7th Annual International IEEE EMBS Conference on Neural Engineering
Montpellier, France, 22 - 24 April, 2015

A novel extreme learning machine for dimensionality reduction on finger movement classification using sEMG

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

Abstract— **Projecting a high dimensional feature into a low-dimensional feature without compromising the feature characteristic is a challenging task. This paper proposes a novel dimensionality reduction constituted from the integration of extreme learning machine (ELM) and spectral regression (SR). The ELM in the proposed method is built on the structure of the unsupervised ELM. The hidden layer weights are determined randomly while the output weight is calculated using the spectral regression. The flexibility of the SR that can take labels into consideration leads a new supervised dimensionality reduction called SRELM. Generally speaking, SRELM is an unsupervised system in term of ELM yet it is a supervised system in term of dimensionality reduction. In this paper, SRELM is implemented in the finger movement classification based on electromyography signals from two channels. The experimental results show that the SRELM can enhance the performance of its predecessor, spectral regression linear discriminant analysis (SRDA) because it has better class separability than SRDA. In addition, its performance is better than principal component analysis (PCA) and comparable to uncorrelated linear discriminant analysis (ULDA).**

I. Introduction

Extreme learning machine (ELM) is a single hidden layer feedforward networks (SLFNs) that applies a random projection in the hidden layer. Meanwhile, the output weight of ELM is calculated analytically using a least square method. It means no iterative training in ELM. As a result, ELM training is very fast compared to traditional SLFNs that uses gradient descent algorithm. Interestingly, despite employing random projection, ELM performance outperforms backpropagation neural network in most cases, either classification or regression problems [1] [2].

The ELM has succeeded to be applied in many applications. As a classifier, it has been implemented in various applications such as myoelectric pattern recognition [3] [4], and character recognition [5] [6]. In addition, it played an importance role in face recognition [7], cancer detection [8], and protein structure prediction [10]. As for a regressor, ELM has proved its benefits in the physical parameter estimation [9] and electrical power system [10].

In addition to the classifier and regressor, ELM can be implemented in dimensionality reduction. G Huang *et al*. [11] have developed an unsupervised extreme learning for unsupervised dimensionality reduction. Its characteristic is similar to principal component analysis (PCA) that reduces the feature dimension with an unknown label. In fact, if the label is available, the dimensionality reduction method can be improved much. For this reason, linear discriminant analysis (LDA) has been proposed to take the labels into account. In many cases, LDA performs better than PCA except in a small number of data [12].

Up to the best author's knowledge, ELM has not been developed as a supervised dimensionality reduction. This paper proposes a new ELM for supervised dimensionality reduction applied to myoelectric pattern recognition. This paper employs spectral regression [13] to calculate the output weight of ELM instead of least square method as in the original ELM. Spectral regression is a spectral analysis of the Laplacian graph solved by least square regression. Its produces eigenvector to project the input space to the output space. This new ELM is called as spectral regression extreme learning machine (SRELM). The SRELM is similar to unsupervised ELM (US-ELM) proposed by G. Huang et al. [11] in the way it utilizes the obtained eigenvector to project the hidden layer output to the output layer. The different is on the calculation of the eigenvector and the label involvement.

The paper provides three main contributions. Firstly, it proposes a new and first ELM for supervised dimensionality reduction, i.e. SRELM. Secondly, it presents a new model of linear discriminant analysis (LDA) for myoelectric pattern recognition system. And the last, it proves experimentally that SRELM enhances the class separability of spectral regression discriminant analysis (SRDA) [14], another model of LDA that uses the spectral regression.

II. Method

A. Extreme Learning Machine (ELM)

Extreme learning machine (ELM) is essentially a learning mechanism intended for SLFNs. Instead of adjusting the hidden weight like in a standard SLFNS, ELM sets the hidden layer weights randomly and calculates the output weights analytically.

For N arbitrary distinct samples $\{(x_i, t_i)\}_{i=1}^{N}$ where $x_i \in \mathbf{R}^n$ and $t_i \in \mathbf{R}^m$, the output of SLFNs with K hidden nodes is

$$f(\mathbf{x}_i) = \sum_{j=1}^{K} \beta_j G(\mathbf{a}_j, b_j, \mathbf{x}_i) = \mathbf{h}(\mathbf{x}_i)\boldsymbol{\beta} = t_i, \quad i = 1,\ldots,N \tag{1}$$

where f is an output of ELM, G is a hidden layer output, $\mathbf{h}(\mathbf{x}_i) \in \mathrm{R}^{\mathrm{NxK}}$ is a matrix of hidden layer output and $\boldsymbol{\beta} \in \mathrm{R}^{\mathrm{Kxm}}$ is a matrix of output weight. To solve the output weights, ELM minimize the sum of squared losses of prediction error as follows:

$$\begin{aligned} &\text{Minimize} \quad : L_{ELM} = \tfrac{1}{2}\|\boldsymbol{\beta}\|^2 + C\,\tfrac{1}{2}\sum_{i=1}^{N}\|e\|^2 \\ &\text{Subject to} : \mathbf{h}(\mathbf{x}_i)\boldsymbol{\beta} = t_i^T - e_i^T \qquad i = 1,\ldots,N \end{aligned} \tag{2}$$

If we substitute the constraint into the objective function, we will obtain:

Khairul Anam is with the University of Jember Indonesia. Now he is a PhD student at University of Technology Sydney.

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).

978-1-4673-6389-1/15/$31.00 ©2015 IEEE

$$\text{Minimize} \quad : L_{ELM} = \tfrac{1}{2}\|\boldsymbol{\beta}\|^2 + C\,\tfrac{1}{2}\sum_{i=1}^{N}\|\mathbf{T}-\mathbf{H}\boldsymbol{\beta}\|^2 \tag{3}$$

where $\mathbf{H}$ = $[\mathbf{h}(\mathbf{x}_1),\ldots,\ \mathbf{h}(\mathbf{x}_N)^T]^T \in R^{NxK}$ and $\mathbf{T} \in R^{Nxm}$. Gradient of (3) with respect to β to zero gives:

$$\nabla L_{ELM} = \boldsymbol{\beta} + C\mathbf{H}^T(\mathbf{T}-\mathbf{H}\boldsymbol{\beta}) = 0 \tag{4}$$

Equation (4) gives two solutions of $\boldsymbol{\beta}$ subject to the **H**. If **H** has more rows than coulomns then :

$$\boldsymbol{\beta} = \left(\mathbf{H}^T\mathbf{H} + \frac{I_K}{C}\right)^{-1}\mathbf{H}^T\mathbf{T} \tag{5}$$

where **I** is an identity matrix of dimension *K*. In the contrary, if **H** has more coulomn than rows, then

$$\boldsymbol{\beta} = \mathbf{H}^T\left(\mathbf{H}\mathbf{H}^T + \frac{I_N}{C}\right)^{-1}\mathbf{T} \tag{6}$$

where **I** is an identity matrix of dimension *N*.

B. Spectral Regression Extreme Learning Machine (SRELM)

To modify ELM for dimensionality reduction, we consider unknown labels for ELM. In another word, we employ unsupervised extreme machine learning as explained in [11]. Therefore, the objective function in (2) is modified as:

$$\begin{aligned}&\text{Minimize} \quad : L_{ELM} = \tfrac{1}{2}\|\boldsymbol{\beta}\|^2 + \lambda\,\tfrac{1}{2}\mathrm{Tr}(\mathbf{F}^T\mathbf{L}\mathbf{F})\\ &\text{Subject to} : \boldsymbol{f}_i = \mathbf{h}(\mathbf{x}_i)\boldsymbol{\beta} \qquad i = 1,\ldots,N\end{aligned} \tag{7}$$

where L_{ELM} is the objective function, **F** is a matrix of the output, ***file*** is the output of ELM and L is a graph Laplacian. By substituting the constraint to the objective function, we have

$$\begin{aligned}&\text{Minimize} \quad : L_{ELM} = \tfrac{1}{2}\|\boldsymbol{\beta}\|^2 + \lambda\,\tfrac{1}{2}\mathrm{Tr}(\boldsymbol{\beta}^T\mathbf{H}^T\mathbf{L}\boldsymbol{\beta}\mathbf{H})\\ &\text{Subject to} \quad : \boldsymbol{\beta}^T\mathbf{H}^T\mathbf{L}\boldsymbol{\beta}\mathbf{H} = \mathbf{I}_m\end{aligned} \tag{8}$$

As proved in [11], the optimal solution of (8) is the solution of the generalized eigenvalue problem:

$$\left(\mathbf{I}_L + \lambda\mathbf{H}^T\mathbf{L}\mathbf{H}\right)\mathbf{u} = \gamma\mathbf{H}^T\mathbf{L}\mathbf{H}\mathbf{u} \tag{9}$$

The spectral graph analysis assumes that the map of a graph to real line y, as a linear function

$$\mathbf{y}=\mathbf{H}\mathbf{u} \tag{10}$$

As a result, Eq. (9) can be formulated as

$$\left(\mathbf{I}_L + \lambda\mathbf{H}^T\mathbf{L}\mathbf{H}\right)\mathbf{u} = \gamma\mathbf{H}^T\mathbf{L}\mathbf{y} \tag{11}$$

According to the spectral regression theory [13] [14], the optimal y can be obtained by minimizing

$$\sum_{i,j}(y_i - y_j)2\mathbf{W}_{ij} = 2\mathbf{y}^T\mathbf{L}\mathbf{y} \tag{12}$$

where L = D – W is a graph Laplacian, D is a diagonal matrix whose elements are $D_{ii} = \Sigma_j W_{ji}$, and W is a symmetric N x N matrix which is a pairwise similarity between two data points. N is the number of samples. Equation (11) can be also optimized by solving the maximum eigenvalue eigen-problem [13]:

$$\mathbf{W}\mathbf{y}=\lambda\mathbf{D}\mathbf{y} \tag{13}$$

In addition, in the spectral regression algorithm, we can include label consisting c classes. The solution for **u** will contain **c-1** solutions as described in [13].

In summary, the solution to (11) is done in two steps. Firstly solves the eigenvalue problem in (13). Secondly finds **u** with satisfies **Hu** = **y** using:

$$\mathbf{u} = \arg\min_u\left(\sum_{i=1}^{N}\left(\mathbf{u}^T h(x_i) - y_i\right)^2 + \alpha\sum_{j=1}^{L}u_j\right) \tag{14}$$

where α is a regression parameter and u_j is the component of **u**. Finally

$$\boldsymbol{\beta} = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{c-1}] \in R^L. \tag{15}$$

Theoretically, the integration of ELM and SR, called SRELM, results in another variation of LDA. The ELM projects the input feature to a random feature. Then, the spectral regression projects the random feature to the reduced and meaningful feature for the classifier. In the structure of ELM, the SR provides values for the output weights. Interesting characteristic of SRELM is, it is an unsupervised method on the side of ELM structure, but it is a supervised one on the side of dimensionality reduction.

C. Finger movement classification

We tested the performance of SRELM in myoelectric pattern recognition system for classifying ten finger movements. EMG signals were recorded at 2000 kHz sampling frequency as in [15]. The signals were collected from flexor policis longus and flexor digitorium superificialis muscles (fig.1). Eight subjects, two females and six males aged 24-60 years old, participated in the experiment. All subjects were normally limbed with no muscle disorder. In addition, the subject's arm fixed at a specific position to avoid the effect of position changing on EMG signals.

This work extracted features from two EMG channel plus one channel formed from summation of the two channels. Signals were extracted every 100 ms in length of 100 ms using time domain (TD) and autoregressive (AR) features. It involved mean absolute value (3 features), zero waveform lengths (3 features), slope sign changes (3 features), number of zero crossings (3 features), and sample skewness (3 features). In addition, some parameters from Hjorth-time domain parameters (9 features) and 6th order autoregressive model parameters (18 features) were included. The total

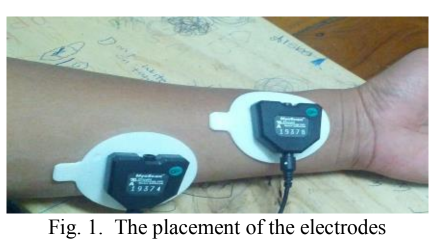


Fig. 1. The placement of the electrodes

number of features extracted is 42.

SRELM, as LDA, reduces the dimension of features from 42 to c-1 features in which c is the number of classes. Then, the performance of SRELM was compared with other dimensionality reduction methods such as uncorrelated linear discriminant analysis (ULDA) [16]. Another comparison was also conducted with spectral regression dimensionality reduction (SRDA) and orthogonal fuzzy discriminant analysis (OFNDA) [17]. In addition, principal component analysis (PCA) and unsupervised extreme learning machine (USELM) got involved. The trial without dimensionality reduction (baseline) was also considered in the comparison.

In addition, various classifiers will utilize projected features of SRELM to identify individual and combined finger movements. Those classifiers are AW-ELM [18] (adaptive wavelet ELM), RBF-ELM (radial basis function ELM), SVM (support vector machine), kNN (k-nearest neighbourhood) and LDA. Different classes will be tested starting from five up to ten classes. The ten classes consist of five individual finger movements, i.e. thumb (T), index (I), middle (M), ring (R), little (L). The other movements are combined finger movements consisted of thumb-index (T–I), thumb-middle (T–M), thumb-ring (T–R), thumb-little (T–L) and the hand close (HC).

III. Result and Discussion

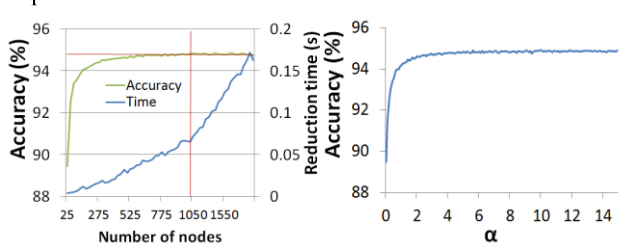
A. Parameter optimization

The parameter optimization of SRELM influences the performance of the system. It contains two main parameters, i.e. the number of hidden nodes (a part of the ELM parameter) and alpha α (a part of the regression coefficient of the spectral regression). As in many feedforward neural networks, the number of hidden nodes is a trivial parameter that is not easy to determine. Therefore, we varied the number of hidden nodes and selected the optimal one by considering the accuracy and the reduction time.

The fig. 2a presents the experimental result. As shown by the intersection of two red lines in the fig. 2a, 1000-nodes in the hidden layer is the optimum number. As for alpha (α), fig. 2b shows that the accuracy is guaranteed good if the alpha is more than 4. In this paper, we simply select alpha = 10.

B. Class number experiments

In this experiment, we tested the performance of SRELM in reducing feature dimension for different classes, ranging from five up to ten classes. Table 1 presents the classes involved in the experiment. The SRELM's performance is compared to other well-known methods such as ULDA,



(a) (b)

Figure 2. Experiment of the number of nodes (a) and alpha (b)

TABLE I. Various classes involved in the experiment

#Classes	Classes
5	T, I, M, R, L
6	T, I, M, R, L, T-I
7	T, I, M, R, L,T-I, T-M
8	T, I, M, R, L, T-I, T-M, T-R
9	T, I, M, R, L, T-I, T-M, T-R, T-L
10	T, I, M, R, L, T-I, T-M, T-R, T-L, HC

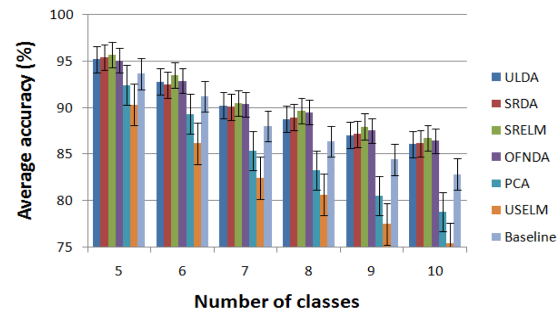


Figure 3. The performance of SRELM and others across eight subjects

SRDA, ONFDA, PCA, USELM and baseline. In addition, all experiments employed AW-ELM as a classifier. Fig 3 presents the experimental result.

As shown in Fig. 3, the accuracy of the system is decreasing as the number of classes is increasing. The trends happen to all methods. The characteristic of SRELM is comparable to the state-of-the-art of linear discriminant analysis. All supervised dimensionality reductions such as ULDA, SRDA, SRELM, and OFNDA attain similar accuracy in all class numbers. Even, SRELM is better than all methods tested. Its accuracy is ranging from 95.67 % to 86.73 % from 5 to 10 classes of movement.

SRELM and SRDA employ same spectral regression. Nevertheless, SRELM has more precise class separation than SRDA as shown in fig. 4. The class separability of SRELM is similar to ULDA and OFNDA (fig. 4). Class separability helps the classifier in classifying the movement type. Hence, the accuracy of the system using SRELM exceeds the others (fig. 3). Despite having better class separation performance, SRELM takes more processing time than SRDA, as seen in fig. 5. However, its processing time is still reasonable and

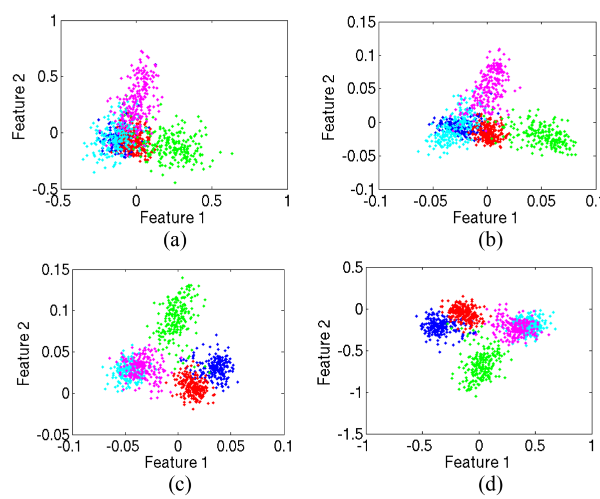


Figure 4. Scatter plot of two first discriminant data using (a). SRDA, (b). SRELM, (c) ULDA and (d) OFNDA on five-class trial from the first subject

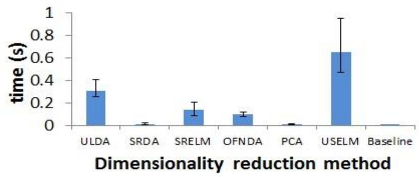


Figure 5. Processing time consumed by some dimensionality reduction methods

less than ULDA's processing time.

C. *Classifier experiments*

Various types of classifiers were involved to test the performance of SRELM. Fig. 6 presents the result. It indicates that SRELM works well across five classifiers. It has similar performance to other LDA extension. In addition, it outperforms PCA and the baseline. Analysis of variance (ANOVA) test was also conducted to find out the true comparison of SRELM and other methods. Fig. 7 presents the ANOVA test for p is set at 0.05. It indicates that there is a significant difference between the performance of SRELM and PCA, USELM, and Baseline (p<0.05). On the other hand, the performance of SRELM is comparable to the well-known dimensionality reduction LDA and its variation (p>0.05).

IV. Conclusion

This paper proposes a new extreme learning machine and at the same time a new dimensionality reduction called SRELM for finger movement classification. The experiment results show that SRELM is comparable to ULDA and OFNDA and better than SRDA. Moreover, it has better class separability than SRDA. SRELM can also work well on different classifiers and various numbers of classes with an average accuracy ranging from 95.67 % to 86.73 % from 5 to 10 classes of movement across eight subjects using two EMG channels.

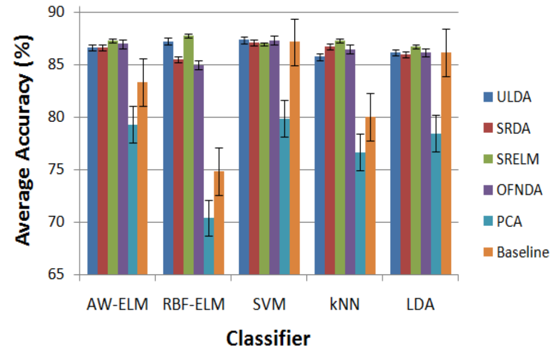


Figure 6. SRELM performance of different classifiers across eight subjects. For abbreviations, see section 2C.

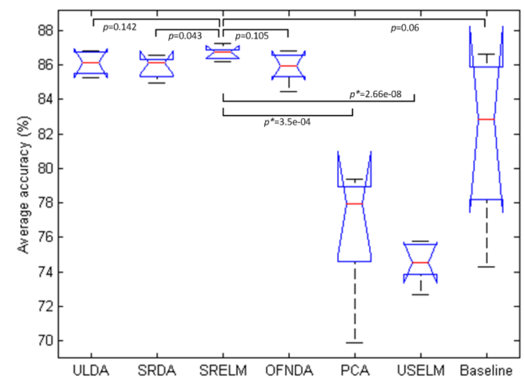


Figure 7. Anova test of SRELM and other methods across eight subjects

References

[1] 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.

[2] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing,* vol. 70, pp. 489-501, 12// 2006.

[3] C.-J. Lin and M.-H. Hsieh, "Classification of mental task from EEG data using neural networks based on particle swarm optimization," *Neurocomputing,* vol. 72, pp. 1121-1130, 2009.

[4] 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.

[5] B. P. Chacko, V. V. Krishnan, G. Raju, and P. B. Anto, "Handwritten character recognition using wavelet energy and extreme learning machine," *International Journal of Machine Learning and Cybernetics,* vol. 3, pp. 149-161, 2012.

[6] W. Zheng, Y. Qian, and H. Lu, "Text categorization based on regularization extreme learning machine," *Neural Computing and Applications,* vol. 22, pp. 447-456, 2013/03/01 2013.

[7] A. A. Mohammed, R. Minhas, Q. Jonathan Wu, and M. A. Sid-Ahmed, "Human face recognition based on multidimensional PCA and extreme learning machine," *Pattern Recognition,* vol. 44, pp. 2588-2597, 2011.

[8] S. Saraswathi, S. Sundaram, N. Sundararajan, M. Zimmermann, and M. Nilsen-Hamilton, "ICGA-PSO-ELM Approach for Accurate Multiclass Cancer Classification Resulting in Reduced Gene Sets in Which Genes Encoding Secreted Proteins Are Highly Represented," *IEEE/ACM Transactions on Computational Biology and Bioinformatics,* vol. 8, pp. 452-463, 2011.

[9] K. Javed, R. Gouriveau, and N. Zerhouni, "SW-ELM: A summation wavelet extreme learning machine algorithm with a priori parameter initialization," *Neurocomputing,* vol. 123, pp. 299-307, 1/10/ 2014.

[10] A. H. Nizar, Z. Y. Dong, and Y. Wang, "Power Utility Nontechnical Loss Analysis With Extreme Learning Machine Method," *IEEE Transactions on Power Systems,* vol. 23, pp. 946-955, 2008.

[11] G. Huang, S. Song, J. N. Gupta, and C. Wu, "Semi-supervised and unsupervised extreme learning machines," 2014.

[12] A. M. Martínez and A. C. Kak, "Pca versus lda," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 23, pp. 228-233, 2001.

[13] D. Cai, X. He, and J. Han, "Spectral regression: A unified approach for sparse subspace learning," in *Seventh IEEE International Conference on Data Mining (ICDM)*, 2007, pp. 73-82.

[14] 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.

[15] K. Anam and A. A. Al-Jumaily, "Swarm-based extreme learning machine for finger movement recognition," in *Biomedical Engineering (MECBME), 2014 Middle East Conference on*, 2014, pp. 273-276.

[16] J. Ye, R. Janardan, Q. Li, and H. Park, "Feature reduction via generalized uncorrelated linear discriminant analysis," *IEEE Transactions on Knowledge and Data Engineering,* vol. 18, pp. 1312-1322, 2006.

[17] R. N. Khushaba, A. Al-Ani, and A. Al-Jumaily, "Orthogonal Fuzzy Neighborhood Discriminant Analysis for Multifunction Myoelectric Hand Control," *IEEE Trans. Biomed. Eng.,* vol. 57, pp. 1410-1419, 2010.

[18] K. Anam and A. Al-Jumaily, "Adaptive Wavelet Extreme Learning Machine (AW-ELM) for Index Finger Recognition Using Two-Channel Electromyography," in *Neural Information Processing*, 2014, pp. 471-478.