Novel Classification System for Classifying Cognitive Workload Levels under Vague Visual Stimulation

R. Mahmoud[1], T. Shanableh[1], I. P. Bodala[3], N. Thakor[3], H. Al-Nashash[2],

[1] Department of Computer Science and Engineering, American University of Sharjah, UAE
{g00060845, tshanableh}@aus.edu
[2] Department of Electrical Engineering, American University of Sharjah, UAE
hnashash@aus.edu
[3] Singapore Institute for Neurotechnology, National University of Singapore
indu.iitbbs@gmail.com , sinapsedirector@gmail.com

***Abstract*— This paper presents a novel method for classifying four different levels of cognitive workload. The workload levels are generated using visual stimuli degradation. EEG signals recorded from 16 subjects were used for workload classification. The proposed solution includes preprocessing of EEG signals and feature extraction based on statistical features. This is followed by variable selection using stepwise regression and multiclass linear classification. The presented method achieved an average classification accuracy of 93.4%. The effect of EEG channel selection on the classification accuracy is also investigated. In comparison to the existing work, we show that the proposed solution is more accurate and computationally less demanding.**

***Index Terms*—Channel selection, cognitive workload, EEG, stepwise regression.**

I. Introduction

Cognitive workload is the amount of mental effort being used in the working memory while performing a mental task. Workload levels are affected by many factors including the requirements of the task, the environment in which the task is being performed and the perceptual capabilities and skills of the performer [1]. For example unskilled or novices are expected to experience more cognitive workload during the implementation of a new task than those who are more familiar with it [2]. Stimuli type such as visual and auditory, determines the conscious perception. The level of stimulus saliency has a great effect on the amount of mental resources needed to process information. Additionally, anticipation or having a priori knowledge of the stimulus can influence the time to perceive information, and consequently reduce the amount of cognitive workload placed on the brain [3].

There is a serious need to accurately quantify the level of cognitive workload. Many jobs are greatly influenced by extremely high or low cognitive workload. These jobs include military, clinical, industrial, computer-based assistance [4], or even driving and gaming. An optimum level of cognitive workload should be maintained while performing these tasks. Deviation from the optimum workload may lead to reduction of cognitive efficiency. Consequently, this will result in degradation in performance [5]. Such degradation can affect the memory, the learning process and decision-making process. Internal and external factors, such as the level of noise in the environment, can affect the current cognitive state and workload [6].

Accurately measuring cognitive workload can be a challenging task. In general, there are four main methods for assessing cognitive workload: analytical, subjective, behavioral and physiological methods [7]. The analytical measures are based on modeling of workload simulation. Subjective methods on the other hand rely on the subjects themselves for rating the different mental tasks. In the behavioral methods, the user performance can help in determining and assessing the cognitive state. Metrics used in these methods include reaction time and accuracy. In comparison with the other suggested methods, physiological methods are found to be objective and have less interference with the main task [8]. One very important physiological means for assessing cognitive workload is using the Electroencephalogram (EEG) signals. EEG reflects the voltages produced by the ionic currents of the brain neurons [9]. Many applications related to understanding or assessing brain

Mahmoud, R., Shanableh, T., Bodala, I.P., Thakar, N., & Al-Nashash, H. (2017). Novel classification system for classifying cognitive workload levels under vague visual stimulation. IEEE Sensors Journal, July, DOI 10.1109/JSEN.2017.2727539 July, 2017.

functionalities use EEG signals. Examples of these applications include sleep disorders, controlling the process of anesthesia, coma, epilepsy and brain death [10]. In addition, EEG signals are very sensitive to changes and variations in alertness and attention.

EEG signals were used by Zarjam *et al.* in [11] to classify seven levels of cognitive workload. The signals contaminated with electrooculogram (EOG) artifacts were discarded. The signals were then filtered with a band-pass filter with frequency bandwidth of (0.5-30Hz). The classification was carried out using Artificial Neural Networks (ANN), and by applying discrete wavelet transform (DWT) for feature extraction. Baldwin and Penaranda in [12], used EEG signals in classifying three different working-memory tasks. After visually inspecting the EEG signals to remove noisy channels, the signals from the remaining channels were filtered using a (0.1- 70 Hz) band-pass filter. The feature vectors consisted of the spectral power calculated from five frequency bands, and ANN classifiers were used for classification. In [13], K Yu *et al.* used Bilinear Common Spatial Patterns (BCSP) feature extraction method to classify four levels of cognitive workload. Linear probabilistic support vector machine (LIBSVM) classifiers were adopted for classification. Signals preprocessing included a band-pass filter of (0.3-40 Hz), and a second-order blind identification based method for EOG artifact removal. The classification accuracy of this method reached around 87%. However, a higher classification accuracy would be preferred. Wireless EEG signals was explored in [14] to assess memory workload. The workload levels of n-back tasks were classified for 9 subjects. The system proposed in that paper included automatic artifacts removal, feature extraction, feature scaling, feature selection and classification. Four groups of feature extraction techniques were employed: spectral power density, statistical features (mean, variance, skewness, and kurtosis), morphological features (curve length, number of peaks and average non-linear energy), and time–frequency features based on four-level DWT. Wavelet entropy was employed for decreasing the dimensionality of the features. In [14], the classification system was based on a support vector machine, which resulted in an average classification accuracy of 82%.

In our proposed work, we present a novel method for classifying four workload levels using EEG signals. The proposed method includes preprocessing of the EEG signals, feature extraction and classification. The objective of preprocessing is to include only the frequency components, which increase classification accuracy. Additionally, any signal contamination due to eye movements is removed. The second step is extracting feature vectors followed by variable selection to retain important features variables. The final step is feature modelling using a linear classifier.

The rest of this paper is organized as follows. Section II, reviews the experimental protocol followed by a review of existing methods used for feature extraction and classification. In section III, the proposed solution for preprocessing, feature extraction, variable selection and classification is presented. Also, channel selection is explained in this section as an optional step in the classification system. The experimental results are then presented in section IV. Finally, the conclusion of this work is provided in section V.

II. METHOD

A. *Experimental Protocol*

The experimental setup and data collection were performed as explained in [13]. The experiment included 16 healthy subjects. Before the experiment starts, each subject is tested for color blindness and has to go through dominant eye test. It is also insured that all participants are not on medication and have not experienced any neurological or cardiovascular diseases. Also, the participants did not suffer from any psychiatric disorders or hypertension. The subjective NASA Task Load Index (NASA-TLX) questionnaire is completed by every participant before and after each cognitive workload level throughout the experiment. NASA-TLX is an assessment method based on different subscales that act as sources of workload demand. These subscales are: Mental Demands, Physical Demands, Temporal Demands, Own Performance, Effort and Frustration [15].

The experiment is conducted using a 24″ monitor to display the visual stimuli. The total time for the experiment including the preparation time is around 90 minutes. Four levels of cognitive workload are tested, with each level lasting for about 10 minutes. During that time, and while the experiment is being performed, EEG signals are recorded from 62 channels with sampling rate of 512 Hz. Two additional channels are collected from the vertical electrooculogram (EOG) and the electrocardiogram (ECG). EEG signals are referenced to linked ears and grounded to the forehead.

For each workload level, the subjects are asked to identify the human face image displayed in gray-level on the monitor as the target, by pressing the letter ‘Q’. Otherwise, if the image is for anything other than a human face (non-target), subjects should press the lfetter ‘P’. The difficulty level of each task increase from level 1 as the easiest to level 4 as the most difficult cognitive task.

Figure 1 illustrates the sequence of the visual stimuli. Each trial starts by displaying a fixation cross (+) that lasts for 500 ms, after that a digit (Digit 1) is displayed for 300 ms, then another digit (Digit 2) is displayed for 300 ms. Then, an image (256×256) is displayed for 300 ms. Finally, a maximum time

Mahmoud, R., Shanableh, T., Bodala, I.P., Thakar, N., & Al-Nashash, H. (2017). Novel classification system for classifying cognitive workload levels under vague visual stimulation. IEEE Sensors Journal, July, DOI 10.1109/JSEN.2017.2727539 July, 2017.

window of 3000 ms is left for the subjects to respond (Response) as fast as possible by pushing letters ‘Q’ or ‘P’. The next trial is then initiated immediately after the response. The previous sequence is repeated for all levels. Appropriate markers are also simultaneously recorded with the various experiment events.

Figure 2 provides an example for each of the 4 cognitive workload levels. Following, is a detailed description of these levels:

In Level 1: the target is simply an image of a human face. The target can easily be identified from the image itself regardless of the values of the previous digits.

In Level 2: the target is only the human-face image that is preceded by two numbers that are both either odd or even.

In Level 3: the target is the same as the previous one, however, the signal to noise ratio (SNR) of the images is reduced to 0 dB. It is hypothesized that this will increase the level of cognitive workload imposed on the subjects.

Lastly, Level 4: is similar to the previous two levels, however, the SNR of the images is decreased further to -5dB. The low SNR is needed to increase the cognitive workload placed on the subjects.

The trials presented to a subject are grouped based on the cognitive workload level. Samples of the four workload levels are explained and presented to each subject in sequential order before the experiment starts. In each workload level, a total of 210 different stimuli sequences or trials containing only 30 target sequences are displayed. However, it should be stress that each workload level is presented to the subject in random order. This is important as to avoid any adaptation that could affect the perception of the subjects during the experiment. Note that randomization of workload levels presented to the subjects is not possible because the requirements for identifying the target differs according to the underlying level. In addition, the subject has to fill in the NASA-TLX questionnaire after each workload level experiment.

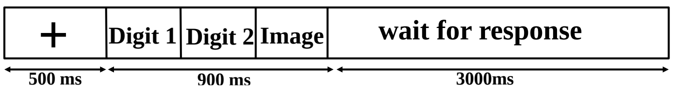

Figure 1.The sequence of visual stimuli

The recorded EEG data is collected from 62 different channels referenced to the two channels recorded from the ears. One EOG channel, and one ECG channel are also simultaneously recorded. All the signals are collected and sampled using the ANT amplifier (ANT B.V., Enschede, Netherlands).

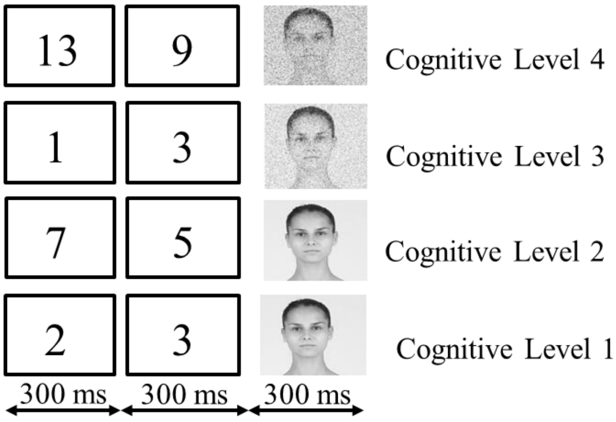

Figure 2. Sample targets for each of the 4 workload levels

Before data recording begins, the subject’s head and earlobes are cleaned with some alcohol. The cap is then manually positioned according to the 10-20 system. This is then followed by filling the electrodes with conductive gel such that the impedance of each electrode is less than 10 KΩ. The two earlobe reference electrodes are then attached and gel is injected. The EOG two electrodes are connected above and below the subject’s left eye. Some baseline EEG and EOG signals are recorded for few minutes before the experimental paradigm begins.

B. Existing Feature Extraction and Classification

BCSP is a method based on conventional Common Spatial Patterns (CSP) method [16]. In [13], BCSP method was applied to extract features based on both spatial and temporal projections, from epoch segments starting from the onset of the image to 500 ms after the onset of the image. BCSP aims to maximize the power ratio of the two objective functions expressed as:

$$\max_{W,V} \frac{det\left(W^T X_+ V V^T {X_+}^T W\right)}{det\left(W^T X_- V V^T {X_-}^T W\right)} \tag{1}$$

$$\max_{W,V} \frac{det\left(W^T X_- V V^T {X_-}^T W\right)}{det\left(W^T X_+ V V^T {X_+}^T W\right)} \tag{2}$$

Where ***W*** and ***V*** are spatial and temporal projections, respectively. ***Xc*** is the EEG data of condition c which includes both target epochs and non-target epochs and *det(•)* is the determinant operator.

Both of the objective functions were updated iteratively, until the ratio value of the objective functions converges. The first two spatial projections and first four temporal projections obtained from each objective function were kept during the iteration and were finally used for feature extraction. The reason

Mahmoud, R., Shanableh, T., Bodala, I.P., Thakar, N., & Al-Nashash, H. (2017). Novel classification system for classifying cognitive workload levels under vague visual stimulation. IEEE Sensors Journal, July, DOI 10.1109/JSEN.2017.2727539 July, 2017.

for this is explained in details in [17]. Four features (two from each objective function) were finally obtained from the BCSP. The total number of feature variables for each epoch is 24.

Linear probabilistic LIBSVM classifiers were used for classification. The extracted features were used for classification using six classifiers built for all possible pairwise combinations of the four conditions. 80% of the entire data was used for training the classifiers while the remaining 20% was used for testing. An average single trial accuracy of above 80% was achieved for all four levels.

C. Proposed Method

In this section we present a novel classification algorithm that produces a higher classification accuracy than what is published in the literature. Figure 3 illustrates the general procedure of the proposed solution. The first step is the preprocessing of EEG signals using filtering techniques. The next step is channel selection. Such a step is considered optional as feature extraction and model generation are typically applied to all EEG channels. However, in the experimental results section we show that a subset of the channels is sufficient to generate high classification results. Channel selection is followed by our proposed feature extraction method, which is based on Discrete Wavelet Transform (DWT) and statistical measures. The dimensionality of the feature vectors is then reduced using stepwise regression. Lastly, a simple linear classifier is used to model and predict the cognitive workload levels.

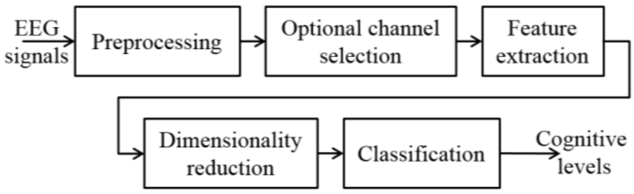

Figure 3. Proposed cognitive workload classification solution

The collected signals are filtered to remove the low frequency drift or high frequency noise using the Hamming windowed sinc FIR filter with a pass-band of (0.1-60 Hz) [18]. The data segment (epoch) of each sequence/trial was extracted from the filtered signals. The time window of such an epoch starts from the onset of Digit 1 to 500 ms after the onset of the image, i.e. 1100 ms in total. The reference EOG signals recorded earlier are then used to in a second filter to remove any undesirable eye blink EOG artifacts using a second-order blind identification (SOBI) based method.

Channel selection is the process of selecting the channels that are most relevant to the process of cognitive workload classification. It reduces the processing and data acquisition complexity. It also improves the overall performance by reducing utilization of unnecessary channels [19]. In this work, we study the effect of channel selection on the overall classification accuracy. We apply the channel selection approach reported in [20]. As illustrated in Figure 3, channel selection is performed prior to feature extraction and classification. In the experimental results section we apply the channel selection algorithm to our data and visualize the locations of the selected channels.

In this section we review the channel selection algorithm proposed by [20] for completeness. The reviewed work proposed a channel selection algorithm with application to person authentication. The data set contains EEG recordings from 50 subjects performing six mental tasks. Feature extraction was applied to individual channels based on the PSD of six 8 Hz frequency bands. The channels were ranked based on a stability criterion. The stability of each channel is calculated using:

$$\boldsymbol{s_i} = \boldsymbol{b_i} - \boldsymbol{w_i} \tag{3}$$

The stability of channel i is defined as the between-subject distance b_i, and within-subject distance w_i. To find the latter distance we start by computing the average Mahalanobis distance between the means of the feature vectors of the cognitive workload levels for the same channel and subject. This is repeated for all subjects and the overall average distance is referred to as within-subject distance. Such a quantity measures the distance between the means of the feature vectors during performing tasks with different workloads within the same subject and same channel. The between-subject distance b_i is calculated by finding the Mahalanobis distance between the means of the feature vectors of the same cognitive workload level and channel but for different subjects. The average over all cognitive workload levels is then computed. Such a quantity measures the distance between the means of the feature vectors of the same channel during performing a task with a specific cognitive workload but in different subjects [20].

D. Feature Extraction

Figure 4 illustrates the process of feature extraction used in our proposed solution.

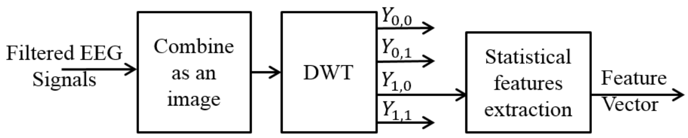

Figure 4. Proposed feature extraction process

The filtered signals from the 62 channels are stored in a matrix and treated as a 2-D image, then, a discrete 2-D wavelet transform (DWT) is applied. We used the Haar wavelet transformation for its simplicity and speed of computation. Applying the DWT on the EEG signals results in four frequency subbands. These subbands contain the approximation coefficients matrix ($\mathbf{Y_{0,0}}$) and the detail coefficients matrices of

Mahmoud, R., Shanableh, T., Bodala, I.P., Thakar, N., & Al-Nashash, H. (2017). Novel classification system for classifying cognitive workload levels under vague visual stimulation. IEEE Sensors Journal, July, DOI 10.1109/JSEN.2017.2727539 July, 2017.

the horizontal ($\mathbf{Y}_{0,1}$), vertical ($\mathbf{Y}_{1,0}$) and diagonal ($\mathbf{Y}_{1,1}$) edges. With a one level wavelet transformation, the dimensions of each subband is half of the input, which is 62/2 rows and 257/2 columns. The number of columns pertains to the length of each epoch.

Statistical features are then extracted from the vertical edges matrix ($\mathbf{Y}_{1,0}$). This matrix is chosen because it represents the temporal dynamics of the EEG signals.
The extracted features included:

i. Spatial and temporal means of the $\mathbf{Y}_{1,0}$ EEG subband.
ii. Spatial and temporal standard deviations of the $\mathbf{Y}_{1,0}$ EEG subband
iii. Spatial and temporal entropy of the $\mathbf{Y}_{1,0}$ EEG subband.
iv. The spatial covariance matrix of the $\mathbf{Y}_{1,0}$ EEG subband.

Since the covariance matrix is symmetrical, then only the values above the main diagonal are retained and represented as a vector.

The outcome of feature extraction process is a set of feature vectors that represent the EEG data. To ensure that the different feature variables are on similar scales, we use the z-score normalization. For a feature variable X the z-score of data point x_i is defined as:

$$\boldsymbol{z_i} = \frac{(\boldsymbol{x_i} - \boldsymbol{\mu_X})}{\boldsymbol{\sigma_X}} \qquad (4)$$

In terms of implementation, the z-scores are applied to the training dataset only. The resultant means and standard deviations of which are then used to normalize the test dataset.

E. Variable Selection

The dimensionality of the resultant feature vectors is detailed in TABLE 1. The total length is 945 variables. Such dimensionality is considered high and may affect the performance of the classifier if not enough feature vectors are available in the training phase. Therefore, we used stepwise regression to retain the most important variables and potentially reduce dimensionality as a by-product. The reason we use stepwise regression is that it produces an initiative result, such that only feature variables that are needed for discriminating the four cognitive workload levels are retained. In the experimental results section we list the feature variables that are retained for each subject.

TABLE 1
DESCRIPTION OF THE FEATURE VECTOR CONTENT AND SIZE

Feature	No. of variables
Mean of subband $\mathbf{Y}_{1,0}$ columns	129
Standard deviation of subband $\mathbf{Y}_{1,0}$ columns	129
Mean of subband $\mathbf{Y}_{1,0}$ rows	31
Standard deviation of subband $\mathbf{Y}_{1,0}$ rows	31
Entropy of subband $\mathbf{Y}_{1,0}$ columns	129
Entropy of subband $\mathbf{Y}_{1,0}$ rows	31
Spatial covariance matrix	465
Total	945

Figure 5 illustrates the process of variable selection. Stepwise regression is applied to the feature vectors of the training dataset. The output from the stepwise regression is the indices of the retained feature variables. These indices are stored and used to reduce the number of features in the test dataset.

Stepwise regression is a method of variable selection, but it can also be used for dimensionality reduction as proposed by [21]. The stepwise regression procedure includes forward selection and backward elimination steps. Forward selection starts with the simplest model of all (i.e. one feature variable), and adds suitable variables one at a time until the "best" model is reached. Backward elimination works with the most general model, and drops variables one at a time until the "best" model is reached. Stepwise, on the other hand, is a combination of both forward and backward methods, where variables can be dropped and added.

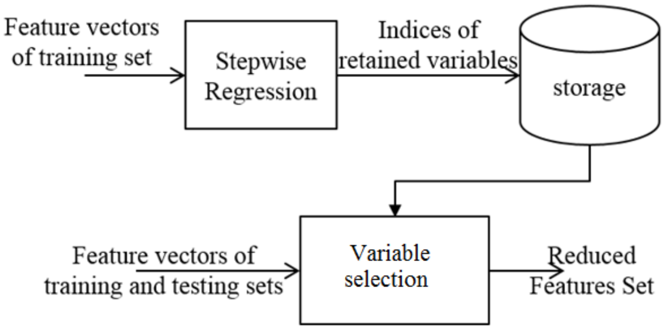

Figure 5. Variable selection using stepwise regression

For a set of variables x_1, x_2... x_k that belongs to cognitive level y, f_{in} is the F-random variable for adding a variable to the model and f_{out} is the value of the F-random variable for removing a variable from the model. The stepwise regression is defined by the following steps [22]:

In the first step, all variables are examined one by one, to generate a one-variable model in the form of:

$$\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{\theta_0} + \boldsymbol{\theta_1 x_1} \qquad (5)$$

Where $h(x)$ denotes the hypothesis that the included variables are needed for the classification of the cognitive level. x_1 is one of the k variables that results in the highest F-statistics. In the second step, the remaining k-1 variables are examined to choose the second best variable such that the model in (6) gives the best classification result. Here, x_2 is added such that its F-statistic $\boldsymbol{f_2}$ is greater than f_{in}.

$$\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{\theta_0} + \boldsymbol{\theta_1 x_1} + \boldsymbol{\theta_2 x_2} \qquad (6)$$

f_2 is calculated by:

Mahmoud, R., Shanableh, T., Bodala, I.P., Thakar, N., & Al-Nashash, H. (2017). Novel classification system for classifying cognitive workload levels under vague visual stimulation. IEEE Sensors Journal, July, DOI 10.1109/JSEN.2017.2727539 July, 2017.

$$f_2 = \frac{SS_R(\theta_1|\theta_2\theta_0)}{MS_E(x_1, x_2)} \tag{7}$$

SS_R denotes the regression sum squares error and MS_E denotes the mean square error. After x_2 is chosen, the algorithm rechecks if x_1 is to be removed. This is done by comparing f_1 to f_{out}. If f_1 is less than f_{out}, x_1 is removed from the model. Where f_1 is calculated using (8).

$$f_1 = \frac{SS_R(\theta_2|\theta_1\theta_0)}{MS_E(x_2, x_1)} \tag{8}$$

In the third step, for the remaining (k-2) variables, the third best variable is included by testing 3-variables model such that the resulting model has the best classification result. The stepwise regression algorithm continues until no more variables can be included or removed from the model.

F. *Classification*

In our proposed solution, we use a simple linear classifier, as it proved to be efficient with the proposed feature extraction method. In linear classifiers, the class label is determined by a linear combination of features with predetermined weights. The weights are the attributes of the model, thus the hypothesis or the predicted class will be:

$$\boldsymbol{h(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \ldots + \theta_N x_N} \tag{9}$$

Where θ_j's are the attributes of the model. For k classes we can define the input matrix X as:

$$\boldsymbol{X = [X_1, X_2, \ldots X_k]^T} \tag{10}$$

Where X_i is a matrix with all the feature vectors belonging to class i. With M features, the size of X_i is $N_i X M$. The optimum set of weights for the i^{th} class is defined as:

$$\boldsymbol{\theta_i^{opt} = argmin_{\theta_i} \|X\theta_i - y_i\|_p} \tag{11}$$

With:

$$\boldsymbol{y_i = [\,0_{N_1}, 0_{N_2}, \ldots 1_{N_i}, \ldots 0_{N_k}]^T} \tag{12}$$

Where y_i is the ideal output vector for the i^{th} class, which should be all zeros except when the input feature vector belongs to the same class. N_i is the number of training examples of the i^{th} class. The $\|.\|_p$ operator is the second norm (p=2). The closed form solution for (11) that gives the optimum weights is defined as:

$$\boldsymbol{\theta_i^{opt} = (\sum_{j=1}^{K} X_j^T X_j)^{-1} X_i^T 1_i} \tag{13}$$

Where 1_i is the target vector of class i, which is comprised of zeros and ones, in the manner as the ideal output vector. The result of the training phase is four sets of weights $\boldsymbol{\{\theta_1, \theta_2, \theta_3 and\ \theta_4\}}$, a set for each cognitive level. It should be noted that the dimensionality of the feature vectors can cause numerical instability when computing the model weights in equation (13). More specifically, higher feature dimensionality can increase the condition number of the feature matrix leading to an ill-conditioned matrix that produces an unreliable set of model weights. In [23], it was reported that for non-linear problems, the feature vectors can be expanded into higher orders using polynomial expansion prior to the use of equation 13. However, in this work we use the first order only without feature vector expansion to avoid numerical instability. The condition number averaged over all subjects without polynomial expansion was 57.2. However, the condition number increased significantly to 1.3973E+18 and 7.1542E+18 when the feature matrix was expanded to the second and third polynomial orders respectively. In [24], it was mentioned that ridge regression tends to produce more stable set of weights, however this was not required in our work since the feature matrices are not ill-conditioned.

In the testing phase, the features of the testing sets are extracted, normalized and reduced in dimension. As explained above, the variable selection uses the indices generated by the stepwise regression applied to the training dataset. Each feature vector from the testing set $\boldsymbol{X_t}$ is multiplied using dot product to calculate the set of scores $\{s_i\}$ as follows:

$$\boldsymbol{s_i = X_t \theta_i^{opt}} \tag{14}$$

It is important to note that such a multi-class classifier discriminates different cognitive workload levels against each other. The classifier results in 4 set of weights one for each cognitive workload level.

The process of model generation and classification is illustrated in

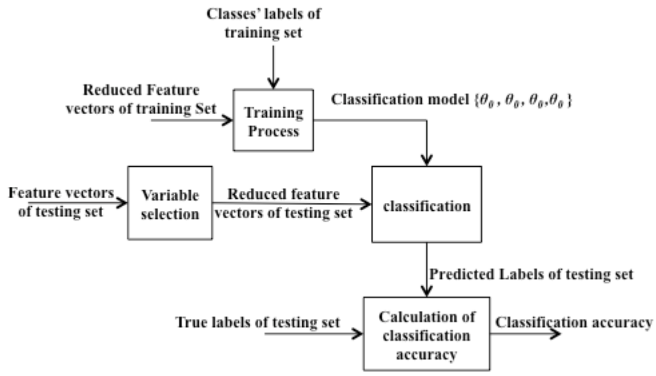

Figure 6. Based on the reduced feature vectors of the training set and their corresponding class labels, the training process generates the classification model. Such a model contains a

Mahmoud, R., Shanableh, T., Bodala, I.P., Thakar, N., & Al-Nashash, H. (2017). Novel classification system for classifying cognitive workload levels under vague visual stimulation. IEEE Sensors Journal, July, DOI 10.1109/JSEN.2017.2727539 July, 2017.

weight set for each of the different levels of cognitive workload. Features are extracted from the testing dataset using the same process explained in Figure 4. These vectors are then reduced in dimensionality using the indices of retained variables generated by the stepwise regression procedure. Linear classifier is then used for classifying the testing feature vectors using the generated models. The true labels of the testing feature vectors are then used to calculate the classification accuracy.

Prior to presenting the classification results in the next sections, we summarize the steps of the proposed cognitive workload classification method:

1. The EEG signals are pre-processed using band-pass filter (0.1-60 Hz) and SOBI filter.
2. The signals are combined into an image of dimensions 62x257.
3. Statistical features are extracted from the $Y_{1,0}$ subband of the DWT image.
4. Feature vectors are normalized using the z-score method.
5. Stepwise regression is applied to reduce the dimensionality of the feature vectors.

A linear classifier is used for classification.

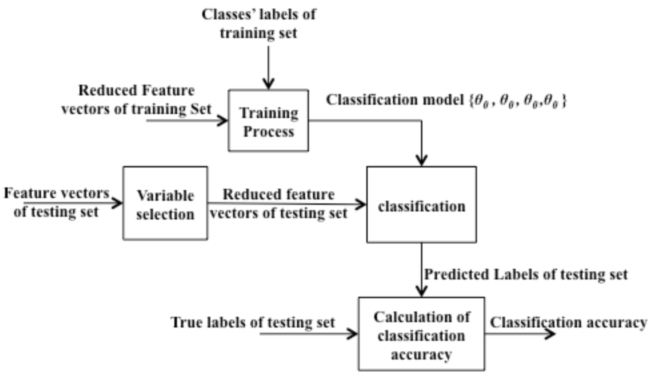

Figure 6. Block diagram of the proposed classification system

III. Experimental Results

In the experiments to follow, each epoch contains 62 EEG channels. Only the EEG signals recorded during the response period (500 ms) were considered in the analysis. Hence, each training or testing epoch contains 62 channels with 257 samples. Similar to the experimental setup used in [13], a randomly selected 80% of the data is used in the training phase; the remaining 20% of the data is used for testing. The process of selecting the training data set is repeated 5 times, generating 5 different sets, and the average and standard deviations of the classification results are calculated and reported. The exact numbers of feature vectors for subjects 1-16 are as follows: 840, 840, 840, 839, 840, 840, 808, 839, 840, 830, 840, 840, 840, 838, 840 and 840 respectively. The stepwise regression procedure is used in our proposed solution for variable selection and dimensionality reduction as a by-product. The output of the procedure is the indices of retained features. Different number of feature variables are retained for different subjects, the minimum number of variables is 33, the maximum is 75 and the average is 56 variable. The exact numbers of retained features per subject are 63, 42, 73, 70, 61, 57, 33, 60, 57, 54, 46, 69, 55, 43, 75 and 43. TABLE 2 provides the stepwise regression output for all subjects. As reported in TABLE 2, all features are extracted from the $Y_{1,0}$ sub-band. The total number of columns and rows are 129 and 31 respectively, and the total number of variables in the covariance matrix is 465. The last row in the table presents the average number of variables retained per feature variable. It is shown that most variables are retained from the covariance matrix and standard deviations. Fewer variables are retained form the means and entropy.

TABLE 2
FEATURES RETAINED BY THE STEPWISE REGRESSION

Sub-ject	**Mean**		**Standard deviation**		**Entropy**		**Cov. matrix**
	Cols 129	Rows 31	Cols 129	Rows 31	Cols 129	Rows 31	465
1	5	2	9	4	0	1	44
2	0	0	4	5	0	0	31
3	11	4	8	9	0	0	45
4	6	0	2	5	0	2	51
5	0	0	2	7	0	0	52
6	3	0	9	7	0	0	38
7	0	0	0	3	0	0	30
8	0	1	1	3	4	0	52
9	4	3	3	6	1	0	42
10	6	0	7	4	0	0	34
11	4	0	3	7	0	0	32
12	0	3	4	9	0	0	56
13	2	0	3	9	1	1	36
14	2	1	6	6	0	0	29
15	8	1	11	10	0	0	45
16	2	0	4	5	0	0	31
Avg	**3.3**	**1**	**4.8**	**6.2**	**0.4**	**0.3**	**41**

The corresponding classification results are presented in Figure 7 for both the existing work [13] and the proposed solution. The x-axis represents the 16 subjects while the y-axis represents the average classification rate from the five runs. The error bars represent the standard deviations of the five runs per

Mahmoud, R., Shanableh, T., Bodala, I.P., Thakar, N., & Al-Nashash, H. (2017). Novel classification system for classifying cognitive workload levels under vague visual stimulation. IEEE Sensors Journal, July, DOI 10.1109/JSEN.2017.2727539 July, 2017.

subject. As shown in the figure, the lowest classification accuracy is 84.9%, while the highest is 98.8%. The average classification accuracy of the proposed solution for all subjects is 93.4%. Whereas the lowest, highest and average classification accuracies of the reviewed work are 80.9%, 91.5% and 87.6% respectively.

In addition, we performed statistical analysis on the classification results presented in Figure 7. We first ran a paired t-test for the null hypothesis:

- Null hypothesis: Mean_ existing - Mean_ proposed = 0

we got a very small value $p < 0.05$. This indicates a significant difference between the mean values of the existing and proposed methods. We then tested the alternative hypothesis:

- Alternative: Mean_ existing -Mean_ proposed < 0

The t-statistics value was -5.61 with value $p < 0.05$. We conclude that at 5% significance level, the data provide enough evidence to support the hypothesis that, on average, the new proposed method provides higher classification rate than the existing methods.

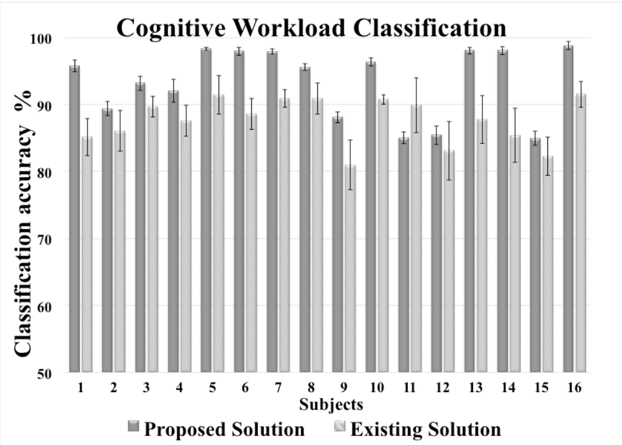

Figure 7. Cognitive workload classification results of the 16 subjects

As mentioned in Section II, a linear classifier is used for model generation and classification. Nonetheless, we have also experimented with a nonlinear classifier using a second order polynomial expansion, which resulted in an average classification result of 91.9%. This is an indication that the linear classifier is suitable for the task at hand.

Figure 8 illustrates the average classification accuracy for each of the four levels, for both the proposed solution and the reviewed work reported in [13].

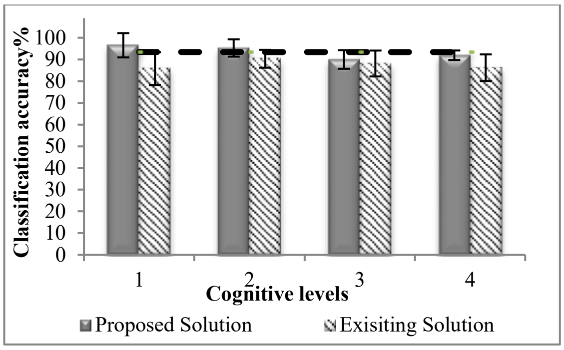

Figure 8. Cognitive workload classification results of the 4 cognitive levels

The error bars represents the standard deviation of the classification accuracy resulting from the five runs of the experiment. It can be seen that the overall average classification accuracy of the proposed method, is 93.4%. The average is shown in the dashed horizontal line in the figure. The lowest classification accuracy is 90% for Level 3, and the highest is 96% for Level 1. In general, the accuracies of Level 1 and Level 2 are higher than those of Level 3 and Level 4. This indicates that Level 1 and Level 2 are more separable than the other two levels. From the figure, it is clear that the proposed solution results in higher classification rate for each level than the existing work. Again, the variation between the five runs per subject which is represented by the error bars is also lower than the existing work.

The confusion matrices of the existing work and the proposed method are presented in Figure 9. The presented results are the average of all subjects. Considering the cases where the probability of confusion is more than 5%, in the existing work, Level 1 can be confused with Level 4 and Levels 2-3 can be confused with Level 1. Whereas in the proposed solution, Level 4 can be confused with Level 3 only. In summary, the number of cognitive workload levels involved in classification confusion is reduced from 4 to 2 by using the proposed solution.

		Predicted Cognitive level			
		L1	**L2**	**L3**	**L4**
True cognitive level	**L1**	85.5%	3.8%	4.4%	**5.7%**
	L2	**5.6%**	89.9%	3.9%	3.5%
	L3	**5.2%**	3.6%	88%	3.3%
	L4	3.7%	2.7%	3.7%	87.5%

(a) Reviewed work

		Predicted Cognitive level			
		L1	**L2**	**L3**	**L4**
True cognitive	**L1**	96.4%	2.5%	1.2%	1.5%
	L2	1.2%	95.2%	2.5%	2.1%
	L3	1.2%	1.7%	90.4%	4.3%

Mahmoud, R., Shanableh, T., Bodala, I.P., Thakar, N., & Al-Nashash, H. (2017). Novel classification system for classifying cognitive workload levels under vague visual stimulation. IEEE Sensors Journal, July, DOI 10.1109/JSEN.2017.2727539 July, 2017.

	L4	1.2%	1.1%	**5.9%**	92.1%

(b) Proposed solution

Figure 9. Cognitive levels classification confusion matrix

As mentioned in Section II, the $Y_{1,0}$ subband of the 2D Haar transformation was used for feature extraction. However we have also experimented with other 2 subbands (i.e $Y_{0,1}$,and $Y_{1,1}$). The average classification results over the 16 subjects, are 77.3% and 92.2%. This is a further indication that the $Y_{1,0}$ subband is most suitable for feature extraction.

We have also experimented with 2-Level 2D-HAAR in which the approximation subband, $Y_{0,0}$, goes through a second level of 2D–HAAR transformation. We then applied the feature extraction and classification solutions using the 3 resultant subbands. The classification accuracies were 65.4%, 87.9% and 86.2%. This gives an indication that 1-Level 2D-HAAR transformation is most suitable for the proposed feature extraction and classification solutions.

As mentioned in Section II-D, we have used the Haar wavelet transformation in feature extraction. It is worth mentioning that we have also experimented with other wavelet transforms like db3, db4 and db5. We found that using the Haar filter results in the highest classification accuracy. More specifically, by using various wavelet transformation filters, the average classification rates are 93.3%, 90.5%, 82% and 72.6% for Haar, db3, db4 and db5 filters respectively. It is clear that by using longer wavelet filters, each EEG sample is affected by many neighboring samples in the transformation process. This has an adverse effect on the true value of the EEG readings and therefore results in reduced classification accuracy.

The computational time of the proposed method and the work presented in [13] are also measured. TABLE 3 *TIME COMPARISON BETWEEN THE EXISTING WORK AND THE PROPOSED* presents the time required by each solution to perform feature extraction and classification using 5-fold cross-validation for the 16 subjects. The time computation for the proposed solution also includes stepwise regression

These measurements are conducted using MATLAB R2012a on an Intel core-i7 processor and 8.00 GB memory computer. It is shown in TABLE 3, that the proposed solution is much faster than the reviewed work. This is due to simple statistical features used in combination with a linear classifier. This is a clear and major improvement of the proposed algorithm over the existing methods.

TABLE 3
TIME COMPARISON BETWEEN THE EXISTING WORK AND THE PROPOSED SOLUTION

Method	Time in sec
Reviewed work	6515.25
Proposed method	784.75

To study the effect of channel selection on cognitive workload classification, we start with an intuitive approach in which EEG channels are grouped based on their locations with respect to the different brain lobes. The cerebral cortex of the human brain consists of four lobes: frontal, parietal, occipital and temporal. The locations are illustrated in Figure 10 as the striped, dotted, white and grey electrodes, respectively. EEG channels in the same region comprise a subset. The Figure shows 16 channels for the Frontal lobe, 9 channels for the Parietal lobe, 10 channels for the Occipital lobe and 6 channels for the Temporal lobe. The Central channels are treated as a separate region (black) that includes 7 channels. As illustrated in Figure 10, there are 7 channels shared between the Frontal and Central regions. These were considered as part of both the Frontal and the Central EEG subsets. Additionally, the 7 channels shared between the Central and the Parietal regions are considered as part of both the Central and the Parietal subsets. Thus, the total number of channels in the Frontal, Parietal, Occipital, temporal and Central regions are 23, 16, 10, 6 and 21, respectively.

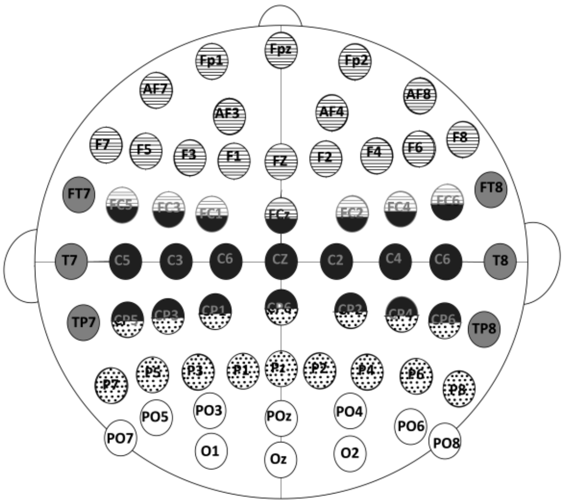

Figure 10. The locations of the EEG electrodes based on the lobes of the human brain

In Figure 11, we repeat the same proposed solution including feature extraction, stepwise regression and classification using each subset of channels individually. The objective of this test is to identify the region of the brain that contributes most to the classification of cognitive levels. The figure shows the average classification rate for all subjects. It is clear that the subset of the Frontal channels achieved the highest classification accuracy of 84.3% using 23 channels, which is the number of channels in the Frontal region. This result is consistent with the decision making functionality of this brain area [25]. This indicates that the EEG channels of the Frontal region are important in cognitive workload classification.

Mahmoud, R., Shanableh, T., Bodala, I.P., Thakar, N., & Al-Nashash, H. (2017). Novel classification system for classifying cognitive workload levels under vague visual stimulation. IEEE Sensors Journal, July, DOI 10.1109/JSEN.2017.2727539 July, 2017.

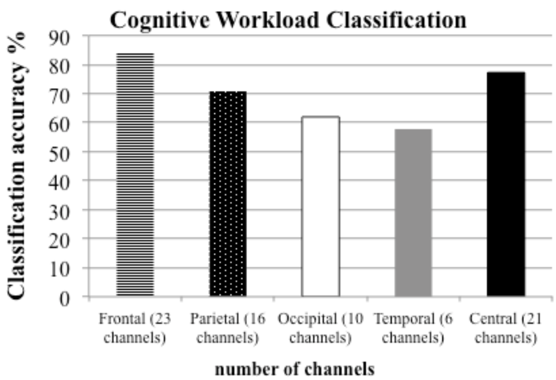

Figure 11. Classification results including channel subsets according the brain regions

The second method of channel selection examined in this work is the one reported in [20] which we introduced in Section III-B above. The ranking of all channels based on this approach is listed in TABLE 4.

TABLE 4
RANKING OF EEG CHANNELS BASED ON STABILITY SCORE

Rank	Ch. name	Rank	Ch. name	Rank	Ch. name
1	AF4	22	Fz	43	FCz
2	FPz	23	FC3	44	P4
3	FP1	24	PO6	45	FC2
4	FP2	25	PO8	46	P5
5	T8	26	C5	47	C1
6	AF7	27	CP6	48	CP3
7	01	28	Oz	49	FC4
8	F5	29	TP8	50	Pz
9	AF3	30	FT8	51	CP1
10	T7	31	PO3	52	P3
11	AF8	32	TP7	53	P2
12	F3	33	PO7	54	Cz
13	F7	34	P8	55	P1
14	C6	35	PO5	56	CP2
15	O2	36	FC1	57	C4
16	F8	37	P7	58	C2
17	F2	38	FC6	59	CPz
18	F4	39	CP5	60	CP4
19	FC5	40	POz	61	CPz
20	PO4	41	C3	62	CP4
21	FT7	42	P6		

Having ranked the EEG channels based on their stabilities using Equation (3), the classification is then carried out using a varying number of channels ranging from 1 to 62. Figure 12 illustrates the effect of channel selection on the classification accuracy. In the figure, the classification results are reported using the proposed solution and the reviewed work. The maximum classification accuracy achieved by the existing work is 87% which requires the complete set of 62 EEG channels. On the other hand, the same classification accuracy is achieved by the proposed solution, with only 15 EEG channels. Additionally, considering the top-ranked 23 EEG channels, the proposed solution achieves 91% classification accuracy, while the existing work results in 72% accuracy.

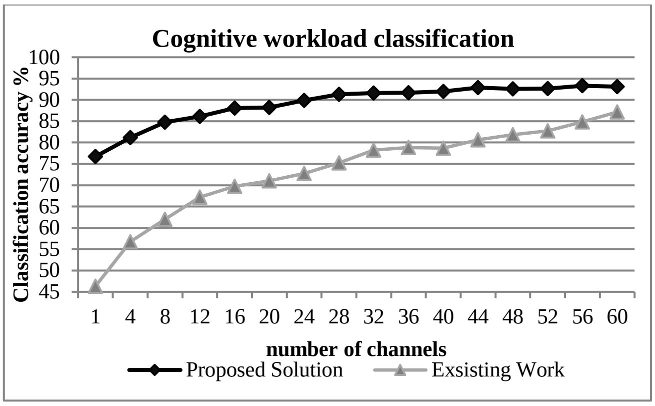

Figure 12. Classification results by varying the number of selected channels ranked by stability

Figure 13 visualizes these highlighted 23 top-ranked channels. It is clear that most of these channels reside in the Frontal region. Interestingly, this confirms the result obtained by the intuitive channel selection approach that we reported in Figure 11 above.

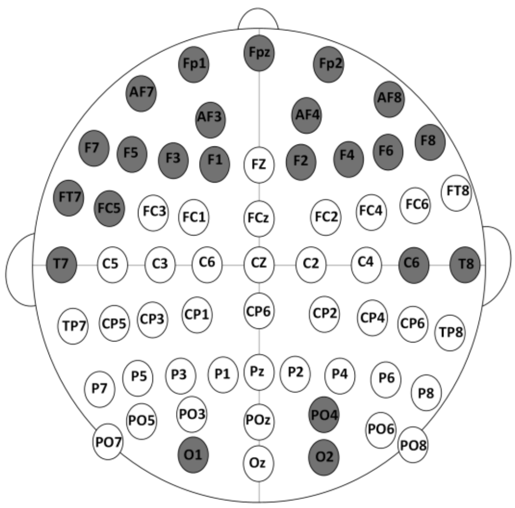

Figure 13. The locations of the top-ranked 23 channels with the highest stability

In summary, a subset of the channels can be used for classifying the level of cognitive workload into 4 classes. This is achieved by sorting channels according to their stability score [20] and using the top-ranked channels for feature extraction and classification. In Figure 12, it is shown that by using 44 out of 62 channels, an average classification accuracy of 93% is achieved. This is very similar to the result achieved by using all

Mahmoud, R., Shanableh, T., Bodala, I.P., Thakar, N., & Al-Nashash, H. (2017). Novel classification system for classifying cognitive workload levels under vague visual stimulation. IEEE Sensors Journal, July, DOI 10.1109/JSEN.2017.2727539 July, 2017.

channels. Likewise by using the top-ranked 23 channels only, the classification accuracy becomes 91%. Again, selecting channels based on their stability score resulted in higher classification accuracy than using channel subsets according to their region in the brain as reported in Figure 11.

IV. Conclusion

Classification of four cognitive workload levels was examined in this work. Statistical features were extracted from the vertical subband of a single level 2-D DWT. The temporal differences contained in the vertical subband are important in cognitive levels classification. Since the dimensionality of the feature vectors was high, stepwise regression was used for variable selection and dimensionality reduction as a by-product. It is also used to retain the important feature variables as well. A linear classifier was then used for model generation and classification. With 16 subjects and 4 cognitive levels, the proposed system resulted in classification accuracy of 93.4%. Also, the proposed solution reduces the classification confusion between the four cognitive levels to the highest two levels only. In terms of computational time, we also found that the proposed system is 8 times faster than the existing work. This increase in computational speedup is due to the simple feature extraction and classification methods used in the proposed solution. The impact of EEG channel selection on cognitive workload classification was also studied in this paper. Using an intuitive approach, it was shown that the channels of the Frontal region are the most important in cognitive levels classification. The EEG channels were also ranked using a stability criterion. Sixteen out of the 23 top-ranked channels are from the frontal region of the brain which is an expected result. Using the top-ranked 23 channels resulted in an adequate classification accuracy of 91%.

References

[1] S. G. Hart, L. E. Staveland, and A. H. a. N. M. Peter, "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research," in *Advances in Psychology*. vol. Volume 52, ed: North-Holland, 1988, pp. 139-183.

[2] J. R. Anderson, C. F. Boyle, R. Farrell, and B. J. Reiser, "Cognitive Principles in the Design of Computer Tutors," DTIC Document1984.

[3] L. Melloni, C. M. Schwiedrzik, N. Muller, E. Rodriguez, and W. Singer, "Expectations Change the Signatures and Timing of Electrophysiological Correlates of Perceptual Awareness," *JOURNAL OF NEUROSCIENCE,* vol. 31, pp. 1386-1396, 2011.

[4] L. Bi, X.-A. Fan, and Y. Liu, "EEG-based brain-controlled mobile robots: a survey," *Human-Machine Systems, IEEE Transactions on,* vol. 43, pp. 161-176, 2013.

[5] P. Chandler and J. Sweller, "The split-attention effect as a factor in the design of instruction," *British Journal of Educational Psychology,* vol. 62, pp. 233-246, 1992.

[6] D. Erdogmus, A. Adami, M. Pavel, T. Lan, S. Mathan, S. Whitlow, *et al.*, "Cognitive state estimation based on eeg for augmented cognition," in *Neural Engineering, 2005. Conference Proceedings. 2nd International IEEE EMBS Conference on*, 2005, pp. 566-569.

[7] K. G. Seeber, "Cognitive load in simultaneous interpreting: Existing theories—new models," *Interpreting,* vol. 13, pp. 176-204, 2011.

[8] J. Frey, C. Mühl, F. Lotte, and M. Hachet, "Review of the use of electroencephalography as an evaluation method for human-computer interaction," *arXiv preprint arXiv:1311.2222,* 2013.

[9] F. L. da Silva, "EEG: origin and measurement," in *EEG-fMRI*, ed: Springer, 2009, pp. 19-38.

[10] B. A. Khalil and K. Misulis, "Atlas of EEG & Seizure Semiology," ed: Elsevier, Butterworth Heinehmann Edition, Philadelphia, 2006.

[11] P. Zarjam, J. Epps, F. Chen, and N. H. Lovell, "Estimating cognitive workload using wavelet entropy-based features during an arithmetic task," *Computers in biology and medicine,* vol. 43, pp. 2186-2195, 2013.

[12] C. L. Baldwin and B. Penaranda, "Adaptive training using an artificial neural network and EEG metrics for within-and cross-task workload classification," *NeuroImage,* vol. 59, pp. 48-56, 2012.

[13] K. Yu, I. Prasad, H. Mir, N. Thakor, and H. Al-Nashash, "Cognitive workload modulation through degraded visual stimuli: a single-trial EEG study," *JOURNAL OF NEURAL ENGINEERING,* vol. 12, 2015.

[14] S. Wang, J. Gwizdka, and W. A. Chaovalitwongse, "Using Wireless EEG Signals to Assess Memory Workload in the n-Back Task," *IEEE Transactions on Human-Machine Systems,* vol. PP, pp. 1-12, 2015.

[15] T. NASA, "Website http://humansystems. arc. nasa. gov/groups," ed: TLX.

[16] 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, pp. 41-56, 2008.

[17] H. Ramoser, J. Muller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement," *Rehabilitation Engineering, IEEE Transactions on,* vol. 8, pp. 441-446, 2000.

[18] A. Delorme and S. Makeig, "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *Journal of neuroscience methods,* vol. 134, pp. 9-21, 2004.

[19] T. Alotaiby, F. E. A. El-Samie, S. A. Alshebeili, and I. Ahmad, "A review of channel selection algorithms for EEG signal processing," *EURASIP Journal on Advances in Signal Processing,* vol. 2015, pp. 1-21, 2015.

[20] S. Altahat, M. Wagner, and E. Martinez Marroquin, "Robust electroencephalogram channel set for person authentication," in *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on*, 2015, pp. 997-1001.

[21] T. Shanableh and K. Assaleh, "Feature modeling using polynomial classifiers and stepwise regression," *Neurocomputing,* vol. 73, pp. 1752-1759, 2010.

[22] D. C. Montgomery and G. C. Runger, *Applied statistics and probability for engineers*: John Wiley & Sons, 2010.

Mahmoud, R., Shanableh, T., Bodala, I.P., Thakar, N., & Al-Nashash, H. (2017). Novel classification system for classifying cognitive workload levels under vague visual stimulation. IEEE Sensors Journal, July, DOI 10.1109/JSEN.2017.2727539 July, 2017.

[23] K.-A Toh, Q.-L. Tran and D. Srinivasan, "Benchmarking a Reduced Multivariate Polynomial Pattern Classifier," IEEE Trans. on pattern analysis and machine intelligence, 26(6), JUNE 2004
[24] M.H. Kutner, C.J. Nachtsheim, and John Neter, "Applied Linear Regression Models," McGraw Hill/Irwin Series: Operations and Decision Sciences, 4th Edition, 2004.
[25] J. M. Fuster, *Prefrontal cortex*: Springer, 1988.

Mahmoud, R., Shanableh, T., Bodala, I.P., Thakar, N., & Al-Nashash, H. (2017). Novel classification system for classifying cognitive workload levels under vague visual stimulation. IEEE Sensors Journal, July, DOI 10.1109/JSEN.2017.2727539 July, 2017.

Mahmoud, R., Shanableh, T., Bodala, I.P., Thakar, N., & Al-Nashash, H. (2017). Novel classification system for classifying cognitive workload levels under vague visual stimulation. IEEE Sensors Journal, July, DOI 10.1109/JSEN.2017.2727539 July, 2017.