Cerebral lateralization assessment: an explainable deep learning approach with channel attention mechanism

Marco A. Formoso, Juan E. Arco, Andrés Ortiz, John Q. Gan, and I. Rodríguez-Rodríguez

***Abstract*— In recent years, cross-frequency coupling (CFC) has emerged as a valuable tool in the study of a wide range of cognitive processes due to the strong evidence of its functional role in neural computation and communication. CFC computed from electroencephalography (EEG) signals provides powerful information for detecting certain neurological conditions associated with atypical cerebral lateralization. The use of deep learning (DL) in this context offers several advantages, including improved scalability and adaptability to individual variability. However, it presents several significant challenges related to the limited availability of labelled samples and the high-dimensional and noisy nature of EEG data, which can lead to overfitting, poor generalization, and temporal and spatial variability between subjects. In this work, we propose a novel deep learning approach to reveal lateralization patterns based on inter-hemispheric functional differences via CFC. To overcome the challenges associated to the use of DL in this context, we propose the use of synthetic signals for pre-training the neural network that computes a specific type of CFC, phase-amplitude coupling (PAC), and a symmetric architecture for evaluating inter-hemispheric differences. Finally, our model incorporates a custom attention layer designed to learn the most relevant information across different EEG channels and its relative importance, further enhancing its ability to detect subtle hemispheric differences and providing the necessary explainability for clinical applications. The results demonstrate a good classification performance (AUC up to 0.85) in assessing lateralization, providing explainable insights into the mechanisms of the disorder. This may aid in early detection and provide a better understanding of the neural basis associated with this condition.**

***Index Terms*— Cerebral lateralization, EEG, Phase-amplitude coupling, Deep learning, Dyslexia.**

This work was submitted for review on XX. This research is part of the PID2022-137461NB-C32, PID2022-137629OA-I00 and PID2022-137451OB-I00 projects, funded by the MCIN/AEI/10.13039/501100011033 and by MCIN/AEI/10.13039/501100011033 by "ESF Investing in your future" grant to Marco A. Formoso. I. Rodríguez-Rodríguez would like to thank Plan Andaluz de Investigación, Desarrollo e Innovación (PAIDI), Junta de Andalucía, Spain.

Marco A. Formoso (e-mail: marco.a.formoso@ic.uma.es), Juan E. Arco, Andrés Ortiz, and I. Rodríguez-Rodríguez are with the Communications Engineering Department, University of Málaga, Málaga 29071, Spain, and also with the Andalusian Research Institute in Data Science and Computational Intelligence (DASCI), Granada 18014, Spain.

Juan E. Arco is with the Department of Signal Theory, Telematics and Communications, University of Granada, Granada 18010, Spain, and with the Andalusian Research Institute in Data Science and Computational Intelligence (DASCI), Granada 18014, Spain.

John Q. Gan is with the School of Computer Science and Electronic Engineering, University of Essex, Colchester CO4 3SQ, UK.

I. Introduction

CEREBRAL lateralization refers to the functional specialization of one hemisphere for cognitive functions that are separable from those supported by the other. It has been an important research topic since Broca and Wernicke first described left-hemispheric motor control and language processing in humans during the 19th century [1]. This idea of a simple dichotomy between halves of the cognitive whole has evolved into a better understanding of brain organization and function. Modern neuroimaging techniques, including structural and functional magnetic resonance imaging (MRI), in combination with methods from cognitive neuroscience have begun to demonstrate that hemispheric lateralization is a multi-faceted process that may not only vary according to the individual but also change during development. However, there is evidence that challenges the old idea that language and logical reasoning are performed in the left hemisphere, while spatial awareness and emotional processing take place in its right-counterpart. The work of Allen et al. [2] shows a more complex picture of hemisphere interaction, with cognitive functions typically dependent on distributed networks involving both sides of the brain. This new perspective is having radical consequences on the way we study brain function, cognitive development and how to approach neurological disorders.

The study of differences in cerebral lateralization requires the acquisition of functional information. This can be carried out by electroencephalography (EEG), which measures the brain's electrical activity with high temporal resolution. Thus, EEG allows us to study brain dynamics and recognize specific oscillatory patterns linked with different cognitive processes such as those involved in learning processes [3]. Moreover, it is necessary to compute specific features to transform the information contained in raw EEG recordings into knowledge with clinical, neurological and psychological value. Where traditional approaches to assessing cerebral lateralization often rely on predefined features, costly neuroimaging techniques, or indirect behavioral measures, we propose a method using a deep learning architecture that directly extracts relevant patterns from raw EEG data. This approach enhances interpretability by identifying the most informative channels, offering a more objective and accessible framework for understanding language-related neural mechanisms.

One of the most relevant methods to infer the functional activity of the brain consists of studying cross-frequency cou-

pling (CFC), which refers to the statistical association between different frequency components of brain oscillations. There is strong evidence of its functional role in neural computation, communication, and cognition [4], making it an invaluable tool when analyzing brain activity. The most commonly observed CFC phenomena involve the modulation of higher frequency amplitudes by lower frequency oscillation phases. This is known as phase-amplitude coupling (PAC), and has a plausible biological basis that can be attributed to three main factors [4]: 1) Excitability fluctuations (low-frequency oscillations create periods of increased and decreased neural excitability, influencing the amplitude of higher frequency activity), 2) Low-frequency oscillations may modulate synaptic integration, affecting the generation of high-frequency activity, and 3) Network interactions (different frequency bands may originate from distinct neural populations, with low-frequency oscillations coordinating the activity of high-frequency generating networks).

Thus, PAC allows the exploration of how low-frequency oscillation phases modulate high-frequency oscillation amplitudes to provide valuable insights into neural coordination over temporal and spatial scales [5]. This is important for understanding both local activity as well as how different brain regions work together during cognitive tasks (long-range PAC) [6]. There are, however, several factors that make the computation of PAC challenging. The use of traditional methods such as phase-locking value (PLV), mean vector length (MVL), or modulation index (MI) is affected by noise and artifacts, leading to possible incorrect results [7]. Additionally, the computation of PAC can be challenging due to the complexity and multi-dimensionality of EEG. Models based on artificial intelligence have been developed to address these issues. Deep neural networks, such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), and long short-term Memory (LSTM) networks, are more feasible when processing EEG signals due to their ability to perform automatic feature extraction and capture intricate temporal dependencies within datasets [8]. Furthermore, other architectures have recently been used for emotion detection from multiple physiological signals, including brainwave signals like EEG, such as stacked autoencoders (SAEs) and RNNs based on LSTM [8].

However, applying deep learning techniques to EEG data presents challenges. The limited availability of labeled EEG samples, due to the complexity and cost of data acquisition, hinders the training of robust deep learning models. Additionally, EEG signals are inherently high-dimensional and noisy, with significant temporal and spatial variability between subjects. These factors can lead to overfitting, poor generalization, and difficulties in capturing the complex neural dynamics. To overcome these challenges, we propose the use of synthetic EEG data to pre-train our neural network. By generating synthetic signals that mimic the statistical properties of real EEG data and include controlled phase-amplitude coupling (PAC) characteristics, we create a large dataset for training.

This paper investigates how lateralization patterns change in subjects with learning disorders (specifically, dyslexia, one of the most common ones) using a Siamese neural network (SNN) architecture. Siamese networks can be used to measure the similarity between two inputs [9], and are here combined to design a new architecture that computes distinct patterns of cerebral lateralization. Our working hypothesis is that dissimilarities exist in brain activity patterns between the two hemispheres of dyslexic subjects. Moreover, sequential information was evaluated using long short-term memory networks, making it suitable for EEG analysis.

Another relevant aspect of this work is that all EEG channels are inputted into the network simultaneously. This leads to a reduction in the computational cost when calculating the PAC between each pair of channels since the network automatically identifies the most critical channels. Our framework also addresses some of the challenges associated with PAC computation by generating synthetic data [10]. Specifically, our method minimizes the necessity for extensive pre-processing and analytical signal calculation in PAC computation because the network already performs these tasks inherently. The use of a Siamese network allows the automatic location of the critical brain regions and key dyslexia-control group differential patterns in time [9]. Additionally, we have designed a specific attention layer to weigh the channel importance during the learning process, producing vital interpretative resources relevant to clinical trials, and helping researchers to identify those brain regions that are more informative during diagnosis. Moreover, this also contributes to enhance the accuracy of the model.

The main contributions of this work can be summarized as follows:

1) A complete deep-learning framework to quantitatively assess cerebral lateralization patterns in terms of synchronization differences via PAC estimation.
2) A neural architecture is devised to compute PAC directly from phase and amplitude signals (hereafter, PAC network).
3) A learning transfer paradigm to train the PAC network based on synthetic coupled signals with the same statistical properties of EEG. This allows to overcome the issue of the limited number of EEG samples during the training of the PAC network.
4) A Siamese architecture containing PAC network blocks eventually fine-tuned with real EEG signals to figure out differential patterns between controls and dyslexic subjects, revealing lateralization differences in language processing.
5) An innovative module following the attention principle to quantify the relevance of each EEG channel and to improve the computational efficiency of the paradigm.
6) A novel application context in which lateralization patterns are used to detect anomalies in cognitive functions associated with language processing in children with dyslexia.

The rest of the paper is structured as follows: the next section presents the previous works on the analysis of EEG signals and the alternatives used in the field of neuroscience along with their advantages and drawbacks. Section III describes the database used in this work and the methodology:

from the design and training of the deep neural architecture to how explainability is addressed. Section IV presents the results obtained in the different experiments, whereas Section V provides a detailed explanation of the implications of these results. The final conclusions are contained in Section VI.

II. Related works

Much progress has been made in analyzing EEG signals in recent decades, from simple frequency analysis to more elaborate signal processing techniques. Advances have presented an opportunity to explore brain functions underlying cognition. PAC represents a crucial alternative for studying how different brain waves interact. Specifically, PAC provides insights into how slow oscillations influence the intensities of high-frequency waves that are thought to be essential for coherence in neural activities across different temporal and spatial scales. This shift from basic filtering to complex signal processing has been enabled by a variety of other methodological advances in contexts such as cognitive neuroscience and structural brain imaging that were not possible before.

PAC analysis has proven valuable in understanding various neurological disorders. In schizophrenia, altered theta-gamma PAC, particularly in specific cortical regions, suggests disrupted neural circuits [11]. fMRI studies in obsessive-compulsive disorder (OCD) have revealed abnormal PAC in frontostriatal regions, explaining difficulties with impulse control and emotional regulation. Alzheimer's disease, particularly in early stages, exhibits decreased low- and high-frequency PAC, notably theta-gamma coupling, in the hippocampus, a key region for memory [12]. Parkinson's disease research highlights PAC's importance in studying pathological basal ganglia oscillations linked to motor symptoms. Furthermore, studies on dyslexia have implicated abnormal functional connectivity in brain networks associated with reading and language, emphasizing the role of brain connections in this disorder [13].

CFC methods have been employed to investigate phonological processing in language-related disorders. Studies have shown that individuals with dyslexia exhibit impaired synchronization between alpha and gamma brainwaves in language processing regions, contributing to reading and writing difficulties. Traditional CFC measures, such as Phase-Locking Value (PLV), Mean Vector Length (MVL), and Modulation Index (MI), while informative, have limitations [7]. These methods can be sensitive to noise, leading to inaccurate results in the presence of artifacts. Moreover, their interpretation can be complex, especially when analyzing signals with multiple frequency components at varying amplitudes [7].

Deep learning methods often require extensive training data to build robust models, but clinical EEG studies typically lack large datasets. Additionally, variability in brain development, experimental paradigms, and movement artifacts complicate data analysis [14]. Synthetic data offer a potential solution by reducing model bias and preventing overfitting [15]. Methods such as autoregressive models and Fourier transforms simulate EEG spectral features [10], while generative adversarial networks (GANs) create artificial neuroelectric signals resembling genuine EEG waveforms [8]. Despite their potential, these methods rarely integrate with deep learning frameworks for PAC estimation. Our approach bridges this gap by combining synthetic data generation with a PAC neural network, addressing the challenges of limited EEG data and traditional PAC methods. To ensure validity, the statistical properties of synthetic data must be compared with real EEG data [16].

Another challenge when calculating PAC is the multichannel nature of EEG data, which means that computing the PACs between all the possible channel combinations constitutes a high computational cost. In this context, attention mechanisms are a powerful tool to help the model identify the most critical parts of an electroencephalographic recording, which improves explainability and performance [9]. Therefore, it is possible to apply this mechanism in EEG data both spatially and temporally. This allows the identification of the most discriminative channels and time points for a given task and excludes the remaining combinations, considerably reducing the computational cost. These attention-based methods are also useful for selecting features that are relevant in the identification of changes across individuals and trials.

Other models such as those based on CNNs have been also effective in the processing of EEG signals, especially in the field of brain computer interfaces (BCI) [17]. Their ability to learn hierarchical features allows them to capture spatiotemporal patterns [18]. One of the most representative examples is EEGNet, which has been successfully used in numerous applications such as the evaluation of the P300 potential or the study of visual evoked potentials [19]. This network was also employed to distinguish ADHD subtypes and normal controls with an accuracy of 83% [20]. Furthermore, [21] introduced deep and shallow CNN architectures to classify EEG motor imagery across participants in a multi-paradigm BCI problems. The automatic feature extraction performed by CNNs allows raw EEG data to be fed into the network with slight preprocessing [8]. On the other hand, RNNs, specially built with LSTMs, constitute a feasible alternative for processing EEG data, since these models can handle long-term dependencies [8].

Siamese networks are symmetric architectures with shared weights, designed to compare two inputs by mapping them to a common feature space and measuring their similarity. Trained on pairs, they perform well in detecting differences and are robust to class imbalance. These architectures have innovative applications in neuroimaging, such as identifying subtle differences between brain hemispheres [9]. They can be used in longitudinal studies to quantify disease progression and evaluate the relationship between structural and functional changes in cognitive decline [22]. For example, a Siamese architecture has been proposed for fusing MRI and PET, improving diagnostic ability [23]. However, challenges in applying deep learning to neuroscience, particularly EEG data, include scarcity of labeled data, high individual variability, and the need for clinical explainability, which remain unaddressed.

III. Materials and Methods

A. Database

The Leeduca research group of the University of Malaga provided the EEG data used in this study. 48 subjects were

TABLE I: Demographics of the subjects.

Group	Male/Female	Mean Age (Months)
Control	17/15	94.1 ± 3.3
Dyslexia	7/9	95.6 ± 2.9

selected from 700 subjects aged 4 to 7 years in 20 different primary schools (Junta de Andalucía) according to clinical diagnosis of dyslexia by educational psychologists along with family risk assessment using the ATLAS method [24]. Clinical diagnosis also includes the criteria of -1.5 standard deviations from the mean in different reading performance metrics. Database was homogenized to avoid biases and only right-handed and native Spanish-speaking matched in socio-economic status subjects were selected [14]. Subjects with comorbidities with other neurodevelopmental disorders and sensory (hearing or visual) disturbances were discarded. Table I shows the demographic statistics of the database.

The data used in this work were obtained with the informed and written consent of the legal guardians of each child and the study was approved by the Medical Ethics Committee of the University of Málaga (ref. CEUMA 16-2020-H), in accordance with the provisions of the Declaration of Helsinki of the World Medical Association. EEG signals were acquired at 500 Hz of sampling frequency using a Brainvision actiCHamp Plus, equipped with a 32-channel amplifier and active electrodes (actiCAP, Brain Products GmbH, Germany) arranged following the standarized 10-20 system as shown in Fig. 1. Stimuli were non-interactive, consisted of band-limited white noise, modulated in amplitude at 4.8 Hz, 16 Hz, and 40 Hz, presented during 150 second each. These frequencies were determined by linguistic psychologists to correspond to the production frequencies of phonological units in human speech (Spanish speakers) [14] to match the production frequency of the different phonological units in Spanish.

To improve signal quality and mitigate skull conduction effects, we applied preprocessing steps, including filtering to remove artifacts (eye blinks, muscle activity) and enhance the signal-to-noise ratio. The Cz electrode was used as a reference, and bipolar VEOG and HEOG channels were created to be able to remove EOG components by means of Independent Component Analysis (ICA). Zero-phase filtering with a 0.1 Hz low cut-off, 120 Hz high cut-off, and a 50 Hz notch filter was applied to suppress power line noise. Then baseline correction was performed, data were re-referenced using a common average reference to reduce volume conduction influence. An expert neurophysiologist reviewed all signals to ensure minimal remaining artifacts or noise.

B. Overall method description

As explained in the introduction, we propose a complete architecture to compute differential lateralization patterns from multichannel EEG signals. The proposed approach consists of a symmetric (Siamese) neural architecture trained to detect local synchronization differences between hemispheres by means of PAC statistics. Thus, the first building block is the PAC network, designed to compute the PAC value directly

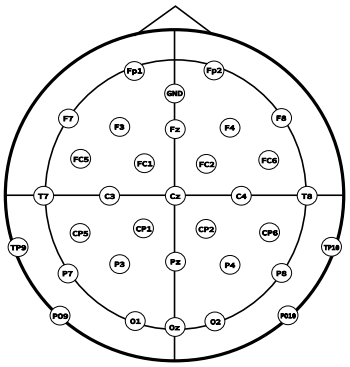

Fig. 1: Electrode montage in the extended 10–20 system used in the experiments. It consists of 32 channels plus GND, whereas the Cz is used as reference.

from two EEG time series representing the phase and amplitude signals of a channel, respectively. This network is trained using synthetic signals with the same statistical properties of real EEG signals, generated with a controlled coupling strength. After the training, two identical PAC networks with shared networks are *transferred* into a Siamese architecture. A final classification block at the top layer provides the classification outcome (control / dyslexia in our case) from the differential patterns coded in the weights of the previous layers.

Additionally, the high computational cost associated with this multichannel scenario was mitigated by designing a first layer that resembles the attention principle, in order to weigh the channels according to their relative importance. This layer eventually provides explainability, i.e. information regarding the channels used by the network to detect lateralization differences between controls and dyslexic subjects.

C. Synthetic EEG data generation

Studies based on EEG usually have a very limited sample size given the difficulty in the acquisition of this kind of data, which causes problems related to replicability and overfitting in scenarios where machine learning methods are applied. The recent interest in using deep learning approaches has made this situation even worse, since these alternatives require more data to obtain models that can generalize their results to unseen samples. To address this issue, we proposed a novel framework to generate synthetic data, an alternative that offers solutions to common problems such as inter-individual differences in brain responses or artifacts present in recordings. The generation of synthetic signals is based on simulating the spectral and temporal characteristics of real EEG signals whose statistical properties are further modified to obtain CFC. Previous studies have established that neural oscillations across different frequencies follow a specific pattern according to the power law [25]. Specifically, EEG signals are similar to colored noise, that can be generated as follows:

$$P(f) \sim \frac{1}{f^{\beta}} \tag{1}$$

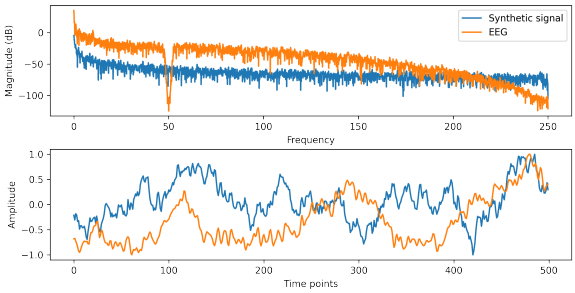

Fig. 2: Colored noise signal with $\beta = 2$ along with an EEG signal. Both signals follow the $1/f$-like powerlaw (the higher the frequency, the weaker the amplitude).

where $P(f)$ is the signal power as a function of the frequency f, and β is the parameter that determines the type of colored noise (Fig. 2). For EEG signals, β typically $\in [1, 2]$, corresponding to pink noise and brown noise, respectively [15].

These synthetic EEG-like signals are generated following the approach described by Kramer and Eden [16], which enables the creation of time series with controllable cross-frequency coupling. Specifically, brown noise is filtered into low and high frequency bands to replicate the characteristic power distribution of real EEG, and the high-frequency amplitude is modulated according to the low-frequency phase. This validated technique [16] produces synthetic signals that closely resemble genuine EEG, capturing key phase-amplitude coupling properties and allowing precise adjustments in coupling strength and noise levels. This method reduces the need for extensive preprocessing and addresses the high computational cost associated with traditional PAC computation. Additionally, it enhances the model's ability to generalize from synthetic to real data, improving performance despite the limited size of real EEG datasets. Additionally, synthetic data generation allows us to simulate various coupling scenarios.

Synthetic signal coupling: generated signals are filtered into specific frequency bands using finite impulse response (FIR) filters. Pairs of signals at different frequencies are generated: a low-frequency signal (1-30 Hz), which provides the phase, and a high-frequency signal (50-120 Hz), to modulate the amplitude. To simulate PAC, the amplitude of the high-frequency signal is modulated based on the phase of the low-frequency signal. This is achieved by applying a Hanning window centered on specific points in the phase of the low-frequency signal [10]. A Hanning window of length L, centered at point C, can be defined as:

$$w_{C,L}(n) = \begin{cases} 0.5 - 0.5\cos\left(\frac{2\pi(n-C+L/2)}{L-1}\right), \\ \qquad C - L/2 \leq n < C + L/2 \\ 0, \quad \text{otherwise} \end{cases} \tag{2}$$

where L is the total number of samples in the window and n is the current sample index, $n \in [0, L-1]$.

The duration of the Hanning window L is calculated as follows:

$$L = 0.5 * \frac{f_s}{f_p} \tag{3}$$

where f_s is the sample rate and f_p is the frequency of the phase signal. The strength of coupling, α, is controlled by adjusting the height of the Hanning window: the greater the window amplitude, the greater the coupling power. A factor in the range $[0 - 1]$ is usually employed, where 0 denotes no coupling and 1 indicates maximum coupling. To simulate the characteristics of the EEG recording conditions, Brownian noise was added to the generated signals (Algorithm 1).

The generation algorithm creates two different signals: a low frequency filtered phase signal and a high frequency filtered amplitude signal with a value of β =2. The algorithm detects the instantaneous phase points from the phase signal and generates a Hanning window whose duration corresponds to the duration of the signal. Consequently, the Hanning window is applied to the amplitude signal and then multiplied by the desired coupling strength at each phase point. Finally, extra brown noise is inserted into the two signals [16].

Algorithm 1 PAC signal generation

Require: Input signal $s[n]$, coupling strength α, coupling phase θ_c, phase signal frequency f_p, amplitude signal frequency f_a, sampling rate f_s, signal duration T (s)

Generate colored noise $s[n]$ with $\beta = 2$
$s_p[n] \leftarrow$ LPF (Low pass filter)$\{s[n], f_p\}$
$s_a[n] \leftarrow$ HPF (High pass filter)$\{s[n], f_a\}$
$\eta[n] \leftarrow s_a[n]$
$\Phi_c \leftarrow \{n \mid \angle s_p[n] = \theta_c\}$
$h[n] \leftarrow 0$ for all $n, 0 \leq n < T * f_s$
for each $n_c \in \Phi_c$ **do**
    $h[n] \leftarrow \alpha \cdot w_{C,L}[n],\ C = n_c,\ L = 0.5 * \frac{f_s}{f_p}$
end for
$s_a[n] \leftarrow s_a[n] \cdot (1 + h[n]) + \eta[n]$
return $s_p[n],\ s_a[n]$

The generation of synthetic data offers different benefits: 1) precise handling of signals with desired characteristics; 2) significant expansion of dataset dimensions; 3) creation of different settings for systematic evaluation of the performance of algorithms for PAC estimation; 4) ability to model variables such as the variability between sessions or among individuals that have a high influence in the subsequent analyses; and 5) use of pre-training models before the fine-tuning phase when real signals are used.

To ensure the quality of the generation process, the coupling value provided by our method was compared to a well-known statistical alternative, the mean vector length (MVL) [7], that is mathematically described as follows:

$$MVL = \left| \frac{\sum_{t=1}^{N} a_t e^{i(\theta_t)}}{N} \right|, N = T * f_s \tag{4}$$

where N is the total number of data points, a_t is the amplitude at the t-point of one signal, and θ_t is the phase angle at the t-point of another signal. This process is repeated with 5000

TABLE II: MVL values for different coupling strengths.

Coupling Strength α	MVL
0.0	0.0004 ± 0.000
0.25	0.0054 ± 0.000
0.5	0.0223 ± 0.003
0.75	0.0334 ± 0.004
1.0	0.0447 ± 0.005

pairs of signals for every different coupling force α in 0, 0.25, 0.5, 0.75 and 1.0. The results in Table II show that the coupling strength α of our method and the MVL are highly correlated, providing a robust method to build synchronized signals.

D. PAC estimation with deep learning

PAC measurements such as MVL, PLV, and MI have traditionally been used to study cognitive processes by means of brain synchronization in EEG and MEG [7]. However, they are effective for well-defined, low-noise signals, whereas these measures present significant limitations when applied to real EEG signals given the inherent complexity and the amount of noise that these signals usually have. The accurate computation of PAC in real signals faces challenges such as the extensive preprocessing pipeline and the need for reducing the amount of noise. Additionally, interpreting these measurements in the context of complex brain signals can be problematic, which has motivated the search for more robust and adaptable approaches. Our solution to these problems is a novel neural network framework that can determine PAC signal power directly from the original signals without the processing steps usually employed in these scenarios. This novel design is based on a Siamese architecture formed by two branches of identical configuration. The first branch receives as input the high-frequency part of the signal associated with amplitude information, while the second branch processes low-frequency content related to phase information. Each branch is formed by a 1D CNN and a series of LSTM layers, as shown in Fig. 3. The 1D CNN layer is tailored to deal with data sequences, like time series or text data [19]. To do so, filters or kernels are slided through its input stream to identify patterns, thereby enabling temporal dependencies or local structures within data to be discovered. We set the size of the convolution kernel at 20 with a stride of 3, whereas the number of output channels was set to 1. Thus, the output was computed by using the following equation:

$$y_t = \sum_{k=0}^{K-1} w_k x_{t-k} + b \tag{5}$$

where K refers to the kernel size, w_k is the kernel weight at position k, x_{t-k} is the input at position $t-k$, and b denotes the bias term.

After each convolutional layer, a MaxPool operation was performed to reduce the spatial dimensions of feature maps, which enhances computational efficiency and help focus on the most relevant features. Then, an LSTM module was implemented. LSTMs are particularly effective in addressing the problem of gradient fading faced by traditional recurrent neural networks [26]. Based on a series of memory cells, recurrent layers with specialized architecture can selectively remember or forget information over time, controlled by three gates: input, forget, and output. Mathematically, we can define the state of the cell s_t, the hidden state (output) h_t, the forget gate $g_{forget,t}$, the input gate $g_{input,t}$, and the output gate $g_{output,t}$, as follows:

Forget Gate:

$$g_{forget,t} = \sigma(W_{forget} x_t + U_{forget} h_{t-1} + b_{forget}) \tag{6}$$

Input Gate:

$$g_{input,t} = \sigma(W_{input} x_t + U_{input} h_{t-1} + b_{input}) \tag{7}$$

Output Gate:

$$g_{output,t} = \sigma(W_{output} x_t + U_{output} h_{t-1} + b_{output}) \tag{8}$$

Cell Input Activation:

$$\tilde{s_t} = \tanh(W_{cell} x_t + U_{cell} h_{t-1} + b_{cell}) \tag{9}$$

Cell State Update:

$$s_t = g_{forget,t} * s_{t-1} + g_{input,t} * \tilde{s_t} \tag{10}$$

Hidden (Output) State Update:

$$h_t = g_{output,t} * \tanh(s_t) \tag{11}$$

where σ is the sigmoid activation function, $\tanh$ is the hyperbolic tangent activation function, W, U, and b are trainable weight matrices and bias vectors for each gate operation, $*$ represents the element-wise multiplication, and x_t is the input at time step t.

Our architecture uses two LSTM layers with a hidden size of 400 units. The hidden state maintains the short- and long-term contextual information of the sequences, capturing the internal state of the network at a specific time step and propagating this information to the next step. We use this hidden state to represent the EEG sequence since its context is more crucial for us than forecasting the subsequent part of the signal.

Finally, the hidden states of both branches are concatenated and fed into a fully connected layer, consisting of two linear layers resulting in an embedding size of 50. This embedding is then passed to two other linear layers leading to a single value that represents the strength of the PAC between the input signals. This structure is beneficial for PAC estimation and reduces the need for complicated preprocessing steps that can introduce biases given their subjective nature [19]. The use of deep learning for PAC estimation is a significant advancement in EEG signal analysis, providing a robust and efficient alternative to conventional techniques, especially in disorders like dyslexia that involves examining how various brain waves interact to understand the disorder [27].

E. Attention layer

The attention layer proposed in this work represents a significant advancement in the efficient preprocessing of multi-channel data. Specifically, this module handles multiple EEG channels simultaneously and selects only the most relevant ones for the distinction of dyslexic and control groups [23]. The primary purpose of the attention layer is to process

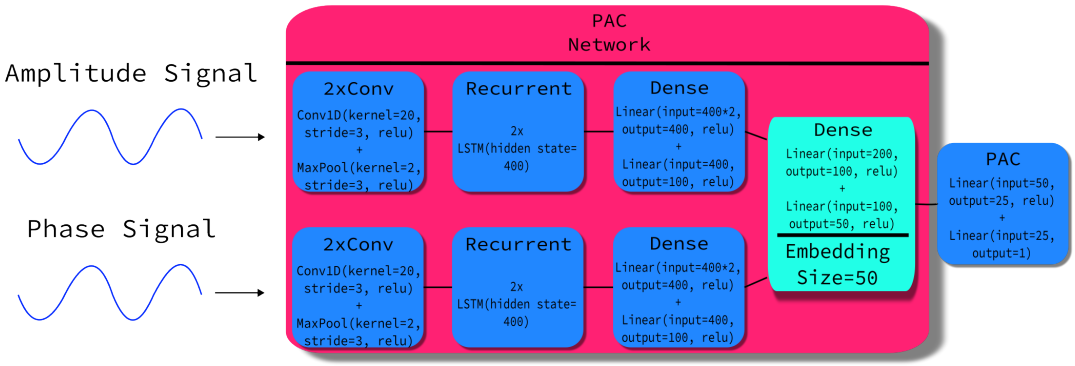

Fig. 3: Network Architecture for PAC calculation.

the collective and multiple inputs and assign them a weight according to their relevance. Based on the contribution of each individual channel, an output channel is then generated. Thus, the benefit of this method is the preservation of essential information within every channel while significantly reducing the computational burden.

A matrix X of size $C \times S$ is received by the input layer, where C represents the number of channels and S is the temporal length of the EEG signal. Each channel is then subjected to one-dimensional convolutions and a pooling operation. A kernel size of 20 is used in the convolution layers, while in average pooling layers, sizes of 5 and 2 are set. Then, we used a dense layer with 40 hidden units, whose output values were converted using a Softmax activation function. This operation provides a normalized weight vector for each individual channel, so that the length of the weight vector sums to one, ensuring a valid probability distribution across all channels. Subsequently, these weights serve as coefficients for determining the importance of each original channel [23], and they allow the interpretation of the results in neurological terms. Specifically, the weights reveal the brain regions where discrepancies in language processing emerge when comparing dyslexics and controls [28].

The process is explained in much more detail in the following equations. Let the input tensor to layer 1 be denoted as $\boldsymbol{X} \in \mathbb{R}^{C \times S}$ where C is the number of channels, and S is the length of the input sequence. The output of this first layer, $\boldsymbol{Y}_1$, would be computed as follows:

Layer 1:

Convolution Layer 1:

$$\boldsymbol{Y}_1 = \text{Conv1D}(\boldsymbol{X}, W_1, b_1), \tag{12}$$

where W_1 and b_1 are the weights and biases of the first convolution layer with kernel size $k_1 = 20$.

Activation and Average Pooling Layer 1:

$$\boldsymbol{Y}_2 = \text{AvgPool1D}(\text{ELU}(\boldsymbol{Y}_1), k_2), \tag{13}$$

where $k_2 = 5$ is the kernel size of the average pooling layer, and ELU refers to the Exponential Linear Unit function.

Layer 2:

Convolution Layer 2:

$$\boldsymbol{Y}_3 = \text{Conv1D}(\boldsymbol{Y}_2, W_2, b_2), \tag{14}$$

where W_2 and b_2 are the weights and biases of the second convolution layer with kernel size $k_1 = 20$.

Activation and Average Pooling Layer 2:

$$\boldsymbol{Y}_4 = \text{AvgPool1D}(\text{ELU}(\boldsymbol{Y}_3), k_3), \tag{15}$$

where $k_3 = 2$ is the kernel size of the average pooling layer.

Dense Layers and Attention Mechanism:

Flatten output and fully connected layer 1:

$$\boldsymbol{Y}_5 = \text{Flatten}(\boldsymbol{Y}_4), \tag{16}$$

Dense Layer 1:

$$\boldsymbol{Y}_6 = W_3 \boldsymbol{Y}_5 + b_3, \tag{17}$$

where W_3 and b_3 are the weights and biases of the first dense layer, transforming into a vector of size 40.

Activation and Dense Layer 2:

$$\boldsymbol{Y}_7 = \text{ELU}(\boldsymbol{Y}_6), \tag{18}$$

$$\boldsymbol{Y}_8 = W_4 \boldsymbol{Y}_7 + b_4, \tag{19}$$

where W_4 and b_4 are the weights and biases of the second dense layer, transforming into a vector of size M (number of channels).

Softmax Activation:

$$\boldsymbol{A} = \text{Softmax}(\boldsymbol{Y}_8), \tag{20}$$

where $\boldsymbol{A}$ represents the attention mechanism.

Attention Application:

Attention Application:

$$\boldsymbol{Z} = \sum_{m=1}^{M} \boldsymbol{A}_m \cdot \boldsymbol{X}_m \tag{21}$$

where $\boldsymbol{Z}$ represents the result of applying the attention $\boldsymbol{A}$ to the original input $\boldsymbol{X}$ through element-wise multiplication followed by summation over the channel dimension.

Output:

The final outputs of the network are:

$$(\boldsymbol{A}, \boldsymbol{Z}) \tag{22}$$

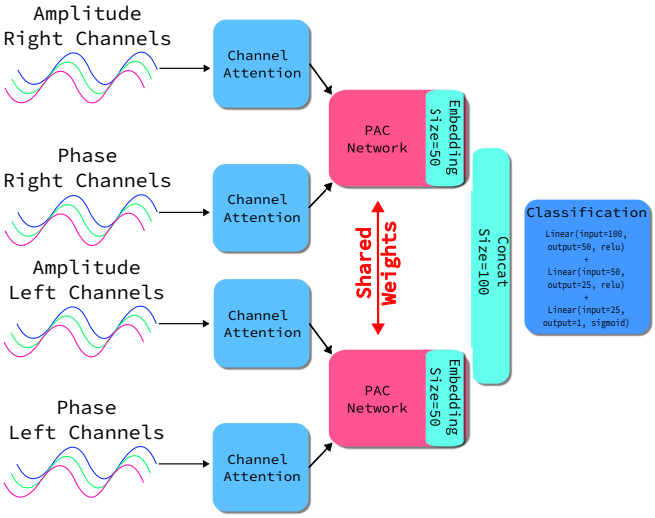

Fig. 4: Siamese network architecture. These networks have in its core the pre-trained PAC network but without its final layer.

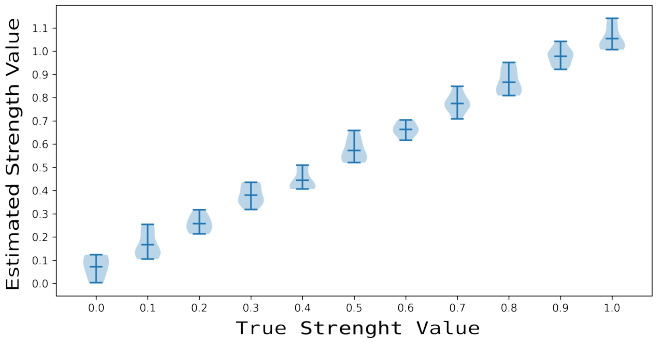

Fig. 5: Validation results for the PAC network on synthetic data. The X axis shows the true value of the signal strength and the Y axis shows the value predicted by the network.

The architecture based on an attention layer offers significant advantages in the analysis of EEG signals. First, it eliminates the need to compute PAC for each possible pair of channels, significantly reducing the computational cost. Second, it self-learns the relevant channels for the subsequent classification task (deeply discussed in next section), which in practice helps to perform built-in feature selection. This is of vital importance for explainability purposes since those channels with highest weight values are the most relevant for identifying dyslexic children.

F. Siamese network to assess lateralization patterns

Siamese networks are symmetric architectures designed to compare two inputs by mapping them to a common space and measuring their similarity or difference. Thus, the architecture proposed in this work is formed by the two previously mentioned modules: the PAC network and the attention block. It is important to note that the proposed architecture is based on comparing differences in the brain activity patterns between the left and right hemisphere, a very relevant aspect in the study of learning disabilities [29]. To do so, a Siamese network was designed, starting with the initial division of the EEG channels into two subsets correspoding with the two hemispheres. The signals are then filtered to the desired frequency bands, with one band providing the phase signals, and another at a higher frequency providing the amplitude signals.

For each hemisphere, two attention layers are implemented: one focused on the fast band and another one on the slow band. These attention layers, as described above, identify and weigh the most relevant channels within each band and hemisphere [23]. The result of each attention layer is a single, integrative signal that summarizes the most relevant information from the corresponding channels. The integrative signals from the attention layers are fed into the PAC network, whose weights have been pre-trained and locked. With this transfer learning approach it is possible to take advantage of the knowledge acquired by the PAC network during the training phase with synthetic data [18]. Specifically, two identical PAC network blocks, each sharing the same pre-trained weights, are incorporated into the Siamese structure. This ensures the retention of the deep representations learned during the synthetic data training, allowing the network to effectively capture inter-hemispheric differences without re-engineering features.

The embeddings obtained from the PAC networks of each hemisphere are concatenated. They contain critical information related to variations in neural activities across the two hemispheres, and thus can act as a higher-level feature extractor [9]. Then, these fused embeddings entered the classification process to test whether our model can identify individuals who have dyslexia by using only these representations [22]. The classifier was formed by three fully connected layers with Leaky ReLU activation functions for the first two layers and a sigmoid activation function in its output layer. This architecture ensures the distinction between controls and dyslexic patients by analyzing subtle differences in their brain activity patterns (see Fig. 4 for a schematic diagram of the complete framework).

G. Network training and validation

PAC network: The training of the PAC network involves creating pairs of artificial signals that mimic the attributes of actual EEG signals, as explained in section III-C. Phase-related signals are produced within the 1-20 Hz frequency range, while amplitude-related signals occur between 50-120 Hz. Each signal has a bandwidth of 4 Hz and is sampled at a rate of 500 Hz [10]. The signals are generated with random coupling strength ($\alpha \in [0, 1]$) and random coupling phase ($\theta \in [0, \pi]$), with both variables uniformly distributed within their respective ranges. These signals are continuously generated throughout the PAC Network training phase, resulting in a virtually infinite dataset of coupling signals [16]. In order to optimize this model, the mean squared error (MSE) is employed as the minimization criterion, and is mathematically defined as:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{23}$$

where n denotes the number of samples, y_i represents the actual values, and $\hat{y}_i$ refers to the model predictions. These values correspond to the true coupling strength (α) and the predicted output from the network. This measure serves as a good indicator of the difference between what this system thinks and what happens. When the training finished, an extensive validation process was conducted on a dataset of 1000 pairs of artificially generated signals to assess the generalization ability of the PAC network. Similar to the previous dataset, these signals were randomly generated based on α and θ values, although in this case, the α values were uniformly distributed in the range of [0,1], in steps of 0.1.

To visualize the outcomes of this validation procedure, the coupling strength values predicted by the network were compared with those calculated analytically (see Fig. 5). A close correlation emerged between model estimates and actual values when these values were graphically displayed. Observed deviations were maintained within an acceptable range, suggesting the network could accurately predict coupling strengths in artificial signals. Such errors are considered acceptable for actual EEG signal analysis. Thus, it seems clear that the trained PAC network can be successfully used as a dependable instrument for evaluating the phase-amplitude coupling, even in experimental settings with more complex and variable EEG signals.

Fine-tuning with Real EEG Data: After pre-training the PAC network on synthetic data, we fine-tune it using the real EEG signals. The phase and amplitude components extracted from the real EEG recordings are fed into the PAC network. We use the PAC values computed via traditional methods (MVL) from the real EEG data as reference targets. The network's output PAC estimates are compared to these reference values, and the weights are adjusted using backpropagation to minimize the mean squared error loss function (eq. (23)), being this time y_i the reference PAC values from traditional computation, and $\hat{y}_i$ are the PAC estimation from the network.

Siamese network: For the training of the Siamese network, experiments were performed using recordings of all available stimuli (4.8, 16, and 40 Hz). This means that a separate model was trained for each combination of stimulus and pair of low/high-frequency bands. The loss function used was the binary cross-entropy, defined as follows:

$$BCE = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(p_i) + (1-y_i)\log(1-p_i)] \quad (24)$$

where N is the total number of samples, y_i is the actual value of the sample (0 or 1), and p_i is the predicted probability of belonging to class 1.

In order to determine whether the model is valid for multiple situations and to prevent overfitting, we used a *k*-fold cross-validation technique. After training the model five times (5-fold), results were computed as the average of the outcome of each individual iteration. The performance were computed according the following metrics: sensitivity, specificity, balanced accuracy, AUC, and precision. Regarding the statistical significance, it was computed using permutation tests to randomize the labels 1000 times, building an empirical distribution based on the resulting AUC obtained [30]. The samples from this distribution are compared with the actual measure of performance (the one obtained when the actual labels are used). Results are considered significant it the percentage of samples from the empirical distribution that surpass the actual AUC is lower than 5%, which corresponds to the widely-established threshold of $p < 0.05$. Thus, the score employed to evaluate the statistical significance was computed as follows:

$$s = \frac{\sum_{i=1}^{N} I(|A_i| \geq |A|)}{N} \quad (25)$$

where s is the score, N is the total number of permutations, A is the observed AUC value, A_i is the AUC value for the $i-$th permutation, and I is the indicator function.

IV. Results

This study developed a novel framework for evaluating cerebral lateralization using cross-frequency coupling. The primary goal was to create a robust tool to address challenges in DL frameworks. We assessed the methodology's predictive power and robustness using auditory stimuli at 4.8, 16, and 40 Hz. Initially, synthetic signals pre-trained the PAC neural network to mitigate data scarcity, enhancing generalizability. The network was then trained with original data, filtered into Delta, Theta, Alpha, Beta, and Gamma bands for the Siamese network. Training on a server with 2 × Intel Xeon E5-2640 v4 CPUs and an NVIDIA GeForce RTX 3090 GPU took about 120 hours, with an additional 40 hours for fine-tuning. Post-training, the model's inference phase is less computationally intensive, allowing routine clinical use on standard hardware.

Table III summarizes the results for each combination of auditory stimuli and brainwave pairs, including alpha (9-12 Hz), beta (13-35 Hz), delta (1-4 Hz), theta (5-8 Hz), and gamma (35-120 Hz). These brainwaves, associated with distinct mental states, were matched to the gamma band, except for gamma-gamma. For the 4-Hz stimulus, the beta-gamma combination achieved the best performance with an AUC of 0.74. The 16-Hz stimulus yielded an AUC of 0.82 for the beta-gamma combination, showing superior performance. Notably, maximum performance varied with different band couplings depending on the stimulus. The highest performance was with the 40-Hz stimulus, achieving an AUC of 0.85 for the alpha-gamma combination, similar to the 16-Hz context.

Once evaluated the performance, it is important to mention the statistical significance of our findings. As Table III shows, we computed it for the three scenarios evaluated, but only in the coupling between bands where maximum performance in terms of AUC was obtained. The reason for this decision was the high computational cost of this type of operation. We must mention not only that the three contexts reach significance, but the low *p*-value found in all scenarios. In fact, this value was lower than 10^5 in all cases, demonstrating that all of them surpassed by far the conventional statistical threshold 0.05. This proves that our findings are clearly valid and not obtained just by pure chance. Fig. 6 depicts the empirical distribution obtained from the 5000 permutations carried out, as well as

TABLE III: Results obtained for different input stimuli and brainwaves. The * denotes no permutation test were done for that band combination.

Stim	Coupling	AUC	Sensitivity	Specificity	Balanced Accuracy	Precision	p-value
4.8	Alpha-Gamma	0.68 ± 0.15	0.63 ± 0.37	0.59 ± 0.48	0.61 ± 0.36	0.52 ± 0.41	*
	Beta-Gamma	**0.74 ± 0.19**	**0.67 ± 0.17**	**0.60 ± 0.19**	**0.64 ± 0.14**	**0.71 ± 0.14**	$< 10^{-5}$
	Delta-Gamma	0.61 ± 0.34	0.90 ± 0.13	0.37 ± 0.46	0.64 ± 0.35	0.53 ± 0.29	*
	Theta-Gamma	0.69 ± 0.28	0.80 ± 0.16	0.60 ± 0.49	0.70 ± 0.38	0.72 ± 0.34	*
16	**Alpha-Gamma**	**0.82 ± 0.16**	**0.83 ± 0.15**	**0.59 ± 0.18**	**0.71 ± 0.12**	**0.69 ± 0.12**	$< 10^{-5}$
	Beta-Gamma	0.76 ± 0.31	0.57 ± 0.29	0.97 ± 0.04	0.77 ± 0.20	0.77 ± 0.39	*
	Delta-Gamma	0.70 ± 0.34	0.60 ± 0.33	0.77 ± 0.39	0.69 ± 0.36	0.66 ± 0.42	*
	Theta-Gamma	0.69 ± 0.23	0.60 ± 0.33	0.74 ± 0.37	0.67 ± 0.33	0.55 ± 0.36	*
40	**Alpha-Gamma**	**0.85 ± 0.14**	**0.77 ± 0.13**	**0.79 ± 0.19**	**0.78 ± 0.11**	**0.83 ± 0.17**	$< 10^{-5}$
	Beta-Gamma	0.81 ± 0.14	0.80 ± 0.12	0.75 ± 0.38	0.78 ± 0.25	0.77 ± 0.26	*
	Delta-Gamma	0.72 ± 0.21	0.80 ± 0.16	0.60 ± 0.49	0.70 ± 0.28	0.72 ± 0.34	*
	Theta-Gamma	0.70 ± 0.22	0.67 ± 0.37	0.57 ± 0.47	0.62 ± 0.26	0.48 ± 0.36	*

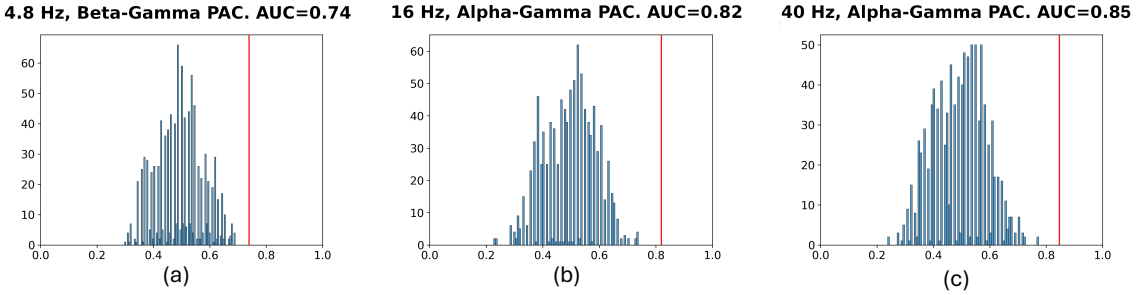

Fig. 6: Best permutation results a) 4.8 Hz, b) 16 Hz and c) 40 Hz) stimuli. Null distributions obtained over 5000 permutations are shown in blue. The red line represents the AUC obtained when using the actual labels.

the actual AUC, i.e. the AUC obtained when the correct labels were used to train the classification framework. We can see that for the three stimuli (4.8, 16 and 40 Hz), none of the labels permutation led to a better result than the one obtained by when using the actual labels, which visually represents what we explained above: results are clearly significant.

In addition to AUC, sensitivity and specificity highlight the model's clinical utility, reflecting its ability to correctly identify dyslexic subjects and exclude controls. While AUC is advantageous for handling imbalanced data and avoiding single-threshold biases, sensitivity and specificity offer a more direct measure of diagnostic value. Moderate AUC standard deviations (around 0.14–0.19) suggest some variation without critically affecting discriminative power. Moreover, statistically significant results indicate that these performance levels are not due to chance, underscoring the method's reliability.

As explained in the method Section (III-E), the attention layer plays an important role in the proposed architecture. First, this layer maximizes the performance of the proposed framework by weighting the contribution of each channel to the computation of the PAC coupling. Second, the attention mechanism allows the detection of the relative importance of each channel according to its weights, which is a mathematical description of their relevance. This resulted in greater explainability, providing additional information about what goes beyond the classification performance. Figure 7 shows a visual representation of the differences between the two classes (controls and dyslexics) in terms of the weights assigned to each channel across the experiments. The topographic maps are divided into phase and amplitude parts for the most discriminative cases found during our experiments, according to Table III. Only significant differences (those with a p-value lower than 0.05) were represented, to figure out the informative patterns related to the lateralization evaluated by our framework. In fact, these lateralization patterns were evaluated as the difference in the weights of each hemisphere. In the 4.8 Hz scenario, we found a greater importance of the channels of the right hemisphere (see Fig. 7 a) both in phase and amplitude terms. This pattern is repeated when using stimuli of 16 Hz, as shown by the darker colors that correspond with the most representative brain areas. However, this asymmetry between hemispheres was the opposite for the 40-Hz stimulus. The topographic map shows a slight imbalance toward the left hemisphere, as different brain rhythms correspond to specific cognitive functions; that is, the higher the difference between the frequency of stimuli, the higher the specific function associated with these brainwaves. This is why the topographic maps when using the 4.8-Hz and 16-Hz stimuli are more similar to the 40-Hz stimulus.

Mentioning the spatial distribution of the weight maps for the three evaluated contexts, in the first scenario (4.8 Hz stimulus), we found high differences in the frontal and temporal regions, which are usually related to speech organization. Regarding the 16-Hz stimuli, significant differences arised in parietal and temporal regions, which are crucial for decoding sounds and speech articulation. Differences in the frontal and parietal regions were identified when processing 40-Hz stimuli as evidence of a lack of fluency in phonological processing.

Table IV compares various studies using EEG with different stimuli as well as imaging data to compute differential patterns

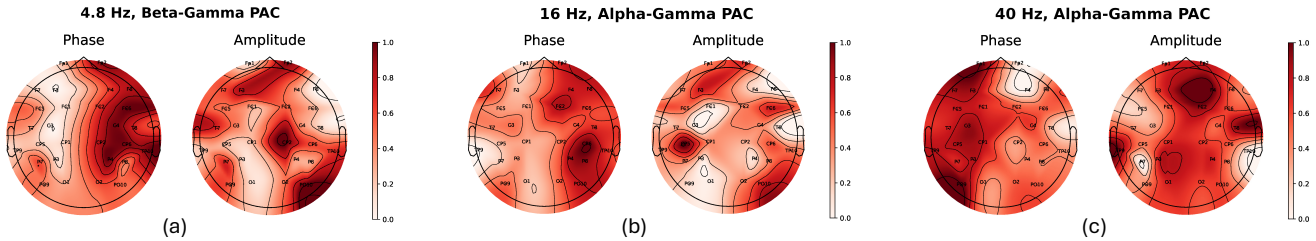

Fig. 7: Topographical maps showing the significant differences ($p<0.05$) between controls and dyslexics in channel importance computed by the attention layer for the a) 4.8 Hz, b) 16 Hz and c) 40 Hz stimuli used in this work.

TABLE IV: Comparison among works implementing different techniques for dyslexia classification. *Data not reported.

Method	Accuracy	Sensitivity	Specificity	AUC	Data
EEG-ERP / SVM (Memory task) [31]	0.79 ± *	0.89 ± *	0.58 ± *	*	Dyslexic 38 / Controls 19
EEG / SVM (Writing Task) [32]	0.72 ± *	0.76 ± *	0.67 ± *	*	Dyslexic 17 / Controls 15
EEG / SVM (Typing Task) [32]	0.78 ± *	0.88 ± *	0.67 ± *	*	Dyslexic 17 / Controls 15
EEG / SVM (Lexical decision) [33]	0.79 ± *	0.79 ± *	0.78 ± *	*	Dyslexic 20 / Controls 20
MRI / SVM [34]	0.80 ± *	0.82 ± *	0.78 ± *	*	Dyslexic 22 / Controls 15
fMRI/ CNN [35]	0.73 ± *	0.75 ± *	0.71 ± *	*	Dyslexic 19 / Controls 19
EEG/ SVM [36]	0.74 ± *	0.75 ± *	0.73 ± *	0.84 ± *	Dyslexic 16 / Controls 32
EEG/ KNN [37]	0.73 ± 0.11	0.76 ± 0.18	0.82 ± 0.15	0.76 ± 0.15	Dyslexic 16 / Controls 32
EEG/ Gradient Tree Boosting [38]	0.79 ± 0.12	0.73 ± 0.17	0.85 ± 0.13	0.79 ± 0.12	Dyslexic 16 / Controls 32
NIRS/ SVM [39]	0.72 ± 0.06	0.71 ± 0.08	0.73 ± 0.06	0.78 ± 0.04	Dyslexic 16 / Controls 32
EEG/ Denoising Autoencoder [40]	0.76 ± 0.09	0.55 ± 0.25	0.84 ± 0.25	0.76 ± *	Dyslexic 16 / Controls 32
EEG/ SVM and Graph Metrics [41]	0.73 ± *	0.72 ± *	0.74 ± *	0.79 ± *	Dyslexic 16 / Controls 32
EEG/ Probabilistic Cross Frequency Coupling [42]	0.79 ± 0.08	*	*	0.82 ± 0.11	Dyslexic 16 / Controls 32
EEG / our method, (4.8 Hz, Beta-100 Hz)	0.71 ± 0.14	0.67 ± 0.17	0.60 ± 0.19	0.74 ± 0.19	Dyslexic 16 / Controls 32
EEG / our method, (4.8 Hz, Beta-100 Hz)	0.71 ± 0.14	0.67 ± 0.17	0.60 ± 0.19	0.74 ± 0.19	Dyslexic 16 / Controls 32
EEG / our method, (16 Hz, Alpha-100 Hz)	0.69 ± 0.12	0.83 ± 0.15	0.59 ± 0.18	0.82 ± 0.16	Dyslexic 16 / Controls 32
EEG / our method, (40 Hz, Alpha-90 Hz)	**0.77 ± 0.13**	**0.79 ± 0.19**	**0.85 ± 0.14**	**0.83 ± 0.17**	**Dyslexic 16 / Controls 32**

to classify control and dyslexic subjects. In addition, some of these studies employed the same dataset that we used in this work; thus, bias in the comparison with the performance of the proposed method was minimized. Our method yields classification results comparable to those obtained in other works, with a similar sample size. A key feature of our study is the exclusive use of passive auditory stimuli, which sets it apart from other approaches. This framework minimizes bias from mood or other factors affected by voluntary responses, unlike studies such as [31]–[33] that employ reading or writing tests. These methods used more complex experimental setups and require controlling for factors like the emotional and relaxation states. [43] acquired brain activity in resting state, which most closely resembles our experimental setup. However, our methodology can be categorized as a semi-resting state, as it incorporates auditory stimuli without requiring subject intervention. To provide a context, Table III includes comparisons with non-deep learning classification methods, such as SVM [41] and purely statistical approaches [42], despite the intrinsic difficulty of comparing our end-to-end architecture—capable of autonomously extracting and classifying complex features—to methods dependent on preselected features. Moreover, although this work focuses on dyslexia, the underlying PAC-based lateralization framework can be extended to other neurological conditions, including schizophrenia or epilepsy, provided that well-curated datasets enable the identification of consistent lateralization markers. Other studies, such as [34], [35], focused on detecting structural and functional patterns using structural MRI and functional MRI.

V. Discussion

This work presents a novel approach for studying brain activity through cerebral lateralization and signal coupling to identify differential patterns in functional activity using PAC, which helps understand interactions between brain rhythms [4]. In EEG analysis, two main analyses are typically performed: local CFC, which studies coupling between different frequency bands within the same electrode, and inter-electrode coupling, which examines how the phase of a low-frequency band signal in one electrode affects the amplitude of a high-frequency band signal in another [5]. This second approach has been used to study interactions between different brain regions. Both methods provide valuable insights into neural coordination over various periods and spatial locations, enhancing our understanding of brain information processing [44].

This work proposes a deep learning framework for PAC estimation, along with a method to overcome the usual drawbacks in EEG processing as well as a symmetric architecture, whose branches are fed with EEG channels (divided by hemisphere) to evaluate differences between them. Our work is based on several studies that support the theory of hemispheric differences in the coordination of brain function in dyslexic individuals [29]. Finally, the classification decision is obtained and compared with the actual diagnosis (control/dyslexic) in order to evaluate the performance of the proposed architecture.

A key contribution of this work lies in its novel approach to EEG channel selection. Effective AI model training requires isolating relevant information from noise, ensuring that the model learns meaningful associations between features and

labels. Traditional methods often evaluate all possible channel pairs, selecting only the most informative ones for subsequent classification. This exhaustive approach incurs significant computational overhead. To address this limitation, we introduce an attentional block that enables the network to learn the relevance of each channel directly. This mechanism dynamically weights the contribution of each channel, effectively fusing them into a single informative representation. By learning these weights, the network efficiently identifies and prioritizes the most crucial channels, significantly reducing computational burden while maintaining high performance.

It is crucial to note that our results focus on interhemispheric asymmetry, rather than overall brain activity. While EEG studies analyze activity across multiple electrodes, conclusions are often drawn at a regional level based on anatomical atlases. Although EEG has limitations in spatial resolution, our attention layer helps identify the most informative channels within the recorded data. Studies on dyslexia have highlighted altered interhemispheric communication in affected individuals [29]. Our findings underscore the importance of evaluating hemispheric contributions independently. The proposed Siamese network, designed to account for these interhemispheric differences, demonstrates high classification performance, highlighting its suitability for this type of analysis. This approach is highly relevant for understanding the brain basis of dyslexia and paves the way for developing new methodologies that leverage cerebral lateralization in the study of this and other neurological conditions.

This study addresses a critical challenge in deep learning for EEG analysis: the scarcity of large training datasets. To overcome this limitation, we generated synthetic data simulating the behavior of Phase-Amplitude Coupling (PAC). We employed a pre-training strategy, utilizing the synthetic data to initialize a neural network before fine-tuning it on the actual EEG recordings. Subsequent validation using a permutation scheme yielded highly significant results (low p-values), indicating the robustness and efficacy of our classification framework. The successful application of our PAC network to real EEG data demonstrates the effectiveness of our pre-training approach with synthetic signals.

While our results underscore the effectiveness of the deep learning approach, it is important to recognize PAC estimation's inherent vulnerabilities to noise and artifacts. Traditional methods often require extensive preprocessing and expert-tuned parameters to mitigate these issues. In contrast, our end-to-end framework leverages pre-training on synthetic datasets to learn stable phase-amplitude coupling patterns, enhancing noise resilience before fine-tuning on real EEG data. This strategy reduces reliance on human expertise, allows for data-driven channel weighting, and can reveal nuanced, non-linear relationships between frequency bands that conventional techniques may overlook. As a result, our method provides a more robust, scalable, and potentially more accurate approach to PAC estimation.

Our results demonstrate optimal performance within the alpha/gamma band, aligning with prior research indicating altered alpha wave activity in dyslexia [45], [46]. Previous studies have observed altered alpha power distribution in dyslexic individuals, linking alpha activity not only to spatial attention but also to temporal organization of stimuli. Notably, one study found a significant correlation between alpha power lateralization and speech rate, impacting speech comprehension [46]. These findings underscore the crucial role of alpha oscillations in auditory information processing and language comprehension, particularly relevant to dyslexia. Our results thus corroborate existing literature, demonstrating the significance of the proposed framework for investigating brain EEG signals in the context of dyslexia

A key strength of our method is its ability to provide spatial information via an attentional block, quantifying the relevance of each electrode and thus the informativeness of its corresponding brain region. This offers crucial insights into dyslexic brain function, valuable for both clinicians and neuroscientists. Our results demonstrate anomalies in dyslexic brain lateralization and interhemispheric asymmetry across different frequency bands [47]. The identified relevant channels align with functional brain networks identified in [13], exhibiting segregated neural processing consistent with the small-worldness property observed in brain networks [48]. Notably, our method detected lateralization differences in the frontal lobe, crucial for speech production and comprehension [49], and in the primary auditory cortex, where interhemispheric interaction is crucial for phonological processing [50]. Importantly, our focus was on exploring differential lateralization patterns associated with language processing in dyslexia, rather than achieving high classification rates. This methodology has broad potential applications in studying other neurological conditions and disorders beyond the specific case of language processing presented here.

PAC at 4.8 Hz is crucial for prosodic processing, influencing rhythmic perception and speech organization. Dyslexic individuals exhibit altered PAC in right frontal and temporal regions, likely reflecting compensatory mechanisms [51]. Observed amplitude anomalies in temporal and occipital regions, critical for auditory-visual integration, support the Temporal Sampling Theory, suggesting impaired rhythmic perception due to atypical neural oscillations [51]. Notably, electrodes F3 and F7, located near Broca's area, demonstrate differential PAC weighting in dyslexic subjects.

PAC at 16 Hz stimulus relates to phonological perception at the syllable level. Significant differences in parietal and temporal regions, crucial for speech sound decoding and phonological segmentation, are observed [52]. Alpha band alterations disrupt neural synchronization; temporoparietal amplitude differences reflect sensorimotor coordination difficulties crucial for speech articulation, consistent with altered auditory-motor connectivity linked to reading automation challenges [53]. At this syllabic level, F7 and F3 exhibit distinct importance patterns, suggesting differential frontal-lobe channel engagement.

40 Hz modulation supports phonemic processing, requiring rapid neural synchronization [54]. Frontal and parietal differences suggest impaired high-frequency auditory information synchronization. Dyslexics exhibit deficient gamma oscillations, hindering rapid phoneme integration [55]. Frontal and temporal amplitude differences indicate impaired rapid auditory-visual integration essential for phonemic perception

[56]. These anomalies underscore dyslexics' difficulties in processing and articulating language sounds, consistent with gamma dysfunction in phonological processing [57]. Altered weighting patterns in electrodes like FC5 and FC3 near language-related areas further support this interpretation.

VI. Conclusions

In this work, we propose a deep learning framework for the assessment of cerebral lateralisation through interhemispheric CFC (PAC in our case) coupling differences, along with a methodology to overcome the usual problems in using deep learning architectures to learn patterns from EEG signals. On the one hand, a neural network architecture is developed to estimate interchannel PAC of EEG signals. This network is trained with synthetic coupled signals. Generating synthetic EEG data for pre-training our PAC network is a significant innovation of this work. It addresses the critical challenge of limited EEG data availability and improves PAC estimation within our deep learning framework. Once this network is able to replicate the PAC value of EEG signals, we use the transfer learning paradigm to build a Siamese network, aimed at estimating interhemispheric differences. Siamese networks present advantages in this context of limited samples and class imbalancing by learning similarities between pairs. The proposed methodology enables us to analyze the interactions between brain regions and to study the differences in how the left and right hemispheres process stimuli. Valuable insights can be obtained in the underlying mechanisms of language processing. The use of a symmetrical network architecture (Siamese) allows the extraction of relevant information from cerebral Lateralization that is highly informative for the classification of EEG signals.

We recognise that cranial conduction limits the spatial resolution of EEG, affecting precise localisation, but this has been mitigated by signal processing and the use of the attention layer. The results of this work show promising performance in classifying dyslexic and control subjects, while significant differences in the PAC strength between brain regions were observed. This suggests that dyslexic individuals exhibit altered patterns of brain activity compared to controls, and these differences can be used to accurately diagnose and understand dyslexia. In addition, it is noteworthy that the proposed methodology also addresses problems related to limited EEG data. By generating synthetic signals to train the PAC network, the dependency of large datasets is reduced, making the approach more accessible and practical. Moreover, the inclusion of an attention layer in the Siamese network reduces the computational burden and allows an efficient channel selection.

Future applications of our method extend beyond dyslexia. Neurological conditions with atypical lateralization, such as epilepsy, schizophrenia, and motor disorders, could benefit from this approach. Additionally, integrating our framework into Brain-Computer Interface systems may improve accuracy by leveraging interhemispheric differences, leading to more effective and personalized neurorehabilitation tools.

References

[1] G. Nasios, E. Dardiotis, and L. Messinis, "From Broca and Wernicke to the Neuromodulation Era: Insights of Brain Language Networks for Neurorehabilitation," *Behavioural Neurology*, vol. 2019, no. 1, p. 9894571, 2019.

[2] K.-A. Allen and R. van der Zwan, "The myth of the left-vs right-brain learning," *International Journal of Innovation, Creativity and Change*, vol. 5, no. 1, pp. 189–200, 2019.

[3] A. Ortiz, P. López, J. L. Luque, F. J. Martínez-Murcia, D. Aquino-Britez, and J. Ortega, "An anomaly detection approach for dyslexia diagnosis using eeg signals," in *International Work-Conference on the Interplay Between Natural and Artificial Computation*. Springer, 2019, pp. 369–378.

[4] R. T. Canolty and R. T. Knight, "The functional role of cross-frequency coupling," *Trends in Cognitive Sciences*, vol. 14, no. 11, pp. 506–515, 2010.

[5] R. T. Canolty, C. F. Cadieu, K. Koepsell, R. T. Knight, and J. M. Carmena, "Multivariate phase–amplitude cross-frequency coupling in neurophysiological signals," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 1, pp. 8–11, 2011.

[6] M. Bonnefond, S. Kastner, and O. Jensen, "Communication between brain areas based on nested oscillations," *eNeuro*, vol. 4, no. 2, 2017.

[7] M. J. Hülsemann, E. Naumann, and B. Rasch, "Quantification of phase-amplitude coupling in neuronal oscillations: comparison of phase-locking value, mean vector length, modulation index, and generalized-linear-modeling-cross-frequency-coupling," *Frontiers in Neuroscience*, vol. 13, p. 573, 2019.

[8] J. Chen, P. Zhang, Z. Mao, Y. Huang, D. Jiang, and Y. Zhang, "Accurate eeg-based emotion recognition on combined features using deep convolutional neural networks," *IEEE Access*, vol. 7, pp. 44 317–44 328, 2019.

[9] J. E. Arco, A. Ortiz, D. Castillo-Barnes, J. M. Górriz, and J. Ramírez, "Ensembling shallow siamese architectures to assess functional asymmetry in alzheimer's disease progression," *Applied Soft Computing*, vol. 134, p. 109991, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1568494623000091

[10] T. T. K. Munia and S. Aviyente, "Time-frequency based phase-amplitude coupling measure for neuronal oscillations," *Scientific Reports*, vol. 9, pp. 1–15, 08 2019.

[11] S. Hirano, A. Nakhnikian, Y. Hirano, N. Oribe, S. Kanba, T. Onitsuka, M. Levin, and K. M. Spencer, "Phase-amplitude coupling of the electroencephalogram in the auditory cortex in schizophrenia," *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, vol. 3, no. 1, pp. 69–76, 2018.

[12] R. Goutagny, G. Ning, C. Cavanagh, J. Jackson, J.-G. Chabot, R. Quirion, S. Krantic, and S. Williams, "Corrigendum: Alterations in hippocampal network oscillations and theta-gamma coupling arise before aβ overproduction in a mouse model of alzheimer's disease," *European Journal of Neuroscience*, vol. 38, no. 9-10, 2013.

[13] N. J. Gallego-Molina, A. Ortiz, F. J. Martínez-Murcia, I. Rodríguez-Rodríguez, and J. L. Luque, "Assessing functional brain network dynamics in dyslexia from fnirs data." *International Journal of Neural Systems*, pp. 2 350 017–2 350 017, 2023.

[14] A. De Vos, S. Vanvooren, J. Vanderauwera, P. Ghesquière, and J. Wouters, "A longitudinal study investigating neural processing of speech envelope modulation rates in children with (a family risk for) dyslexia," *Cortex*, vol. 93, pp. 206–219, 05 2017.

[15] B. J. He, J. M. Zempel, A. Z. Snyder, and M. E. Raichle, "The temporal structures and functional significance of scale-free brain activity," *Neuron*, vol. 66, no. 3, pp. 353–369, 2010.

[16] M. Kramer and U. Eden, "Assessment of cross-frequency coupling with confidence using generalized linear models," *Journal of Neuroscience Methods*, vol. 220, no. 1, pp. 64–74, 2013.

[17] V. Lawhern, A. Solon, N. Waytowich, S. Gordon, C. Hung, and B. Lance, "Eegnet: A compact convolutional network for eeg-based brain-computer interfaces," *Journal of Neural Engineering*, vol. 15, 11 2016.

[18] S. U. Amin, M. Alsulaiman, G. Muhammad, M. A. Bencherif, and M. S. Hossain, "Multilevel weighted feature fusion using convolutional neural networks for eeg motor imagery classification," *IEEE Access*, vol. 7, pp. 18 940–18 950, 2019.

[19] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "Eegnet: a compact convolutional neural network for eeg-based brain–computer interfaces," *Journal of Neural Engineering*, vol. 15, no. 5, p. 056013, 2018.

[20] A. Vahid, A. Bluschke, V. Roessner, S. Stober, and C. Beste, "Deep learning based on event-related eeg differentiates children with adhd from healthy controls," *JCM*, vol. 8, p. 1055, 2019.

[21] K. Roots, Y. Muhammad, and N. Muhammad, "Fusion convolutional neural network for cross-subject eeg motor imagery classification," *Computers*, vol. 9, p. 72, 2020.

[22] M. D. Li, K. Chang, B. Bearce, C. Y. Chang, A. J. Huang, J. P. Campbell, J. M. Brown, P. Singh, K. V. Hoebel, D. Erdoğmuş *et al.*, "Siamese neural networks for continuous disease severity evaluation and change detection in medical imaging," *NPJ Digital Medicine*, vol. 3, no. 1, p. 48, 2020.

[23] J. E. Arco, A. Ortiz, N. J. Gallego-Molina, J. M. Górriz, and J. Ramírez, "Enhancing multimodal patterns in neuroimaging by siamese neural networks with self-attention mechanism," *International Journal of Neural Systems*, vol. 33, no. 04, p. 2350019, 2023.

[24] A. Giménez, J. L. Luque, M. López-Zamora, and M. Fernández-Navas, "A self-report questionnaire on reading-writing difficulties for adults.[autoinforme de trastornos lectores para adultos (atlas)]," *Anales de Psicología/Annals of Psychology*, vol. 31, no. 1, pp. 109–119, 2015.

[25] K. J. Miller, L. B. Sorensen, J. G. Ojemann, and M. den Nijs, "Power-law scaling in the brain surface electric potential," *PLOS Computational Biology*, vol. 5, no. 12, pp. 1–10, 12 2009.

[26] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber, "Lstm: A search space odyssey," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 10, pp. 2222–2232, 2016.

[27] M. A. Formoso, A. Ortiz, F. J. Martinez-Murcia, N. Gallego, and J. L. Luque, "Detecting phase-synchrony connectivity anomalies in eeg signals. application to dyslexia diagnosis," *Sensors*, vol. 21, no. 21, p. 7061, 2021.

[28] I. Rodríguez-Rodríguez, A. Ortiz, N. J. Gallego-Molina, M. A. Formoso, and W. L. Woo, "Eeg interchannel causality to identify source/sink phase connectivity patterns in developmental dyslexia," *International Journal of Neural Systems*, vol. 33, no. 04, p. 2350020, 2023.

[29] F. Richlan, "Developmental dyslexia: dysfunction of a left hemisphere reading network," *Frontiers in human neuroscience*, vol. 6, p. 120, 2012.

[30] P. Golland and B. Fischl, "Permutation tests for classification: Towards statistical significance in image-based studies," in *Information Processing in Medical Imaging*, C. Taylor and J. A. Noble, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2003, pp. 330–341.

[31] I. I. Andreadis, G. A. Giannakakis, C. Papageorgiou, and K. S. Nikita, "Detecting complexity abnormalities in dyslexia measuring approximate entropy of electroencephalographic signals," in *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2009, pp. 6292–6295.

[32] P. Perera, H. Harshani, M. F. Shiratuddin, K. W. Wong, and K. Fullarton, "Eeg signal analysis of writing and typing between adults with dyslexia and normal controls," *International Journal of Interactive Multimedia and Artificial Intelligence*, 2018.

[33] A. Frid and L. M. Manevitz, "Features and machine learning for correlating and classifying between brain areas and dyslexia," *arXiv preprint arXiv:1812.10622*, 2018.

[34] P. Tamboer, H. Vorst, S. Ghebreab, and H. Scholte, "Machine learning and dyslexia: Classification of individual structural neuro-imaging scans of students with and without dyslexia," *NeuroImage: Clinical*, vol. 11, pp. 508–514, 2016.

[35] S. Zahia, B. Garcia-Zapirain, I. Saralegui, and B. Fernandez-Ruanova, "Dyslexia detection using 3d convolutional neural networks and functional magnetic resonance imaging," *Computer Methods and Programs in Biomedicine*, vol. 197, p. 105726, 2020.

[36] N. J. Gallego-Molina, A. Ortiz, F. J. Martínez-Murcia, and I. Rodríguez-Rodríguez, "Unraveling dyslexia-related connectivity patterns in eeg signals by holo-hilbert spectral analysis," in *Artificial Intelligence in Neuroscience: Affective Analysis and Health Applications*, J. M. Ferrández Vicente, J. R. Álvarez-Sánchez, F. de la Paz López, and H. Adeli, Eds. Springer International Publishing, 2022, pp. 43–52.

[37] M. A. Formoso, A. Ortiz, F. J. Martínez-Murcia, D. A. Brítez, J. J. Escobar, and J. L. Luque, "Temporal phase synchrony disruption in dyslexia: Anomaly patterns in auditory processing," in *Artificial Intelligence in Neuroscience: Affective Analysis and Health Applications*, J. M. Ferrández Vicente, J. R. Álvarez-Sánchez, F. de la Paz López, and H. Adeli, Eds. Springer International Publishing, 2022, pp. 13–22.

[38] N. J. Gallego-Molina, A. Ortiz, M. A. Formoso, F. J. Martínez-Murcia, and W. L. Woo, "Enhancing neuronal coupling estimation by nirs/eeg integration," in *Artificial Intelligence for Neuroscience and Emotional Systems*, J. M. Ferrández Vicente, M. Val Calvo, and H. Adeli, Eds. Springer Nature Switzerland, 2024, pp. 24–33.

[39] J. E. Arco, N. J. Gallego-Molina, P. J. López-Pérez, J. Ramírez, J. M. Górriz, and A. Ortiz, "Extracting heart rate variability from nirs signals for an explainable detection of learning disorders," in *Artificial Intelligence for Neuroscience and Emotional Systems*, J. M. Ferrández Vicente, M. Val Calvo, and H. Adeli, Eds. Springer Nature Switzerland, 2024, pp. 118–127.

[40] F. J. Martinez-Murcia, J. M. Gorriz Sáez, A. Ortiz, J. Ramírez Pérez De Inestrosa, P. J. Lopez-Abarejo, M. López Zamora, and J. L. Luque, "Eeg connectivity analysis using denoising autoencoders for the detection of dyslexia," 5 2020. [Online]. Available: https://hdl.handle.net/10481/80026

[41] N. J. Gallego-Molina, A. Ortiz, F. J. Martínez-Murcia, M. A. Formoso, and A. Giménez, "Complex network modeling of eeg band coupling in dyslexia: An exploratory analysis of auditory processing and diagnosis," *Knowledge-Based Systems*, vol. 240, p. 108098, 2022.

[42] D. Castillo-Barnes, N. J. Gallego-Molina, M. A. Formoso, A. Ortiz, P. Figueiredo, and J. L. Luque, "Probabilistic and explainable modeling of phase–phase cross-frequency coupling patterns in eeg. application to dyslexia diagnosis," *Biocybernetics and Biomedical Engineering*, vol. 44, no. 4, pp. 814–823, 2024.

[43] Z. Rezvani, M. Zare, G. Žarić, M. Bonte, J. Tijms, M. Van der Molen, and G. Fraga González, "Machine learning classification of dyslexic children based on eeg local network features," *BioRxiv*, p. 569996, 2019.

[44] T. E. Özkurt, "Statistically reliable and fast direct estimation of phase-amplitude cross-frequency coupling," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 7, pp. 1943–1950, 2012.

[45] C. Huang and P. Lin, "Effects of symbol component on the identifying of graphic symbols from eeg for young children with and without developmental delays," *Applied Sciences*, vol. 9, p. 1260, 2019.

[46] M. Wöstmann, B. Herrmann, B. Maess, and J. Obleser, "Spatiotemporal dynamics of auditory attention synchronize with speech," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 113, pp. 3873–3878, 2016.

[47] I. Rodríguez-Rodríguez, A. Ortiz, N. J. Gallego-Molina, M. A. Formoso, and W. L. Woo, "Eeg interchannel causality to identify source/sink phase connectivity patterns in developmental dyslexia," *International Journal of Neural Systems*, vol. 33, no. 04, p. 2350020, 2023.

[48] O. Sporns and C. J. Honey, "Small worlds inside big brains," *Proceedings of the National Academy of Sciences*, vol. 103, no. 51, pp. 19 219–19 220, 2006.

[49] E. Zaccarella, G. Papitto, and A. D. Friederici, "Language and action in broca's area: Computational differentiation and cortical segregation," *Brain and Cognition*, vol. 147, p. 105651, 2021.

[50] M. Gavrilescu, S. Rossell, G. Stuart, T. Shea, H. Innes-Brown, K. Henshall, C. McKay, A. Sergejew, D. Copolov, and G. Egan, "Reduced connectivity of the auditory cortex in patients with auditory hallucinations: A resting state functional magnetic resonance imaging study," *Psychological Medicine*, vol. 40, pp. 1149 – 1158, 2010.

[51] U. Goswami, "A temporal sampling framework for developmental dyslexia," *Trends in cognitive sciences*, vol. 15, no. 1, pp. 3–10, 2011.

[52] G. Schulte-Körne and J. Bruder, "Clinical neurophysiology of visual and auditory processing in dyslexia: A review," *Clinical neurophysiology*, vol. 121, no. 11, pp. 1794–1809, 2010.

[53] B. A. Shaywitz, S. E. Shaywitz, B. A. Blachman, K. R. Pugh, R. K. Fulbright, P. Skudlarski, W. E. Mencl, R. T. Constable, J. M. Holahan, K. E. Marchione *et al.*, "Development of left occipitotemporal systems for skilled reading in children after a phonologically-based intervention," *Biological psychiatry*, vol. 55, no. 9, pp. 926–933, 2004.

[54] P. Tallal, "Improving language and literacy is a matter of time," *Nature Reviews Neuroscience*, vol. 5, no. 9, pp. 721–728, 2004.

[55] K. Lehongre, F. Ramus, N. Villiermet, D. Schwartz, and A.-L. Giraud, "Altered low-gamma sampling in auditory cortex accounts for the three main facets of dyslexia," *Neuron*, vol. 72, no. 6, pp. 1080–1090, 2011.

[56] H. Luo, Y. Wang, D. Poeppel, and J. Z. Simon, "Concurrent encoding of frequency and amplitude modulation in human auditory cortex: Meg evidence," *Journal of neurophysiology*, vol. 96, no. 5, pp. 2712–2723, 2006.

[57] J. A. Hämäläinen, H. K. Salminen, and P. H. Leppänen, "Basic auditory processing deficits in dyslexia: systematic review of the behavioral and event-related potential/field evidence," *Journal of learning disabilities*, vol. 46, no. 5, pp. 413–427, 2013.