Mental Workload Estimation using Wireless EEG Signals

Quadri Adewale[1] and George Panoutsos[2][a]
[1]Montreal Neurological Institute, McGill University, Montreal, Canada
[2]Automatic Control & Systems Engineering, University of Sheffield, Sheffield, U.K.

Keywords: Electroencephalogram (EEG), Mental Workload, Cross-task, Cross-subject, Cross-session, Wireless EEG Headset, Domain Adaptation, N-Back Task, Mental Arithmetic Task.

Abstract: Previous studies have demonstrated the applicability of electroencephalogram (EEG) in estimating mental workload. However, developing reliable models for cross-task, cross-subject and cross-session classifications of workload remains a challenge. In this study, we used a wireless Emotiv EPOC headset to evaluate workload in eight subjects and two mental tasks, namely n-back, and arithmetic tasks. 0-back and 2-back tasks, and 1-digit and 3-digit additions were employed as low and high workloads in the n-back and arithmetic tasks, respectively. Using power spectral density as features, a signal processing and feature extraction framework was developed to classify workload levels. Within-session accuracies of 98.5% and 95.5% were achieved in the n-back and arithmetic tasks, respectively. To facilitate real-time estimation of workload, a fast domain adaptation technique was applied to achieve a cross-task accuracy of 68.6%. Similarly, we obtained accuracies of 80.5% and 76.6% across sessions, and 74.4% and 64.1% across subjects, in n-back and arithmetic tasks, respectively. Although the number of participants is limited, this framework generalised well across subjects and tasks, and provides a promising approach towards developing subject and task-independent models. It also shows the feasibility of using a consumer-level wireless EEG headset in cognitive monitoring for real-time estimation of workload in practice.

1 INTRODUCTION

Brain-computer interface (BCI) is mainly applied to aid disabled persons by using the brain signals for communication and control while bypassing auxiliary muscles or nerves (Wolpaw et al., 2002). However, BCI is now used in healthy subjects in an application called Passive BCI (Zander et al., 2010). Passive BCI can be used to obtain information about a user's level of workload, mental state, or attentiveness. This application can help to improve a vehicle driver's performance, prevent accidents in systems and industries and ensure attentiveness of security officers in surveillance systems (Mueller et al., 2008; Venthur et al., 2010; Welke et al., 2009).

Although there is no universal definition of mental workload (Cain, 2007; Zander et al., 2010), workload can be viewed as the result of the interaction between work demands and human capacity (Hart and Staveland, 1988). As the workload increases, the task demand approaches the upper limit of human ability. Physiological correlates of workload have been established in many literatures. Some of these measures include heart rate (Brookings et al., 1996), blood pressure (Theorell et al., 1988), Electromyogram (EMG) (Mehta & Agnew, 2012) and EEG. Even though there seem not to be a best physiological indicator of workload, some studies showed EEG to be more promising compared to other indicators (Hogervorst et al., 2014; Taylor et al., 2010).

In mental workload classification, a variety of machine learning methods have been employed. Some of these methods include support vector machine (SVM) (Wang et al., 2016), artificial neural network (ANN) (Baldwin and Penaranda, 2012) and hierarchical Bayes model (Wang et al., 2012). SVM finds more application because it generalizes well and handles high-dimensional data (Burges, 1998). However, due to the nonstationary nature of EEG signals, the performance of algorithms degrades when the training and test data are taken from different sessions and subjects. Hence such algorithms need to

[a] https://orcid.org/0000-0002-7395-8418

Adewale, Q. and Panoutsos, G.
Mental Workload Estimation using Wireless EEG Signals.
DOI: 10.5220/0010251302000207
In *Proceedings of the 14th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2021) - Volume 4: BIOSIGNALS*, pages 200-207
ISBN: 978-989-758-490-9

be trained or adapted for every user and session (personalised and bespoke models). While such data modelling attempts are useful, not being able to use previously elicited models in applications for other users is a weakness in terms of developing multi-user software tools and algorithms.

Some attempts have been made to overcome this performance degradation. EEG source localization and functional connectivity estimation ware applied to classify workload across tasks (Dimitrakopoulos et al., 2017). Similarly, a combination of deep recurrent network and 3D convolutional neural network was used to learn detailed features for cross-task classification (Zhang et al., 2019). Other studies proposed domain adaptation and transfer learning to overcome the shifts in data distribution across different subjects (Albuquerque et al., 2019; J. Zhang et al., 2017). However, these studies considered either cross-task or cross-subject classifications separately. Moreover, some of them used many electrodes for recording the EEG signals which reduces the comfortability of using EEG headsets in practical and online scenarios.

We address these issues by (i) applying a single framework to overcome cross-session, cross-subject and cross-task performance degradations (ii) using a consumer-level wireless EEG headset with just 14 channels. We developed a simple signal processing and feature extraction technique to facilitate practical and real-time application. The model was tested across eight (8) subjects in two different types of task – n-back task and arithmetic task. We then applied a fast domain adaptation paradigm called Adaptive Subspace Feature Machine (ASFM) (Chai et al., 2017) to improve the model performance across sessions and tasks. We compared the results from ASFM with those of SVM. Some subjective and performance indices of mental workload were also used to verify that the experimental design reflects different levels of workload.

2 METHOD

2.1 Subjects

Eight (8) subjects (6 males and 2 females) participated in the EEG experiment which was held at the Physiological Signal Processing Laboratory, Department of Automatic Control and Systems Engineering, University of Sheffield. Participants were aged between 19 and 30 (Mean = 25 ± 3 years). All subjects were right-handed, reported normal or corrected-to-normal vision, and had no history of any fatigue-related disorder. The experiment was performed in accordance with the University's ethics guidelines, and participants gave written informed consent.

2.2 Experimental Design

N-back task and arithmetic task were employed in this study and each task had two difficulty levels. The two tasks have been extensively used to induce workload demands (Dimitrakopoulos et al., 2017; S. Wang et al., 2016; Zarjam et al., 2012). In the n-back task, 0-back and 2-back tasks were used to represent low and high workloads, respectively. As shown in Figure 1a, for the 0-back condition, the target letter is 'X'. For the 2-back condition, participant decides if the letter displayed currently is same as the letter displayed two sequences earlier. Hence, the participant updates his memory by memorizing two previous letters as the sequence progresses. In both task levels, the participant presses the appropriate key to indicate if the letter is a target or not.

In the arithmetic tasks, participants are required to perform arithmetic operations without any aid such as pen and paper or calculator. The answer from every arithmetic operation is memorized and retrieved after some seconds when an answer is displayed. If the number displayed on the screen is the correct result from the last arithmetic operation, then such number is the target number (T), else it is a non-target (NT). 1-digit addition was used for low workload level and 3-digit addition for high workload level. The 3-digit addition is shown in Figure 1b.

To perform the experiment, the participants' attention is focused on a cross on the screen for 30 seconds without any movement or much eye blinking. Then, participants perform 5-minute blocks each of the 0-back task, 2-back, 1-digit arithmetic and 3-digit arithmetic tasks. To remove time-dependent confounding effects, 3 participants were asked to repeat the experiment after a week, while the tasks were presented in a counterbalanced order. The participant took a break after every task block and rated each task on an RSME scale (Zijlstra, 1993) based on the perceived expended effort in solving that task. The scale ranges from 0 to 150 in increasing order of perceived effort expended.

While performing the tasks, the EEG signals of participants were recorded using a wireless Emotiv EPOC neuroheadset (EMOTIV, 2013). The Emotiv EPOC headset uses 14 electrodes with two additional electrodes for referencing (DRL) and noise cancellation (CMS). All the available electrodes were used in this study.

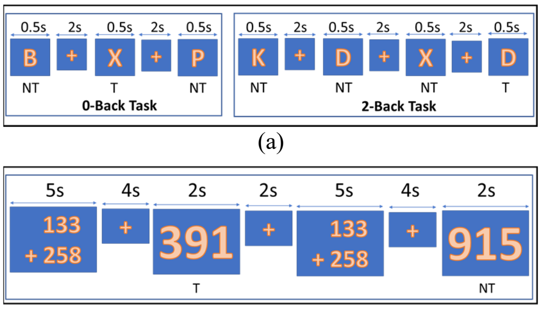


Figure 1: Experimental tasks used to evaluate workload levels. (a) N-back tasks. (b) 3-digit arithmetic task.

2.3 Data Analysis

2.3.1 Data Acquisition and Pre-processing

The EEG data were sampled at 128Hz. All the 14 channels were used, with the reference electrode attached to the left mastoid. Each raw EEG measurement was imported into MATLAB and data corresponding to the fourteen channels were extracted. A bandpass filter of 1.5-40Hz was applied to remove high frequency noise and low frequency DC components. The bandpass filtering was done in in two directions to avoid phase shift or distortion of the EEG data. With the aid of the markers set during the EEG recording, the epochs corresponding to each task were extracted.

2.3.2 Feature Extraction

The filtered data were divided into 4-second blocks with 2-second overlaps between adjacent blocks. The data in each block was normalised for zero mean as shown in (1) below.

$$\boldsymbol{x}_{normal} = \boldsymbol{x} - \boldsymbol{x}_{mean} \tag{1}$$

where **x** is the whole data in a 4-second block, $\boldsymbol{x}_{mean}$ is the mean of the data in such block and $\boldsymbol{x}_{normal}$ is the normalised data in the block. The power spectral density (PSD) in each normalised block was computed using Welch's method with 1-second Hamming window and 50% overlap. Windowing was necessary to reduce signal leakage, and the overlaps allow for smooth transition between windows. In each block, the power spectral densities of eight frequency bands were computed thus: 4-8Hz, 8-12Hz, 12-16Hz, 16-20Hz, 20-24Hz, 24-28Hz, 28-32Hz, 32-36Hz. For each frequency band, the root-mean-square (RMS) value was calculated as follows:

$$RMS = \frac{\|PSD\|}{\sqrt{l_p}} \tag{2}$$

where ||PSD|| is the Euclidean length of the PSD in a frequency band and l_p is the length of the PSD vector. With 14 channels and 8 frequency bands, 112 (14 × 8) features were generated.

2.3.3 Data Classification

An SVM (with a linear or RBF kernel) was used for classifying the workload levels in the n-back and arithmetic tasks. In addition, the performance of the SVM was investigated for cross-session, cross-task, and cross-subject classifications. ASFM was also applied and the results were compared with those obtained from SVM.

ASFM was proposed in (Chai et al., 2017) as a fast domain adaptation technique for EEG-based emotion recognition to overcome the performance degradation when EEG data are sampled from different subjects or sessions. The nonstationary nature of EEG and variability of brain dynamics with individuals and age cause a mismatch between the marginal and conditional distributions of the source domain (training data) and target domain (testing data). In other words, if there is a source domain X_s with label Y_s and a target domain X_t with label Y_t, ASFM formulates a new feature to reduce the marginal distribution mismatch between P_sX_s and P_tX_t, and conditional distribution mismatch between $P_s(Y_s|X_s)$ and $P_t(Y_t|X_t)$.

First, a Subspace Alignment (Fernando et al., 2013) is used to unify the marginal distribution of the source and target domains through principal component analysis (PCA). The eigenvectors form the new subspaces Z_s and Z_t for the source and target, respectively. A linear transformation is then obtained to map Z_s to Z_t. If there is a transformation B, an objective function can be formulated to align the subspaces as follow:

$$min\ \ (\|Z_sB - Z_t\|_F^2) \tag{3}$$

where $\|.\|_F^2$ is Frobenius norm. The above equation can be rewritten as:

$$\begin{aligned} \mathrm{F}\,(\mathrm{B}) = \ & \|Z_s^T Z_s B - Z_s^T Z_t\|_F^2 \\ & = \|B - Z_s^T Z_t\| \end{aligned} \tag{4}$$

where T denotes transpose. Hence, the objective function is minimised when $B^* = Z_s^T Z_t$. Then, the subspace can be transformed thus:

$$Z_{Tran} = Z_s Z_s^T Z_t \quad (5)$$

To reduce the marginal distribution mismatch between $P_s(X_s)$ and $P_t(X_t)$, $X_s Z_{Tran} = X_s Z_s Z_s^T Z_t$ and $X_t Z_t$ are then computed.

Next, the conditional distributions in $X_s Z_{Tran}$ and $X_t Z_t$ are adapted by obtaining the probability for an input target in the transformed subspace and moving the target to the training set. In other words, the discrepancy between $P_s(Y_s|X_s Z_{Tran})$ and $P_t(Y_t|X_t Z_t)$ is reduced. If $X_s Z_{Tran}$ is denoted as L, then the transformed source domain can be represented as $\{l_1, l_2 \ldots l_n\}$ with labels $\{y_1, y_2 \ldots y_n\}$. Logistic regression is applied to compute the conditional distribution of the source. Probabilistic model of logistic regression is given as:

$$P(y|l;w) = \frac{1}{1 + \exp(-yw^T l)} \quad (6)$$

where $y = \pm 1$ and w is a weight vector that can be learnt with gradient descent algorithm. The conditional distribution of the source domain can then be written according to Equation (2.34) as follows:

$$\begin{aligned} &P_s(Y_s = 1|L;w) \\ &= P_s(Y_s = 1|X_s Z_{Tran};w) \\ &= \frac{1}{1 + \exp(-w^T X_s Z_{Tran})} \end{aligned} \quad (7)$$

On the other hand, the conditional distribution $P_t(Y_t|X_t Z_t)$ of the target domain cannot be exactly calculated since the target data are not labelled. A solution is assumed that $P_t(Y_t|X_t Z_t) \approx P_s(Y_s|X_s Z_{Tran};w)$. In an iterative manner, logistic regression could be used to estimate $P_s(Y|X_t Z_t;w)$ to indicate the level of certainty that $X_t Z_t$ belongs to the predicted label y. Consequently, a confidence level c is defined to determine the target samples that would be moved to the training set.

$$c\,(Z_T) = \begin{cases} 1 & if\ \ P_s(Y|X_t Z_t;w) > \tau \\ 0 & otherwise \end{cases} \quad (8)$$

where τ is a threshold between 0 and 1. Samples with confidence level of 1 in the target set are moved to the training set and the conditional distribution $P_s(Y_s|X_s Z_{Tran};w)$ is recomputed. As the process is repeated in more iterations, the marginal distribution discrepancy between the source and target domain is reduced.

3 RESULTS

3.1 Subjective Measure (RSME)

The mental workload perceived by subjects increased with memory load, with average rating of 44 for 0-back task and 86 for 2-back task on the RSME scale. Paired-samples t-test showed that the two task levels were significantly different (t(7) = -9.361, p<0.05). The average ratings for the 1-digit and 3-digit arithmetic tasks were 42 and 73 respectively, and the two task levels differed significantly (t(7) = -4.47, p<0.05). The averages of both the n-back tasks and arithmetic tasks confirmed that our experimental design provides two discriminative levels of workload (low and high).

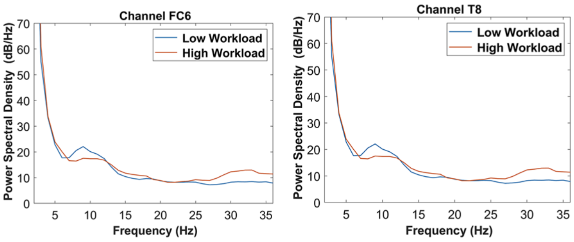
3.2 Performance Measures

Average response time increased with workload from 0-back 547.9ms (0-back) to 853.9ms (2-back). Due to non-normality, we used Wilcoxon signed-rank test and found a significant difference in response times of the two workload levels (p = 0.012; p < 0.05). Response time on the arithmetic task also increased with workload level from 972.5ms (1-digit) to 1251.8ms (3-digit). The difference was statistically significant (t(7) = -4.773, p = 0.002; p < 0.05), implying a significant interaction between the speed of performance of a task and workload.

The average accuracy of response to stimuli significantly degraded (t(7) = 3.399, p = 0.011; p < 0.05) as the workload increased from 0-back (98.2%) to 2-back (91.4%). Increase in workload from 1-digit to 3-digit arithmetic also resulted in significant decrease in average accuracy from 92.5% to 78.5% (p = 0.048; p < 0.05). The results from both tasks confirmed the expected difference between the difficulty levels of low and high mental workloads.

3.3 Variation of EEG Spectral Power

Grand averages of spectral powers across all the eight subjects for some brain regions are shown in Figure 2. In consonance with previous studies, alpha power (8-12Hz) decreased with workload across all the electrodes, gamma power (>25Hz) increased with workload, and theta power (4-7Hz) increased with workload. Similar to the findings in (S. Wang et al., 2016), increase in power with workload was observed in the high beta band (20-25Hz), especially at the frontal sites (AF4 and FC6). Furthermore, the effect of workload on spectral power is prominent in the gamma band across all electrodes. The results support the use of EEG spectral power as a feature for estimating mental workload.



(a) (b)

Figure 2: Grand averages of spectral power vary with workload across frequency bands. (a) Power spectra in frontal region. (b) Power spectra in temporal region.

3.4 Classification Accuracies

3.4.1 Within-session Classification

The EEG obtained from a subject in an experiment session was used to train and test the accuracy of the model for such subject in a 10-fold cross-validation. Figure 3a shows the performance of the SVM (with linear kernel) in classifying the workload levels for the two types of task. The algorithm classified the two levels of workload in n-back task with an average accuracy of 98.5% (SD = 2.1%) as against a mean accuracy of 95.5% (SD = 4.1%) for the two workload levels in arithmetic task. The average accuracies are close to the 98.6% (0-back vs 2-back) and 94.2% (1-digit vs 2-digit multiplication) obtained in (Hwang et al., 2014) using same Emotiv EPOC headset. About 100% accuracy was also reported by (Wang et al., 2016) using Emotiv headset for 0-back vs 2-back tasks.

The highest and lowest accuracies achieved for the n-back task were 100% and 93.5%, respectively. The arithmetic task produced 100% and 88.4% as the highest and lowest accuracies, respectively. The classification accuracies in the n-back and arithmetic tasks were significantly different $p=0.028$ ($p<0.05$). The difference in accuracy for the two tasks could be because be the two levels of workload in the arithmetic tasks have more similarities than those of the n-back tasks. Hence, there are likely more common features in the 1-digit and 3-digit subtasks which makes it less easy for the algorithm to discriminate between the two arithmetic workload levels. It could also be that there are more cross-subject variabilities in the arithmetic task than the n-back task, therefore, the model could generalise better for the latter task. As shown in the results, accuracies of the two tasks varied between subjects. Very high accuracies were achieved for subjects **P05** and **P08**. These discrepancies point to the variation of brain dynamics with individuals and age; hence, a model may not generalise well across subjects and therefore require tuning for every user. However, the model developed in this work generalised across many subjects without individual-based tuning.

3.4.2 Cross-session Classification

Due to the nonstationary nature of EEG signals, the performance of a model degrades if the training and test data are from different sessions or times. As a result, training is often repeated for every session. To test the performance of the model across different training sessions, three participants were asked to repeat the tasks after seven days. Then, the data from the first day were used for training while the data from the eighth day were used for testing. Here, SVM and ASFM were applied for classification and compared against each other as shown in Figure 3b.

The performance of SVM degraded when the trainings from the previous experiment session were used to classify data obtained many days later without retraining. The accuracy of SVM, without any feature adaptation, reduced to as low as 43.9% (below 50%) in one of the cases. Conversely, the use of ASFM, a domain adaption technique, achieved high average cross-session accuracies of 76.6% (SD = 2.5%) and 80.5% (SD=16%) in the arithmetic and n-back tasks, respectively.

ASFM reduced the marginal and conditional distribution mismatch of EEG data across two different experimental sessions. This result suggests that the model with ASFM could be used for a subject at every session without retraining. ASFM was first applied for emotion recognition using differential entropy as features (Chai et al., 2017). In that work, it achieved a cross-session accuracy of 75.1% (SD = 7.7%). Our work has however shown that it can be successfully applied to mental workload using power spectral density as features.

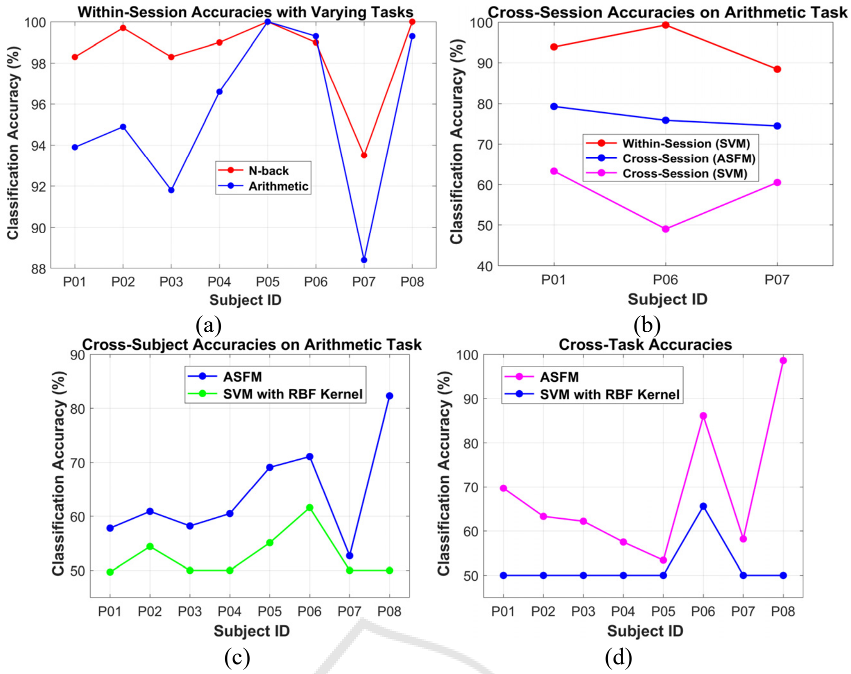


Figure 3: Classification accuracies of the model (a) Within-session classification accuracy using SVM with linear kernel. (b) Cross-session performance on arithmetic task. (c) Cross-subject performance on arithmetic task. (d) Cross-task classification accuracies.

Table 1: Average classification accuracies.

	Within-Session Acc. (%)		**Cross-Session Acc. (%)**		**Cross-Subject Acc. (%)**		**Cross-Task Acc. (%)**
	N-Back	Arithmetic	N-Back	Arithmetic	N-Back	Arithmetic	
SVM	98.5 (SD = 2.1)	95.5 (SD = 4.1)	63 (SD = 18.5)	57.6 (SD = 6.2)	60.4 (SD = 20.5)	52.6 (SD = 4.2)	52 (SD = 5.5)
ASFM	---	---	80.5 (SD = 16)	76.6 (SD = 2.5)	74.4 (SD = 13)	64.1 (SD = 9.5)	68.6 (SD =15.8)

3.4.3 Cross-subject Classification

To evaluate the effect of variability of brain dynamics across subjects, the model was evaluated for cross-subject performance. Leave-one-subject-out classification method was applied by using data from one subject for testing and the data from the remaining seven subjects for training. The procedure was repeated eight times so that data from every subject was used for testing. To limit the size of the training data, only about 60-second data window (60 samples) was selected from each subject for inclusion in the training set. Hence, the training set contained 420 samples. In the test set, the whole 5-minute length of data from a subject was used. Furthermore, the kernel of the SVM was changed to RBF kernel because the linear kernel could not find a linear hyperplane for one of the cases. SVM with RBF kernel was compared against ASFM as shown in Figure 3c. SVM achieved a mean classification accuracy of 60.4% (SD = 20.5%) and 52.6% (SD = 4.2%) on n-back and arithmetic tasks, respectively. ASFM improved the cross-subject accuracies to 74.4% (SD =13%) and 64.1% (SD = 9.5%) in the n-back and arithmetic tasks, respectively.

Even though using a non-linear kernel can improve performance of SVM or even find a solution where using linear kernel is infeasible, SVM without feature adaptation is limited in capturing the cross-human variability that exists in brain dynamics. Such limitation is observed in subject **P01** where the model performance deteriorated below the average level. The results show that feature adaptation with ASFM can mitigate the effect of subject variability on model performance.

3.4.4 Cross-task Classification

Cross-task performance of the model was examined by training on n-back tasks and classifying on arithmetic tasks. The result of the cross-task classification is shown in Figure 3d. SVM with RBF kernel provided an average accuracy of 52% (SD = 5.5%) while ASFM yielded a higher average accuracy of 68.6% (SD = 15.8%). The deterioration in performance could be attributed to the difference in absolute workload levels in the two tasks. For example, low workload level in the n-back task (0-back) may not be equivalent to low workload level in the arithmetic task (1-digit). This effect is also observable in the differences of average subjective ratings on the RSME scale presented earlier. Besides, the underlying brain dynamics resulting from performing the n-back tasks could be different from those of the arithmetic tasks. Nevertheless, the use of ASFM as a feature adaptation technique reduced the mismatch between the different workload types. The classification results are summarised in Table 1.

4 CONCLUSION

This work proposed a robust modelling technique for online estimation of mental workload using a 14-channel wireless EEG headset. The subjective and performance measures indicated that the experimental design provided discriminative workload levels. Using SVM with linear kernel, the model could classify workload levels in more than one type of task without requiring subject or task adaptation. Furthermore, a domain adaptation technique, ASFM, was used to overcome the variabilities that exist across subjects, experimental sessions, and tasks. ASFM showed better performance than SVM (with RBF kernel) in the presence of these variabilities. ASFM – to the best of our knowledge – has not been used in estimating workload before. However, it was successfully applied in this work and yielded good performance in cross-subject, cross-session and cross-task classifications of workload. This research provided a promising framework for estimating mental workload across subjects, sessions, and tasks. It also shows the feasibility of developing models that would not require retraining or recalibration when there are changes in users, sessions, or types of task.

In this preliminary study, only 8 subjects were included in the trials for performance evaluation, and 3 subjects for cross-session classification. Based on the very promising results obtained, it is recommended that a larger study is conducted with more participants to establish the generalisation and robustness of the proposed method.

In addition, this work has used two separate types of task to estimate workload, more tasks can be designed to further investigate the generalisability of the model across different tasks. In addition, multi-class workload levels can be used instead of the two-class workload levels to capture more levels of workload such as ‘very low’, ‘very high’, etc. Validation on many tasks and workload levels can facilitate the development of a task-independent model for within-task and cross-task classification in practical settings. The model can also be tested in real-time when the subjects are performing cognitive tasks. Ultimately, this research work highlights the potential for the creation of a robust online cognitive monitoring system for assessing mental workload in practical situations.

REFERENCES

Albuquerque, I., Monteiro, J., Rosanne, O., Tiwari, A., Gagnon, J. F., & Falk, T. H. (2019). Cross-subject statistical shift estimation for generalized electroencephalography-based mental workload assessment. *Conference Proceedings - IEEE International Conf. on Systems, Man and Cybernetics*. https://doi.org/10.1109/SMC.2019.8914469

Baldwin, C. L., & Penaranda, B. N. (2012). Adaptive training using an artificial neural network and EEG metrics for within- and cross-task workload classification. *NeuroImage*, *59*(1), 48–56. https://doi.org/10.1016/j.neuroimage.2011.07.047

Brookings, J. B., Wilson, G. F., & Swain, C. R. (1996). Psychophysiological responses to changes in workload during simulated air traffic control. *Biological Psychology*, *42*(3), 361–377. https://doi.org/10.1016/0301-0511(95)05167-8

Burges, C. J. C. (1998). A Tutorial on Support Vector Machines for Pattern Recognition. *Data Mining and Knowledge Discovery*, *2*(2), 121–167. https://doi.org/10.1023/A:1009715923555

Cain, B. (2007). A Review of the Mental Workload Literature. *Defence Research and Development Toronto (Canada)*, *1998*, 4-1-4–34. http://www.dtic.mil/cgi-bin/GetTRDoc?Location=U2&do c=GetTRDoc.pdf&AD=ADA474193

Chai, X., Wang, Q., Zhao, Y., Li, Y., Liu, D., Liu, X., & Bai, O. (2017). A fast, efficient domain adaptation technique for cross-domain electroencephalography (EEG)-based emotion recognition. *Sensors (Switzerland)*, *17*(5), 1–21. https://doi.org/10.3390/s17051014

Dimitrakopoulos, G., Kakkos, I., Dai, Z., Lim, J., Bezerianos, A., & Sun, Y. (2017). Task-independent mental workload classification based upon common multiband EEG cortical connectivity. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *4320*(c), under revision. https://doi.org/10.1109/TNSRE.2017.2701002

EMOTIV. (2013). *EMOTIV Epoc - 14 Channel Wireless EEG Headset*. EMOTIV Inc.

Fernando, B., Habrard, A., Sebban, M., & Tuytelaars, T. (2013). Unsupervised visual domain adaptation using subspace alignment. *Proceedings of the IEEE International Conference on Computer Vision*, 2960–2967. https://doi.org/10.1109/ICCV.2013.368

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *Advances in Psychology*, *52*(C), 139–183. https://doi.org/10.101 6/S0166-4115(08)62386-9

Hogervorst, M. A., Brouwer, A. M., & van Erp, J. B. F. (2014). Combining and comparing EEG, peripheral physiology and eye-related measures for the assessment of mental workload. *Frontiers in Neuroscience*, *8*(OCT). https://doi.org/10.3389/fnin s.2014.00322

Hwang, T., Kim, M., Hwangbo, M., & Oh, E. (2014). Comparative analysis of cognitive tasks for modeling mental workload with electro encephalogram. *Conf. Proceedings : Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference*, *2014*, 2661–2665. https://doi.org/10.1109/EMB C.2014.6944170

Mehta, R. K., & Agnew, M. J. (2012). Influence of mental workload on muscle endurance, fatigue, and recovery during intermittent static work. *European Journal of Applied Physiology*, *112*(8), 2891–2902. https://doi.org/10.1007/s00421-011-2264-x

Mueller, K. R., Tangermann, M., Dornhege, G., Krauledat, M., Curio, G., & Blankertz, B. (2008). Machine learning for real-time single-trial EEG-analysis: From brain-computer interfacing to mental state monitoring. *Journal of Neuroscience Methods*, *167*(1), 82–90. https://doi.org/10.1016/j.jneu meth.2007.09.022

Taylor, G., Reinerman-Jones, L., Cosenzo, K., & Nicholson, D. (2010). Comparison of Multiple Physiological Sensors to Classify Operator State in Adaptive Automation Systems. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *54*(3), 195–199. https://doi.org/10.1177 /154193121005400302

Theorell, T., Perski, A., Akerstedt, T., Sigala, F., Ahlberg-Hultén, G., Svensson, J., & Eneroth, P. (1988). Changes in job strain in relation to changes in physiological state. A longitudinal study. *Scandinavian Journal of Work, Environment & Health*, *14*(3), 189–196. https://doi.org/10.5 271/sjweh.1932

Venthur, B., Blankertz, B., Gugler, M. F., & Curio, G. (2010). Novel applications of BCI technology: Psychophysiological optimization of working conditions in industry. *Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics*, 417–421. https://doi.org/10.1109/ ICSMC.2010.5641772

Wang, S., Gwizdka, J., & Chaovalitwongse, W. A. (2016). Using Wireless EEG Signals to Assess Memory Workload in the n-Back Task. *IEEE Transactions on Human-Machine Systems*, *46*(3), 424–435. https://doi.org/10.1109/THMS.2015.2476818

Wang, Z., Hope, R. M., Wang, Z., Ji, Q., & Gray, W. D. (2012). Cross-subject workload classification with a hierarchical Bayes model. *NeuroImage*, *59*(1),64–69 https://doi.org/10.1016/j.neuroimage.2011.07.094

Welke, S., Juergensohn, T., & Roetting, M. (2009). Single-trial detection of cognitive processes for increasing traffic safety. *Proceedings of the 21st International Technical Conference on the Enhanced Safety of Vehicles Conference (ESV).*, 1–10.

Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G., & Vaughan, T. M. (2002). Brain-computer interfaces for communication and control. *Clinical Neurophysiology : Official Journal of the International Federation of Clinical Neurophysiology*, *113*(6), 767–791. https://doi.org/ 10.1016/S1388-2457(02)00057-3

Zander, T., Kothe, C., Jatzev, S., & Gaertner, M. (2010). Enhancing Human-Computer Interaction with Input from Active and Passive Brain-Computer Interfaces. *Brain-Computer Interfaces*, 149–178. https://doi.org /10.1007/978-1-84996-272-8

Zarjam, P., Epps, J., Lovell, N. H., & Chen, F. (2012). Characterization of memory load in an arithmetic task using non-linear analysis of EEG signals. *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, 3519–3522. https://doi.org/10.1109/ EMBC.2012.6346725

Zhang, J., Wang, Y., & Li, S. (2017). Cross-subject mental workload classification using kernel spectral regression and transfer learning techniques. *Cognition, Technology and Work*. https://doi.org/ 10.1007/s10111-017-0425-3

Zhang, P., Wang, X., Zhang, W., & Chen, J. (2019). Learning Spatial-Spectral-Temporal EEG Features With Recurrent 3D Convolutional Neural Networks for Cross-Task Mental Workload Assessment. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*. https://doi.org/10.1109/TNSRE.2018 .2884641

Zijlstra, F. R. (1993). Efficiency in work behaviour: A design approach for modern tools. *Delft University Press*, *January 1993*, 1–186. https://doi.org/90-6275-918-1