Title

A Novel Self-Supervised Learning Method for Sleep Staging and its Pilot Study on Patients with Disorder of Consciousness

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 46(0)

Authors

Li, Jingcong
Chen, Quanlin
Pan, Jiahui
et al.

Publication Date

2024

Peer reviewed

A Novel Self-Supervised Learning Method for Sleep Staging and its Pilot Study on Patients with Disorder of Consciousness

Jingcong Li (lijingcong@hotmail.com)

Quanlin Chen (2023024247@m.scnu.edu.cn)

Jiahui Pan (panjiahui@m.scnu.edu.cn)

Haiyun Huang (huanghaiyun@m.scnu.edu.cn)[1]

School of Software, South China Normal University, Wanjin Lu, Shishan Town, Nanhai District, Foshan, 528225, Guangdong Province, China

Abstract

Sleep staging holds significant importance in clinical medicine, aiding in the diagnosis of various disorders related to sleep and cognition. However, manually annotating a large amount of sleep data is time-consuming and labor-intensive, making it difficult to achieve. Efficiently utilizing these unannotated data poses a challenging task. We propose a novel self-supervised learning method with **T**emporal-**s**plit **C**ontrastive and **E**lectrode **A**utoencoder (TsC-EA) for sleep staging. We demonstrate that our method achieves state-of-the-art performance in self-supervised learning on SleepEDF and MASS-SS3. Moreover, experimental results indicate that our method can surpass the performance of supervised learning methods using only 10% of labeled data. Additionally, we explore the application of self-supervised learning in patients with disorder of consciousness. It can assist in diagnosing the severity of DoC through analysis of sleep staging. Staging the sleep patterns of patients with disorders of consciousness can help in diagnosing the severity of their condition.

Keywords: Self-supervised Learning; Sleep Staging; Disorder of Consciousness; Temporal-split Contrast; Autoencoder

Introduction

Sleep staging using EEG is crucial for diagnosing sleep disorders(Carskadon & Rechtschaffen, 2011). Specifically, it helps identify diseases by comparing sleep patterns in patients with disorders of consciousness (DoC) to those in healthy individuals(Pan et al., 2021). Traditional manual sleep staging by doctors is time-intensive. However, machine learning(Adnane, Jiang, & Yan, 2012), especially deep learning(Supratak et al., 2017), has revolutionized this process by automating it, though its success largely depends on access to large, well-annotated datasets. Acquiring such data in clinical environments is often difficult due to resource constraints and privacy laws, making the use of unlabeled data in automated systems a significant hurdle.

One of the limitations of deep learning methods is its reliance on large amounts of data with high-quality labels. Training a deep learning model with insufficient labeled data is a formidable challenge. In recent times, self-supervised learning has emerged as a promising approach to circumvent this limitation. This paradigm leverages unlabeled data to construct auxiliary tasks, facilitating the extraction of latent feature. After pre-training, only a small amount of labeled data is needed for fine-tuning to achieve good performance. Self-supervised learning, well-studied in computer vision(Chen et al., 2020) and natural language processing(Devlin et al., 2018), doesn't easily apply to sleep signal analysis. Moreover, certain data augmentation strategies that are universally accepted within the realm of image analysis may not be directly transferable to the task of sleep signal discrimination, as they could potentially compromise the integrity of the analytical process. The experiments conducted later in the article demonstrated that relying solely on data augmentation yielded unsatisfactory results.

To address the aforementioned challenges, we propose a novel self-supervised learning method for sleep staging. The contributions of this research are delineated as follows:

1) We propose a novel self-supervised learning method (termed Temporal-split Contrastive and Electrode Autoencoder, TsC-EA) to integrate the temporal and spatial features of sleep EEG signals.

2) Compared with state-of-the-art self-supervised learning approaches, the proposed method achieves better performance with accuracies even close to supervised learning in scenarios with lacking labeled data.

3) Based on the proposed self-supervised learning method, a pilot study on DoC patients is conducted to alleviate the lack of data in DoC patients.

Related Works

Self-supervised Learning

In self-supervised learning, contrastive learning methods like SimCLR(Chen et al., 2020) and MoCo(K. He et al., 2020) have marked a significant milestone, each with unique features such as momentum-based negative sample generation (K. He et al., 2020) or projection head that improve performance(Chen et al., 2020). These methods, which treat augmented versions of the same sample as positive pairs and different sample as negative pairs, have enhanced the quality of learned representations, especially in image analysis.

Alongside these contrastive learning, generative models have carved out their niche within self-supervised learning. The use of generative models for self-supervised training has garnered significant attention due to its simplicity and efficiency. While masked autoencoder like Context Encoders lag behind supervised methods(Pathak et al., 2016), Vision

[1]Corresponding author

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

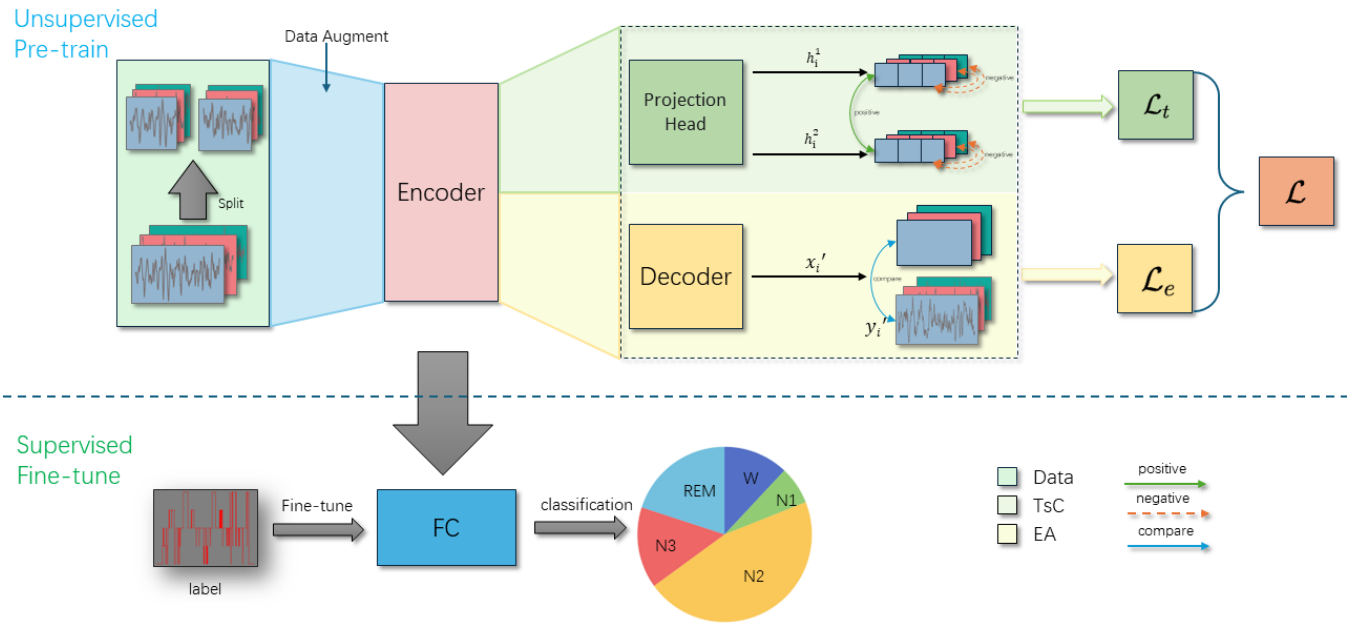

Figure 1: Architecture of the proposed TsC-EA.

Transformer (ViT) based autoencoder have seen success(Bao et al., 2021; K. He et al., 2022), inspired by language models like BERT and GPT(Devlin et al., 2018; Radford et al., 2019). The CIM technique uses the Beit model to create corrupted images for better autoencoder training, thereby facilitating the efficacious training of autoencoder within convolutional frameworks(Fang et al., 2022). The foregoing studies underscore the pivotal role of suitably designed generative tasks in bolstering self-supervised learning via autoencoder. Despite these advancements, research on using autoencoder for self-supervised learning in sleep staging is limited, with current models showing less effective pre-training (Banville et al., 2019; Xiao et al., 2021).

Self-supervised learning has achieved great success in the field of biosignal. For instance, the CLOCS framework(Kiyasseh, Zhu, & Clifton, 2021) employs the temporal stability inherent in cardiac signals to categorize distinct temporal segments and leads from an identical subject as positive pairs, while designating signals from disparate subjects as negative pairs. Similarly, the CLISA framework(Shen et al., 2023) identifies positive pairings predicated on the congruence of emotional responses elicited in different subjects while observing the same cinematic content. These methodologies exploit the intrinsic properties of biological signals to enhance the robustness of learned representations. If a method can be designed based on the biological properties of sleep EEG, it will effectively enhance the application of self-supervised learning in the field of sleep.

Self-supervised learning in sleep staging

Within the specialized domain of sleep stage classification, the application of self-supervised learning techniques remains comparatively underexplored. Two auxiliary tasks called relative positioning (RP) and temporal shuffling (TS) were proposed for self-supervised sleep staging(Banville et al., 2019), which based on the stability of sleep. Moreover, the SleepDPC framework (Xiao et al., 2021) adopts two approach, incorporating both predictive and discriminative tasks to facilitate training.

Notwithstanding the ingenuity of these methods, they are not without limitations. A significant reliance on the fine-tuning of hyperparameters is evident, or there persists a discernible performance discrepancy when benchmarked against fully supervised learning paradigms. Moreover, these techniques predominantly concentrate on the temporal dynamics of sleep patterns, frequently neglecting the spatial aspects in sleep stages. In contrast, our approach simultaneously takes into account both the temporal and the spatial feature, and achieve excellent performance.

Methods

To deal with discernible performance discrepancy in the existing self-supervised learning methods for sleep EEG staging, we propose a novel method. This section fully explains the parts of the proposed framework. Time and spatial features are extracted through modules Temporal-split Contrastive and Electrode Autoencoder. As depicted in the Figure 1, the initial step encompasses the application of a multitude of augmentative transformations to the dataset, the aug-

mented data will be processed by our two modules to extract temporal and spatial feature. Subsequently, a Temporal-split Contrastive module is integrated into the framework, tasked with the comparative analysis of the preceding and following halves of a data sample to reduce their distance. We further introduce the Electrode Autoencoder generation module, which is responsible for employing signals acquired from a electrode to extrapolate and synthesize corresponding signals for an another electrode, thereafter evaluating the similarity between the synthesized and authentic signals. The aforementioned auxiliary task is designed based on the temporal and electrode congruence of the sleep signals(Carskadon & Rechtschaffen, 2011; Berry et al., 2012) as shown in Figure 2.

Temporal-split Contrastive

We introduce the Temporal-split Contrastive module which is designed to investigate latent features associated with the temporal dynamics inherent in sleep data. Temporal-split Contrastive is more capable of finding the essence of sleep stage compared to methods relying on data augmentation. The prevailing classification system for sleep stages typically delineates a stage duration ranging from several minutes to tens of minutes, an attribute that has been extensively examined in the literature(Carskadon & Rechtschaffen, 2011). Our method involves independently encoding the anterior and posterior segments of a data sample, which are then projected through a nonlinear transformation mechanism termed the 'projection head'(Chen et al., 2020). Usually, instances derived from identical temporal segments of a sample are categorized as positive pairs, while disparate samples within the same batch are considered negative pairs for contrastive learning. Owing to the inherent stability of sleep patterns, it is probable that segments extracted from an individual sample correspond to an identical sleep phase. We employ the cosine similarity metric to ascertain the proximity between segments and apply the cross-entropy loss for optimization purposes.

For a given batch comprising N input samples, we bisect an individual sample into anterior and posterior segments, thereby generating contexts x_1 and x_2. Each segment is subjected to a distinct set of data augmentations. We then derive representations $z_1 = f(x_1)$ and $z_2 = f(x_2)$, where f denotes the encoding function. Subsequent to a nonlinear transformation executed by the projection head, we procure outputs h_1 and h_2 utilizing a multilayer perceptron (MLP). The tuple (h_1, h_2) is regarded as a positive pair, emblematic of the anterior and posterior segments of the same sample. The rest of the samples within the mini-batch, amounting to $(2N-2)$, are treated as negative pairs. This leads to the formulation of the context contrastive loss, which is designed to augment the similarity between a sample and its positive pairs while concurrently diminishing the similarity with negative pairs within the same mini-batch.

For a given context h_i^1, its similarity to h_i^2 is normalized by dividing it by the aggregate of its similarities to all $(2N-1)$ samples within the batch, which includes both positive and $(2N-2)$ negative samples. Then, the Temporal Contrastive Loss function $\mathcal{L}_t$, is articulated as follows:

$$\ell(h_i^1, h_i^2) = -log \frac{epx(sim(h_i^1, h_i^2)/\ \tau)}{\sum_{k=1}^{2N} \mathbb{1}_{[k \neq i]} epx(sim(h_i^1, h_k)/\ \tau)} \quad (1)$$

$$\mathcal{L}_t = \frac{1}{2N} \sum_{k=1}^{N} [\ell(2k-1, 2k) + \ell(2k, 2k-1)] \quad (2)$$

$$sim(u, v) = u^T v / ||u|| ||v|| \quad (3)$$

where $sim(\cdot)$ denotes the dot product between the pre-similarity scores, $\mathbb{1}$ is an indicator function to evaluate that $k \neq i$, τ is a temperature parameter, and h_k are negative pairs.

Electrode Autoencoder

Choosing challenging tasks to generate data that can aid in using autoencoder for self-supervised learning(K. He et al., 2022). In accordance with the guidelines delineated by the American Academy of Sleep Medicine (AASM), polysomnographic signals exhibit both commonalities and disparities across varied electrodes. Certain characteristics are ubiquitous across all electrodes, while some manifest more prominently within specific cerebral regions(Berry et al., 2012). Synthesizing the data frome another electrode is a challenging task for autoencoder generation.

In the proposed method, an Electrode Autoencoder Generation module is designed to enhance the model's capacity for discerning spatial discrepancies. This module employs a strategy that utilizes the simultaneous reconstruction of data from an alternative electrode to apprehend spatial characteristics. Data procured from a single electrode undergoes dimensional diminution via an Encoder, followed by the utilization of a Decoder for reconstruction. The Decoder's architecture comprises deconvolutional layers and upsampling processes. They form the autoencoder.

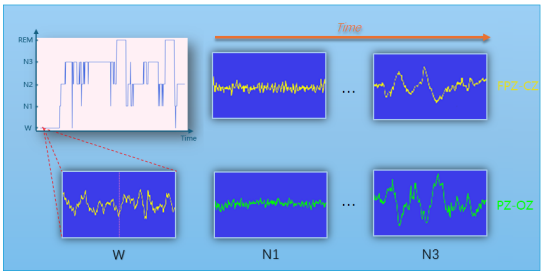

Figure 2: Examples of sleep EEG in different stages on electrodes FPZ-CZ and PZ-OZ. The commonalities between the preceding and following halves of a data sample. Within the same sleep stage, there are commonalities and variabilities in EEGs across different electrodes.

To distill latent spatial features inherent in sleep EEG, we incorporate a generative task in the Electrode Autoencoder module. Given a latent representation z, the autoencoder fabricates the corresponding representation of an alternate electrode through the Decoder. In this method, the augmented data x from a electrode is initially transformed to $z = f(x)$

to elicit feature extraction. Subsequently, z is reconstituted to yield $x' = g(z)$, where g signifies the Decoder function. The Mean Squared Error metric is employed to quantify the congruence between the generated sample and the contemporaneous sample from an alternate Electrode, which denoted as y.

$$\mathcal{L}_e = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - x_i'\right)^2 \tag{4}$$

Finally, after scaling both loss functions to the same order of magnitude, they are added together to obtain the total loss function. Once pretraining is complete, the Projection Head and Decoder are discarded, and fully connected layers are employed in their place for outputting classification results.

$$\mathcal{L} = \lambda_1 \mathcal{L}_t + \lambda_2 \mathcal{L}_e \tag{5}$$

In the formulas, the parameters λ_1 and λ_2 represent hyperparameters.

Experiment

We designed three sets of experiments to demonstrate the effectiveness of our method:

1) We conducted experiments using a small number of labels for training on both the Sleep-EDF and MASS-SS3 datasets.

2) We performed unsupervised training on the Sleep-EDF dataset and subsequently fine-tuned the model on the MASS-SS3 dataset using only a small number of labels.

3) We performed unsupervised training on the Sleep-EDF dataset and subsequently fine-tuned the model on the DoC dataset.

Datasets

1) **SleepEDF**: The SleepEDF dataset contains 41 polysomnographic (PSG) recordings from 20 healthy individuals, including EEG, EOG, EMG, and ECG signals at a 100Hz sampling rate for EEG, segmented into 30-second epochs(Kemp et al., 2000). It originally followed the R&K guidelines, distinguishing REM from NREM sleep, with NREM divided into stages N1, N2, N3, and N4. To conform to the updated AASM standards, we modified the dataset in line with Deepsleepnet's method(Supratak et al., 2017) by combining N3 and N4 into a single N3 stage and focusing on the 30 minutes directly preceding and following the sleep period, thereby aligning the data with contemporary AASM guidelines(Berry et al., 2012).

2) **MASS-SS3**: The Mass-SS3 dataset encompasses PSG recordings from 62 subjects, featuring EEG, EOG, EMG, and ECG signals. The EEG is sampled at a rate of 256Hz, with a page size of 30 seconds. The dataset annotations adhere to the AASM standards, categorizing sleep into REM and NREM stages. The NREM stage further includes N1, N2, and N3 sub-stages(O'reilly et al., 2014).

3) **DoC Dataset**: The DoC dataset encompasses overnight signals from seven subjects with DoC, including EEG signals. The EEG signals were sampled at a rate of 500 Hz with a page size of 30 seconds.Given that the criteria for sleep staging in patients with DoC are still under investigation(Pan et al., 2021), we adopted the AASM standards for staging, categorizing sleep into REM and NREM stages, with the latter comprising N1, N2, and N3 stages. By assessing the differences between the DoC in the same sleep stage and that of normal individuals, it can assist in determining whether a patient is in an Vegetative State (VS) or a Minimally Conscious State (MCS). The smaller the difference compared to normal individuals, the better the recovery tends to be.

Experiment Details

In the few-label(1%) training experiment 1), to align with the actual situation, we employed leave-one-subject-out (LOSO) cross validation approach. We initially conducted pre-training using the training set with labels removed. For consistency in validation and comparison, we utilized the CNN-1d as the backbone (Wang, Yan, & Oates, 2017). During the pre-training phase, we used the Adam optimizer with a weight decay of 1e-4, learning rate of 1e-4, and trained for 200 epochs. Upon completion of pre-training, we retained the weights and added a fully connected layer, fine-tuning with 1% of the training set. We used the Adam optimizer for fine-tuning with a weight decay of 1e-4 and a learning rate of 1e-4. We set exponential decay of 0.9 which was applied after 20 epochs, and training ceased when accuracy plateaued for 30 consecutive epochs. We calculated the mean and standard deviation of the results obtained from multiple experiments.

In the transfer learning experiment 2), we transfer the weights trained on Sleep-EDF to MASS-SS3. Here, we divided the Sleep-EDF dataset into 90% for training and 10% for testing. The training strategy mirrored that of the few-label training experiment. After training, we saved the weights. Likewise, the MASS-SS3 dataset was split into training (90%) and validation (10%) sets. We fine-tuned the model on 1% of labeled data from the training set, following the parameters set in the few-label training experiment. We repeated this fine-tuning process with five randomly chosen fixed seeds, and calculated the mean and standard deviation of the results obtained from multiple experiments.

In the experiment 3), the weights pre-trained on the Sleep-EDF dataset were loaded and transferred to classify DoC patients' sleep stages. In the experiments, we partitioned the local data using a five-fold cross-validation method. The training strategy was similar to the few-label training experiment. The results of five-fold cross-validations were averaged, and standard deviation was calculated.

The code was implemented using PyTorch 2.1 and trained on an NVIDIA GeForce RTX 4070 GPU.

Experimental Results and Discussion

To assess the effectiveness of our method, we compared it with the following approaches in the few-label training experiment.In the following method, SleepDPC employ 10% of labeled data. Other methods employ 1% of labeled data. CPC and TS-TCC have not been validated using the leave-one-subject-out cross-validation method(Eldele et al., 2023).

1) **Supervised Method**: The method of directly using 1% of labels for supervised training(Wang et al., 2017).

2) **CPC**:A self-supervised learning approach for prediction using autoregressive models(Oord, Li, & Vinyals, 2018). This method has been proven to perform well in sleep staging(Eldele et al., 2023).

3) **SimCLR**:A self-supervised Learning method based on data augmentation(Chen et al., 2020).

4) **TS-TCC**:A self-supervised Learning method integrating autoregressive model prediction and data augmentation contrast(Eldele et al., 2023).

5) **SleepDPC**:A self-supervised Learning approach for sleep stage classification using predictive and discriminative contrastive coding(Xiao et al., 2021).

As shown in Table 1, SimCLR exhibits suboptimal performance in the context of sleep, which underscores the inapplicability of many inductive biases from computer vision to sleep signal analysis. CPC and TS-TCC achieve the third and second best results. But CPC, TS-TCC, and SleepDPC focus predominantly on temporal features, overlooking spatial characteristics. This highlights the efficacy of our approach in capturing both temporal and spatial dimensions concurrently. These results indicate that our method outperforms the current state-of-the-art.

Table 1: Accuracies% of the proposed method and baselines in few-label (1%) experiments.

Method	SleepEDF	MASS-SS3
Supervised(Wang et al., 2017)	60.1±2.6	69.2±1.0
CPC(Oord et al., 2018)	74.7±0.2	-
SimCLR(Chen et al., 2020)	66.3±0.9	69.4±0.5
TS-TCC(Eldele et al., 2023)	75.8±0.3	-
SleepDPC(Xiao et al., 2021)	70.1±0.8	-
TsC-EA(Ours)	79.8±0.4	77.9±0.3

In the transfer learning experiment, we compared our method with (1) Supervised: Trained only on MASS-SS3, (2) SimCLR, (3) Transfer: models pretrained on Sleep-EDF and then fine-tuned on MASS-SS3. Our results are in Table 2. The results demonstrate that our method surpasses SimCLR and outperforms the transfer methods relying on labels.

Table 2: Accuracies% of the proposed method and baselines in transfer learning experiment (SleepEDF→MASS-SS3).

Method	Accuracy
Supervised(Wang et al., 2017)	67.5±0.7
Transfer(Wang et al., 2017)	70.6±0.7
SimCLR(Chen et al., 2020)	70.1±1.1
TsC-EA(Ours)	76.0±0.2

In the sleep staging experiment on DoC patients, our results are in Table 3. Due to the absence of sleep spindles and K-complexes, which are typically scarce in patients with DoC, we were compelled to classify stages lacking distinctive N1 or N3 features as N2 (Pan et al., 2021). The lack of electrooculography data further compromised the accuracy of our REM stage classification. We processed the data in two separate batches: one retaining all original data (Raw) and the other excluding the contentious N2 and REM stages (Processsed). In the first set of data, our method showed capacity to classify different sleep stages though with low accuracies. This phenomenon may be attributed to the ambiguity of the N2 staging criteria for DoC patients, leading to a significant difference between N2 stages in normal individuals and DoC patients.

Existing research indicates a lack of standardized N2 sleep stage criteria for DoC patients (Sebastiano et al., 2018), complicating feature learning and classification for models. Transfer learning and SimCLR underperformed compared to direct training in our experiments, likely due to the distinct sleep patterns of DoC patients, particularly those in a Vegetative State (VS), which differ markedly from healthy subjects (Pan et al., 2021). This discrepancy leads to a mismatch between the training and target data. In contrast, our method yields more adaptable and generalizable features.

Table 3: Accuracies% of the proposed method and baselines in the experiment with DoC

Method	Raw	Processed
Supervised(Wang et al., 2017)	60.3±0.1	61.2±0.8
Transfer(Wang et al., 2017)	58.5±0.1	57.0±0.7
SimCLR(Chen et al., 2020)	57.54±0.1	57.6±0.4
TsC-EA(Ours)	60.6±0.2	65.2±0.5

We explore the effect of label ratio on accuracy by using the SleepEDF dataset, and the results are shown in Figure 3. Our study shows that using just 5% of labeled data, our approach nearly matches the performance of fully supervised training. And with 10% labeled data, our approach exceeds fully supervised training. This underscores our approach's efficiency in using minimal labeled data to achieve robust learning. However, accuracy gains level off after 50% data labeling due to the limitations of our model, which doesn't incorporate additional sleep staging inputs like electrooculography

and electromyography. Future work will enhance our model with multimodal feature fusion and leverage larger datasets to boost accuracy further.

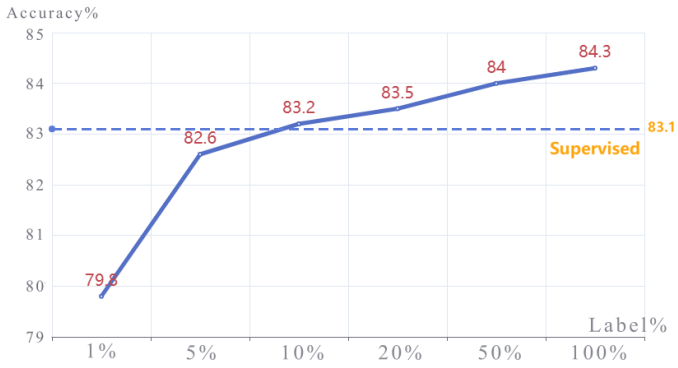

Figure 3: The effect of label ratio on accuracy.

The superior performance of our method in all the experiments underscores its efficacy and suggests its potential for achieving better results compared to existing approaches.

We visualized the severity of the patients and the performance of the algorithm using UMAP. Among them, healthy individuals came from the sleepEDF dataset, and MCS and VS from the DoC dataset. As shown in the Figure 4, when the subject is a normal person, the clustering effect is the best, with the wakefulness (W) period and other sleep cycles being most distinct, and the other sleep cycles also being somewhat separable. When the patient is assessed as MCS, the clustering effect is worse, but still shows some clustering, with wakefulness and other sleep cycles having some distinguishability; wakefulness is concentrated on the left side of the figure, while other sleep cycles are relatively to the right. When the patient is assessed as VS, there is almost no discernible clustering, with all points evenly scattered across the graph. This is consistent with related research; when the patient is assessed as MCS, features such as spindle waves and slow waves mostly disappear, but some sleep patterns are still maintained, therefore wakefulness and other sleep cycles still show some separability. When the patient is assessed as VS, they are considered to have possibly lost sleep patterns and circadian rhythm functions, retaining only sleep behavior, making it difficult to distinguish even between wakefulness and other sleep stages(Sebastiano et al., 2018; Pan et al., 2021).

Our work still has some limitations. Our exploration of applying self-supervised learning to sleep staging in patients with DoC is not yet comprehensive enough. Future research endeavors will concentrate on exploring domain adaptation and transfer to patients with DoC(Z. He, Zhong, & Pan, 2022; Pan et al., 2023), as well as on the aggregation of relevant data and the pursuit of further investigative efforts to drive progress in the field of medical automation.

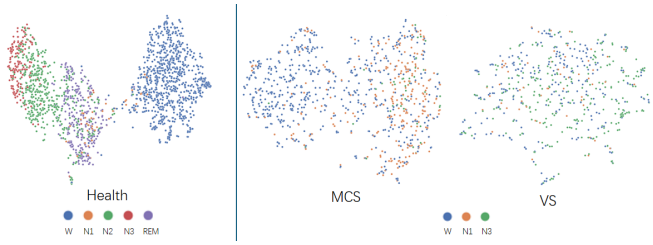

Figure 4: UMAP visualization of subjects from different types.

Conclusion

In this paper, we propose a novel framework for sleep self-supervised learning. The introduced framework learns by capturing both temporal and spatial variations and invariances in sleep signals. Experimental results demonstrate the efficacy of our framework, show its ability to learn meaningful features from sleep signals during the pre-training phase. We also conducted research on sleep staging in patients with DoC, demonstrating the practical value of our approach. By staging the sleep of patients with DoC, the severity of DoC can be diagnosed. In the self-supervised learning of sleep staging domains, our approach attains state-of-the-art results.

Acknowledgments

This work was supported by the STI 2030-Major Projects under grant 2022ZD0208900, National Natural Science Foundation of China under Grant 62306120, Guangdong Basic and Applied Basic Research Foundation under grant 2023A1515140100.

References

Adnane, M., Jiang, Z., & Yan, Z. (2012). Sleep–wake stages classification and sleep efficiency estimation using single-lead electrocardiogram. *Expert Systems with Applications*, *39*(1), 1401–1413.

Banville, H., Albuquerque, I., Hyvärinen, A., Moffat, G., Engemann, D.-A., & Gramfort, A. (2019). Self-supervised representation learning from electroencephalography signals. In *2019 IEEE 29th International Workshop on Machine Learning for Signal Processing (MLSP)* (pp. 1–6).

Bao, H., Dong, L., Piao, S., & Wei, F. (2021). Beit: Bert pre-training of image transformers. *arXiv preprint arXiv:2106.08254*.

Berry, R. B., Brooks, R., Gamaldo, C. E., Harding, S. M., Marcus, C., Vaughn, B. V., et al. (2012). The AASM manual for the scoring of sleep and associated events. *Rules, Terminology and Technical Specifications, Darien, Illinois, American Academy of Sleep Medicine*, *176*, 2012.

Carskadon, M. A., & Rechtschaffen, A. (2011). Monitoring and staging human sleep. *Principles and practice of sleep medicine*, *5*, 16–26.

Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020). A simple framework for contrastive learning of visual repre-

sentations. In *International conference on machine learning* (pp. 1597–1607).

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Eldele, E., Ragab, M., Chen, Z., Wu, M., Kwoh, C.-K., & Li, X. (2023). Self-Supervised Learning for Label- Efficient Sleep Stage Classification: A Comprehensive Evaluation. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *31*, 1333-1342. doi: 10.1109/TNSRE.2023.3245285

Eldele, E., Ragab, M., Chen, Z., Wu, M., Kwoh, C.-K., Li, X., & Guan, C. (2023). Self-Supervised Contrastive Representation Learning for Semi-Supervised Time-Series Classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *45*(12), 15604-15618. doi: 10.1109/TPAMI.2023.3308189

Fang, Y., Dong, L., Bao, H., Wang, X., & Wei, F. (2022). Corrupted image modeling for self-supervised visual pretraining. *arXiv preprint arXiv:2202.03382*.

He, K., Chen, X., Xie, S., Li, Y., Dollár, P., & Girshick, R. (2022). Masked autoencoders are scalable vision learners. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 16000–16009).

He, K., Fan, H., Wu, Y., Xie, S., & Girshick, R. (2020). Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 9729–9738).

He, Z., Zhong, Y., & Pan, J. (2022). An adversarial discriminative temporal convolutional network for EEG-based cross-domain emotion recognition. *Computers in biology and medicine*, *141*, 105048.

Kemp, B., Zwinderman, A., Tuk, B., Kamphuisen, H., & Oberye, J. (2000). Analysis of a sleep-dependent neuronal feedback loop: the slow-wave microcontinuity of the EEG. *IEEE Transactions on Biomedical Engineering*, *47*(9), 1185-1194. doi: 10.1109/10.867928

Kiyasseh, D., Zhu, T., & Clifton, D. A. (2021). Clocs: Contrastive learning of cardiac signals across space, time, and patients. In *International conference on machine learning* (pp. 5606–5615).

Oord, A. v. d., Li, Y., & Vinyals, O. (2018). Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*.

O'reilly, C., Gosselin, N., Carrier, J., & Nielsen, T. (2014). Montreal Archive of Sleep Studies: an open-access resource for instrument benchmarking and exploratory research. *Journal of sleep research*, *23*(6), 628–635.

Pan, J., Liang, R., He, Z., Li, J., Liang, Y., Zhou, X., ... Li, Y. (2023). ST-SCGNN: A Spatio-Temporal Self-Constructing Graph Neural Network for Cross-Subject EEG-Based Emotion Recognition and Consciousness Detection. *IEEE Journal of Biomedical and Health Informatics*, 1-12. doi: 10.1109/JBHI.2023.3335854

Pan, J., Wu, J., Liu, J., Wu, J., & Wang, F. (2021). A systematic review of sleep in patients with disorders of consciousness: from diagnosis to prognosis. *Brain Sciences*, *11*(8), 1072.

Pathak, D., Krahenbuhl, P., Donahue, J., Darrell, T., & Efros, A. A. (2016). Context encoders: Feature learning by inpainting. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2536–2544).

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I., et al. (2019). Language models are unsupervised multitask learners. *OpenAI blog*, *1*(8), 9.

Sebastiano, D. R., Visani, E., Panzica, F., Sattin, D., Bersano, A., Nigri, A., ... Franceschetti, S. (2018). Sleep patterns associated with the severity of impairment in a large cohort of patients with chronic disorders of consciousness. *Clinical Neurophysiology*, *129*(3), 687–693.

Shen, X., Liu, X., Hu, X., Zhang, D., & Song, S. (2023). Contrastive Learning of Subject-Invariant EEG Representations for Cross-Subject Emotion Recognition. *IEEE Transactions on Affective Computing*, *14*(3), 2496-2511. doi: 10.1109/TAFFC.2022.3164516

Supratak, A., Dong, H., Wu, C., & Guo, Y. (2017). DeepSleepNet: A model for automatic sleep stage scoring based on raw single-channel EEG. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *25*(11), 1998–2008.

Wang, Z., Yan, W., & Oates, T. (2017). Time series classification from scratch with deep neural networks: A strong baseline. In *2017 International joint conference on neural networks (IJCNN)* (pp. 1578–1585).

Xiao, Q., Wang, J., Ye, J., Zhang, H., Bu, Y., Zhang, Y., & Wu, H. (2021). Self-supervised learning for sleep stage classification with predictive and discriminative contrastive coding. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 1290–1294).