Applied Soft Computing xxx (2012) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

Perceptual pain classification using ANFIS adapted RBF kernel support vector machine for therapeutic usage

Maryam Vatankhah[a], Vahid Asadpour[b,*], Reza Fazel-Rezai[c]

[a] Azad University, Mashhad Branch, Iran
[b] Department of Electrical Engineering, Sadjad Institute of Higher Educations, Iran
[c] Department of Electrical Engineering, University of North Dakota, USA

ARTICLE INFO

Article history:
Received 5 November 2011
Received in revised form 7 July 2012
Accepted 24 November 2012
Available online xxx

Keywords:
Electroencephalogram
EEG
Pain
Adaptive network fuzzy inference system
Support vector machine

ABSTRACT

Diagnosing pain mechanisms is one of the main approaches to improve clinical treatments. Especially, the detection of existence and/or level of pain can be vital when verbal information is not present for instant for neonates, disabled persons, anesthetized patients and also animals. Various researches have been performed to locate and classify pain, however, no consistent result has been achieved. The aim of this study is to show a strict relation between electroencephalogram (EEG) signal features and perceptual pain levels and to clarify the relation of classified signal to pain origin. Cortical regions on scalp were assigned based on an evolutional method for optimized alignment of electrodes that improve the clinical monitoring results. The EEG signals were recorded during relax condition and variety of pain conditions. Specific spectral features which are studied to show consistency with dynamical characteristic of EEG signals were combined with non-linear features including approximate entropy and Lyapunov exponent to provide the feature vector. Evolutionary optimization method was used to reduce the features space dimension and computational costs. A hybrid adaptive network fuzzy inference system (ANFIS) and support vector machine (SVM) scheme was used for classification of pain levels. ANFIS optimizer is used to fine tune the non-linear alignment of kernels of SVM. The results show that pain levels can be differentiated with high accuracy and robustness even for few recording electrodes. This research verifies the hypothesis that electrical variations of brain patterns can be used for determination of pain levels. The proposed classification method provided up to 95% accuracy.

1. Introduction

Our understanding of brain mechanisms of chronic pain conditions has expanded significantly over the past 5–10 years [1]. With the advent of neuroimaging studies, e.g. positron emission tomography (PET) and functional magnetic resonance imaging (fMRI), on human chronic pain disorders and animal models, the cortical representations of persistent pain conditions have begun to be understood and related to origins of neural mechanisms [2]. A few EEG-changes have been repeatedly observed during processing of tonic pain. Lopes [3] and also Steriade and coworkers [1] studied that a tonic experimental pain stimulus will cause a decrease in alpha and an increase in gamma amplitudes in the cortical areas. The findings have been verified by a variety of studies [4]. Cold-press has been mainly used to induce pain to the subjects [5]; furthermore, injections of capsaicin into the skin [6] and into the muscle [7], hypertonic injection of saline into the muscle [8] and cuff-pressure [9] have been used. Experimental comparing of tonic skin and muscle pains directly complies with the result that besides considerable similarities in EEG patterns. Tonic muscle pain resulted into a stronger beta II activity [10]. These results have been obtained without any consistency and even did not allow for generalization across physical stressors.

The stage appears to be set for further progress. However, there are still doubts to the specificity of the described EEG changes for pain [3]. The changes in EEG patterns might simply be due to sensory processing in general and to all cognitive and affective processes linked to sensory processing. The lack of clear answers to these questions is partly a consequence of the methods used for EEG analysis, which do not allow for sufficient experimental control. None of the previous studies used EEG processing to classify pain level and this research is the first attempt to apply pattern recognition to recognize the pain mechanism. In the present study, a new approach classification and recognition of pain for diagnosing purposes is presented.

The aim of this research is to outline the difference between "no pain", "pain sense" and "pain" conditions. A kernel based Support

* Corresponding author. Tel.: +98 511 6029411/2/3.
E-mail addresses: vatankhah@aum.ac.ir (M. Vatankhah), asadpour@aut.ac.ir (V. Asadpour), reza@engr.und.edu (R. Fazel-Rezai).

1568-4946/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.asoc.2012.11.032

Vector Machine (SVM) classifier is used that discriminant hyperplanes identify the desired classes [11]. The selected hyperplane is the one that maximizes the distance from the nearest training points. Maximizing the margin is known to increase the generalization capabilities. An Adaptive Neuro-Fuzzy Interface System (ANFIS) optimizer was used to fine tune the hyperplanes of SVM classifier. ANFIS learns features in the data set and adjusts the system parameters according to a given error criterion [12]. Successful implementations of ANFIS in biomedical engineering have been reported, for classification and data analysis [13], but because of its complex algorithm, it may not being converge for datasets with large dimension. The results show that hybrid ANFIS-SVM has the best performance for classifying nonlinear features.

2. Data collection

Fifteen young, healthy, right-handed, and drug-free volunteers (average age: 28, 7 males and 8 females) were selected to participate in the experiment and those who suffered from any acute and chronic disease were excluded. The volunteers were given a detailed explanation of the procedure. Complete EEG recordings were obtained from 12 electrodes based on 10/20 electrode standard. EEG recording electrodes were affixed at frontal (F3, F4), central (C3, C4), temporal (T3, T4), parietal (P3, P4) and occipital (O1, O2) locations in both hemispheres. Referential electrodes were bilaterally connected to the two earlobes and then averaged. Impedance was controlled to be less than 5 kohm in each lead. EEG signals were sampled at 200 Hz sampling rate and filtered using an elliptic filter with band-width of 0.5–30 Hz.

Sessions of EEG recording took place in a sound-attenuated room with dimmed luminescence. Subjects were seated comfortably in an upright chair. After fixation of EEG leads and recording of 1 min of resting EEG activity, which allowed for screening for pathological EEG patterns, the individual pain thresholds of the subjects were determined. Ethical and safety aspects considered due to possible pain enforcements to volunteers and required considerations and contractions used [14]. In the cold pain condition, subject placed his left hand in a bucket of ice water till the pain is sensed. Volunteer takes back his hand after about 5 s. Due to ethical considerations, volunteers could avoid the experiment in case of inconvenience. The mean ± standard deviation temperatures at beginning and end of cold pain duration were 4.3 ± 0.8 and $4.5 \pm 2.1\,^{\circ}C$, respectively. The subjects were told to take their hand out of the water if they became uncomfortable. The subjects told the perceived intensity rate of the cold pain verbally by a scale from 0 to 10 (no pain is defined by 0, pain is defined by a level of 1–10).

3. Feature extraction

Raw EEG signals were down sampled to the rate of 100 Hz to form a unique sampling and reduce the computational cost. It is shown that pain levels could be differentiated with high accuracy. In spite of high accuracy, the high computation costs and the duration of algorithm's running could be compromised. The quality of recorded data should also be verified on the classification method. In the following sections, the features which are used to compose feature vectors are discussed.

3.1. Energy ratio features

EEG is usually interpreted by the frequency content of the observed data sequence. The frequency domain analysis gives a clear picture of changes. Because of frequency changes during pain, energy ratios between different EEG sub-bands were computed for each channel during pain. It was shown that there would be pain related changes in the EEG at specific electrode locations [7,10]. To discover the brain rhythms during pain, alpha (8–13 Hz), beta (13–35 Hz), delta (0–4 Hz), and theta (4–8 Hz) band energy ratios of spectrogram $SPEC(t,f)$ at time t and frequency f were calculated as shown below [6]. They show the total energy of each defined spectral band relative to total signal energy. Time resolution of 10 ms and frequency resolution of 0.15 Hz was selected:

$$\alpha = \frac{\int_8^{13} SPEC(t,f)df}{\int_0^{35} SPEC(t,f)df} \quad (1)$$

$$\beta = \frac{\int_{13}^{35} SPEC(t,f)df}{\int_{13}^{35} SPEC(t,f)df} \quad (2)$$

$$\delta = \frac{\int_0^{4} SPEC(t,f)df}{\int_0^{35} SPEC(t,f)df} \quad (3)$$

$$\theta = \frac{\int_4^{8} SPEC(t,f)df}{\int_0^{35} SPEC(t,f)df} \quad (4)$$

3.2. Approximate entropy

Approximate entropy is a recently formulated family of parameters and statistics quantifying regularity (orderliness) in serial data [4]. It has been used mainly in the analysis of heart rate variability [7,8], endocrine hormone release pulsatility [9], estimating regularity in epileptic seizure time series data [10] and estimating the depth of anesthesia [15]. Approximate entropy assigns a non-negative number to a time series, with larger values corresponding to more complexity or irregularity in the data.

EEG signal represents regular and uniform pattern during synchronized cooperative function of cortical cells. This pattern results to low entropy values. In contrast concentric functions and higher levels of brain activity lead to high values of entropy. Shannon entropy H is defined as:

$$H = \sum_{i=1}^{N} P_i \log_2 P_i \quad (5)$$

in which P_i is the average probability that amplitude of ith frequency band of brain rhythm be greater than times standard deviation and N is the total number of frequency bands. H is 0 for a single frequency and 1 for uniform frequency distribution over total spectrum. Because of the non-linear characteristics of EEG signals, approximate entropy can be used as a powerful tool in the study of the EEG activity. In principle, the accuracy and confidence of the entropy estimate improve as the number of matches of length and $m+1$ increases. Although m and r are critical in determining the outcome of approximate entropy, no guidelines exist for optimizing their values. For this study, approximate entropy was estimated with $m=3$ and $r=0.25$ based on an investigation on original data sequence. Therefore, we are provided one dimensions of feature vector.

3.3. Fractal dimension

Fractal dimension emphasizes the geometric property of basin of attraction. These dimension show geometrical property of attractors and is also computed very fast [15]. Our goal was to associate each 5-s segment data as a trial to its corresponding class. To do this, features were extracted from each 1-s segment with 50% overlap, and sequence of 9 extracted features were considered as the feature vector of a 5-s segment, which was to be modeled and classified.

Please cite this article in press as: M. Vatankhah, et al., Perceptual pain classification using ANFIS adapted RBF kernel support vector machine for therapeutic usage, Appl. Soft Comput. J. (2012), http://dx.doi.org/10.1016/j.asoc.2012.11.032

In Higuchi's algorithm, k new time series are constructed from the signal $x(1), x(2), x(3), \ldots, x(N)$, under study is [16]:

$$x_m^k = \left\{ x(m), x(m+k), x(m+2k), \ldots, \left(m + \left[\frac{N-m}{k} \right] k \right) \right\} \quad (6)$$

in which $m = 1, 2, \ldots, k$ and k indicate the initial time value, and the discrete time interval between points, respectively. For each of the k time series x_m^k the length $L_m(k)$ is computed by:

$$L_m(k) = \frac{\sum_i |x(m+ik) - x(m+(i-1)k)|(N-1)}{[(N-m)/k]k} \quad (7)$$

in which N is the total length of the signal x. An average length is computed as the mean of the k lengths $L_m(k)$ (for $m = 1, 2, \ldots, k$). This procedure is repeated for each k ranging from 1 to k_{max}, obtaining an average length for each k. In the curve of $\ln(L(k))$ versus $\ln(1/k)$, the slope of the best fitted line to this curve is the estimate of the fractal dimension.

3.4. Lyapunov exponent

Lyapunov exponents are a quantitative measure for distinguishing among the various types of orbits based upon their sensitive dependence on the initial conditions, and are used to determine the stability of any steady-state behavior, including chaotic solutions [17]. The reason why chaotic systems show aperiodic dynamics is that phase space trajectories that have nearly identical initial states will separate from each other at an exponentially increasing rate captured by the so called Lyapunov exponent. In this study, the Lyapunov exponents estimated from the observed time series [18]. This is defined as follows: consider two (usually the nearest) neighboring points in phase space at time 0 and at time t, distances of the points in the ith direction being $||\delta X_i(0)||$ and $||\delta X_i(t)||$, respectively. The Lyapunov exponent is then defined by the average growth rate λ_i of the initial distance [19]

$$\frac{||\delta X_i(t)||}{||\delta X_i(0)||} = 2^{\lambda_i}(t \to \infty)$$

or

$$\lambda_i = \lim_{t\to\infty} \frac{1}{t} \log_2 \frac{||\delta X_i(t)||}{||\delta X_i(0)||} \quad (8)$$

Generally, Lyapunov exponents can be extracted from observed signals in two different ways [20]. The first method is based on the idea of following the time evolution of nearby points in the state space [15]. This method provides an estimation of the largest Lyapunov exponent only. The second method is based on the estimation of local Jacobi matrices and is capable of estimating all the Lyapunov exponents [18]. Vectors of all the Lyapunov exponents for particular systems are often called their Lyapunov spectra. This method was used for Lyapunov vector extraction in this work. An optimized size of 7 is considered for the vector which led to the best mean classification rate on the SVM classifier.

4. Classification

The SVM has been used extensively for classification of EEG signals [21]. It is shown that EEG signal has separable intrinsic vectors which could be used in SVM classifier. SVM classifiers use discriminant hyperplanes for classification. The selected hyperplanes are those that maximize the margin of classification edges. The distance from the nearest training points usually measured based on a non-linear kernel to map the problem to a linear solvation space [22]. A RBF kernel based SVM is proposed here that Lagrangian optimization is performed using an adjustable ANFIS algorithm. It will be shown that due to conceptual nature of pain for patients this proposed method leads to adjustable soft decision classification.

4.1. Proposed classification using nonlinear SVM with RBF kernel

Training the SVM is a quadratic optimization problem in which the hyperplane is defined as [23]:

$$y_i(w\Phi(x_i, y_i) + b) \geq 1 - \xi_i, \quad \xi_i \geq 0,\ i = 1, \ldots, l,\ j = 1, \ldots, m \quad (9)$$

in which x_i is input vector, b is bias, w is adapted weights, ξ_i is class separation, $\Phi(x_i, y_j)$ is the mapping kernel, l is the number of training vectors, j is the number of output vectors, and y_i is desired output vector. The weight parameters should be achieved so that the margin between the hyperplane and the nearest point to be maximized. The only free parameter, C, in SVMs controlled the trade-off between the maximization of margin and the amount of misclassifications. Optimization of Eq. (9) yields to optimum w which is the answer of problem. It could be done using Lagrange multipliers defined as [11]:

$$L(w, b, \alpha, \mu) = \frac{1}{2}||w||^2 + C\sum_{i=1}^{l} \xi_i - \sum_{i=1}^{l}\sum_{i=1}^{m} \alpha_i\alpha_j(y_j(w\Phi(x_iy_i) + b - 1 + \xi_i)) - \sum_{i=1}^{l} \mu_i\xi_i \quad (10)$$

in which $C > \alpha_i$, $\alpha_j \geq 0$ and $\mu \geq 0$ for $i = 1, \ldots, l$ and $j = 1, \ldots, m$ and C is the upper bond for Lagrange coefficient. The coefficient C is a representation of error penalty so that higher values yield to bigger penalties. Karush–Kuhn–Tucher conditions lead to optimization of Lagrange multipliers [24]:

$$\frac{\partial L}{\partial W_v} = W_v - \sum_i \alpha_i y_i x_{iv} = 0 \quad (11)$$

$$\frac{\partial L}{\partial b} = -\sum_i \alpha_i y_i = 0$$

$$\frac{\partial L}{\partial \xi_i} = C - \alpha - \mu_i$$

subject to

$$\mu_i(w\Phi(x_i, y_j) + b - 1 + \xi_i) \geq 0, \quad \xi_i \geq 0,\ \alpha_i \geq 0$$

and,

$$\alpha_i(y_i(w\Phi(x_i, y_j) + b - 1 + \xi_i)) = 0 \quad (12)$$

Usually nonlinear kernels provide classification hyperplanes that cannot be achieved by linear weighting [25]. Using appropriate kernels, SVM offered an efficient tool for flexible classification with a highly nonlinear decision boundary. Radial Basis Function (RBF) kernel was used in this work which is defined as:

$$\Phi(x, y) = \exp\left(\frac{-||x-y||^2}{2\sigma^2} \right) \quad (13)$$

in which σ is the standard deviation. The proposed feature extraction method is depicted in Fig. 1. The outputs are fed to the adjustable ANFIS described in next section. Two parameters had to be selected beforehand: the trade-off parameter C and the kernel standard deviation σ. They could be optimized for an optimal generalization performance in the traditional way, by using an independent test set or n-fold cross-validation. It has been suggested that the parameters could be chosen by optimizing the upper bound of the generalization error solely based on training data [25]. The fraction of support vectors, i.e. the quotient between the number of support vectors and all training samples, gave an upper bound on the leave-one-out error estimate because the resulting decision function changed only when support vectors were omitted. Therefore, a low fraction of support vectors could be used as a criterion for the parameter selection.

Please cite this article in press as: M. Vatankhah, et al., Perceptual pain classification using ANFIS adapted RBF kernel support vector machine for therapeutic usage, Appl. Soft Comput. J. (2012), http://dx.doi.org/10.1016/j.asoc.2012.11.032

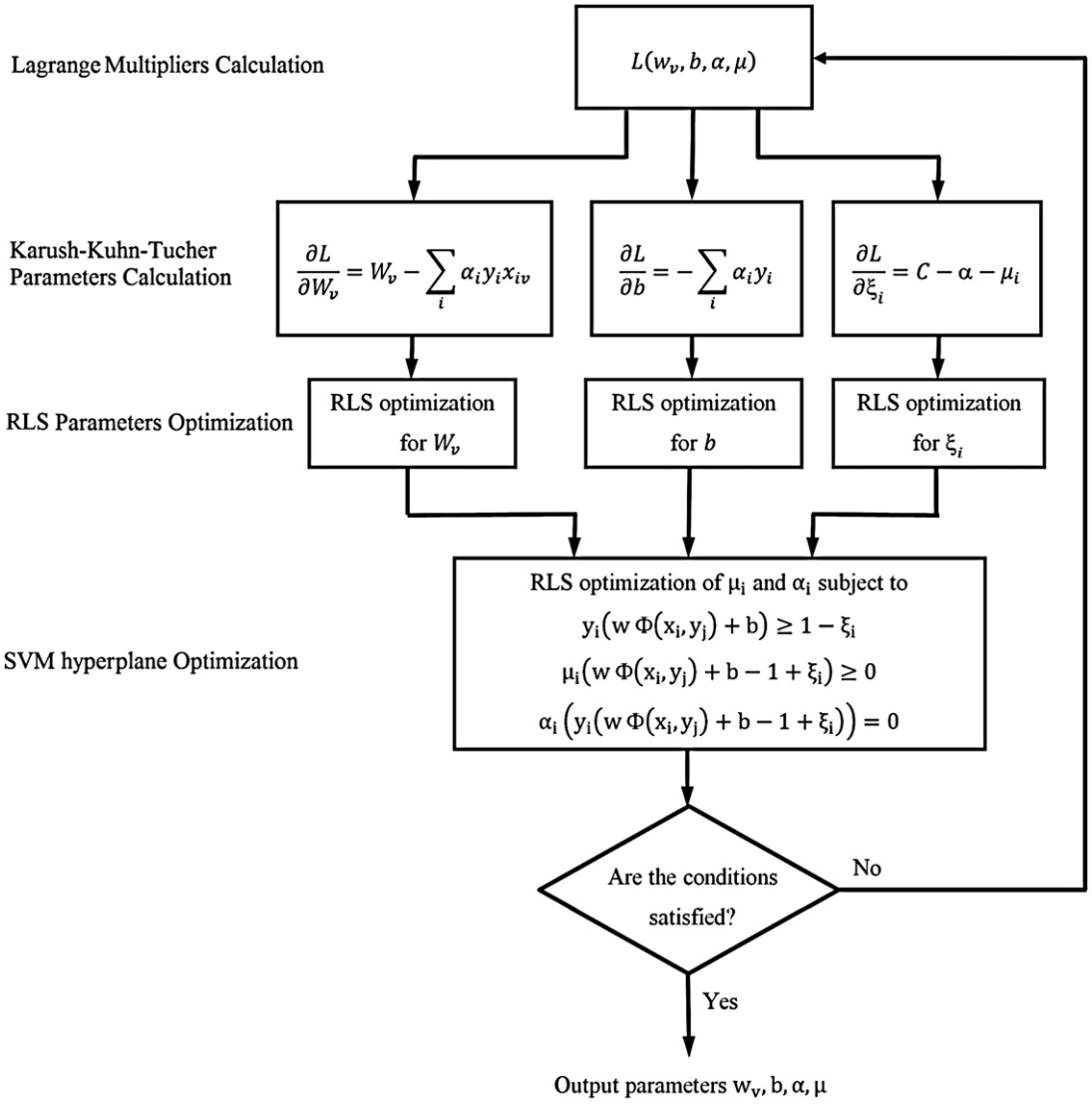

Fig. 1. Block diagram of RBF kernel SVM classification system. The optimization of SVM parameters are performed in: Lagrange multipliers, Karush–Kuhn–Tucher optimization, and SVM hyperplane optimization.

4.2. Adjustable ANFIS optimization

An ANFIS system is used with Sugeno fuzzy model for fine adjustment of SVM classification kernels. Such framework makes the ANFIS modeling more systematic and less reliant on expert knowledge and therefore facilitates learning and adjustment. The ANFIS architecture for implementation of these two features is shown in Fig. 2. In this figure a circle indicates a fixed node, whereas a square indicates an adaptive node. In the first layer, all the nodes are adaptive nodes. The outputs of layer 1 are the fuzzy membership grade of the inputs, which are given by [26]:

$$O_i^1 = \mu_{A_i}(x), \quad i = 1, 2 \quad \text{and} \quad O_i^1 = \mu_{B_{i-2}}(x), \quad i = 3, 4 \qquad (14)$$

O_i^1 is the ith output of layer 1, $\mu_{A_i}(x)$ and $\mu_{B_i-2}(x)$ are type A and type B arbitrary fuzzy membership functions of nodes i and $i-2$, respectively. In the second and third layer, the nodes are fixed nodes. They are labeled M and N respectively, indicating they perform as

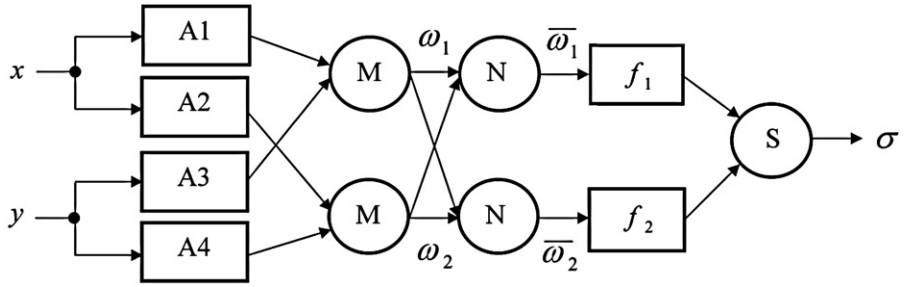

Fig. 2. ANFIS network architecture.

Please cite this article in press as: M. Vatankhah, et al., Perceptual pain classification using ANFIS adapted RBF kernel support vector machine for therapeutic usage, Appl. Soft Comput. J. (2012), http://dx.doi.org/10.1016/j.asoc.2012.11.032

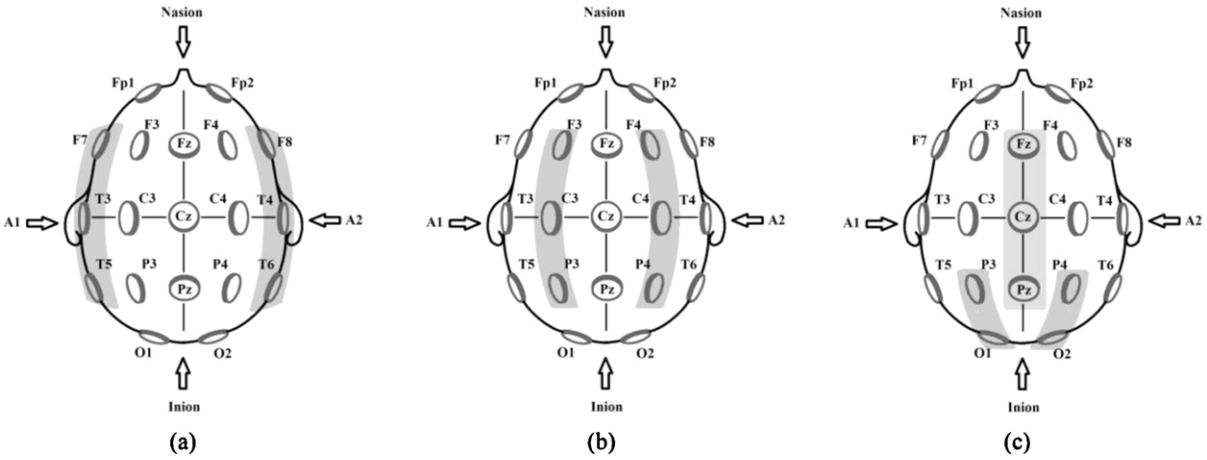

Fig. 3. Montages (a) I, (b) II, and (c) III for electrodes based on 10/20 standard.

a simple multiplier. The outputs of these layers can be represented as:

$$O_i^2 = w_i = \mu_{A_i}(x)\mu_{B_i}(x), \quad i = 1, \ldots, 4 \tag{15}$$

$$O_i^3 = \bar{w}_i = \frac{w_i}{w_i + w_{i+1}} \quad i = 1, \ldots, 4 \tag{16}$$

which are the so-called normalized firing strengths. In the fourth layer, the nodes are adaptive nodes. The output of each node in this layer is simply the product of the normalized firing strength and a first order polynomial for the first order Sugeno model. The outputs of this layer are given by:

$$O_i^4 = \bar{w}_i f_i = \bar{w}_i(p_i x + q_i y + r_i) \quad i = 1, \ldots, 4 \tag{17}$$

in which f_i is the firing rate, p_i is the x scale, q_i is the y scale, and r_i is the bias for ith node. In the fifth layer, there is only one single fixed node that performs the summation of all incoming signals:

$$O_i^5 = \sum_{i=1}^{2} \bar{w}_i f_i = \sum_{i=1}^{2} \frac{w_i f_i}{w_i + w_{i+1}} \tag{18}$$

It can be observed that there are two adaptive layers in this ANFIS architecture, namely the first layer and the fourth layer. In the first layer, there are three modifiable parameters $\{a_i, b_i, c_i\}$, which are related to the input membership functions. These parameters are the so-called premise parameters. In the fourth layer, there are also three modifiable parameters $\{p_i, q_i, r_i\}$, pertaining to the first order polynomial. These parameters are so-called consequent parameters.

The task of the learning algorithm for this architecture is to tune all the above-mentioned modifiable parameters to make the ANFIS output match the training data. When the premise parameters of the membership function are fixed, the output of the ANFIS model can be written as:

$$\sigma = (\bar{w}_1 x)p_1 + (\bar{w}_1 x)q_1 + (\bar{w}_1 x)r_1 + (\bar{w}_2 x)p_2 + (\bar{w}_2 y)q_2 + (\bar{w}_2)r_2 \tag{19}$$

This is a linear combination of the modifiable consequent parameters p_1, q_1, r_1, p_2, q_2 and r_2. When the premise parameters are fixed the least squares method is used easily to identify the optimal values of these parameters after adjustment of ANFIS weights using SVM. When the premise parameters are not fixed, the search space becomes larger and the convergence of the training becomes slower.

A hybrid algorithm combining the least squares method and the gradient descent method was adopted to identify the optimal values of these parameters [27]. The hybrid algorithm is composed of a forward pass and a backward pass. The least squares method (forward pass) was used to optimize the consequent parameters with the premise parameters fixed. Once the optimal consequent parameters are found, the backward pass starts immediately. The gradient descent method (backward pass) was used to adjust optimally the premise parameters corresponding to the fuzzy sets in the input domain. The output of the ANFIS was calculated by employing the consequent parameters found in the forward pass. The output error was used to adapt the premise parameters by means of a standard back propagation algorithm. It has been shown that this hybrid algorithm is highly efficient in training the ANFIS [19].

5. Results

EEG variation during pain and no-pain condition is studied in various research works [28]. Our results show that the EEG pattern changes significantly during pain and relaxation; specifically alpha and theta changes have remarkable therapeutic content and could be used for cognitive pain level classification [29]. Cognitive levels of pain named “no pain”, “pain sense”, and “pain” are labeled as level I, level II and level III, respectively. In the following experiments, 9 subjects were used for training and 4 subjects were used for test. Differentiation between levels of pain is completely based on subjective recognition after sophisticated duration of training. Implementation of proposed methods performed using MATLAB R2010a software.

Preliminary tests showed that no pain and pain sense could be differentiated more distinctive than pain sense and pain and more exact labeling could be achieved. Arrangement of cortical electrodes was challenging because its optimum selection could make the use of method more adaptable and comfortable for subjects and physicians. Three different montages of electrodes namely montage I, montage II, and montage III were used based on clinical experiments [30] and their classification results were compared to assign the best electrode arrangement. Fig. 3 illustrates the above-mentioned montages of electrodes. As it could be seen the entire region of sensor cortex and motor cortex is examined by these electrode sets. Table 1 shows the classification rates for ANFIS-SVM

Table 1
ANFIS-SVM classification rate for three types of montages.

	Montage I	Montage II	Montage III
Classification rate for ANFIS-SVM	78%	81%	88%

Please cite this article in press as: M. Vatankhah, et al., Perceptual pain classification using ANFIS adapted RBF kernel support vector machine for therapeutic usage, Appl. Soft Comput. J. (2012), http://dx.doi.org/10.1016/j.asoc.2012.11.032

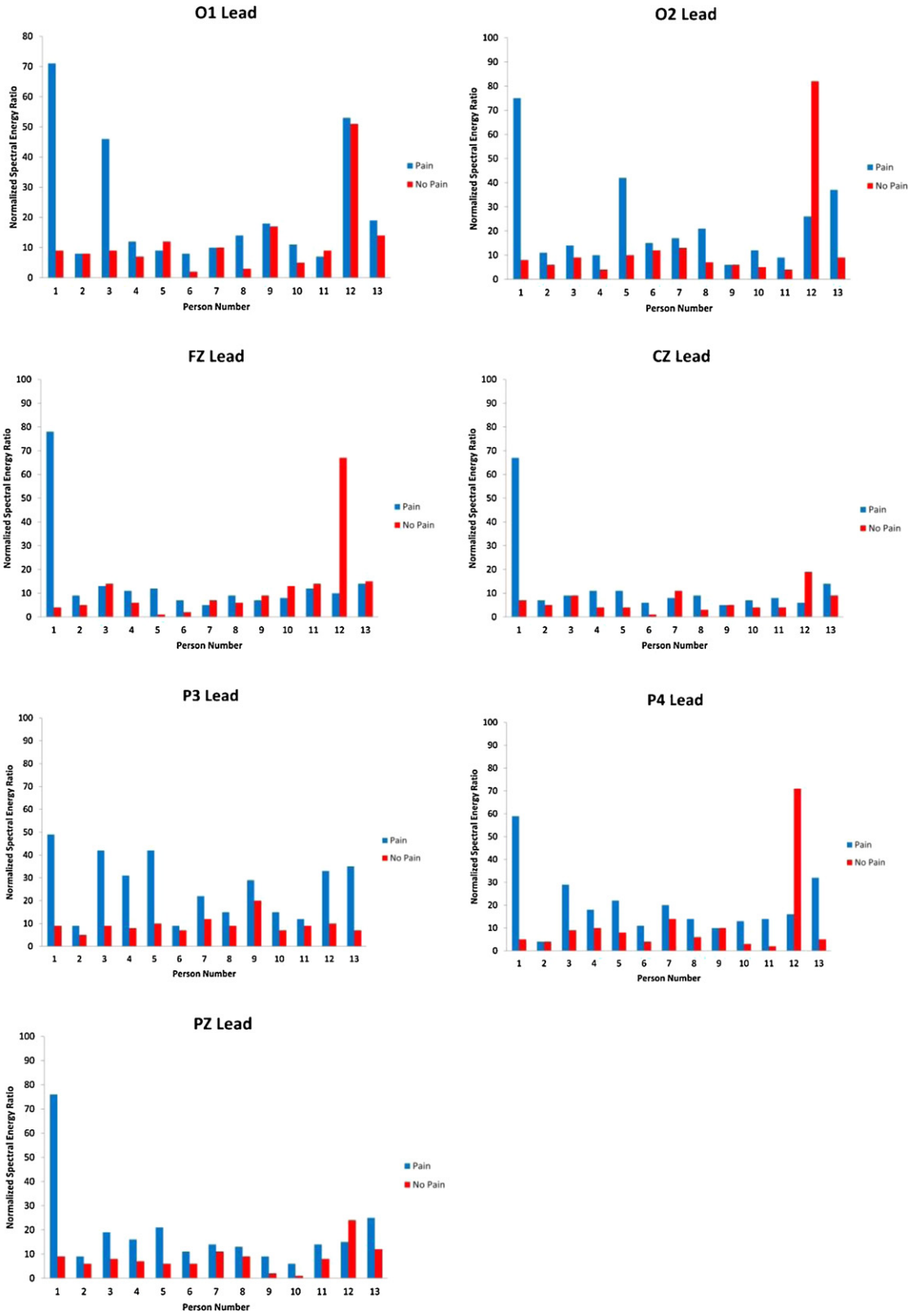

Fig. 4. Alpha band average power (vertical axis with rational units) of montage III electrode set for 13 subjects (horizontal axis) in which solid line is for no-pain and dashed line is for pain condition.

using these electrode sets. Montage III resulted to the best classification and is considered as the electrode set for pain classification in this research. The solid theoretical foundations of the SVM allow us to optimize several parameters of Kernel function using analytical methods. The ANFIS optimization, therefore, results to the optimized results. The over fitting is cleverly avoided by controlling the tradeoff between the training error minimization and the learning capacity of the decision functions. Finally, the decision function parameters can be easily updated because they depend on the SVs only.

Please cite this article in press as: M. Vatankhah, et al., Perceptual pain classification using ANFIS adapted RBF kernel support vector machine for therapeutic usage, Appl. Soft Comput. J. (2012), http://dx.doi.org/10.1016/j.asoc.2012.11.032

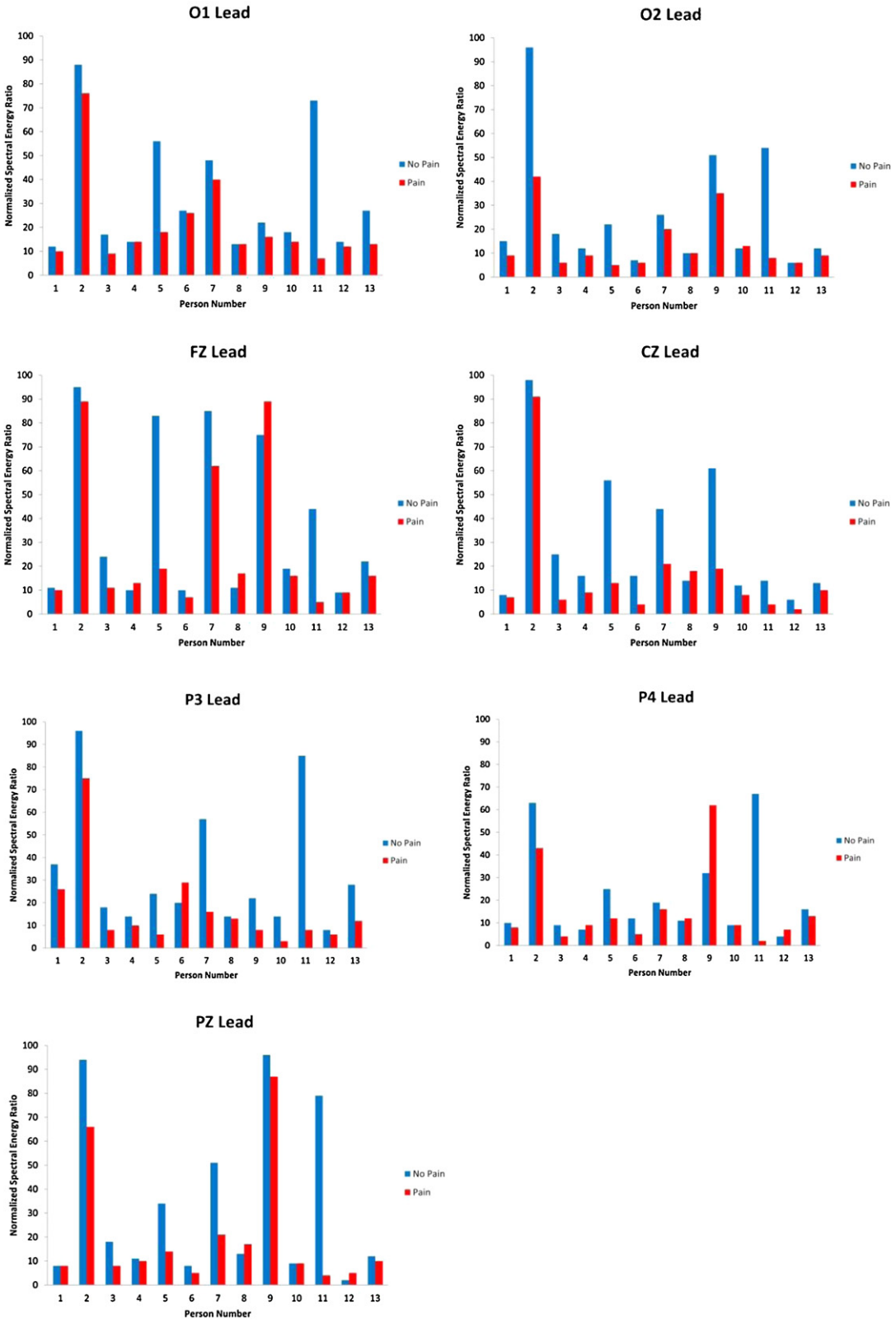

Fig. 5. Delta band average power (vertical axis with rational units) of montage III electrode set for 13 subjects (horizontal axis). Solid line for no-pain and dashed line for pain condition.

Alpha average power for electrodes with montage III is depicted in Fig. 4 in which pain and no-pain conditions are separated. Differentiation is obvious through visual investigations, however, there is some non-uniformity of variations especially for some special subjects. For instance, subject 12 shows different pattern compared to other subject especially for O2, FZ, CZ and P4 electrodes. Generally speaking the alpha average power could be a distinguishing criterion for the problem as it has been shown in the

Please cite this article in press as: M. Vatankhah, et al., Perceptual pain classification using ANFIS adapted RBF kernel support vector machine for therapeutic usage, Appl. Soft Comput. J. (2012), http://dx.doi.org/10.1016/j.asoc.2012.11.032

Table 2
Classification rates of SVM and ANFIS SVM for reduced and not-reduced features for differentiating pain and no-pain.

	Features	SVM	ANFIS-SVM
Not-reduced features	Standard deviation, theta ratio, alpha ratio, entropy, Lyapanov, and fractal dimension	90%	95%
Reduced features	Theta ratio, alpha ratio, and entropy	83%	87%

Table 3
Classification rate for differentiating between pain and no-pain.

	Spectral features	Non-linear features
SVM	75%	89%
ANFIS-SVM	80%	93%

researches [9]. Fig. 5 depicts delta average power for montage III electrodes for pain and no-pain conditions. Differentiation is much more obvious in this frequency band and it also is in compliance with other researches [7,8]. Approximate entropy, Lyapunov exponent and fractal dimension are other features used which are known as non-linear features in this work. It could be seen that approximate entropy is much more robust than spectral domain features and can provide meaningful criterion even for subjects that spectral domain features does not provide differentiation.

Some features are defined for EEG analysis as it mentioned in Section 3. It is feasible to assign the appropriate features that result to best classification rates and rank them using a certain criterion. Using a set of high rank features leads to reduction of computational cost and preventing mismatch of data labeling. A genetic algorithm observation based on ANFIS-SVM classifier reveals that "theta ratio", "alpha ratio", and "entropy" lead to best classification rates. Classification rates of pain and no-pain for reduced and not-reduced features are shown in Table 2 for SVM and ANFIS-SVM classification. As the results for SVM and ANFIS-SVM show the identification rate reduces 7% and 8% for reduced features, respectively. This reduction happens for identification based on three most effective features, and it is obvious that the effect of "standard deviation" and "fractal dimension" could not be neglected. The accuracy of 95% is achieved for ANFIS-SVM proposed method using not-reduced features.

Another evaluation was performed on feature space to assign the most appropriate feature types for ANFIS-SVM classification. The features were classified as spectral features namely, "theta ratio", "alpha ratio", and nonlinear features namely, "entropy", "standard deviation" and "fractal dimension". Table 3 shows the results for differentiating pain and pain sense and Table 4 shows the results for differentiating pain sense and pain. It is obvious that non-linear features result up to 17% improvement for SVM and 14% improvement for ANFIS-SVM classification. Non-linear features are much more effective than spectral features. In the remaining part of this work, all of features were used to achieve high classification rates.

The number of fuzzy rules is effective on the accuracy of classification. Fig. 6 shows the accuracy of ANFIS-SVM classification versus the number of fuzzy rules. The accuracy remains almost the same for the number of fuzzy rules greater than 9 which is used for the rest of discussion. It is a compromising between computational cost and accuracy. This selection does not affect accuracy more than 1%.

Table 4
Classification rate for differentiating between pain and intolerable pain.

	Spectral features	Non-linear features
SVM	68%	73%
ANFIS-SVM	70%	75%

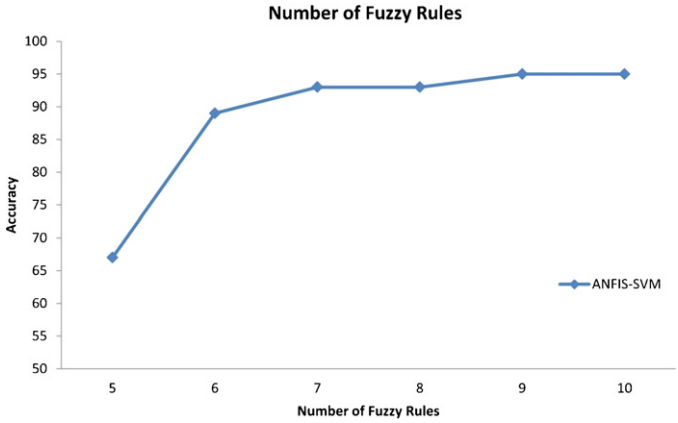

Fig. 6. Accuracy of ANFIS-SVM method versus number of fuzzy rules.

Accuracy of SVM and ANFIS-SVM classifications versus window length M is compared for Lyapunov, Fractal Dimension, Approximate Entropy and Rational Energy features in Figs. 7 and 8, respectively. It could be seen that Lyaponuv features lead to best results for large values of M. SVM classification for Lyapunov features degrades Fractal Dimension for values of M less than 500. It could be due to the method used for Lyapunov feature extraction. As mentioned in Section 3, part D, Jacobi matrices were used in this method. The dimension of these matrices reduces for small values of M and the ranks of matrices reduce with higher rate. It leads to rank insufficiency and provides overlapping features. This reduces the accuracy of classification method, consequently. The highest accuracy of 95% was achieved for ANFIS-SVM classification with M equal to 700. This value was used in the remaining of the paper.

Robustness of the algorithm was also verified using false acceptance and false rejection parameters depicted in Figs. 9 and 10, respectively. False rejection rates show lower values than false acceptance which is appropriate for monitoring and medical care applications. False acceptance and false rejection value of Lyapunov feature reduce to 12% and 4%, respectively. It is shown that non-linear boundary is able to separate pain and no-pain classes. The adjustability of ANFIS classifier using SVM kernels provides accurate classification which could not be achieved for non-adjustable or linear classifier.

As mentioned before, 9 subjects used for training and 4 subjects for test in the experiments. To evaluate the independence of

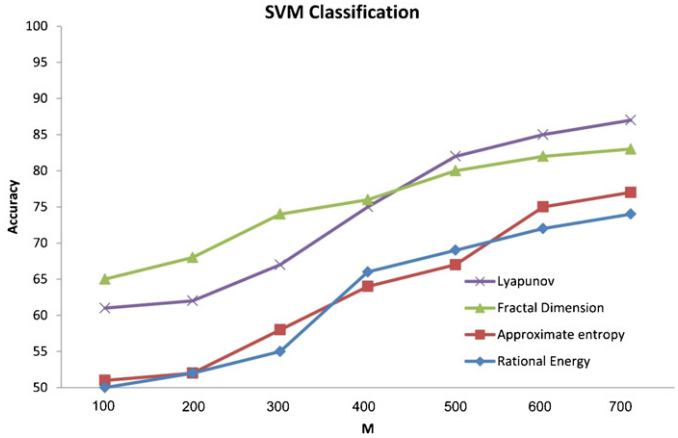

Fig. 7. Accuracy of SVM classification versus window length M for Lyapunov, fractal dimension, approximate entropy, and rational energy features.

Please cite this article in press as: M. Vatankhah, et al., Perceptual pain classification using ANFIS adapted RBF kernel support vector machine for therapeutic usage, Appl. Soft Comput. J. (2012), http://dx.doi.org/10.1016/j.asoc.2012.11.032

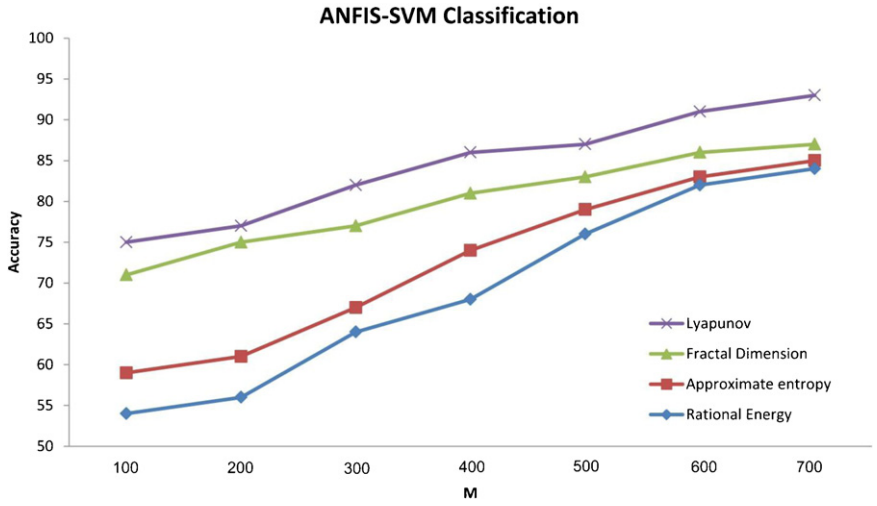

Fig. 8. Accuracy of SVM classification versus window length *M* for Lyapunov, fractal dimension, approximate entropy, and rational energy features.

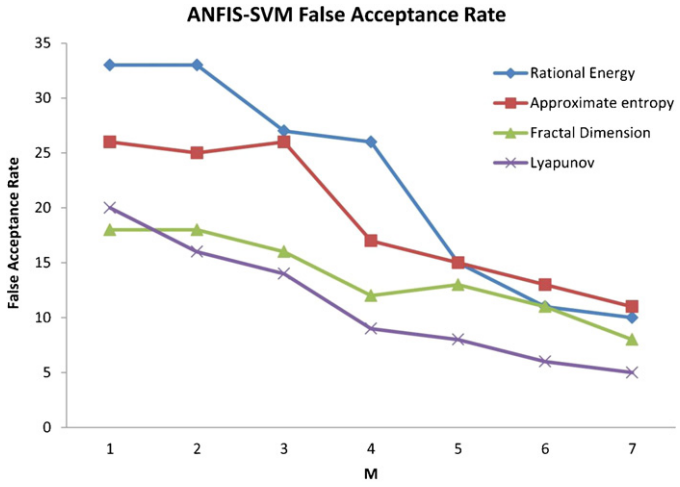

Fig. 9. False acceptance rate of SVM classification versus window length *M* for Lyapunov, fractal dimension, approximate entropy, and rational energy features.

results to the subjects, the final test is done with train and test subjects equal to 12 and 1 person(s) respectively and also for 1 and 12 person(s), respectively. The results are shown in Table 5 for spectral and non-linear features. It is obvious that the difference is 6 and 7 percent for spectral and non-linear features, respectively. The accuracy of 82% was achieved for 1–12 test and train subjects and 75% for 12–1 test and train subjects. Robustness is remarkably high and does not depend on the training subjects.

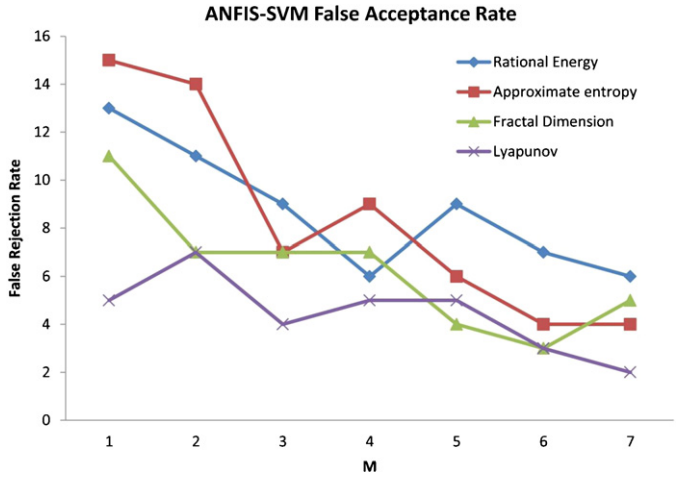

Fig. 10. False rejection rate of SVM classification versus window length *M* for Lyapunov, fractal dimension, approximate entropy, and rational energy features.

Table 5
Evaluation of independence of results to the subjects.

Train to test ratio for ANFIS-SVM classification	Spectral features	Non-linear features
One to twelve	78%	82%
Twelve to one	72%	75%

6. Discussion

Distinguishing pain involved brain regions can improve the curing methods. A lot of research works have been done and studied pain diagnosis. These studies used different methods for inducing pain, but there is not a consistent result. Cold pain is used in this paper and it is shown that intelligent signal processing methods can make difference between three levels of pain. According to EEG sub-band frequency changes during pain, extracting energy ratio through each band would help differentiating three levels of pain. It was shown that although the EEG patterns during pain sense and pain are very similar to each other but differentiating between these levels is possible and the lower accuracy rate shows this reality.

A classification method base on ANFIS adapted SVM proposed in this paper. It showed that non-linear features improve classification rate as an effective component. Through the feature space constructed using approximate entropy and fractal dimension in addition to conventional spectral features, different stages of EEG signals can be recognized from each other expressly. The successful implementations of ANFIS-SVM for EEG signal classification were reported in this context. The results confirmed that this method provides better performance in datasets with lower dimension. This performance achieved for RBF kernel of SVM modified by ANFIS. This research can be a strong base for improved methods in the field of pain cognition and analysis for therapeutic applications.

References

[1] M. Steriade, P. Gloor, R.R. Llinas, Basic mechanisms of cerebral rhythmic activities, Electroencephalography and Clinical Neurophysiology 76 (1990) 481–508.
[2] A.C.N. Chen, P. Rappelsberger, O. Filz, Topology of EEG coherence changes may reflect differential neural network activation in cold and pain perception, Brain Topography 11 (1998) 125–132.

Please cite this article in press as: M. Vatankhah, et al., Perceptual pain classification using ANFIS adapted RBF kernel support vector machine for therapeutic usage, Appl. Soft Comput. J. (2012), http://dx.doi.org/10.1016/j.asoc.2012.11.032

[3] F.L. Silva, Neural mechanisms underlying brain waves: from neural membranes to networks, Electro Encephalography in Clinical Neurophysiology 79 (1991) 81–93.
[4] S. Ferracuti, S. Seri, D. Mattia, G. Cruccu, Quantitative EEG modifications during the cold water pressor test: hemispheric and hand differences, International Journal of Psychophysiology 17 (1994) 261–268.
[5] L. Michels, M. Moazami-Goudarzi, D. Jeanmonod, Correlations between EEG and clinical outcome in chronic neuropathic pain: surgical effects and treatment resistance, Brain Imaging and Behavior 5 (4) (2011) 329–348.
[6] S. Shao, K. Shen, K. Yu, E.P.V. Wilder-Smith, X. Li, Frequency-domain EEG source analysis for acute tonic cold pain perception, Clinical Neurophysiology 123 (10) (2012) 2042–2049.
[7] R.R. Nir, A. Sinai, R. Moont, E. Harari, D. Yarnitsky, Tonic pain and continuous EEG: prediction of subjective pain perception by alpha-1 power during stimulation and at rest, Clinical Neurophysiology 123 (3) (2012) 605–612.
[8] P.F. Chang, N. Arendt, A.C.N. Chen, Dynamic changes and spatial correlation of EEG activities during cold pressor test in man, Brain Research Bulletin 57 (5) (2002) 667–675.
[9] R. Nir, A. Sinai, E. Raz, E. Sprecher, D. Yarnitsky, Pain assessment by continuous EEG: association between subjective perception of tonic pain and peak frequency of alpha oscillations during stimulation and at rest, Brain Research 1344 (2010) 77–86.
[10] K. Roberts, A. Papadaki, C. Gonçalves, M. Tighe, D. Atherton, R. Shenoy, D. McRobbie, P. Anand, Contact heat evoked potentials using simultaneous EEG and fMRI and their correlation with evoked pain, BMC Anesthesiology 8 (1) (2008) 8.
[11] S. Shi-Yun, S. Kai-Quan, J.O. Chong, E. Wilder-Smith, L. Xiao-Ping, Automatic EEG artifact removal: a weighted support vector machine approach with error correction, IEEE Transactions on Biomedical Engineering 56 (2) (2009) 336–344.
[12] K. Majumdar, Human scalp EEG processing: various soft computing approaches, Applied Soft Computing 11 (8) (2011) 4433–4447.
[13] A. Yardimci, Soft computing in medicine, Applied Soft Computing 9 (3) (2009) 1029–1043.
[14] P. Wolpe, Treatment, enhancement, and the ethics of neurotherapeutics, Brain and Cognition 50 (2002) 387–395.
[15] I.D. Guler, Adaptive neuro-fuzzy inference system for classification of EEG signals using wavelet coefficients, Neuroscience Methods 148 (2005) 113–121.
[16] A. Anier, T. Lipping, S. Mel, S. Hovilehto, Higuchi fractal dimension and spectral entropy as measures of depth of sedation in intensive care unit, in: Annual International Conference of the IEEE Engineering in Medicine and Biology Society, San Francisco, 2004, pp. 526–529.
[17] P.N. Sandeep, D. Shiau, J.C. Principe, L.D. Iasemidis, P.M. Pardalos, W.M. Norman, P.R. Carney, K.M. Kelly, J.C. Sackellares, An investigation of EEG dynamics in an animal model of temporal lobe epilepsy using the maximum Lyapunov exponent, Experimental Neurology 216 (1) (2010) 115–121.
[18] E. Derya, Recurrent neural networks employing Lyapunov exponents for analysis of ECG signals, Expert Systems with Applications 37 (2) (2010) 1192–1199 (Original Research Article).
[19] E. Derya, Lyapunov exponents/probabilistic neural networks for analysis of EEG signals, Expert Systems with Applications 37 (2) (2010) 985–992.
[20] J. Lia, Y. Chena, W. Zhangb, Y. Tiana, Computation of Lyapunov values for two planar polynomial differential systems, Applied Mathematics and Computation 204 (1) (2008) 240–248.
[21] E.D. Übeyli, Least squares support vector machine employing model-based methods coefficients for analysis of EEG signals, Expert Systems with Applications 37 (1) (2010) 233–239.
[22] C. Cheng, R.L. Tutwiler, S. Slobounov, Automatic classification of athletes with residual functional deficits following concussion by means of EEG signal using support vector machine, IEEE Transactions on Neural Systems and Rehabilitation Engineering 16 (4) (2008) 327–335.
[23] J.S. Taylor, N. Cristianini, Support Vector Machines and Other Kernel-Based Learning Methods, Cambridge University Press, London, 2000.
[24] B. Vahdani, S.H. Iranmanesh, S.M. Mousavi, M. Abdollahzade, A locally linear neuro-fuzzy model for supplier selection in cosmetics industry, Applied Mathematical Modeling 36 (10) (2012) 4714–4727.
[25] N. Cristianini, J. Shawe-Taylor, Support Vector and Kernel Machines, Cambridge University Press, London, 2001.
[26] J.S. Jang, ANFIS: adaptive-network-based fuzzy inference system, IEEE Transactions on Systems, Man, and Cybernetics 23 (1993) 665–685.
[27] S. Zhan-Li, A. Kin-Fan, C. Tsan-Ming, Neuro-fuzzy inference system through integration of fuzzy logic and extreme learning machines, IEEE Transactions on Systems, Man, and Cybernetics 37 (5) (2007) 1321–1331.
[28] R. Dowman, H. Barkan, V. Thadani, D. Roberts, Human intracranially-recorded cortical responses evoked by painful electrical stimulation of the sural nerve, NeuroImage 34 (2007) 743–763.
[29] L.G. Larrea, M. Frot, M. Valeriani, Brain generators of laser-evoked potentials: from dipoles to functional significance, NeuroPhysiological Clinic 33 (2003) 279–292.
[30] A.C.N. Chen, P. Rappelsberger, Brain and human pain: topographic EEG amplitude and coherence mapping, Brain and Human Pain 7 (1994) 129–140.

Please cite this article in press as: M. Vatankhah, et al., Perceptual pain classification using ANFIS adapted RBF kernel support vector machine for therapeutic usage, Appl. Soft Comput. J. (2012), http://dx.doi.org/10.1016/j.asoc.2012.11.032