Wavelet-Augmented Self-Supervised Learning for Accurate Classification of Cognitive Workload

Nooshin Taheri Chatrudi[1], Will Clegern[2], Robert Hager[3], Lonnie Nelson[2], Hassan Ghasemzadeh[1]

***Abstract*—Wearable EEG (electroencephalogram) systems have demonstrated potential in epilepsy monitoring, sleep assessment, and determining cognitive workload to improve human decision-making. However, analyzing EEG signals is challenging due to their non-stationary nature and susceptibility to noise. In particular, achieving high accuracy in machine learning tasks requires large amounts of labeled data, which is difficult to obtain due to the time-consuming and labor-intensive nature of data labeling. To address these challenges, we propose a self-supervised learning (SSL) approach for cognitive workload classification using wavelet-based augmentations of EEG signals. First, two augmentations per channel are generated, and their wavelets are computed. The visual representations of these wavelets are then fed to the SSL pretext phase as contrastive pairs to pre-train the model. Finally, the pre-trained model is fine-tuned for workload classification using small amounts of labeled EEG data. Experimental results on the EEG During Mental Arithmetic Tasks (EEGMAT) dataset show that our method outperforms the state-of-the-art supervised models. Notably, our model achieves an accuracy of 99.5% with only 50% of the labeled data, demonstrating the effectiveness of our approach in scenarios with limited labeled data availability. Furthermore, the proposed approach achieves an accuracy of 98.6% in the leave-one-subject-out analysis.**

***Index Terms*—EEG classification, cognitive workload, self-supervised learning, contrastive learning, wavelet transform**

I. Introduction

Electroencephalography (EEG) is a non-invasive, affordable method that measures the brain's electrical activity using electrodes on the scalp. EEGs are able to track in-home monitoring of brain activity in real time. Non-invasive EEG systems are used to study various brain functions, such as cognitive workload classification [1]. Cognitive workload is defined as the amount of mental resources needed by a person to perform a cognitive task, which can determine the human workload capacity. Overloading capacity workload can negatively impact the performance and productivity of individuals [2].

Extracting robust and relevant features [3], particularly in wearable computing where real-time monitoring is critical, is challenging because the EEG signals are generally noisy and non-stationary [4]. Machine learning methods have been employed to identify discriminative features that represent intrinsic data patterns [5], [6]. Prior research proposed a method for cognitive performance detection using entropy-based features extracted from EEG signals, employing Support Vector Machine (SVM) and KNN classifiers [7]. Moreover, deep learning models such as convolutional neural networks (CNN), recurrent neural networks (RNN), and deep belief networks (DBN) have been proposed in EEG classification tasks [8]. Additionally, a shared spatial map network (SSMN) that abstracts EEG data representations from different individuals and electrode layouts is proposed to extract features and use deep learning techniques to classify mental workload [9].

Labeling EEG recordings is challenging due to the complicated nature of brain processes, the need for human experts, and the presence of noise, such as human movements and eye blinks. Moreover, combining various datasets is infeasible because the experimental setups across various studies are different [10]. To address these issues, a self-supervised learning (SSL) method is proposed to learn representations of the input data from unlabeled data in an unsupervised manner. Unlike supervised learning, which depends on large volumes of labeled data, and unsupervised learning, which identifies patterns without labels, SSL creates artificial labels from the data itself. Therefore, SSL could be a promising approach in wearable systems where data labeling is costly [11].

The process of designing an SSL model contains two phases, including a pretext task and a downstream task. In the pretext task, a model is trained to extract meaningful representations from unlabeled data. In the downstream task, the pre-trained model is used to classify the dataset with insufficient labeled data [10]. Contrastive learning is a well-known method for designing self-supervised learning in which augmented input data are fed into the model to identify the differences between input pairs. Hence, it reveals general and robust features across different tasks [12].

One of the prominent contrastive learning methods is SimCLR (Simple Framework for Contrastive Learning of Visual Representations) [13]. SimCLR, proposed specifically for images, improves the learning process by maximizing the similarity between different augmentations of the same data sample using a contrastive loss. Mohsenvand *et al.* [11] modified the SimCLR method to adapt to EEG signals. A contrastive learning method was presented to train a channel-wise feature extractor for learning representations from EEG signals. Furthermore, the authors evaluated a set of augmentation techniques specifically for EEG data and assessed their effectiveness on a classification task.

In our proposed method, we first develop two augmentation techniques for the EEG recordings of each channel, followed by computing the scalograms of the augmented signals. Scalograms can be viewed as 2D image representations of continuous wavelet transform (CWT). CWT is a critical

[1]Arizona State University, Phoenix, AZ, USA
[2]Washington State University, Pullman, WA, USA
[3]Preveal Technologies, Spokane, WA
emails:{ntaheric,Hassan.Ghasemzadeh}@asu.edu,
{will.clegern,lonnie.nelson}@wsu.edu, rnhager@comcast.net

method that effectively extracts features from non-stationary signals [14]. To the best of our knowledge, no prior work has developed SSL for EEG wavelet-based classification. Our aim in the pretext phase is to maximize the similarity between these two wavelet images for each channel using contrastive loss to extract robust features of the EEG data. We use the representations extracted from the pretext phase for specific classification tasks in the downstream task. The main contributions of this paper are as follows: (1)We propose a novel wavelet-based augmentation for use in self-supervised learning for EEG signal classification by creating two signal augmentations and computing their scalograms; (2) We introduce a method to learn feature representations from EEG scalograms in a self-supervised way; (3) We use contrastive learning on scalograms to enhance classification accuracy while minimizing the need for large labeled datasets.

II. Proposed Method

The overview of our proposed method for classifying EEG signals based on SSL is shown in Fig. 1. We applied our proposed method on the public EEGMAT dataset [12], which is described in the sections III-A. As shown in Fig. 1, our proposed method consists of two steps, including a pretext task and a downstream task. These steps are described in Section II-A and Section II-B, respectively.

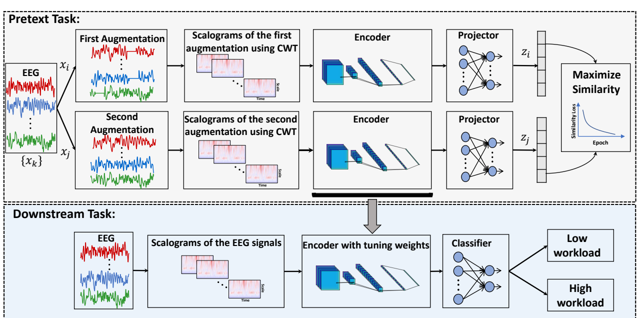

Fig. 1: Proposed approach consisting of pretext and downstream tasks.

A. Pretext

During the pretext task step, a robust representation of the EEG data is extracted for use in a downstream task. The objective of this step is to maximize the similarity between two augmentations of the same channel by utilizing a contrastive loss [15]. Each step of the pretext task is illustrated in Fig. 1 and explained in more details below.

1) EEG Augmentation: Augmentation plays a crucial role in contrastive learning as long as the meaningfulness of the EEG records is preserved. While augmenting images is perceivable to the human eye, the augmentation of EEG records requires expert verification. Mohsenvand et al. [11] assess several EEG record augmentation techniques within a specific range recommended by neurologists to maintain the physiological relevance of the records. Among all the techniques, zero-masking and amplitude scaling perform significantly better in extracting useful features for downstream tasks. Amplitude scaling adjusts the overall amplitude of the EEG signal, with recommended scaling between 0.5 to 2 times. Zero-masking involves replacing a segment of the EEG signal with zeros, recommended between 0 and 150 data points.

These two augmentation methods are employed in our study, and an example of each on the EEG data is shown in Fig. 2. During this step, each channel is analyzed individually, and two augmentations are computed for each channel. These two augmented data points are then considered positive pairs.

(a) Zero-masking (b) Amplitude-scaling

Fig. 2: An example of two augmentation methods on EEG.

2) Computing scalogram using continuous wavelet transform (CWT): A scalogram is a visual representation of the continuous wavelet transform. The Morlet wavelet with a scale of 129 is used in this study. After computing two augmentations for each channel, the scalograms of both the original and augmented data are computed as color images with a size of $3 \times 224 \times 224$.

3) Encoder: In this step, we employ a Convolutional Neural Network (CNN) to extract representation vectors from the scalogram images. The architecture of the designed CNN, as shown in Figure 3, consists of three Conv-2D layers followed by two dense layers. Consequently, by the end of this step, we have extracted 128 salient features from the images.

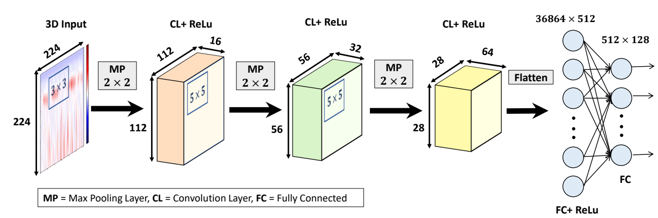

Fig. 3: The encoder Architecture. The kernel size for each layer is shown inside the boxes in blue.

4) Projector: The salient features of the encoder are fed to the projector to collapse them into 64 features. The projector architecture is shown in Figure 4, which consists of two dense layers with a ReLU activation function in between.

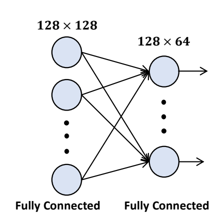

Fig. 4: The projector Architecture

5) Contrastive Loss for Pretext stage: Our method uses the NT-Xent (normalized temperature-scaled cross entropy) loss function from Chen et al. [13]. For a set $\{x_k\}$ with a positive pair x_i and x_j, the goal is to identify x_j within $\{x_k\}_{k \neq i}$ for a given x_i. Let z_i and z_j be the projector outputs for x_i and x_j. The NT-Xent loss for this pair is:

$$\ell_{i,j} = -\log\left(\frac{\exp(\text{sim}(z_i, z_j)/\tau)}{\sum_{k\neq i}^{2N}\exp(\text{sim}(z_i, z_k)/\tau)}\right)$$

where $\text{sim}(u, v)$ is the cosine similarity between u and v, and τ is a temperature scaling parameter. The final loss is the average $\ell_{i,j}$ over all positive pairs, considering (i, j) and (j, i).

B. Downstream

In the downstream task, we fine-tune the pre-trained model specifically for EEG classification. The encoder's weights, learned during the pretext task, are used to initialize the model in this phase. We replace the projector with a classifier, adjusting the output size of the final layer to match the number of classes; in this case, an output size of 2 for classifying EEG data as low workload or high workload. Then the model is fine-tuned to achieve high accuracy.

III. Experiments and Results

We compared our proposed method with that of two state-of-the-art classifier methods, [7] and [9], using the EEGMAT dataset in order to highlight the significance of our proposed method for providing better performance. In the following sections, we will describe the EEG dataset used, our experimental setups, and the outcome results.

A. Dataset

The publicly available dataset [12] includes EEG recordings, sampled at 500 Hz, from 36 participants before and during intense cognitive activity, using 19 channels. Participants first undertake a three-minute resting EEG with eyes closed, followed by a four-minute mental serial subtraction task. The dataset includes EEG records from the resting phase and the initial minute of the task. Data were filtered with high-pass (0.5 Hz), low-pass (45 Hz), and power line notch (50 Hz) filters, and artifacts were removed using Independent Component Analysis (ICA).

For our analysis, we selected the first 60 seconds of data from the unloaded trial to indicate a low mental workload(LMW) and the entire data from the task trial to indicate a high mental workload(HMW).

B. Experimental setups

Because the dataset consists of recordings from 36 subjects with 19 channels where each subject has two recordings corresponding to the resting and task trials, the dataset has a total of $1,368$ instances. The scalograms of the entire length of each data instance with the duration of 60 sec. were computed. These samples are split into 70%, 15%, and 15% of total samples for train, validation, and test sets, respectively, to use in our proposed SSL model. Augmentation is performed only on the training samples to avoid data leakage. The number of epochs is set to 50 for both the pretext and downstream tasks. The batch size is set to 10 for training data. We used the Adam optimizer with a learning rate of 3×10^{-4} for the pretext section and SGD with a learning rate of 1×10^{-3} and momentum of 0.9 for the downstream task. The baseline model is made up of the encoder and the classifier. The performance metrics for evaluating the models' results include Accuracy, Precision, Recall, and F1_Score.

C. Results and Discussion

Table I presents the performance of state-of-the-art methods and our proposed methods in terms of test accuracy and F1-score. The best performances of the methods proposed by Sharma *et al.* [7] and SSMN [9] on the EEGMAT dataset are shown in this table. We implemented 10-fold cross-validation for our baseline model, as it is used in other methods. The results of our SSL model on the test dataset (15% of the total data) are shown in table I. We can conclude that the test accuracy of our baseline model outperforms the model of Sharma *et al.* by about 5% and significantly surpasses the SSMN method in both test accuracy and F1-score. Moreover, our SSL method achieves more than 99% in test accuracy and F1-score, highlighting its effectiveness in improving workload classification based on EEG data.

TABLE I: Performance comparison of state-of-the-art works classifiers with our proposed method

Reference	Classifier	Test accuracy	F1-score
Sharma *et al.* [7]	SVM	94.00%	-
SSMN [9]	CNN	74.6%	74.6%
Our Baseline	CNN	98.98%	98.94%
Our SSL	CNN	99.52%	99.54%

Besides improving accuracy, the self-supervised learning method aims to offer better results even when there is a limited amount of labeled data. To investigate the performance of the proposed self-supervised learning method, we fine-tuned the model by leveraging different percentages of labeled data, assuming only this percentage of total data has been labeled. The results of performance metrics on test data when the SSL and baseline model are trained with different percentages of labeled data are shown in Figure 5. By noting this figure, we can conclude that the SSL model, compared with the baseline model, offers better performance in all terms of test accuracy, F1-Score, recall, and precision when the models are trained on different percentages of labeled data. We can see in Figure 5a that the SSL model achieves an accuracy of about 98% with only 20% of labeled data, while it reaches 99.52% accuracy when fine-tuning with the entire dataset. Moreover, the SSL model can achieve the same accuracy as the baseline model with a significantly lower percentage of label data, e.g., SSL achieves around 99% (see pointers ❶) with only 30% of labeled data, whereas the baseline model reaches this accuracy by entire labeled data (see pointer ❷). Furthermore, the accuracy of the baseline model dropped more quickly when the percentages of labeled data were less than 50% in comparison with the SSL model. These results highlight the effectiveness of our proposed SSL method in improving classification performance for EEG data, demonstrating its potential for real-world applications where labeled data is limited or hard to obtain.

Figure 6 shows the training and validation loss and accuracy of the SSL and baseline model over 50 epochs. The SSL

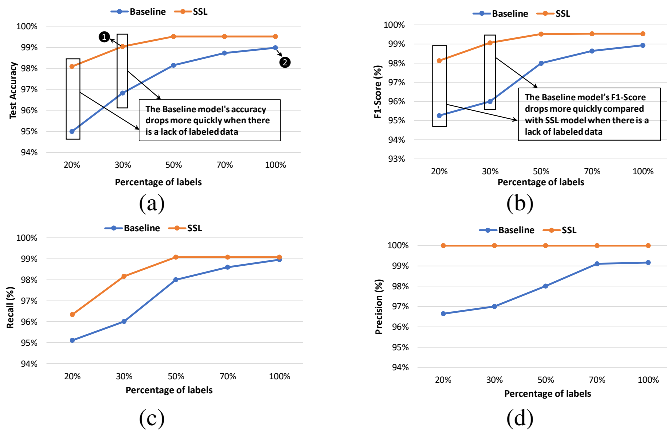

Fig. 5: Performance metrics for different percentages of labeled data: (a) Accuracy,(b) F1-score,(c) Recall,(d) Precision

model converged after around 18^{th} epochs, with training and validation accuracy approaching 100% and the losses converging, while the baseline model converged after 36^{th} epochs. This indicates that the SSL model learns more effectively and generalizes better to unseen data. Faster learning in SSL is due to its method of using unlabeled data to pre-train the model. So, the pretext phase acts like weight initialization, giving the model a better starting point and improving training accuracy.

We also evaluated the performance of our SSL-based technique for deployment in the real world to classify the mental workload of each subject based on the total channel recordings. For this purpose, we performed a leave-one-subject-out analysis, where data from each subject were iteratively considered test data, and the remaining data were used for training the model. Each subject has 19 channel recordings for two workloads: LMW and HMW. The model predicted the class for each channel as either LMW or HMW. A majority voting was then used to classify the mental workload of the subject. The overall accuracy was 98.6%.

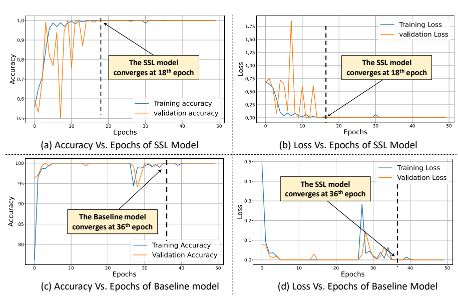

Fig. 6: Training and Validation Curves over epochs (a) and (c) Accuracy. (b) and (d) Loss.

IV. Conclusions

This paper presents a novel wavelet-based self-supervised learning technique for cognitive workload classification. Our method utilizes wavelet transforms and contrastive learning to enhance feature extraction. The proposed approach achieved a test accuracy of 99.52% and an F1-score of 99.54% on the EEGMAT dataset. This suggests that our method outperforms the state-of-the-art models on similar machine learning tasks. Moreover, the proposed SSL method achieves performance metrics (i.e., test accuracy, F1-score, recall, and precision) above 90% when the downstream model is trained using a small portion of the available training data. Hence, the performance of our model remains relatively high even with small amounts of labeled data. This observation indicates the potential of our approach for use in practical applications in wearable health monitoring and real-time workload detection systems where data collection and labeling are costly.

Acknowledgment

This work was supported in part by grant P30AG073107 from the National Institute on Aging, part of the National Institutes of Health.

References

[1] Y. Zhang and Y. Shen, "Parallel mechanism of spectral feature-enhanced maps in eeg-based cognitive workload classification," *Sensors*, vol. 19, no. 4, p. 808, 2019.

[2] Y. Zhou, S. Huang, Z. Xu, P. Wang, X. Wu, and D. Zhang, "Cognitive workload recognition using eeg signals and machine learning: A review," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 14, no. 3, pp. 799–818, 2021.

[3] R. Saeedi, B. Schimert, and H. Ghasemzadeh, "Cost-sensitive feature selection for on-body sensor localization," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, 2014, pp. 833–842.

[4] P. Zarjam, J. Epps, and N. H. Lovell, "Beyond subjective self-rating: Eeg signal classification of cognitive workload," *IEEE Transactions on Autonomous Mental Development*, vol. 7, no. 4, pp. 301–310, 2015.

[5] R. R. Azghan, N. C. Glodosky, R. K. Sah, C. Cuttler, R. McLaughlin, M. J. Cleveland, and H. Ghasemzadeh, "Personalized modeling and detection of moments of cannabis use in free-living environments," in *2023 IEEE 19th International Conference on Body Sensor Networks (BSN)*, 2023, pp. 1–4.

[6] A. Mamun, C.-C. Kuo, D. W. Britt, L. D. Devoe, M. I. Evans, H. Ghasemzadeh, and J. Klein-Seetharaman, "Neonatal risk modeling and prediction," in *2023 IEEE 19th International Conference on Body Sensor Networks (BSN)*. IEEE, 2023, pp. 1–4.

[7] L. D. Sharma, R. K. Saraswat, and R. K. Sunkaria, "Cognitive performance detection using entropy-based features and lead-specific approach," *Signal, Image and Video Processing*, vol. 15, no. 8, pp. 1821–1828, 2021.

[8] F. Movahedi, J. L. Coyle, and E. Sejdić, "Deep belief networks for electroencephalography: A review of recent contributions and future outlooks," *IEEE journal of biomedical and health informatics*, vol. 22, no. 3, pp. 642–652, 2017.

[9] Z. Yin, J. Tang, J. Zhang, Z. Wang, J. Zhang, Y. Wang, and B. Zhang, "Generic mental workload measurement using a shared spatial map network with different eeg channel layouts," *IEEE Transactions on Instrumentation and Measurement*, 2024.

[10] M. H. Rafiei, L. V. Gauthier, H. Adeli, and D. Takabi, "Self-supervised learning for electroencephalography," *IEEE Transactions on Neural Networks and Learning Systems*, 2022.

[11] M. N. Mohsenvand, M. R. Izadi, and P. Maes, "Contrastive representation learning for electroencephalogram classification," in *Machine Learning for Health*. PMLR, 2020, pp. 238–253.

[12] I. Zyma, S. Tukaev, I. Seleznov, K. Kiyono, A. Popov, M. Chernykh, and O. Shpenkov, "Electroencephalograms during mental arithmetic task performance," *Data*, vol. 4, no. 1, p. 14, 2019.

[13] H. Banville, O. Chehab, A. Hyvärinen, D.-A. Engemann, and A. Gramfort, "Uncovering the structure of clinical eeg signals with self-supervised learning," *Journal of Neural Engineering*, vol. 18, no. 4, p. 046020, 2021.

[14] S. S. P. Sadam and N. Nalini, "Epileptic seizure detection using scalogram-based hybrid cnn model on eeg signals," *Signal, Image and Video Processing*, vol. 18, no. 2, pp. 1577–1588, 2024.

[15] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *International conference on machine learning*. PMLR, 2020, pp. 1597–1607.