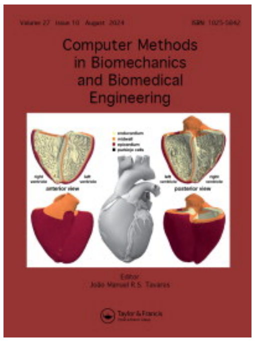

Taylor & Francis
Taylor & Francis Group

Computer Methods in Biomechanics and Biomedical Engineering

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/gcmb20

Classification of mental workload with EEG analysis by using effective connectivity and a hybrid model of CNN and LSTM

MohammadReza Safari, Reza Shalbaf, Sara Bagherzadeh & Ahmad Shalbaf

To cite this article: MohammadReza Safari, Reza Shalbaf, Sara Bagherzadeh & Ahmad Shalbaf (31 Jul 2024): Classification of mental workload with EEG analysis by using effective connectivity and a hybrid model of CNN and LSTM, Computer Methods in Biomechanics and Biomedical Engineering, DOI: 10.1080/10255842.2024.2386325

To link to this article: https://doi.org/10.1080/10255842.2024.2386325

Published online: 31 Jul 2024.

COMPUTER METHODS IN BIOMECHANICS AND BIOMEDICAL ENGINEERING
https://doi.org/10.1080/10255842.2024.2386325

Classification of mental workload with EEG analysis by using effective connectivity and a hybrid model of CNN and LSTM

MohammadReza Safari[a], Reza Shalbaf[a], Sara Bagherzadeh[b] and Ahmad Shalbaf[c]

[a]Institute for Cognitive Science Studies, Tehran, Iran; [b]Department of Biomedical Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran; [c]Department of Biomedical Engineering and Medical Physics, School of Medicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran

ABSTRACT

Estimation of mental workload from electroencephalogram (EEG) signals aims to accurately measure the cognitive demands placed on an individual during multitasking mental activities. By analyzing the brain activity of the subject, we can determine the level of mental effort required to perform a task and optimize the workload to prevent cognitive overload or underload. This information can be used to enhance performance and productivity in various fields such as healthcare, education, and aviation. In this paper, we propose a method that uses EEG and deep neural networks to estimate the mental workload of human subjects during multitasking mental activities. Notably, our proposed method employs subject-independent classification. We use the "STEW" dataset, which consists of two tasks, namely "No task" and "simultaneous capacity (SIMKAP)-based multitasking activity". We estimate the different workload levels of two tasks using a composite framework consisting of brain connectivity and deep neural networks. After the initial preprocessing of EEG signals, an analysis of the relationships between the 14 EEG channels is conducted to evaluate effective brain connectivity. This assessment illustrates the information flow between various brain regions, utilizing the direct Directed Transfer Function (dDTF) method. Then, we propose a deep hybrid model based on pre-trained Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) for the classification of workload levels. The accuracy of the proposed deep model achieved 83.12% according to the subject-independent leave-subject-out (LSO) approach. The pre-trained CNN + LSTM approaches to EEG data have been found to be an accurate method for assessing the mental workload.

ARTICLE HISTORY
Received 19 September 2023
Accepted 25 July 2024

KEYWORDS
Mental workload; EEG; brain connectivity; convolutional neural networks (CNN); long short-term memory (LSTM)

1. Introduction

Mental workload (MWL) is a concept that refers to how hard the brain is working to meet task demands (Mingardi et al. 2020) also defined as the number of mental resources needed to perform a task (das Chakladar et al. 2020). It is a complex, dynamic, person-specific, and non-linear construct that is multidimensional (Longo et al. 2022). Measuring MWL is widely used in various fields, such as helping researchers and practitioners understand the fluctuations in human performance (Sevcenko et al. 2021), identifying sources of error and improving performance in various industries, including medicine (Byrne 2011), estimating the fatigue index experienced by professional drivers, and evaluating mental health and education effectiveness (Bannert 2003).

There are several ways to measure MWL, including subjective measures, performance measures, and psychological measures. Subjective measures rely on self-reporting by the individual performing the task. A common method is to use a questionnaire asking subjects to rate the difficulty of the task (Byrne 2011). There are some well-known indices like the National Aeronautics and Space Administration (NASA) Task Load Index (NASA-TLX) (Hart and Staveland 1988) and Subjective Workload Assessment Technique (SWAT) (Reid and Nygren 1988). This method is based on the assumption that individuals can accurately report their MWL (Ouwehand et al. 2021). Subjective measures are easy to administer and can provide valuable insights into how individuals perceive their own MWL. However, they are subject to bias and may not accurately reflect the actual MWL (Schnotz and Christian 2007). Measuring MWL through performance measures involves assessing the performance of individuals on a task as an indirect measure of their MWL (Sevcenko et al. 2021). Finally, the psychophysiological measures method involves

CONTACT Ahmad Shalbaf shalbaf@sbmu.ac.ir

measuring MWL by analyzing physiological signals such as heart rate variability, electroencephalography (EEG), and functional near-infrared spectroscopy (fNIRS) (Chen et al. XXXX) that among these signals, EEG for its fast data acquisition and convenient usage has been widely used in the study of MWL (Zhu et al. 2021). Measuring MWL through EEG offers several benefits including assessment in real-time, without subject biasing, portable methods, high temporal resolution, and being noninvasive (Zhou et al. 2022).

Traditionally, machine learning methods based on EEG have been used to classify MWL classes using various algorithms, like support vector machine (SVM) (Sciaraffa et al. 2019; Dimitrakopoulos et al. 2017; Mazher et al. 2017; Almogbel et al. 2018; Yu et al. 2015; Zarjam et al., 2011a; Walter et al. 2013; Zarjam et al., 2011b; So et al. 2017), and linear discriminative analysis (LDA) (Dehais et al. 2019; Aricò et al. 2016; Roy et al. 2016; Kakkos et al. 2019). These methods involve feature extraction, feature selection, and classification of EEG signals to measure MWL. Recent studies have explored the use of deep learning algorithms and other advanced techniques to improve the accuracy and reliability of MWL assessment (Jiao et al. 2018; Zhang and Shen 2019; Qiao and Bi 2020; Kwak et al. 2020; Zhang et al. 2019; Siddhad et al. 2022). These methods offer more objective and accurate measures of MWL and can provide valuable insights into cognitive demands and fluctuations in human performance. These algorithms are able to automatically extract significant features and classify them directly from the data, without the need for manual feature extraction. Deep learning algorithms have become more effective than traditional machine learning algorithms due to the availability of larger training data and greater computer processing power in recent years (LeCun et al. 2015). Recently, in Siddhad et al. (2022) a deep learning model for assessing MWL is proposed. They have used a Transformer network to classify MWL from raw EEG data. Moreover, in Khanam et al. (2023) a framework is proposed that utilizes discrete wavelet transform (DWT) to decompose the EEG signal for extracting the non-stationary features of task-wise EEG signals. Furthermore, a SVM is implemented to classify the task from the DWT-based extracted features. The proposed methodology has been implemented on an EEG dataset that captured three levels of cognitive load from the n-back test (Shin et al. 2018). The linear SVM achieved the average classification accuracy of 77.20 ± 6.63 and 87.89 ± 7.3 for 3-class and 2-class approaches, respectively. Another study Karacan et al. (2023), aims to classify the mental workload level of MS patients as low, medium, or high from EEG signals during cognitive tasks in computer and virtual reality environments and to compare them with a healthy group performing the same tasks. The mental workload level of 45 volunteers is estimated by using EEG signals and the NASA-Raw Task Load Index questionnaire results in 3 cognitive tasks in computer and virtual reality environments. The three-level mental workload classification accuracy in MS patients with the SVM classifier is 96.08% and 94.12% for computer and virtual reality environments, respectively. For healthy volunteers, classification accuracy is 95.24% and 94.05% in computer and virtual reality environments, respectively. In addition Mohanavelu et al. (2022), aimed to analyze the cognitive workload of fighter pilots during different flight phases using physiological signals such as ECG and EEG. The researchers employed classification algorithms including LDA, SVM, and k-Nearest Neighbour (k-NN) to classify the pilots' cognitive workload. The results showed that the LDA and SVM with an accuracy of 75% were found to be more consistent classifiers compared to the k-NN classifier with an accuracy of 60%. Also, another study (Raufi and Longo 2022) aimed to assess the impact of alpha-to-theta and theta-to-alpha band ratios on creating models capable of discriminating self-reported perceptions of mental workload. This study used the STEW dataset and results indicate high classification accuracy of those models trained with the high-level features extracted from the alpha-to-theta ratios and theta-to-alpha ratios, demonstrating the richness of information in the temporal, spectral, and statistical domains extracted from these EEG band ratios for the discrimination of self-reported perceptions of mental workload.

Moreover, in another study, a composite framework has been proposed consisting of a Grey Wolf Optimizer for selecting the most relevant features and a hybrid deep learning model based on Bidirectional Long Short-Term Memory (BLSTM) and Long Short-Term Memory (LSTM) for classification (das Chakladar et al. 2020).

In this research paper, we present three main contributions. Firstly, we use a method that is proposed in Bagherzadeh et al. (2022) for the purpose of detecting schizophrenia by converting EEG data into three-dimensional spatial-time color images using effective brain connectivity. This approach allows us to represent EEG signals in a visually informative manner, capturing both spatial and temporal information in order to feed the model. Secondly, we design a hybrid model that combines a pre-trained CNN with LSTM layers. This hybrid model leverages the strengths of both architectures, enabling more effective feature extraction and capturing temporal dependencies in the EEG data. Lastly, we use subject-independent (SI) classification, which aims to develop a model that can accurately classify EEG signals regardless of the individual subject. This approach has the potential to enhance the generalizability and applicability of EEG-based classification models. By incorporating these three innovations, our research contributes to the advancement of EEG analysis and provides valuable insights into the assessment of mental workload.

2. Material and methods

2.1. Participants and EEG recording

We have used the Simultaneous Task EEG Workload (STEW) dataset which is an open-access dataset consisting of raw EEG data from 48 male subjects from the university's graduate population who participated in a multitasking workload experiment utilizing the SIMKAP multitasking test (Lim et al. 2018). The experiment consisted of two parts: First, data was collected for 2.5 min while subjects were at rest, or "No task," which was labeled as "low" MWL. Second, subjects performed the SIMKAP test while their EEG was recorded, and the final 2.5 min were used as the workload condition, labeled as 'high' MWL. The EEG signals were obtained using the Emotiv EPOC EEG headset, featuring a 16-bit A/D resolution, 128 Hz sampling frequency, and 14 channels (AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, and AF4) based on the 10–20 international electrode placement system, along with two reference channels (common mode sense (CMS), driven right leg (DRL)). The STEW dataset is valuable for studying multitasking workload and analyzing brain activity during different cognitive tasks. Researchers can use this dataset to develop and evaluate algorithms and models for mental workload classification and prediction.

2.2. Preprocessing

We followed the preprocessing pipeline that is suggested in the database-providing paper (Lim et al. 2018). The pipeline is as below:

1. Applying a high-pass filter to the initial data at a frequency of 1 Hz to eliminate low-frequency disturbances. These disturbances may originate from various sources, such as sweat on the scalp, head movements, movement of electrode cables, or gradual changes in the EEG readings over extended periods.
2. Eliminating line noise, which arises from electrical devices like power cables that produce electromagnetic fields, disrupting the EEG recordings.
3. Utilizing artifact subspace reconstruction (ASR) to identify and remove any abnormal disturbances or anomalies present in the EEG data automatically.
4. Re-referencing the data to average to convert the readings from a set or standard reference to an "average reference." This method is recommended, particularly when the electrode setup encompasses almost the entire scalp.

In this pipeline applying ASR which is a non-stationary method to remove large-amplitude artifacts (Mullen et al. 2015) is very important because data has many large amplitude artifacts. Figure 1 provides a sample of data before and after this preprocessing pipeline. All steps in preprocessing are done by the EEGLAB toolbox (version 2021.0) in MATLAB software (version 2019a).

2.3. Convert EEG data to image

There are some ways to convert EEG data to images such as Continuous Wavelet transform (CWT) Short-Time Fourier Transform (STFT) (Chaudhary et al. 2019) and connectivity-based methods (Bagherzadeh et al. 2022). There are some brain connectivity measures for measuring brain connectivity and we've chosen the direct Directed Transfer Function (dDTF) due to its high accuracy in various fields such as emotion recognition (Bagherzadeh et al. 2022) and major depressive disorder diagnosis in previous studies (Shahabi et al. 2023). dDTF estimates direct causal relations between signals based on the frequency domain of conditional Granger-Causality. It is a modified version of the Directed Transfer Function (DTF) and is used to model effective brain connectivity. The dDTF method allows for the determination of information flow direction among brain structures, providing insights into the causal influence of one neural region over others. The dDTF from channel j to channel i at frequency f is estimated by Equation (1) (Korzeniewska et al. 2003):

$$dDTF_{ij} = \frac{|H_{ij}(f)|^2}{\sum_f \sum_{k=1}^{M} |H_{ik}(f)|^2} \times \frac{\hat{S}_{ij}(f)}{\sqrt{\hat{S}_{ii}(f)\hat{S}_{jj}(f)}} \tag{1}$$

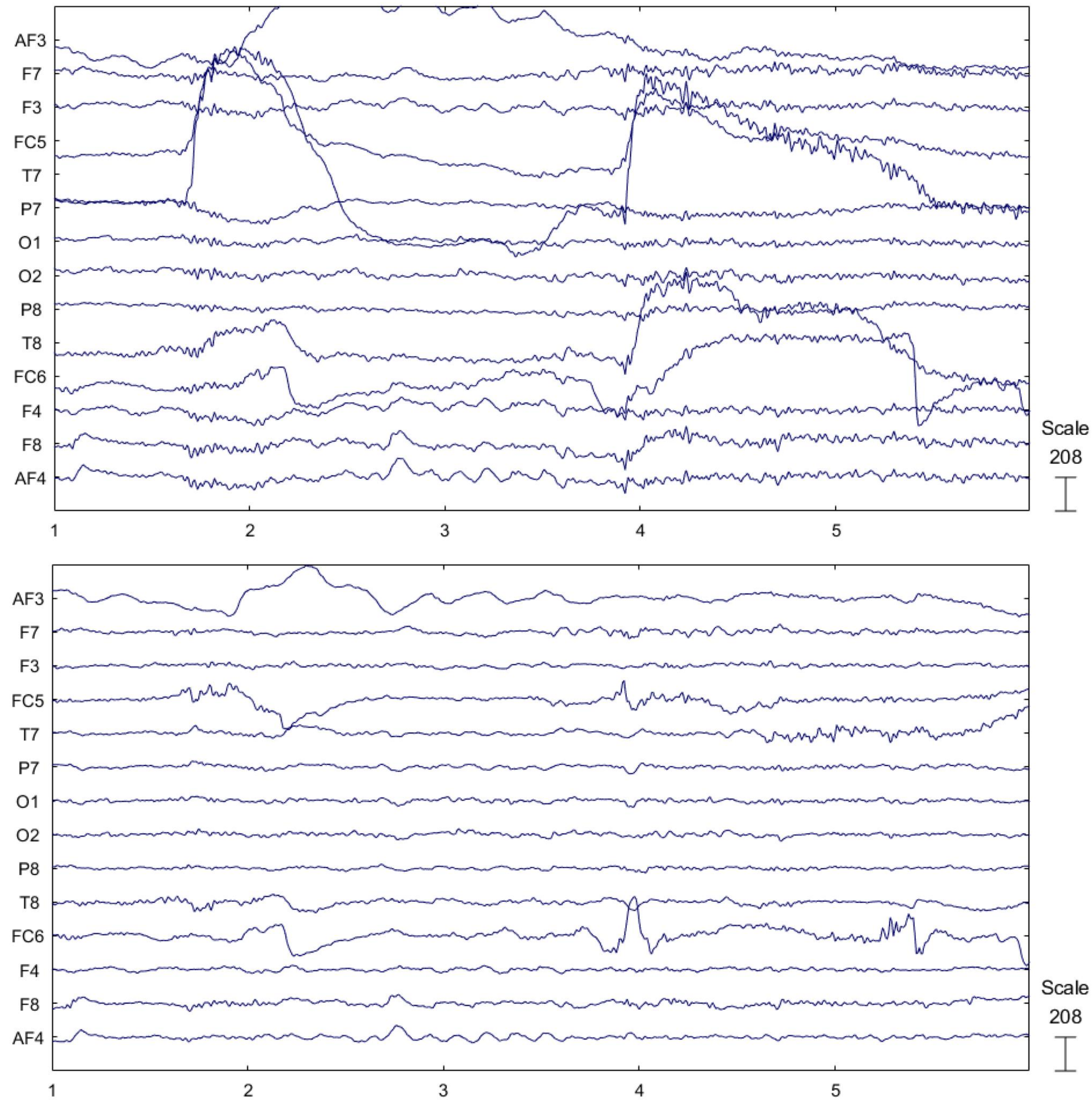

Figure 1. EEG signal before and after preprocessing. The top panel shows the raw EEG signal, while the bottom panel shows the same signal after preprocessing. The preprocessing steps applied to the signal are described in the methods section.

$$S(f) = H(f) \sum H(f)^* \tag{2}$$

where $H(f)$ is the transfer matrix of the system at a specific frequency and $S(f)$ is the spectral density matrix. We extract five frequency ranges for the dDTF measure by averaging the frequency spectrum as follows: delta (2–4 Hz), theta (4–8 Hz), alpha (8–13 Hz), beta (13–32 Hz), and gamma (32–50 Hz). For generating dDTF images, we primarily select three key parameters: window length, move step size, and model order. Subsequently, we apply this window to the EEG data within each frequency band. Ultimately, using dDTF measures, we obtain a connectivity image corresponding to each step of the window. This image is represented as an $n \times n$ matrix, where n is equal to the number of channels. All steps for dDTF measurement are done in MATLAB software *via* the Source Information Flow Toolbox (SIFT) version 0.1a (Mullen 2010).

2.4. Pre-trained CNN models

Convolutional Neural Networks (CNNs) are a specialized type of neural network architecture designed to process data with a grid-like topology, such as images. CNN models are composed of two main parts: the feature extraction component and the classification component. The feature extraction component includes convolutional layers, pooling layers, and fully connected layers. Convolutional layers extract high-level features from the input image while typically added after convolutional layers. Its main purpose is to reduce the spatial dimensions (width and height) of the feature maps while preserving the depth (number of channels). Fully connected layers prepare the extracted features to be classified by the softmax layer (Mullen 2010). CNNs have good properties such as generalization and flexibility, which improve classification accuracy. In previous studies, three pre-trained CNN models, Densenet121, VGG-16, and ResNet-50, that are trained on the ImageNet dataset, have been utilized in various fields and have achieved high accuracy (Bagherzadeh et al. 2022; Mullen 2010; Shahabi and Shalbaf 2023; Shahabi et al. 2023).

ImageNet is a large visual database designed for use in visual object recognition software research. It contains over 14 million images that have been hand-annotated to indicate what objects are pictured, and in at least one million of the images, bounding boxes are also provided. The images are organized and labeled in a hierarchy according to the WordNet hierarchy, with more than 20,000 categories, and each category consists of several hundred images. The dataset is designed for developing computer vision algorithms, and it is used to train large-scale object recognition models.

In this study, the mentioned models (Densenet121, VGG-16, and ResNet-50) were chosen due to their established success in previous studies to improve the accuracy of mental workload classification from EEG signals. DenseNet121 is a convolutional neural network architecture that utilizes dense connections between layers through dense blocks, where each layer obtains additional inputs from all preceding layers and passes on its own feature maps to all subsequent layers. In a traditional feed-forward CNN, each convolutional layer except the first one receives the output of the previous convolutional layer and produces an output feature map that is then passed on to the next convolutional layer. In contrast, in DenseNet121, each layer is connected directly with every other layer. The architecture consists of a series of dense blocks, each containing multiple convolutional layers. Each dense block takes the output of the previous block as input, as well as the outputs of all the previous blocks. This creates a dense connectivity pattern between all the layers of the network, allowing information to flow more efficiently through the network. VGG16 is a simple and sequential convolutional neural network architecture that consists of convolutional layers, pooling layers, and fully connected dense layers. It has 13 convolutional layers, 5 pooling layers, and 3 fully connected dense layers and takes an input image of dimensions $224 \times 224 \times 3$. ResNet-50 is a variant of the ResNet model that can be divided into input pre-processing, configuration blocks, and a fully connected layer. Different versions of the ResNet architecture use a varying number of configuration blocks at different levels. ResNet-50 achieved high accuracy on the ImageNet dataset and is widely used for image classification tasks (Mohanty et al. 2022). These three models were implemented in the Keras package with the TensorFlow backend and Python programming. The aim of the study was to use these models to classify mental workload based on EEG signals.

2.5. LSTM models

LSTM is a type of Recurrent Neural Network (RNN) that is designed to solve the vanishing gradient problem that traditional RNNs struggle with. LSTMs can retain information for longer periods and are capable of learning long-term dependencies, which is useful for certain types of prediction that require the network to retain information over longer time periods. LSTMs are widely used in complex problem domains such as machine translation and speech recognition. They are capable of processing and predicting entire sequences of data, making them ideal for processing time-series data like analyzing EEG data in areas like detecting MWL (das Chakladar et al. 2020).

2.6. The proposed model

The pre-trained CNN-LSTM hybrid model combines the capacity to extract deep specific features from pre-trained CNNs and investigate time dependency using LSTM (Bagherzadeh et al. 2022). This combination has been traditionally employed in various fields, such as major depressive disorder diagnosis and schizophrenia detection, with high accuracy and successful outcomes (Bagherzadeh et al. 2022; Shahabi et al. 2023; Shahabi and Shalbaf 2023). In the proposed model, whose structure is shown in Figure 2, initially, every 5 sequential dDTF images with a size of $224 \times 224 \times 3$ form an input unit with dimensions (5, 224, 224, 3) to be fed into the CNN-LSTM model. The TimeDistributed layer is then applied to CNN

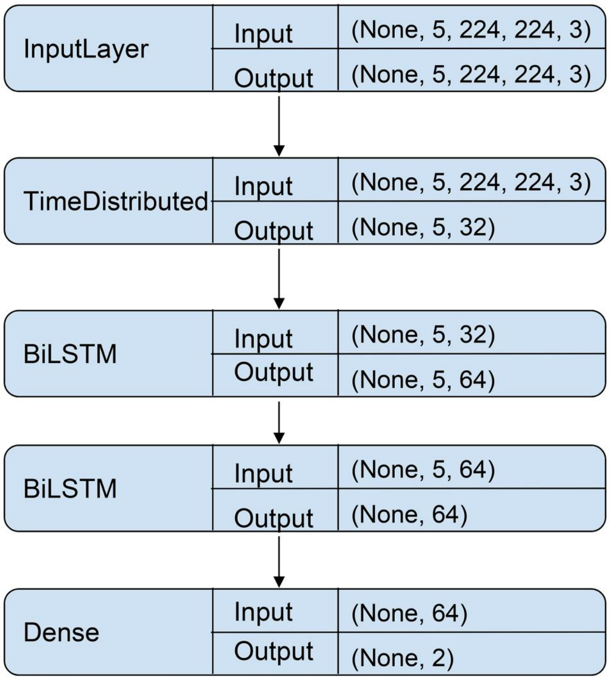

Figure 2. Block diagram of the hybrid pre-trained CNN-LSTM model. This figure shows the shape of input and output of each layer in the model.

models to handle all 5 images across the temporal dimension. Bidirectional LSTM layers analyze the most advantageous features from each input sample by examining it in both directions of the temporal dimension. The first bidirectional LSTM layer generates 64 features from each input sample. The output of the second bidirectional LSTM layer with 64 neurons then proceeds to the dense layer with the softmax activation function to classify MWL.

2.7. Data splitting and training

In this study, we used the SI approach to improve the validity and generalization of the proposed model, and to achieve this purpose we applied the leave-subject-out (LSO) method. We randomly selected 6 subjects as test group and trained the model with 42 other subjects and this process was repeated 5 times. Finally, the average and standard deviation of the 5 iterations were taken as the final result. By using this method, we ensured that the model was tested with different subjects who didn't see any data from them in training sessions, so the final result has high validity. Also in each iteration, the training data was split into training and validation data. The validation data is needed to check that the model isn't overfitted during the training session.

In this study, all trainable parameters of the pre-trained CNN model are fine-tuned by dDTF images with Stochastic gradient descent (SGD) as an optimizer (Ruder 2016), categorical cross-entropy as the loss function, and learning rate equal to 1e-4. All training processes are done in Python programming language and run on Google Colab with 25GB RAM and NVIDIA T4 GPU.

2.8. Model evaluation

To evaluate the model, 20% of the training data is split for validation data in each iteration so the model is trained on 80% of training data and evaluated on 20% of training data in 50 epochs, and after that, we calculate and plot the accuracy and loss of the model on both training and validation data. Finally, we pick the best result between all epochs based on the minimum validation loss as the final result of the iteration. In addition, we have used six performance measures that are computed by Equations (3)–(7) as below (Sokolova and Lapalme 2009):

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{3}$$

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{5}$$

$$Specificity = \frac{TN}{TN + FP} \tag{6}$$

$$False\ Rejection\ Rate\ (FRR) = \frac{FN}{FN + TP} \tag{7}$$

$$False\ Acceptance\ Rate\ (FAR) = \frac{FP}{FP + TN} \tag{8}$$

3. Results

The EEG data from each subject's 14 channels were pre-processed using the EEGLAB toolbox in MATLAB software (version 2019a). Subsequently, the EEG signals were transformed into brain connectivity images using the dDTF method. dDTF images were built from every 6-s window of 14 channels of each subject per 5 frequency bands (delta, theta, alpha, beta, and gamma) with model order equal to 6. A window was moved in EEG signals with a step size equal to 4 s. In Figure 3, we have presented some samples of dDTF images extracted from both 'high' and "low" MWL classes related to subject 38, categorized by frequency band. It is apparent from this figure that the brain connectivities are generally lower when MWL is low compared to when MWL is high. This trend was consistently observed across all subjects. Horizontal axes and vertical axes represent channels. Considering 150 s of EEG signals, 6 s as

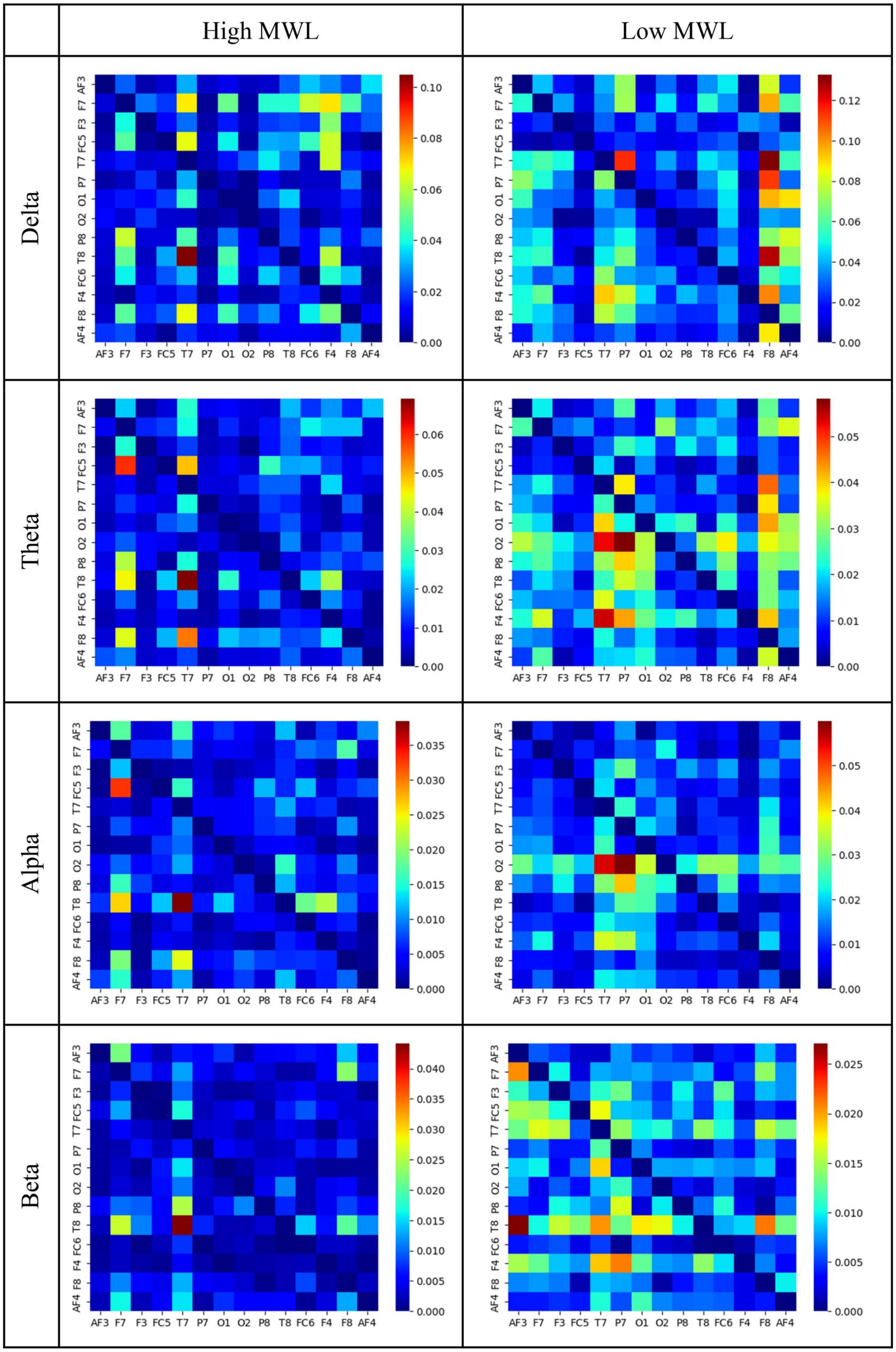

Figure 3. dDTF images in each frequency band of subject 38 in two high and low mental workload groups.

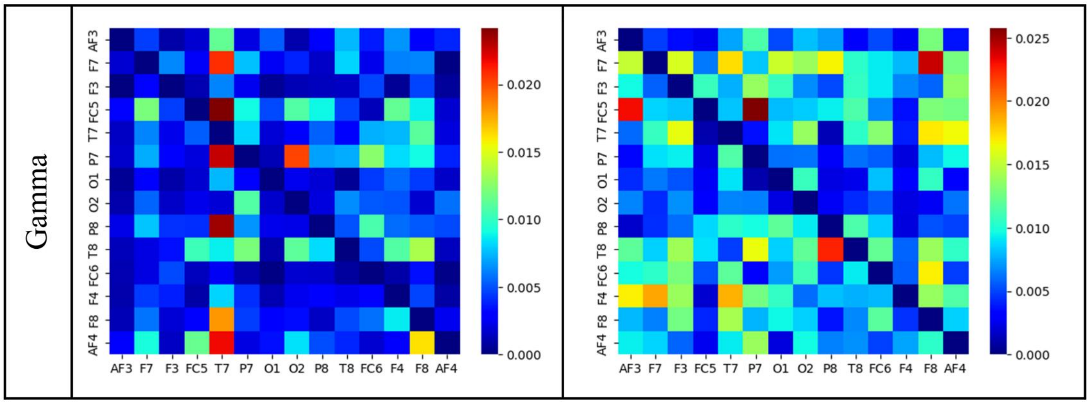

Figure 3. Continued.

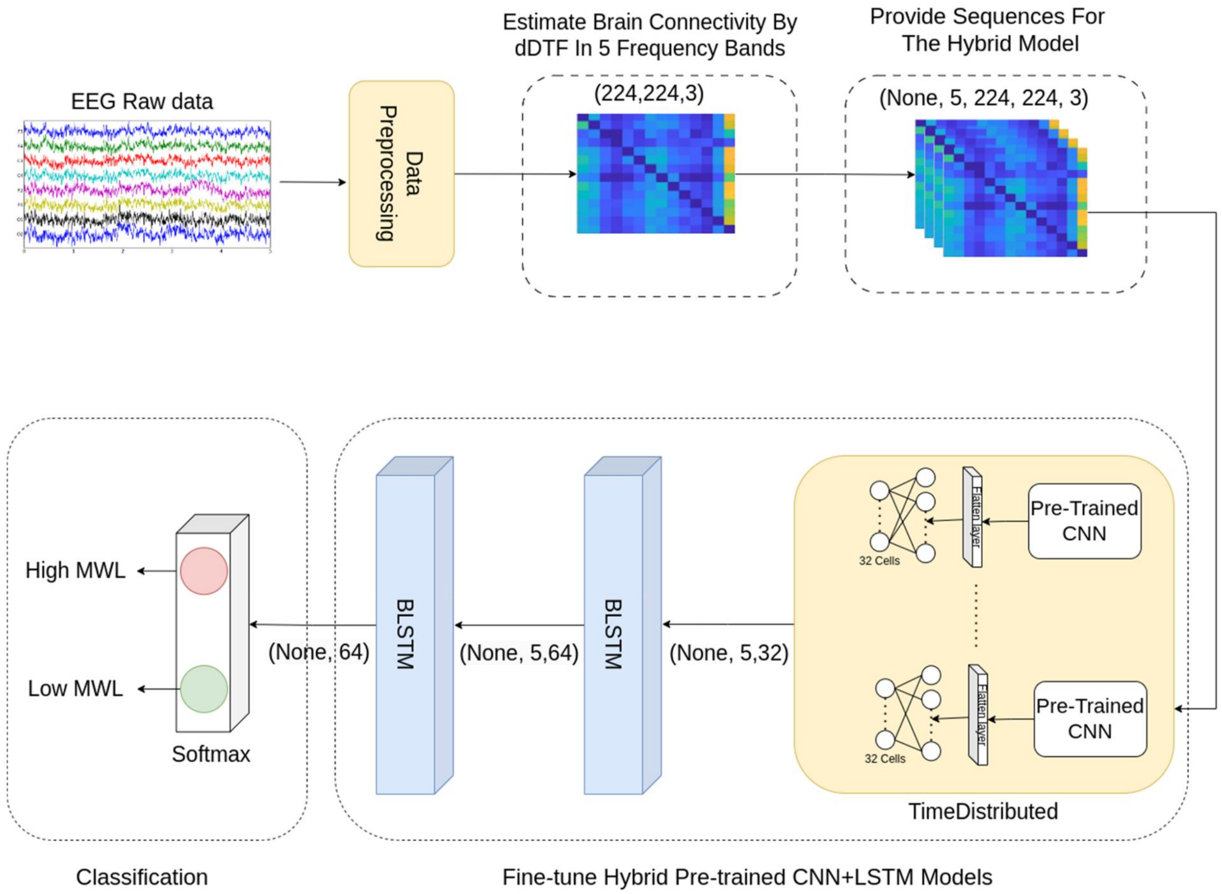

Figure 4. Block diagram illustrating the proposed method for the MWL detection from multichannel EEG using hybrid CNN-LSTM models.

window size, and 4 s as window step, we achieve 37 images per EEG data, so with two classes and 42 and 6 subjects as the training and test group, we got 3108 and 444 images as training and test data, respectively. Pre-trained CNN models require input images with specific sizes, therefore, each dDTF image is resized to $224 \times 224 \times 3$. In order to create spatial-time images to feed the hybrid CNN + LSTM model, we move a window in images with a size equal to 5 and a step equal to 1. So finally, we got 2772 and 393 video data for training and test data respectively and each video data consisting of 5 dDTF images. All of these steps are shown in Figure 4.

After feeding training data to the models, the results for training and validation data were extracted for each frequency band in three hybrid models. All accuracy and loss curves of training and validation data in all three hybrid models for each frequency band are shown in Figures 5 and 6, respectively. As we mentioned before, to prevent overfitting, we

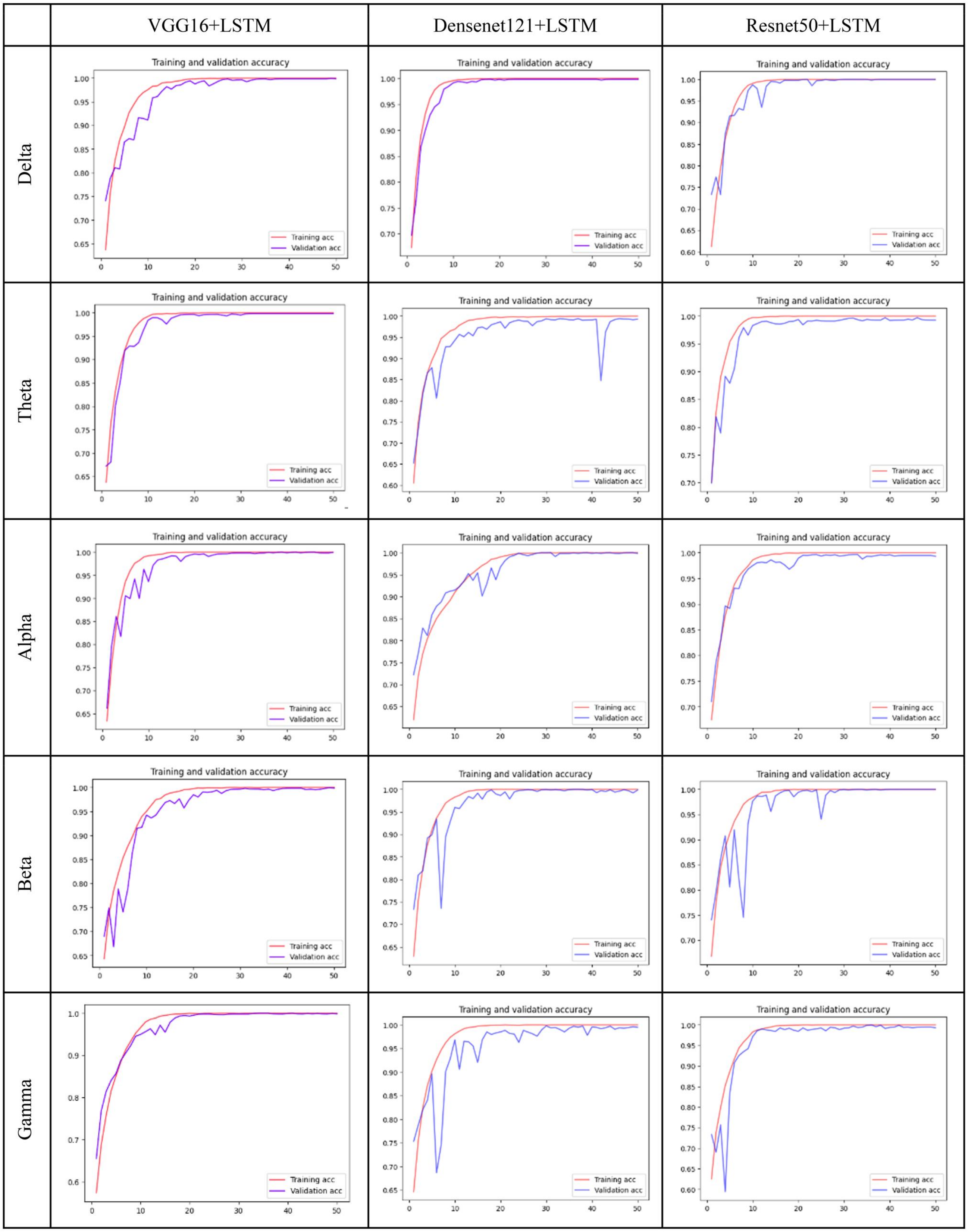

Figure 5. The accuracy curve of hybrid pre-trained CNN-LSTM models for training (red) and validation (blue) dDTF images in every brain frequency band.

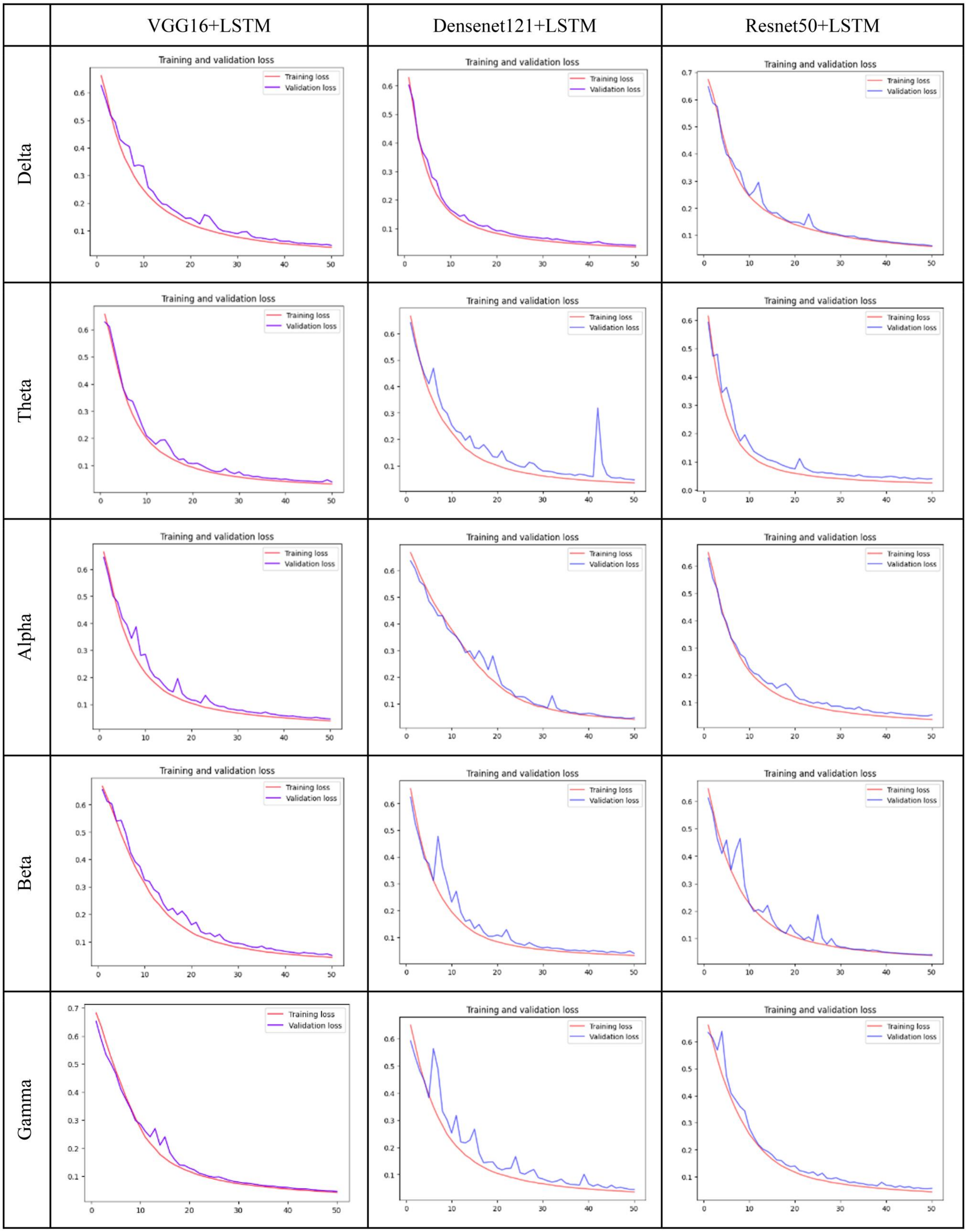

Figure 6. The loss curve of hybrid pre-trained CNN-LSTM models for training (red) and validation (blue) dDTF images in every brain frequency band.

reserved a portion of the data for validation, and by analyzing the accuracy and loss curves of validation data, it was confirmed that no overfitting occurred. For all training sessions, the learning rate, batch size, and epoch were equal to 1e-4, 10, and 50, respectively. We reached these hyperparameters through a process of trial and error. After the training session in each iteration, the models were tested by the test data that were split by the SI method. After getting all the results, we reported the mean and standard deviation for each frequency band in Table 1. Also, the confusion matrix for each model in the alpha frequency band which based on Table 1 has the best performance between all frequency bands, is shown in Figure 7.

The best model for predicting the MWL classes based on accuracy, precision, specificity, sensitivity, FAR, and FRR is the ResNet50 + LSTM which is trained with images extracted from the alpha frequency band. This model predicted the 'low' MWL class very well and got a total accuracy of 83.1%. Overall, it's clear from the results that the alpha frequency band in all three proposed models was the best frequency to classify MWL classes in the proposed method.

4. Discussion

In this study, we proposed a framework to detect MWL based on the effective brain connectivity images calculated with dDTF and the hybrid pre-trained CNN + LSTM model. The proposed framework achieved 83.1% accuracy and 85% sensitivity in predicting MWL classes by using the ResNet50 + LSTM model that was fine-tuned with dDTF images extracted from the alpha frequency band. Traditionally, shown in Marinescu et al. (2018), Puma et al. (2018), and Xie et al. (2016) that alpha and theta frequency bands are the best indicators for predicting MWL, this study verifies that because in the proposed framework, the alpha frequency band is the most helpful indicator for predicting MWL classes in the proposed framework. Additionally, we can see from the results that between the three hybrid pre-trained CNN + LSTM models that we used, the Resnet50 was better than VGG16 + LSTM and Densenet121 + LSTM based on all metrics and

Table 1. Performance metrics (mean ± std) for Resnet50 + LSTM, VGG16 + LSTM, and Densenet121 + LSTM models on every brain frequency band for subject-independent classification of MWL. The metrics include accuracy, precision, specificity, sensitivity, FAR, and FRR.

Model	Freq. band	Accuracy (%)	Precision (%)	Specificity (%)	Sensitivity (%)	FAR (%)	FRR (%)
Resnet50 + LSTM	Delta	65.1±6.6	63.6±6.2	67±7	71.4±5.9	33±7	36.4±6.2
	Theta	66.9±7	63.2±5	74.4±11.7	80.8±9.7	25.6±11.7	36.8±5
	Alpha	83.1±6.3	85±11.4	84.4±9.7	83.2±14	15.6±9.7	15±11.4
	Beta	71.4±7.1	67.2±7	79.4±5.1	85.8±2.8	20.6±5.1	32.8±7
	Gamma	64.5±7.2	61±5.3	71.6±11.6	80.6±8.6	28.4±11.7	39±5.3
VGG16 + LSTM	Delta	62.6±3.5	61±2.6	64.6±4.7	70±5.2	35.4±4.7	39±2.6
	Theta	67.8±3.7	64.6±2.8	72.8±5.9	78.4±5.5	27.2±5.9	35.4±2.8
	Alpha	72.3±7	69.6±6.7	76.2±7.7	80±6	23.8±7.7	30.4±6.6
	Beta	70.3±5.5	67.2±4.9	74.8±6.4	79.4±5.6	25.2±6.4	33±4.8
	Gamma	69.8±5.5	66.2±4.9	75.2±6.1	80.8±4.7	24.8±6.1	33.8±4.9
Densenet121 + LSTM	Delta	60±2.2	60±2.3	60±2.6	61.4±3.2	40±2.3	40±2.3
	Theta	69.6±6.4	67.2±4.6	73.4±9.6	76±11.1	26.8±9.6	32.8±4.6
	Alpha	73.9±6.3	68.8±5.6	82.6±7.4	87.6±5.1	17.4±7.4	31.2±5.6
	Beta	65.5±4.5	64.6±3.8	66.8±5.9	68.8±6.3	33.2±5.93	35.4±3.8
	Gamma	71.8±4.7	66.4±3.5	83±8.5	88.4±6	17±8.5	33.6±3.5

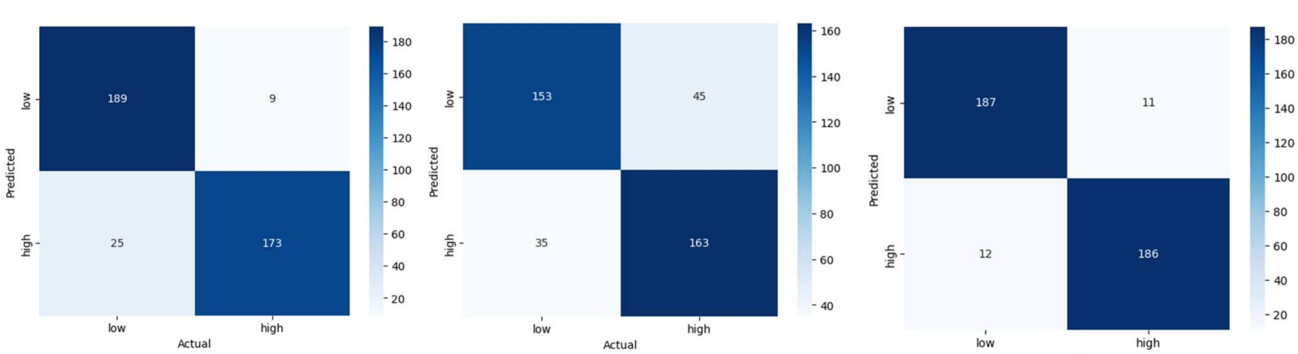

Figure 7. Confusion matrix for the three hybrid models in the alpha frequency band, which achieved the highest accuracy among all frequency bands. Resnet50-LSTM (left) – VGG16-LSTM (Middle) – Densenet121 + LSTM (right).

Table 2. Comparisons of the prediction accuracy (%) of subject-independent experiments among the various methods.

References	Method	SI* Accuracy (%)	SD* Accuracy (%)
(Zhu et al. 2021)	SVM	–	89.6
(das Chakladar et al. 2020)	Random Forest	–	83
(das Chakladar et al. 2020)	CNN + LSTM	–	85.21
(Kingphai and Moshfeghi 2021)	BLSTM + LSTM	–	91.15
(Siddhad et al. 2023)	Transformer network	–	95.28
(Taori et al. 2022)	Variable length frame and deep BLSTM-LSTM	–	100
(Fan et al. 2022)	EEG-Tnet	82.78	99.82
(Ji et al. 2023)	DRNA-Net	60.55	–
(Pandey et al. 2020)	k-NN	61.08	–
(Pandey et al. 2020)	MLP	58.68	–
(Pandey et al. 2020)	LSTM	57.3	–
(Pandey et al. 2020)	CNN + LSTM	57.19	–
(Pandey et al. 2020)	Random Forest	58.52	–
Our model	**Pre-trained CNN + LSTM**	**83.1**	–

between all frequency bands. Also, the most stability in the accuracy of predicting MWL among all iterations is provided by the Densenet121 + LSTM model at the delta frequency band. The accuracy of this model in the delta frequency band is 60% with a standard deviation of 2.2%.

In Table 2 we provide a comparison between all studies that proposed a framework to predict MWL classes from the STEW dataset. In most studies on EEG data classification, subject-dependent (SD) classifications are often considered more accurate due to the ability to obtain EEG data for each individual during the training phase (Fan et al. 2022) and the correlation between EEG channels, so they are not generalized models. Also, in the Brain-Computer Interface (BCI) area, SI classification enables the development of BCIs that different individuals can use without the need for individual calibration. Additionally, SI classification enhances the usability and practicality of BCIs (Zhang et al. 2023). Totally, the proposed framework has the advantage of having state-of-the-art accuracy in a subject-independent approach, but the disadvantage is that it requires high processing resources due to the use of transfer learning and the high number of learnable parameters. The use of brain effective connectivity analysis through dDTF and a combined deep learning model has significantly advanced the field of MWL assessment. This method enhances the precision of MWL assessment and also facilitates the development of more robust and interpretable models for MWL assessment.

In this research, there were several limitations that need to be acknowledged. Firstly, the data size was limited due to the use of the STEW dataset, which only contains EEG data from 48 male subjects. Secondly, the lack of data variety due to the STEW dataset only containing male subjects may limit the model's ability to generalize to other genders or populations. Lastly, the processing resources were limited, which prevented the exploration of different models such as Efficientnetb0 or Inception which are powerful models (Bagherzadeh et al. 2022; Mullen 2010). If the processing resources were not limited, it would have been possible to try different models and hyperparameters and compare their performances. Despite these limitations, the proposed model achieved promising results.

In future research, we plan to extend our work by using different pre-trained CNN models and hyperparameters to select the optimum combination that improves the estimation accuracy of MWL. Additionally, we aim to build a new database that includes more occupational groups and all genders to address the limitation of data variety in the STEW dataset. This will enhance the generalization ability of our model and improve its practical applicability. Researchers conducting future studies may find it advantageous to broaden their dataset by incorporating a more diverse range of participants, including those from various genders and demographic backgrounds. This expansion could enhance the robustness and generalizability of the research findings. Furthermore, investigating a wider variety of models beyond those utilized in the current study could also provide valuable insights.

5. Conclusion

In this research, we proposed a framework for predicting MWL classes that contains a hybrid CNN + LSTM and converting EEG raw data to effective brain connectivity images by dDTF measure. We examined three pre-trained CNN models (Densenet121, Resnet50, and VGG16) as the CNN base of the hybrid model. Our framework could achieve state-of-the-art accuracy equal to 83.1% in the SI classification.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This research is financially supported by "Shahid Beheshti University of Medical Sciences" (Grant No 43004523).

References

Almogbel MA, Dang AH, Kameyama W. 2018. Cognitive workload detection from raw EEG-signals of vehicle driver using deep learning. In International Conference on Advanced Communication Technology (ICACT). pp. 1–6.

Aricò P, Borghini G, Flumeri GD, Colosimo A, Babiloni F. 2016. A passive brain-computer interface application for the mental workload assessment on professional air traffic controllers during realistic air traffic control tasks. Progr Brain Res. 228:295–328.

Bagherzadeh S, Maghooli K, Shalbaf A, Maghsoudi A. 2022. Emotion recognition using effective connectivity and pre-trained convolutional neural networks in EEG signals. Cogn Neurodyn. 16(5):1087–1106. doi: 10.1007/s11571-021-09756-0.

Bagherzadeh S, Shahabi MS, Shalbaf A. 2022. Detection of schizophrenia using hybrid of deep learning and brain effective connectivity image from electroencephalogram signal. Comput Biol Med. 146:105570. doi: 10.1016/j.compbiomed.2022.105570.

Bannert M. 2003. Managing cognitive load—recent trends in cognitive load theory. Learn Instr. 12(1):139–146. doi: 10.1016/S0959-4752(01)00021-4.

Byrne A. 2011. Measurement of mental workload in clinical medicine: a review study. Anesth Pain Med. 1(2):90–94. doi: 10.5812/kowsar.22287523.2045.

Chaudhary S, Taran S, Bajaj V, Sengur A. 2019. Convolutional neural network based approach towards motor imagery tasks EEG signals classification. IEEE Sensors J. 19(12):4494–4500. doi: 10.1109/JSEN.2019.2899645.

Chen S, Epps J, Chen F. A comparison of four methods for cognitive load measurement. In Proceedings of the 23rd Australian Computer-Human Interaction Conference (OzCHI '11). Association for Computing Machinery, New York, NY, USA, 76–79. doi: 10.1145/2071536.2071547.

das Chakladar D, Dey S, Roy PP, Dogra DP. 2020. EEG-based mental workload estimation using deep BLSTM-LSTM network and evolutionary algorithm. Biomed Signal Process Control. 60:101989. doi: 10.1016/j.bspc.2020.101989.

Dehais F, Duprès A, Blum S, Drougard N, Scannella S, Roy R, Lotte F. 2019. monitoring pilot's mental workload using ERPs and spectral power with a six-dry-electrode EEG system in real flight conditions. Sensors. 19(6):1324. doi: 10.3390/s19061324.

Dimitrakopoulos GN, Kakkos I, Dai Z, Lim J, Desouza J, Bezerianos A, Sun Y. 2017. Task-independent mental workload classification based upon common multiband EEG cortical connectivity. IEEE Trans Neural Syst Rehabil Eng. 25(11):1940–1949. doi: 10.1109/TNSRE.2017.2701002.

Fan C, Hu J, Huang S, Peng Y, Kwong S. 2022. EEG-TNet: an end-to-end brain computer interface framework for mental workload estimation. Front Neurosci. 16:869522. doi: 10.3389/fnins.2022.869522.

Hart SG, Staveland LE. 1988. Development of NASA-TLX (task load index): results of empirical and theoretical research. Adv Psychol. 52:139–183.

Ji Z, Tang J, Wang Q, Xie X, Liu J, Yin Z. 2023. Cross-task cognitive workload recognition using a dynamic residual network with attention mechanism based on neurophysiological signals. Comput Methods Programs Biomed. 230:107352. doi: 10.1016/j.cmpb.2023.107352.

Jiao Z, Gao X, Wang Y, Li J, Xu X, Haojun L. 2018. Deep convolutional neural networks for mental load classification based on EEG data. Pattern Recognit. 76:582–595. doi: 10.1016/j.patcog.2017.12.002.

Kakkos I, Dimitrakopoulos GN, Gao L, Zhang Y, Qi P, Matsopoulos GK, Thakor N, Bezerianos A, Sun Y. 2019. Mental workload drives different reorganizations of functional cortical connectivity between 2D and 3D simulated flight experiments. IEEE Trans Neural Syst Rehabil Eng. 27(9):1704–1713. doi: 10.1109/TNSRE.2019.2930082.

Karacan SŞ, Saraoğlu HM, Kabay SC, Akdağ G, Keskinkılıç C, Tosun M. 2023. EEG-based mental workload estimation of multiple sclerosis patients. SIViP. 17(7):3293–3301. doi: 10.1007/s11760-023-02547-6.

Khanam F, Hossain AA, Ahmad M. 2023. Electroencephalogram-based cognitive load level classification using wavelet decomposition and support vector machine. Brain-Computer Interfaces. 10(1):1–15. doi: 10.1080/2326263X.2022.2109855.

Kingphai K, Moshfeghi Y. 2021. On EEG preprocessing role in deep learning effectiveness for mental workload classification. *Human Mental Workload: Models and Applications: 5th International Symposium, H-WORKLOAD 2021, Virtual Event, November 24–26, 2021, Proceedings 5*, pp. 81-98. doi: 10.1007/978-3-030-91408-0_6.

Korzeniewska A, Mańczak M, Kamiński M, Blinowska KJ, Kasicki S. 2003. Determination of information flow direction among brain structures by a modified directed transfer function (dDTF) method. J Neurosci Methods. 125(1-2):195–207. doi: 10.1016/s0165-0270(03)00052-9.

Kwak Y, Kong K, Song WJ, Min BK, Kim SE. 2020. Multilevel feature fusion with 3D convolutional neural network for EEG based workload estimation. IEEE Access. 8:16009–16021. doi: 10.1109/ACCESS.2020.2966834.

LeCun Y, Bengio Y, Hinton GE. 2015. Deep learning. Nature. 521(7553):436–444. doi: 10.1038/nature14539.

Lim W, Sourina O, Wang L. 2018. STEW: simultaneous task EEG workload dataset. IEEE Trans Neural Syst Rehabil Eng. 26(11):2106–2114. doi: 10.1109/TNSRE.2018.2872924.

Longo L, Wickens CD, Hancock G, Hancock PA. 2022. Human mental workload: A survey and a novel inclusive definition. Front Psychol. 13:883321. doi: 10.3389/fpsyg.2022.883321.

Marinescu AC, Sharples S, Ritchie AC, Sánchez López T, McDowell M, Morvan HP. 2018. Physiological parameter response to variation of mental workload. Hum Factors. 60(1):31–56. doi: 10.1177/0018720817733101.
Mazher M, Aziz AA, Malik AS, Amin HU. 2017. An EEG-based cognitive load assessment in multimedia learning using feature extraction and partial directed coherence. IEEE Access. 5:14819–14829. doi: 10.1109/ACCESS.2017.2731784.
Mingardi M, Pluchino P, Bacchin D, Rossato C, Gamberini L. 2020. Assessment of implicit and explicit measures of mental workload in working situations: implications for industry 4.0. Applied Sciences. 10(18):6416. doi: 10.3390/app10186416.
Mohanavelu K, Poonguzhali S, Janani A, Vinutha S. 2022. Machine learning-based approach for identifying mental workload of pilots. Biomed Signal Process Control. 75: 103623. doi: 10.1016/j.bspc.2022.103623.
Mohanty N, Pradhan M, Reddy AVN, Kumar S, Alkhayyat A. 2022. Integrated design of optimized weighted deep feature fusion strategies for skin lesion image classification. Cancers (Basel). 14(22):5716. doi: 10.3390/cancers14225716.
Mullen T. 2010. Source information flow toolbox (SIFT). Swartz center for computational neuroscience, California, San Diego.
Mullen TR, Kothe CAE, Chi YM, Ojeda A, Kerth T, Makeig S, Jung T-P, Cauwenberghs G. 2015. Real-time neuroimaging and cognitive monitoring using wearable dry EEG. IEEE Trans Biomed Eng. 62(11):2553–2567. Nov doi: 10.1109/TBME.2015.2481482.
Ouwehand K, Kroef AVD, Wong J, Paas F. 2021. Measuring cognitive load: are there more valid alternatives to Likert rating scales? Front Educ. 6:702616. doi: 10.3389/feduc.2021.702616.
Pandey V, Choudhary DK, Verma V, Sharma G, Singh R, Chandra S. 2020. Mental workload estimation using EEG. In 2020 Fifth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN). Bangalore: IEEE; 83–86. doi: 10.1109/ICRCICN50933.2020.9296150.
Puma S, Matton N, Paubel PV, Raufaste É, El-Yagoubi R. 2018. Using theta and alpha band power to assess cognitive workload in multitasking environments. Int J Psychophysiol. 123:111–120. doi: 10.1016/j.ijpsycho.2017.10.004.
Qiao W, Bi X. 2020. Ternary-task convolutional bidirectional neural turing machine for assessment of EEG-based cognitive workload. Biomed Signal Process Control. 57:101745. doi: 10.1016/j.bspc.2019.101745.
Raufi B, Longo L. 2022. An evaluation of the EEG alpha-to-theta and theta-to-alpha band ratios as indexes of mental workload. Front Neuroinform. 16:861967. doi: 10.3389/fninf.2022.861967.
Reid GB, Nygren TE. 1988. The subjective workload assessment technique: a scaling procedure for measuring mental workload. Adv Psychol. 52:185–218.
Roy RN, Charbonnier S, Campagne A, Bonnet S. 2016. Efficient mental workload estimation using task-independent EEG features. J Neural Eng. 13(2):26019. doi: 10.1088/1741-2560/13/2/026019.
Ruder S. 2016. An overview of gradient descent optimization algorithms. CoRR.
Schnotz W, Christian K. 2007. A reconsideration of cognitive load theory. Educ Psychol Rev. 19(4):469–508. doi: 10.1007/s10648-007-9053-4.
Sciaraffa N, Aricò P, Borghini G, Flumeri GD, Florio AD, Babiloni F. 2019. On the use of machine learning for EEG-based workload assessment: algorithms comparison in a realistic task. In: Human mental workload: models and applications. 170–185.
Sevcenko N, Ninaus M, Wortha F, Moeller K, Gerjets P. 2021. Measuring cognitive load using in-game metrics of a serious simulation game. Front Psychol. 12:572437. doi: 10.3389/fpsyg.2021.572437.
Shahabi MS, Shalbaf A, Nobakhsh B, Rostami R, Kazemi R. 2023. Attention-based convolutional recurrent deep neural networks for the prediction of response to repetitive transcranial magnetic stimulation for major depressive disorder. Int J Neural Syst. 33(2):2350007. doi: 10.1142/S0129065723500077.
Shahabi MS, Shalbaf A, Rostami R, Kazemi R. 2023. A convolutional recurrent neural network with attention for response prediction to repetitive transcranial magnetic stimulation in major depressive disorder. Sci Rep. 13(1): 10147. doi: 10.1038/s41598-023-35545-2.
Shahabi MS, Shalbaf A. 2023. Prediction of treatment outcome in major depressive disorder using ensemble of hybrid transfer learning and long short-term memory based on EEG signal. IEEE Trans Cogn Dev Syst. 15(3): 1279–1288. doi: 10.1109/TCDS.2022.3207350.
Shin J, von Lühmann A, Kim D-W, Mehnert J, Hwang h-j, Müller K-R. 2018. Simultaneous acquisition of EEG and NIRS during cognitive tasks for an open access dataset. Sci Data. 5(1):180003. doi: 10.1038/sdata.2018.3.
Siddhad G, Gupta A, Dogra D, Roy P. 2022. Efficacy of transformer networks for classification of raw EEG data. *arXiv preprint arXiv* arXiv. 2022:2202.05170.
Siddhad G, Gupta A, Dogra D, Roy P. 2022. Efficacy of transformer networks for classification of EEG data. Biomed Signal Process Control. 87:105488. doi: 10.1016/j.bspc.2023.105488.
So WKY, Wong SWH, Mak JN, Chan RHM, Emmanuel M. 2017. An evaluation of mental workload with frontal EEG. PLoS One. 12(4):e0174949. doi: 10.1371/journal.pone.0174949.
Sokolova M, Lapalme G. 2009. A systematic analysis of performance measures for classification tasks. Informat Process Manage. 45(4):427–437. doi: 10.1016/j.ipm.2009.03.002.
Taori T, Gupta SS, Gajre S, Manthalkar R. 2022. Cognitive workload classification: towards generalization through innovative pipeline interface using HMM. Biomed Signal Process Control. 78:104010. doi: 10.1016/j.bspc.2022.104010.
Walter C, Schmidt S, Rosenstiel W, Gerjets P, Bogdan M. 2013. Using cross-task classification for classifying workload levels in complex learning tasks. In: 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction. pp. 876–881.
Xie J, Xu G, Wang J, Li M, Han C, Jia Y. 2016. Effects of mental load and fatigue on steady-state evoked potential based brain computer interface tasks: a comparison of

periodic flickering and motion-reversal based visual attention. PLoS One. 11(9):e0163426. doi: 10.1371/journal.pone.0163426.

Yu K, Prasad I, Mir H, Thakor N, Al-Nashash H. 2015. Cognitive workload modulation through degraded visual stimuli: a single-trial EEG study. J Neural Eng. 12(4):46020. doi: 10.1088/1741-2560/12/4/046020.

Zarjam P, Epps J, Chen F. 2011a. Characterizing working memory load using EEG delta activity. In Proceedings of the 19th European Signal Processing Conference (EUSIPCO). August; pp. 1554–1558.

Zarjam P, Epps J, Chen F. 2011b. Spectral EEG features for evaluating cognitive load. In: International Conference of the IEEE IEEE Engineering in Medicine and Biology Society (EMBS), Boston, Massachusetts USA. August 30–September 3, pp. 3841–3844. doi: 10.1109/IEMBS.2011.6090954.

Zhang H, Ji H, Yu J, Li J, Jin L, Liu L, Bai Z, Ye C. 2023. Subject-independent EEG classification based on a hybrid neural network. Front Neurosci. 17:1124089. doi: 10.3389/fnins.2023.1124089.

Zhang P, Wang X, Chen J, You W, Zhang W. 2019. Spectral and temporal feature learning with two-stream neural networks for mental workload assessment. IEEE Trans Neural Syst Rehabil Eng. 27(6):1149–1159. doi: 10.1109/TNSRE.2019.2913400.

Zhang Y, Shen Y. 2019. Parallel mechanism of spectral feature-enhanced maps in EEG-based cognitive workload classification. Sensors. 19(4):808. doi: 10.3390/s19040808.

Zhou Y, Huang S, Xu Z, Wang P, Wu X, Zhang D. 2022. Cognitive workload recognition using EEG signals and machine learning: a review. IEEE Trans Cogn Dev Syst. 14(3):799–818. doi: 10.1109/TCDS.2021.3090217.

Zhu G, Zong F, Zhang H, Wei B, Liu F. 2021. Cognitive load during multitasking can be accurately assessed based on single channel electroencephalography using graph methods. IEEE Access. PP. 1-1. 9:33102–33109. doi: 10.1109/ACCESS.2021.3058271.