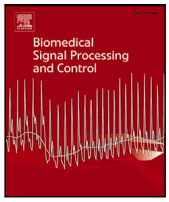
Biomedical Signal Processing and Control 94 (2024) 106284

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc



Mental workload estimation with electroencephalogram signals by combining multi-space deep models

Hong-Hai Nguyen, Ngumimi Karen Iyortsuun, Seungwon Kim, Hyung-Jeong Yang, Soo-Hyung Kim *

Artificial Intelligence Convergence, Chonnam National University, Gwangju, 61186, South Korea

ARTICLE INFO

Keywords:
Electroencephalogram
Mental workload
Temporal convolutional networks
Time–frequency domains

ABSTRACT

The human brain is perpetually active, operating during both work and rest. Excessive mental activity, termed overload, can detrimentally impact health. Advances in predicting mental health conditions aim to prevent severe consequences and improve overall well-being. Consequently, mental status estimation has gained considerable research attention due to its potential benefits. Among the various signals used to assess mental state, the electroencephalogram, with its wealth of brain-related information, is widely employed by researchers. In this paper, we categorize mental workload into three states (low, middle, and high) and estimate a continuum of mental workload levels. Our method leverages information from multiple spatial dimensions (time and frequency domain) to achieve optimal results. In our approach, focusing on the time domain, we utilize Temporal Convolutional Networks to capture temporal information. Meanwhile, in the frequency domain, we implement two distinct branches aimed at acquiring varied information regarding Power Spectral Density by utilizing different lengths of the Fast Fourier Transform. Each branch consists of two blocks, known as the Multi-Dimensional Residual Block, representing a novel architecture by combining residual blocks. The integration of these two domains yields significant results compared to individual estimates in each domain. Our approach achieved a 74.98% accuracy in the three-class classification, surpassing the provided data results at 69.00%. Specially, our method demonstrates efficacy in estimating continuous levels, evidenced by a corresponding Concordance Correlation Coefficient result of 0.629. The combination of time and frequency domain analysis in our approach highlights the exciting potential to improve healthcare applications in the future.

1. Introduction

Mental workload (MWL) refers to brain activities, which are the number of resources in the human brain. The level of the resource can be changed when the human is thinking or performing a task [1]. Because the mental workload resource of a person is limited, the ability to process a lot of information at once will be limited [2,3]. Humans will be bored if they do easy work when their MWL is low and their MWL is high when they do complicated tasks. However, a high MWL is not suitable for their health [4] and can impair their abilities for memory, communication, activity, etc. For jobs with high MWL, such as doctors, soldiers and pilots, the brain sometimes has serious accidents [5]. Furthermore, understanding how the human brain works in daily activities and tasks is an essential area of neuroergonomics research. Therefore, the MWL estimation helps to observe and help people at work and evaluate their work system, which can be improved in the future [6]. EEG signals are captured by positioning electrodes onto the scalp. These signals represent the biological activity of individual brain cells or specific groups of cells, which are transmitted through the cerebral cortex and scalp, ultimately forming the EEG recording. EEG signals that record brain activities reflect physiological and pathological functions of the hemisphere or of the whole brain related to clinical symptoms, supplementing diagnosis and monitoring treatment, called a clinical electroencephalogram. By providing a reliable EEG can contribute to predicting and preventing work-related risks. Diagnostic applications in MWL estimation can be created to address various real-world issues in healthcare, education, and smart traffic, among others [7,8].

Three metrics were used to measure MWL, namely subject, performance, and physiological measures [6,9–11]. The traditional method to measure the context is through a set of various questions, such as

* Corresponding author.
E-mail addresses: hhnguyen@jnu.ac.kr (H.-H. Nguyen), karen@jnu.ac.kr (N.K. Iyortsuun), Seungwon.Kim@jnu.ac.kr (S. Kim), hjyang@jnu.ac.kr (H.-J. Yang), shkim@jnu.ac.kr (S.-H. Kim).

https://doi.org/10.1016/j.bspc.2024.106284
Received 16 November 2023; Received in revised form 6 March 2024; Accepted 23 March 2024
Available online 1 April 2024

the Aeronautics and Space Administration Task Load Index (NASA-TLX) [12] or Assessment Technique (SWAT) [13]. Performance measurement involves measuring the person's performance during the task with an increasing workload. Both subject and performance measures are taken after the task is completed, making them prone to bias. On the contrary, physiological measurements can continuously record information about the workload and do not affect the performance of the main task. Therefore, physiological signals can be used to evaluate the effectiveness of mental workload.

Physiological measurement uses a variety of biosignals such as electroencephalogram (EEG), Galvanic Skin Response (GSR), Heart Rate Variability (HRV), Skin Temperature (Temp), R-R intervals (RR) and electrocardiogram (ECG). Among such signals, EEG is widely used for the estimation of MWL [14] as the information capacity in EEG is higher than other physiological signals [15]. However, most researchers do not release their datasets, which makes it challenging to compare methods. Fortunately, Lim et al. [16] provided an open dataset with a large number of participants and MWL levels. In this study, we conducted experiments on this dataset to estimate the MWL. Our work comprises classification (low, middle and high) and continuous levels estimation for MWL. Regarding continuous perceptual estimation, which represents a novel task that has not been explored previously. Thus, we proceeded with additional experiments utilizing the same dataset, scaling the label values to fit within the range of [0, 1]. Specifically, we proposed a Multi-Dimensional Residual Block to learn frequency domain information, and we combined time and frequency domain information based on multimodal fusion, which significantly improved the results.

In conclusion, our contributions are summarized as follows:

- Proposed Multi-Dimensional Residual Block to learn information about the frequency domain.
- Combined time and frequency domains based on multimodal fusion. Specifically, we conducted various experiments to demonstrate the combination method that proved to be efficient.
- Solved to classification and continuous levels estimation for mental workload. In this study, we proposed a method for estimating the continuum of cognitive estimation, a task not addressed by previous methods.

The next parts of our paper are presented in the following sections: related work is introduced in Section 2, the proposed method is described in Section 3, the experiments in Section 4 and the discussion in Section 5 and finally, our work concludes in Section 6.

2. Related work

Traditional Machine learning and Deep learning are widely used methods to classify MWL. Handcrafted features usually combine with traditional methods, while deep learning automatically extracts features for classification. Wang et al. [6] employed Support Vector Machine (SVM) to classify different levels of MWL, including low workload (0-back) and high workload (1, 2, 3-back), as well as distinguishing between different levels of workload such as 1-back and 2-back (80%) and 1-back and 3-back (84%). The authors extracted a range of features based on statistical, signal power, morphological, and frequency features for binary classification. Ladekar et al. [17] utilized K-nearest neighbors (KNN) to classify four levels of cognitive load by decomposing the EEG signal into ten subbands and combining it with ensemble subspace KNN. Their method achieved an accuracy of 83.65% for four classes. However, the authors performed the experiment on a small sample size of less than 2000, limiting the generalizability of their findings. Gómez et al. [2] used SVM, KNN, and random forest (RF), which are combined with features extraction from frequency bands, the energy and the entropy of the Discrete Wavelet Transform (DWT), Hjorth parameters, Fractal dimension, Detrended fluctuation analysis, and Lempel–Ziv complexity. Zammouri et al. [18] used PSD and archived an accuracy of 79% in the theta band. The alpha band got 78% of the average accuracy. There are two levels of cognitive load used to classify. The authors showed that theta and alpha powers decrease with the increasing level of mental task in the central and posterior locations, respectively. The research methods employ the Machine Learning technique for classification. To achieve desirable outcomes, it necessitates the calculation and appropriate selection of features. However, the drawback of these methods is that they do not incorporate continuous information of the time or frequency domain, which rely on handcrafted features.

Automatic feature extraction of the deep learning model is better than handcrafted features. In their work, Yang et al. [19] achieved a classification accuracy of 92% for MWL by using SDAE (Stacked Denoising Autoencoder) to preserve local information in the EEG. They also applied an ensemble classifier to classify binary MWL (low and high) by extracting PSD features (theta, alpha, beta, and gamma). Qiao et al. [20] proposed a Ternary Task Convolutional Bidirectional Neural Turing Machine (TT-CBNTM) for the classification of cognitive states, which are four levels of cognitive load. The TT-CBNTM learns temporal information and preserves spectral and spatial information of EEG. Moreover, the TT-CBNTM can effectively avoid overfitting. Although the method gives quite high results (96.3%), the authors performed it on a small sample (2600 samples). Kwak et al. [21] achieved a binary classification accuracy of 93.9% using a combination of a three-dimensional convolutional neural network (3D CNN) and a long short-term memory model (LSTM). They obtained 3D EEG images from the 1D EEG signal, and the 3D-CNN learned spatial and spectral information from these images, while the LSTM extracted temporal attention from subframes of the EEG image. Deep learning techniques have the capability to extract features automatically. However, previous studies have predominantly used frequency domain information as input for the models. As a result, the outcomes indicate that the potential of time-domain information has not been fully utilized.

In the field of emotion and motor imagery, a 3D representation of the EEG signal has been used. To maintain spatial information on the electrodes, Zhao et al. [22] converted the 2D array of EEG signals into 3D. They applied a multi-branch 3D-CNN to classify motor imagery using the 3D representation, which can be extended to other related areas such as cognitive load and seizure detection. Similarly, Liu et al. [23] utilized 3D representation to classify motor stages while preserving spatial and temporal information, by passing the 3D representation through a three-branch 3D-CNN. Jia et al. [24] used two 3D representations (spectral feature and temporal feature) to complement the significant difference between features. For EEG emotion recognition, a spatial–spectral–temporal based attention 3D dense network was proposed, which performed well on two datasets.

We propose a fusion of the time and frequency domains to leverage the richness of information available in both spaces for the classification task. It is noteworthy that most existing methods are focused on classification, and a regression approach for the continuous estimation of MWL remains relatively less explored. In this study, we adopt a regression framework for continuous MWL estimation. Furthermore, we exploit the benefits of 3D representation to facilitate the learning of complex relationships among the features extracted from EEG signals.

3. Proposed method

In this section, we present a deep learning model for estimating MWL based on both time and frequency domain information. We provide an overview of our method in Fig. 1. Additionally, We describe the 3D (three-dimensional) representation of PSD into 3D block as depicted in Fig. 2. Finally, we propose a new architecture, the Multi-Dimensional Residual Block (MDRB), illustrated in Fig. 4.

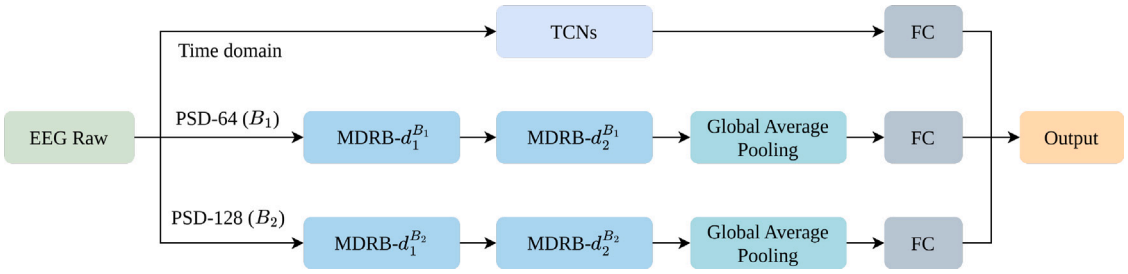


Fig. 1. System overview: Combined time and frequency domains for mental workload estimation.

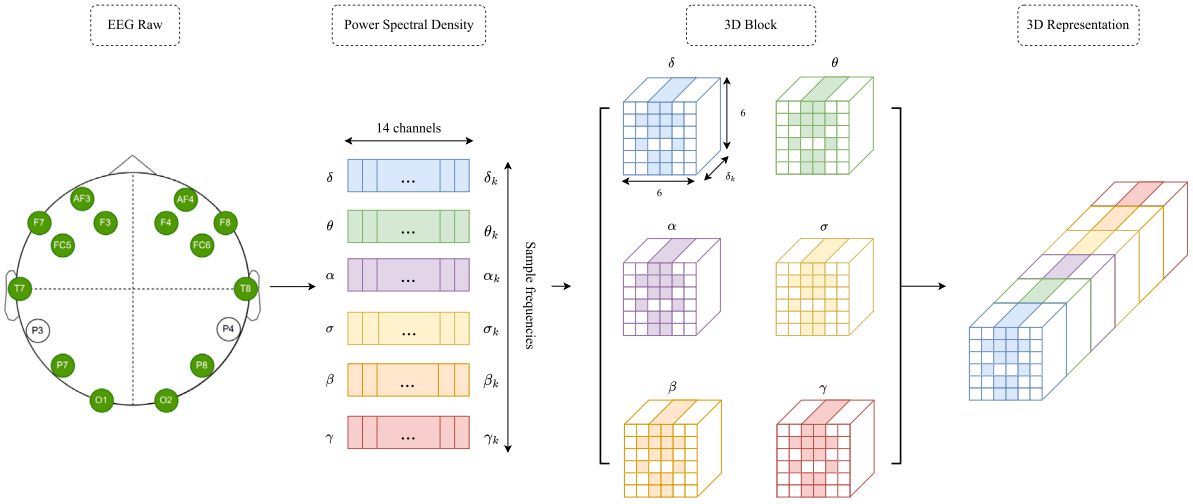


Fig. 2. Overview of converting EEG signals to a 3D representation.

3.1. 3D representation block

We define $S^n = [E_1^n, E_2^n, \dots, E_W^n]^T \in \mathbb{R}^{W \times C}$ denote the n^{th} sample, where $n \in 1, 2, \dots, N$ and N is the number of samples. Here, W represents the window size, C represents the number of channels in the EEG devices, and $E_i^n = [e_i^1, e_i^2, \dots, e_i^C] \in \mathbb{R}^C$ is the EEG signal at timestamp i, where $i \in 1, 2, \dots, W$. For analysis, we convert EEG signals from the time domain (S^n) to the frequency domain using Welch's method to obtain the Power Spectral Density (PSD). Each frequency band carries information about different brain activities and states. Therefore, frequencies link changes in brain activity with various stages of brain waves. Initially, Hans Berger identified two bands: alpha (α) and beta (β). Subsequently, researchers introduced additional bands such as delta (δ), theta (θ), sigma (σ), and gamma (γ). In this study, we extract six frequency bands: delta (0.5–4 Hz), theta (4–8 Hz), alpha (8–12 Hz), sigma (12–16 Hz), beta (16–30 Hz), and gamma (30–45 Hz). The extracted frequency bands depend on the window length of the Fast Fourier Transform (FFT). Therefore, the length of values in each band may differ, but the number of channels remains the same (14 channels). We define PSD for the six bands as follows:

$$P^n = [P^n_{\delta_1}, P^n_{\delta_2}, \dots, P^n_{\delta_k}, P^n_{\theta_1}, P^n_{\theta_2}, \dots, P^n_{\theta_k}, P^n_{\alpha_1}, P^n_{\alpha_2}, \dots, P^n_{\alpha_k}, P^n_{\sigma_1}, P^n_{\sigma_2}, \dots, P^n_{\sigma_k}, P^n_{\beta_1}, P^n_{\beta_2}, \dots, P^n_{\beta_k}, P^n_{\gamma_1}, P^n_{\gamma_2}, \dots, P^n_{\gamma_k}]^T \in \mathbb{R}^{K \times C} \quad (1)$$

where $K = \delta_k + \theta_k + \alpha_k + \sigma_k + \beta_k + \gamma_k$ is the total length of frequency of the bands; $\delta_k, \theta_k, \alpha_k, \sigma_k, \beta_k, \gamma_k$ is the length of the frequency band $\delta, \theta, \alpha, \sigma, \beta, \gamma$ respectively, $P^n_{b_k} = [p^1_{b_k}, p^2_{b_k}, \dots, p^C_{b_k}] \in \mathbb{R}^C$ is the signal frequency band b at the frequency k^{th}.

Zhao et al. [22] proposed a method for converting EEG data from 2D to 3D representation. The maximum number of horizontal and vertical electrodes is selected as the width and height of the matrix, respectively, based on the electrode positions. Fourteen signal positions ($P^n_{b_k}$) are then mapped onto the 2D matrix with a size of 6 × 6. If an electrode position is empty, it is filled with a value of zero. The resulting 2D matrix is shown in Eq. (2) and illustrated in Fig. 3 (b).

$$f(X_j) = \begin{bmatrix} 0 & 0 & AF3_j & AF4_j & 0 & 0 \\ 0 & F7_j & F3_j & F4_j & F8_j & 0 \\ 0 & 0 & FC5_j & FC6_j & 0 & 0 \\ 0 & T7_j & 0 & 0 & T8_j & 0 \\ 0 & 0 & P7_j & P8_j & 0 & 0 \\ 0 & 0 & O1_j & O2_j & 0 & 0 \end{bmatrix} \quad (2)$$

where X_j is the signal frequency of the band at the frequency j^{th}.

The extracted frequency bands step is shown in Fig. 2 as the Power Spectral Density step. From each band, we convert 2D matrix to 3D block with the size $6 \times 6 \times b_k$. There are six blocks from six frequency bands (the 3D Block step as Fig. 2). We stack them as a larger 3D block (the 3D representation step, illustrated in Fig. 2). The height and width dimension keeps relationships between channels, and the depth dimension maintains information about the frequency and continuity of the bands. We define 3D block as $M^n = [M^n_1, M^n_2, \dots, M^n_K] \in \mathbb{R}^{6 \times 6 \times K}$, where $M^n_k = f(P^n_{b_k})$, b_k is the band b at frequency k^{th}.

3.2. MDRB: Multi-Dimensional Residual Block

We propose a Multi-Dimensional Residual Block (MDRB) to capture the relationship between channels and frequency bands, as illustrated in Fig. 4. The input to the MDRB is a 3D block of size $6 \times 6 \times K$, where K is the depth of the 3D input. The MDRB has two branches. In the first branch, the input is passed through a 3D convolution with a filter size of $h \times w \times d$, where h, w, and d are the height, width, and depth of the filter, respectively. The aim is to learn information about the relationships between channels and frequency simultaneously. In the second branch, the input passes through two 3D convolutions. The first convolution extracts information on the frequency bands using a

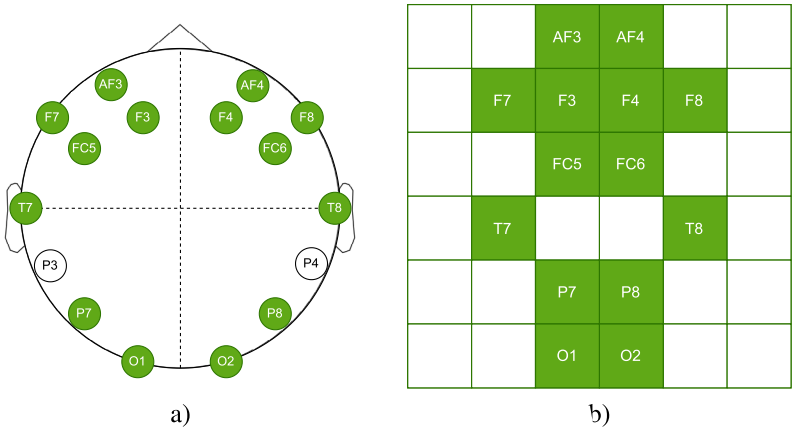


Fig. 3. Mapping of channels positions to a 2D matrix.

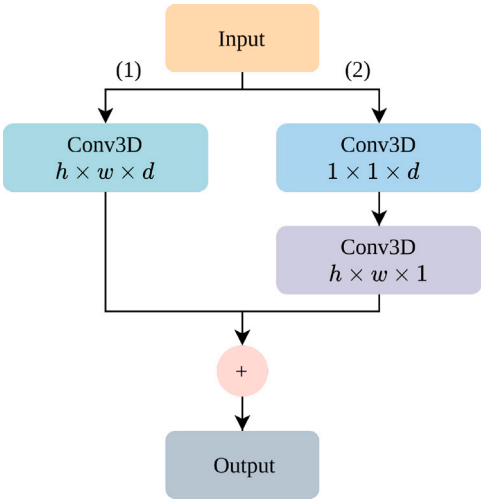


Fig. 4. Multi-Dimensional Residual Block.

filter size of $1 \times 1 \times d$. The second convolution extracts the relationships between channels using a filter size of $h \times w \times 1$. Finally, we compute the average of the outputs of both branches to obtain information from every observation of both branches.

3.3. Combined time and frequency domain

Our proposed model aims to combine the time and frequency domains, as depicted in Fig. 1. To learn temporal information about the time domain, we utilize Temporal Convolutional Networks. For the frequency domain, we have two branches: PSD-64 (B_1) and PSD-128 (B_2), corresponding to two 3D blocks extracted from different lengths of FFT, computed by Welch's method. Each branch has a different depth, $d_{in}^{B_1}$ and $d_{in}^{B_2}$, respectively. We have PSD-64 (B_1) and PSD-128 (B_2) branches with the lengths of FFT 64 and 128, respectively. For each branch, the 3D block input passes 2 MDRBs, which have a different depth length of the MDRB, d_1^i, d_2^i, respectively, where i is the PSD branch, $i \in \{B_1, B_2\}$. The value of MDRB depth (MDRB-$d_j^i, j \in \{1,2\}$) depends on the depth of the input block (d_{in}^i). The MDRB-d_1^i has a depth d_1^i which is calculated by getting the ceiling function of dividing by 2 as Eq. (3), d_2^i is calculated as Eq. (5).

$$d_1^i = \lceil \frac{d_{in}^i}{2} \rceil \tag{3}$$

$$d_{out}^i = \lfloor \frac{d_{in}^i + 2 \times pad - dil \times (d_1^i - 1) - 1}{str} + 1 \rfloor \tag{4}$$

$$d_2^i = d_{in}^i - d_{out}^i \tag{5}$$

where pad is padding ($pad = 0$), dil is dilation ($dil = 1$), and str is convolution stride ($str = 1$).

The Global Average Pooling was applied to the MDRB output for down sampling, and then the fully connected layer was attached to obtain the prediction. The mean of the predictions (3 branches) returns the prediction for the classification and estimation levels of MWL.

4. Experiments

4.1. Dataset

The STEW (Simultaneous Task EEG Workload) EEG database [16] includes 48 participants. Each participant needs to participate in two tasks, the 'No' task and the SIMKAP (Simultaneous capacity) task. For the 'No' task, participants do not perform any task and require a comfortable position with their eyes open. For the second task, participants need to solve the SIMKAP task, a test about a commercial psychological created by Schuhfried GmbH to evaluate a person's multitasking ability and stress. Both tasks were recorded for 3 min for the EEG signal. The first and last 15 s have been removed to eliminate the effects between tasks. A rating scale of 1–9 was used to assess MWL levels for each participant. The STEW dataset contains EEG data collected from the Emotiv headset. The device established 128 Hz for sampling frequency, 16-bit A/D resolution, and fourteen channels (AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4) illustrated in Fig. 3(a) by 10–20 international system.

4.2. Mental workload estimation

We conducted experiments for classification and prediction of continuous levels. The classification task splits the rating scale into 3 classes of low (lo) workload (1–3 rating), middle (mi) workload (4–6 rating), and high (hi) workload (7–9 rating). The continuous level prediction task scales the MWL levels into the range [0–1] based on the rating scale from 1 to 9. The window length is 512 and the shift 128 for fourteen channels, which is also the parameter (window and shift size) that the STEW data used for their experiments. The dataset comprises forty-eight participants but ignores three because there are not available ratings for these participants. Therefore, 36 participants (80%) were used for training and 9 (20%) were used for testing. In Table 1, we describe in detail the number of samples in each class for classification and regression. We can see that the ratio of the three classes is 1.73: 1: 1.05 on the training and testing set.

Table 1
Number of samples use the window size 512 Hz and the shift size 128 Hz.

	Score	Training (N_{train})	Testing (N_{test})	Dimension of input ($W \times C$)
Classification	1–3	4851	1323	512 × 14
	4–6	2793	588	
	7–9	2940	735	
Regression	1–9	10 584	2646	512 × 14

4.3. Experimental setup

We used the TensorFlow [25] framework to train and test our method. Our proposed method is trained in 150 epochs and uses early stopping to obtain the best model. The Adam [26] optimizer was employed with a learning rate of 0.001 for classification and 0.0001 for estimation of continuous levels. The TCNs [27] applied with the kernel size is 2, the number of filters used for convolution layers is 128, 2 stacks of residual block were applied, and the dilation list is $[1, 2, 4, 8]$. The classification of three classes is a multi-label classification, so the softmax function was used. The sigmoid function was also utilized to get the output estimation for continuous levels of MWL. We specifically describe the parameters of each branch of the model in Table 2, Table 3, Table 4 respectively for the time domain branch, the B_1 branch, and the B_2 branch.

In this study, categorical cross-entropy loss is used for classification. The loss function is calculated as follows:

$$CE = -\sum_{i=1}^{N} y_i.log(\hat{y}_i) \tag{6}$$

And we used MSE (mean squared error) as a loss function for continuous levels estimation.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{7}$$

where N is the number of samples, y_i is the label of i^{th} and $\hat{y}_i$ is the value prediction of i^{th}.

Additionally, we use the sensitivity, specificity, precision, negative predictive to evaluate the classification. The evaluations used are as follows:

$$Sensitivity = \frac{TP}{TP + FN} \tag{8}$$

$$Specificity = \frac{TN}{TN + FP} \tag{9}$$

$$Precision = \frac{TP}{TP + FP} \tag{10}$$

$$Negative\ predictive = \frac{TN}{TN + FN} \tag{11}$$

where TP is true positive, TN is true negative, FP is false positive and FN is false negative. We used the Concordance Correlation Coefficient (ρ_c) function to evaluate for the estimation task. The ρ_c function is defined as:

$$\rho_c = \frac{2\rho\sigma_{\hat{Y}}\sigma_Y}{\sigma_{\hat{Y}}^2 + \sigma_Y^2 + (\mu_{\hat{Y}} - \mu_Y)^2} \tag{12}$$

where μ_Y is the mean of the label Y, $\mu_{\hat{Y}}$ is the mean of the prediction $\hat{Y}$, $\sigma_{\hat{Y}}$ and σ_Y are the corresponding standard deviations, ρ is the Pearson correlation coefficient between $\hat{Y}$ and Y.

4.4. Results

Table 5 presents the results of our proposed method. We have split the dataset into an 8:2 ratio, following the methodology of Lim et al. [16], to compare our results with their best single trial. The comparison shows that our proposed method outperforms Lim et al. [16] with an accuracy of 74.98% compared to 69.0%. We conducted separate and combined branch experiments, where the bold font defines the best performance. We also conducted seven experiments for each task that corresponds to the combination of 3 branches (Time, B_1, and B_2). It is noteworthy that combining of the branches gives better results than applying an individual branch. For the classification task, we obtained the best results 74.98% for combining time and multi-frequency (B_1 and B_2) domain, which is better than the previous work [16] by extracting features of the delta, theta, alpha, and beta bands, 69.00% (28 features from PSD) and 69.20% (56 features from PSD) obtained, respectively. Furthermore, by combining the branch B_1 and branch B_2 without the time branch, we obtained 70.79%, is better than Lim et al. [16], 69.00% respectively. It can be seen that the information from the frequency domain is helpful for MWL classification by combination.

We used the Concordance Correlation Coefficient (ρ_c) metric to evaluate continuous estimation in our study. It should be noted that our use of ρ_c as a metric is relatively rare. Therefore, we recognize the need for further clarification and explanation of the results to enhance the reader's comprehension. ρ_c is a statistical measure widely employed in machine learning to evaluate the agreement between two measurement methods or observers for continuous-level estimation tasks, such as stress or valence/arousal estimation. Our proposed method achieved a ρ_c value of 0.629 for the continuous level estimation task. It is important to consider the specific context and application when interpreting the ρ_c values as high or low. A ρ_c value between 0.5 and 0.7 generally suggests moderate agreement, while a value above 0.7 indicates strong agreement. Therefore, the ρ_c value of 0.629 obtained in our study can be interpreted as demonstrating moderate agreement.

Fig. 5 shows the confusion matrix of Lim et al. [16](a) and our method (b). Figure (a) achieved the best for the 'lo' class, but the 'mi' and 'hi' classes are not good, as the only reached 0.46 and 0.31 respectively, which is less than 0.5 threshold. Our method improves for 'mi' and 'hi' classes, 0.52 and 0.79, respectively. In particular, our method obtained a significant result in the 'hi' class, which is 2.5 times greater than [16] ($0.31 < 0.79$). Generally, we have good results for classification, but our method is not really outstanding for the 'mi' class, with the result being much lower than the 'lo' and 'hi' classes. This shows our proposed focus on the 'lo' and 'hi' classes.

In our experimentation, we also varied the shift size when processing the training data. Specifically, we tried shift sizes of 128, 256, 384, and 512, which correspond to window overlaps of 75%, 50%, 25%, and 0%, respectively. The corresponding results are presented in Table 6. Our findings indicate that a shift size of 128 results in better accuracy than the other shift sizes (256, 384, and 512).

We conducted experiments to compare the performance of using only one of the two branches in the MDRB. The results are summarized in Table 7. We observed that using both branches in combination yields the best result, with an accuracy of 74.98% achieved with a shift value of 128. On the other hand, using only branch (1) or branch (2) results in lower accuracy, with 57.56% and 59.18%, respectively. For continuous labels, using an individual branch only achieves half the accuracy of the combined branches ($0.323 < 0.629$), indicating that combining information from both branches is effective. Branch (1) captures an overview of information on both channels and frequency, while branch (2) extracts detailed information in order of frequency and then channels.

5. Discussion

Our method utilizes both time and frequency domain information to extract valuable features that contribute to accurate classification and regression results. The experimental results demonstrate the effectiveness of combining these two domains, achieving good classification accuracy (74.09%) and a concordance correlation coefficient (ρ_c) of 0.629 for regression. Our innovation lies in combining the time and

Table 2
The parameter of time domain branch.

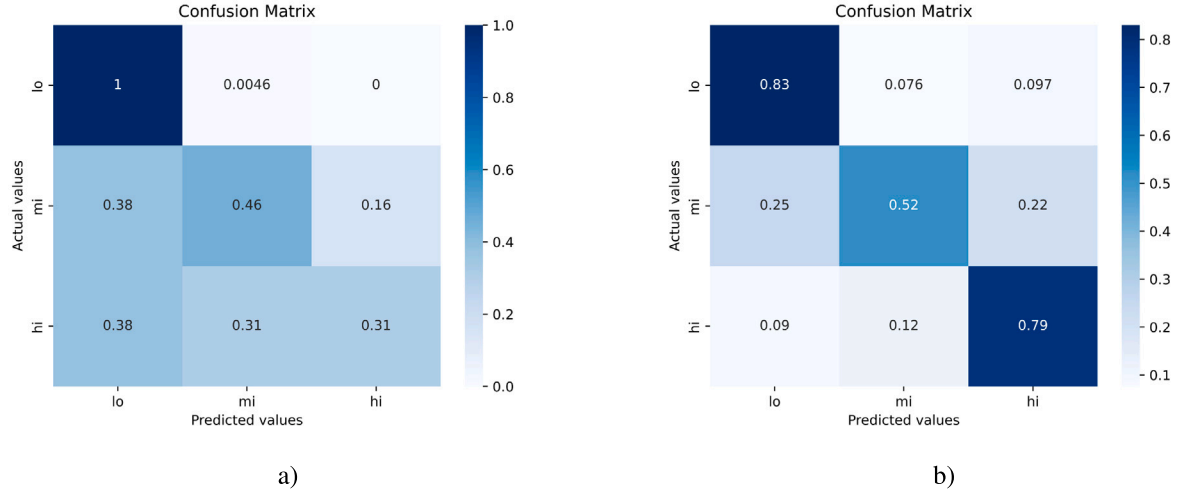
	Input	Output	Kernel size	Dilations	Number of stacks	Window size (W)	Channels (C)
Input layer	512 × 14	–	–	–	–	512	14
TCNs	512 × 14	128	2	[1, 2, 4, 8]	2	512	14
Dense	128	number of class	–	–	–	–	–

Table 3
The parameter of B_1 branch.

	Input	Output	h	w	d	Filters	K
Input layer	6 × 6 × 22 × 1	–	–	–	–	–	–
MDRB-$d_1^{B_1}$	6 × 6 × 22 × 1	4 × 4 × 12 × 44	3	3	11	44	22
MDRB-$d_2^{B_1}$	4 × 4 × 12 × 44	2 × 2 × 1 × 44	3	3	12	44	22
Global Average Pooling 3D	2 × 2 × 1 × 44	–	–	–	–	–	–
Dense	44	number of class	–	–	–	–	–

Table 4
The parameter of B_2 branch.

	Input	Output	h	w	d	Filters	K
Input layer	6 × 6 × 45 × 1	–	–	–	–	–	–
MDRB-$d_1^{B_2}$	6 × 6 × 45 × 1	4 × 4 × 23 × 90	3	3	23	90	45
MDRB-$d_2^{B_2}$	4 × 4 × 23 × 90	2 × 2 × 1 × 90	3	3	23	90	45
Global Average Pooling 3D	2 × 2 × 1 × 90	–	–	–	–	–	–
Dense	90	number of class	–	–	–	–	–



a) b)

Fig. 5. Confusion matrix for classification: (a) Lim et al. [16] with 28 features, (b) Our proposed method.

Table 5
MWL estimation: The results compare with previous work by accuracy and ρ_c where Acc_3 is the classification for 3 classes.

	Feature	Acc_3 (%)	ρ_c
Lim et al. [16]	28 features from PSD	69.00	–
	56 features from PSD	69.20	–
Ours	Time	56.92	0.341
	B_1	51.10	0.524
	B_2	62.89	0.510
	Time + B_1	34.32	0.609
	Time + B_2	36.28	0.477
	B_1 + B_2	70.79	0.525
	Time + B_1 + B_2	**74.98**	**0.629**

frequency domains, which is an improvement over the previous work that used separate domain experiments. Additionally, our proposed method incorporates a Multi-dimensional Residual Block (MDRB) to extract information on channel relationships and frequency domain efficiently for the deep learning model. Furthermore, we conducted a novel experiment on the regression problem, which yielded good results and has the potential to contribute to the development of continuous-level mental workload state activity prediction systems.

Although we have achieved good results in MWL estimation, we have not thoroughly tested each channel or band. Our experiment involved using all EEG channels (14 channels) without performing specific experiments on each channel to evaluate their contributions to the MWL estimator process. In the frequency domain, we extracted information from six channels (delta, theta, alpha, sigma, beta, and gamma) and used them as a whole, but we did not perform separate experiments for each channel to assess the importance of each band in comparison to the effectiveness of combining them. These limitations need to be addressed in future research.

The STEW dataset has been widely used by researchers since its release in 2018. Table 8 provides the authors who used the STEW dataset for their study. However, there are variations in the segment inputs used by different authors, making it difficult to compare objectively such as Lim et al. [16] used a slide window of 512 with a shift of 128 as a baseline to perform experiments, the authors [9,28] used complete data duration, the authors [29] used more than 3000 time points, and

Table 6
The results for test set with change of shift on train set.

Shift	128		256		384		512	
	ρ_c	Acc (%)	ρ_c	Acc (%)	ρ_c	Acc (%)	ρ_c	Acc (%)
Time	0.341	56.92	0.349	32.65	0.231	42.86	0.243	40.59
B_1	0.524	51.10	0.527	**54.27**	0.417	58.16	0.449	55.97
B_2	0.510	62.89	0.417	50.60	0.365	**65.99**	0.408	61.38
Time + B_1	0.609	34.32	**0.53**	27.78	0.302	27.70	0.310	27.44
Time + B_2	0.477	36.28	0.341	27.51	0.355	27.63	**0.534**	27.51
B_1 + B_2	0.525	70.79	0.303	51.25	0.285	50.30	0.352	**62.43**
Time + B_1 + B_2	**0.629**	**74.98**	0.514	49.66	**0.446**	65.61	0.315	47.51

Table 7
Experiments of the MDRB if the branch (1) or the branch (2) is used.

Shift	Only branch (1)		Only branch (2)		Combine branch	
	Acc_3 (%)	ρ_c	Acc_3 (%)	ρ_c	Acc_3 (%)	ρ_c
128	48.56	0.024	56.16	**0.323**	**74.98**	**0.629**
256	43.84	0.194	53.17	0.110	49.66	0.514
384	**57.56**	**0.320**	46.03	0.090	65.61	0.446
512	50.98	0.094	**59.18**	0.003	47.51	0.315

Table 8
Segment of previous work.

	Year	Segment input	Train-Test data (%)	
			Train	Test
Lim et al. [16]	2018	Sliding window of 512 with shift of 128	80	20
Chakladar et al. [9]	2020	Complete data duration	–	–
Mohdiwale et al. [28]	2020	Complete data duration	80	20
Zhu et al. [29]	2021	Greater than 3000 time points	50	50
Taori et al. [14]	2022	Using 117 millisecond non-overlap	80	20

Table 9
Performance measures from the confusion matrix.

	Our proposed	Lim et al. [16]
Sensitivity	**74.98**	69.00
Specificity	**87.49**	84.50
Precision	**74.98**	69.00
Negative predictive	**87.49**	84.50

the authors [14] used 117 millisecond non-overlap. In this study, we followed the input provided in the data, using a window size of 512 and a shift size of 128 for an objective comparison. We demonstrated that our method of classifying the three classes yielded higher results than the baseline (74.98% > 69.00%), and also achieved good accuracy in other assessments such as sensitivity, specificity, precision, negative predictive value (Table 9).

6. Conclusion

This paper presents our experiments on the STEW dataset, where we propose a method that addresses both the classification and the estimation of continuous levels of MWL. Our proposed method can be applied to other EEG data problems that involve labels, such as classification and estimation of continuous levels. Additionally, we introduce a Multi-Dimensional Residual Block (MDRB), which utilizes 3D residual blocks to learn channel relationships and information bands of the frequency domain. To improve the performance of the model, we combine the frequency and time domains. We use TCNs to learn temporal information for the time domain and combine it with two branches of the frequency domain. Our experimental results show that our proposed model outperforms the baseline of the STEW dataset. We also demonstrate that our multimodal fusion approach, based on multi-domain, leads to better results as each domain contributes differently to the MWL estimation process. However, we acknowledge the limitation of our model in classifying the middle stage of mental workload, which we aim to improve in the future. We plan to conduct experiments on other types of estimation, such as sleep stages, emotions, depression, etc., for EEG data.

CRediT authorship contribution statement

Hong-Hai Nguyen: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Validation, Visualization, Writing – original draft, Writing – review & editing. **Ngumimi Karen Iyortsuun:** Writing – review & editing. **Seungwon Kim:** Funding acquisition, Supervision, Writing – review & editing. **Hyung-Jeong Yang:** Funding acquisition, Supervision, Writing – review & editing. **Soo-Hyung Kim:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Visualization, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government(MSIT) (RS-2023-00219107). This work was also supported by Institute of Information & communications Technology Planning & Evaluation (IITP), South Korea under the Artificial Intelligence Convergence Innovation Human Resources Development (IITP-2023-RS-2023-00256629) grant funded by the Korea government(MSIT).

References

[1] Neville Moray, Mental Workload: Its Theory and Measurement, vol. 8, 2013.

[2] Luis Cabañero Gómez, Ramón Hervás, Iván González, Vladimir Villarreal, Studying the generalisability of cognitive load measured with EEG, Biomed. Signal Process. Control 70 (2021) 103032.

[3] Paul A Kirschner, John Sweller, Richard E Clark, PA Kirschner, RE Clark, Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, in: Based Teaching Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching,(November 2014), 2010, pp. 37–41.

[4] Pengbo Zhang, Xue Wang, Junfeng Chen, Wei You, Weihang Zhang, Spectral and temporal feature learning with two-stream neural networks for mental workload assessment, IEEE Trans. Neural Syst. Rehabil. Eng. 27 (6) (2019) 1149–1159.

[5] Pengbo Zhang, Xue Wang, Weihang Zhang, Junfeng Chen, Learning spatial–spectral–temporal EEG features with recurrent 3D convolutional neural networks for cross-task mental workload assessment, IEEE Trans. Neural Syst. Rehabil. Eng. 27 (1) (2018) 31–42.

[6] Shouyi Wang, Jacek Gwizdka, W. Art Chaovalitwongse, Using wireless EEG signals to assess memory workload in the *n*-back task, IEEE Trans. Hum.-Mach. Syst. 46 (3) (2015) 424–435.

[7] Georgios N Dimitrakopoulos, Ioannis Kakkos, Zhongxiang Dai, Julian Lim, Joshua J deSouza, Anastasios Bezerianos, Yu Sun, Task-independent mental workload classification based upon common multiband EEG cortical connectivity, IEEE Trans. Neural Syst. Rehabil. Eng. 25 (11) (2017) 1940–1949.

[8] Maria Bannert, Managing cognitive load—recent trends in cognitive load theory, Learn. Instr. 12 (1) (2002) 139–146.

[9] Debashis Das Chakladar, Shubhashis Dey, Partha Pratim Roy, Debi Prosad Dogra, EEG-based mental workload estimation using deep BLSTM-LSTM network and evolutionary algorithm, Biomed. Signal Process. Control 60 (2020) 101989.

[10] Jacek Gwizdka, Distribution of cognitive load in web search, J. Am. Soc. Inf. Sci. Technol. 61 (11) (2010) 2167–2187.

[11] Julien Cegarra, Aline Chevalier, The use of tholos software for combining measures of mental workload: Toward theoretical and methodological improvements, Behav. Res. Methods 40 (4) (2008) 988–1000.

[12] Sandra G. Hart, Lowell E. Staveland, Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research, vol. 52, 1988, pp. 139–183.

[13] Gary B. Reid, Thomas E. Nygren, The subjective workload assessment technique: A scaling procedure for measuring mental workload, in: Advances in Psychology, vol. 52, 1988, pp. 185–218.

[14] Trupti J Taori, Shankar S Gupta, Suhas S Gajre, Ramchandra R Manthalkar, Cognitive workload classification: Towards generalization through innovative pipeline interface using HMM, Biomed. Signal Process. Control 78 (2022) 104010.

[15] Maarten A. Hogervorst, Anne-Marie Brouwer, Jan B.F. Van Erp, Combining and comparing EEG, peripheral physiology and eye-related measures for the assessment of mental workload, Front. Neurosci. 8 (2014) 322.

[16] W.L. Lim, O. Sourina, Lipo P. Wang, STEW: Simultaneous task EEG workload data set, IEEE Trans. Neural Syst. Rehabil. Eng. 26 (11) (2018) 2106–2114.

[17] Mahesh Y Ladekar, Shankar S Gupta, Yashwant V Joshi, Ramchandra R Manthalkar, EEG based visual cognitive workload analysis using multirate iir filters, Biomed. Signal Process. Control 68 (2021) 102819.

[18] Amin Zammouri, Abdelaziz Ait Moussa, Yassine Mebrouk, Brain-computer interface for workload estimation: Assessment of mental efforts in learning processes, Expert Syst. Appl. 112 (2018) 138–147.

[19] Shuo Yang, Zhong Yin, Yagang Wang, Wei Zhang, Yongxiong Wang, Jianhua Zhang, Assessing cognitive mental workload via EEG signals and an ensemble deep learning classifier based on denoising autoencoders, Comput. Biol. Med. 109 (2019) 159–170.

[20] Weizheng Qiao, Xiaojun Bi, Ternary-task convolutional bidirectional neural turing machine for assessment of EEG-based cognitive workload, Biomed. Signal Process. Control 57 (2020) 101745.

[21] Youngchul Kwak, Kyeongbo Kong, Woo-Jin Song, Byoung-Kyong Min, Seong-Eun Kim, Multilevel feature fusion with 3d convolutional neural network for eeg-based workload estimation, IEEE Access 8 (2020) 16009–16021.

[22] Xinqiao Zhao, Hongmiao Zhang, Guilin Zhu, Fengxiang You, Shaolong Kuang, Lining Sun, A multi-branch 3D convolutional neural network for EEG-based motor imagery classification, IEEE Trans. Neural Syst. Rehabil. Eng. 27 (10) (2019) 2164–2177.

[23] Tianjun Liu, Deling Yang, A three-branch 3D convolutional neural network for EEG-based different hand movement stages classification, Sci. Rep. 11 (1) (2021) 1–13.

[24] Ziyu Jia, Youfang Lin, Xiyang Cai, Haobin Chen, Haijun Gou, Jing Wang, Sst-emotionnet: Spatial-spectral-temporal based attention 3d dense network for eeg emotion recognition, in: Proceedings of the 28th ACM International Conference on Multimedia, 2020, pp. 2909–2917.

[25] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, et al., Tensorflow: Large-scale machine learning on heterogeneous distributed systems, 2016, arXiv preprint arXiv:1603.04467.

[26] Diederik P. Kingma, Jimmy Ba, Adam: A method for stochastic optimization, 2014, arXiv preprint arXiv:1412.6980.

[27] Philippe Remy, Temporal convolutional networks for keras, 2020, https://github.com/philipperemy/keras-tcn,

[28] Samrudhi Mohdiwale, Mridu Sahu, G.R. Sinha, Varun Bajaj, Automated cognitive workload assessment using logical teaching learning-based optimization and PROMETHEE multi-criteria decision making approach, IEEE Sens. J. 20 (22) (2020) 13629–13637.

[29] Guohun Zhu, Fangrong Zong, Hua Zhang, Bizhong Wei, Feng Liu, Cognitive load during multitasking can be accurately assessed based on single channel electroencephalography using graph methods, IEEE Access 9 (2021) 33102–33109.