Wavelet Based Feature Extraction for Multi-Model Ensemble Approach for Mental Workload Classification Using EEG

Fiza Parveen[a] and Arnav Bhavsar[b]
Indian Institute of Technology Mandi, Mandi, India

Keywords: Mental Workload Classification, XGBoost, Hand Crafted Feature Extraction, Stationary Wavelet Transform, Ensemble Majority Voting.

Abstract: Mental workload is a crucial aspect of cognitive processing as it reflects how much of our working memory is engaged. Studying n-back tasks of varying complexity, has been a popular way to explore the relationship between mental workload and EEG patterns. However there is still scope of improvement in achieving good performance in such a mapping. In this work, we address the classification of EEG patterns corresponding to different n-back tasks. We use publicly available n-back dataset, comprising 0-back, 2-back, and 3-back tasks to represent low, medium, and high levels of mental workload, respectively. We use wavelet-based signal decomposition technique to compute multi-resolution representation having both time and frequency patterns. This is followed by extracting a variety of hand crafted feature. We train different XGBoost models for two level and three level mental workload classification. Furthermore, we employ ensemble techniques at different levels to better categorize EEG signals. Our approach also involves finding channels that are most significant for classification of highly complex 2-back and 3-back task EEG data.

1 INTRODUCTION

Mental workload analysis is a key consideration in understanding human interaction with tasks and activities, indicating how much of our working memory is engaged when processing new information. Cognitive tasks like mathematical problem-solving, memory testing, and simulated real-world tasks contribute to determining cognitive load by assessing processing speed, working memory, attention, and ability to manage demanding tasks. Complex tasks need more cognitive effort and resources, which leads to a higher perceived workload (Bläsing and Bornewasser, 2021).

Mental workload assessment is often done using subjective and adhoc approaches, such as questionnaires and performance indicators, or by using neuronal electrophysiological activity signatures. Techniques like electroencephalography (EEG), electrocorticography (ECoG), and intracortical neuron recording are widely used. EEG is risk-free, affordable, and capable of monitoring significant neural activity over the entire cerebral cortex. Common cognitive assessments include n-back, visual search, and simultaneous capacity (SIMKAP), etc. (Leon-Dominguez et al., 2015), (Pang et al., 2020). N-back tasks involve remembering a certain number of previous stimuli, with higher values of ‘n’ implying greater complexity. Measuring N-back tasks is useful for experimental WM research, especially when dealing with higher cognitive demands, such as fluid intelligence, under high workload conditions (Miller et al., 2009).

In this work, we use publicly available EEG data (Shin et al., 2018), acquired during n-back tasks of varying difficulty to understand how the human brain responds and exhibits varied EEG patterns in various cognitive contexts. Classifying EEG signals for different mental workload conditions is a complex task due to non-stationary and time-varying characteristics. Further, a variety of domains for feature extraction methods may be considered such as time, frequency, and time-frequency domain features (Sharmila and Geethanjali, 2020). Also, different signal decomposition algorithms, like EMD and wavelets, can be used to split signals into multiple components. The extracted features are used in machine learning models, such as CNNs, SVM, XGBoost etc., which are used to classify the EEG signals (Amin et al., 2015). Thus, overall the proposed study

[a] https://orcid.org/0009-0003-2672-4656
[b] https://orcid.org/0000-0003-2849-4375

Parveen, F. and Bhavsar, A.
Wavelet Based Feature Extraction for Multi-Model Ensemble Approach for Mental Workload Classification Using EEG.
DOI: 10.5220/0012381300003657

In *Proceedings of the 17th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2024) - Volume 1*, pages 770-777
ISBN: 978-989-758-688-0; ISSN: 2184-4305

contribute in the following way:

- Our approach stands out due to its realism, presenting results at the subject level rather than at the smaller segmented windows (which is often followed in many EEG studies). Also we provided a generalized workflow across various subjects and various sessions.
- We explore the application of multiple binary and tertiary classification models, each differentially tailored to accurately classify distinct mental workload levels in EEG signals. We device a number of ways in which mental workload among different workload levels can be classified such as 0-back vs 2-back, 0-back vs 3-back, 2-back vs 3-back, 0-back vs 2-back and 3-back combined, and 0-back vs 2-back vs 3-back.
- We use Stationary Wavelet Transform (SWT) to decompose complex EEG signals, followed by different combinations of handcrafted features, encompassing both time and frequency domain characteristics extracted after SWT.
- Our framework involves three diffedrent ways of majority voting for different stages/subtasks, which we term as differential ensembling.
- To enhance classification accuracy in one of the more challenging settings, we implement channel selection techniques, ensuring use of the most relevant EEG channels for a precise analysis of mental workload patterns.

2 RELATED WORK

As indicated earlier, mental workload assessment comes in different flavours with different tasks associated with task-specific datasets. In this section, we review the relevant literature that work on n-back task dataset (Shin et al., 2018) for mental workload classification methods and techniques, which to our knowledge, is rather limited.

In this study (Shao et al., 2021), pre-processed n-back data is decomposed into various components using ICEEMDAN to obtain IMFs. In each IMF the motif ratio of different scales are calculated individually. Parameters with statistical difference are selected through t-test. 927 features were selected across 0, 2 and 3 back task, used for training Bi-LSTM model for three class mental workload classification.

In (Salimi et al., 2019), preprocessed data is divided into small instances of 1.1 seconds, a spectogram is calculated for every channel. For each subject, 28 CNNs (1 for each channel) were trained and validated using 28 channel spectrograms. Validation accuracy of 28 models were used to rank the corresponding channels for each subject and the ensemble classifier, consisting consisted of five CNNs, performed the best.

In another study (Khanam et al., 2023), task wise EEG signal is decomposed using discrete wavelet transform (DWT). Furthermore, a support vector machine (SVM) is trained on DWT based features for 0-back vs 2-back + 3-back (binary) and 0-back vs 2-back vs 3-back (tertiary) classification.

Importantly, the studies mentioned above follow diverse experimental procedures in terms of how they use the data. In this work, we evaluate the model performance on 9-fold cross-validation dataset to support our claim of generalization over various sessions. We train a single model for all the subjects proving it to be more realistic approach. The channel selection is done on the basis of evaluation on 9-fold cross-validation dataset, making it more reliable.

3 DATA DESCRIPTION

26 healthy right-handed individuals took part in this study (Shin et al., 2018) aged 17 to 33, with an average age of 26.1 years. The study consisted of three sessions of the n-back task series, 0-back, 2-back, and 3-back tasks, presented in a balanced sequence. The tasks were presented in a counterbalanced sequence, which means they were organized as follows: 0→2→3→2→3→0→3→0→2. Each participant performed 9 n-back tasks in each session, starting with a 2-second instruction, followed by a 40-second task period and a 20-second rest period. Tasks began and ended with a 250-millisecond beep, a 250-millisecond stop, and a fixation cross during the rest interval. A random one-digit number was presented every 2 seconds, with 20 trials in each series, with targets appearing with a 30 chance and non-targets accounting for the remaining 70.

In the 0-back task, participants were asked to press a 'target' button or a 'non-target' button (related to the number 7 and 8 respectively) to ensure their involvement in an experiment. In 2-back and 3-back tasks, participants were asked to press the 'target' button if the displayed number matched the number presented two or three places within the sequence, or hit the 'non-target' button if no match was found. The fixation cross was presented for 20 seconds during the rest phase. Each participant performed 180 trials for each n-back task, which was the result of a combination of 20 trials in each of the 3 series and 3 sessions.

EEG data was collected using a multichannel

BrainAmp EEG amplifier from Brain Products GmbH in Gilching, Germany. The data was recorded at a frequency of 1,000 Hz, but then downsampled to 200 Hz. Thirty active EEG electrodes were secured using a stretchy cloth cap from EASYCAP GmbH in Herrsching am Ammersee, Germany, using the worldwide 10-5 system (Oostenveld and Praamstra, 2001). Fp1, Fp2, AFF5h, AFF6h, AFz, F1, F2, FC1, FC2, FC5, FC6, Cz, C3, C4, T7, T8, CP1, CP2, CP5, CP6, Pz, P3, P4, P7, P8, POz, O1, O2, TP9 (used as reference), and TP10 (used as ground) electrodes had been used. After then, the collected data was downsampled to 200 Hz.

4 METHODOLOGY

In this section, we discuss our approach in detail. We first cover the data pre-processing followed by the windowing details. In section 4.3, we discuss feature extraction techniques. We discuss our classification framework in section 4.4, and ensemble majority voting in section 4.5.

4.1 Pre-Processing

Data preprocessing is a key component in enhancing the quality of EEG signals. Noise and distortions in EEG data can have a considerable influence on the accuracy and reliability of analytical models. EEG signals have distinct properties in terms of frequencies, spatial patterns, and correlations with various brain states. Delta (0.5 to 4 Hz), theta (4 to 8 Hz), alpha (8 to 13 Hz), and beta (13 to 30 Hz) are common frequency bands, each correlating to various stages of brain activity, such as profound sleep or awake. We follow the subsequent steps to clean the data (Parveen and Bhavsar, 2023).

- Reference electrodes are often used to record EEG signals. We used average referencing in this study, which includes determining the mean value for each channel and subtracting it from all data points linked with that channel.
- EEG signals contain a variety of frequency components that reflect broad behavioral patterns in neurons. We used a 6th order Butterworth bandpass filter (an infinite impulse response or IIR filter) in the 1-60 Hz range to eliminate unwanted frequency components. This filter effectively removes low-frequency drifts and high-frequency noise, covering a range from delta to gamma frequencies. We also used a 50 Hz notch filter to mitigate line noise interference.
- EEG signals can contain artifacts, which are signals unrelated to brain activity, originating from sources like eye blinks or muscle movements. Removal of artifacts is a critical step in EEG analysis. In our study, we employed the ADJUST algorithm (Mognon et al., 2011), which implements independent component analysis (ICA) to separate EEG signals into independent components, each representing different sources of brain activity, some of which may contain artifacts.

4.2 Windowing

Following preprocessing, we applied sliding windows with a size of 800 samples and an overlap of 200 samples across the entire EEG signal length for all subjects. 37 windows with shape 28 x 800 are contributed by each subject. The motivation behind choosing 800 time samples is to capture the mental activity for all levels of n-back task. As per the experiment, every other stimuli appears 2 seconds after the first stimuli. By taking 4 seconds of data on account, our aim is to analyse the EEG pattern on how the subject proceeds with new information while retaining the previous knowledge. We get a total of 7696 training and 962 test windows for each of the three n-back tasks, comprising data from all sessions and all subjects.

4.3 Feature Extraction

In this study, we employed the SWT with a Daubechies 4 (db4) mother wavelet of order 3 to decompose windowed EEG signals into six wavelets, representing different levels of decomposition, providing a comprehensive view of both high and low-frequency components. SWT helps detect transient events and changes in brain activity, making it crucial for assessing mental workload using EEG signals. Time domain data captures temporal transients and spatial variations, while frequency domain features reveal spectrum patterns representing diverse cognitive states associated with varying levels of mental workload. Following SWT, we calculated various time and frequency domain features, including mean, standard deviation, skewness, kurtosis, and Hjorth mobility and complexity in order to obtain average amplitude, variability in signal values, asymmetry, shape of a probability distribution, unpredictability and mobility of EEG signals respectively (Safi and Safi, 2021). We also analyze the Power Spectral Density (PSD) within delta, theta, alpha, beta, and gamma frequency bands of EEG signals in order to trigger adaptive responses in real-time systems (Welch, 1967). The features are calculated on each window of 800 time samples belonging to a particular channel as mentioned in section 4.2. Computing the 6 time-domain features across all channels of every wavelet we get a total of 1008 time-domain features (6x28x6=1008). For frequency-domain features, we calculated PSD on 5 frequency bands as mentioned above on every channel of every wavelet and get a total of 840 features (5x28x6).

4.4 Model Description

As we employ a rich set of handcraftted features mentioned above, we employ the popular XGBoost (Extreme Gradient Boosting) framework for classification. It is an extremely effective and extensively used machine learning algorithm that is well-known for its performance in classification problems including a variety of features (Chen and Guestrin, 2016). It uses ensemble learning methods to develop a robust predictive model by combining the predictions of many decision trees. This ensemble strategy improves the accuracy and resilience of the model. It employs the gradient boosting approach, which reduces errors from prior models by altering the weights of individual data points. Over time, this iterative technique results in a more refined and accurate model.

It augments a feature relevance analysis, which aids in determining which features are most important in the classification process. In this study, we use Xgboost model for classification with different combinantion of parameters for different mental workload conditions. Table-1, shows parameter description for multiple 2 & 3 class classification models.

Table 1: XGBoost hyperparameters.

parameters	0vs2	0vs3	2vs3	0vs2+3
max depth	3	3	3	3
learning rate	0.1	0.1	0.9	0.9
alpha	0.01	0.01	0.01	0.01
lambda	1.0	1.0	1.0	1.0

4.5 Ensemble Majority Voting

In this work, we use three different majority voting techniques specific to task requirement. Majority voting on windows is applied on all binary and tertiary classification strategies to get predictions at subject level. Ensemble majority voting on channel-specific models is used for 2-back vs 3-back classification, along with majority voting on windows, as shown in Figure-1. For three class classification, we apply majority voting on different labels predictions for each subject after performing majority voting on windows, as shown in Figure-2 .

For binary classification task 2-back vs 3-back, 28 individual models were trained, each corresponding to one of the 28 EEG channels. Channel-specific models that demonstrated promising performance were selected, and majority voting on windows was applied solely to predictions from these models. Ensemble majority voting is employed for multi-class classification using binary classifiers. For classification between 0-back, 2-back, and 3-back, three separate models were used, with test windows fed to these models. For any odd window there is a possibility that it get label 0 on model-1, label 2 on model-2 and label 1 on model 3. In such condition it is difficult to decide the class label as different model predicted different label. To avoid that, firstly we applied majority voting on windows to get the most frequent labels on a set of 37 windows belonging to each individual and following that we apply majority on class labels.

5 EXPERIMENTS

In this study, we analyze data on three mental workload levels: 0-back, 2-back, and 3-back, representing low, medium, and high mental workload levels. The data is organized into small 4-second windows, creating a structured data format of dimensions 26x9x28x800 for each task. The data is partitioned into training and testing datasets in an 9:1 ratio, with eight out of 9 series assigned to training and one series reserved for testing. The EEG data underwent sliding window processing, resulting in 7696 windows for training and 962 windows for testing for each mental workload class.

5.1 Binary Classification

5.1.1 0-back vs 2-back

For 0-back vs 2-back data, we decomposed the training and test windows into 6 wavelets using 3rd level SWT. Following that, we calculated time-domain features as mentioned in section 4.3. The best results were achieved with mean, standard deviation, skewness, kurtosis, hjorth mobility and complexity on each channel of each sample for all the wavelets. We then train XGBoost model on training samples using the extracted 1008 time-domain features and predict class labels on 9-fold cross-validation test samples. We employ majority voting on windows to categorise mental workload levels based on the most frequently assigned category across all windows.

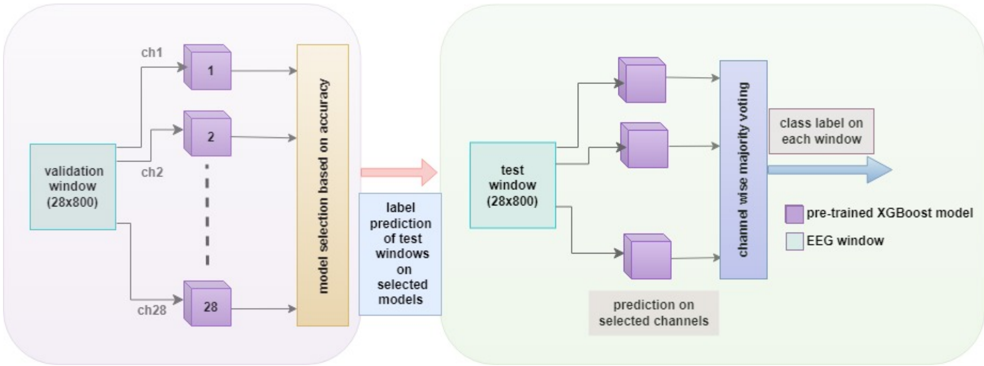

Figure 1: Ensemble model for 2-back vs 3-back classification.

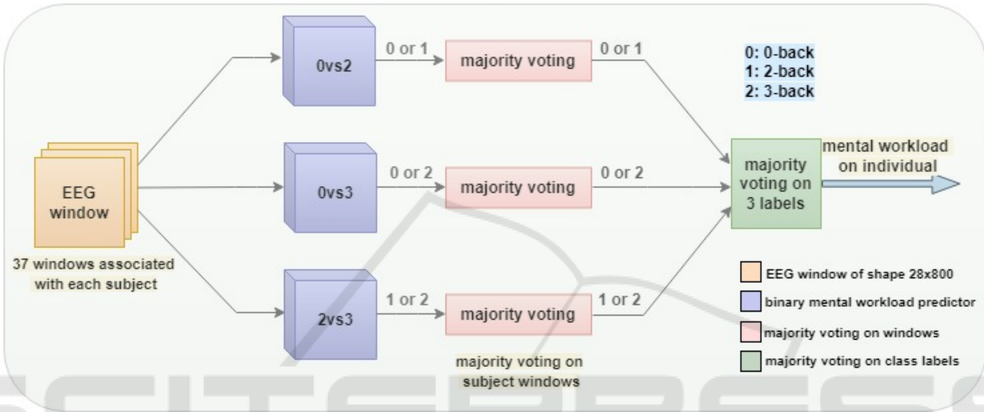

Figure 2: Workflow of tertiary classifier using various binary classifiers.

5.1.2 0-back vs 3-back

In the classification task of 0-back and 3-back EEG windows, following the SWT technique, we obtain the best results with Power Spectral Density (PSD) features of the input data using Welch's method (Welch, 1967) on different frequency bands namely delta (0.5Hz-4Hz), theta (4Hz-8Hz), alpha (8Hz-13Hz), beta (13Hz-30Hz) and gamma (30Hz- 45Hz) on each channel of every sample for all the 6 wavelets. As above, to classify the mental workload levels, we apply a majority voting approach to the windows, for assigning label to a complete signal.

5.1.3 2-back vs 3-back

The 2-back vs 3-back task presents a challenge due to the similarity of EEG channels, making it difficult to perform classification using a single XGBoost model. To address this, we adopted a greedy fine-grained approach, training 28 distinct XGBoost models, each dedicated to one of the 28 EEG channels. These models were rigorously validated using a separate validation dataset. To ensure model reliability, a benchmark of over 60% accuracy across all 9 cross-validation datasets was established. The training data was further subdivided into training and validation sets, allocating 6 out of 8 series to training and 2 to validation. Channel-specific models with over 60% accuracy on all cross-validation datasets were selected to predict labels on test data. For the 2-back versus 3-back task, the training data was divided into 5772 training and 1924 validation windows. A 3rd level SWT was performed on these windows, resulting in 6 decomposed wavelets. Time domain features were computed over every channel of each sample, normalized across all windows, and fed to the ensemble network.

5.1.4 0-back vs 2-back and 3-back Combined

A separate XGBoost model was created for the 0-back vs 2-back and 3-back (combined) classification using time domain features. Feature selection was achieved using Principal Component Analysis (PCA), with the top 50 features selected and fed into the XGBoost model for binary classification. To avoid the effect of class imbalance as 0 back has 7696 training windows and 2-back and 3-back combined has 15392 training windows, we employed class weight balance while training the model. We then fed 9-fold cross-validation test data to get the prediction on windows followed by majority voting on windows to get the label on subjects level.

5.2 Tertiary Classification

For the three-class mental workload classification, predictions were made for each test window using the 0-back vs 2-back, 0-back vs 3-back, and 2-back vs 3-back selected channel specific models. This process resulted in three labels per window, corresponding to each model. A majority voting approach was then applied to groups of 37 samples drawn from all 2886 windows spanning the three n-back tasks, generating 3x78 predictions. Here, 78 predictions represented all 26 subjects for each of the three n-back tasks. The most frequent label was selected among the three labels for each subject across all 78 predictions, ultimately yielding mental workload classifications at the subject level. We repeated this process for all 9 cross validation datasets.

6 RESULTS

Table-2 presents classification results for two and three class combinations, showing high accuracy across most cross validation sets, except for a few cases. We also provide median accuracy, in addition to mean, considering noisy signals and subjective variations in EEG data. The best accuracy for 0-back vs 2-back classification was 90.38% on the 6th cross validation dataset, with mean and median accuracies of 83.96 and 84.61, respectively. For two class classification (0-back vs 3-back) with low vs high mental workload, the best accuracy is 98.07% with mean and median accuracies of 88.44% and 88.45%, which is expected to be better than 0-back vs 2-back. After channel-specific model selection for 2-back vs 3-back, mean and median accuracies are 81.83% and 84.61%, respectively.

Table-3 shows comparision of result with and without this ensemble approach. In Table-3, we compare the results on the single XGBoost model for all the channels, with the proposed ensemble model. The mean and median accuracies show a great improvement from 61.52% and 67.30% to 81.83% and 84.61% respectively. This clearly shows the usefulness of the channel-specific model selection.

Table-4 shows the frequency of channel specific model performing well on the 9-fold cross validation dataset. Several combinations of channels were considered as shown in Table-5, including those that appeared more than three times, those that appeared more than four times, and those that appeared more than five times. Furthermore, we explored a specific combination involving channels 4, 8, and 11 which are associated with channels AFz, FC2 and Cz respectively, each of which had appeared more than five times.

We tried another combination of two class classification in which we use 0-back as one label and combined data from 2-back and 3-back as another label. On this particular classification model, the best accuracy came to be 93.58% . The mean and median accuracies are 87.314% and 87.17% respectively.

For three class classification, that is 0-back vs 2-back vs 3-back, we use combination of 0-back vs 2-back model, 0-back vs 3-back model and selected channel specific 2-back vs 3-back models corresponding to channel 4, 8 and 11. The best mean and median accuracies are 75.90% and 79.48% respectively.

7 DISCUSSION AND COMPARISONS

We being this section by noting the following salient points on the proposed framework.

- We have provided a generalized workflow across various subjects and various sessions.
- The results are at signal level rather than on EEG segments which seems a more realistic and practical approach.
- We explore differentially tailored multiple binary and tertiary classification models to accurately classify different mental workload levels.
- We explore hand crafted time-domain and frequency-domain features on SWT to get the most relevant features for classification.
- We devised three different majority voting frameworks for different stages of classification.
- Our model is tailored to encompass all subjects, eliminating the need to train individual models for each subject. This approach not only optimizes efficiency but showcase the robustness and generalization of our model.

As indicated in section 2, the methods in (Shao et al., 2021), (Salimi et al., 2019), (Khanam et al., 2023) have been reported on the n-back dataset (Shin et al., 2018) used in this work. However, none of these have followed a standard protocol for experimentation. Hence, in all fairness, a direct comparison of results cannot be done. Notwithstanding this, here we provide a discussion about different aspects of all the methods, including ours.

Table 2: Mental workload classification accuracy across different levels.

Cross-validation	0vs2	0vs3	2vs3	0vs2+3	0vs2vs3
set-1	86.50%	82.60%	100%	88.45%	82.00%
set-2	84.61%	94.23%	100%	84.61%	88.46%
set-3	76.92%	86.50%	98.07%	85.89%	84.60%
set-4	82.69%	88.45%	96.15%	84.61%	84.60%
set-5	88.46%	94.23%	78.84%	91.02%	79.48%
set-6	90.38%	86.53%	50.00%	93.58%	60.25%
set-7	84.61%	90.38%	84.61%	91.02%	79.40%
set-8	82.69%	98.07%	80.76%	87.17%	75.64%
set-9	78.80%	75.00%	48.07%	79.48%	48.71%
Mean	83.96%	88.44%	81.83%	87.314%	75.9%
Median	84.61%	88.45%	84.61%	87.17%	79.48%

Table 3: Comparision on classification accuracy of 2-back vs 3-back data before and after applying ensemble approach.

Cross-validation	single model for 28 channels	ensemble model with selected channels
set-1	100%	100%
set-2	92.30%	100%
set-3	71.15%	98.07%
set-4	44.23%	96.15%
set-5	50.00%	78.84%
set-6	05.70%	50.00%
set-7	67.30%	84.61%
set-8	82.69%	80.76%
set-9	40.38%	48.07%
Mean	61.52%	81.83%
Median	67.30%	84.61%

Table 4: Channels and number of times they resulted in higher than 60% accuracy on validation dataset.

Channel Name	No. of times the model associated with the channel has accu. $> 60\%$
0 (Fp1), 23 (P7), 25 (POz), 27 (O2)	3
7 (FC1), 13 (C4), 21 (P3)	4
8 (FC2), 26 (O1)	5
11(Cz)	6
4 (AFz)	9

In (Shao et al., 2021), the methodology lacks clarity on train-test splits, indicating a window-level classification. On the other hand, we have also used EEG windows, but via majority voting, the classification result is on level of complete signal, which seems a more realistic approach. (Salimi et al., 2019) share a comparable train-validation split, focusing solely on 0-back vs 2-back classification with channel selection. However, there is no clear evidence regarding the consistency of selected channels across all subjects and sessions. Our study enhances clarity, selecting channels based on 9-fold cross-validation dataset and then presenting results on separate test data, demonstrating consistency across all subjects and sessions. (Khanam et al., 2023) also has a lack of clarity about the training and test data split. While they include channel analysis, it is not conclusive in assessing the significance of channels across different trials. We provide results on most relevant channels performing well across all subjects and sessions.

8 CONCLUSION

In this study, we propose a machine learning approach for EEG based mental workload classification, wherein we employ several features based on wavelet based representation, followed by using ensemble models at various levels. Further, we demonstrate classification for various cases involving binary and multi-class scenarios, allowing for a more comprehensive understanding of mental workload.

Considering non-stationary nature of EEG, we divided the data into smaller windows. To aggregate the results at the subject level, we use a majority voting system, which synthesizes the outcomes from these smaller windows into an overall assessment of the subject's mental workload. To tackle the challenging 2 vs 3 back case, the ensemble technique that combine models specific to individual channels.

Our research has also revealed the significance of specific channels in high-level mental workload classification. These findings can shed light on the importance of certain physiological markers in understanding and predicting cognitive load.

Table 5: Accuracy on 2vs3 back task with different combination of channel specific models.

Cross validation	channels with frequency > 3	channels with frequency > 4	channels with frequency > 5	with channel 4,8and11
set-1	100%	100%	100%	100%
set-2	90.38%	59.60%	50.00%	100%
set-3	73.07%	75.00%	76.92%	98.07%
set-4	65.38%	94.23%	100%	96.15%
set-5	50.00%	48.07%	57.69%	78.84%
set-6	13.46%	50.00%	50.00%	50.00%
set-7	61.53%	86.50%	51.92%	84.61%
set-8	100%	73.07%	100%	80.76%
set-9	50.00%	50.00%	50.00%	48.07%
Mean	67.09%	70.71%	70.72%	81.83%
Median	65.38%	73.07%	57.69%	84.61%

REFERENCES

Amin, H. U., Malik, A. S., Ahmad, R. F., Badruddin, N., Kamel, N., Hussain, M., and Chooi, W.-T. (2015). Feature extraction and classification for eeg signals using wavelet transform and machine learning techniques. *Australasian physical & engineering sciences in medicine*, 38:139–149.

Bläsing, D. and Bornewasser, M. (2021). Influence of increasing task complexity and use of informational assistance systems on mental workload. *Brain Sciences*, 11(1):102.

Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794.

Khanam, F., Hossain, A. A., and Ahmad, M. (2023). Electroencephalogram-based cognitive load level classification using wavelet decomposition and support vector machine. *Brain-Computer Interfaces*, 10(1):1–15.

Leon-Dominguez, U., Martín-Rodríguez, J. F., and León-Carrión, J. (2015). Executive n-back tasks for the neuropsychological assessment of working memory. *Behavioural brain research*, 292:167–173.

Miller, K., Price, C., Okun, M., Montijo, H., and Bowers, D. (2009). Is the n-back task a valid neuropsychological measure for assessing working memory? *Archives of Clinical Neuropsychology*, 24(7):711–717.

Mognon, A., Jovicich, J., Bruzzone, L., and Buiatti, M. (2011). Adjust: An automatic eeg artifact detector based on the joint use of spatial and temporal features. *Psychophysiology*, 48(2):229–240.

Oostenveld, R. and Praamstra, P. (2001). The five percent electrode system for high-resolution eeg and erp measurements. *Clinical neurophysiology*, 112(4):713–719.

Pang, L., Fan, Y., Deng, Y., Wang, X., and Wang, T. (2020). Mental workload classification by eye movements in visual search tasks. In *2020 13th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, pages 29–33. IEEE.

Parveen, F. and Bhavsar, A. (2023). Attention based 1d-cnn for mental workload classification using eeg. In *Proceedings of the 16th International Conference on PErvasive Technologies Related to Assistive Environments*, pages 739–745.

Safi, M. S. and Safi, S. M. M. (2021). Early detection of alzheimer's disease from eeg signals using hjorth parameters. *Biomedical Signal Processing and Control*, 65:102338.

Salimi, N., Barlow, M., and Lakshika, E. (2019). Mental workload classification using short duration eeg data: an ensemble approach based on individual channels. In *2019 IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 393–398. IEEE.

Shao, S., Wang, T., Song, C., Su, Y., Wang, Y., and Yao, C. (2021). Fine-grained and multi-scale motif features for cross-subject mental workload assessment using bi-lstm. *Journal of Mechanics in Medicine and Biology*, 21(05):2140020.

Sharmila, A. and Geethanjali, P. (2020). Evaluation of time domain features on detection of epileptic seizure from eeg signals. *Health and Technology*, 10:711–722.

Shin, J., Von Lühmann, A., Kim, D.-W., Mehnert, J., Hwang, H.-J., and Müller, K.-R. (2018). Simultaneous acquisition of eeg and nirs during cognitive tasks for an open access dataset. *Scientific data*, 5(1):1–16.

Welch, P. (1967). The use of fast fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Transactions on audio and electroacoustics*, 15(2):70–73.