Multi-Tier Platform for Cognizing Massive Electroencephalogram

Zheng Chen[1][†] , **Lingwei Zhu**[2][†] , **Ziwei Yang**[2] and **Renyuan Zhang**[2][*]

[1]Osaka University, Japan
[2]Nara Institute of Science and Technology, Japan

chen.zheng.es@osaka-u.ac.jp, lingwei.andrew.zhu@gmail.com,
{yang.ziwei.ya3, rzhang}@is.naist.jp,

Abstract

An end-to-end platform assembling multiple tiers is built for precisely cognizing brain activities. Being fed massive electroencephalogram (EEG) data, the time-frequency spectrograms are conventionally projected into the episode-wise feature matrices (seen as tier-1). A spiking neural network (SNN) based tier is designed to distill the principle information in terms of spike-streams from the rare features, which maintains the temporal implication in the nature of EEGs. The proposed tier-3 transposes time- and space-domain of spike patterns from the SNN; and feeds the transposed pattern-matrices into an artificial neural network (ANN, Transformer specifically) known as tier-4, where a special spanning topology is proposed to match the two-dimensional input form. In this manner, cognition such as classification is conducted with high accuracy. For proof-of-concept, the sleep stage scoring problem is demonstrated by introducing multiple EEG datasets with the largest comprising 42,560 hours recorded from 5,793 subjects. From experiment results, our platform achieves the general cognition overall accuracy of 87% by leveraging sole EEG, which is 2% superior to the state-of-the-art. Moreover, our developed multitier methodology offers visible and graphical interpretations of the temporal characteristics of EEG by identifying the critical episodes, which is demanded in neurodynamics but hardly appears in conventional cognition scenarios.

1 Introduction

Studying neurophysiological processes is an important step toward understanding the brain. As a staple brain imaging tool, electroencephalography (EEG) reveals neuronal dynamics non-invasively with millisecond precision from the scalp surface [Sabbagh *et al.*, 2019; Varatharajah *et al.*, 2017; Hosseini *et al.*, 2021], rendering it important not only in fundamental research such as pathophysiology [Qu *et al.*, 2020]

---

† indicates joint first authors.
∗ corresponding author.

and psychiatry [Li *et al.*, 2017], but also in a variety of applications such as and Brain-Machine Interfaces (BCIs) [Collins and Frank, 2018]. The cognition of EEGs is a typical multi-tier task essentially, which is conducted by the capture, feature representation, coding, and pattern recognition etc. From the signal processing point of view, the EEG data is usually massive, redundant, and noisy [Hosseini *et al.*, 2021]. Moreover, the principle characteristics of any specific EEG is carried in the time domain. Above traits lead to the demands of proper methodologies in each phase (or say tier) for processing EEGs.

Yielding EEG signals cognizable by models [Sabbagh *et al.*, 2019], pre-processing the specific waveforms or graphoelements in both of time- or other domain (power of rhythms for instance, seen as the space-domain in the following context) is needed. Namely, the main aspiration of feature representation tier lies on projecting the massive and noisy signals onto a lower dimensional subspace with the hope to predominantly contain original neuronal signals, and then extracting the features related to different rhythms and discrete neurophysiological events in EEG [Varatharajah *et al.*, 2017]. Then, various applications are achieved by using the well pre-processed features through pattern recognition models [Phan *et al.*, 2019]. Although improving the spatio-temporal (ST) feature-representation technologies helps to increase the general quality of service (QoS) [Kayser *et al.*, 2009], the coordination and synthesis between pre-processing and pattern recognition tiers appear great impact on overall performances [Mijatovic *et al.*, 2021]. Glancing several image-processing like efforts of EEG cognitions, where the deep convolutional neural networks (CNNs) were implemented for recognizing ST patterns from EEGs, the classification of some brain activities is performed [Eldele *et al.*, 2021]. Unfortunately, the poor QoS indicates that EEGs cannot be treated as images since the temporal features dissipate in the plain artificial neural networks (ANNs). A reasonable substitution is employing inherently temporal based models such as recurrent neural networks (RNNs) to perform ST pattern recognitions [Pathak *et al.*, 2021]. However, the global-term perspective of RNN conflicts against the instantaneous-sensitivity of brain activities, which leads to the loss of cognition quality [Qu *et al.*, 2020]. Escaping from the RNNs and CNNs, high-end platforms are demanded.

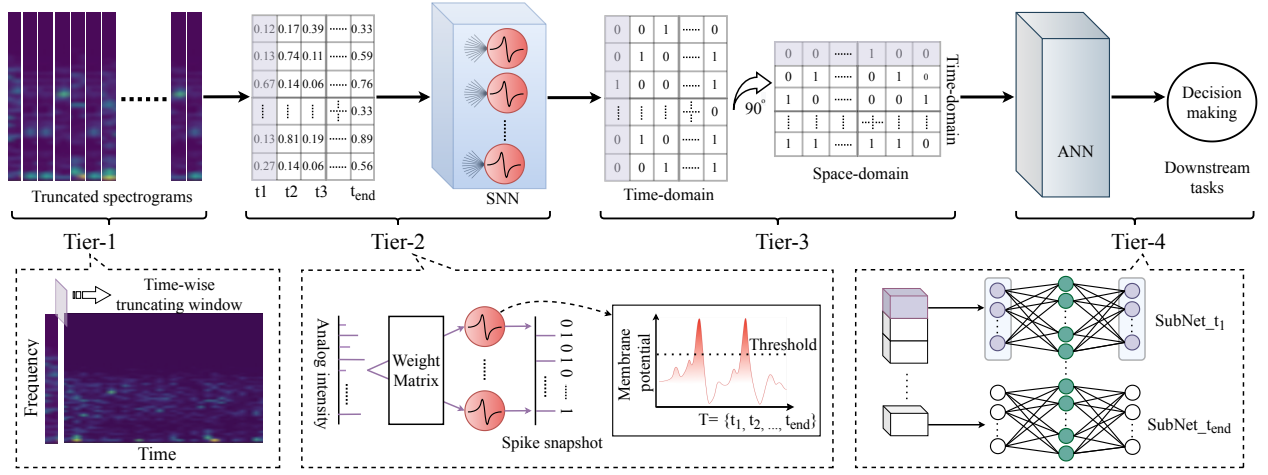In this work, all tiers of the EEG cognition are re-developed

Figure 1: System overview of the proposed platform

and coordinated as a general purpose EEG cognizing platform. A state-of-the-art (SOTA) EEG feature extractor is employed for translating specific EEG samples into ST feature matrices as tier-1. The proposed spiking neural network (SNN) based stage (seen as tier-2) codes ST features in the term of spike series. Through the transposing in tier-3, spike patterns are fed to a special analog neural network (ANN) for various recognition tasks, where the entire network is spanned to a set of sub-networks by the attention mechanism. The case study of sleep stage scoring problem is demonstrated by our platform for proof-of-concept. Introducing a variety of datasets with the largest comprising 42,560 hours recorded from 5,793 subjects, the proposed platform achieves the overall scoring precision of 87%, which is 2% superior to SOTA. In addition, various mutations of tier-variations are quantitatively and visibly evaluated to optimize the platform.

2 Proposed Method

The entire platform consists of four tiers as illustrated in Figure 1. The frontend tier-1 is a plain SOTA layer to extract the features of EEGs. A specific EEG is converted to a two-dimensional map (or say matrix), where the vertical and horizontal indicate the frequency band (read as "spatial" as above) and the episodes in time-sequence, respectively. The tier-2 is a pseudo-SNN for pre-coding the analog features into spike patterns. An episode from a specific EEG sample drives one single snapshot (known as time-step in SNN theory) of a spike pattern thru this pseudo-SNN. This mechanism totally differs from the conventional SNN, where one entire sample fans out multiple snapshots for recognition. Thus, the spiking-analog hybrid of federated learning is proposed over tier-2, tier-3, and tier-4 as following. In tier-3, the spike patterns are transposed and snapshot-wisely spanned into multiple sub-networks in tier-4. Finally, the analog attention model is employed for backend recognitions.

2.1 Tier-1: Time-Frequency Representation

To reveal information in both time- and space-domains, tier-1 of the proposed platform generates a conventional log-power spectrogram for each input EEG sample [Phan *et al.*, 2021]. The procedure of the spectrogram transformation first applies the Discrete Fourier Transform with a moving window to the input signal, which is given by:

$$X(k) = \sum_{n=0}^{N-1} x(n)\omega(n)e^{-\frac{2\pi i}{N}kn}, \qquad (1)$$

where N is the length of Hamming window function $\omega(n)$. k is the frequency. A log-power representation is then defined by:

$$S_{\log}(k) = \log(|X(k)|), \quad \forall t. \qquad (2)$$

Elements of the spectrogram represent the intensities at different frequency resolutions of corresponding time duration. Each spectrogram viewed as a time-frequency matrix is then normalised into a grey-scale intensity image scaled between $[0, 1]$:

$$S(k) = \frac{S_{\log}(k) - \min(S_{\log}(k))}{\max(S_{\log}(k) - \min(S_{\log}(k)))}. \qquad (3)$$

$S(k)$ is *cut* sequentially for every second to yield episodes named as truncated spectrograms $S(k,t)$, where t denotes the truncation for t-th second. The truncated spectrograms with analog intensity are then fed into the SNN tier-2 for further pre-coding the ST information.

2.2 Tier-2: Pseudo-Spiking Neural Network

To tackle the non-differentiable issue of SNN activation units, we follow the well-known iterative leaky integrate-and-fire (LIF) model [Wu *et al.*, 2019]. LIF can be succinctly described by the following relation by focusing on one timestep $t + 1$ and the $n + 1$-th layer of the network:

$$u_{t+1,n+1} = \tau u_{t,n+1}(1 - o_{t,n+1}) + \sum_j w_{j,n} o_{t+1,n}^j, \quad (4)$$

where the subscript t denotes t-th timestep and n denotes n-th layer within the architecture. Hence $u_{t+1,n+1}$ is the membrane potential of neurons at the $n + 1$-th layer and timestep

$t+1$. τ is the decay rate of membrane potential. o_{t+1}^j denotes the j-th firing spike with weight $w_{j,n}$ from the previous layer at time step $t+1$ Given the above information, $\sum_j w_j o_{t+1,n}^j$ is the pre-synaptic input and its information carrier is accumulated from the n-th layer. Hence, the neuron will output a fire spike and u_{t+1} is the reset to u_0 ($u_0 = 0$) ,when u_{t+1} reaches the firing threshold u_{th}:

$$o_{t+1,n+1} \begin{cases} 1 & \text{if } u_{t+1,n+1} > u_{th} \\ 0 & \text{otherwise} \end{cases} \qquad (5)$$

The spike fired by $o_{t+1,n+1}$ will propagate forward and activate next layer's neurons.

Conventionally, SNN iterates t times over each sample s to produce a time-indexed sequence of samples and errors $(s_1, \ldots, s_t)$, $(\epsilon_1, \ldots, \epsilon_t)$. The output sequence $(s_1, \ldots, s_t)$ is unsaved and typically only the average $\bar{s}$ is computed. The errors $(\epsilon_1, \ldots, \epsilon_t)$ are averaged to yield the *timestep-average* ϵ for backpropagating ϵ of sample s. However, one cannot backtrack individual contribution of each timestep from the average of all timesteps. Moreover, EEG data are inherently time-indexed, which renders such iterates ambiguous to apply. We propose a pseudo-SNN to handle the episode-wise truncated spectrograms $S(k,t) = (\tilde{s}_1, \ldots, \tilde{s}_t)$ that contain spatial information for every second, i.e. $\tilde{s}_t$ contains information from the t-th second. Every $\tilde{s}_i, 1 \le i \le t$ is then fed into the SNN layer for 1 rather than t times forward pass. The resultant snapshot z_i and error $\tilde{\epsilon}_i$ are appended to the sequence $(z_1, \ldots, z_i)$ and $(\tilde{\epsilon}_1, \ldots, \tilde{\epsilon}_i)$, respectively. Here, $\tilde{\epsilon}_i$ denotes error associated with $\tilde{s}_i$ rather than with i-th iteration of whole spectrogram s. The sequence of errors are averaged to yield *sample-average* error $\tilde{\epsilon}$ for backpropagation.

We argue that leveraging the truncated spectrograms has at least three advantages over the conventional method:

1. since EEG data are inherently time-indexed, timestep-average can be regarded as a special case of *sample-average* of truncated spectrograms by not performing truncation and increasing the number of forward passes.

2. spike snapshots allow for more efficient and finer-grind feature extraction, significantly boosting performance of downstream tasks.

3. pre-coding phase based on the inherent EEG time naturally models the cumulate/fire process of SNN neurons. We conjecture such similarity might be essential for improved feature distillation and performance.

The second and last advantages are demonstrated by our experimental results in Sections 4.1 and 4.2, respectively .

2.3 Tier-3: Preserving Spanning Topology

At tier-2 of the proposed architecture, we independently precode the analog features into a spike snapshot for each truncated spectrogram. The sequence of spike snapshots can be viewed as a matrix $Z = [z_1, \ldots, z_t]$ to be fed into the latter ANN layer for refining SNN spike features. Matrix Z preserves the spanning topology obtained from the SNN layer tier-2 and might benefit significantly the latter construction of feature subspaces, which stands as a sharp contrast to the conventional method that outputs an average vector $\bar{s}$. From an EEG perspective, maintaining Z helps *dynamically* capture and accumulate the ST characteristics. To better excavate information associated with snapshots, a transpose operation is required for latter processing in tier-4.

2.4 Tier-4: Attention For Truncated Spectrograms

Feature outputs from conventional SNN are fed into an ANN layer (or a subsequent softmax layer) for downstream tasks such as classification. However, to do so the feature output should be compressed, e.g., reshaping to a vector for input to a fully connected layer. Such compression would destroy the independence between snapshots. Hence, a network that can individually process and summarize snapshots is called for. Specifically, the recently popular attention-based mechanism excavates relevant relationship from input matrices [Dosovitskiy *et al.*, 2021]. As such, an attention layer naturally suits our platform for processing and tracking individual contribution from the sequence of truncated spectrograms, while each truncated spectrogram is accepted by a sub-network to maintain the feature space.

The input of tier-4 is the linear projection of transposed input Z^T onto higher dimensional feature spaces. The attention layer calculates the relevance of rows of Z^T and maps the relevance to the ground truth by three matrices: query Q, key K, and value V. Specifically, the relevance is computed as:

$$A = \sigma(\frac{QK^T}{\sqrt{d}}) \cdot V, \qquad (6)$$

where $\sigma(\cdot)$ denotes row-wise softmax function. $\sqrt{d}$ denotes a normalization-like scale that is applied to each Q-K computation. The result matrix A records the relevance score calculated as the weighted average of the rows of V, with the weights corresponding to the softmax probabilities. Motivated by the recent success in visual tasks, tier-4 comprises an addition task-based indicator, such as class-token referring to [Dosovitskiy *et al.*, 2021] for details, for further decision-making such as classification.

3 Experimental Setting

3.1 Datasets

As a proof-of-concept, we examine the efficacy of the proposed platform on the sleep stage scoring problem. Specifically, we compare the proposed platform against state-of-the-art algorithms on several authoritative datasets: (i) Sleep Heart Health Study (SHHS) Database; (ii) Sleep-EDF Database. The SHHS dataset is the largest public sleep dataset comprising 42,560 hours recorded from 5,793 subjects [Chen *et al.*, 2021].

3.2 Comparison

We compare comprehensively against state-of-the-art algorithms which were based on various architectures. We especially pay attention to the very recent work [Phan *et al.*, 2021] whose sleep stage classification results based on multi-source signals are considered as very close to human experts. It is also helpful to compare with [Eldele *et al.*, 2021; Fiorillo *et al.*, 2021; Qu *et al.*, 2020] who only exploited the EEG signal.

	Method		Wake	N1	N2	N3	REM	*Acc*
[Karimzadeh *et al.*, 2018]	*EEG + KNN*	*Pre*	0.89	0.55	0.75	0.84	0.86	
		Re	0.81	0.58	0.68	0.54	0.78	0.83
		F1	0.85	0.56	0.71	0.66	0.81	
[Eldele *et al.*, 2021]	*EEG + CNN*	*Pre*	0.90	0.31	**0.87**	0.87	0.80	
		Re	0.83	0.37	0.86	**0.87**	0.83	0.85
		F1	0.86	0.33	0.87	0.87	0.82	
Proposed platform	*EEG + proposal*	*Pre*	**0.95**	0.36	**0.87**	**0.89**	0.80	
		Re	**0.94**	0.34	**0.90**	0.86	0.78	0.87
		F1	**0.94**	0.35	**0.88**	**0.88**	0.79	
[Biswal *et al.*, 2018]	*EEG, Resp, EMG + RCNN*	*Pre*	0.90	**0.69**	0.84	0.80	0.79	
		Re	0.81	**0.67**	0.78	0.76	0.74	0.87
		F1	0.85	**0.68**	0.81	0.78	0.76	
[Pathak *et al.*, 2021]	*EEG, EOG, EMG + CNN, bi-LSTM*	*Pre*	0.92	0.31	0.83	0.84	**0.88**	
		Re	0.92	0.50	0.84	0.67	**0.89**	0.85
		F1	0.92	0.40	0.84	0.76	**0.89**	
[Phan *et al.*, 2021]	*EEG, EOG, EMG + GRU, LSTM*	*Pre*	-	-	-	-	-	
		Re	-	-	-	-	-	**0.89**
		F1	0.92	0.50	0.88	0.85	0.88	

Table 1: Performance obtained by proposed platform and existing works using same SHHS database.

Figure 2: Overall accuracy comparison for the ablation choices defined in Representation Ablation, Section 3.3.

3.3 Ablation Study

Representation Ablation. Extensive ablation studies are conducted to demonstrate how the proposed platform performs when removing specific components. Since we advocate for truncating spectrogram in the time-domain, we compare the proposed method against:

1. space-domain truncated spectrograms;
2. proposed platform with truncated spectrograms replaced to spike train [Kayser *et al.*, 2009], with tier-2;
3. conventional SNN iterates and averaging without truncation as introduced in Section 2.2;
4. conventional SNN iterates with averaging time-domain truncated spectrograms.

In Section 4, we refer to the above ablation choices as space-domain truncated spectrograms (`SDTS`); platform with spike train input (`SpikeTrain`); conventional SNN iterates and averaging (`NoTS`); average time-domain truncated spectrograms (`ATDTS`). The proposed method is named `Proposal`.

Architecture Ablation. Tier-2 plays the central role of extracting spiking features from EEG. It is enlightening to study how the performance varies by removing/replacing the SNN layer. We compare with:

1. proposed platform without tier-2;
2. same with `SpikeTrain`. but without tier-2;
3. an LSTM layer replacing the attention layer;
4. SG/NEO spike detection [Xu *et al.*, 2021] + attention.

To ease reading, we refer to the above ablation choices as `NoTier2`, `SpikeTrainNoTier2`, `LSTM` and `SG`, `NEO`. The proposed method is named `Proposal`. The comparison results against SOTA algorithms are available in Section 4.1, and we present the ablation study results in Section 4.2. Details of the experiments are available in our online Appendix*.

---

*https://doi.org/10.6084/m9.figshare.19682961.v1

	Method		Wake	N1	N2	N3	REM	*Acc*
[Qu *et al.*, 2020]	*EEG + CNN*	*Pre*	0.86	0.37	0.84	0.84	0.83	
		Re	0.88	0.36	0.87	0.79	0.78	0.80
		F1	0.87	0.37	0.85	0.81	0.80	
[Fiorillo *et al.*, 2021]	*EEG + CNN*	*Pre*	0.90	0.48	0.80	0.87	0.79	
		Re	0.93	**0.44**	0.85	0.73	0.73	0.84
		F1	0.91	**0.46**	0.83	0.79	0.76	
Proposed platform	*EEG + proposal*	*Pre*	0.92	0.49	**0.88**	**0.90**	**0.85**	
		Re	**0.93**	0.30	**0.90**	0.85	0.84	**0.86**
		F1	**0.92**	0.40	**0.89**	**0.87**	**0.85**	
[Phan *et al.*, 2019]	*EEG, EOG + CNN*	*Pre*	0.79	**0.55**	0.88	0.85	0.75	
		Re	0.75	0.32	0.87	**0.87**	**0.91**	0.82
		F1	0.77	0.42	0.87	0.86	0.83	
[Korkalainen *et al.*, 2020]	*EEG, EOG + CNN*	*Pre*	**0.93**	0.45	0.87	0.78	**0.85**	
		Re	0.90	0.32	0.86	0.76	0.83	0.84
		F1	**0.92**	0.38	0.86	0.77	0.84	

Table 2: Performance obtained by proposed platform and existing works using same Sleep-EDF database.

Figure 3: Overall accuracy comparison for the ablation choices defined in Architecture Ablation, Section 3.3.

4 Results

Section 4.1 presents the comparison between the proposal against existing methods, followed by ablation studies on both the data representation as well as architecture in Section 4.2. Visualization for interpretability is shown in Section 4.3. For statistical significance, all results are averaged over 10 random seeds.

4.1 Comparison

In this section we show the results for the largest SHHS and Sleep-EDF datasets and present complete results in Appendix due to the page limit.

SHHS Dataset. We first compare the proposed platform against existing work that also solely leverages EEG data [Eldele *et al.*, 2021; Karimzadeh *et al.*, 2018]. From Table 1 it is visible that the proposed platform significantly outperformed those existing methods in terms of classification accuracy for all stage. Compared to [Eldele *et al.*, 2021] that leveraged CNN to extract spatio-temporal features, the *precision, recall, F1* (Pre, Re, F1 in short) of the proposed platform for the stage wake has been greatly improved to 0.95, 0.94, 0.94, respectively. Similar trend holds also for stage N2 or N3 as well. The superior performance is consistent with our

conjecture that the platform and truncated spectrograms enabled better capturing the relative phase of spatio-temporal rhythms. Such rhythms were proved by experiments to be not only complementary to time- and space-domain features but also essential in that they constitute a complete description for the underlying neuronal activity. The improvement is significant even compared with other methods exploiting multi-source signals such as [Pathak *et al.*, 2021; Phan *et al.*, 2021; Biswal *et al.*, 2018]. Compared with [Pathak *et al.*, 2021], the proposed platform attained lower Pre, Re, F1 scores for the rapid eye movement (REM) stage. This was due to the electrooculogram (EOG) signal used in their study: it is generally received that EOG signal directly reflects the characteristics of REM. Classification for the N1 stage is difficult even for human experts. On this problem, [Biswal *et al.*, 2018] achieved higher scores than all other methods. We conjecture this was due to the large training set they used: their training set contained 16,000 subjects.

Sleep-EDF Dataset. Similar conclusion holds also for the Sleep-EDF dataset: the proposed platform attained highest overall accuracy. For existing works exploiting sole EEG data [Fiorillo *et al.*, 2021; Qu *et al.*, 2020], significantly higher scores for all stages except N1 were achieved.
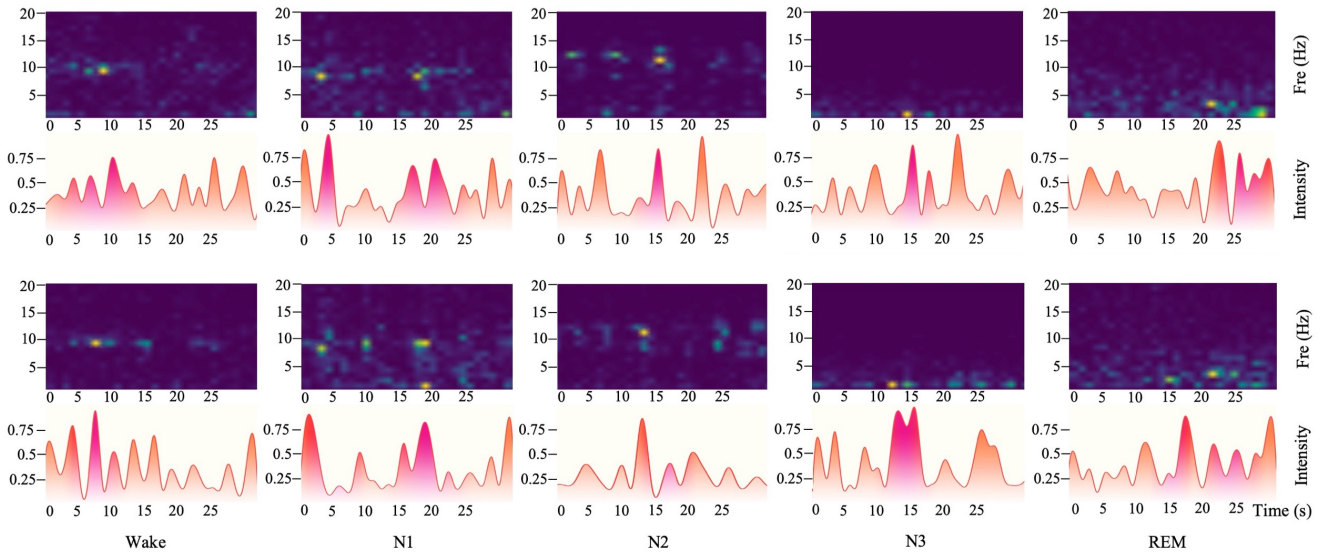
Figure 4: Input-output correspondence visualization by feeding the EEG spectrogram (upper) into the proposed platform to obtain intensity graphs (lower). Shaded areas (from shallow to deep) indicate relevance of truncated spectrogram to the stage classification result.

The slightly better performance of [Korkalainen *et al.*, 2020; Phan *et al.*, 2019] for stage REM was due to their additional EOG signal as introduced above. As a summary, it might be safe to put that the proposed method established a new EEG-based SOTA on the SHHS dataset. For the Sleep-EDF dataset, the proposed platform achieved better results than existing methods and hence established an overall SOTA.

4.2 Ablation Study

Representation Ablation Results. Recall that the ablation study choices and their abbreviations were defined in Section 3.3, and the results are plotted in Figure 2. By comparing the proposed method (`Proposal`) with space-domain truncated spectrograms (`SDTS`), we see `Proposal` consistently outperformed `SDTS` by a large margin, demonstrating the time-domain truncated spectrograms successfully matched the firing process of SNN neurons. Result of replacing the input data representation from truncated spectrograms to spike train is shown as `SpikeTrain`, which outperformed other methods except `Proposal`, demonstrating our representation can better extract underlying characteristics. This result is consolidated by `NoTS` which fed the entire EEG spectrogram instead of truncated spectrograms, incurring around 30% accuracy drop. The averaging of SNN output turned out to not destructive for the temporal information as can be seen from the curves of `ATDTS`, which performed the worst in all cases.

Architecture Ablation Results. The ablation study choices and their abbreviations were defined in **Architecture Ablation**. The results are plotted in Figure 3. `Proposal` again consistently outperformed other methods. However, it is interesting to see the classic and simple method `NoTier2` achieved accuracy only 10% worse than `Proposal` and slightly outperformed `LSTM`. This suggests that combining the binary spike sequence with our proposed platform might be a promising future direction. On the other hand, choices

without the SNN layer `SpikeTrainNoTier2`, `SG`, `NEO` performed significantly poorer, even at the presence of spike detection mechanism.

4.3 Visualization

Illustrating the correspondence between input and output is enlightening for further understanding the mechanism by which the proposed method cognizes brain activities. We feed EEG spectrograms into the proposed platform and plot the resulting intensity output in Figure 4, motivated by [Chefer *et al.*, 2021]. It is visible from the figure that the proposed method successfully captured the *highlights* in EEG and represented them as shaded peaks, in which the peaks were the *relevance to the result* perceived by the model, and colors from deep to shallow represented the extent of match to the ground truth. The correspondence shown in Figure 4 partly illustrated the superior performance of the proposed method.

5 Conclusion and Future Work

This paper presented a four-tiers-based platform for cognizing massive EEG. We proposed a pseudo-SNN for pre-coding spatio-temporal information, which is preserved in the latter attention layer for further spike pattern recognition. The spiking-analog hybrid of federated learning between tiers has not been considered in any published literature to the best of our knowledge. Extensive experiments on a variety of datasets confirmed that our proposal established a new SOTA for EEG-based methods. The superior results shed a light on how one should encode, extract and process EEG features.

Interesting future directions include investigating how performance of the platform could be further improved by more advanced encoding of EEG data. Another direction might be to verify the effectiveness of the proposed platform in more fundamental neurobiological problems such as seizures or Alzheimer's disease.

References

[Biswal *et al.*, 2018] Siddharth Biswal, Haoqi Sun, Balaji Goparaju, M Brandon Westover, J. Sun, and Matt Bianchi. Expert-level sleep scoring with deep neural networks. *Journal of the American Medical Informatics Association : JAMIA*, 25, 11 2018.

[Chefer *et al.*, 2021] Hila Chefer, Shir Gur, and Lior Wolf. Transformer interpretability beyond attention visualization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 782–791, 2021.

[Chen *et al.*, 2021] Zheng Chen, Ziwei Yang, et al. An end-to-end sleep staging simulator based on mixed deep neural networks. In *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 848–853, 2021.

[Collins and Frank, 2018] Anne G. E. Collins and Michael J. Frank. Within- and across-trial dynamics of human eeg reveal cooperative interplay between reinforcement learning and working memory. *Proceedings of the National Academy of Sciences*, 115(10):2502–2507, 2018.

[Dosovitskiy *et al.*, 2021] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, and Others. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations*, pages 1–12, 2021.

[Eldele *et al.*, 2021] Emadeldeen Eldele, Zhenghua Chen, Chengyu Liu, Min Wu, Chee-Keong Kwoh, Xiaoli Li, and Cuntai Guan. An attention-based deep learning approach for sleep stage classification with single-channel eeg. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 29:809–818, 2021.

[Fiorillo *et al.*, 2021] Luigi Fiorillo, Paolo Favaro, and Francesca Dalia Faraci. Deepsleepnet-lite: A simplified automatic sleep stage scoring model with uncertainty estimates. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 29:2076–2085, 2021.

[Hosseini *et al.*, 2021] Mohammad-Parsa Hosseini, Amin Hosseini, and Kiarash Ahi. A review on machine learning for eeg signal processing in bioengineering. *IEEE Reviews in Biomedical Engineering*, 14:204–218, 2021.

[Karimzadeh *et al.*, 2018] Foroozan Karimzadeh, Reza Boostani, Esmaeil Seraj, and Reza Sameni. A distributed classification procedure for automatic sleep stage scoring based on instantaneous electroencephalogram phase and envelope features. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 26(2):362–370, 2018.

[Kayser *et al.*, 2009] Christoph Kayser, Marcelo A. Montemurro, Nikos K. Logothetis, and Stefano Panzeri. Spike-phase coding boosts and stabilizes information carried by spatial and temporal spike patterns. *Neuron*, 61(4):597–608, 2009.

[Korkalainen *et al.*, 2020] Henri Korkalainen, Juhani Aakko, Sami Nikkonen, Samu Kainulainen, Akseli Leino, Brett Duce, Isaac O. Afara, Sami Myllymaa, Juha Töyräs, and Timo Leppänen. Accurate deep learning-based sleep staging in a clinical population with suspected obstructive sleep apnea. *IEEE Journal of Biomedical and Health Informatics*, 24(7):2073–2081, 2020.

[Li *et al.*, 2017] Yitong Li, michael Murias, samantha Major, geraldine Dawson, Kafui Dzirasa, Lawrence Carin, and David E Carlson. Targeting eeg/lfp synchrony with neural nets. In *Advances in Neural Information Processing Systems*, volume 30, pages 1–11, 2017.

[Mijatovic *et al.*, 2021] Gorana Mijatovic, Yuri Antonacci, Tatjana Loncar-Turukalo, Ludovico Minati, and Luca Faes. An information-theoretic framework to measure the dynamic interaction between neural spike trains. *IEEE Transactions on Biomedical Engineering*, 68(12):3471–3481, 2021.

[Pathak *et al.*, 2021] Shreyasi Pathak, Changqing Lu, Sunil Belur Nagaraj, Michel van Putten, and Christin Seifert. Stqs: Interpretable multi-modal spatial-temporal-sequential model for automatic sleep scoring. *Artificial Intelligence in Medicine*, 114:102038, 2021.

[Phan *et al.*, 2019] Huy Phan, Fernando Andreotti, Navin Cooray, Oliver Y. Chén, and Maarten De Vos. Joint classification and prediction cnn framework for automatic sleep stage classification. *IEEE Transactions on Biomedical Engineering*, 66(5):1285–1296, 2019.

[Phan *et al.*, 2021] Huy Phan, Oliver Y. Chen, Minh C. Tran, Philipp Koch, Alfred Mertins, and Maarten De Vos. Xsleepnet: Multi-view sequential model for automatic sleep staging. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1–12, 2021.

[Qu *et al.*, 2020] Wei Qu, Zhiyong Wang, Hong Hong, Zheru Chi, David Dagan Feng, Ron Grunstein, and Christopher Gordon. A residual based attention model for eeg based sleep staging. *IEEE Journal of Biomedical and Health Informatics*, 24(10):2833–2843, 2020.

[Sabbagh *et al.*, 2019] David Sabbagh, Pierre Ablin, Gael Varoquaux, Alexandre Gramfort, and Denis A. Engemann. Manifold-regression to predict from meg/eeg brain signals without source modeling. In *Advances in Neural Information Processing Systems*, volume 32, pages 1–10, 2019.

[Varatharajah *et al.*, 2017] Yogatheesan Varatharajah, Min Jin Chong, Krishnakant Saboo, and Others. Eeg-graph: A factor-graph-based model for capturing spatial, temporal, and observational relationships in electroencephalograms. In *Advances in Neural Information Processing Systems*, volume 30, pages 1–10, 2017.

[Wu *et al.*, 2019] Yujie Wu, Lei Deng, Guoqi Li, Jun Zhu, Yuan Xie, and L.P. Shi. Direct training for spiking neural networks: Faster, larger, better. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 1311–1318, 2019.

[Xu *et al.*, 2021] Zhendi Xu, Tianlei Wang, Jiuwen Cao, Zihang Bao, Tiejia Jiang, and Feng Gao. Bect spike detection based on novel eeg sequence features and lstm algorithms. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 29:1734–1743, 2021.