ. Ant colony optimization-based feature selection method for surface electromyography signals classification. Comput Biol Med 2012;42:30-38.

28. Karnan M, Thangavel K, Sivakuar R, Geetha K. Ant colony optimization for feature selection and classification of microcalcifications in digital mammograms. Advanced Computing and Communications. Surathkal, Heidelberg: ADCOM; 2006.

29. Khushaba, RN, AlSukker A, Al-Ani A, Al-Jumaily A. Intelligent artificial ants based feature extraction from wavelet packet coefficients for biomedical signal classification. ISCCSP; 2008; Malta: Heidelberg.

30. Price GW, Lee JW, Garvey C, Gibson N. Appraisal of sessional EEG features as a correlate of clinical changes in an rTMS treatment of depression. Clin EEG Neurosci 2008;39:131-138.

31. Micoulaud-Franchi JA, Richieri R, Cermolacce M, Loundou A, Lancon C, Vion-Dury J. Parieto-temporal alpha EEG band power at baseline as a predictor of antidepressant treatment response with repetitive transcranial magnetic stimulation: a preliminary study. J Affect Disord 2012;137:156-160.

32. Kito S, Hasegawa T, Koga Y. Cerebral blood flow ratio of the dorsolateral prefrontal cortex to the ventromedial prefrontal cortex as a potential predictor of treatment response to transcranial magnetic stimulation in depression. Brain Stimul 2012;5:547-553.

33. Richieri R, Boyer L, Farisse J, Colavolpe C, Mundler O, Lancon C, et al. Predictive value of brain perfusion SPECT for rTMS response in pharmacoresistant depression. Eur J Nucl Med Mol Imaging 2011;38: 1715-1722.

34. Coutin-Churchman P, Anez Y, Uzcategui M, Alvarez L, Vergara F, Mendez L, et al. Quantitative spectral analysis of EEG in psychiatry revisited: drawing signs out of numbers in a clinical setting. Clin Neurophysiol 2003;114:2294-2306.

35. Khodayari A, Reilly JP, Hasey GM, deBruin H, MacCrimmon DM. Diagnosis of psychiatric disorders using EEG data and employing a statistical decision model. International Conference of the IEEE EMBS; 2010; Buenos Aires: Heidelberg.

36. Khushaba, RN, AlSukker A, Al-Ani A, Al-Jumaily. A combined ant colony and differential evolution feature selection algorithm. Ant Colony Optimization and Swarm Intelligence Lecture Notes in Computer Science 2008;5217:1-12.

37. Chiang Y. The application of ant colony optimization for gene selection in microarray-based cancer classification. Proceedings of the 7th International Conference on Machine Learning and Cybernetics; 12-15 July 2008: Kunming.

38. Leuchter AF, Cook IA, Lufkin RB, Dunkin J, Newton TF, Cummings JL, et al. Cordance: a new method for assessment of cerebral perfusion and metabolism using quantitative electroencephalography. Neuroimage 1994;1:208-219.

39. Leuchter AF, Uijtdehaage SH, Cook IA, O'Hara R. Mandelkern M. Relationship between brain electrical activity and cortical perfusion in normal subjects. Psychiatry Res 1999;90:125-140.

40. Khodayari A, Reilly JP, Hasey GM, deBruin H, MacCrimmon D. Using pretreatment electroencephalography data to predict response to transcranial magnetic stimulation therapy for major depression. International Conference of the IEEE EMBS; 2011; Boston: Heidelberg.

41. Micheli-Tzanakou E. Neural Networks in Biomedical Signal Processing. In: Bronzino JD, Editor. Medical Devices and Systems. Boca Raton: CRC Press LLC, 2000, p.7.1-7.14.

42. Fausett L. Fundamentals of Neural Networks Architectures, Algorithms, and Applications. Englewood Cliffs, NJ: Prentice Hall; 1994.

43. Haykin S. Neural Networks: A Comprehensive Foundation. New York:

Macmillan; 1994.

44. Z Hong. Recognition of epileptic EEG signals based on BP neural networks. Sci Technol Info2009;35:18-23.

45. Lek S, Guegan J. Artificial neural networks as a tool in ecological modeling, an introduction. Ecol Model 1999;120:65-73.

46. Subbotin S. Modifications of ant colony optimization method for feature selection. CADSM 2007; February 20-24, 2007; Polyana: Ukraine.

47. Dorigo M, Gambardella LM. Ant colony system: a cooperative learning approach to the traveling salesman problem. IEEE Trans Evol Comput 1997;1:53-66.

48. Haupt RL, Haupt SE. Practical Genetic Algorithms, 2nd Edition. New York, NY: John Wiley & Sons, Inc; 2004.

49. Khodayari-Rostamabad A, Hasey GM, Maccrimmon DJ, Reilly JP, de Bruin H. A pilot study to determine whether machine learning methodologies using Pre-treatment electroencephalography can predict the symptomatic response to clozapine therapy. Clin Neurophysiol 2010; 121:1998-2006.

50. Sabeti M, Boostani R, Katebi SD, Price GW. Selection of relevant features for EEG signal classification of schizophrenic patients. Biomed Signal Process Control 2007;2:122-134.

51. Hyun TK, Kim BY, Park EH, Kim JW, Hwang EW, Han SK, et al. Computerized recognition of Alzheimer disease-EEG using genetic algorithms and neural network. Future Gener Comp Syst 2005;21: 1124-1130.

52. Arns M, Drinkenburg WH, Fitzgerald PB, Kenemans JL. Neurophysiological predictors of non-response to rTMS in depression. Brain Stimul 2012;5:569-576.

53. Ocak H. Optimal classification of epileptic seizures in EEG using wavelet analysis and genetic algorithm. Signal Process 2008;88:1858-1867.

54. Bursa M, Lhotska L. Ant colony cooperative Strategy clustering in Electrocardiogram and electroencephalogram data clustering. Sicily, Italy: International Workshop on nature Inspired Cooperative Strategies for optimization; 2007.

55. Atyabi A, Luerssen M, Fitzgibbon S, Powers D. The impact of PSO based dimension reduction in EEG study. Brain Informatics Lecture Notes in Computer Science; 2012; Macau: Heidelberg.

56. Li YJ, Fan FY. Classification of Schizophrenia and depression by EEG with ANNs. Proceedings Int. Conf. of the IEEE Eng. New York: In Medicine and Biology Society; 2005.

57. Brakemeier EL, Wilbertz G, Rodax S, Danker-Hopfe H, Zinka B, Zwanzger P,?et al. Patterns of response to repetitive transcranial magnetic stimulation (rTMS) in major depression: replication study in drug-free patients. J Affect Disord 2008;108:59-70.

58. Flotzinger D, Pregenzer M, Pfurtscheller G. Feature selection with distinction sensitive learning vector quantisation and genetic algorithms. IEEE World Congress on Computational Intelligence; 1994; Orlando: USA.

59. Garrett D, Peterson DA, Anderson CW, Thaut MH. Comparison of linear, nonlinear, and feature selection methods for EEG signal classification. IEEE Trans Neural Syst Rehabil Eng 2003;11:141-144.

60. Chen B, Chen L, Chen Y. Efficient ant colony optimization for image feature selection. Signal Process 2013;93:1566-1576.