Applied Soft Computing 30 (2015) 83–93

Contents lists available at ScienceDirect

Applied Soft Computing

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

Facial neuromuscular signal classification by means of least square support vector machine for MuCI

Mahyar Hamedi [a,b,*], Sh-Hussain Salleh [a], Alias Mohd Noor [a]

[a] Centre for Biomedical Engineering, Transportation Research Alliance, Universiti Teknologi Malaysia, Skudai, Malaysia
[b] Faculty of Bioscience and Medical Engineering, Universiti Teknologi Malaysia, Skudai, Malaysia

ARTICLE INFO

Article history:
Received 13 December 2013
Received in revised form 20 January 2015
Accepted 22 January 2015
Available online 30 January 2015

Keywords:
Facial neuromuscular signal
Electromyogram
Least square support vector machine
Facial gesture recognition
Muscle computer interface
Classification

ABSTRACT

Facial neuromuscular signal has recently drawn the researchers' attention to its outstanding potential as an efficient medium for Muscle Computer Interface (MuCI) applications. The proper analysis of such electromyogram (EMG) signals is essential in designing the interfaces. In this article, a multiclass least-square support vector machine (LS-SVM) is proposed for classification of different facial gestures EMG signals. EMG signals were captured through three bi-polar electrodes from ten participants while gesturing ten different facial states. EMGs were filtered and segmented into non-overlapped windows from which root mean square (RMS) features were extracted and then fed to the classifier. For the purpose of classification, different models of LS-SVM were constructed while tuning the kernel parameters automatically and manually. In the automatic mode, 48 models were formed while parameters of linear and radial basis function (RBF) kernels were tuned using different optimization techniques, cost functions and encoding schemes. In the manual mode, 8 models were shaped by means of the considered kernel functions and encoding schemes. In order to find the best model with a reliable performance, constructed models were evaluated and compared in terms of classification accuracy and computational cost. Results reported that the model including RBF kernel which was tuned manually and encoded by one-versus-all scheme provided the highest classification accuracy (93.10%) and consumed 0.98 s for training. It was indicated that automatic models were outperformed since they required too much time for tuning the parameters without any meaningful improvement in the final classification accuracy. The robustness of the selected LS-SVM model was evaluated through comparison with Support Vector Machine, fuzzy C-Means and fuzzy Gath-Geva clustering techniques.

1. Introduction

Human computer interaction (HCI) is a promising approach which offers various opportunities noticeably in designing assistive technologies for daily life tasks. Such systems aim to provide a reliable pathway between people with severe motor dysfunction and the world around them. Common means of HCI that have been widely employed are human hands, voice, head and gaze [1,2]. Biosignal-based HCIs have been recently proposed which are executed by voluntary signals generated from either a brain activity known as brain computer interaction (BCI) or a body muscle called muscle computer interaction (MuCI). It is reported that, BCI is preferred when the use of MuCIs is not feasible [3]. There are numerous studies on the potential of MuCI systems including multifunction prosthesis [4–6], power exoskeleton control [7], electric-powered wheelchairs [8,9], robotic control [10] and grasping control [11].

Most of the mentioned interfaces are designed based on hand, wrist and finger muscle movements; however, these communication channels are unserviceable for individuals with crucial disabilities like those who cannot even move their neck. Hence, MuCI systems based on facial muscle movements have been proposed [12]. Face contains complicated muscles capable of generating delicate movements (facial gestures/expressions) that carry different information and can be detected through superficial sensors. According to [3], the myoelectric signal activity pattern of different facial gestures is diverse and this phenomenon has motivated researchers to classify such patterns in order to recognize various facial gestures. Ang et al. [13] introduced an EMG-based

* Corresponding author at: Centre for Biomedical Engineering, Transportation Research Alliance, Universiti Teknologi Malaysia, Skudai, Malaysia.
Tel.: +60 147730290.
E-mail addresses: hamedi.mahyar@ieee.org (M. Hamedi), hussain@fke.utm.my (S.-H. Salleh), alias@mail.fkm.utm.my (A.M. Noor).

http://dx.doi.org/10.1016/j.asoc.2015.01.034
1568-4946/

recognition system of the three facial expressions happiness, anger and sadness through three pairs of electrodes. They extracted four different features: mean, standard deviation, Root Mean Square (RMS) and the power spectrum density (PSD) from the rectified filtered EMGS. They reported 94.44% recognition accuracy by utilizing a minimum distance classifier. The use of facial neuromuscular activities to control an electrically powered wheelchair was investigated by Kim et al., [14]. Hidden Markov Model (HMM) was applied to classify the extracted linear prediction coefficients (LPCs) features from eye blink, clenching left, right and both molars. Finally, a mean accuracy of 96.8% was obtained from handicapped and healthy participants. Gibert et al. [15] proposed a recognition system based on six basic expressions: anger, surprise, disgust, happiness, sadness and neutral. They used eight pairs of bipolar surface sensors to capture the EMGs. Absolute value features were extracted from low-passed filtered signals and used as the main characteristics to project each expression by simple Gaussian model. They calculated the Bhattacharyya distance between the Gaussian models to measure the separability of expressions in classification and a mean accuracy of 92% was reached. Recognition of facial movements during unvoiced speech was investigated through four EMG recording channels by Arjunan and Kumar [16]. They tried to identify the unspoken vowel based on the normalized integral moving RMS (MRMS) values of EMGs. They finally stated that the features of surface EMG signals were suitable for characterizing muscle activation during unvoiced speech and subtle gestures. In [12], an interface was designed for a hands-free control system by means of multi-channel facial EMGs to control a virtual robotic wheelchair. Three pairs of bipolar electrodes were employed to acquire the considered facial gestures: smile, frown and pulling up lip corners. They reported 89.75%-100% classification accuracy while mean absolute value (MAV) and support vector machine (SVM) were adopted for feature extraction and classification respectively. Their work was expanded by Rezazadeh et al. [17] to control a virtual interactive tower crane through five different facial gestures. They reached 92.6% recognition accuracy by extracting RMS features and classifying them with subtractive fuzzy c-means (SFCM) clustering method. In their recent study [18], they attempted to classify eight gestures using SFCM plus adaptive neuro-fuzzy inference system (ANFIS). By considering only EMG signals, 93.02% discrimination accuracy was attained.

In our previous studies we focused on developing this technology. To decrease the rate of computational load during processing, we proposed a two channel-based facial gesture recognition system [19] where 90.8% accuracy was obtained by FCM for recognizing five gestures. A comparative study was carried out where FCM outperformed SVM in classifying eight facial expressions recorded by three bipolar channels [20]. A multipurpose interface was designed through ten facial gestures for human machine interface (HMI) applications and 90.41% discrimination rate was achieved by FCM. Moreover, the best combinations of facial gestures were introduced for different applications with two to ten control commands [21]. The effectiveness of six different types of facial EMG features namely integrated EMG (IEMG), variance, wave length (WL), mean absolute value slope (MAVS), MAV and RMS were compared and evaluated for the classification of ten gestures where RMS outperformed the other features [22]. In [23], the effectiveness of conventional Multilayer Perceptron (MLP) and RBF Neural Networks were investigated and compared in order to classify the facial myoelectric signals. Although RBF could provide more reliable performance in both terms of accuracy and speed, the obtained accuracy was lower than previous studies. Recently, a versatile elliptic basis function neural network (VEBFNN) classifier was studied for recognizing the facial gestures [24]. Results indicated that, despite very low computational cost for training the network, classification accuracy was not improved.

According to the mentioned experiments, as long as the maximum number (ten) of facial gestures is concerned, FCM provided a high level of accuracy for classification with a slow performance whereas VEBFNN performed very fast with a lower accuracy. However, a promising classifier should meet the major pre-requisite of real-time MuCI systems which is a reliable trade-off between speed and accuracy. This requires the implementation of a more efficient learning agent capable of classifying complex facial neuromuscular patterns. Different nonlinear methods have been proposed to classify different states of neuromuscular signal activities among which SVM has shown its robustness in coping with the randomness and non-stationary essence of EMGs [25]. The most important feature of SVM is that optimization problems are intrinsically convex and have no local minima which is derived from employing Mercer's conditions on the characterization of the kernels [26]. Solving quadratic optimization problems in SVM requires high computational cost; thus, a modified version called least-square support vector machine (LS-SVM) was proposed by Suykens and Vandewalle [27]. In this method, equality constraints are used instead of inequality ones employed in the SVM. It is stated that the efficiency underlying the induction of LS-SVM classifier significantly depends on the appropriate selection of hyperparameters values [28]. This has been addressed by introducing a range of methods such as cross-validated model selection, extensive grid search and heuristic optimization rules [29,30]. Despite the reported success of employing LS-SVM in numerous researches such as [31–34], to the best of our knowledge the capability of LS-SVM classifier has not yet been examined in terms of EMG analysis, especially for the classification of different facial neuromuscular signals.

In this paper, a multiclass LS-SVM was proposed to classify ten different facial gestures EMGs. Since the efficiency of this method depends significantly on the appropriate selection of kernel hyperparameters, a range of scenarios were investigated to tune the parameters manually and automatically with respect to different performance measures. LS-SVM models were constructed by considering various kernels, optimization methods, multiclass encoding schemes and cost functions. These models were evaluated and compared in order to find the one with the best performance in both terms of classification accuracy and computational cost. This investigation also revealed the most efficient optimization technique and encoding scheme of LS-SVM for facial EMG classification in this study. Finally, the suggested model that provided the best result was evaluated and compared with other popular classifiers.

The remaining parts of this paper are outlined as follows. Next Section presents summarized background information of LS-SVM. Methodologies used for data acquisition, experiments, analysis and evaluation are described in Section 3. Subsequently, the experimental results and statistical findings are investigated and the impact of different kernels, optimization functions and encoding schemes are explored. Finally, in Section 5, conclusions are reported.

2. Least square support vector machines (LS-SVMs) – background

SVM is a nonparametric machine learning algorithm which is originally designed for binary classification [35]. Considering the given training data sets $\{(x_i, y_i)\}_{i=1}^{N}$ with $y_i \in \{\pm 1\}$, SVM maps the input patterns $x_i \in \Re^m$ into a higher dimensional feature space by means of some nonlinear mapping function, $\phi : \Re^m \rightarrow \Re^n$. SVM finds a classification function that separates data classes with the maximum margin through a hyperplane, $w^T\phi(x) + b$ in $\Re^n$ where w is formed by a linear mixture of the set of nonlinear data transformations $w = \sum_{i=1}^{N} \beta_i y_i \phi(x_i)$ and $b = -0.5\,\langle w, x_r + x_s\rangle$ in which β_i is the Lagrange multiplier, x_r and x_s are data points near the

optimal hyperplane known as "support vectors" from each class satisfying $\beta_r > 0$, $y_r = -1$; $\beta_s > 0$, $y_s = 1$. In the conventional SVM, optimal separating hyperplane is obtained by considering margin maximization and training error minimization through solving the quadratic programming (QP) $\min_{w,b}(1/2)(w^T w) + \gamma\sum_{i=1}^{n}\vartheta_i$ subject to the constraints $y_i(w \times \varphi(x_i) + b) \geq 1 - \vartheta_i$, for $i = 1, ..., n$. Where γ is a regularization hyperparameter and ϑ_i are slack variables measuring the error between y_i and the actual SVM output. In order to facilitate solving QP problem, the use of kernel functions are proposed. A kernel function enables the operation to be carried out in the input space rather than in the high dimensional feature space. Various kernel functions deliver different feature spaces and therefore different generalization susceptibility of the resultant classifier. In this paper, linear kernel $K(x_i, x_j) = x_i.x_j^T$ and RBF kernel $K(x_i, x_j) = \exp(-\left\|x_i - x_j\right\|^2/2\sigma^2)$ are considered where σ denotes the kernel parameter [36].

For multiclass classification, the problem reforms into a set of binary classification problems by considering the multiple hyperplane separations where input vectors and training labels are specified as $\{x_i, y_i^c\}_{i=1,c=1}^{i=n,c=C}$, in which n and C are the training pattern index and the number of classes respectively. There are various multiclass categorization techniques and encoding schemes introduced to address the multiclass problems. Four popular methods are one-vs-one (OVO), one-vs-all (OVA), Minimum Output Coding (MOC), and Error Correcting Output Coding (ECOC). OVO applies SVMs to the selected pairs of classes where $C(C-1)/2$ hyperplanes separate the classes. OVA creates C hyperplanes which discriminate each class from the rest. Encoding schemes represent each class M_c, $c = 1, ..., C$, by a unique binary output code-word $M_c \in \{-1, +1\}^L$ of L bits. Then, L binary classifiers are trained to discriminate two opposing subsets with different output bits. MOC is applied to solve the multiclass problem using the minimum number of bits (L) to encode up to 2^L classes while ECOC uses redundant bits to encode each class. There are several decoding schemes for assigning the multiclass label to a new input x. In this paper, Hamming distance decoding [37] is used which computes the output vector $y \in \{-1, +1\}^L$ as the sign of the outputs of the L binary discriminants. The class label M_c is then assigned to the corresponding code-word m_c with minimal Hamming distance to the output vector y.

Solving quadratic optimization problem during the training stage in SVM causes a high degree of computational cost. To address this issue, a least squares type of SVM is introduced by Suykens and Vandewalle [27]. In this method, the least squares cost function with equality restraints is considered in order to obtain an optimal solution through solving a set of linear equations to be used for training. LS-SVMs are closely related to regularization networks [38] and Gaussian processes [39] but additionally emphasize and exploit primal-dual interpretations. Multiclass LS-SVM is trained by $\min_{w_k,b_k,\vartheta_{i,k}}(1/2)\sum_{k=1}^{m}w_k^T w_k + (\gamma/2)\sum_{i=1}^{n}\sum_{k=1}^{m}\vartheta_{i,k}^2$ where m is the number of classes. Detailed information about LS-SVM can be found in [27,40].

3. Methods and materials

3.1. Subject preparation and electrode placement

Facial EMGs have small amplitude and they are contaminated by external and internal factors like motion artifacts, eye movement, brain activity and recording device itself. In addition, they are inherently non-stationary and their characteristics vary within different muscle movements and subjects even during the isotonic muscle movements. Moreover, signal amplitude of different subjects which is the major component of EMGs for classification vary based on active motor units and their activation rate. So, in order to record high quality EMGs with desirable signal to noise ratio (SNR) and to minimize inter-subject variations, several considerations must be taken into account. In this study, at first, the positions of electrodes on participant's face were cleaned from any sweat and dust by using alcohol pad to minimize the impedance between electrodes and skin and remove the motion artifacts. Then, the lead wires were secured via adhesive tape and sufficient numbers of electrodes were placed on the most optimum face sites to capture all facial gestures EMGs with high amplitude [41]. In addition, to eliminate the common noise, bipolar electrodes configuration was applied where signals between the two electrodes were amplified differentially with respect to the reference electrode. Thus, as shown in Fig. 1(a) six electrodes were used for recording signals through three channels simultaneously. Channel 1 was placed on Frontails muscle (above the eyebrows), channel 2 and 3 were located on right and left Temporalis muscle and one electrode was positioned on the boney part of left wrist as the ground. There was a distance of about 2 cm between electrodes in each channel. As another solution for inter-subject variations, EMGs were recorded in consecutive sections by keeping the electrodes on the same location [42].

3.2. System setup and data acquisition protocol

The protocol of this experiment was approved by the Universiti Teknologi Malaysia Human Ethics Research Committee. In the present experiment, facial EMGs were recorded via BioRadio 150 (Clevemed) and they were sampled at ~1000 Hz using a 12 bit A/D converter. EMGs were passed through a notch (50 Hz) and high-pass filters with a cut-off frequency of 0.1 Hz in order to remove undesirable artifacts from user movements and inference line noises. A band-pass filter within the range of 30–450 Hz was also employed to envelope the most significant spectrum of signals. Ten healthy volunteers including five male and five female in the age range of 20–37 participated in this study. The facial gestures considered were smiling with both sides of the mouth, smiling with left side of the mouth, smiling with right side of the mouth, opening the mouth (saying 'a' in the word apple), clenching the molars, gesturing 'notch' by raising the eyebrows, frowning, closing both eyes, closing the right eye and closing the left eye (Fig. 1). All subjects were asked to perform each facial gesture five times lasting for 2 s with 5 s rest between to reduce the effect of muscle exhaustion. Since for each gesture, 10 (5×2) s was informative and EMGs were captured through three channels simultaneously, a three dimensional data set $[Ch_1; Ch_2; Ch_3]_{3\times10}^T$ was achieved. Therefore, for each subject ten (number of gestures) data sets of 3×10 s were collected for further processing.

3.3. Data segmentation

Feeding all conditioned signals directly to a classifier is not practical because of the enormous amount of data and some non-informative EMGs. Therefore, signals must be mapped into lower dimension information vectors (feature vectors) to highlight the most important properties of EMGs. Accordingly, data need to get segmented into a sequence of time portions prior to feature estimation. It is reported that, a short segment results in bias and variance in feature extraction whereas a long one leads to high computational cost and may fail real-time operations [25]. In this study, filtered signals were segmented into non-overlapped windows with 256 ms length and therefore 39 segments were obtained ($10{,}000 \div 256 \approx 39$) for each gesture in each channel. Fig. 2 illustrates the first four segments of clenching molars in channel one.

Fig. 1. (a) Electrode positions, (b) smiling with both sides of the mouth, (c) smiling with right side of the mouth, (d) smiling with left side of the mouth, (e) open the mouth like saying 'a' in 'apple', (f) clenching the molars, (g) raising the eyebrows, (h) closing both eyes, (i) closing right eye, (j) closing left eye, (k) frowning.

3.4. Feature extraction

Feature extraction tends to highlight the most important properties of EMGs to avoid processing a huge amount of data. A suitable feature should carry proper information to discriminate facial gestures and have low computational complexity to be applicable for real-time MuCI applications. Features have direct role on the classifiers performance in both terms of accuracy and computational cost; efficient features facilitate the training procedure and provide a feature space with maximum margin within different class patterns. Many studies have examined various types of time-domain (TD), frequency-domain (FD), and time-scale features for MuCI control systems. Literature shows that there are some restrictions when analyzing facial EMGs through their spectrums since they contain similar frequency components. Therefore, EMGs cannot be processed either by FD or time-frequency distribution algorithms for facial gesture recognition [20,43]. More appropriate characteristics of facial EMGs are TD ones as they are easy to compute and they work base on signal amplitudes. Amplitude is affected by the number of active motor units (muscle fibers + motor neuron) and their activation rate, action potential resulting from different muscle movements, signaling source and innervation ratio of muscles [44]. So, facial EMG amplitude represents activation level, signal energy and contraction duration of the muscles involved in forming different facial gestures. It is reported that EMG power usually identified by RMS TD feature [45] is an excellent measure since in a stable force and non-fatiguing contraction it gives the best estimation of amplitude when a signal is formed as a Gaussian random process. The high potential of this type of feature for facial gesture recognition has already been examined in [19,21]. The efficiency of this feature was also compared to other popular TD features MAV, MAVS, IEMG, WL and signal variance [22]. It was indicated that RMS resulted in maximum classification accuracy as it provided more discriminative information and delivered more clean boundaries within different facial gestures patterns. Therefore, in this study, RMS was computed and extracted from each signal segment which resulted in 39 RMSs for each gesture in each channel. By considering three channels, the feature vector was constructed as $[Ch_1; Ch_2; Ch_3]^T_{3\times 39}$ for each gesture and consequently $[Ch_1; Ch_2; Ch_3]^T_{3\times 390}$ for each subject. In order to have more separable feature vectors, log transform was employed on extracted features to spread the concentrated features [46].

3.5. Classification

In order to classify and discriminate the RMSs extracted from the facial neuromuscular activities, LS-SVM technique was

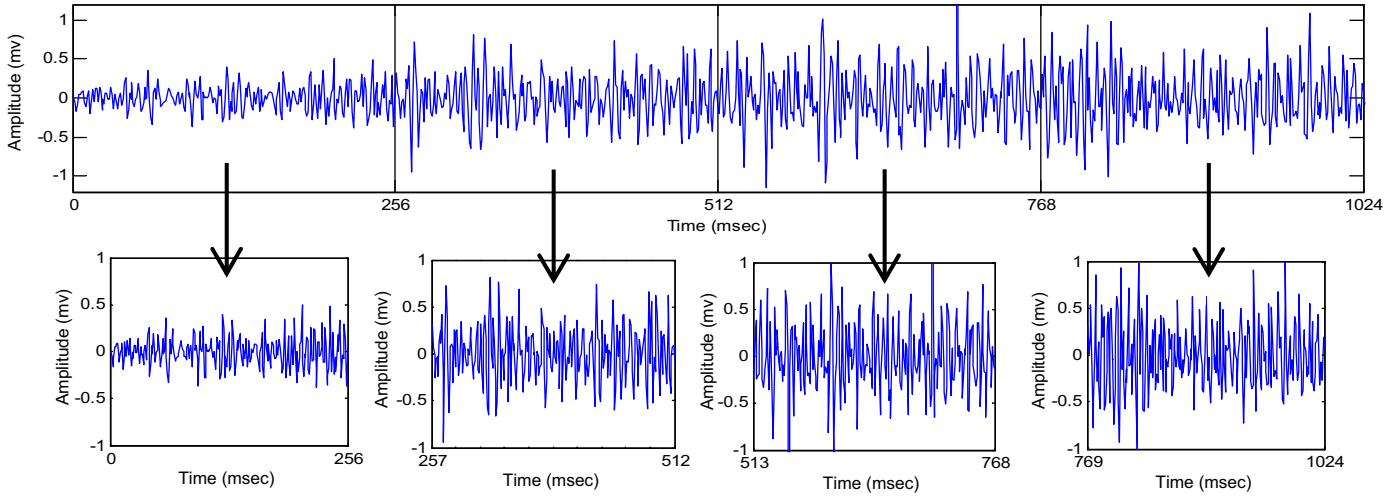

Fig. 2. Segmentation of the filtered signal (the first four segments of clenching molars in channel one).

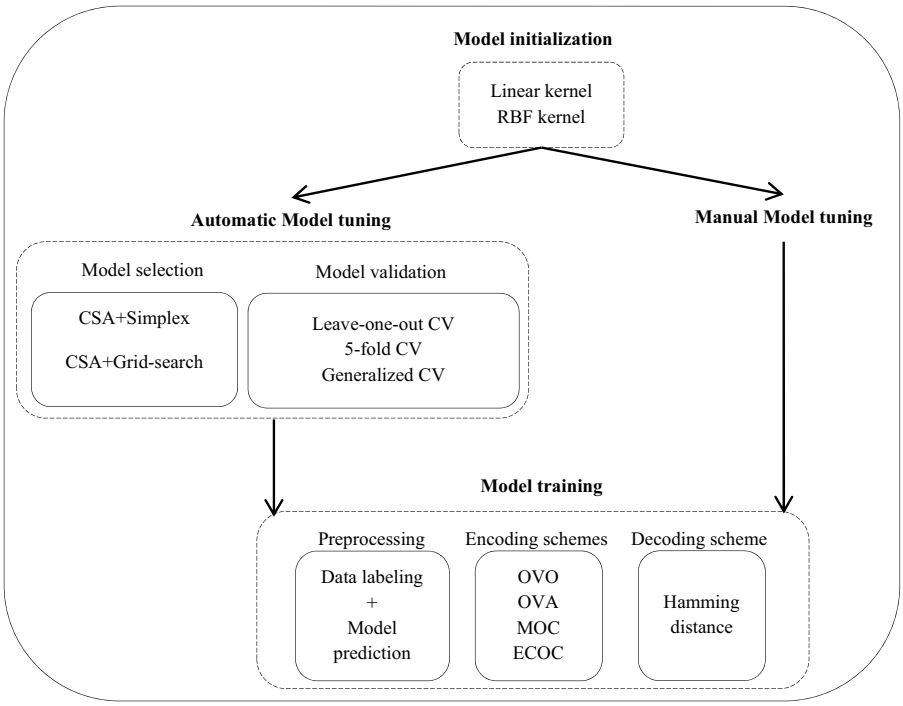

Fig. 3. Stages of LS-SVM model structure.

implemented in this study. As stated earlier, LS-SVM is a non-parametric supervised method with a robust theory for solving non-linear classification problems. Since EMG data are non-stationary and their characteristics vary with time, LS-SVM may perform better in classification of different facial gestures than the conventional supervised classifiers. In order to investigate the behavior and efficiency of LS-SVM on facial gesture classification, different models of this classifier were constructed and examined so as to find a desirable one. As depicted in Fig. 3, the models parameters were at first initialized; then they were tuned manually and automatically; and finally the model training was performed. This procedure was carried out in MATLAB 8.1 by means of LS-SVMlab toolbox version 1.7 [47] as the core of a multiclass LS-SVM classifier. The experimental platform was composed of Intel core i7 CPU, 1.73 GHz with 8GB physical memory.

3.5.1. Model initialization

In this step, LS-SVM models architecture were initialized by setting the type of kernel function, Linear or RBF. Both of these kernels contain regularization (γ) parameter which determines the trade-off between training error minimization and smoothness of the estimated function. Low and high γ emphasize minimizing the model complexity and good fit of the training data points respectively. Smoothing (σ^2) is another RBF kernel parameter which is commonly the squared bandwidth and a larger σ^2 indicates stronger smoothing [48]. These parameters must be optimally tuned in order to deliver accurate performance for LS-SVM which is explained in the next section.

3.5.2. Model tuning

As stated earlier, LS-SVM model parameters need to be tuned properly so as to select the most optimum ones that can deliver the best performance; consequently, some kind of model selection (parameter search) must be done. Global optimization methods are the well-known techniques to estimate the tuning parameters but very slowly. Due to the observed shortcomings, ensuring convergence to a global optimum might be impractical therefore faster convergence techniques could be more suitable solutions. Accordingly, Simulated Annealing (SA) algorithms [49,50] and other heuristic based techniques have been introduced. A typical method is grid-search which searches over a manually specified subset of parameters and performs a cost function for each of the grid values to select the best ones. Another method is Simplex which is well-defined for problems with derivatives that may not be known [51] and can be used when the number of hyper-parameters is quite small. This method finds a local minimum of a function starting from an initial point. The local minimum is located via the Nelder–Mead simplex algorithm which does not require any gradient information. However, this technique is a heuristic search method that can converge to non-stationary points [52]. These two methods often get trapped in poor optima because of the speed-up procedures. A better alternative for tuning parameters is an extension to the well-known SA algorithm [50] with variance control, Coupled Simulating Annealing (CSA) [53]. This method aims to escape from local optima and improve the solution quality without compromising the speed of convergence too much. As a main difference with SA, CSA presents a new form of acceptance probabilities functions to be employed on ensemble of optimizers. This technique considers several current states which are coupled together by their energies in their acceptance function. Moreover, parallelism is an intrinsic feature of CSAs and it is reported that they are more efficient than multi-start gradient descent optimization [54]. Another benefit of CSA is that it controls the variance of the acceptance probabilities via the acceptance temperature. This results in improved optimization efficiency since it reduces the algorithm sensitivity to the initialization parameters while guiding the optimization process to quasi-optimal runs.

In this article, model parameter tuning was carried out manually and automatically. In the former, a wide range of model parameters (γ, σ^2) values were tested manually to pick those which resulted in the best performance. In the latter, the model parameters (γ, σ^2) were tuned in two steps to overcome the drawbacks of the mentioned single optimization techniques. At first, good initial start values were determined by means of CSA with the use of five

Table 1
Different models of LS-SVM considering (a) automatic (b) manual modes of parameter tuning.

No.	Kernel	2nd Optimization	Cost function	Encoding scheme	No.	Kernel	2nd Optimization	Cost function	Encoding scheme
(a)									
1	Linear	Simplex	LOOCV	OVO	25	RBF	Simplex	LOOCV	OVO
2	Linear	Grid Search	LOOCV	OVO	26	RBF	Grid Search	LOOCV	OVO
3	Linear	Simplex	KCV	OVO	27	RBF	Simplex	KCV	OVO
4	Linear	Grid Search	KCV	OVO	28	RBF	Grid Search	KCV	OVO
5	Linear	Simplex	GCV	OVO	29	RBF	Simplex	GCV	OVO
6	Linear	Grid Search	GCV	OVO	30	RBF	Grid Search	GCV	OVO
7	Linear	Simplex	LOOCV	OVA	31	RBF	Simplex	LOOCV	OVA
8	Linear	Grid Search	LOOCV	OVA	32	RBF	Grid Search	LOOCV	OVA
9	Linear	Simplex	KCV	OVA	33	RBF	Simplex	KCV	OVA
10	Linear	Grid Search	KCV	OVA	34	RBF	Grid Search	KCV	OVA
11	Linear	Simplex	GCV	OVA	35	RBF	Simplex	GCV	OVA
12	Linear	Grid Search	GCV	OVA	36	RBF	Grid Search	GCV	OVA
13	Linear	Simplex	LOOCV	MOC	37	RBF	Simplex	LOOCV	MOC
14	Linear	Grid Search	LOOCV	MOC	38	RBF	Grid Search	LOOCV	MOC
15	Linear	Simplex	KCV	MOC	39	RBF	Simplex	KCV	MOC
16	Linear	Grid Search	KCV	MOC	40	RBF	Grid Search	KCV	MOC
17	Linear	Simplex	GCV	MOC	41	RBF	Simplex	GCV	MOC
18	Linear	Grid Search	GCV	MOC	42	RBF	Grid Search	GCV	MOC
19	Linear	Simplex	LOOCV	ECOC	43	RBF	Simplex	LOOCV	ECOC
20	Linear	Grid Search	LOOCV	ECOC	44	RBF	Grid Search	LOOCV	ECOC
21	Linear	Simplex	KCV	ECOC	45	RBF	Simplex	KCV	ECOC
22	Linear	Grid Search	KCV	ECOC	46	RBF	Grid Search	KCV	ECOC
23	Linear	Simplex	GCV	ECOC	47	RBF	Simplex	GCV	ECOC
24	Linear	Grid Search	GCV	ECOC	48	RBF	Grid Search	GCV	ECOC

No.	Kernel	Encoding scheme	No.	Kernel	Encoding Scheme	No.	Kernel	Encoding Scheme	No.	Kernel	Encoding Scheme
(b)											
1	Linear	OVO	3	Linear	MOC	5	RBF	OVO	7	RBF	MOC
2	Linear	OVA	4	Linear	ECOC	6	RBF	OVA	8	RBF	ECOC

multiple starters [55] and then Simplex or Grid-search were performed using the previous result as the start value for fine-tuning (Fig. 3). This procedure led to more optimal tuning parameters and hence better performance. The search limits of the CSA method were set to [exp(−10), exp(10)]. Finally, the performance of the models was examined through three cost functions Leave-one-out cross-validation (LOOCV), K-fold CV (KCV) where $K = 5$, and Generalized CV (GCV) in order to validate the models and select the parameters which delivered the optimum model performance.

3.5.3. Model training

In this study, the whole RMS feature set ($[Ch_1; Ch_2; Ch_3]^T_{3\times 390}$) was shuffled and divided into about 75% and 25% for training ($[Ch_1; Ch_2; Ch_3]^T_{3\times 300}$) and testing ($[Ch_1; Ch_2; Ch_3]^T_{3\times 90}$) stages respectively. Training the LS-SVM model was conducted in three steps (Fig. 3) as follows: at first, the preprocessing assigned a label to the train feature set of each class. For instance, label '1′ was assigned to the first 30 RMSs (indicating class #1, clenching the molars), label '2' for the second 30 RMSs (indicating class #2, raising the eyebrows) and so forth. Then, they were rescaled to the binary labels −1 and +1. After that, RMSs served to predict the LS-SVM model and multi-class problem was decomposed in a set of binary classification problems using encoding techniques OVO, OVA, MOC and ECOC. These encoding schemes resulted in corresponding codebooks to represent each class. Finally, to detect the classes of a disturbed encoded signal given the corresponding codebook, Hamming distance (HD) function was employed to decode the generated code-words into original form.

3.5.4. Model performance evaluation

For classification, different LS-SVM models were formed by considering various methods in each step. This paper aimed to find the best model with the best performance. As shown in Table 1, for automatic and manual modes 48 and 8 models were made respectively. Methods used for preprocessing and decoding scheme in all models were the same and CSA was used as the first optimization technique in all automatic models. The models of automatic mode included orderly kernel function, second optimization technique, cost function and encoding scheme whereas in the manual mode only kernel function and encoding scheme varied in the models.

In order to evaluate the behavior and effectiveness of constructed models for facial gesture recognition, two main performance metrics classification accuracy and computational cost were considered. The accuracy of testing session was computed and finally considered as the classification performance. Moreover, the time consumed during tuning and training stages was measured as the computational cost index. Tuning time was only considered for the models in automatic mode while training time was calculated for both automatic and manual models. All mentioned experiments were statistically analyzed to interpret the results.

4. Results and discussion

This section presents the results achieved by several experiments. At first, the classification accuracy obtained by each model (automatic and manual) averaged over all subjects is reported. Accordingly, the effect of different kernels, optimization techniques and encoding schemes are examined and compared. Consequently, the most accurate model of each mode is identified. The next part investigates the computational load consumed by each model during tuning and training stages and the fastest model of each mode is reported. In the third part, the distributional characteristics of classification accuracies obtained by the selected models across all subjects are inspected in form of box-plots. Finally, the best LS-SVM model is assessed through comparison with SVM, FCM and Fuzzy Gath-Geva clustering (GGFC) methods.

Table 2
Classification accuracy of each model averaged over all subjects in automatic mode.

No.	Acc (%)	No.	Acc (%)	No.	Acc (%)	No.	Acc (%)	No.	Acc (%)
1	86.61 ± 2.21	11	11.38 ± 5.01	21	25.50 ± 4.22	31	83.30 ± 2.85	41	37.76 ± 4.10
2	85.57 ± 2.53	12	13.40 ± 4.41	22	15.65 ± 4.95	32	81.15 ± 2.90	42	17.85 ± 5.45
3	86.68 ± 1.92	13	34.40 ± 3.11	23	30.00 ± 4.56	33	84.45 ± 2.95	43	85.50 ± 3.12
4	84.39 ± 3.56	14	35.50 ± 2.50	24	22.20 ± 4.20	34	83.34 ± 2.12	44	84.45 ± 3.08
5	86.59 ± 2.25	15	30.00 ± 3.38	25	87.75 ± 2.34	35	53.72 ± 3.14	45	83.36 ± 2.79
6	86.43 ± 276	16	31.10 ± 3.10	26	85.97 ± 2.55	36	50.00 ± 4.30	46	73.35 ± 3.21
7	15.35 ± 4.50	17	33.25 ± 4.10	27	88.90 ± 2.16	37	86.55 ± 2.25	47	44.40 ± 4.90
8	11.60 ± 4.90	18	29.80 ± 3.72	28	88.25 ± 2.48	38	86.61 ± 2.40	48	46.60 ± 4.19
9	14.85 ± 4.67	19	24.40 ± 4.15	29	70.00 ± 3.10	39	84.41 ± 2.54		
10	15.17 ± 3.89	20	33.30 ± 3.24	30	78.80 ± 3.23	40	85.53 ± 2.38		

4.1. Classification accuracy

4.1.1. Automatic kernel parameter tuning

In this experiment the classification accuracies achieved by each model considered in Table 1(a) are presented and discussed. Table 2 provides the average of classification accuracies over all subjects while tuning the kernel parameters automatically. This Table indicates that model #27 including RBF kernel function, Simplex optimization techniques and encoding scheme OVO outperformed all other models as it reached the highest degree of accuracy, 88.9%. This means that this model included the most efficient and powerful components for classification of ten facial gestures. In contrast, model #11 constructed by Linear kernel, Simplex optimization function and OVA encoding scheme gained the lowest accuracy at 11.38%.

As can be seen, the results in Table 2 are diverse due to the use of different kernel functions, optimization techniques, cost functions and encoding schemes. The effect of Linear and RBF kernel functions on classification are compared in Fig. 4(a) (each kernel is considered in 24 models). According to this figure, the overall classification performance was significantly higher when using RBF kernel. There were only three models in which Linear kernel outperformed RBF such as the model constructed by grid search, GCV and MOC (i.e. models 18 in Table 2). The impact of two optimization techniques Simplex and Grid-search is inspected and the results are depicted in Fig. 4(b). Classification results indicate that both of these functions had almost similar effect on the parameter tuning session however in some cases Simplex was a bit more efficient. The influence of applying different encoding schemes OVO, OVA, MOC and ECOC on multiclass classification of ten facial gestures is shown in Fig. 4(c). According to Table 1(a), 12 different models were made for each encoding scheme where Linear and RBF kernels were respectively used in the first and the second six models. This graph illustrates that OVO performed better than others for ten-class facial EMG problem as it provided the highest accuracies. A notable point is that the performances of other three schemes were substantially improved when RBF kernel was used which again emphasizes the usefulness of RBF on classification accuracy.

4.1.2. Manual kernel parameters tuning

As stated earlier in Table 1(b), eight models were constructed for LS-SVM to classify facial gestures while tuning the kernel parameters manually. Table 3 reports the average classification accuracy over all subjects for each model. It can be seen that the model #6

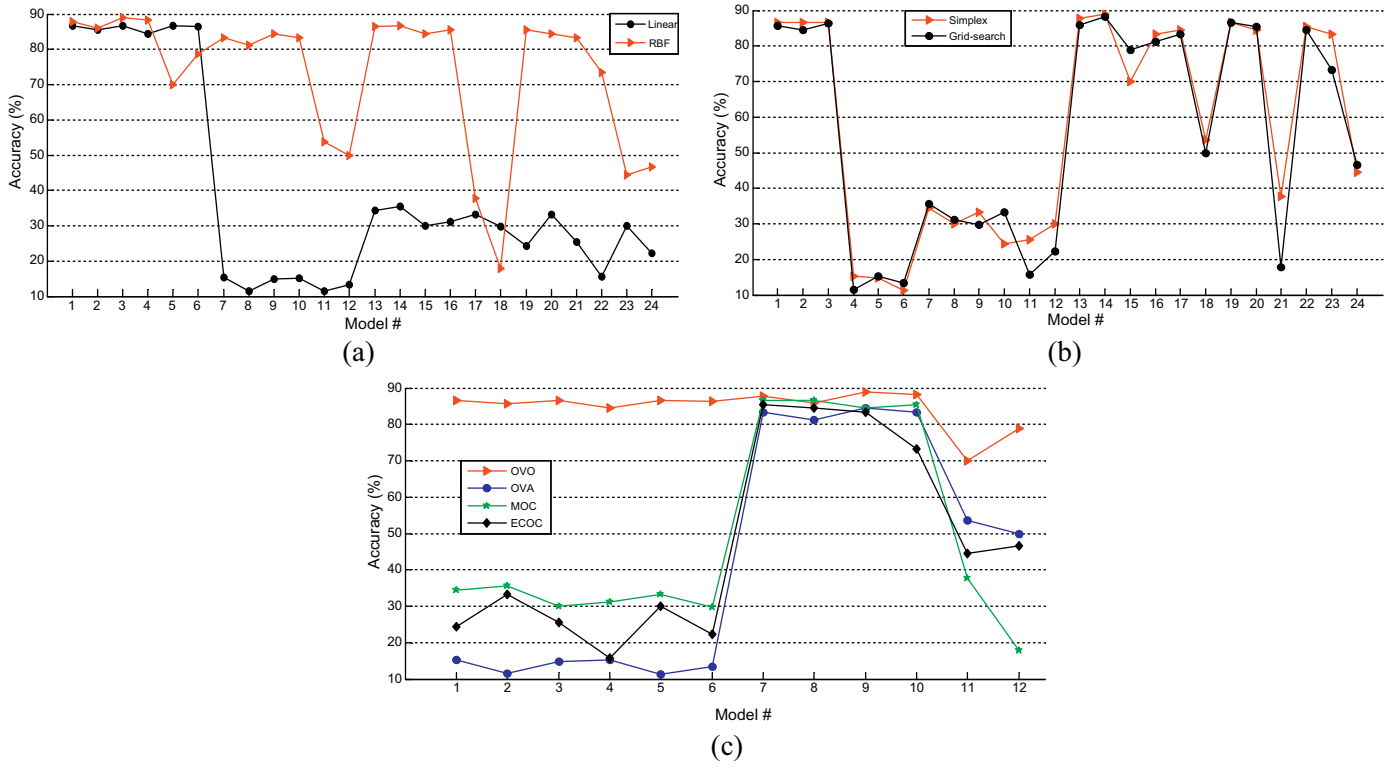

Fig. 4. Comparing the effect of (a) linear and RBF kernels (b) simplex and grid-search optimization techniques (c) different encoding schemes on classification accuracy of all models in automatic mode.

Table 3
The performance of models that their parameters were tuned manually averaged over all subjects in terms of classification accuracy and computational cost.

No.	ACC (%)	Total time (s)	No.	ACC (%)	Total time (s)
1	20.81 ± 5.30	0.53	5	88.75 ± 2.26	0.57
2	15.40 ± 5.92	0.86	6	93.10 ± 1.30	0.98
3	38.75 ± 4.36	0.72	7	91.62 ± 2.01	0.87
4	49.96 ± 3.95	0.68	8	90.51 ± 2.10	0.79

containing RBF kernel and OVA encoding scheme outperformed others by obtaining the highest accuracy rate of 93.1%. In contrast, model #2 including Linear kernel and OVA provided the lowest degree of accuracy (15.40%). In addition, the robustness of RBF in manual mode was confirmed as it delivered much higher classification accuracy in all models. This table also indicates that OVA multiclass encoding scheme outperformed others as it resulted in the best performance.

According to the results reported in this section, RBF kernel offered much better performance than Linear. This is because RBF kernel defines a much larger function space. Moreover, this kernel gives an access to all analytic functions which can deliver much better hyperplanes for different patterns whereas a linear kernel uses linear functions which are really impoverished. Poorer performance of Grid-search optimization technique might be due to its disadvantages such as restriction of the desired number of tuning parameters in a model due to the combinatorial explosion of grid points and incapability of assuring the overall generated solution quality [56,57]. According to previous studies there is no unique encoding scheme for multiclass problems that always performs the best and it completely depends on the application and the number of classes and training samples [58]. Some researchers concluded that OVO is more practical, especially when dealing with more classes [59] while others claimed that OVA is as accurate as other methods if the parameters are well tuned [60]. This statement is in accordance with our results in this study since when the model parameters were tuned automatically OVO resulted in better classification accuracy and through manual tuning OVA performed better. However, results proved that hyperparameters were tuned better in the manual mode.

4.2. Computational load

The purpose of this experiment was to investigate the computational load during tuning the parameters and training the LS-SVM models.

4.2.1. Automatic kernel parameter tuning

Table 4 presents tuning, training and the total time consumed by each model averaged over all subjects. This table indicates that the minimum total time (about 4.45 s) was consumed by model #3 which included Linear kernel, Simplex optimization technique and OVO encoding scheme while the maximum time was 54.32 s for model #32 composed of RBF kernel, Grid-search and OVA techniques. According to these results, the complexity of the methods employed in model #3 was considerably lower than the ones which formed model #32.

As can be observed in Table 4, different kernel functions, optimization methods and encoding schemes influenced the time consumed during tuning and training stages. Fig. 5(a) investigates the impact of Linear and RBF kernel on the total time consumed by each model. The total time was expected to be increased when applying RFB kernel due to its additional smoothing parameter in comparison with Linear kernel where just regularization parameter needed to be tuned. However, Fig. 5(a) demonstrates that using different kernel functions did not affect the computational load meaningfully since no specific manner was observed. The impact of Simplex and Grid-search optimization techniques is studied in Fig. 5(b). Apparently, Simplex required lower computational cost since the total time consumed by most of the models was lower when using this method. However, in some models there were no noticeable differences between these two optimization techniques and both performed similarly. The effectiveness of the encoding schemes on computational load averaged over all subjects for each model is shown in Fig. 5(c). Obviously, OVO is the quickest method to encode ten facial gestures whereas OVA performed too slowly. Besides, the models formed by MOC and ECOC consumed almost the same total time.

4.2.2. Manual kernel parameters tuning

In this part the computational load consumed by each model considered in Table 1(b) is explored. In the manual parameter tuning mode, the total time is equal to training time since the tuning time is zero for all models. The total time averaged over all subjects for each model can be seen in Table 3. It is obvious that the total time did not vary significantly among these models and they only needed less than a second to be trained. However, the model including Linear kernel and OVO encoding scheme performed as the fastest model with 0.53 s. In addition, it is worth noting that

Table 4
The tuning, training and total time consumed by each model averaged over all subjects in automatic mode of parameter tuning.

No.	Tuning time	Training time	Total time	No.	Tuning time	Training time	Total time	No.	Tuning time	Training time	Total time
1	5.97	2.01	7.98	17	17.11	0.67	17.78	33	17.97	1.74	19.71
2	8.92	1.63	10.55	18	17.99	0.88	18.87	34	23.58	1.70	25.28
3	3.14	1.31	4.45	19	19.69	0.92	20.61	35	11.97	2.90	14.87
4	8.30	1.40	9.70	20	23.19	0.96	24.15	36	8.38	3.07	11.45
5	5.39	2.37	7.76	21	4.63	0.92	5.55	37	22.74	0.90	23.64
6	7.87	1.91	9.78	22	6.23	1.21	7.44	38	38.17	0.96	39.13
7	37.12	1.34	38.46	23	22.42	1.37	23.79	39	6.81	1.12	7.93
8	45.50	1.49	46.99	24	21.71	0.98	22.69	40	10.49	0.96	11.45
9	10.11	1.38	11.49	25	5.82	3.15	8.97	41	6.04	1.29	7.33
10	13.70	1.65	15.35	26	9.32	2.96	12.28	42	5.68	1.60	7.28
11	38.21	1.57	39.78	27	3.31	3.40	6.71	43	27.82	1.31	29.13
12	40.30	1.29	41.59	28	6.04	3.63	9.67	44	51.74	1.40	53.14
13	14.70	0.90	15.60	29	2.26	3.86	6.12	45	8.36	1.02	9.38
14	12.96	0.73	13.69	30	2.30	2.90	5.20	46	13.26	1.57	14.83
15	3.96	0.89	4.85	31	27.64	1.73	29.37	47	7.82	1.41	9.23
16	5.10	0.84	5.94	32	52.62	1.70	54.32	48	14.51	1.96	16.47

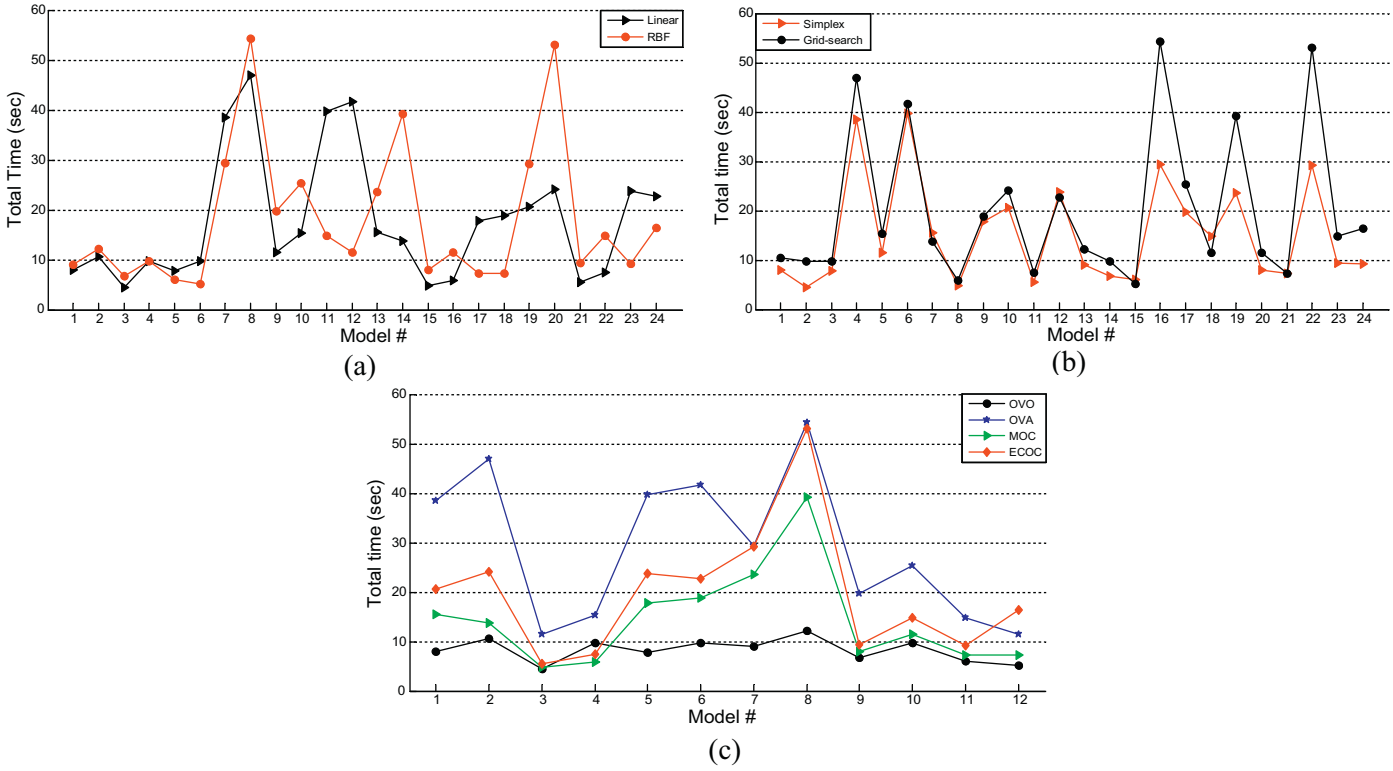

Fig. 5. Comparing the effect of (a) linear and RBF kernels (b) simplex and grid-search optimization techniques (c) different encoding schemes on total time consumed by each model in automatic mode.

the model which reached the highest accuracy consumed 0.98 s for training.

The experimental results reported in this part showed that all manual models needed less time compared to the automatic models. The minimum time consumed in automatic and manual modes was 4.45 s (model #3) and 0.53 s (model #1) respectively. It was shown that among all factors, encoding schemes played the major role in imposing the computational load on the system. Our results indicated that OVO strategy was substantially faster in both modes of automatic and manual whereas OVA performed as the slowest scheme.

4.3. Statistical analysis

According to the results stated in previous sections, the most accurate automatic and manual models were #27 (Table 2) and #6 (Table 3) with 88.9% and 93.1% classification accuracy respectively. In this experiment the distributional characteristics of these models across all subjects are evaluated and compared (Fig. 6). As can be seen, model #6 outperformed model #27 since the overall range of classification accuracy was greatly higher. Interquartile ranges indicate lower degree of dispersion when data of different subjects were classified by manual mode of LS-SVM. Model #6 is formed in a short box which specifies the close classification accuracies gained for all subjects. In contrast, longer spread of classification accuracies in model #27 reveals its high sensitivity over different subjects. Regardless of the outlier in model #6, symmetric boxes for both models point out that the accuracies obtained for different subjects were split evenly at the median. In conclusion, this analytical comparison proved a better performance of the chosen manual model across different subjects as its median and upper whisker are also placed above the automatic model.

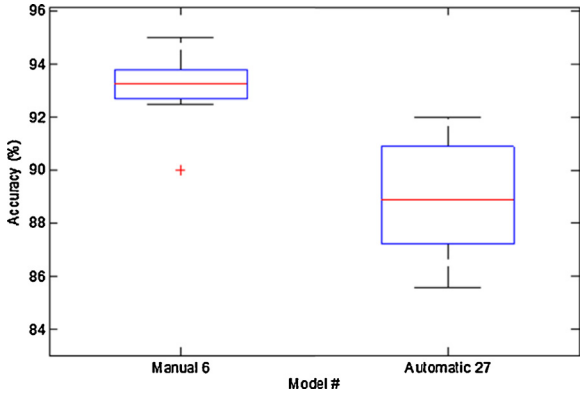

Fig. 6. Analytical comparisons of selected LS-SVM models over all subjects.

4.4. Method evaluation by a comparative study

The final experiment evaluates the robustness of the best model of LS-SVM (manual model #6) through comparison with popular classifiers SVM, FCM and GGFC in terms of classification accuracy and computational cost. Results in Table 5 indicate that LS-SVM outperformed other classifiers as it obtained the highest accuracy. It is also shown that LS-SVM performed almost as fast as SVM (about

Table 5
Classification accuracy and training time by LS-SVM, SVM, FCM and GGFC classifiers averaged over all subjects.

Classifier	Accuracy (%)	Training time (s)
LS-SVM	93.10 ± 1.30	0.98
SVM	85.50 ± 2.85	0.89
FCM	90.41 ± 3.12	2.43
FGGC	91.82 ± 2.71	0.72

0.9 s) but it consumed more time than FGGC for classifying the facial gestures EMGs. As mentioned before, the vital criterion in MuCI real-time applications is a reliable trade-off between accuracy and speed. Therefore, LS-SVM is still the preferred algorithm since it met this requirement by providing robust accuracy and satisfactory speed.

In this article, the capability of LS-SVM to classify ten different facial gesture EMGs was examined and the robustness of this method was approved. The results also indicated that the manual tuning of the kernel parameters was a better alternative because of higher classification accuracy and lower computational cost which satisfied the requirements of MuCI real-time applications. On the other hand, the automatic mode of LSSVM not only required too much time but was also unable to improve the system performance. Besides, RBF kernel function was preferred to Linear as it incredibly enhanced the system performance. Although OVO performed a bit faster, OVA was more suitable to solve multiclass facial gesture classification by delivering the highest degree of classification accuracy in the manual mode; so, it was selected as the best encoding scheme. As a fine parameters tuning step in automatic mode, optimization technique Simplex performed better than Grid-Search. In comparison with our previous study in which the same number of classes was considered [24], the system performance was improved since the classification accuracy increased about 6% and training time remained below 1 s.

5. Conclusion

In this paper, a neuromuscular-based facial gesture recognition system was proposed as an interacting pathway to be used in MuCI technology. EMG signals were recorded from ten subjects while gesturing ten different facial states. The RMS features were extracted from filtered signals to characterize the most important properties of facial EMGs. LS-SVM was employed to classify the facial EMG features while an in-depth investigation was carried out in order to find the best model of this classifier. For this purpose, automatic and manual modes of Linear and RBF kernel parameters tuning were explored. In the former, Simplex and Grid-search optimization techniques, OVO, OVA, MOC and ECOC encoding schemes and KCV, GCV and LOOCV cost functions were utilized to construct 48 models. In the latter, 8 models were formed by using the considered kernel functions and encoding schemes. All these models were examined to find the one that provided a reliable trade-off between classification accuracy and computational load. The experimental results showed that the manual LS-SVM model including RBF kernel and OVA encoding technique performed as the best model by providing 93.10% accuracy while consuming 0.98 s for training. The reliability and robustness of our finding was validated through being compared with SVM, FCM and GGFC classifiers.

As mentioned, facial EMG characteristics vary within different muscle movements and different subjects. Although the former is a beneficial phenomenon and offers a possibility of classification of different gestures, the latter causes crucial problems in the analysis of EMG and interpretation of the results even during the isotonic muscle movements. In this study, inter-subject variation was reduced and partially controlled through some precautions such as consecutive recording sections by keeping the electrodes over the same location. In future, other techniques will be implemented and examined to enhance the facial signal quality and diminish the variability of EMGs like robust denoising and signal normalization. Furthermore, the usefulness of LS-SVM for online facial neuromuscular signal classification will be investigated for real-time MuCI applications. In addition, facial EMG signals from disabled subjects will be studied.

Acknowledgements

This research project is supported by Center of Biomedical Engineering (CBE), Transport Research Alliance, Universiti Teknologi Malaysia Research University Grant (GUP Tier 1: Q.J130000.2545.04H21 and Q.J130000.2545.06H28) and funded by Ministry of Higher Education (MOHE).

References

[1] I. Moon, M. Lee, J. Ryu, M. Mun, Intelligent robotic wheelchair with EMG-, gesture-, and voice-based interfaces, in: Proceedings of 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003), vol. 4, IEEE, 2003, pp. 3453–3458, http://dx.doi.org/10.1109/IROS.2003.1249690.

[2] I.M. Rezazadeh, M. Firoozabadi, H. Hu, S.M.R.H. Golpayegani, Co-adaptive and affective human–machine interface for improving training performances of virtual myoelectric forearm prosthesis, IEEE Trans. Affect. Comput. 3 (3) (2012) 285–297, http://dx.doi.org/10.1109/T-AFFC.2012.3.

[3] A. Ferreira, R.L. Silva, W.C. Celeste, T.F. Bastos Filho, M. Sarcinelli Filho, Human–machine interface based on muscular and brain signals applied to a robotic wheelchair, J. Phys. 90 (1.) (2007), http://dx.doi.org/10.1088/1742-6596/90/1/012094.

[4] K. Englehart, B. Hudgin, P.A. Parker, A wavelet-based continuous classification scheme for multifunction myoelectric control, IEEE Trans. Biomed. Eng. 48 (3) (2001) 302–311, http://dx.doi.org/10.1109/10.914793.

[5] K. Englehart, B. Hudgins, A robust, real-time control scheme for multifunction myoelectric control, IEEE Trans. Biomed. Eng. 50 (7) (2003) 848–854, http://dx.doi.org/10.1109/TBME.2003.813539.

[6] P. Shenoy, K.J. Miller, B. Crawford, R.P. Rao, Online electromyographic control of a robotic prosthesis, IEEE Trans. Biomed. Eng. 55 (3) (2008) 1128–1135, http://dx.doi.org/10.1109/TBME.2007.909536.

[7] H. Yan, R.P. Han, Y. Wang, J. Chi, Controlling a powered exoskeleton system via electromyographic signals, in: 2009 IEEE International Conference on Robotics and Biomimetics (ROBIO), IEEE, 2009, pp. 349–353, http://dx.doi.org/10.1109/ROBIO.2009.5420670.

[8] J.S. Han, Z. Zenn Bien, D.J. Kim, H.E. Lee, J.S. Kim, Human–machine interface for wheelchair control with EMG and its evaluation, in: Proceedings of the 25th Annual International Conference on Engineering in Medicine and Biology Society, vol. 2, IEEE, 2003, pp. 1602–1605, http://dx.doi.org/10.1109/IEMBS.2003.1279672.

[9] I. Moon, M. Lee, J. Chu, M. Mun, Wearable EMG-based HCI for electric-powered wheelchair users with motor disabilities, in: Proceedings of the 2005 IEEE International Conference on Robotics and Automation (ICRA 2005), IEEE, 2005, pp. 2649–2654, http://dx.doi.org/10.1109/ROBOT.2005.1570513.

[10] B. Crawford, K. Miller, P. Shenoy, R. Rao, Real-time classification of electromyographic signals for robotic control, in: Proceedings of the National Conference on Artificial Intelligence, vol. 20(2), AAAI Press/MIT Press, Menlo Park, CA/Cambridge MA/London, 2005, pp. 523–528.

[11] M.C. Carozza, G. Cappiello, G. Stellin, F. Zaccone, F. Vecchi, S. Micera, P. Dario, On the development of a novel adaptive prosthetic hand with compliant joints: experimental platform and EMG control, in: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2005), IEEE, 2005, pp. 1271–1276, http://dx.doi.org/10.1109/IROS.2005.1545585.

[12] S.M.P. Firoozabadi, M.A. Oskoei, H. Hu, A human–computer interface based on forehead multi-channel bio-signals to control a virtual wheelchair, in: Proceedings of the 14th Iranian Conference on Biomedical Engineering (ICBME), 2008, pp. 272–277.

[13] L.B.P. Ang, E.F. Belen, R.A. Bernardo Jr., E.R. Boongaling, G.H. Briones, J.B. Coronel, Facial expression recognition through pattern analysis of facial muscle movements utilizing electromyogram sensors, in: 2004 IEEE Region 10 Conference on TENCON 2004, vol. 100, IEEE, 2004, pp. 600–603, http://dx.doi.org/10.1109/TENCON.2004.1414843.

[14] K.H. Kim, J.K. Yoo, H.K. Kim, W. Son, S.Y. Lee, A practical biosignal-based human interface applicable to the assistive systems for people with motor impairment, IEICE Trans. Inf. Syst. 89 (10) (2006) 2644–2652, http://dx.doi.org/10.1093/ietisy/e89-d.10.2644.

[15] G. Gibert, M. Pruzinec, T. Schultz, K. Stevens, Enhancement of human computer interaction with facial electromyographic sensors, in: Proceeding of the 21st Annual Conference of the Australian Computer Human Interaction Special Interest Group on Design Open 247 OzCHI 09, Melborn, Australia, ACM Press, 2009, pp. 1–4, http://dx.doi.org/10.1145/1738826.1738914.

[16] S. Arjunan, D.K. Kumar, Recognition of facial movements and hand gestures using surface electromyogram (sEMG) for HCI based applications, in: 9th Biennial Conference of the Australian Pattern Recognition Society on Digital Image Computing Techniques and Applications, IEEE, 2007, pp. 1–6, http://dx.doi.org/10.1109/DICTA.2007.4426768.

[17] I.M. Rezazadeh, X. Wang, R. Wang, M. Firoozabadi, Toward affective handsfree human-machine interface approach in virtual environments-based equipment operation, in: Training 9th International Conference on Construction Applications on Virtual Reality (CONVR2009), Sydney, Australia, 2009.

[18] I.M. Rezazadeh, M. Firoozabadi, H. Hu, S.M.R. Hashemi Golpayegani, A novel human-machine interface based on recognition of multi-channel facial

bioelectric signals, Australas. J. Phys. Eng. Sci. Med. 34 (4) (2011) 497–513, http://dx.doi.org/10.1007/s13246-011-0113-1.
[19] M. Hamedi, I.M. Rezazadeh, M. Firoozabadi, Facial gesture recognition using two-channel biosensors configuration and fuzzy classifier: a pilot study, in: International Conference on Electrical, Control and Computer Engineering (INECCE), Pahang, Malaysia, 2011, pp. 338–343, http://dx.doi.org/10.1109/INECCE.2011.5953903.
[20] M. Hamedi, Sh-H. Salleh, T.S. Tan, A. Kamarul, Surface electromyography-based facial expression recognition in bipolar configuration, J. Comput. Sci. 7 (9) (2011) 1407–1415, http://dx.doi.org/10.3844/jcssp.2011.1407.1415.
[21] M. Hamedi, Sh-H. Salleh, T.S. Tan, A. Kamarul, A. Jalil, C. Dee-Uam, C. Pavaganun, P.P. Yupapin, Human facial neural activities and gesture recognition for machine interfacing applications, Int. J. Nanomed. 6 (2011) 3461–3472, http://dx.doi.org/10.2147/IJN.S26619.
[22] M. Hamedi, Sh-H. Salleh, A.M. Noor, T.S. Tan, A. Kamarul, Comparison of different time-domain feature extraction methods on facial gestures' EMGs, in: Progress in Electromagnetics Research Symposium Proceedings, 27–30 March, KL, Malaysia, 2012, pp. 1897–1900, http://dx.doi.org/10.1186/1475-925X-12-73.
[23] M. Hamedi, Sh-H. Salleh, M. Astaraki, A.M. Noor, A.R.A. Harris, Comparison of multilayer perceptron and radial basis function neural networks for EMG-based facial gesture recognition, in: The 8th International Conference on Robotic, Vision Signal Processing & Power Applications, Springer, Singapore, 2014, pp. 285–294, http://dx.doi.org/10.1007/978-981-4585-42-2_33.
[24] M. Hamedi, Sh-H. Salleh, M. Astaraki, A.M. Noor, EMG-based facial gesture recognition through versatile elliptic basis function neural network, Biomed. Eng. 12 (1) (2013) 73, http://dx.doi.org/10.1186/1475-925X-12-73, Online.
[25] M.A. Oskoei, H. Hu, Support vector machine-based classification scheme for myoelectric control applied to upper limb, IEEE Trans. Biomed. Eng. 55 (8) (2008) 1956–1965, http://dx.doi.org/10.1109/TBME.2008.919734.
[26] B. Schölkopf, A. Smola, Learning with Kernels, The MIT Press, Cambridge, 2002.
[27] J.A. Suykens, J. Vandewalle, Least squares support vector machine classifiers, Neural Process. Lett. 9 (3) (1999) 293–300, http://dx.doi.org/10.1023/A:1018628609742.
[28] O. Chapelle, V. Vapnik, O. Bousquet, S. Mukherjee, Choosing multiple parameters for support vector machines, Mach. Learn. 46 (1–3) (2002) 131–159, http://dx.doi.org/10.1023/A:1012450327387.
[29] V. Cherkassky, Y. Ma, Practical selection of SVM parameters and noise estimation for SVM regression, Neural Netw. 17 (1) (2004) 113–126, http://dx.doi.org/10.1016/S0893-6080(03)00169-2.
[30] F. Friedrichs, C. Igel, Evolutionary tuning of multiple SVM parameters, Neurocomputing 64 (2005) 107–117, http://dx.doi.org/10.1016/S0925-2312(05)00029-9.
[31] A. Sengur, Multiclass least-squares support vector machines for analog modulation classification, Exp. Syst. Appl. 36 (3) (2009) 6681–6685, http://dx.doi.org/10.1016/j.eswa.2008.08.066.
[32] V. Mitra, C.J. Wang, S. Banerjee, Text classification: a least square support vector machine approach, Appl. Soft Comput. 7 (3) (2007) 908–914, http://dx.doi.org/10.1016/j.asoc.2006.04.002.
[33] E.D. Übeyli, Least squares support vector machine employing model-based methods coefficients for analysis of EEG signals, Exp. Syst. Appl. 37 (1) (2010) 233–239, http://dx.doi.org/10.1016/j.eswa.2005.04.011.
[34] C.A. Lima, A.L. Coelho, M. Eisencraft, Tackling EEG signal classification with least squares support vector machines: a sensitivity analysis study, Comput. Biol. Med. 40 (8) (2010) 705–714, http://dx.doi.org/10.1016/j.compbiomed.2010.06.005.
[35] V. Vapnik, The Nature of Statistical Learning Theory, Springer-Verlag, New York, 1995.
[36] C. Cortes, V. Vapnik, Support-vector networks, Mach. Learn. 20 (3) (1995) 273–297, http://dx.doi.org/10.1023/A:1022627411411.
[37] T.G. Dietterich, G. Bakiri, Solving multiclass learning problems via error-correcting output codes, J. Artif. Intell. Res. 2 (1995) 263–286.
[38] T. Evgeniou, M. Pontil, T. Poggio, Regularization networks and support vector machines, Adv. Comput. Math. 13 (1) (2000) 1–50, http://dx.doi.org/10.1023/A:1018946025316.
[39] G. Wahba, Spline models for observational data, in: CBMS-NSF Regional Conference Series in Applied Mathematics, vol. 59, SIAM PA, 1990.
[40] D. Tsujinishi, S. Abe, Fuzzy least squares support vector machines for multiclass problems, Neural Netw. 16 (5) (2003) 785–792, http://dx.doi.org/10.1016/S0893-6080(03)00110-2.
[41] I.M. Rezazadeh, M. Firoozabadi, H. Hu, M.R. Hashemi Golpayegani, Determining the surface electrodes locations to capture facial bioelectric signals, Iran. J. Med. Phys. 7 (2010) 65–79.
[42] S. Day, Important Factors in Surface EMG Measurement, Bortech Biomedical Ltd., Calgary, 2002.
[43] K.G. Sawarkar, Analysis and inference of EMG using FFT, in: Proceeding of SPIT-IEEE Colloquium and International Conference, Mumbai, India, 2007.
[44] F. Andris, Biomechanics of the Upper Limbs Mechanics Modelling and Musculoskeletal Injuries, 1st ed., CRC Press, London, UK, 2004.
[45] M. Rosenblum, Y. Yacoob, L.S. Davis, Human expression recognition from motion using a radial basis function network architecture, IEEE Trans. Neural Netw. 7 (5) (1996) 1121–1138, http://dx.doi.org/10.1109/72.536309.
[46] K. Momen, S. Krishnan, T. Chau, Real-time classification of forearm electromyographic signals corresponding to user-selected intentional movements for multifunction prosthesis control, IEEE Trans. Neural Syst. Rehabil. Eng. 15 (4) (2007) 535–542, http://dx.doi.org/10.1109/TNSRE.2007.908376.
[47] LS-SVMlab version 1.7. http://www.esat.kuleuven.be/sista/lssvmlab/
[48] K. De Brabanter, P. Karsmakers, F. Ojeda, C. Alzate, De J. Brabanter, K. Pelckmans, B. De Moor, et al., LS-SVMlab Toolbox User's Guide, 2011.
[49] L. Ingber, Very fast simulated re-annealing, J. Math. Comput. Model. 12 (1989) 967–973, http://dx.doi.org/10.1016/0895-7177(89)90202-1.
[50] S. Rajasekaran, On simulated annealing and nested annealing, J. Glob. Optim. 16 (2000) 43–56, http://dx.doi.org/10.1023/A:1008307523936.
[51] J.A. Nelder, R. Mead, A simplex method for function minimization, Comput. J. 7 (1965) 308–313, http://dx.doi.org/10.1093/comjnl/7.4.308.
[52] M.J. Powell, On search directions for minimization algorithms, Math. Program. 4 (1) (1973) 193–201, http://dx.doi.org/10.1007/BF01584660.
[53] S. Xavier de Souza, J.A.K. Suykens, J. Vandewalle, D. Bollé, Coupled simulated annealing, IEEE Trans. Syst. Man Cybern. B 40 (2) (2010) 320–336, http://dx.doi.org/10.1109/TSMCB.2009.2020435.
[54] J.A.K. Suykens, J. Vandewalle, B. De Moor, Intelligence and cooperative search by coupled local minimizers, Int. J. Bifurc. Chaos 11 (8) (2001) 21332144, http://dx.doi.org/10.1142/S0218127401003371.
[55] S. Xavier de Souza, J.A.K. Suykens, J. Vandewalle, D. Bollé, Coupled simulated annealing, IEEE Trans. Syst. Man Cybern. B 40 (2) (2010) 320–335, http://dx.doi.org/10.1109/TSMCB.2009.2020435.
[56] K.P. Bennett, J. Hu, X. Ji, G. Kunapuli, J.S. Pang, Model selection via bilevel optimization, Int. Joint Conf. Neural Netw. (2006) 1922–1929, http://dx.doi.org/10.1109/IJCNN.2006.246935.
[57] C.-M. Huang, Y.-J. Lee, D.K.J. Lin, S.-Y. Huang, Model selection for support vector machines via uniform design, Comput. Stat. Data Anal. 52 (1) (2007) 335–346, http://dx.doi.org/10.1016/j.csda.2007.02.013.
[58] J. Milgram, M. Cheriet, R. Sabourin, "One against one" or "one against all": which one is better for handwriting recognition with SVMs? in: Tenth International Workshop on Frontiers in Handwriting Recognition, 2006.
[59] C.W. Hsu, C.J. Lin, A comparison of methods for multiclass support vector machines IEEE Trans. Neural Netw. 13 (2) (2002) 415–425, http://dx.doi.org/10.1109/72.991427.
[60] R. Rifkin, A. Klautau, In defense of one-vs-all classification, J. Mach. Learn. Res. 5 (2004) 101–141, http://dx.doi.org/10.1109/72.991427.