Swarm-wavelet based Extreme Learning Machine for Finger Movement Classification on Transradial Amputees

Khairul Anam, *Student Member, IEEE* , Adel Al-Jumaily, *Senior Member, IEEE*

***Abstract*— The use of a small number of surface electromyography (EMG) channels on the transradial amputee in a myoelectric controller is a big challenge. This paper proposes a pattern recognition system using an extreme learning machine (ELM) optimized by particle swarm optimization (PSO). PSO is mutated by wavelet function to avoid trapped in a local minima. The proposed system is used to classify eleven imagined finger motions on five amputees by using only two EMG channels. The optimal performance of wavelet-PSO was compared to a grid-search method and standard PSO. The experimental results show that the proposed system is the most accurate classifier among other tested classifiers. It could classify 11 finger motions with the average accuracy of about 94 % across five amputees.**

I. Introduction

ELM is a vast improvement of feed-forward neural networks, which remarkably save the training time by omitting an iterative learning process. In ELM, the hidden node weights and biases are determined randomly while the output weights are calculated analytically. Therefore, the training time is enormously fast compared to the traditional neural networks. ELM method has been used in a wide range of applications [1].

Nevertheless, the hidden node parameters, the input weights and biases which are arbitrarily defined, result in a non-optimal system. Some efforts have done to deal with such an optimization problem. Self-adaptive evolutionary ELM (SAE-ELM) [1], and PSO-ELM [2] are a number of methods developed to optimize the hidden node parameters.

ELM is not merely working on a node style. A kernel system can be incorporated in ELM by replacing the node processing structure with a kernel function. This kernel-based ELM is considered as a special type of least-square-support vector machine (LS-SVM) without using output bias [3]. Similar to the node based ELM, the kernel based ELM faces an optimization problem too. The efficacy of the ELM greatly depends on the optimum combination of the kernel parameters [4]. The popular grid-search algorithm which is simple and direct has been used to search the optimal kernel parameters for several years [5]. However, the exhaustive grid search on a large number of the parameter space may result in time consuming process.

Khairul Anam is with University of Jember Indonesia. Now he is a PhD student at University of Technology Sydney, P.O. Box 123 Broadway, NSW 2007 Australia (Phone: (+61)295147962; fax: - ; e-mail: Khairul.Anam@student.uts.edu.au).

Adel Al-Jumaily is an associate professor with School of Electrical, Mechanical and Mechatronic Systems, University of Technology, Sydney (e-mail: Adel.Al-Jumaily@uts.edu.au).

A particle swarm optimization (PSO) algorithm can be a promising solution for optimizing the kernel parameters in the kernel-based ELM. Nevertheless, PSO tends to get trapped in the local minima. One solution that can used to deal with a local minimal problem in PSO is by mutating the swarm particles using a wavelet function [6]. This work proposes swarm-wavelet based ELM, an optimization of the kernel-based ELM using hybridization of PSO and wavelet.

On the other hand, in reality, the amputees experience different limb amputation. As a result, it is difficult to place electrodes on the forearm for EMG signal acquisition. This circumstance can be solved by involving a limited number of electrodes without compromising the classification performance. Recently, the use of two EMG channels in the myoelectric-pattern-recognition system has been investigated. Khushaba *et al*. [7] utilized two EMG channels to recognize ten finger movements with accuracy of approximately 92 %. Moreover, Anam *et al*. [5] succeeded to classify ten finger motions by accuracy of roughly 98 %.

However, those pattern-recognition systems were implemented only on the able-bodied subjects. To the author's knowledge, no one has employed two EMG channels on the amputees. Utilizing a small number of EMG channels on the amputee is a challenge. For that reason, a powerful classifier should be developed and implemented. This work proposes a swarm-wavelet-based ELM for classification of the finger motions on the amputee subjects.

II. Methods

A. Proposed Method

This work utilized a state-of-the art of pattern recognition method for EMG signal. It comprised data acquisition process, filtering and windowing on the collected data, and feature extraction using a combination of time-domain features and autoregressive parameters. Moreover, the system also consisted of dimensionality reduction using spectral regression discriminant analysis (SRDA), an extension of linear discriminant analysis (LDA), and classification using swarm-wavelet ELM.

B. Data Collection

This work utilized data collected in [8] recorded from five transradial amputees aged 25-35 years old. The demographic of the amputees is presented in Table I. Eleven pairs of self-adhesive Ag-AgCl electrodes forming 11 electrode pairs were located on the forearm of the amputee subjects with different levels of transradial amputation [8].

978-1-4244-7929-0/14/$26.00 ©2014 IEEE

TABLE I. DEMOGRAPHIC OF THE AMPUTEE INVOLVED IN THE EXPERIMENT

ID	Age (year)	Missing hand	Dominant hand	Stump length (cm)	Stump circumference (cm)	Time since amputation (year)
A1	25	Left	Right	13	27	4
A2	33	Left	Right	18	24	6
A3	27	Left	Right	16	23	4
A4	35	Left	Right	23	26	8
A5	29	Left	Right	24	26	7

Data were recorded using a custom-built multichannel EMG acquisition device developed in [8]. It consists of a 1000-gain-factor amplifier for each channel and two analog filters (a fourth-order Butterworth low-pass filter with the cut-off frequency of 450 Hz and a second-order Butterworth high-pass filter with a cut-off frequency of 10 Hz). Furthermore, the interface employed a USB data acquisition device (USB-6210 of National Instruments) with sample rate of 2000 Hz and 16-bit resolution. In addition, two digital filters, a pass-band frequency 20-450 Hz and a fifth-order Butterworth notch filter at 50 Hz were also implemented. Acquired EMG signals were stored and displayed in a PC that run LABVIEW software from National Instruments.

The amputee subjects were asked to imagine performing eleven individual finger movements plus one rest state (R). The individual finger movements consisted of a thumb abduction (Ta), thumb flexion (Tf), index finger flexion (If), middle finger flexion (Mf), ring finger flexion (Rf), little finger flexion (Lf), thumb finger extension (Te), index finger extension (Ie), middle finger extension (Me), ring finger extension (Re), and a little finger extension (Le).

C. Features extraction

Sixteen features were extracted from Time domain features (TD) and Autoregressive features (AR). The combination comprises mean absolute value (MAV), mean absolute value slope (MAVS), zero crossings (ZC), slope sign changes (SSC), waveform length (WL), sample skewness (SS), root mean square (RMS), Hjorth time domain parameters (HTD), and six-order autoregressive (AR) model.

All features extracted from all EMG channels are concatenated to create a large feature set. As a result, the dimension of the feature set is enormous. Then the number of features were reduced using Spectral Regression Discriminant Analysis (SRDA) [9]. In SRDA, the feature set is reduced and projected to c-1 features where c is the number of classes. Furthermore, aforementioned features were segmented by using a sliding window with the length of 200 ms shifted by 25 ms. The period was selected in order to meet the optimal window length [10] and the optimal controller delay time [11].

D. Extreme Learning Machine

Huang *et al*. [4] presented ELM as a generalization of single-hidden-layer feed-forward networks (SLFNs). One of the attractive features of the ELM is that the hidden layer does not need to be tuned, and its nodes implement a random computational process which is independent of the training data. In training mode, the aim of ELM is to attain the smallest training error and norm of output weights, which is different from traditional learning algorithm of SLFNs. In addition, if the feature mapping of the hidden node is unknown to the user, several kernel functions can be used [4].

For N samples $\{(x_j, y_j)\}_{j=1}^{N}$ where input x_j = $[x_{j1}, x_{j2}, ..., x_{jm}]^T$ and target y_j = $[y_{j1}, y_{j2}, ..., y_{jn}]^T$, the output of a standard SLFN with M hidden neurons is

$$f_j(x) = \sum_{i=1}^{M} \beta_i g(\mathbf{w}_i \cdot x_j + b_i) = \mathbf{G}\beta \quad j = 1, ..., N \quad (1)$$

where $\mathbf{w}_i = [w_{i1}, ..., w_{im}]^{\mathrm{T}}$ denotes the vector of the weight linking the *i*th hidden neuron and the input neurons. Moreover, $\beta_i = [w_{i1}, ..., w_{in}]^{\mathrm{T}}$ defines the weight vector of the *i*th hidden neuron, $f_j(x) = [f_{j1}, ..., f_{jm}]^{\mathrm{T}}$ is the output vector of SLFN, b_i is the threshold of the *i*th hidden neuron and *g(x)* is the activation function of the hidden node. The right part of equation 1 is the compact form of SLFN output where

$$\mathbf{G} = \begin{bmatrix} g(w_1 \cdot x_1 + b_1) & \cdots & g(w_M \cdot x_1 + b_M) \\ \vdots & \vdots & \vdots \\ g(w_1 \cdot x_N + b_1) & \vdots & g_M(w_M \cdot x_N + b_M) \end{bmatrix}_{NxM} \quad (2)$$

and

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_M^T \end{bmatrix}_{Mxm} \quad (3)$$

In ELM, the input weights w_i and the biases b_i are assigned randomly while the output weights are calculated analytically using the following equation:

$$\beta = \mathbf{G}^{\dagger}\mathbf{T} \quad (4)$$

where $\mathbf{G}^{\dagger}$ is the Moore-Penrose generalized inverse of the matrix G. Furthermore, Huang *et al.* [4] also introduced the optimization problem of ELM for the multi-class classifier in such a way that the equation (4) becomes:

$$\boldsymbol{\beta} = \mathbf{G}^{\mathrm{T}}\left(\frac{\mathbf{I}}{C} + \mathbf{G}\mathbf{G}^{\mathrm{T}}\right)^{-1}\mathbf{T} \quad (5)$$

where C is a user-specified parameter. Eventually, the output function of SLFN in the equation (1) can be modified to become:

$$f(x) = \mathbf{g(x)}\boldsymbol{\beta} = \mathbf{g(x)}\mathbf{G}^{\mathrm{T}}\left(\frac{\mathbf{I}}{C} + \mathbf{G}\mathbf{G}^{\mathrm{T}}\right)^{-1}\mathbf{T} \quad (6)$$

where g(x) is a feature mapping (hidden layer output vector) which can be presented by:

$$\mathbf{g(x)} = [G(a_1, b_1, x), \cdots, G(a_L, b_L, x)] \quad (7)$$

In (7), G is a non-linear piecewise continuous function such as a sigmoid, hard limit, Gaussian, and multi quadratic function. If the feature mapping h(x) is not known, a kernel function can be used to represent h(x). Then, the equation (6) would be:

$$f(x) = \begin{bmatrix} K(\mathrm{x},\mathrm{x}_1) \\ \vdots \\ K(\mathrm{x},\mathrm{x}_M) \end{bmatrix}^{\mathrm{T}} \left(\frac{\mathbf{I}}{\mathrm{C}} + \Omega_{ELM} \right)^{-1} \mathbf{T} \quad (8)$$

where

$$\Omega_{ELM} = \mathbf{G}\mathbf{G}^{\mathrm{T}} : \Omega_{ELM\,i,j} = g(\mathrm{x}_i)\cdot g(\mathrm{x}_j) = K(\mathrm{x}_i,\mathrm{x}_j) \quad (9)$$

In the above equation, K is a kernel function like a radial basis function as shown in (10).

$$K(\mathbf{x}_i,\mathbf{x}_j) = \exp\left(-\gamma \left\|\mathbf{x}_i - \mathbf{x}_j\right\|^2\right) \quad (10)$$

E. Particle Swarm Optimization with wavelet mutation

In PSO, a swarm of particles moves in an n-dimensional search space of the possible solution of the problem. A position and a velocity represent a particle in the swarm. Some generations that are generated to update the particle's positions and velocities explore the promising domain to find the best solutions which spread throughout the swarm. The parameter adaptations are given by:

$$\vec{x}_i(t+1) = \vec{x}_i(t) + \vec{v}_i(t+1) \quad (11)$$

$$\vec{v}_i(t+1) = \varphi\vec{v}_i(t) + c_1\cdot r_1\cdot(\vec{p}_i(t) - \vec{x}_i(t)) + c_2\cdot r_2\cdot(\vec{g}(t) - \vec{x}_i(t)) \quad (12)$$

where $\vec{p}_i$ denotes the best previous (local) position and $\vec{g}$ denotes the best global position. Moreover, t represents the generation and φ is inertia weight. Other two parameters, c_1 and c_2 are acceleration constants that are weighted by r_1 and r_2, a random function in the range of [0-1]. According to [12], total number of c_1 and c_2 should exceed 4 to assure the convergence. Following the work of [12], the c_1 and c_2 are set at 2.05 while φ is 0.9. In addition, the optimization was done until 150 generations with 30 particles in each generation.

The wavelet mutation in PSO was proposed by [6]. A mutation chance is driven by a mutation probability $p_m \in$ [0 1]. If $x_i(t)$ is selected to be mutated then a new position is given by:

$$\vec{x}_i(t) = \begin{cases} \vec{x}_i(t) + \sigma\times(par^i_{\max} - \vec{x}_i(t)) & \text{if} \quad \sigma > 0 \\ \vec{x}_i(t) + \sigma\times(\vec{x}_i(t) - par^i_{\min}) & \text{if} \quad \sigma \le 0 \end{cases} \quad (13)$$

where σ is a Morlet wavelet function, par_{max} and par_{min} are the maximum and minimum position, respectively.

$$\sigma = \frac{1}{\sqrt{a}} e^{-\left(\frac{\alpha}{a}\right)^2/2} \cos(5\left(\frac{\alpha}{a}\right)) \quad (14)$$

The equation of "a" in the Morlet wavelet is determined by the following equation:

$$a = e^{-\ln(h)\times\left(1-\frac{t}{T}\right)^{\theta wm} + \ln(h)} \quad (15)$$

According to [13], α is randomly generated while p_m is 0.1, h=1000, and θwm is 1.

The objective of the optimization using wavelet PSO is to find the best kernel-based ELM parameters which minimize the classification error of the finger motion recognition. A 2-fold cross-validation was employed to measure the error. Moreover, the fitness function of particle $\vec{x}$ is defined by

$$f(\vec{x}) = \frac{\text{Number of uncorrect samples}}{\text{Total number of testing samples}} x100\% \quad (16)$$

The optimization was implemented in Radial basis function kernel only along with the parameter's range of C is $[2^{-20}, 2^{20}]$, and γ is $[2^{-20}, 2^{20}]$.

III. Result and Discussion

Three experiments using three different methods were performed. These three methods are the grid-search method, PSO, and PSO with wavelet mutation. Classification accuracy is employed to investigate their performances in classifying twelve finger-motion classes on five amputee subjects sing two EMG channels. Figure 1 presents the experimental result.

Figure 1 shows that, on average, the ELM that was optimized using the wavelet-PSO (swarm-wavelet-ELM) achieved the best performance compared to the ELM optimized utilizing PSO (swarm-ELM) and the grid-search method (grid-search-ELM). Moreover, swarm-ELM attained similar accuracy to swarm-wavelet-ELM on all amputees except on the amputee S1 and S3. In these two amputees, the swarm-ELM is less accurate than the swarm-wavelet-ELM and grid-search-ELM. Probably, the PSO on the swarm-ELM trapped on the local minima. Figure 2 gives clearer information about this assumption. Figure 2 shows that, after 30th generation, the PSO did not change the fitness value. Meanwhile, the wavelet mutation helped the PSO to avoid the local minima.

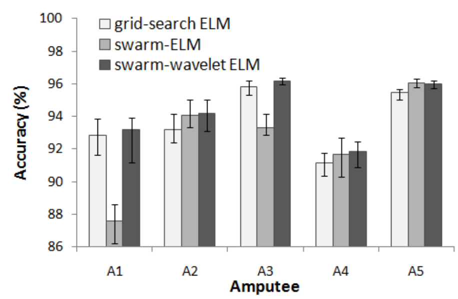

Fig 1. Average classification accuracy of three different ELM methods

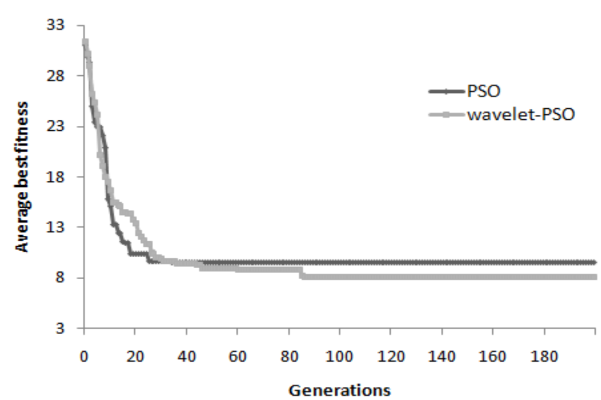

Fig 2.Average best fitness of PSO and wavelet-PSO across five amputees

A statistical test on the accuracy using one-way ANOVA (p was set at 0.05) was also done. The performance of the swarm-wavelet-ELM is significantly different from swarm-ELM ($p<0.05$). In addition, the grid-search ELM attained the average accuracy which is significantly similar to the swarm-wavelet-ELM ($p>0.05$). Although the grid-search-ELM and swarm-wavelet-ELM are statistically similar, their average accuracy is different. The swarm-wavelet-ELM achieved the average accuracy of 94.27 %, while the grid-search-ELM attained the average accuracy of 93.69. As for the swarm-ELM, it attained the average accuracy of 92.55 %.

In addition, the classification performance in regards to the finger motion were performed too. As shown in Figure 3, the swarm-ELM classified the flexion motions with the average accuracy more than 90%. In contrast, it identified the extension motions with the average accuracy less than 90%. As for the swarm-wavelet-ELM, similar to the swarm-ELM, it recognized the flexion motions better than the extension motions, but with accuracy better than the swarm-ELM.

The confusion matrix in Table 2 provides information about the misclassified finger motions. According to the Figure 3, the swarm-wavelet-ELM poorly classified the little finger extension (Le), middle finger extension (Me), and ring finger extension (Re). Me was mostly misclassified to the thumb abduction (Ta) and middle finger flexion (Mf). Furthermore, the system mostly misclassified the little finger extension (Le) to Re and vice versa. Although the misclassified motions were present, arguably the swam-wavelet-ELM has succeeded in recognizing different finger motions on five amputee subjects with accuracy of about 94% in which it is a promising result compared to others [8].

IV. CONCLUSION

The proposed pattern-recognition system, which employs PSO mutated using a wavelet function to optimize the kernel-based ELM, was able to recognize eleven imagined finger motions on five transradial amputees with the high accuracy of 94.27 % even though it employed only two EMG channels. The proposed system performed better than the grid-search-ELM and the standard-PSO-ELM. This promising result encourages the real-time implementation to verify the capability of the proposed system.

ACKNOWLEDGMENT

The authors would like to thank Ali H Al-Timmy [8] for sharing EMG data.

REFERENCES

[1] J. Cao, Z. Lin, and G.-B. Huang, "Self-adaptive evolutionary extreme learning machine," *Neural processing letters,* vol. 36, pp. 285-305, 2012.

[2] Y. Xu and Y. Shu, "Evolutionary extreme learning machine–based on particle swarm optimization," in *Advances in Neural Networks-ISNN 2006*, ed: Springer, 2006, pp. 644-652.

[3] W. Zong, H. Zhou, G.-B. Huang, and Z. Lin, "Face recognition based on kernelized extreme learning machine," in *Autonomous and Intelligent Systems*, ed: Springer, 2011, pp. 263-272.

[4] G. B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," *IEEE Trans. Syst. Man Cybern. B, Cybern.,* vol. 42, pp. 513-529., 2012.

[5] K. Anam, R. N. Khushaba, and A. Al-Jumaily, "Two-channel surface electromyography for individual and combined finger movements," in *Proc. 35th Ann. Int. Conf. IEEE-EMBS Eng. Med. Biol. Soc.*, 2013, pp. 4961-4964.

[6] S.-H. Ling, H. H. Iu, K. Y. Chan, H.-K. Lam, B. C. Yeung, and F. H. Leung, "Hybrid particle swarm optimization with wavelet mutation and its industrial applications," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics,* vol. 38, pp. 743-763, 2008.

[7] R. N. Khushaba, S. Kodagoda, M. Takruri, and G. Dissanayake, "Toward improved control of prosthetic fingers using surface electromyogram (EMG) signals," *Expert Syst. App.,* vol. 39, pp. 10731-10738, September 2012.

[8] A. H. Al-Timemy, G. Bugmann, J. Escudero, and N. Outram, "Classification of Finger Movements for the Dexterous Hand Prosthesis Control With Surface Electromyography," *IEEE J. Biomed. Health Inf.,* vol. 17, pp. 608-618, 2013.

[9] D. Cai, X. He, and J. Han, "SRDA: An efficient algorithm for large-scale discriminant analysis," *IEEE Trans. Knowl. Data Eng.,* vol. 20, pp. 1-12, 2008.

[10] L. H. Smith, L. J. Hargrove, B. A. Lock, and T. A. Kuiken, "Determining the optimal window length for pattern recognition-based myoelectric control: balancing the competing effects of classification error and controller delay," *IEEE Trans. Neural Syst. Rehabil. Eng.,* vol. 19, pp. 186-192, Apr 2011.

[11] T. R. Farrell and R. F. Weir, "The optimal controller delay for myoelectric prostheses," *IEEE Trans. Neural Syst. Rehabil. Eng.,* vol. 15, pp. 111-118, 2007.

[12] B. F. de Souza, A. de Carvalho, R. Calvo, and R. P. Ishii, "Multiclass SVM model selection using particle swarm optimization," in *Sixth International Conference on Hybrid Intelligent Systems, 2006. HIS'06*, 2006, pp. 31-31.

[13] S.-H. Ling, H. Iu, F. H.-F. Leung, and K. Y. Chan, "Improved hybrid particle swarm optimized wavelet neural network for modeling the development of fluid dispensing for electronic packaging," *IEEE Transactions on Industrial Electronics,* vol. 55, pp. 3447-3460, 2008.

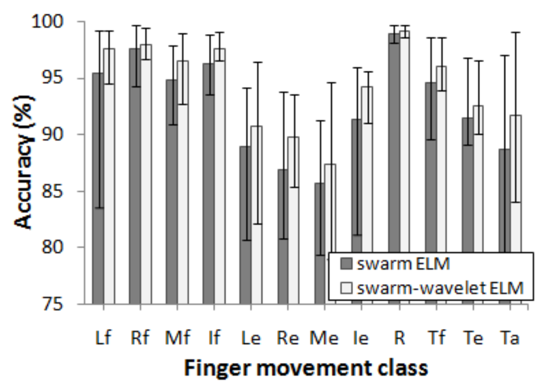

Fig 3. The accuracy of different finger motions across five amputees

TABLE II. THE CONFUSION MATRIX OF THE CLASSIFICATION RESULTS OF SWARM-WAVELET ELM AVERAGED FOR FIVE AMPUTEES (UNITS : %)

		Intended Task											
		Lf	Rf	Mf	If	Le	Re	Me	Ie	R	Tf	Te	Ta
Classified Task	Lf	98.2	0.5	0.1	0.0	0.0	0.0	0.1	0.3	0.0	0.4	0.0	0.3
	Rf	0.8	98.4	0.6	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0
	Mf	0.2	0.7	95.8	0.3	0.3	0.8	0.9	0.6	0.0	0.0	0.3	0.2
	If	0.2	0.1	0.1	97.7	0.3	0.3	0.2	0.1	0.0	0.7	0.1	0.3
	Le	0.0	0.0	0.4	0.3	90.1	4.6	1.8	0.2	0.0	0.6	1.0	0.9
	Re	0.1	0.0	0.7	0.2	3.8	89.8	2.1	0.2	0.0	0.6	0.7	1.7
	Me	0.2	0.0	1.3	0.3	2.4	3.1	88.6	1.3	0.0	0.4	0.7	1.8
	Ie	0.1	0.0	0.7	0.3	0.2	0.2	1.0	94.8	0.1	0.2	1.7	0.8
	R	0.1	0.0	0.0	0.2	0.1	0.0	0.0	0.1	99.1	0.3	0.0	0.0
	Tf	0.1	0.0	0.0	1.3	0.9	0.5	0.2	0.2	0.1	96.0	0.4	0.2
	Te	0.0	0.1	0.3	0.1	1.2	0.9	0.7	2.0	0.0	0.7	92.8	1.1
	Ta	0.0	0.0	0.3	0.4	1.2	2.3	1.1	0.3	0.0	0.2	1.2	93.0