Masking Covariance for Common Spatial Pattern as Feature Extraction

H. Asyraf[1], M. I. Shapiai[1, 2], N. A. Setiawan[3], W. S. N. S. Wan Musa[1]
[1]*Electronic System Engineering, MJIIT.*
[2]*Centre of Artificial Intelligence and Robotics (CAIRO), UTM Kuala Lumpur, Malaysia.*
[3]*Faculty of Engineering, Universitas Gadjah Mada Yogyakarta, Indonesia.*
hafizulasyraf.ahmad@gmail.com

***Abstract*—Brain Computer Interface (BCI) requires accurate and reliable discrimination of EEG's features. One of the most common method used in feature extraction for BCI is a common spatial pattern (CSP). CSP is a technique to distinguish two opposite features by computing the spatial pattern of the measured EEG channels. The existing CSP is known to be prone to the over-fitting problem. Covariance estimation is an important process in obtaining spatial pattern using CSP. The empirical covariance estimation ignores the spurious information between channels of EEG device. This may cause inaccurate covariance estimation, which results to lower accuracy performance. In this study, a masking covariance matrix is introduced based on the functionality of brain region. The addition of masking covariance is to improve the performance of CSP. Features obtained through features extraction is then used as the input to Extreme Learning Machine (ELM). Comparisons between features of conventional CSP and with the addition of masking covariance are visually observed using the collected EEG signals using EMOTIV. The performance accuracy of the proposed technique has offered slight improvement from 0.5 to 0.5567. The obtained results are then discussed and analyzed in this study. Therefore, by introducing masking covariance matrix, the performance of the existing CSP algorithm can be improved.**

***Index Terms*—Masking Weight; Common Spatial Pattern; Covariance Filter; Extreme Learning Machine.**

I. Introduction

Brain computer interfaces (BCI) is now being aggressively studied with the aim to translate brain activities into a readable command where this command is later used for triggering actuator in real life. Our brain activity can be monitored by using Electroencephalogram (EEG) device where this device consisting of electrodes to acquire EEG signal on their respective region. Traditionally, the process of monitoring brain activity required subject to undergo surgical implantation to place the electrode on the surface or within the depth of the brain. Modernization has now allowed the signal to be acquired by only attaching electrodes on the scalp without doing any surgery.

However, the capabilities of fetching EEG signal without surgical implantation brings drawback in the quality of the signal. The distance from the neuron will diminish the EEG signal property that we would like to observe while mixing the signal from another region (noise) [1]. Therefore, the signal cannot be used directly in BCI application. Overcoming this problem lead to the study of numerous feature extractions such as Common Spatial Pattern (CSP) [2]. Laplacian filter [3] and common average reference [4] to obtain only usable feature to distinguish the EEG signal into its category. This paper will be focusing on the implementation of CSP method to extract the relevant feature from EEG Signal.

CSP method is an algorithm commonly used to extract the most significant discriminative information from EEG signals. It was first suggested for binary classification of EEG trials [5]. CSP algorithm performs covariance estimation process where it computes the projection of most differing power or variance ratios in feature space by using the spatial filter and the projections are calculated by simultaneous diagonal of covariance matrices of two classes [2]. The importance of this process is to obtain the correlation of each channel with others. Most of the time, the significant features are obtained from the first few most discriminative filters. These features will be used to differentiate between two opposite features we want to observe.

Even though CSP is capable of extracting the features by applying a discriminative filter. The objective of the process is not to directly input the relevant features obtain through CSP process into classifier but to amplify the EEG signal with most significant features while minimizing the EEG signal with weak features so information received by classifier will be clearer. Often a classifier will receive input by computing the variance of these features together with collected EEG earlier [2]. CSP is also known to be highly sensitive to noise and prone to over-fitting. To address this issue, regularized CSP is proposed as in [6, 7].

In order to increase the accuracy of CSP analysis, the algorithm is further refined by introducing few addition on the original algorithm. In paper [5], Haiping Lu took a different approach by regularizing the covariance matrix estimation in common spatial pattern extraction. The proposed algorithm is consisting of two regularization parameters where the first parameter controls the shrinkage of a subject specific covariance matrix towards a "generic" covariance matrix to improve the estimation stability based on the principle of generic learning [8]. The second regularization parameter controls the shrinkage of the sample based covariance matrix estimation towards a scaled identity matrix to account for the biased due to limited number of samples [5]. Besides regularizing covariance matrix approach, in paper [9], Allesandro Balzi has implement the Importance Weighted Estimator to the covariance estimation (used as input by every CSP variant) rather than modify the CSP itself that make the covariance estimation more robust to non-stationary signal, as a result the method brings to improvement in classification accuracy.

Moving differently, to overcome the over-fitting information on CSP algorithm, we believed that by introducing the masking weight covariance, CSP performance can be improved. The idea is based on the functionality of brain's love region. Therefore, the masking covariance role is to reduce the relationship of the channel that is not mapped within the same functionality region. To design the masking covariance, the knowledge regards on how the EEG signal is arranged is required because it involves predefining matrix based on the channel allocation on the signal data. The wanted region will be defined by allocating maximum covariance while others are lower.

The feature map of the data can be self-organized referring to the matrix setup. In this case, it will be easier for the classifier to respond on the topographic map corresponds to a particular feature of the input pattern.

II. EMPLOYED TECHNIQUE

A. *Common Spatial Pattern*

The EEG signal obtained through EMOTIV device form $14 \times T$ data where 14 is the number of channels for EMOTIV and T represents samples per channel. The covariance of the signal can be calculated from:

$$C = \frac{EE^{\mathrm{T}}}{trace(EE^{\mathrm{T}})} \tag{1}$$

where E is the EEG signal and T denotes the transpose operator. $trace$ is the sum of a diagonal element of its content. The spatial pattern of each class (left and right motor imagery) will be distributed individually by calculating the average trials of each group.

The spatial pattern is then masked by the previously declared masking weight covariance. The masked spatial pattern for each class can be computed as follow:

$$C_{M1} = W \times C_1 \tag{2}$$

W is the masking weight declared earlier and $C_1$is the covariance of signal class 1 (thinking left). This shall be applied on both class and obtain the composite spatial covariance from (3).

$$C_T = \bar{C}_{M1} + \bar{C}_{M2} = U_0 \sum U_0^T \tag{3}$$

where U_0 is the matrix of eigenvectors and $\sum$ is the sum of diagonal matrix of eigenvalues. The whitening transformation matrix is:

$$P = \sum^{-\frac{1}{2}} U_0^T \tag{4}$$

which equalized the variances in the space spanned by U_c. Now all eigenvalues of $P\bar{C}_1P'$ are equal to 1.

$$S_1 = P\bar{C}_1P' \qquad S_2 = P\bar{C}_2P' \tag{5}$$

Now $S_1$and S_2 share common eigenvectors where:

$$S_1 = U{\textstyle\sum}_1 U^T \qquad S_2 = U{\textstyle\sum}_2 U^T \tag{6}$$

$$\textstyle\sum_1 + \sum_2 = I \tag{7}$$

The eigenvectors with the largest eigenvalues for S_1 have the smallest eigenvalues for S_2 and vice versa. The transformation of whitened EEG onto the eigenvectors corresponding to the largest eigenvalues in $\sum_1$and $\sum_2$ is optimal for separating variance in two signal matrices. Thee projection matrix W is denoted as:

$$W = U^T P \tag{8}$$

With projection matrix W. The original EEG can be transformed into uncorrelated components:

$$Z = WX \tag{9}$$

Z can be seen as EEG source components including common and specific components of different tasks. The original EEG X can be reconstructed by:

$$X = W^{-1} Z \tag{10}$$

where W^{-1} is the inverse matrix of W. The columns of W^{-1} are spatial patterns, which can be considered as EEG source distribution vectors. The first and last columns of W^{-1} are the most important spatial patterns that explain the largest variance of one task and the smallest variance of the other.

B. *Feature Extraction*

The features used for classification are obtained by decomposing the EEG as (6). For each direction of imagery movement, the variances of only small number m of signals most suitable for discrimination are used for the construction of the classifier. The signal $Z_p$that maximizes the difference of variance of left versus right motor imagery EEG are the ones that are associated with the largest eigenvalues $\sum_1$and $\sum_2$. These signals are the m first and last rows of Z due to the calculation of W.

$$f_p = \log\left(\frac{var(Z_p)}{\sum_{i=1}^{2m} var(Z_p)}\right) \tag{11}$$

The feature vectors f_p of left and right trials are used to calculate a linear classifier [5] [10]. The log-transformation serves to approximate normal distribution of the data.

For proper estimation of the classification accuracy, the data set of each subject is divided into a training and testing set. The training set is used to calculate a classifier, which is used to classify the testing set. This training procedure is repeated using a random weight generated by ELM and also taking a different portion of training and testing set.

C. Classifier

The evaluation of both training and testing accuracy is done using ELM model. By providing the pattern produced by feature extraction at (10), ELM will create a hyper plane between two features and distinguish which side represents which information. The model of ELM can be depicted as Figure 1.

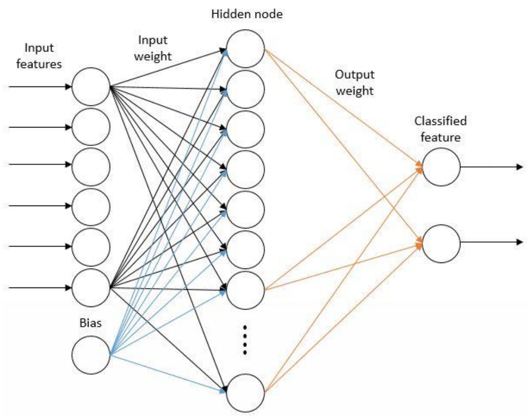

Figure 1: ELM Model

From the model in Figure 1, the computation output is obtained by using following equation. The operation of ELM can be summarized as follows [11]:

1) Obtain features set x_t, activation function $g(x)$ and number of hidden node N
2) Randomly assign input weight w_i and bias b_i
3) Calculate the hidden layer output matrix
4) Calculate the output weight β

The computation can be derived by (12):

$$f(x) = sig(\sum_{i=1}^{L} \beta i G(ai, bi, x)) \tag{12}$$

III. METHODOLOGY

This section will discuss about the conventional CSP method in computing covariance matrix estimation and later the implementation of the masked weight into the CSP. The performance of both methods is then compared using ELM by evaluating the accuracy of both methods. Most of ELM part is referred in [11, 13].

A. EEG Signal Acquisition

Firstly, EEG signals are acquired by using EMOTIV Epoch (14 channels). Subjects were briefed about how the experiment was going to be conducted. A short video was played during the acquisition period to stimulate subjects on thinking of moving either right hand or left hand. Subjects were sitting on the conducive chair while the other part of their body remains static. In this experiment, the process of thinking left hand and right hand is set to be alternating so that subjects will active switching their thoughts after short rest between these two sets. The experiment setup is depicted as Figure 2.

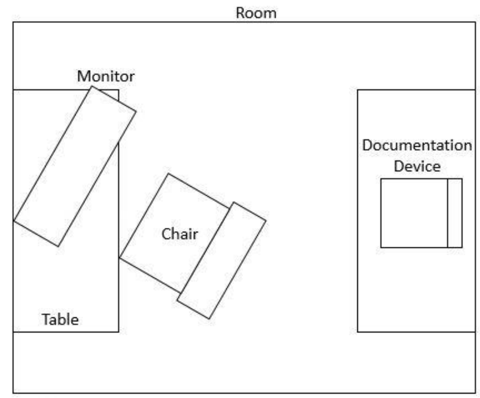

Figure 2: Experiment setup

B. EEG Signal Information

The sampling rate of EMOTIV device had been set to 128Hz. A session length lasts for 53 seconds where early preparation last for 5 seconds, each cue took 3 seconds and rest between cues took 2 seconds. The collective acquisition consisting of 160 trials for training samples and 120 trials for testing samples where each subject undergo 10 trials each (5 left hand and 5 right hand). This process had gone through a little modification from [8] where visual cue is reduced from 3.5 seconds to 3 seconds.

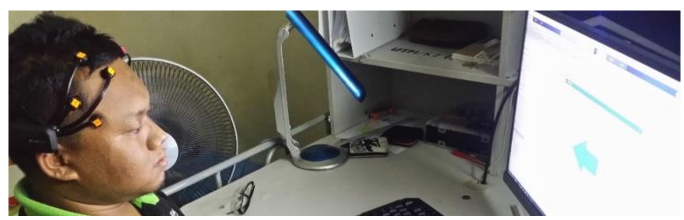

Figure 3: Data acquisition process

C. Masking Weight Covariance

The aim of this experiment is to improve the design of spatial filters which lead to new time series whose variances are optimal for discrimination of two class of EEG related with left and right motor imagery. Most of the parts, the analysis is using CSP procedure however a little modification on calculating normalized covariance takes place. This method is called masking weight, is introduced by defining a correlation between each electrode and maximize the selected electrodes we believed giving us the most necessary information regarding the conducted experiment. In this case, the defined masking weight is as in Figure 4.

The empty space in the matrix represents less important electrodes where the value assigned in this experiment is distributed equally. In this experiment, we believed that the most significant information can be retrieved from channels AF3, AF4, F7, F8, F3, F4, FC5 and FC6. These regions are called as the functional region where it's provides the information omitted by motor cortex. The selected region is depicted as Figure 5.

	AF3	F7	F3	FC5	T7	P7	O1	O2	P8	T8	FC6	F4	F8	AF4
AF3	1													1
F7		1	1	1							1	1	1	
F3		1	1	1							1	1	1	
FC5		1	1	1							1	1	1	
T7					1					1				
P7						1			1					
O1							1	1						
O2							1	1						
P8						1			1					
T8					1					1				
FC6		1	1	1							1	1	1	
F4		1	1	1							1	1	1	
F8		1	1	1							1	1	1	
AF4	1													1

Figure 4: Masking weight pattern

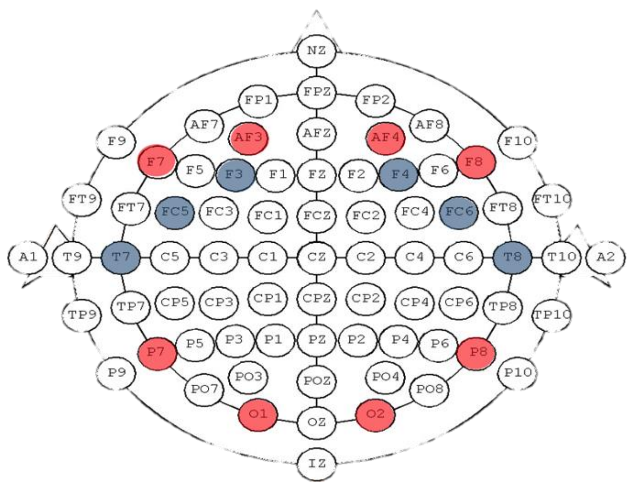

Figure 5: Selected region

D. EEG Signal Preprocessing

Signal preprocessing was done to remove unnecessary information within the signal such as preparation and rest duration. In addition, the signal will be filtered using bandpass filter to remove unnecessary frequency band, which may carry artifact in the signal [14]. In this experiment, a Butterworth bandpass filter is used to accept only frequency ranging from 8Hz to 30Hz [15]. This frequency band was chosen because it encompasses the alpha and beta frequency bands, which have been shown to be most important for movement classification [5, 14].

IV. Experiment, Results and Discussion

A. Parameter Setting

Few parameters need to be stressed on during implementation of ELM. In this case, ELM is set to use classifier mode instead of regression because the dataset is going to be split based on its homogeneity, a dependent variables of the features. The number of hidden nodes were tested with iterative number and found 15 nodes are sufficient to optimize the algorithm performance [17]. Sigmoid is used for the activation function based on [17]

B. Feature Observation

For calculation of the spatial filters, each trial is split into different time segment of 3 seconds length where is contains 384 time sample for each trial. The features produce through the stated process are evaluated by eyes before sending it through ELM. Two graphs are plotted where the top represents the features obtain through conventional CSP approach while the graph at the bottom represents the approach.

Each of masked weight value is tested for 5 times and the average of them is compared with default CSP approach. At first, a few analysis on a number of hidden nodes is done by tested, which hidden nodes is best to be used in ELM model. Then activation function can be any nonlinear piecewise continuous functions, where in this case, sigmoid is used [18].

Tables 1 and 2 represent the average training and testing accuracy over 5 times and show the comparison between performance of CSP and CSP with addition masking covariance (CSP-MC). These calculations were done few trials with variable weight in ELM model. The results from the trials are used to obtain the mean, standard deviation, maximum and minimum of the overall data. In this case, the bold value represents the best performance as compared to the other method. Among all the masking weight covariance set up, testing accuracy of Masking Weight Covariance (MWC) 0.5 produces the highest average accuracy as compared to others value and also conventional CSP. The maximum and minimum value of MWC 0.5 is slightly higher than conventional CSP as well. From these results, 0.5 is the best value in weight distribution on less relevant region.

Table 1
Comparison Between Training Accuracy of CSP and CSP-MC

Method		average	std-dev	max	min
CSP		0.6259	0.0361	0.7000	0.5667
CSP-MC	0.1	0.6125	**0.0243**	0.6500	0.5750
	0.2	0.6133	0.0246	0.6417	0.5583
	0.3	0.6325	0.0295	0.6833	**0.5917**
	0.4	0.6425	0.0438	0.6917	0.5667
	0.5	0.6325	0.0461	**0.7083**	0.5667
	0.6	0.6267	0.0280	0.6583	0.5833
	0.7	0.6283	0.0281	0.6667	0.5750
	0.8	**0.6408**	0.0394	0.6917	0.5583
	0.9	0.6300	0.0255	0.6667	**0.5917**

Table 2
Comparison Between Testing Accuracy of CSP and CSP-MC

Method		average	std-dev	max	min
CSP		0.5000	0.0509	0.5667	0.4333
CSP-MWC	0.1	0.5082	0.0713	**0.6167**	0.4000
	0.2	0.5007	0.0651	0.6500	0.4000
	0.3	0.4517	**0.0215**	0.4833	0.4167
	0.4	0.5150	0.0506	**0.6167**	0.4500
	0.5	**0.5567**	0.0589	0.5833	0.4500
	0.6	0.5100	0.0486	0.5833	0.4500
	0.7	0.5382	0.0368	**0.6167**	**0.4833**
	0.8	0.4450	0.0485	0.5167	0.3667
	0.9	0.0550	0.0497	0.6000	0.4333

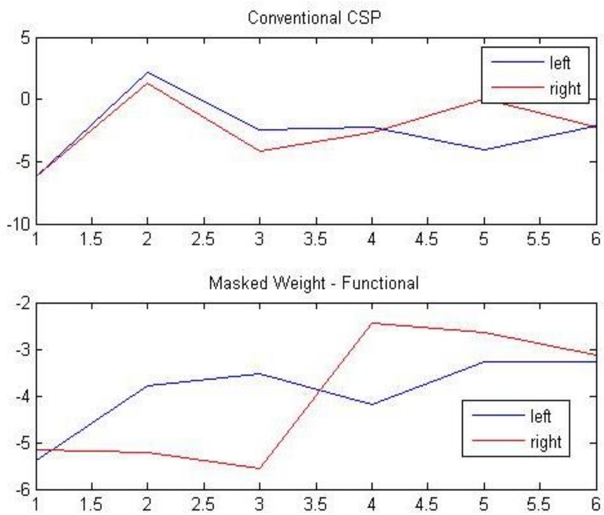

Figure 6: Pattern improvement by masked weight

Figure 6 shows that the pattern produced after computation (10) for single trial has significantly changed. This is because others channels, which provide signal believed to be event related desynchronize (ERD) is reduced by masking the weight on covariance of the CSP approach. On visual inspection, the first three channels provide higher information about imagery movement of the left hand while the last 3 channels provide higher information imagery movement of right hand.

V. Conclusion

Conventional CSP algorithm computes the covariance by calculating the relationship between each channel directly. However, by introducing masked weight, monitoring the exact channels, which providing the needed information. Masking weight allows 2 channels correlate to each other by either on full scale information for relevant channels and minimize scale information on non-relevant channels.

The resulting pattern of the features obtained through CSP algorithm is inspected manually for most trials to see if there is a significant difference between conventional approach and addition of masking weight. It appears that the pattern produced by addition masking weight has slightly changed the pattern for most trials.

Current experiment concludes that by choosing 0.5 as the masking weight for covariance will produce the best testing accuracy for most of the times. In this paper, the approach of introducing masked weight is distributed evenly on all less important channels. We believe by having a more proper distribution, further improvement in the result accuracy can be obtained.

Acknowledgment

The author would like to thank Dr. Parvaneh Shahbanzade post-doctoral at Malaysia Japan International Institute of Technology for her contribution on guiding, correcting and monitoring the progress of writing this paper. The author also would like to use this opportunity to thank all subjects who undergo data acquisition even though the process of setting up the device take quite a long time. Finally, thanks to Universiti Teknologi Malaysia for funding this research project through a research university Grant (Q.K130000.2643.10J98) titled "Classification of Brain Signal for Normal Person and Stroke Patients for Application in Brain Computer Interface".

References

[1] M. Seeck, F. Lazeyras, C.M Michel, O. Blanke, C.A Gericke, J. Ives, J. Delavelle, X. Golay, C.A Haenggelli, N. de Tribolet, and T. Landis, "Non-Invasive epileptic focus localization using EEG triggered functional MRI and electromanetic tomography," *Electroencephalography and Clinical Neurophysiology*, vol. 106, pp. 508-512, June. 1998.

[2] H. Lu, K. N. Plataniotis and A. N. Venetsanopoulos, "Regularized Common Spatial Pattern with Generic learning for EEG Signal Classification," *IEEE Transactions on Biomedical Engineering*, vol. 57, pp. 2936-2946, 2010

[3] M. Aubry, S. Paris, S. W. Hasinoff, J. Kautz, and F. Durand, "Fast local laplacian filters: Theory and applications," *ACM Transactions on Graphics*, vol. 33, no. 167, Aug. 2014.

[4] G. Fein, J. Raz, F. F. Brown and E. L. Merrin, "Common reference coherence data are confounded by power and phase effects," *Electroencephalography and clinical neurophysiology*, vol. 69, pp. 581-584, 1988.

[5] H. Ramoser, J. Muller and G. Pfurtsceller, "Optimal Spatial Filtering of single trial EEG during imagined hand movement," *IEEE Trans Rehab Eng*, vol. 8, pp. 441-446, 2000.

[6] F. Lotte, and C. Guan, "Regularizing common spatial patterns to improve BCI designs: unified theory and new algorithms," *Biomedical Engineering, IEEE Transactions*, vol. 58, pp. 355-362, 2011.

[7] H. Wang, Q. Tang, and W. Zheng, "L1-norm-based common spatial patterns," *Biomedical Engineering, IEEE Transactions*, vol. 59, pp. 653-662, 2012.

[8] Y. Wang, S. Gao and X. Gao, "Common Spatial Pattern Method for Channel Selelction in Motor Imagery Based Brain-computer Interface," *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, Shanghai, 2005, pp. 5392-5395.

[9] A. Balzi, F. Yger, and M. Sugiyama, "Importance-weighted covariance estimation for robust common spatial pattern," *Pattern Recogniton Letters*, vol. 68, pp. 139-145, 2015.

[10] R. O. Duda, P. E. Hart, and D. G. Stork, "Pattern classification and scene analysis 2nd ed." *ed: Wiley Interscience*, 1995.

[11] G. B. Huang, Q. Y. Zhu, and C. K. Siew, "Extreme learning machine: theory and applications," *Neurocomputing*, vol. 70, no. 1, pp. 489-501, 2006.

[12] M. Ayman, Y. Rubiyah, and S. Mohd Ibrahim "Extreme Learning Machine for Structured Output Space," unpublished.

[13] S. Ding, H. Zhao, Y. Zhang, and R. Nie, "Extreme learning machine: algorithm, theory, and applications," *Artif Intell Rev*, vol. 44, pp. 103-115, 2015.

[14] R. Agarwal, J. Gotman, D. Flanagan, and B. Rosenblatt, "Automatic EEG analysis during long-term monitoring in the ICU," *Electroencephalography and clinical Neurophysiology*, vol. 107, no. 1, pp. 44-58, 1998.

[15] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K. R. Muller, "Optimizing spatial filters for robust EEG single-trial analysis," *Signal Processing Magazine, IEEE*, vol. 25, no. 1, pp. 41-56, 2008.

[16] G. Pfutscheller, C. Neuper, D. Flotzinger and M. Pregenzer, "EEG-based discrimination between imagination of right and left hand movement," *Electroenc. Clin. Neurophys.*, vol. 103, no. 5, pp. 1-10, 197.

[17] Y. Huang, and D. Lai, "Hidden node optimization for extreme learning machine," *AASRI Procedia*, vol. 3, pp. 375-380, 2012.

[18] M. Ayman, Y. Rubiyah, S. Mohd Ibrahim, "Extreme learning machine for structured output space," unpublished.