

Unraveling Temporal Dynamics in Resting-State Data with an Interpretable Siamese Convolutional Neural Network

Sergio Kazatzidis

Student ID: 6409185

Master Specialisation: Cognitive Neuroscience

Date: July 31, 2025

Internship Location: Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, the Netherlands

Supervisor(s):

- Vincent van de Ven, Associate Professor
- Yue-juan Wang, PhD student

Total Word Count: 7061 words

Table of Contents

CONTENTS

I	Introduction	3
I-A	Temporal dynamics	3
I-B	Resting-State EEG	4
I-C	Machine Learning in EEG	4
I-D	Thesis outline	5
II	Method	6
II-A	Relative Positioning	6
II-B	EEGNET	6
II-C	The proposed siamese architecture	6
II-D	Data	6
II-E	Preprocessing	6
II-F	Data analysis	8
III	Results and discussion	8
IV	Future Directions and Limitations	14
V	Conclusion	14
VI	Disclosure of generative AI in the writing process	15
	References	15

Abstract—How time is exactly represented in the brain remains a fundamental and unresolved question. In this study, we employed the Relative Positioning task to train a model to learn temporal dynamics from resting-state EEG data. Specifically, the model was tasked with determining whether two EEG windows were temporally close or distant. Using a Siamese convolutional neural network architecture, we achieved the highest accuracy of 77.22% when training on data filtered above 30 Hz, outperforming the other frequency bands, with notable variability observed across subjects. To interpret the model’s decision-making, we applied Grad-CAM, an explainable AI technique, to visualize the features driving classification. Analysis of the power spectral density of the Grad-CAM weights revealed a focus on signals above 30 Hz, as well as a prominent peak around 2 Hz. Further examination showed that the most important Grad-CAM weights in the 1.5–4 Hz range reflected changes in power modulation and rhythmicity of gamma oscillations. Our findings suggest that this low-frequency peak in the Grad-CAM weights reflects slow fluctuations in the gamma envelope, which likely encode temporal structure and could be used to differentiate temporally close and distant segments in resting-state EEG.

Keywords: EEG-net, resting-state EEG, explainable AI, Grad-CAM, Siamese network

I. INTRODUCTION

A. Temporal dynamics

Time is essential for everyday living, yet it remains largely unclear how the brain represents temporal features such as event order and duration (Wittmann, 2013). In this study, we investigated whether neural activity, as measured by resting-state EEG, contains sufficient information to distinguish the temporal proximity between time segments, and which neural features are involved in making this distinction. Understanding whether resting-state EEG encodes temporal proximity connects with established theories of how the brain reconstructs past time through neural dynamics. The perception of past time is known as retrospective timing. In retrospective timing, the brain reconstructs time estimates by analyzing memory dynamics and neural activity patterns. A similar mechanism is believed to support temporal order judgments, where the sequence of events is taken from the event trajectories associated with their encoding (Tsao et al., 2022). In a temporal order task, participants are required to maintain the sequence in which items appeared. During this task, increased theta oscillations have been observed over prefrontal regions (Hsieh et al., 2011). Another known mechanism for tracking time involves time cells, which represent the flow of time by firing at specific moments within a temporally structured interval (Eichenbaum, 2014). In addition, there are ramping cells, which are neurons that encode an epoch by gradually increasing or decreasing their activity during the experience. Together, ramping cells and time cells provide a primary mechanism for tracking time over intervals ranging from seconds to minutes (Umbach et al., 2020). Beyond memory and episodic timing, temporal organization also plays a critical role in sensory processing. For example, in the auditory domain, cortical oscillations are believed to support temporal alignment with speech. According to Giraud and Poeppel (2012), phase locking between cortical activity and the temporal structure of

speech occurs in distinct frequency bands, which may facilitate the segmentation and comprehension of spoken language.

A promising approach for investigating temporal proximity in neural signals is the Relative Positioning task, introduced by Banville et al. (2021). Originally developed for self-supervised learning in sleep staging, the task relies on the idea that temporally adjacent EEG segments often share similar neural patterns, reflecting the gradual evolution of meaningful brain activity representations. In this task, a model is trained to determine whether two EEG windows were close in time or far apart. In their study, the definitions of ‘close’ and ‘far’ were treated as hyperparameters, selected based on model performance on the sleep staging task. They found that the optimal setting defined ‘close’ as segments within 1 minute of each other, and ‘far’ as segments separated by more than 15 minutes. The learned data representation from the temporal proximity task proved to be meaningful, as it successfully transferred to the sleep staging task. This suggests that the temporal structure captured by the model contains relevant temporal information. Originally applied to sleep EEG, this approach can also be used to explore temporal structure in resting-state data, potentially revealing intrinsic temporal neural dynamics. A compelling interpretation of the Relative Positioning task, originally developed for machine learning models, is to view it as an analogy for the brain’s temporal processing mechanisms. The brain is thought to reconstruct time estimates by analyzing the temporal dynamics of neural activity associated with memory or experience (Tsao et al., 2022). One proposed mechanism involves the formation of unique neural representations for each event, which together form a trajectory of neural activity. To estimate the time between two events, the brain may compare the neural states at those moments, such as by assessing differences in oscillatory patterns or the similarity of neural representations. Events that are close in time tend to exhibit more similar neural signatures while those farther apart show greater dissimilarity. Mentally traversing this neural trajectory allows the brain to evaluate temporal distance using the evolving pattern of neural activity as an internal signal for how close or distant events are in time. One of the possible mechanisms underlying this process is the reactivation of time cells, which fire at specific moments during a temporally structured episode, thereby encoding the flow of time. Sequences of time cell activation are believed to reflect the temporal structure of unique experiences (Eichenbaum, 2014). The Relative Positioning task mirrors this cognitive process. It asks whether two EEG segments are temporally close or distant, relying on a machine learning model to extract and compare neural signatures. In this sense, the model addresses a question analogous to one the brain may solve during memory reconstruction: how far apart in time were two events, based solely on their neural representation. Although individual time cells cannot be detected with EEG due to its limited spatial resolution, EEG can capture large-scale oscillatory dynamics that may reflect the temporal structure of brain activity. By training a model to differentiate between temporally close and distant EEG segments, we tested whether

these broader neural patterns contain sufficient information to support time-based distinctions. This provided a data-driven framework for exploring temporal cognition.

B. Resting-State EEG

To gain a deeper understanding of how the brain processes time, resting-state EEG data could be utilized. In recent years, there has been increasing interest in resting-state EEG, as it captures intrinsic brain activity and has been associated with various mental disorders, cognitive decline, and altered states of consciousness (Van Diessen et al., 2015). For instance, abnormalities in resting-state EEG have been observed in individuals with autism spectrum disorder (Wang et al., 2013). Such findings suggest that resting-state data contain valuable information that can be leveraged for understanding brain function. Another application of resting-state EEG is subject identification, which demonstrates that resting data contain sufficient within-subject consistency and across-subject variability to enable reliable classification. This implies that individual neural signatures are stable over time and can serve as reliable markers of identity (Di et al., 2019). During the eyes-closed resting state, alpha activity is typically dominant in healthy individuals (Barry et al., 2007). Moreover, studies have demonstrated complex interactions between brain rhythms, such as gamma-delta coupling (Canolty et al., 2006). Utilizing magnetoencephalography (MEG), Florin and Baillet (2015) identified high-frequency gamma oscillations that were phase-coupled with slower neural oscillations, demonstrating the existence of cross-frequency interactions in resting-state brain activity. These dynamic and oscillatory interactions may vary over time, potentially aiding in the classification of EEG segments as temporally close or distant. An interesting point to note is that certain oscillations, such as alpha activity, are less prominent in individuals with attention deficit disorder (ADHD), while slower oscillations tend to be more pronounced. These differences may reflect alterations in neural communication (Woltering et al., 2012). Another example is Parkinson's disease, where resting-state activity shows a slowing across the theta, beta, and gamma frequency bands (Bosboom et al., 2006).

In this study, we investigated the temporal dynamics of healthy resting-state EEG and explored whether machine learning can differentiate between EEG window pairs that were temporally close or distant. This study aimed to reveal the temporal patterns that characterize resting-state neural activity.

C. Machine Learning in EEG

To investigate the temporal dynamics, machine learning offers a powerful tool for analyzing complex neural patterns. Deep learning networks have been utilized for numerous EEG signal tasks, including sleep staging (Phan and Mikkelsen, 2022), emotion recognition (Ng et al., 2015), and seizure detection (Ullah et al., 2018). Among these, the most commonly used architectural framework for processing EEG data is the convolutional neural network (CNN) (Craik et al., 2019). However, deep learning models typically require large

datasets, making them less commonly used in smaller-scale studies. This is where EEGNET (Lawhern et al., 2018) could provide a solution. EEGNet was developed primarily for brain-computer interface (BCI) tasks, where the amount of available data is often limited. It is a compact CNN architecture with relatively few parameters, designed to efficiently learn from EEG data with fewer training samples. EEGNet has demonstrated competitive or superior performance compared to other architectures such as ShallowNet and DeepNet (Schirmer et al., 2017) across a range of BCI-related tasks including P300 detection, feedback error-related negativity, and movement-related cortical potentials.

Self-supervised learning is a method of machine learning where the task and labels are not explicitly provided in the data but are generated through a task designed to create training labels. This technique is commonly employed to learn meaningful data representations that can be transferred to downstream tasks (Banville et al., 2021; Kazatzidis and Mehrkanoun, 2024). In this study, however, labels were explicitly assigned to form a supervised dataset, with the objective of analyzing how the model addresses the given task. To this end, a convolutional neural network (CNN) within a Siamese network architecture was utilized, processing pairs of EEG data segments. A convolutional neural network typically consists of convolutional layers, pooling layers, and fully connected layers. While CNNs are well-established in image classification (O'shea and Nash, 2015), they have increasingly been applied to EEG signal processing (Banville et al., 2021; Latifi et al., 2024). The convolutional layers employ learnable kernels that convolve across the input to extract feature maps, capturing relevant spatial or temporal patterns. Pooling layers serve to downsample these features, reducing dimensionality while preserving essential information. Finally, fully connected layers perform classification, similar to traditional artificial neural networks (O'shea and Nash, 2015). Siamese networks, composed of two weight-sharing convolutional neural networks processing paired inputs, are particularly effective for relational learning and similarity assessment between the pair. This architecture has been successfully applied to EEG classification problems involving paired data. For example, Latifi et al. (2024) employed a Siamese network to classify EEG signals by comparing individuals with ADHD to healthy controls, demonstrating the model's effectiveness in pairwise discrimination tasks. In their setup, the network was trained to determine whether two EEG windows originated from the same class, either both ADHD, or from different classes. This approach allowed the model to learn discriminative features relevant to ADHD while accounting for the intra-class variability, particularly within the control group, that often challenges traditional binary classification. The dataset comprised 121 individuals. A Leave One Subject Out Cross Validation strategy was employed in which the model was trained 121 times, each time leaving out one subject for testing while using the remaining 120 subjects for training. For ADHD and control test subjects, the corresponding training sets included 5,017 and 5,074 EEG window pairs respectively, allowing the model

to assess whether the test windows more closely resembled ADHD or control neural patterns. The model achieved a classification accuracy of 99.17%, outperforming conventional classification methods.

Similarly, You et al. (2023) employed a Siamese network combined with convolutional neural networks (CNNs) to classify sleep stages. The model was trained on pairs of EEG segments labeled as either belonging to the same or different sleep stages. The Sleep-EDF dataset (Goldberger et al., 2000) consists of two nights of EEG recordings from 20 participants. The data were segmented into 30-second epochs. Their model achieved an accuracy of 84.9%, outperforming other state-of-the-art machine learning methods in sleep stage classification. Convolutional neural networks are often used as the machine learning models within Siamese architectures for EEG analysis. For instance, Zhang et al. (2022) developed a weighted convolutional Siamese network to assist in decoding motor signals from post-stroke patients, demonstrating the model's effectiveness in capturing complex EEG patterns.

Despite their effectiveness, neural networks often lack transparency, making it challenging to explain their decisions. This lack of interpretability, commonly referred to as the black-box problem, poses a significant challenge in domains where understanding model behavior is critical, such as medical diagnosis, neuroscience, and finance. In such contexts, the explainability of AI is essential, as users may be directly affected by model predictions and must be able to trust and understand the reasoning behind them (Xu et al., 2019). To overcome this limitation, explainable AI (XAI) techniques have been developed to provide insights into the internal workings of these models. The three most utilized methods for explainable AI are LIME, SHAP and CAM (Nguyen et al., 2021). LIME (Local Interpretable Model-agnostic Explanations) is an explainable AI tool, which can be applied to any model of machine learning and can explain the predictions of the models by estimating it locally with a simpler interpretable model, for example a linear model or a decision tree. It examines how changes in the input affect the output, and in doing so, identifies the importance of each component or feature in the input (Ribeiro et al., 2016). An example of this method applied to EEG is a study where LIME was utilized to interpret a model for stroke prediction. The analysis revealed that spectral delta and theta features served as key local predictors for identifying stroke (Islam et al., 2022). Another tool is SHAP (SHapley Additive exPlanations), which explains a model's prediction by calculating each feature's contribution to that prediction. This is achieved by computing Shapley values, a concept from cooperative game theory. Shapley values are calculated by evaluating the model's output across all possible combinations of features and measuring how the prediction changes when a specific feature is added or removed (Lundberg and Lee, 2017). An instance of this method is its application to a classification task for epileptic seizure detection. SHAP values were used to quantify the contribution of each feature, providing insights that could assist medical experts and support improved patient care (Ahmad et al.,

2024). The final Explainable AI tool we will discuss is CAM (Class Activation Mapping). CAM utilizes the activation maps from the final convolutional layers of a Convolutional Neural Network (CNN) to highlight the regions of an image that the model focuses on for a certain classification. This technique helps visualize which parts of the image contribute most to the model's decision, thereby providing insight into the model's attention during prediction (Zhou et al., 2016). An example of CAM applied to EEG is in the context of driver drowsiness detection. A CNN was used to identify shared EEG features across different subjects. CAM revealed biologically interpretable patterns such as alpha spindles and theta bursts, which are indicative of a drowsy state (Cui et al., 2022). However, this method requires a specific convolutional neural network architecture, which is a significant limitation. To address this, Grad-CAM was developed to be applicable to any CNN architecture. The explainability method used in this study was Grad-CAM (Selvaraju et al., 2020), which calculates the importance weights of specific filters within CNN layers. This produces a heatmap that highlights the contribution of input features to the model's decisions. For instance, in a sleep staging task, Grad-CAM was utilized to reveal that the model's predictions relied on physiologically meaningful patterns such as sleep spindles and K-complexes (Vaquerizo-Villar et al., 2023). Grad-CAM has also been applied in the context of emotion recognition using EEG, where it revealed that specific electrodes had a greater influence on the model's classification decisions (Zhang et al., 2023). Grad-CAM can also be applied to other data types such as audio signals. For example, in a speech emotion recognition task, Grad-CAM revealed that the model focuses on specific parts of sentences that most strongly convey emotional content, enabling more accurate emotion classification (Kim and Kwak, 2024). This study aimed to uncover the temporal patterns that characterize resting-state neural activity using a Siamese Convolutional Neural Network in a data-driven manner, given the absence of a clear hypothesis about the specific form of the underlying temporal structure. By utilizing deep learning techniques alongside electroencephalography (EEG) data, we sought to investigate the mechanisms driving the model's classification decisions in a temporal task, employing Grad-CAM to enhance interpretability, and performed a detailed analysis of its important weights through time-frequency representations focused on key temporal regions.

D. Thesis outline

This paper is structured as follows: Section II describes the methods, including the temporal task, EEGNET, the proposed architecture, the data, the preprocessing, and the data analysis. Section III presents the results and discussion. Section IV provides future directions and limitations. Section V concludes the paper, followed by a disclosure on the use of generative AI.

II. METHOD

A. Relative Positioning

The Relative Positioning task, introduced by Banville et al. (2021), is designed for a model to learn temporal relationships in EEG data based on the assumption that EEG segments closer in time display more similar patterns than those further apart. This temporal similarity enables a model to learn to distinguish whether two EEG windows are close in time or far apart. In this study, we define two windows as close if they are separated by at most two time steps (i.e., $|t_i - t'_i| \leq 2$), and as far if they are separated by more than ten time steps (i.e., $|t_i - t'_i| > 10$). In our case, each time step corresponds to 2.5 seconds, reflecting the duration of each EEG window. These parameters were selected to enhance the model's ability to distinguish between the close and far conditions. Given the starting indices t_i and t'_i of the two windows, the binary label y_i is assigned according to the following rule (Kazatzidis and Mehrkanoon, 2024):

$$y_i = \begin{cases} 0, & |t_i - t'_i| > 10 \\ \text{or} \\ 1, & |t_i - t'_i| \leq 2. \end{cases} \quad (1)$$

B. EEGNET

EEGNet (Lawhern et al., 2018) is a compact convolutional neural network with relatively few parameters, designed specifically for EEG signal processing. It effectively learns both temporal and spatial features of EEG data through a combination of depthwise and separable convolutional layers. The architecture, shown in Table I, closely follows the original design, with the exception of the final dropout layer, which is fixed at 0.5 rather than being dynamically adjusted based on within- or across-subject variability. Originally developed for brain-computer interface (BCI) applications, EEGNet is particularly well-suited for scenarios with limited training data.

C. The proposed siamese architecture

To investigate temporal dynamics in resting-state EEG, we employed a Siamese Convolutional Neural Network (CNN) architecture, as illustrated in Figure 1. The network took pairs of EEG windows as input and classified them based on their temporal proximity, whether they are close or distant in time, following the Relative Positioning task (Banville et al., 2021). Each branch of the Siamese network consists of a compact CNN specifically made for EEG signal processing, known as EEGNet (Lawhern et al., 2018). The two branches share weights to ensure consistent feature extraction from each EEG window. The extracted features were subsequently passed through a flattening layer followed by fully connected layers, resulting in a 128-dimensional feature vector. These feature vectors from both branches were then combined using the L1 distance metric to compute a distance vector, which was finally utilized for binary classification. The model was implemented using Keras (Chollet et al., 2015), a high-level neural network framework, and trained on labeled EEG window pairs from resting-state recordings. Once trained, we applied Grad-CAM

TABLE I: EEGNet Architecture as implemented in this study

Block	Layer	Output Shape	Parameters / Notes
1	Input	(61, 625, 1)	EEG input (channels \times time \times 1)
	Conv2D	(61, 625, 8)	Temporal filters: 8, kernel = (1, 61), padding='same'
	Batch Normalization	(61, 625, 8)	–
	ELU Activation	(61, 625, 8)	–
	Depthwise Conv2D	(1, 625, 16)	Spatial-filter, depth multiplier = 2, kernel = (61, 1), padding='valid'
	Batch Normalization	(1, 625, 16)	–
	ELU Activation	(1, 625, 16)	–
	Average Pooling2D	(1, 156, 16)	Pool size = (1, 4)
	Dropout	(1, 156, 16)	p = 0.5
	Dropout	(1, 156, 16)	p = 0.5
2	Separable Conv2D	(1, 156, 16)	kernel = (1, 16), padding='same'
	Batch Normalization	(1, 156, 16)	–
	ELU Activation	(1, 156, 16)	–
	Average Pooling2D	(1, 39, 16)	Pool size = (1, 4)
	Dropout	(1, 39, 16)	p = 0.5
3	Flatten	(624)	$1 \times 39 \times 16$
	Dense	(128)	Fully connected layer, ReLU activation

(Selvaraju et al., 2020) to visualize the timepoints of the input that contributed most to the network's decision. This step enabled us to identify neural features associated with temporal information and provided insight into the specific EEG components that represent the temporal structure.

D. Data

The dataset used in this study is a subset of the LEMON (Leipzig Mind-Brain-Body) project, which includes resting-state EEG data from 203 participants (Babayan, 2020). For this study, data from 42 healthy participants (eyes-closed condition) were selected. Each recording was approximately 8 minutes in duration. EEG data were recorded according to the international 10–10 electrode placement system, using a 61-channel setup referenced to FCz. The ground electrode was placed at the sternum, and electrode impedances were maintained below 5 k Ω . Signals were digitized at a sampling rate of 2500 Hz with a bandpass filter between 0.015 Hz and 1 kHz and an amplitude resolution of 0.1 μ V.

E. Preprocessing

Preprocessing was carried out in two main stages. The first stage was conducted by the original authors (Babayan, 2020) of the LEMON dataset and included several standard EEG cleaning procedures. The raw data were initially downsampled from 2500 Hz to 250 Hz and bandpass

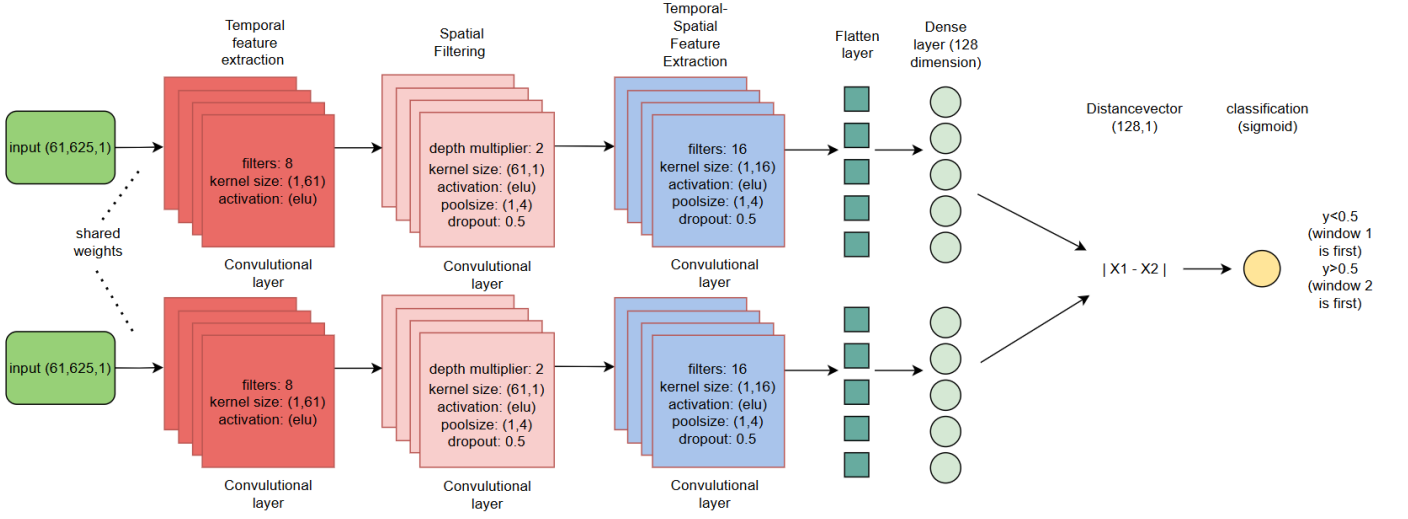


Fig. 1: Architecture of the Siamese EEGNet model. Each branch, with shared weights, processes an EEG window to generate embeddings. The absolute difference between the two embeddings is computed and used to classify whether the windows are temporally close or far apart.

filtered between 1–45 Hz using an 8th-order Butterworth filter, although some residual high-frequency activity above 45 Hz may have remained. Segments exhibiting extreme amplitude fluctuations or bursts of high-frequency noise were manually removed through visual inspection, since they most likely reflect artifacts. Dimensionality reduction was performed using principal component analysis (PCA), retaining components that explained 95% of the variance, with a maximum of 30 components. Independent component analysis (ICA) was then applied using the Infomax algorithm to remove components associated with eye blinks, eye motion, and cardiac artifacts, and the remaining components were transformed back into EEG signals at the electrode level for further analysis.

In the second stage, additional preprocessing was performed in this study using Python (Python Software Foundation, 2023) and the MNE-Python toolbox (Gramfort et al., 2013). Missing channels were interpolated to ensure a consistent 61-channel montage across participants. All 61 EEG channels were retained to preserve the full spatial resolution of the data. This choice aligns with the data-driven nature of the approach, as no prior assumptions were made about which channels would be most informative for uncovering the temporal structure. The participants were randomly shuffled and then divided into training, validation, and test sets using a 60-20-20 split, resulting in 25 participants for training, 8 for validation, and 9 for testing. Subsequently, all EEG windows were z-score normalized on a per-channel basis, as defined in Equation 2, to standardize signal scales across channels and participants. This normalization ensures that the data across subjects and channels are brought to a comparable scale, which may help more stable and effective model learning.

$$z = \frac{x - \mu}{\sigma} \quad (2)$$

where:

- z is the normalized value,
- x is the original data point,
- μ is the mean of the data,
- σ is the standard deviation of the data.

The continuous EEG recordings were segmented into epochs, each lasting 2.5 seconds (corresponding to 625 time points). The number of epochs per subject ranged from 171 to 192. Subsequently, pairs of epochs were created for each subject, with a binary label y assigned to indicate whether two windows were temporally close or far apart. To address the 1/f bias typical of EEG data, the FOOOF algorithm (Donoghue et al., 2020) was used to isolate periodic components for visualization. The resulting power spectral density (PSD), shown in Figure 2, displays a clear alpha peak around 9 Hz, which is characteristic of eyes-closed resting-state EEG.

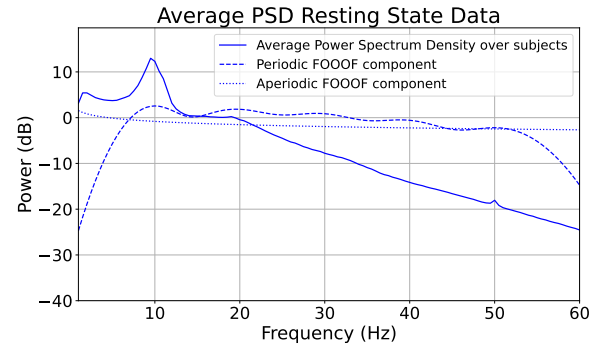


Fig. 2: Average power spectral density (PSD) of eyes-closed resting-state EEG, showing a prominent alpha peak around 9 Hz characteristic of the resting state.

F. Data analysis

The analysis involved the use of the Siamese CNN. The network’s decision-making process was analyzed using Gradient-weighted Class Activation Mapping (Grad-CAM) to identify EEG features critical for temporal proximity judgments. Grad-CAM can be applied to any layer of a CNN, with earlier layers capturing simpler spatial or temporal features, while later layers extract more complex and abstract representations. To investigate which frequency bands contributed most to the model’s classification performance, the EEG data were bandpass filtered into five distinct frequency ranges. These included: (1) the full broadband signal (1–45 Hz), (2) delta/theta band (1–8 Hz), (3) alpha band (8–15 Hz), (4) beta band (15–30 Hz), and (5) gamma band (30+ Hz). Bandpass filtering was performed using a zero-phase finite impulse response (FIR) filter with a Hamming window. By evaluating model accuracy across these frequency-specific conditions, we aimed to identify which oscillatory components most strongly support temporal proximity classification.

Additionally, the EEGNet generated a 128-dimensional feature vector (embedding) for each EEG window. These embeddings can also be used for further analysis (Heusser et al., 2018). These embeddings allowed for further analyses such as clustering to uncover patterns utilizing UMAP (McInnes et al., 2018) and exploring correlation with time.

To better understand which neural features the model relied on, we further investigated the importance scores derived from Grad-CAM. Specifically, we analyzed whether the highly weighted Grad-CAM points share common characteristics and if the same points are repeatedly chosen. For this, we applied K-means clustering to 0.4-second EEG windows surrounding the top 0.01% of Grad-CAM points. Subsequently, a time-frequency analysis using Morlet wavelet transforms was performed to examine power fluctuations around the time points highlighted by Grad-CAM. This analysis focused on the top 0.01% of Grad-CAM weights by extracting 4-second EEG segments centered on these points for further investigation. The goal was to visualize the neural features to which the model was most sensitive. We hypothesized that oscillatory components would emerge from this analysis, based on the premise that resting-state oscillations may contain information relevant for distinguishing temporally proximal from temporally distant EEG segments. Based on preliminary results, additional exploratory analyses were conducted to further investigate the Grad-CAM weights. These included phase-amplitude coupling and gamma envelope power analyses, aimed at clarifying the initial findings.

III. RESULTS AND DISCUSSION

We evaluated the model’s performance on the full frequency range as well as within specific EEG frequency bands. The highest accuracy was observed when using gamma band activity (30+ Hz), with an accuracy of 77.22%, significantly outperforming other bands, as shown in Table II. Interestingly, the model achieved higher accuracy when trained solely on the gamma band (30+ Hz) compared to the full broadband

signal. This suggests that the inclusion of lower-frequency components may introduce noise or irrelevant information, overshadowing the temporally informative features present in the gamma range.

TABLE II: Model Accuracy Across Different EEG Frequency Bands

Frequency Band (Hz)	Accuracy (%)
All	0.6251
1–8 (Delta/Theta)	0.5649
8–15 (Alpha)	0.5731
15–30 (Beta)	0.5686
30+ (Gamma)	0.7722

We present subject-wise classification accuracy across all frequency conditions in Figure 3. As shown, there is considerable variability between subjects in each frequency band. Despite this variability, performance generally remains above chance level for nearly all participants. Interestingly, the model trained on the gamma band (30+ Hz) outperformed all other frequency bands across all subjects. This suggests that the gamma band consistently contains information relevant for distinguishing temporal proximity in EEG signals across individuals. Notably, high performance in one frequency band does not necessarily translate to high performance in another. For example, subject S3 shows strong accuracy in the gamma band (Figure 3e), but performs considerably worse in the full-band condition (Figure 3a). This suggests that while relevant temporal information is present in the broadband signal, it may be obscured by lower-frequency activity, limiting the model’s ability to extract it effectively and subsequently classify the pair correctly.

An additional point of interest arises with subject S4. In the beta band (15–30 Hz), the model performed at chance level (Figure 3d), indicating it was unable to distinguish between close and far temporal windows in this frequency range. However, in the gamma band, which partially overlaps with beta frequencies due to the gradual roll-off of the FIR filter, the model achieved higher accuracy. This implies that either the gamma-specific (30+ Hz) content contained more discriminative temporal information, or that the presence of lower-frequency components in the beta band introduced noise that confounded the model’s ability to learn meaningful patterns from the data.

To further investigate the model’s strong performance in the gamma band, we analyzed the embeddings generated for pairs of EEG windows. Specifically, we computed the Manhattan distance between the embeddings of each pair and assessed its relationship with the actual time difference between the windows. As shown in Figure 4, we observed a significant positive correlation (Spearman’s $\rho = 0.524$, $p < 0.005$), indicating that the model has learned to represent time in its embedding space. Windows that were further apart in time were also embedded farther apart. One possible explanation for the better performance in the gamma band is that gamma activity may contain richer temporal dynamics

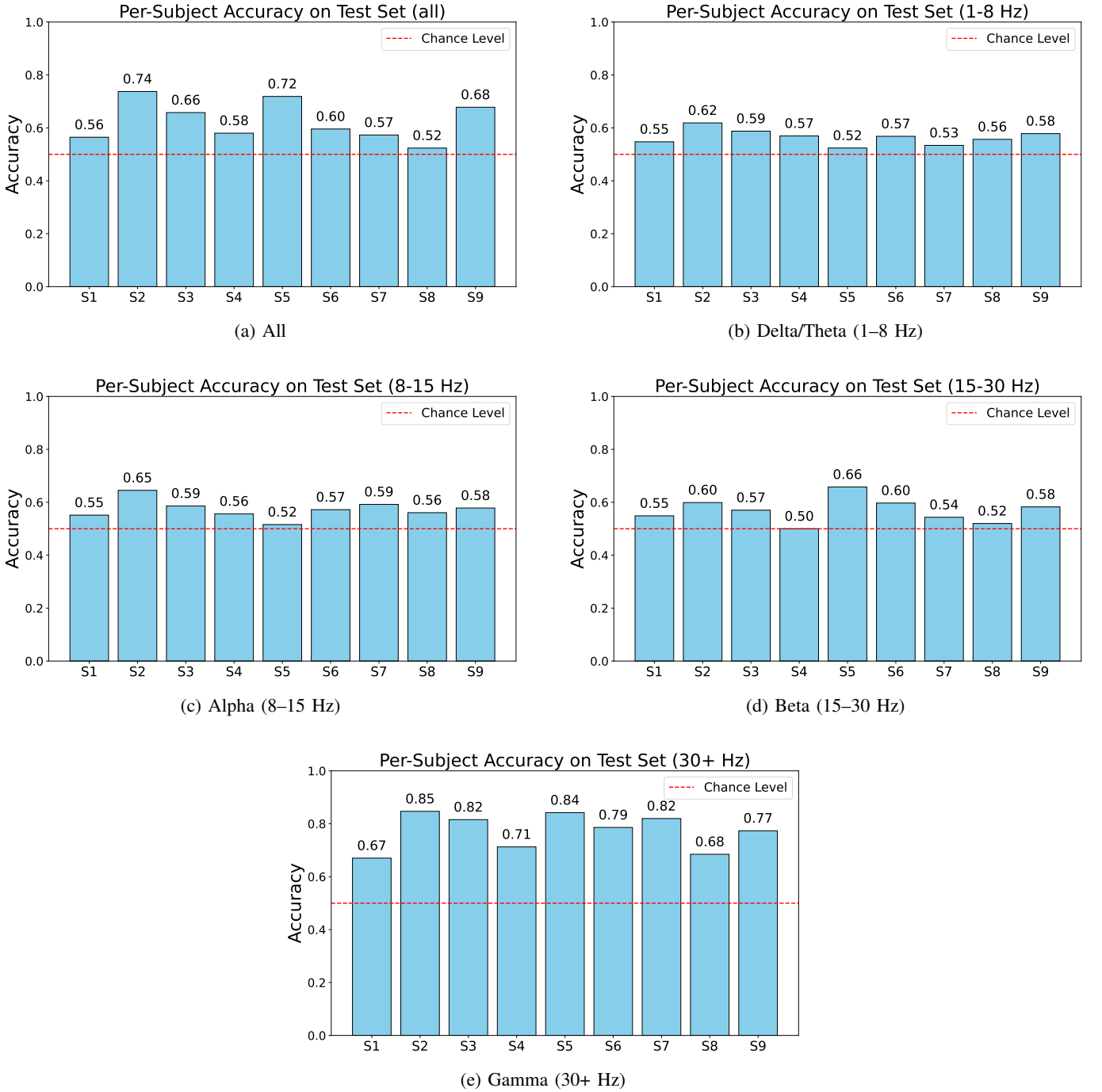


Fig. 3: Subject-wise accuracy for different EEG frequency bands. Each plot shows the performance across participants for a specific band, with variability evident in each condition.

or higher variability, making it easier for the model to detect patterns that distinguish windows from different time points.

To visualize the information contained in the embeddings, UMAP (McInnes et al., 2018) was employed, a technique that projects high-dimensional feature vectors into a two-dimensional space while preserving both local and global data structure. The UMAP algorithm was configured with 15 neighbors, a minimum distance of 0.1, and Euclidean distance

as the distance metric. In this visualization, we plotted the absolute temporal differences between pairs of EEG windows derived from the Relative Positioning task. As illustrated in Figure 5, two distinct clusters emerge, corresponding to “close” and “far” pairs. This separation indicates that the model has learned meaningful representations, enabling it to effectively differentiate temporally close pairs from distant pairs. An interesting observation of the visualization is the

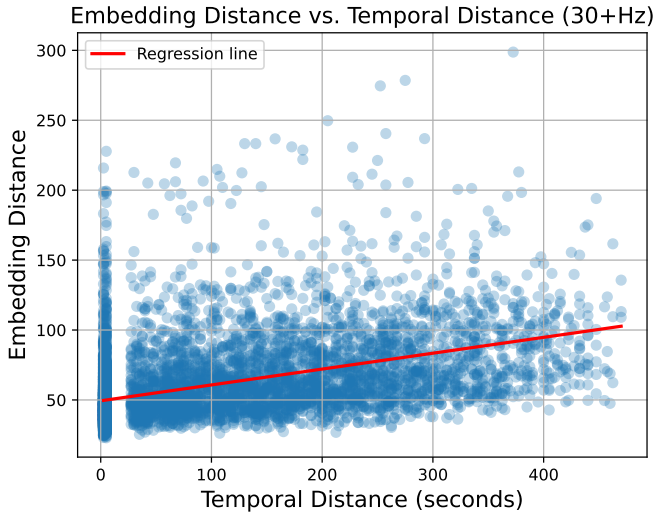


Fig. 4: Relationship between temporal distance and embedding distance of EEG window pairs. The red line represents the regression fit, showing a significant positive Spearman correlation (Spearman’s $\rho = 0.524$, $p < 0.005$), indicating that the model’s embeddings captured temporal information effectively.

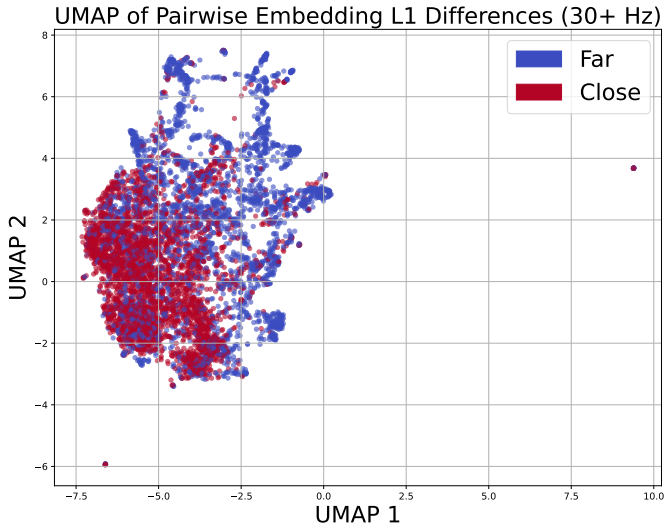


Fig. 5: Visualization of the absolute pairwise embedding differences using UMAP. Distinct clusters are observed for the far and close conditions, indicating clear separation by the model.

branching pattern observed within the “far” cluster. Given UMAP’s ability to preserve global and local structure, these branches suggested that the model leveraged multiple types of information to classify pairs as temporally far apart. This implies that the pairs labeled as “far” are not uniform but instead differ from each other in various ways. In contrast, the “close” condition forms a more cohesive cluster, reflecting that EEG windows with high similarity tend to be temporally near.

The greater variability within the “far” cluster could likely reflect the brain’s dynamic activity, which can diverge along multiple pathways over time. The observed overlap between the ‘far’ and ‘close’ conditions in the 2D UMAP visualization may arise from several factors. First, some overlap may result from misclassified samples. Second, reducing the original 128-dimensional embeddings to two dimensions likely causes information loss, limiting the ability of UMAP to fully separate the conditions. Lastly, it is possible that the neural data corresponding to these conditions are inherently similar or overlapping, reflecting the brain states themselves being difficult to differentiate.

To gain deeper insight into what the model has learned, we applied Grad-CAM, an explainable AI technique that visualizes the importance of specific time points and channels within a convolutional neural network for classification decisions. Specifically, we applied Grad-CAM to the first temporal convolutional layer. This choice was motivated by the fact that this layer is closest to the input, making its activations more interpretable, and because it primarily captures temporal features rather than spatial ones. An example of Grad-CAM applied to a particular channel and time sample is shown in Figure 6.

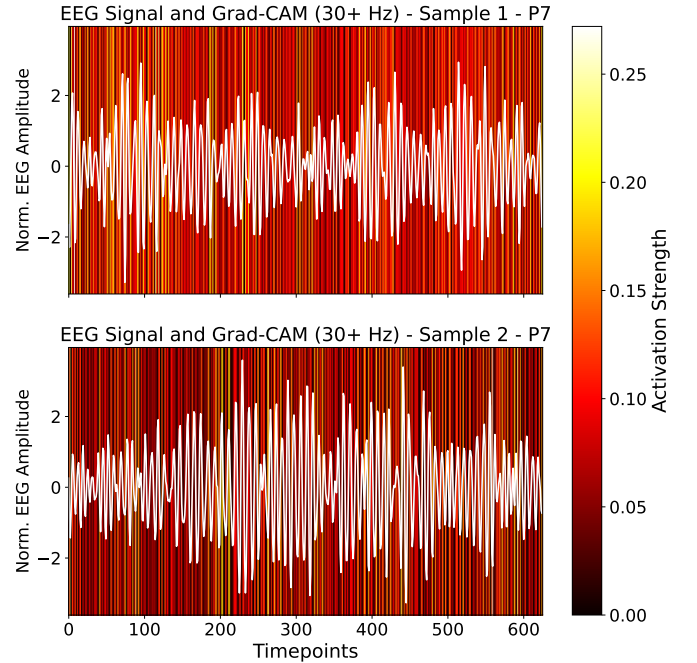


Fig. 6: Example of Grad-CAM visualization on two EEG windows from a single electrode (out of 61 total). The heatmap intensity indicates the importance of each timepoint for the model’s classification decision, with brighter areas reflecting higher importance.

The EEG signal (shown in white) has an associated importance and activation strength at each time point. We applied Grad-CAM to all test data filtered at 30 Hz and above and computed Grad-CAM maps for each EEG window.

Grad-CAM activations were then averaged across the eight temporal filters in the first convolutional layer, as well as across all channels and samples. As illustrated in Figure 7, the model primarily focuses on frequencies present in the input data, showing a prominent peak around 28 Hz and sustained power up to 70 Hz. The presence of Grad-CAM activations outside the specified filter range may result from the gradual transition bands inherent in finite impulse response (FIR) and 8th-order Butterworth filters, which do not produce perfectly sharp frequency cutoffs. Another possibility is that the power spectral density (PSD) of the average Grad-CAM weights reflects the model’s internal feature representations, which can include nonlinear combinations generated within the network. Consequently, even though the input data is band-limited (e.g., filtered below 45 Hz), the internal activations can exhibit richer spectral content at frequencies beyond this range. Therefore, the model’s internal representations, and by extension the averaged Grad-CAM weights, can show activity at these combined frequencies, even if they are not explicitly present in the input signal. Thus, peaks in the Grad-CAM PSD may not directly mirror the input signal but instead arise from learned representations within the model. Moreover, it reveals that certain frequency components within the data are attended to more closely and assigned higher importance by the model, as shown by distinct peaks in the Grad-CAM weights. This indicates that not all frequencies contribute equally to the model’s classification decisions, which suggests that certain brain oscillations exhibit greater distinct changes over time, making them more informative for the model when assessing temporal proximity.

Interestingly, there is also a notable peak around 2 Hz, a frequency absent from the filtered input data, indicating that the model may be capturing additional temporal features beyond the primary gamma-band activity.

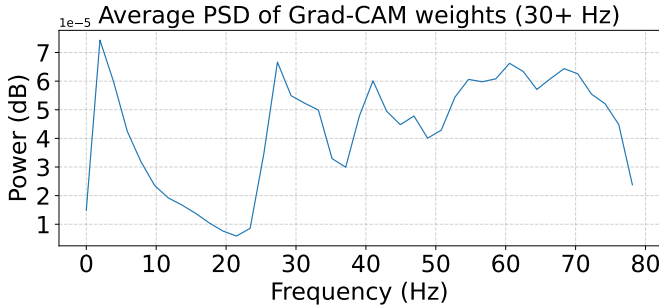


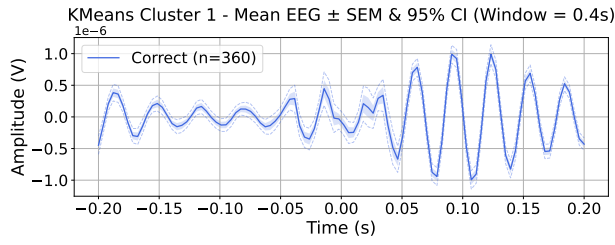
Fig. 7: Average power spectral density of the Grad-CAM weights from the model’s first temporal layer. Power remains relatively stable with peaks between 25 and 60 Hz, and a notable peak at 2 Hz.

To explore this further, we filtered the Grad-CAM weights within the 1.5–4 Hz band. We then identified the top 0.01% of the weights and extracted 0.4-second EEG segments around these points from the 1–45 Hz filtered non-segmented data, as shown in Figure 8. Eight out of nine subjects had these

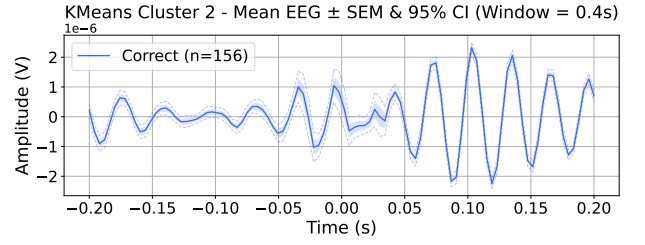
salient points. To characterize these segments, we applied K-means clustering, resulting in four distinct clusters. While one cluster (cluster 3) contained only three samples, the others included over 100 samples each. The averaged waveforms within these clusters exhibited low variability, as indicated by small standard errors of the mean (SEM), suggesting consistent patterns across samples. The oscillatory activity in the clusters was observed predominantly around 30–35 Hz. Each cluster exhibited distinct patterns of power modulation, suggesting that the Grad-CAM weights in the 1.5–4 Hz range correspond to underlying fluctuations in power. Furthermore, waveforms following the 0-second time point demonstrated increased rhythmicity and temporal consistency, whereas greater variability and reduced rhythmic structure were evident prior to this time point. Even in resting-state conditions, intrinsic brain rhythms can spontaneously phase-align or reset, leading to enhanced rhythmic patterns post-event. The 0-second time point, as identified by the prominent Grad-CAM weights, may represent a transient neural event or temporal landmark that the model has learned to prioritize, potentially reflecting brief increases in neural synchrony or power not associated with external stimuli. To further explore this, we extracted 4-second EEG segments centered around the time points corresponding to the top 0.01% of Grad-CAM weights in that frequency band. We then computed the average power from the Morlet wavelet transform across these segments. The results, displayed in Figure 9, show clear power fluctuations at a frequency of approximately 2 Hz, indicating that the model likely detected and relied on this slow rhythmic modulation to distinguish between close and far pairs. Interestingly, alpha-band activity at approximately 10 Hz also appeared to fluctuate around a 2 Hz rhythm. Although the alpha band was filtered out of the input data, this suggests that gamma oscillations may be indirectly influenced by an underlying delta rhythm that also modulates alpha activity. This kind of cross-frequency coupling, where slower oscillations such as delta influence the amplitude of faster rhythms like gamma or alpha, is commonly observed in resting-state EEG. This coupling may provide insights into how communication is coordinated between different brain networks.

One subject, Subject 6, illustrated in Figure 10, exhibited particularly strong 2 Hz power changes. The within-subject variability was often considerably lower than the variability across subjects, suggesting that brain states and oscillatory patterns are more consistent within individuals but differ substantially between individuals. Although variability across subjects was substantial, as reflected by the standard error of the mean, these findings suggest that the model may have detected patterns of phase and amplitude coupling. Such patterns could differ between temporally close and distant EEG window pairs, potentially contributing to the model’s ability to distinguish between them.

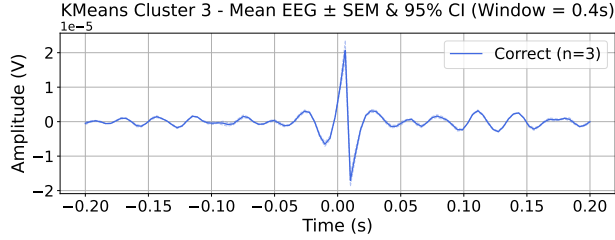
To investigate whether phase–amplitude coupling (PAC) was present in the 4-second EEG windows surrounding the Grad-CAM-derived importance points, we performed a paired-samples t-test comparing PAC values between salient and



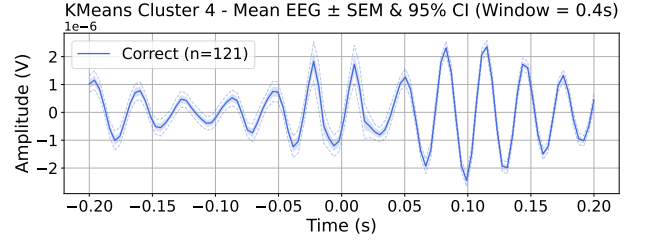
(a) Cluster 1



(b) Cluster 2

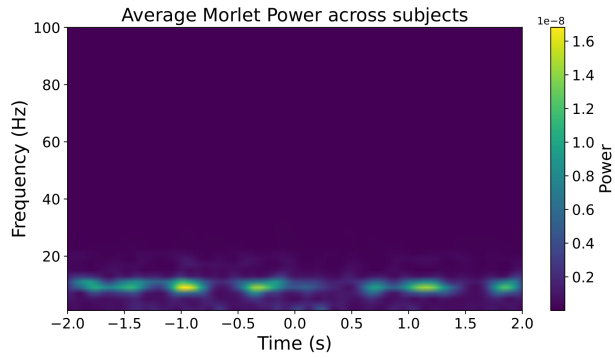


(c) Cluster 3

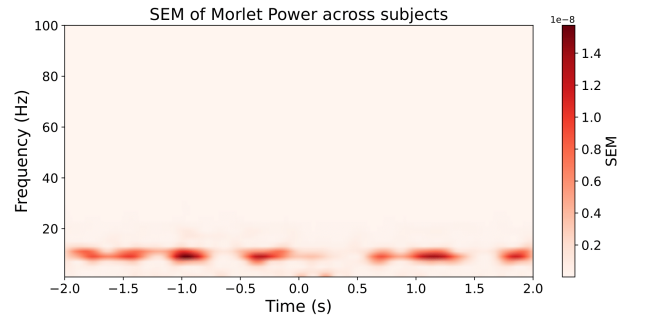


(d) Cluster 4

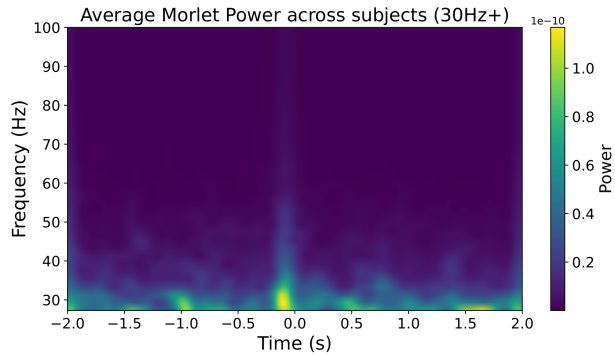
Fig. 8: Average EEG waveforms (30+ Hz) extracted from 0.4-second windows centered on the top 0.01% of Grad-CAM timepoints from the model trained on gamma-filtered data. K-means clustering grouped similar waveforms, revealing distinct power dynamics. Each subfigure corresponds to one identified waveform cluster.



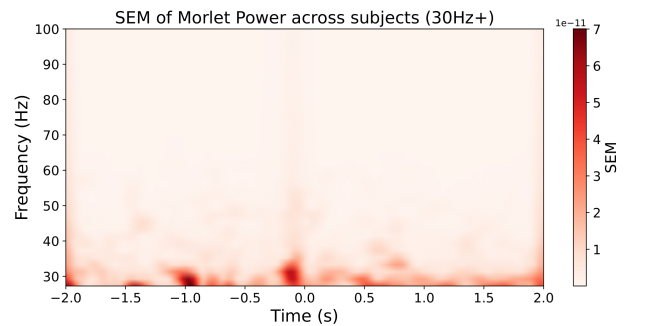
(a) Average Morlet-transformed power showing all frequencies while the Grad-CAM weights are picked from a model trained on 30+ Hz data. Some alpha bursts appear every 2 Hz.



(b) Average Morlet transformed power SEM showing considerable variability between subjects.



(c) Average Morlet power for 30+ Hz filtered data showing bursts of gamma activity around every 2 Hz.



(d) SEM for 30+ Hz Morlet power showing considerable variability between subjects.

Fig. 9: Summary of average Morlet transformed power and corresponding SEM across subjects for the 30+ Hz filtered data analyzed with Grad-CAM weights.

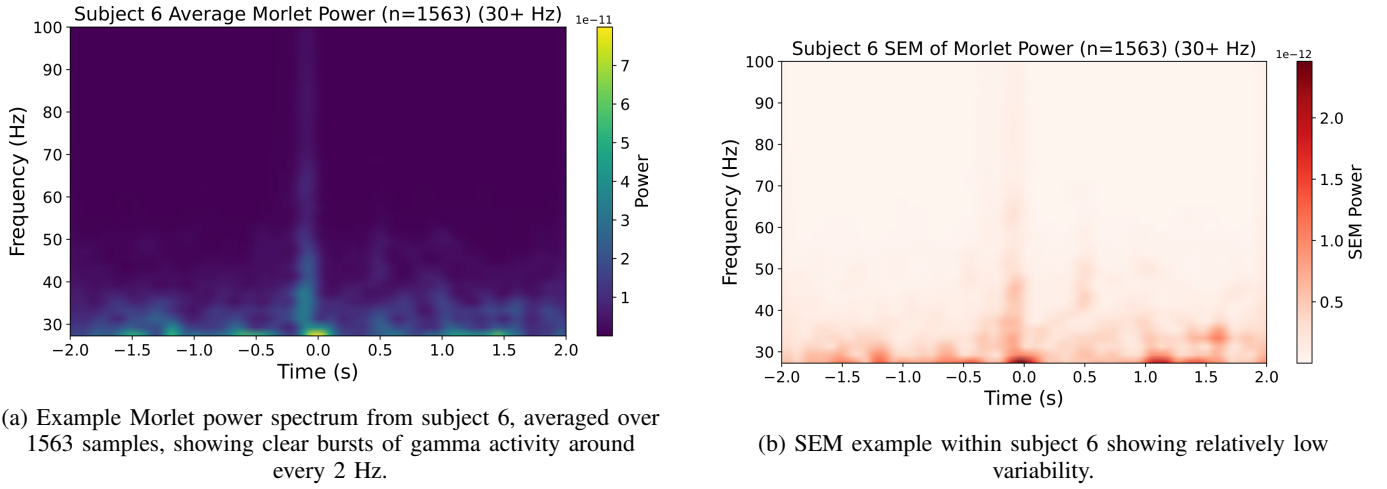


Fig. 10: Example Morlet power and SEM within a single subject (subject 6) for the 30+ Hz filtered data analyzed with Grad-CAM weights.

control conditions. PAC was computed between delta-band phase (1.5–4 Hz) and gamma-band amplitude (25–45 Hz) using 200 permutations to improve robustness. The control windows were selected randomly from non-overlapping regions of the data, with a 0.5-second buffer from salient windows to minimize overlap. We employed the Mean Vector Length (MVL) method, which calculates the length of the average complex vector representing the phase–amplitude relationship. Phase information was extracted using a Hilbert transform on the bandpass-filtered signal, and amplitude was derived by applying a Hilbert transform to the gamma envelope. All PAC computations were performed using the *tensorpac* toolbox (Combrisson et al., 2020).

At the group level, the analysis of standardized zPAC values yielded no statistically significant difference between conditions ($t(7) = -2.045$, $p = 0.0802$). Given the high inter-subject variability, we further examined the effects at the individual level and found that two subjects showed significant differences between the salient and control PAC values. Specifically, Subject 2 showed a difference with $t(29) = -2.35$, $p = 0.0258$, and Subject 8 with $t(41) = -2.34$, $p = 0.0243$. However, both cases had relatively few EEG windows, which may reduce the reliability of the result. Taken together, these findings suggest that while PAC may play a role in some individuals, it is unlikely to be a consistent mechanism used by the model across subjects. Although previous studies have observed PAC between slow and gamma oscillations in MEG data (Florin and Baillet, 2015), PAC does not appear to be significant in the EEG windows surrounding the important Grad-CAM weights compared to control windows in our data. This could indicate that PAC is not directly utilized by the model, or alternatively, that CNN models may leverage cross-frequency relationships or non-phase-locked amplitude modulations that traditional PAC measures fail to capture.

Since no robust PAC effect was found, we conducted a follow-up analysis on the gamma-band envelope itself to

better understand what features the model might be using. Specifically, we examined whether slow fluctuations in the envelope could account for the observed peak in importance weights around 1.5–4 Hz. As shown in Figure 11, the power spectral density (PSD) of the gamma envelope across all windows revealed a clear peak at approximately 1.3 Hz.

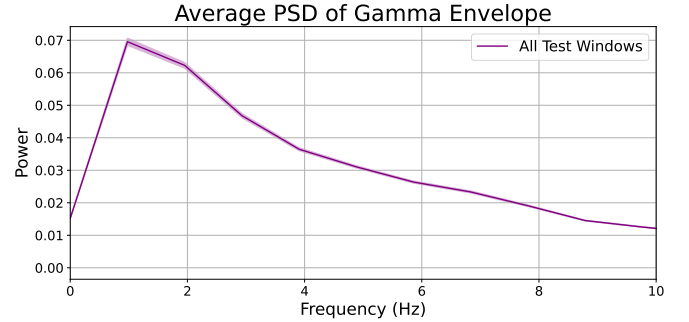


Fig. 11: Power spectral density (PSD) of the gamma band envelope across all windows, showing a consistent low-frequency peak around 1.3 Hz. This suggests slow rhythmic modulation of gamma amplitude.

To determine whether this effect differed between close and far temporal conditions, we performed an independent-samples t-test on the envelope power in the 1.5–4 Hz range. The result shows a significant difference between conditions ($t(6744) = -4.400$, $p = 0.000011$), as visualized in Figure 12.

This suggests that the model may utilize slow amplitude modulations in the gamma band to infer temporal proximity between windows. The Grad-CAM weights in the 1.5–4 Hz range likely highlight time points where differences in the gamma envelope between the two windows are most informative. Although no significant phase-amplitude coupling (PAC) was found, the presence of a strong component around 1.3 Hz in the gamma envelope suggests that slow ampli-

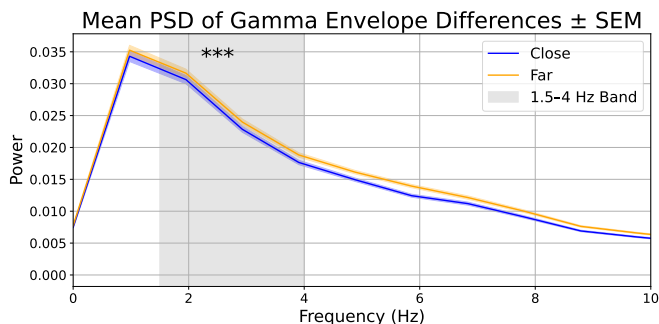


Fig. 12: Difference in gamma envelope power spectral density (PSD) between temporally close and far conditions. The 1.5–4 Hz envelope power difference differs significantly between conditions, suggesting that gamma envelope dynamics vary more with increased temporal distance.

tude modulations exist independently of delta phase. From a neurophysiological perspective, these slow fluctuations in the gamma band envelope may reflect oscillatory dynamics that unfold over time and facilitate communication between brain regions. Gamma oscillations are often considered information packets, and the slower modulation of their amplitude could serve as a stop and go signal, regulating the timing and amount of information transmitted. This type of temporal organization may enhance communication between brain regions, as coherent oscillatory neuronal groups interact most effectively when their windows of excitability, periods during which input and output are optimally aligned, occur simultaneously (Fries, 2005). Moreover, the 2 Hz modulation of gamma power may reflect slow neural processes evolving over time, producing similar gamma envelope patterns in temporally adjacent windows. These fluctuations in gamma amplitude may support the brain’s encoding of temporal structure, segmenting continuous experience into structured temporal units and discrete gamma information packets that underlie perception and cognition.

IV. FUTURE DIRECTIONS AND LIMITATIONS

Several promising directions warrant exploration in future research. One promising avenue is to investigate the model’s different layers to better understand the information each encodes, because spatial correlations between electrodes may also carry temporal information, analyzing these interactions could provide additional insights. Another potential direction is to apply this paradigm to different types of data, such as movie-watching datasets, where oscillatory patterns around important model weights might be linked to specific events or stimuli.

A key advantage of this approach lies in its ability to generate task-specific labels intrinsically, allowing the model to learn from the data without relying on externally provided annotations. This flexibility means the framework can be adapted to address a wide range of research questions by modifying the task the model performs. For example, implementing a Temporal Shuffling task (Banville et al., 2021),

in which the model is required to infer the original temporal order of EEG segments, could provide insight into how the model captures temporal order in the data. Using Grad-CAM to interpret the model’s decisions in this context could reveal precisely which temporal features or time points the model relies on to establish the order of neural activity, providing insights that can be directly related back to underlying brain dynamics. In addition, this framework can be adapted to tasks beyond temporal analysis. For example, one could design a spatial classification task where the model learns to identify the brain region corresponding to each EEG electrode, such as frontal, occipital, parietal, or temporal areas. By applying interpretability techniques like Grad-CAM, this approach could reveal the features the model uses to differentiate signals from distinct cortical regions, providing insights into spatial differences in neural activity. This serves as another example of how the method can uncover meaningful information from EEG data related to both temporal and spatial dimensions.

A limitation of the current study is the relatively small sample size of 42 participants. Increasing the number of participants in future studies could help strengthen and validate the findings presented here. Another limitation is that the reported accuracy results were not validated using permutation testing or bootstrapping, raising the possibility that the model may be overfitting to the specific configuration of training and test subjects. Future work should incorporate such techniques to ensure the generalizability and statistical significance of the results.

V. CONCLUSION

In this paper, we utilized state-of-the-art machine learning techniques to better understand both resting-state EEG data and the model’s internal workings. We found that the model performed best when given gamma-band input, though it still performed above chance with other frequency bands. Interestingly, when provided with the full frequency range, the model was unable to fully extract the information, especially from in the gamma band. Using data filtered at 30+ Hz, we demonstrated that the model learned a meaningful representation of time, as evidenced by a significant correlation between temporal distance and embedding distance. UMAP visualizations further showed that embeddings from temporally close pairs were more tightly clustered than those from distant pairs.

Applying Grad-CAM to the first temporal convolutional layer revealed that, aside from reflecting the power spectrum of the input data, the Grad-CAM power spectral density also exhibited a prominent peak at 2 Hz. Further examination of the EEG data surrounding the key Grad-CAM weights corresponding to this peak showed dynamic changes in power and rhythmic activity around these points. Phase-amplitude coupling between delta and gamma bands was initially tested as a potential cause, but was found to be insignificant. Instead, analysis of the gamma band envelope revealed a distinct peak near 1.3 Hz and a significant difference between temporally close and distant conditions within the 1.5 to 4 Hz range, likely accounting for the Grad-CAM peak near 2 Hz. This slow

modulation of gamma envelope power may reflect a temporal structure relevant for neural information processing.

These findings emphasize the model's capacity to leverage complex neural dynamics and demonstrate how explainable AI techniques, such as Grad-CAM, can provide valuable insights into brain function, potentially advancing our understanding of temporal dynamics in resting-state EEG.

VI. DISCLOSURE OF GENERATIVE AI IN THE WRITING PROCESS

During the preparation of this work, I, Sergio Kazatzidis, used ChatGPT to improve English phrasing. Following the use of this tool/service, I reviewed and edited the content as necessary, and I take full responsibility for the content of this publication.

REFERENCES

- Ahmad, I., Yao, C., Li, L., Chen, Y., Liu, Z., Ullah, I., Shabaz, M., Wang, X., Huang, K., Li, G., et al. (2024). An efficient feature selection and explainable classification method for eeg-based epileptic seizure detection. *Journal of Information Security and Applications*, 80:103654.
- Babayan, A. (2020). Max planck institut leipzig mind-brain-body dataset-lemon. Functional Connectomes Project International Neuroimaging Data-Sharing Initiative. Dataset.
- Banville, H., Chehab, O., Hyvärinen, A., Engemann, D. A., and Gramfort, A. (2021). Uncovering the structure of clinical eeg signals with self-supervised learning. *Journal of Neural Engineering*, 18(4):046020.
- Barry, R. J., Clarke, A. R., Johnstone, S. J., Magee, C. A., and Rushby, J. A. (2007). Eeg differences between eyes-closed and eyes-open resting conditions. *Clinical Neurophysiology*, 118(12):2765–2773.
- Bosboom, J., Stoffers, D., Stam, C., Van Dijk, B., Verbunt, J., Berendse, H., and Wolters, E. C. (2006). Resting state oscillatory brain dynamics in parkinson's disease: an meg study. *Clinical Neurophysiology*, 117(11):2521–2531.
- Canolty, R. T., Edwards, E., Dalal, S. S., Soltani, M., Nagarajan, S. S., Kirsch, H. E., Berger, M. S., Barbaro, N. M., and Knight, R. T. (2006). High gamma power is phase-locked to theta oscillations in human neocortex. *Science*, 313(5793):1626–1628.
- Chollet, F. et al. (2015). Keras. <https://github.com/fchollet/keras>.
- Combrisson, E., Nest, T., Brovelli, A., Ince, R. A., Soto, J. L., Guillot, A., and Jerbi, K. (2020). Tensorpac: An open-source python toolbox for tensor-based phase-amplitude coupling measurement in electrophysiological brain signals. *PLoS computational biology*, 16(10):e1008302.
- Craik, A., He, Y., and Contreras-Vidal, J. L. (2019). Deep learning for electroencephalogram (eeg) classification tasks: a review. *Journal of Neural Engineering*, 16(3):031001.
- Cui, J., Lan, Z., Liu, Y., Li, R., Li, F., Sourina, O., and Müller-Wittig, W. (2022). A compact and interpretable convolutional neural network for cross-subject driver drowsiness detection from single-channel eeg. *Methods*, 202:173–184.
- Di, Y., An, X., He, F., Liu, S., Ke, Y., and Ming, D. (2019). Robustness analysis of identification using resting-state eeg signals. *Ieee Access*, 7:42113–42122.
- Donoghue, T., Haller, M., Peterson, E. J., Varma, P., Sebastian, P., Gao, R., Noto, T., Lara, A. H., Wallis, J. D., Knight, R. T., Shestyuk, A., and Voytek, B. (2020). Parameterizing neural power spectra into periodic and aperiodic components. *Nature neuroscience*, 23(12):1655–1665.
- Eichenbaum, H. (2014). Time cells in the hippocampus: a new dimension for mapping memories. *Nature Reviews Neuroscience*, 15(11):732–744.
- Florin, E. and Baillet, S. (2015). Resting-state activity of the brain reveals spectral dynamics of local cortical networks. *NeuroImage*, 111:1–11.
- Fries, P. (2005). A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends in cognitive sciences*, 9(10):474–480.
- Giraud, A.-L. and Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nature neuroscience*, 15(4):511–517.
- Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L., and Hämäläinen, M. S. (2013). Meg and eeg data analysis with mne-python. *Frontiers in Neuroscience*, 7:267.
- Heusser, A. C., Ezzyat, Y., Shiff, I., and Davachi, L. (2018). Perceptual boundaries cause mnemonic trade-offs between local boundary processing and across-trial associative binding. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 44(7):1075–1090.
- Hsieh, L.-T., Ekstrom, A. D., and Ranganath, C. (2011). Neural oscillations associated with item and temporal order maintenance in working memory. *Journal of Neuroscience*, 31(30):10803–10810.
- Islam, M. S., Hussain, I., Rahman, M. M., Park, S. J., and Hossain, M. A. (2022). Explainable artificial intelligence model for stroke prediction using eeg signal. *Sensors*, 22(24):9859.
- Kazatzidis, S. and Mehrkanoon, S. (2024). A novel dual-stream time-frequency contrastive pretext tasks framework for sleep stage classification. In *2024 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE.
- Kim, T.-W. and Kwak, K.-C. (2024). Speech emotion recognition using deep learning transfer models and explainable techniques. *Applied Sciences*, 14(4):1553.
- Latifi, B., Amini, A., and Nasrabadi, A. M. (2024). Siamese based deep neural network for adhd detection using eeg signal. *Computers in Biology and Medicine*, 182:109092.
- Lawhern, V. J., Solon, A. J., Waytowich, N. R., Gordon, S. M., Hung, C. P., and Lance, B. J. (2018). Eegnet: a compact convolutional neural network for eeg-based brain-computer interfaces. *Journal of Neural Engineering*, 15(5):056013.
- Lundberg, S. M. and Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30.
- McInnes, L., Healy, J., and Melville, J. (2018). Umap: Uni-

- form manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*.
- Ng, H. W., Nguyen, V. D., Vonikakis, V., and Winkler, S. (2015). Deep learning for emotion recognition on small datasets using transfer learning. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, pages 443–449. ACM.
- Nguyen, H. T. T., Cao, H. Q., Nguyen, K. V. T., and Pham, N. D. K. (2021). Evaluation of explainable artificial intelligence: Shap, lime, and cam. In *Proceedings of the FPT AI Conference*, pages 1–6.
- O’shea, K. and Nash, R. (2015). An introduction to convolutional neural networks. *arXiv preprint arXiv:1511.08458*.
- Phan, H. and Mikkelsen, K. (2022). Automatic sleep staging of eeg signals: recent development, challenges, and future directions. *Physiological Measurement*, 43(4):04TR01.
- Python Software Foundation (2023). *Python Language Reference, version 3.12*. Python Software Foundation.
- Ribeiro, M. T., Singh, S., and Guestrin, C. (2016). ” why should i trust you?” explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1135–1144.
- Schirrmeister, R. T., Springenberg, J. T., Fiederer, L. D. J., Glasstetter, M., Eggensperger, K., Tangermann, M., Hutter, F., Burgard, W., and Ball, T. (2017). Deep learning with convolutional neural networks for eeg decoding and visualization. *Human brain mapping*, 38(11):5391–5420.
- Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D. (2020). Grad-cam: visual explanations from deep networks via gradient-based localization. *International Journal of Computer Vision*, 128:336–359.
- Tsao, A., Yousefzadeh, S. A., Meck, W. H., Moser, M. B., and Moser, E. I. (2022). The neural bases for timing of durations. *Nature Reviews Neuroscience*, 23(11):646–665.
- Ullah, I., Hussain, M., and Aboalsamh, H. (2018). An automated system for epilepsy detection using eeg brain signals based on deep learning approach. *Expert Systems with Applications*, 107:61–71.
- Umbach, G., Kantak, P., Jacobs, J., Kahana, M., Pfeiffer, B. E., Sperling, M., and Lega, B. (2020). Time cells in the human hippocampus and entorhinal cortex support episodic memory. *Proceedings of the National Academy of Sciences*, 117(45):28463–28474.
- Van Diessen, E., Numan, T., Van Dellen, E., Van Der Kooij, A. W., Boersma, M., Hofman, D., et al., and Stam, C. J. (2015). Opportunities and methodological challenges in eeg and meg resting state functional brain network research. *Clinical Neurophysiology*, 126(8):1468–1481.
- Vaquero-Villar, F., Gutiérrez-Tobal, G. C., Calvo, E., Álvarez, D., Kheirandish-Gozal, L., Del Campo, F., Gozal, D., and Hornero, R. (2023). An explainable deep-learning model to stage sleep states in children and propose novel eeg-related patterns in sleep apnea. *Computers in Biology and Medicine*, 165:107419.
- Wang, J., Barstein, J., Ethridge, L. E., Mosconi, M. W., Takarae, Y., and Sweeney, J. A. (2013). Resting state eeg abnormalities in autism spectrum disorders. *Journal of Neurodevelopmental Disorders*, 5:1–14.
- Wittmann, M. (2013). The inner sense of time: how the brain creates a representation of duration. *Nature Reviews Neuroscience*, 14(3):217–223.
- Woltering, S., Jung, J., Liu, Z., and Tannock, R. (2012). Resting state eeg oscillatory power differences in adhd college students and their peers. *Behavioral and Brain Functions*, 8(1):60.
- Xu, F., Uszkoreit, H., Du, Y., Fan, W., Zhao, D., and Zhu, J. (2019). Explainable ai: A brief survey on history, research areas, approaches and challenges. In *Natural language processing and Chinese computing: 8th cCF international conference, NLPCC 2019, dunhuang, China, October 9–14, 2019, proceedings, part II 8*, pages 563–574. Springer.
- You, Y., Guo, X., Yang, Z., and Shan, W. (2023). A siamese network-based method for improving the performance of sleep staging with single-channel eeg. *Biomedicines*, 11(2):327.
- Zhang, S., Ang, K. K., Zheng, D., Hui, Q., Chen, X., Li, Y., Tang, N., Chew, E., Lim, R. Y., and Guan, C. (2022). Learning eeg representations with weighted convolutional siamese network: a large multi-session post-stroke rehabilitation study. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 30:2824–2833.
- Zhang, Y., Cui, C., and Zhong, S. (2023). Eeg-based emotion recognition via knowledge-integrated interpretable method. *Mathematics*, 11(6):1424.
- Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., and Torralba, A. (2016). Learning deep features for discriminative localization. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2921–2929.