Decoding the Developing Brain

An EEG-based Functional Connectivity Analysis in Pediatric Multiple Sclerosis

I.L. de Wit

TU Delft Delft University of Technology

Erasmus MC University Medical Center Rotterdam

Decoding the Developing Brain

An EEG-Based Functional Connectivity Analysis in Pediatric Multiple Sclerosis

by

I.L. (Ilse) de Wit

In Partial Fulfilment of the Requirements of
Master of Science in Biomedical Engineering
Track Neuromusculoskeletal Biomechanics
at the Delft University of Technology,
to be defended publicly on Thursday July 10, 2025 at 15:30.

Student number:	5065046	
Project duration:	March 1, 2025 – July 10, 2025	
Thesis advisors:	Dr. ir. M.L. van de Ruit	
	Dr. R. van den Berg	
	M. Verboom, MSc	
Thesis committee:	Dr. ir. M.L. van de Ruit (Chair)	Delft University of Technology
	Dr. R. van den Berg	Erasmus MC
	Prof. dr. ir. A.C. Schouten	Delft University of Technology

TU Delft

Abstract

Children with multiple sclerosis (MS) frequently have reduced visual processing speed, which can affect learning, social interaction and daily functioning. Identification of cognitive impairment at an early stage is essential for timely intervention. This study explores if electroencephalography (EEG)-based functional connectivity (FC) can be utilized as a non-invasive biomarker for visual processing speed in children with MS.

Resting state EEG data and neuropsychological assessments were analyzed from ten children diagnosed with MS. A custom preprocessing pipeline was developed and validated to ensure data quality for FC analysis. Connectivity metrics, including fronto-occipital and interhemispheric connectivity, network efficiency based on minimum spanning tree analysis, and individual alpha peak frequency, were extracted from source-reconstructed EEG. Correlations were tested between these metrics and visual processing speed, measured with the Processing Speed Index of the Wechsler Intelligence Scale for Children.

As hypothesized, trends toward greater network efficiency and stronger interhemispheric and fronto-occipital connectivity were associated with higher visual processing speed. Although none of these associations reached statistical significance, the results support the potential of EEG-based functional connectivity to evolve into a clinically relevant biomarker, enabling early diagnosis, personalized treatment and improved prognosis for children with MS.

Contents

1

Introduction

The processing speed of visual information is a critical cognitive function that evolves throughout childhood and influences essential activities such as reading, learning, and social interaction (Siu and Murphy, 2018, Reigal et al., 2019). Visual processing speed is recognized as a key measure of brain efficiency, reflecting the capacity to perceive, process, and retrieve information (Neubauer and Fink, 2003). Impairments in visual processing speed are commonly observed in various neurodevelopmental and neurological disorders including multiple sclerosis (MS) (Koller, 2012).

MS is a chronic disease of the central nervous system characterized by immune-mediated damage to the myelin sheath surrounding the nerve fibers (Thompson et al., 2018). This protective layer ensures efficient signal transmission between neurons. Inflammatory lesions, demyelination, and eventual axonal damage disrupt communication between the brain and body, leading to various physical and cognitive symptoms. Although MS is typically diagnosed in adults, a subset of patients experience symptom onset during childhood or adolescence. Pediatric MS is associated with early neuroinflammatory activity, which can alter white matter development and interfere with efficient brain communication (Ahmed et al., 2022). Cognitive impairment affects up to one-third of children with MS, with visual processing speed being one of the most frequently affected domains (Portaccio et al., 2021).

The capacity to process visual information rapidly and accurately is determined by efficient communication between different areas of the brain (Johnson, 2003). This communication can be measured using electroencephalography (EEG), a noninvasive technique to record brain activity (Schomer and da Silva, 2017). Electrodes are placed on the scalp to measure the voltage fluctuations resulting from postsynaptic potentials in large populations of synchronously active neurons. EEG provides high temporal resolution and is commonly used to study brain function in both clinical and research settings. Functional connectivity (FC), which can be measured using EEG, represents the synchronization of activity patterns between different areas of the brain over time (Friston et al., 1993). These synchronized activity patterns are associated with the rapid exchange of information across the segregated brain network and are thought to reflect cooperation between specialized brain areas involved in cognitive, sensory, and perceptual processing (Mišić et al., 2014, Bastos and Schoffelen, 2015).

Functional connectivity has been studied in patients with MS using functional magnetic resonance imaging (fMRI). Multiple studies have revealed altered network topologies, such as decreased global efficiency and disruption of interhemispheric communication (Rocca et al., 2022, Tahedl et al., 2018). These changes are believed to reflect the cumulative effects of lesion formation and neurodegeneration. Akbar et al. explored alterations in FC in pediatric MS patients compared to typically developing peers, demonstrating that a loss of the microstructural integrity of white matter is associated with increased functional connectivity in the resting state, potentially reflecting diffuse and compensatory network activation in early stages of the disease (Akbar et al., 2016). EEG and magnetoencephalography (MEG) have been applied less frequently in MS research than fMRI. Available studies have reported differences in resting-state network dynamics between patients and healthy controls (Schoonheim et al., 2013 and Shirani and Mohebbi, 2022). However, these results remain variable, probably because of lesion heterogeneity. Nevertheless, alterations in network topology have emerged as a consistent finding across modalities. Healthy controls often have more efficient and segregated brain networks. Although network topology potentially reflects changes in the brain's capacity to process information

efficiently, these metrics are less frequently analyzed within the patient group. Recent work by Bardel et al. (Bardel et al., 2024) linked FC patterns to motor symptoms and treatment response. However, the role of EEG-based FC in explaining cognitive deficits such as impaired visual processing speed remains underexplored. Particularly in pediatric MS, the potential of EEG-derived connectivity metrics as markers of cognitive function has not yet been systematically investigated.

Understanding the functional connectivity patterns underlying visual processing speed in children with MS may have important clinical implications. Early detection of cognitive impairments, such as reduced visual processing speed, can enable timely intervention to support brain development during critical periods of maturation (Amato et al., 2008). Studies have shown that early treatment in pediatric-onset MS, particularly with high-efficacy therapies, can help preserve neurological function and reduce the risk of progression to higher levels of long-term disability (Sharmin et al., 2024). EEG-based biomarkers could complement clinical assessments by identifying children at risk for cognitive decline before these symptoms become behaviorally apparent (Bardel et al., 2024 and Akbar et al., 2016). Moreover, characterizing individual patterns of disrupted connectivity can inform targeted rehabilitation strategies tailored to a child's specific needs, such as cognitive training (Sumowski et al., 2018). Ultimately, improving diagnostic precision and enabling earlier and more personalized treatment could help slow the progression of symptoms, improve daily functioning, and support academic and social participation in children with MS.

This study aims to identify EEG-based markers of cognitive impairment in pediatric multiple sclerosis by examining the relationship between functional connectivity and visual processing speed. Based on a previous literature review (de Wit, 2025), it is hypothesized that stronger interhemispheric and fronto-occipital connectivity, a more developed alpha frequency profile and higher network efficiency are associated with higher visual processing speed. An automated preprocessing pipeline tailored to pediatric EEG data is needed to ensure a valid estimation of FC metrics, as these metrics are highly sensitive to non-neural artifacts that could mask the small neural signal and lead to incorrect conclusions (Chiarion et al., 2023). By implementing a custom preprocessing pipeline and performing a pilot study focusing on resting state functional connectivity, this study aims to contribute to the development of EEG-based biomarkers that support early diagnosis, personalized treatment strategies, and improved clinical outcomes for children with MS.

2

Methodology

2.1. Participants

The data used in this study is obtained from the longitudinal database of the Child Brain Lab (CBL) at the Erasmus Medical Center in Rotterdam. The dataset consists of children with brain disorders with an age from 0 to 18 years. All participants underwent neuropsychological and physical assessments and EEG measurements at five time points: 0–12 months, 30–42 months, 5–7 years, 9–13 years and 15–18 years. In some cases, multiple measurements were taken for a participant at a time point when the first measurement was of insufficient quality. In this study, all measurements of participants diagnosed with MS from the start of the CBL in May 2023 to May 2025 are included if both EEG and eye-tracking recordings to mark different stages of the measurement were available.

2.2. Data acquisition

2.2.1. EEG recording

The EEG was recorded using the CBL protocol, which uses an interactive avatar on a video screen that guides the participant through the EEG recording. The protocol is adapted for every age group and consists of six tasks: resting state, facial processing, social attention, go/no-go, sequence learning and matrix reasoning. For this study, only the resting state recording is used, which was usually recorded at the beginning of the session.

Two variants are available for the resting state protocol:

- RS1.0: No specific instructions are given before the task to the youngest age group (0-12 months old). During the task, the avatar shows whether the eyes should be open or closed. Ideally, children will mimic this. The EEG technician or caregiver will try to keep the child calm and awake for optimal recording. Approximately five minutes of resting state EEG is recorded for each participant.
- RS2.0: For the other age groups (>12 months old) the avatar explains to the child that in the upcoming test stage, he or she has to open and close their eyes. They close their eyes for 30 seconds and open them for 15 seconds. This cycle is repeated five times. The protocol can be extended if necessary.

For the EEG recording, a 128-channel HydroCel Geodesic Sensor Net (GSN) with two Nuevo 64-channel amplifiers was used. The EEG was recorded at a sample frequency of 1000 Hz and electrode 55 (located near Cz) served as the reference electrode. The time points of the resting state recording and the opening and closing of the eyes were annotated during the recording.

2.2.2. Data quality

After visual inspection, a clinical neurophysiologist assigned a quality score to each EEG recording on a four-point scale ranging from zero to three. A score of 0 indicates that the recording is unusable because of excessive artifacts; 1 indicates that the recording should be used with caution because of severe artifacts; 2 indicates moderate artifacts; and 3 indicates good quality with few artifacts.

2.2.3. Visual processing speed

Visual processing speed was measured using the Processing Speed Index (PSI) total score from the Wechsler Intelligence Scale for Children (WISC-V)(Wechsler, 2014). The WISC-V is part of the neuropsychological assessment that children receive at the CBL. The WISC-V is an intelligence test for children and adolescents aged 6–16 years that indicates overall intelligence. The PSI is one of the five primary indexes of this test and it is measured with two subtests: symbol search and coding.

- Symbol search: The child is asked to say if a certain symbol is present in a row of symbols as fast as they can.

- Coding: The child is asked to translate a number sequence into a sequence of symbols as fast as they can with a translation key which is given prior to the task.

The PSI score used in this study is the sum of the scores of these subtests and is an objective measure of how fast and accurately a child can process visual information and perform such tasks.

2.3. Preprocessing

To ensure sufficient data quality for reliable connectivity analysis, a custom preprocessing pipeline is developed, validated and applied. The first step in the preprocessing pipeline is to prepare the data for ICA decomposition. Only EEG segments during the test conditions are retained. Explanation or practice periods are excluded. Segments containing extreme amplitudes (exceeding 150 µV) and channels identified as bad are also excluded from the ICA fitting. Next, the data is high-pass filtered at 1 Hz and re-referenced to the average. ICA decomposition is then performed using the Infomax algorithm, with the maximum number of components determined by the number of retained channels to effectively separate neural sources from non-neural artifacts. Artifact components are identified using the ICLabel extension and subsequently removed from the original raw data. Following artifact removal, further preprocessing is applied. Specifically, bad channels are identified, after which standard filtering steps are applied: low-pass filtering at 80 Hz, downsampling to 250 Hz, notch filtering at 50 and 100 Hz, and high-pass filtering at 0.1 Hz. The signal is then re-referenced to the average and the previously identified bad channels are interpolated.

This pipeline is constructed using a stepwise approach, comparing concept solutions using established performance metrics to ensure consistency across evaluations and support a robust pipeline development process. A full explanation of the choices made in the development process is provided in Appendix A. The preprocessing pipeline is developed in Python because of the open-source nature of the language and the growing use of Python in EEG research, making it a valuable platform for standardized EEG analysis. Additionally, Python aligns with the existing work of the research group and therefore using Python facilitates integration, reproducibility, and future collaboration.

To evaluate the pipeline before further analysis, several quantitative and functional validation metrics are used. Signal preservation is assessed using the signal error ratio (SER) and the correlation between the cleaned and original signal, whereas artifact removal is evaluated using the artifact-to-residual ratio (ARR) and mutual information between the removed and cleaned signal. In addition, the improvement in data quality is measured and the strength of the Berger effect (the increase in alpha power when the eyes are closed) and visual inspection by an experienced neurophysiologist served as functional validation indicators. A more detailed description of the performance metrics is provided in Appendix B. The validation of the preprocessing is conducted on the full EEG dataset of the CBL, which includes participants across diagnostic groups, to ensure that the evaluation was based on a large and representative sample. Because the current sample is a subset of this dataset, the obtained validation results are considered representative for the data used in this study.

2.4. Data processing

All data processing and analysis are performed using Python. The full scripts and implementation details are available in the GitHub repository (https://github.com/Ilse2001/code-repository-msc-thesis-eeg-fc).

2.4.1. Source reconstruction

Source reconstruction is applied to obtain anatomically meaningful measures of functional connectivity. First, the data is epoched into 1-second segments across the entire resting-state recording. Epochs containing extreme amplitudes above 150 μV are discarded. Source reconstruction is performed using the exact low resolution electromagnetic tomography (eLORETA) method implemented in MNE-Python (Gramfort et al., 2013), which assumes that brain activity is spatially distributed over larger cortical regions rather than originating from a single point. This results in a solution with a lower spatial resolution but improved stability and no systematic localization errors under ideal conditions. The forward model is based on the average Magnetic Resonance Imaging (MRI) template available in the MNE sample data. Although the MRI template may not be fully representative of younger children, it provides an initial step to explore FC patterns. The last input variable, the noise covariance matrix, is estimated from one minute of recording before the actual tests using the shrunk method, which is a regularization technique that combines the covariance estimate from the data with a standard covariance matrix to improve robustness and prevent overfitting. The inverse solution is computed and applied to the resting state recording.

To reduce the dimensionality of the source-level data and allow for anatomical interpretation, time series were extracted for cortical regions defined by the Desikan-Killiany atlas (Desikan et al., 2006). For each region (label), the source estimates were averaged using the mean flip method, which accounts for dipole orientation by aligning the sign of the source signals before averaging. This approach prevents signal cancellation by opposite polarities and results in one representative time series for each region and epoch.

2.4.2. Functional connectivity analysis

The phase lag index (PLI) between all pairs of regions is computed to assess FC. The PLI is selected as the FC metric based on a comprehensive literature review comparing commonly used functional connectivity methods in pediatric EEG research (de Wit, 2025). PLI is chosen because of its robustness to volume conduction, as it only considers consistent phase lags that are non-zero, thereby minimizing the influence of spurious connectivity caused by common sources (Stam et al., 2007). These measures are computed using the available functions in MNE-Python, specifically spectral_connectivity for the PLI. FC matrices are computed separately for eyes open and closed resting state conditions and for the five canonical frequency bands (delta: 1-4 Hz, theta: 4-8 Hz, alpha: 8-12 Hz, beta: 12-25 Hz, gamma: 25-50 Hz) as well as for the broadband spectrum (1-50 Hz) using a multitaper method. Connectivity matrices are computed for each epoch and averaged across epochs to obtain a single connectivity matrix per condition, participant and frequency band. To facilitate region-level analyses, labels are grouped in two ways: first into anatomical units including the cortical lobes (occipital, frontal, parietal and temporal), limbic system and insula, and second by hemisphere. Mean connectivity is computed both between the anatomical regions (lobe-to-lobe) and between and within hemispheres to assess inter- and intrahemispheric connectivity.

As a quality check for the FC analysis, alpha band connectivity across the anatomical regions and hemispheres is compared between the eyes open and eyes closed resting state conditions. This comparison tests whether the FC analysis captures known neurophysiological effects, specifically the Berger effect.

2.4.3. Network analysis

The network topology is examined using minimum spanning tree (MST) analysis. This method allows for an unbiased comparison of the network structure by reducing the graph to its most essential connections (Tewarie et al., 2015). The MST is chosen for its suitability in inter-subject comparisons based on a comprehensive literature review prior to this research comparing the current literature methods and findings on functional connectivity in pediatric populations (de Wit, 2025). For each participant and condition, the inverse of the connectivity matrix (1 – PLI) is used as a distance matrix to construct the MST with the networkx.minimum_spanning_tree function, ensuring an unbiased and cycle-free representation of the network. Graph-theoretical metrics, including diameter and leaf fraction, are extracted from each MST and used to assess potential changes in network efficiency and integration between participants.

2.4.4. Individual alpha peak frequency (iAPF)

The iAPF is visually estimated by a neurophysiologist based on the presence of a stable alpha rhythm in the raw EEG during the eyes-closed condition. The frequency is determined by counting the number of alpha cycles per second in the occipital segments with clearly identifiable rhythmic activity.

2.5. Statistical analysis

2.5.1. Validation of EEG functional connectivity: Berger effect

Paired sample t-tests are used to determine whether these lobe-to-lobe and hemispheric connectivity patterns in the alpha band differ significantly between the eyes open and eyes closed resting state conditions. To minimize the risk of multiple comparisons, the mean connectivity of the occipital lobes with the other lobes and the interhemispheric connectivity in the alpha band are treated as primary outcomes.

2.5.2. Correlations between connectivity and visual processing speed

The relations of iAPF, frontal-occipital connectivity, interhemispheric connectivity and MST diameter and leaf fraction with PSI are tested using Spearman's rank correlation. This analysis is performed for both eyes open and eyes closed resting state (except for the iAPF). Multiple testing is corrected using Bonferroni correction based on the number of potential biomarkers. Because five biomarkers are tested, the adjusted significance threshold is set at $p < 0.01$.

3

Results

3.1. Data characteristics

In total, 13 EEG recordings were available from children diagnosed with MS. The participant characteristics are summarized in Table 3.1.

Participant	Number of measurements	Quality of measurement(s)	Age at most recent measurement (years (months))	PSI
1	2	0, 1	15 (0)	23
2	1	0	17 (1)	17
3	1	1	15 (1)	26
4	1	2	10 (5)	22
5	2	0, 1	17 (2)	21
6	2	0, 2	5 (4)	-
7	1	3	17 (5)	-
8	1	0	17 (4)	16
9	1	1	15 (0)	29
10	1	3	10 (10)	14

Table 3.1: Overview of participant characteristics. Measurement quality was rated by a pediatric neurologist on a 4-point scale: 0 = unusable due to artifacts, 1 = severe artifacts, 2 = moderate artifacts, and 3 = good quality with few artifacts. Multiple values correspond to multiple recordings. PSI scores (Processing Speed Index) were derived from the Wechsler Intelligence Scale for Children – Fifth Edition (WISC-V); a dash indicates that no score was available.

3.2. Preprocessing performance

The average cleaning performance of the preprocessing pipeline is assessed with the following measurements: SER = 2.4 (SD = 3.5), ARR = 17.1 (SD = 14.1), correlation between raw and cleaned data = 0.56 (SD = 0.21), and mutual information between the removed and cleaned signal = 0.56 (SD = 0.26). On average, preprocessing resulted in an absolute increase of 10% in the proportion of good-quality segments per recording. More details on the validation results of the preprocessing pipeline, including visual inspection and quality comparisons, are included in Appendix A, Section A.6.

3.3. Validation of EEG functional connectivity: Berger effect

Paired t-tests revealed no significant difference in occipital alpha connectivity between the eyes open and eyes closed resting state conditions ($t(13) = -0.91$, $p = 0.38$), nor in interhemispheric alpha connectivity ($t(13) = -0.39$, $p = 0.71$).

In a subset of participants with a pronounced Berger effect in the spectral domain (participants 3, 4, 7, and 9), visual inspection suggested a reduction in long-range alpha connectivity during the eyes-open condition, particularly in the occipital lobe. Figure 3.1 shows this pattern for participant 3.

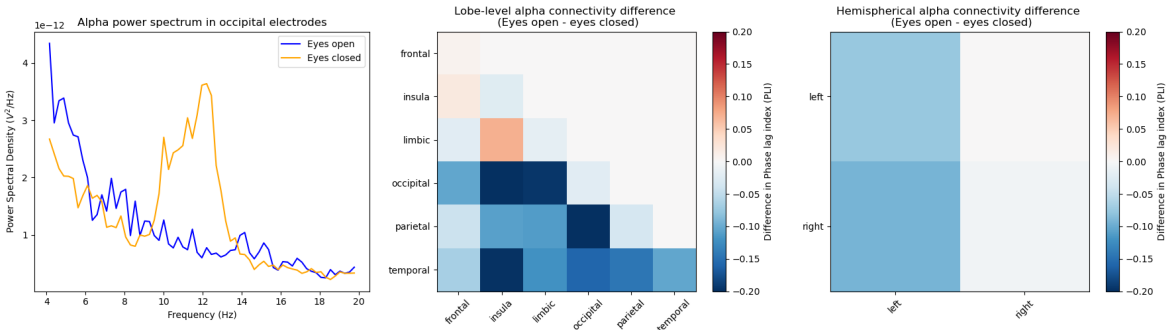

Figure 3.1: Differences in spectral alpha power (left) and connectivity between the lobes (middle) and hemispheres (right) for the eyes open and closed resting state conditions for participant 3.

3.4. Correlations between connectivity and visual processing speed

For eight participants, a PSI score from the WISC-V was available. Five participants were tested at the CBL and scores from three participants were requested from other medical facilities. Figure 3.2 shows scatter plots for both eyes open and eyes closed conditions, illustrating the associations between PSI and four functional connectivity measures: frontal-occipital connectivity, interhemispheric connectivity and diameter and leaf fraction of the MST. The association between iAPF and PSI is shown in Figure 3.3. Two EEG recordings (participants 5 and 8) were of insufficient quality to assess the iAPF and therefore are not included.

Frontal-occipital and interhemispheric connectivity both showed a positive association with PSI in both eyes open and eyes closed conditions. These associations had a higher Spearman's ρ in the eyes closed condition. The MST metrics showed a consistent pattern across the two conditions: higher PSI scores were associated with smaller diameters and higher leaf fractions. No visual trend was observed in the scatter plot illustrating the association between iAPF and PSI.

Spearman correlation coefficients and corresponding p-values are shown in the figures. None of the correlations reached statistical significance after applying the adjusted threshold of $p < 0.01$.

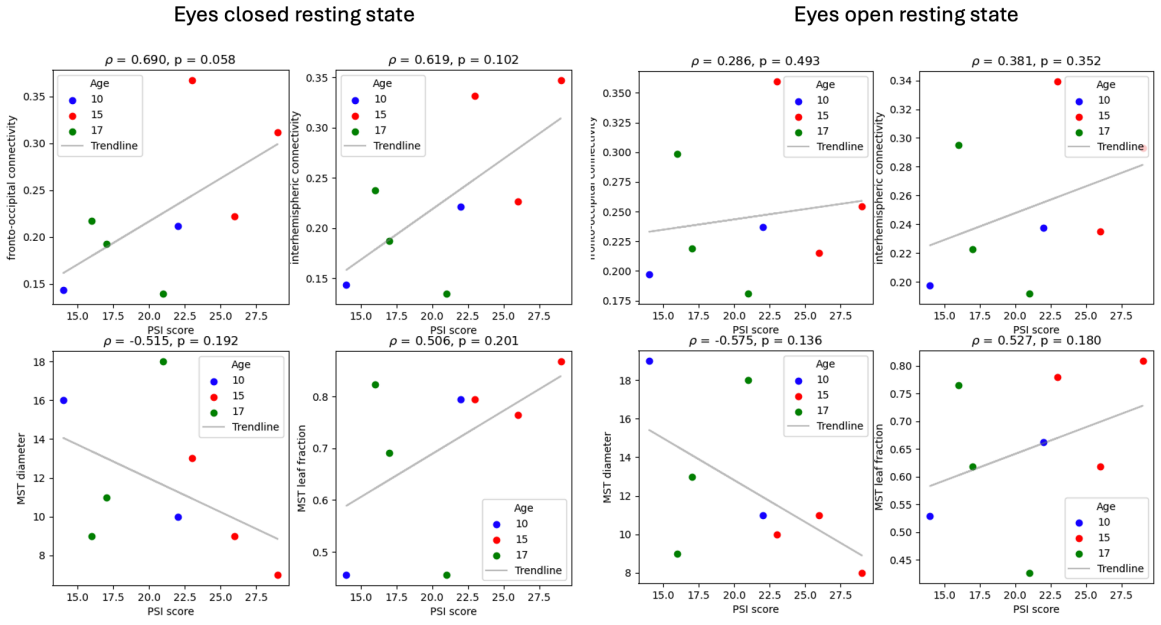

Figure 3.2: Correlations between functional connectivity or network metrics and processing speed index (PSI), measured during eyes-closed and eyes-open resting states. Each scatter plot represents one EEG-derived measure: fronto-occipital connectivity, interhemispheric connectivity, minimum spanning tree (MST) diameter, or MST leaf fraction. Colors indicate participant age (10, 15 and 17 years), and trendlines are included to visualize the association.

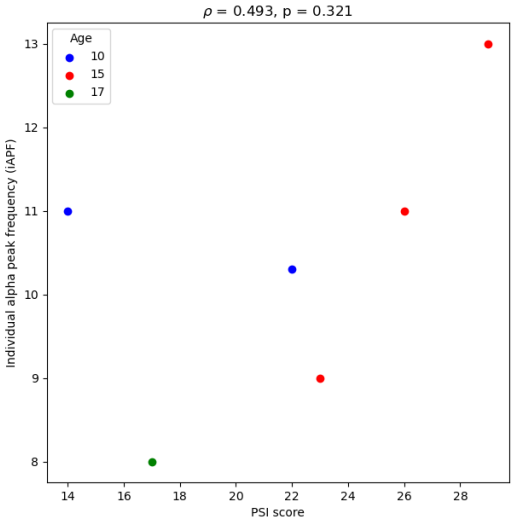

Figure 3.3: Correlation between individual alpha peak frequency (iAPF) and processing speed index (PSI). Each dot represents a participant, colored according to the age group. No clear association was observed.

4 Discussion

4.1. Interpretation of results

This study aimed to identify EEG-based markers of cognitive impairment in pediatric multiple sclerosis by analyzing the relations between resting state FC, network efficiency, and iAPF with visual processing speed. The results suggest that network efficiency metrics, particularly those derived from MST analysis, show the most consistent associations with visual processing speed, indicating their potential as markers for cognitive impairment. While fronto-occipital and interhemispheric FC also showed associations with processing speed, these effects were less robust. In contrast, iAPF was not associated with visual processing speed in the current sample.

To ensure the validity of the EEG analyses, the quality of preprocessing and FC estimation was evaluated prior to the interpretation of the results. The quantitative quality metrics indicated an overall improvement in data quality after preprocessing, with effective artifact removal and sufficient signal preservation. The SER indicated that the neural signal power was adequately retained and the correlation between the original and cleaned signals demonstrated temporal coherence. The ARR confirmed a substantial power reduction of the noise components and the mutual information scores indicated the successful removal of the identified artifacts. Although standardized benchmarks for these metrics are lacking in the literature, the combination of these quantitative results with visual inspection and the strong expression of the inhibitory alpha rhythm in the eyes closed condition supports the reliability of the cleaned data. A more detailed discussion on the development, validation, limitations and implications of the preprocessing pipeline is provided in Appendix A, Section A.7.

Subsequently, the validity of the source-space FC analysis was further supported by the detection of the Berger effect. A decrease in alpha-band FC was observed during the eyes open condition compared to the eyes closed condition, particularly in participants who also showed increased posterior alpha power during eyes-closed. This inhibitory alpha signal is associated with an increase in occipital connectivity that arises when the visual cortex is no longer processing external input and defaults to a diffuse resting rhythm. In contrast, when the eyes are open and visual processing resumes, both alpha power and FC decrease as neural activity becomes more spatially differentiated. These findings indicate that the FC measures were sensitive to meaningful changes in brain state, supporting the validity of their use in subsequent correlation analyses.

Among the connectivity-derived metrics, MST-based network measures showed the most consistent association with visual processing speed. Although not significant, both lower diameter and higher leaf fraction (complementary indicators of network efficiency) showed a trend with higher PSI. This pattern was similar in both eyes open and eyes closed conditions, making the finding more robust. These findings are in line with MEG research in adults with MS, which showed that cognitive impairment was linked to disrupted alpha-band network topology, particularly a loss of hierarchical organization (Tewarie et al., 2015). Additionally, the study reported group differences between patients with MS and healthy controls, suggesting that network topology may not only reflect individual cognitive capacity but also serve as a marker of disease-related neural disruption. Together, these findings support the idea that network-level measures, such as MST-derived diameter and leaf fraction, may be useful biomarkers for identifying and monitoring cognitive impairment in pediatric MS.

Beyond global network topology, the FC between specific brain regions also showed trends in the hypothesized direction, particularly in the eyes-closed condition. Although the associations did not reach statistical significance, stronger interhemispheric and fronto-occipital connectivity showed a trend with faster visual processing. These connectivity measures have also been linked to cognitive performance in the literature. For example, Charlebois-Poirier et al. (Charlebois-Poirier et al., 2025) found that fronto-parietal and anterior–posterior FC patterns were associated with distinct domains of cognitive functioning, including processing speed, in typically developing children. Similarly, Lin et al. (Lin et al., 2020) reported reduced interhemispheric connectivity in adults with MS using fMRI, with strong associations to slower processing speed. This reduction was hypothesized to reflect inflammatory damage to the corpus callosum, a key structure involved in interhemispheric communication. In contrast, Akbar et al. reported increased anterior–posterior FC in pediatric MS compared to healthy controls, which has been interpreted as a compensatory mechanism for disease-related disruptions (Akbar et al., 2016). Although the current study did not include a control group, the observed trend of higher FC in the MS group being associated with faster visual processing may support this interpretation. Children with more effective compensatory FC may process visual information faster. These comparisons with the literature underscore the importance of interpreting FC in specific clinical and methodological contexts. Additional context is essential for meaningful interpretation of observed associations and to better understand the underlying mechanisms. This is especially relevant when aiming to use FC as a biomarker for clinical monitoring or prediction, as its utility depends on a clear understanding of the observed changes. Although the current study did not find statistically significant correlations, the observed trends support the potential of interhemispheric and fronto-occipital FC as promising biomarkers for individual differences in visual processing speed and potentially useful for characterizing cognitive functioning in pediatric MS. Further research is needed to understand the underlying mechanisms and improve the use of FC as a biomarker.

Finally, in contrast to the hypothesis, no consistent association was found between iAPF and visual processing speed. This is notable given that prior research has linked iAPF to cognitive performance, brain maturation, and network organization in healthy individuals (Klimesch, 1999)(Jin et al., 2006)(Grandy et al., 2013). Some studies have reported a modest positive relationship between iAPF and processing speed, but not with broader measures of cognition or intelligence (Lin et al., 2020). In the current sample, variability in iAPF did not correspond to PSI, suggesting that its utility as a cognitive marker may be limited in pediatric MS. This result highlights the ongoing debate in the literature regarding the functional significance of iAPF, particularly in clinical populations. While iAPF remains a theoretically appealing biomarker due to its developmental and neurophysiological underpinnings, the current findings do not support its use as a biomarker for cognitive impairment in pediatric MS.

4.2. Limitations

This study has several limitations that should be considered when interpreting the results. First, the relatively small sample size limits statistical power, particularly for correlation analyses that are sensitive to individual variability and outliers, thereby reducing the generalizability and robustness of the identified biomarkers. Pediatric MS is a heterogeneous condition with variations in disease duration and lesion burden. These sources of inter-individual variability may obscure subtle associations between EEG metrics and cognitive performance. In addition, the small sample size did not allow for the statistical control of potential confounders, such as medication use, fatigue, or symptom severity, which may influence EEG-derived connectivity and cognitive outcomes. Despite these limitations, the study revealed consistent trends, supporting the feasibility of functional connectivity measures as biomarkers of visual processing speed in pediatric MS.

Second, several methodological limitations apply to the processing of EEG data. The preprocessing pipeline was developed and validated using quantitative metrics and functional tests. However, in the absence of a ground truth signal or objective benchmark, determining the optimal balance between artifact suppression and signal preservation remains difficult. In particular, subtle neural signals are at risk of being incorrectly removed or overshadowed by residual noise. Similarly, source reconstruction relied on an adult MRI template, which reduces anatomical specificity, especially in younger children, whose brain structure diverges from adult norms. Although the effect of this mismatch is likely less pronounced in older children and may be less critical for global metrics, such as long-range connectivity, it introduces additional uncertainty in spatial localization. These limitations may affect the accuracy

of functional connectivity measures, specifically fronto-occipital and interhemispheric connections, potentially weakening the observed correlations with cognitive outcomes. However, because this study focuses on global network measures, such influence is expected to be limited. Future research aiming to test more specific hypotheses should prioritize improved preprocessing validation and incorporate individualized anatomical data to enhance the sensitivity and spatial precision of EEG-derived biomarkers in pediatric MS.

Finally, limitations were encountered in the interpretation of correlation results. Although the CBL includes participants aged 0–18 years, pediatric-onset MS is typically diagnosed from early school age onward. The youngest participant in this study was five years old. Cognitive assessment was based on the WISC-V, which is standardized for children aged 6–16 years. Two participants were not in this age range and did not complete the same version of the test. While other Wechsler intelligence scales for different ages are available for additional participants, the subsets scores (used in this study) could not be directly compared. Age correction is not convenient in pediatric populations with neurodevelopmental disorders, as their developmental trajectories may substantially deviate from the norm. As such, a unified metric for visual processing speed across age groups is currently lacking, which limits the generalizability of the observed associations as biomarkers.

4.3. Future research recommendations

While improvements such as increased sample sizes and refined methodological tools are important, the true potential of this research lies in how to expand, integrate and translate functional connectivity measures into a clinically meaningful tool. This pilot study offers technical insights and a glimpse of the broader value of EEG-based biomarkers in understanding and recognizing cognitive impairments in pediatric MS. Importantly, the developed framework could be applicable beyond pediatric MS and may inform biomarker research in other neurodevelopmental disorders.

At this point, it is essential to acknowledge that the findings are exploratory. Robust evidence from longitudinal studies and replication in independent samples is required before any clinical application can be established. In the context of neurodevelopmental disorders, functional connectivity must be shown to be a reliable, specific, and clinically relevant biomarker. Although initial predictive models can be developed without full mechanistic insight, their clinical application remains challenging. Therefore, a necessary and logical next step is to deepen the understanding of the underlying neural mechanisms. This fundamental knowledge can help identify robust and clinically meaningful features, improve explainability of predictive models and support clinical applications.

First, optimizing preprocessing is important because it directly affects the signal quality and interpretability of the connectivity results. The developed preprocessing pipeline is a valuable and scalable starting point for EEG analysis, even beyond this application, but should be further optimized for adaptability and consistency in application to different data types and acquisition protocols. A full explanation of the proposed preprocessing optimization is included in Appendix C. A critical next step is to improve the specificity and interpretation of the current connectivity measures. Although global metrics are robust, future studies should examine how frequency-specific activity, task conditions, and anatomical precision affect the relationship with cognition. Comparisons with healthy controls will be of particular interest in distinguishing developmental variations from MS specific patterns.

Building on the present findings, which demonstrate the feasibility of global connectivity measures in pediatric MS, future research can extend this framework by integrating data from other modalities, such as eye-tracking, structural imaging, or clinical characteristics, which can enrich the feature space and improve the explanatory value of machine learning models. Age, symptom severity, medication use, and comorbidities are potential confounding factors that should be included. Instead of considering these factors as sources of noise, they can be used as informative covariates in predictive models. However, this dataset exemplifies the challenges of high-dimensional, low-sample-size datasets. In such cases, machine learning approaches are particularly vulnerable to overfitting, multicollinearity, and limited generalizability. This highlights the importance of first understanding the underlying mechanisms, which can guide feature reduction within each modality. In addition, it is important to select modeling approaches suitable for high-dimensional data, such as linear regression using the Least Absolute Shrinkage and Selection Operator (LASSO), which penalizes the absolute size of coefficients and can shrink some to zero to support feature selection (Muthukrishnan and Rohini, 2016). By combining these methodological improvements with a deeper understanding of the underlying mechanisms,

functional connectivity research can evolve from exploratory analysis to a clinically useful tool for identifying and tracking cognitive impairment in pediatric MS.

4.4. Conclusion

This study aimed to identify EEG-based markers of cognitive impairment in pediatric MS by examining the relationship between FC and visual processing speed. To achieve this, an automated EEG preprocessing pipeline was developed and applied in a pilot study that focused on resting-state EEG. As hypothesized, faster visual processing was associated with stronger interhemispheric and fronto-occipital connectivity and higher network efficiency. Among these associations, the MST metrics for network efficiency showed the most robust pattern across both eyes-open and eyes-closed conditions. The smaller diameter and higher leaf fraction of the PLI-based MST were associated with a higher visual processing speed. Although the results did not reach statistical significance, they support the value of EEG-based functional connectivity for studying brain–behavior relations and underscore the potential of MST network measures as biomarkers of cognitive impairment in pediatric MS, with implications for early diagnosis, personalized treatment, and improved prognosis.

Bibliography

Ahmed, N. S., AbdAllah, M. A., Nassef, A. M., Mohamed, A. E. A., & Nada, M. A. (2022). Cognitive impairment in paediatric onset multiple sclerosis and its relation to thalamic volume and cortical thickness of temporal lobe by magnetic resonance imaging. *Egypt. J. Neurol. Psychiatr. Neurosurg.*, *58*(1).

Akbar, N., Giorgio, A., Till, C., Sled, J. G., Doesburg, S. M., De Stefano, N., & Banwell, B. (2016). Alterations in functional and structural connectivity in pediatric-onset multiple sclerosis. *PLoS One*, *11*(1), e0145906.

Amato, M. P., Goretti, B., Ghezzi, A., Lori, S., Zipoli, V., Portaccio, E., Moiola, L., Falautano, M., De Caro, M. F., Lopez, M., Patti, F., Vecchio, R., Pozzilli, C., Bianchi, V., Roscio, M., Comi, G., Trojano, M., & Multiple Sclerosis Study Group of the Italian Neurological Society. (2008). Cognitive and psychosocial features of childhood and juvenile MS. *Neurology*, *70*(20), 1891–1897.

Anijärv, T. E., & Campbell, A. J. (2024). EEG-pyline: EEG pipeline in python (v0.1.5). https://doi.org/10.5281/zenodo.13830865

Appelhoff, S., Hurst, A. J., Lawrence, A., Li, A., Mantilla Ramos, Y. J., O'Reilly, C., Xiang, L., Dancker, J., Scheltienne, M., & Bialas, O. (2023). PyPREP: A python implementation of the preprocessing pipeline (PREP) for EEG data.

Bailey, N. W., Biabani, M., Hill, A. T., Miljevic, A., Rogasch, N. C., McQueen, B., Murphy, O. W., & Fitzgerald, P. B. (2023). Introducing RELAX: An automated pre-processing pipeline for cleaning EEG data - part 1: Algorithm and application to oscillations. *Clin. Neurophysiol.*, *149*, 178–201.

Bardel, B., Ayache, S. S., & Lefaucheur, J.-P. (2024). The contribution of EEG to assess and treat motor disorders in multiple sclerosis. *Clin. Neurophysiol.*, *162*, 174–200.

Barry, R. J., Clarke, A. R., Johnstone, S. J., Magee, C. A., & Rushby, J. A. (2007). EEG differences between eyes-closed and eyes-open resting conditions. *Clin. Neurophysiol.*, *118*(12), 2765–2773.

Bastos, A. M., & Schoffelen, J.-M. (2015). A tutorial review of functional connectivity analysis methods and their interpretational pitfalls. *Front. Syst. Neurosci.*, *9*, 175.

Berger, H. (1929). Über das elektrenkephalogramm des menschen. *Archiv f. Psychiatrie*, *87*(1), 527–570.

Bigdely-Shamlo, N., Mullen, T., Kothe, C., Su, K.-M., & Robbins, K. A. (2015). The PREP pipeline: Standardized preprocessing for large-scale EEG analysis. *Front. Neuroinform.*, *9*, 16.

Castellanos, N. P., & Makarov, V. A. (2006). Recovering EEG brain signals: Artifact suppression with wavelet enhanced independent component analysis. *J. Neurosci. Methods*, *158*(2), 300–312.

Charlebois-Poirier, A.-R., Davoudi, S., Lalancette, È., Knoth, I. S., & Lippé, S. (2025). The level of cognitive functioning in school-aged children is predicted by resting EEG directed phase lag index. *Sci. Rep.*, *15*(1), 1531.

Chiarion, G., Sparacino, L., Antonacci, Y., Faes, L., & Mesin, L. (2023). Connectivity analysis in EEG data: A tutorial review of the state of the art and emerging trends. *Bioengineering (Basel)*, *10*(3), 372.

Clayson, P. E., Baldwin, S. A., Rocha, H. A., & Larson, M. J. (2021). The data-processing multiverse of event-related potentials (ERPs): A roadmap for the optimization and standardization of ERP processing and reduction pipelines. *Neuroimage*, *245*(118712), 118712.

de Wit, I. (2025). Eeg functional connectivity analysis in correlation with visual processing speed in children: A literature review on metrics, brain regions and frequency bands.

Debnath, R., Buzzell, G. A., Morales, S., Bowers, M. E., Leach, S. C., & Fox, N. A. (2020). The maryland analysis of developmental EEG (MADE) pipeline. *Psychophysiology*, *57*(6), e13580.

Delorme, A. (2023). EEG is better left alone. *Sci. Rep.*, *13*(1), 2372.

Delorme, A., Truong, D., Pion-Tonachini, L., & Makeig, S. (2024). Automatic EEG independent component classification using ICLabel in python.

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., Buckner, R. L., Dale, A. M., Maguire, R. P., Hyman, B. T., Albert, M. S., & Killiany, R. J. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage*, *31*(3), 968–980.

Fitzgibbon, S. P., DeLosAngeles, D., Lewis, T. W., Powers, D. M. W., Grummett, T. S., Whitham, E. M., Ward, L. M., Willoughby, J. O., & Pope, K. J. (2016). Automatic determination of EMG-contaminated components and validation of independent component analysis using EEG during pharmacologic paralysis. *Clin. Neurophysiol.*, *127*(3), 1781–1793.

Fló, A., Gennari, G., Benjamin, L., & Dehaene-Lambertz, G. (2022). Automated pipeline for infants continuous EEG (APICE): A flexible pipeline for developmental cognitive studies. *Dev. Cogn. Neurosci.*, *54*(101077), 101077.

Friston, K. J., Frith, C. D., Liddle, P. F., & Frackowiak, R. S. (1993). Functional connectivity: The principal-component analysis of large (PET) data sets. *J. Cereb. Blood Flow Metab.*, *13*(1), 5–14.

Gabard-Durnam, L. J., Mendez Leal, A. S., Wilkinson, C. L., & Levin, A. R. (2018). The harvard automated processing pipeline for electroencephalography (HAPPE): Standardized processing software for developmental and high-artifact data. *Front. Neurosci.*, *12*.

Goncharova, I. I., McFarland, D. J., Vaughan, T. M., & Wolpaw, J. R. (2003). EMG contamination of EEG: Spectral and topographical characteristics. *Clin. Neurophysiol.*, *114*(9), 1580–1593.

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L., & Hämäläinen, M. (2013). MEG and EEG data analysis with MNE-Python. *Front. Neurosci.*, *7*, 267.

Grandy, T. H., Werkle-Bergner, M., Chicherio, C., Lövdén, M., Schmiedek, F., & Lindenberger, U. (2013). Individual alpha peak frequency is related to latent factors of general cognitive abilities. *Neuroimage*, *79*, 10–18.

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., Del Río, J. F., Wiebe, M., Peterson, P., … Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357–362.

Hill, A. T., Enticott, P. G., Fitzgerald, P. B., & Bailey, N. W. (2024). RELAX-Jr: An automated preprocessing pipeline for developmental EEG recordings. *Hum. Brain Mapp.*, *45*(14), e70034.

Jin, Y., O'Halloran, J. P., Plon, L., Sandman, C. A., & Potkin, S. G. (2006). Alpha EEG predicts visual reaction time. *Int. J. Neurosci.*, *116*(9), 1035–1044.

Johnson, M. H. (2003). Development of human brain functions. *Biol. Psychiatry*, *54*(12), 1312–1316.

Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: A review and analysis. *Brain Res. Rev.*, *29*(2-3), 169–195.

Koller, H. P. (2012). Visual processing and learning disorders. *Curr. Opin. Ophthalmol.*, *23*(5), 377–383.

Lin, S.-J., Kolind, S., Liu, A., McMullen, K., Vavasour, I., Wang, Z. J., Traboulsee, A., & McKeown, M. J. (2020). Both stationary and dynamic functional interhemispheric connectivity are strongly associated with performance on cognitive tests in multiple sclerosis. *Front. Neurol.*, *11*, 407.

Mišić, B., Fatima, Z., Askren, M. K., Buschkuehl, M., Churchill, N., Cimprich, B., Deldin, P. J., Jaeggi, S., Jung, M., Korostil, M., Kross, E., Krpan, K. M., Peltier, S., Reuter-Lorenz, P. A., Strother, S. C., Jonides, J., McIntosh, A. R., & Berman, M. G. (2014). The functional connectivity landscape of the human brain. *PLoS One*, *9*(10), e111007.

Muthukrishnan, R., & Rohini, R. (2016). LASSO: A feature selection technique in predictive modeling for machine learning. *2016 IEEE International Conference on Advances in Computer Applications (ICACA)*.

Muthukumaraswamy, S. D. (2013). High-frequency brain activity and muscle artifacts in MEG/EEG: A review and recommendations. *Front. Hum. Neurosci.*, *7*, 138.

Neubauer, A. C., & Fink, A. (2003). Fluid intelligence and neural efficiency: Effects of task complexity and sex. *Pers. Individ. Dif.*, *35*(4), 811–827.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Pion-Tonachini, L., Kreutz-Delgado, K., & Makeig, S. (2019). ICLabel: An automated electroencephalographic independent component classifier, dataset, and website. *Neuroimage*, *198*, 181–197.

Portaccio, E., De Meo, E., Bellinvia, A., & Amato, M. P. (2021). Cognitive issues in pediatric multiple sclerosis. *Brain Sci.*, *11*(4), 442.

Reigal, R. E., Barrero, S., Martín, I., Morales-Sánchez, V., Juárez-Ruiz de Mier, R., & Hernández-Mendo, A. (2019). Relationships between reaction time, selective attention, physical activity, and physical fitness in children. *Front. Psychol.*, *10*, 2278.

Rocca, M. A., Schoonheim, M. M., Valsasina, P., Geurts, J. J. G., & Filippi, M. (2022). Task- and resting-state fMRI studies in multiple sclerosis: From regions to systems and time-varying analysis. current status and future perspective. *NeuroImage Clin.*, *35*(103076), 103076.

Schomer, D. L., & da Silva, F. H. L. (2017). *Niedermeyer's electroencephalography: Basic principles, clinical applications, and related fields*.

Schoonheim, M. M., Geurts, J. J. G., Landi, D., Douw, L., van der Meer, M. L., Vrenken, H., Polman, C. H., Barkhof, F., & Stam, C. J. (2013). Functional connectivity changes in multiple sclerosis patients: A graph analytical study of MEG resting state data. *Hum. Brain Mapp.*, *34*(1), 52–61.

Sharmin, S., Roos, I., Malpas, C. B., Iaffaldano, P., Simone, M., Filippi, M., Kubala Havrdova, E., Ozakbas, S., Brescia Morra, V., Alroughani, R., Zaffaroni, M., Patti, F., Eichau, S., Salemi, G., Di Sapio, A., Inglese, M., Portaccio, E., Trojano, M., Amato, M. P., … Italian Multiple Sclerosis and Related Disorders Register and MSBase Study Group. (2024). Disease-modifying therapies in managing disability worsening in paediatric-onset multiple sclerosis: A longitudinal analysis of global and national registries. *Lancet Child Adolesc. Health*, *8*(5), 348–357.

Shirani, S., & Mohebbi, M. (2022). Brain functional connectivity analysis in patients with relapsing-remitting multiple sclerosis: A graph theory approach of EEG resting state. *Front. Neurosci.*, *16*, 801774.

Siu, C. R., & Murphy, K. M. (2018). The development of human visual cortex and clinical implications. *Eye Brain*, *10*, 25–36.

Somers, B., Francart, T., & Bertrand, A. (2018). A generic EEG artifact removal algorithm based on the multi-channel wiener filter. *J. Neural Eng.*, *15*(3), 036007.

Stam, C. J., Nolte, G., & Daffertshofer, A. (2007). Phase lag index: Assessment of functional connectivity from multi channel EEG and MEG with diminished bias from common sources. *Hum. Brain Mapp.*, *28*(11), 1178–1193.

Sumowski, J. F., Benedict, R., Enzinger, C., Filippi, M., Geurts, J. J., Hamalainen, P., Hulst, H., Inglese, M., Leavitt, V. M., Rocca, M. A., Rosti-Otajarvi, E. M., & Rao, S. (2018). Cognition in multiple sclerosis. *Neurology*, *90*(6), 278–288.

Tahedl, M., Levine, S. M., Greenlee, M. W., Weissert, R., & Schwarzbach, J. V. (2018). Functional connectivity in multiple sclerosis: Recent findings and future directions. *Front. Neurol.*, *9*, 828.

Tewarie, P., van Dellen, E., Hillebrand, A., & Stam, C. J. (2015). The minimum spanning tree: An unbiased method for brain network analysis. *Neuroimage*, *104*, 177–188.

Thompson, A. J., Baranzini, S. E., Geurts, J., Hemmer, B., & Ciccarelli, O. (2018). Multiple sclerosis. *Lancet*, *391*(10130), 1622–1636.

Wechsler, D. (2014). *Wechsler intelligence scale for children* (5th ed.). Pearson.

Zhao, L., Zhang, Y., Yu, X., Wu, H., Wang, L., Li, F., Duan, M., Lai, Y., Liu, T., Dong, L., & Yao, D. (2023). Quantitative signal quality assessment for large-scale continuous scalp electroencephalography from a big data perspective. *Physiol. Meas.*, *44*(3).

A

Development of the preprocessing pipeline

Effective preprocessing must remove artifacts without compromising underlying brain activity. However, this balance is difficult to achieve, especially because the true brain signal is unknown. This makes the evaluation of the preprocessing complex because incorrect assumptions can let the meaningful neural signals be overshadowed by noise or remove too much variation as noise and remove the neural signal with it. While automated preprocessing pipelines have been developed in MATLAB for pediatric EEG, no unbiased validation has been performed to conclude the best option and standardize its usage. For Python, only the adult pipeline of PyLine has been developed and used in studies, but no validation has been performed.

To address this challenge, a Python-based preprocessing pipeline tailored to pediatric EEG data is developed using a systematic approach to evaluate and compare preprocessing methods. Key aspects of the development of this preprocessing pipeline included signal preservation and artifact removal checks.

A.1. Introduction of concept pipelines

Different steps of already developed automated preprocessing pipelines are extracted and implemented in Python for comparison as concept solutions. Four currently available automated preprocessing pipelines for pediatric data are MADE, HAPPE, RELAX-JR, and APICE.

- **HAPPE**: The Harvard Automated Preprocessing Pipeline for EEG (Gabard-Durnam et al., 2018)
- **MADE**: The Maryland Analysis of Developmental EEG preprocessing pipeline (Debnath et al., 2020)
- **RELAX-Jr**: The Reduction of Electroencephalographic Artefacts Junior pipeline (Hill et al., 2024) (specifically the MWF wICA implementation is adviced)
- **APICE**: The Automated Pipeline for Infants Continuous EEG (Fló et al., 2022)

In Table A.1 the available preprocessing pipelines are compared for different steps. The common sequence of steps is resampling, filtering, bad channel and epoch rejection, artifact reduction and re-referencing. A comparison of the existing pipelines shows that the main addition to the standard filters currently used by the Child Brain Lab is the rejection of bad channels and artifact reduction.

Table A.1: Comparison of existing pediatric EEG preprocessing pipelines. Abbreviations: HAPPE (Harvard Automated Preprocessing Pipeline), MADE (Maryland Analysis of Developmental EEG preprocessing pipeline), RELAX-Jr (Reduction of Electroencephalographic Artefacts Junior pipeline), APICE (Automated Pipeline for Infants Continuous EEG), FIR (Finite impulse response), SD (standard deviation), FASTER (Fully Automated Statistical Thresholding for EEG artifact Rejection), PREP (preprocessing pipeline), wICA (wavelet enhanced ICA), ICA (Independent component analysis), MWF (multichannel wiener filter), PCA (Principle component analysis), MARA (Multiple Artifact Rejection Algorithm), ADJUST (Automatic EEG artifact Detector based on the Joint Use of Spatial and Temporal features), PICARD (Preconditioned ICA for Real Data).

Pipelines	**HAPPE (Gabard-Durnam et al., 2018)**	**MADE (Debnath et al., 2020)**	**RELAX-Jr(Hill et al., 2024)**	**APICE (Fló et al., 2022)**
Filter	1 Hz high-pass, 250 Hz low pass, CleanLine (50 Hz)	0.3 Hz high-pass, 48 Hz low-pass (FIR)	0.25-80 Hz bandpass, 47-53 Hz notch	0.1 Hz high-pass, 40 Hz low-pass (FIR)
Bad channel rejection	Joint probability >3 SD	FASTER (Hurst, corraltion, variance)	PREP's findNoisyChannels	Multiple rejection cycles with relative thresholds for extreme values
Initial outlying rejection	not applied	not applied	1s epochs using same thresholds as for bad channels	Algorithm dependent (multiple used)
Initial artifact reduction	wICA	ICA weights copied back to original	MWF (muscle, blink, eye movement)	PCA to specific brief events and wICA
Second artifact reduction	MARA to reject artifact components	Adjusted-ADJUST to reject artifact components	ICA with PICARD algorithm and Adjusted-ADJUST used to identify artifact components to apply wICA	Adjusted-ADJUST to reject artifact components
Re-referencing	user's choice	average	robust average	robust average

Multiple criteria are used in pipelines for bad channel rejection. Most of these criteria are implemented in the Preprocessing Pipeline (PREP) (Bigdely-Shamlo et al., 2015). PREP is an automated method originally developed for large-scale EEG preprocessing that identifies and interpolates bad channels and is also available in Python under the name pyPREP (Appelhoff et al., 2023). The choice to use this pipeline was made without extensively comparing individual bad channel rejection criteria to focus efforts on evaluating artifact reduction methods, where greater methodological variation exists between pipelines.

For artifact reduction, the existing pipelines show three main options, which are used as concepts for comparison: the multichannel Wiener filter, independent component analysis (ICA) with automated artifact component rejection and wavelet-enhanced ICA. These three methods are implemented in

Python as three concept pipelines to compare their cleaning performances.

1. The MWF is implemented as described by Somers et al. in 2018 (Somers et al., 2018). The filter is targeted at muscle artifacts because muscle activity spans a broad frequency range and is spatially widespread. Compared to ocular artifacts, muscle artifacts have the highest risk of introducing spurious synchrony across channels in a functional connectivity analysis (Goncharova et al., 2003, Muthukumaraswamy, 2013). Muscle artifacts were identified from the neural signal using a threshold of for the log power log frequency slope of -0.59 (Fitzgibbon et al., 2016).

2. ICA is performed by the ICA function in the preprocessing toolbox of MNE python (Gramfort et al., 2013) with 15 components and the infomax method. IClabel is then used to automatically identify the artifactual components to be rejected (Pion-Tonachini et al., 2019). IClabel was chosen over other classification methods because it provides the most accurate results with the least computational cost (Pion-Tonachini et al., 2019) and is implemented in Python without loss of performance (Delorme et al., 2024).

3. For the wICA, an ICA is first preformed using the FastICA function of the decomposition toolbox of sklearn with 25 components. This function allows the reconstruction of the EEG with filtered sources. Wavelet filtering was then applied as described by Castellanos et al. in 2016 (Castellanos and Makarov, 2006).

All concept preprocessing pipelines start with basic filters, as already applied in the current preprocessing method of the Child Brain Lab. This includes a low-pass FIR filter of 80 Hz after which the data is resampled to a sample frequency of 250 Hz. Then the line noise at 50 and 100 Hz is removed with a notch filter and lastly a high pass FIR filter of 0.1 Hz is applied. All concept preprocessing pipelines end with re-referencing to the average, because this is also done in the existing pipelines and in the currently used method of the CBL.

A.2. Methods for comparison and selection between the concept preprocessing pipelines

All EEG resting state recordings from the CBL (regardless of any diagnosis) are used to compare the performance of the concept pipelines. All EEGs are preprocessed with the three concept preprocessing pipelines and cleaning performance is compared using metrics that quantify signal preservation and artifact removal. First, the quality of the data before and after preprocessing is assessed to evaluate the stability and effect of each pipeline. Furthermore, signal preservation is quantified using the signal error ratio (SER) and the correlation between the cleaned and original signal. Artifact removal is evaluated based on the artifact to residue ratio and the mutual information between the removed and cleaned signal. Lastly, the strength of the Berger effect and visual inspection by a neurologist are included as functional assessments to determine whether the pipelines sufficiently prepare the EEG data for clinically relevant analyses. The performance metrics are described in detail in Appendix B.

The selection process cannot be fully predefined because it involves a trade-off between multiple, partly conflicting criteria. For instance, strong artifact removal may occur at the expense of signal preservation. Therefore, no single metric is decisive. Instead, a balanced evaluation is required across all metrics. Signal preservation is prioritized because maintaining the integrity of the neural signal is essential. However, perfect signal preservation (100% correlation) is neither expected nor desirable, as even the cleanest segments may contain subtle noise or interference. Thus, the goal is to develop a stable preprocessing pipeline that effectively removes artifacts while retaining as much neural signal as possible. Among the signal preservation metrics, visual inspection and the strength of the Berger effect are considered the most important. Although the SER provides a useful quantitative index, its interpretation is less straightforward because of the absence of clear normative thresholds. In contrast, visual inspection allows for direct assessment of whether meaningful brain signals are retained, and the strength of the Berger effect is a task-relevant outcome measure that reflects whether neural responses are preserved after preprocessing.

After selecting the most promising approach, this pipeline is further optimized by adjusting the parameters or combining methods if beneficial. The final optimized pipeline is re-evaluated using the same performance metrics to validate its performance before use in analysis.

A.3. Results of the comparison between the preprocessing pipelines

Seventy EEG resting state recordings from the CBL were preprocessed using MWF, IClabel, and wICA. The scripts of the pipelines are available in a GitHub repository (https://github.com/Ilse2001/code-repository-msc-thesis-eeg-fc).

A.3.1. Signal preservation and artifact removal

Table A.2 and Figure A.1 show the metrics for signal preservation and artifact removal for all the pipelines. The MWF shows the highest signal preservation, but also the lowest artifact removal. The performances overlap between the different pipelines and depend on the file used. However, the means do show a difference. wICA had the best artifact removal, but the lowest signal preservation.

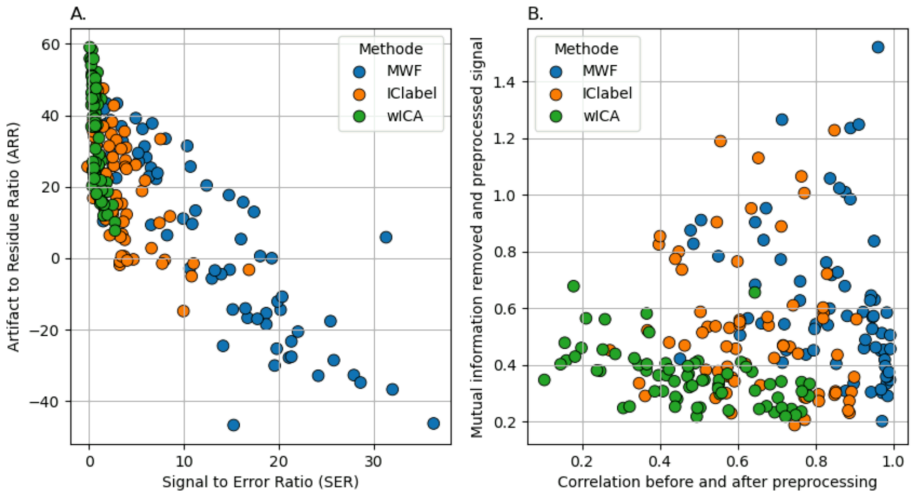

Figure A.1: Trade off between signal preservation and artifact removal in the three concept pipelines for preprocessing. A) Scatter plot of the signal error ratio (SER) against the artifact residue ratio (ARR). B) Scatter plot of the correlation before and after preprocessing and the mutual information of the removed and preprocessed signal.

mean ± SD	SER	ARR	correlation	mutual information
MWF	13.1 ± 8.5	5.0 ± 25.4	0.84 ± 0.15	0.61 ± 0.27
IC-label	3.4 ± 3.0	20.3 ± 15.3	0.65 ± 0.16	0.52 ± 0.25
wICA	0.8 ± 0.6	32.9 ± 14.3	0.48 ± 0.18	0.36 ± 0.10

Table A.2: All the cleaning performance metrics per concept pipeline. Abbriviations: SD (standard deviation), SER (Signal Error Ratio), ARR (Artifact Residue Ratio), MWF (Multichannel Wiener Filter), wICA (wavelet enhanced Independent Component Analysis)

A.3.2. Strength of Berger effect

Figure A.2 shows the average alpha band topographical maps (topomaps) and corresponding power spectral density as a mean over all participants for each preprocessing pipeline (including the current standard without artifact reduction) for both the eyes-open and eyes-closed conditions. All pipelines are able to detect the Berger effect. The results of the Bayesian t-test, which indicates the power with which the pipelines can detect the Berger effect, are shown in Table A.3. wICA, which had the highest artifact removal scores, shows the highest Bayesian factor, indicating the strongest detection of the Berger effect. In the MWF preprocessed data, which showed the lowest artifact removal, the Berger effect is shown with the least strength of the concept pipelines. However, in all pipelines, the difference was significant.

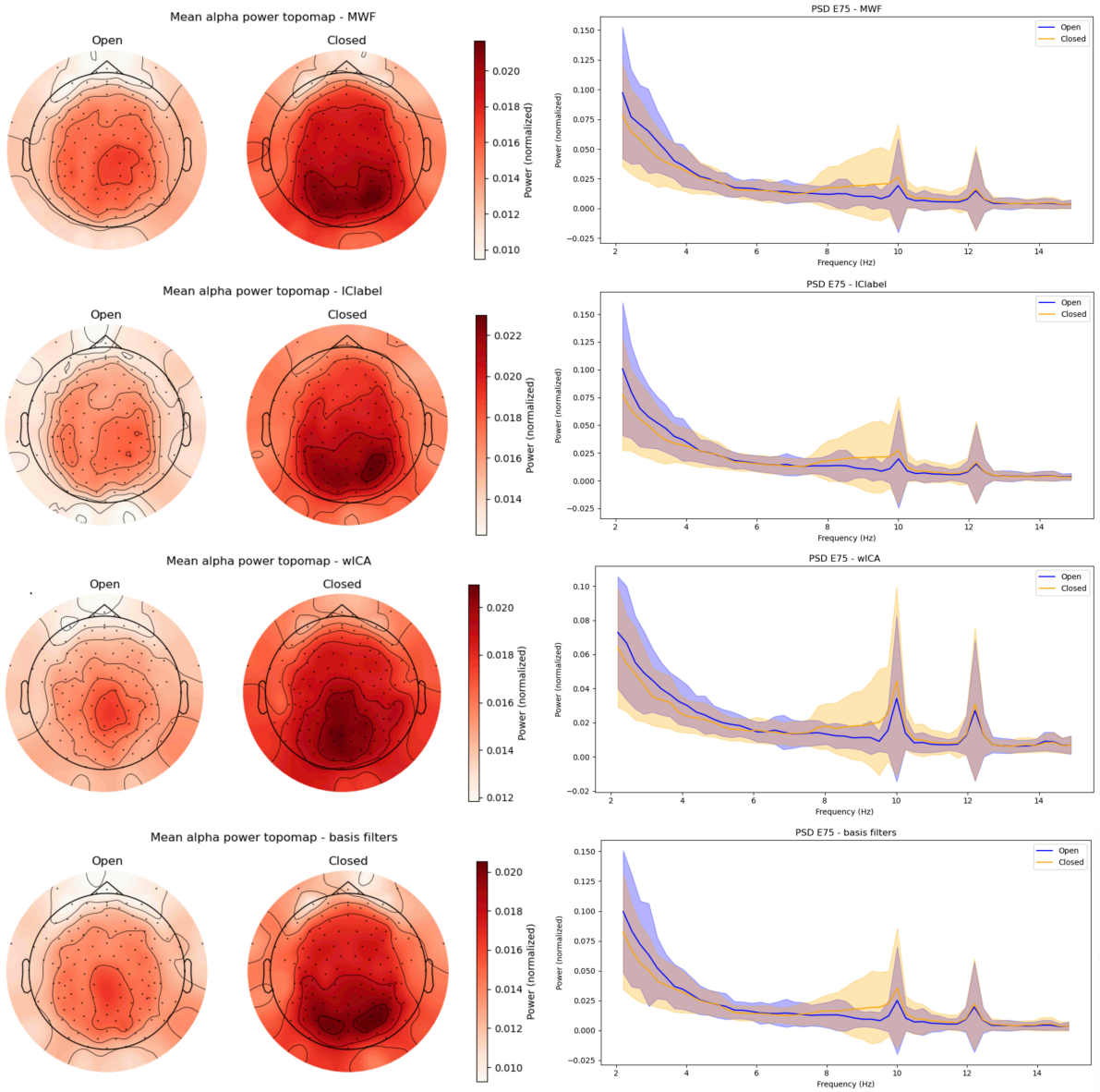

Figure A.2: Topomaps and power spectra of the average alpha power across all participants for eyes open and eyes closed resting state, shown per preprocessing pipeline.

pipelines	BF10	cohens d	p-value
MWF	3589.644	0.651813	0.000005
IC-label	1.108e+04	0.650122	0.000001
wICA	2.608e+04	0.655677	5.641739e-07
Basic filters	2548.04	0.660397	0.000007

Table A.3: Results of the Bayesian t-test for the difference in alpha power between eyes open and closed for all concept pipelines. Abbreviations: BF10 (Bayesian factor), MWF (Multichannel Wiener Filter), wICA (wavelet enhanced Independent Component Analysis)

A.3.3. Stability of quality improvement

In figure A.3 paired scatter plots of the quality improvement compared to the basic filters without artifact reduction are shown for all concept pipelines. wICA shows the most stable and largest improvement in quality for all the EEG recordings. However, all pipelines show a stable improvement in quality. However, with the MWF and IC-label pipelines, it is noticeable in the paired scatter plot that some

specific files are less responsive to preprocessing, while they do improve with the wICA pipeline.

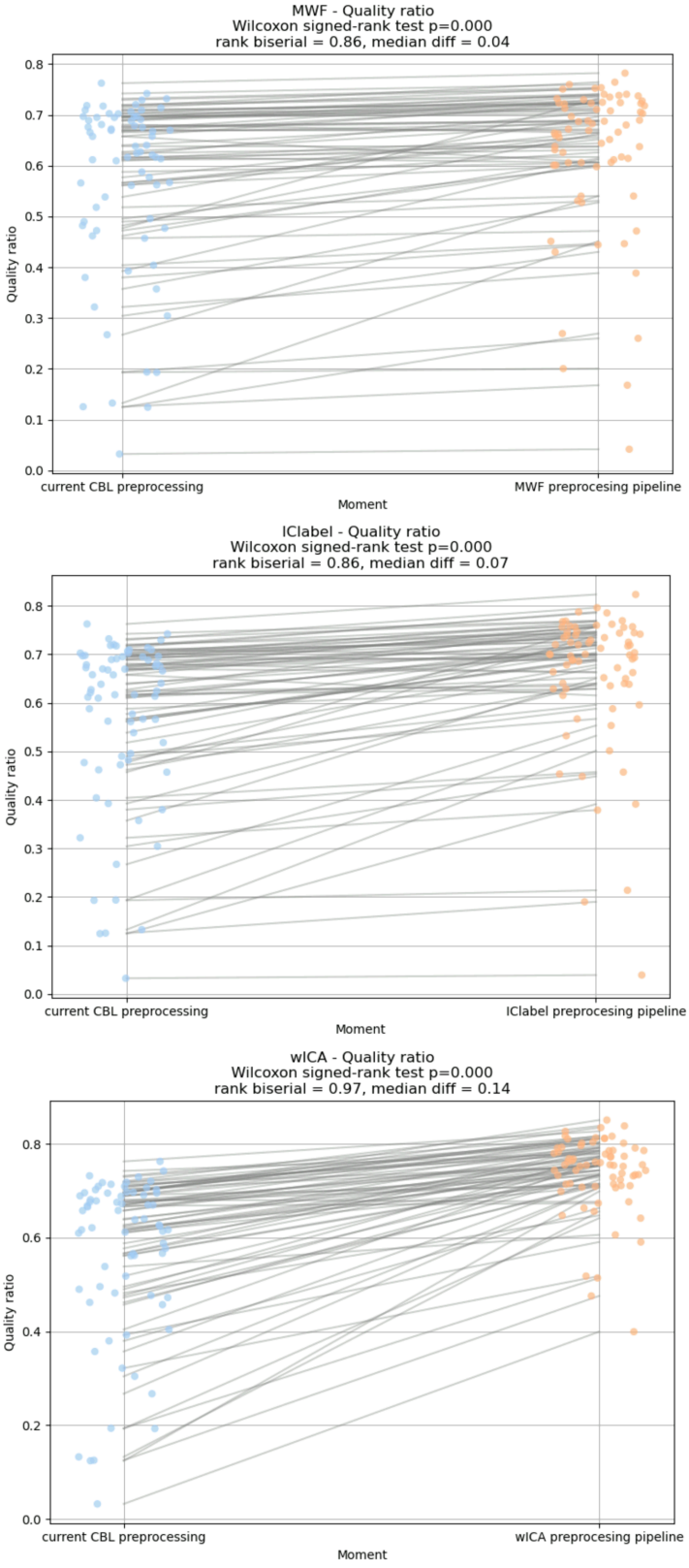

Figure A.3: Paired scatter plots of all the concept preprocessing pipelines to show the difference before and after preprocessing

A.3.4. Visual inspection

In figures A.4 to A.7 a small piece of the same EEG recording is shown when preprocessed with the different concept pipelines. In this fragment, some muscle and blink artifacts are indicated for a good comparison of the cleaning. It is worth noting that the MWF in the concept pipeline is focused on muscle artifacts and not on blinks, so those are not expected to be removed. Additionally, an epileptic spike train is indicated, which should not be removed by the preprocessing pipelines, as this is a neurological signal, although it is similar to an artifact in some aspects.

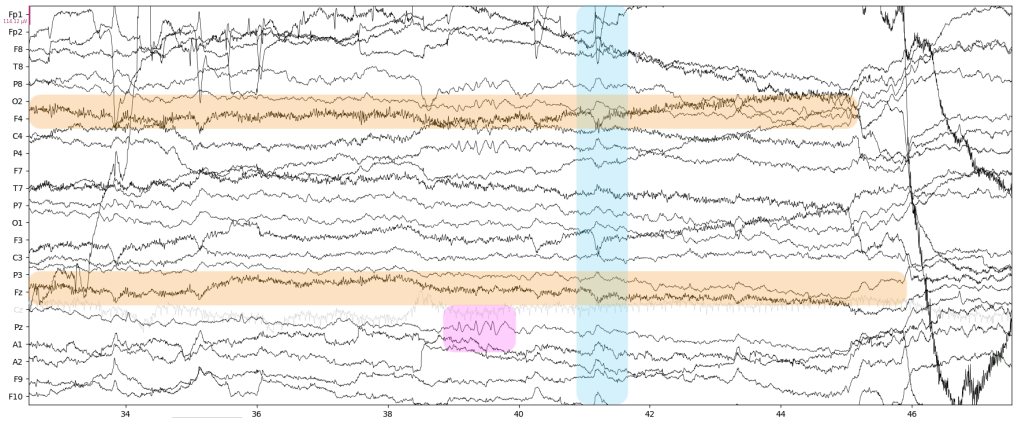

Figure A.4: fragment of recording preprocessed with basic filters, orange boxes are muscle artifacts, blue box is blink artifact and pink box is epileptic spike train

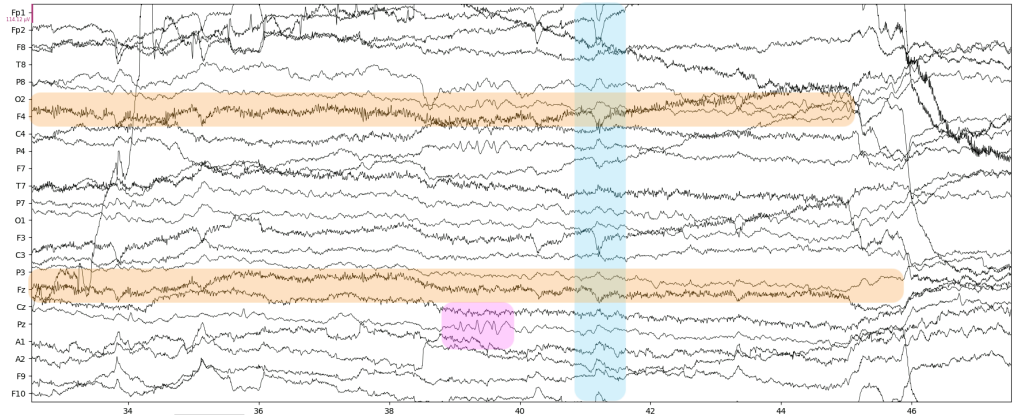

Figure A.5: fragment of recording preprocessed with MWF, orange boxes are muscle artifacts, blue box is blink artifact and pink box is epileptic spike train

A.4. Discussion on selecting most promising concept pipeline for further improvement

The IC-Label pipeline is identified as the most suitable basis for further refinement of the final preprocessing approach. The IC-label pipeline performs well based on visual inspection. It is the second-best option for artifact removal but has better signal preservation than the wICA pipeline. The MWF pipeline is not successful, probably because there are too many artifacts to distinguish between the clean signal and the artifact signal. The wICA pipeline shows the highest artifact removal ratio and the highest strength of the Berger effect, but should not be used in analysis because visual inspection shows that neurological signals such as epileptic spikes are removed. With the wavelet-enhanced method, it is not completely known what is removed exactly. Other literature also highlights that overly aggressive preprocessing can distort the neural signal and reduce data quality. Therefore, a more conservative and well-validated approach is recommended(Delorme, 2023.

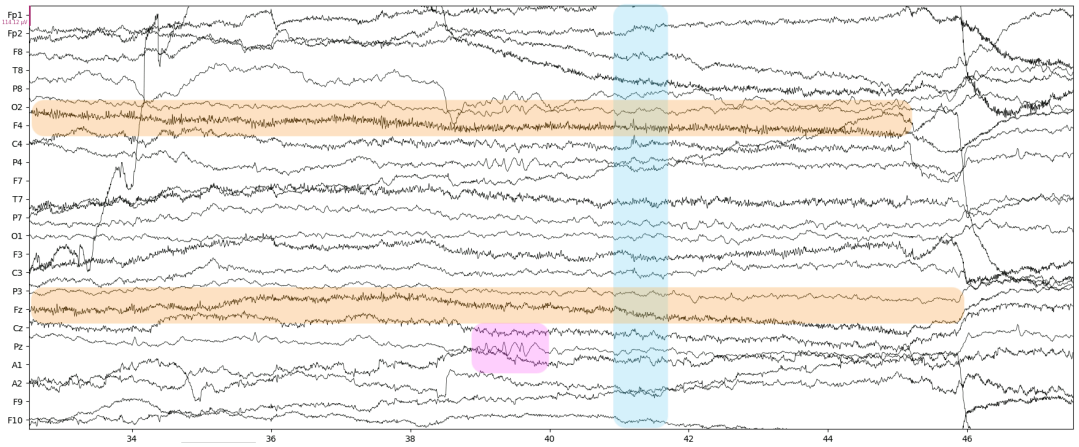

Figure A.6: fragment of recording preprocessed with IClabel, orange boxes are muscle artifacts, blue box is blink artifact and pink box is epileptic spike train

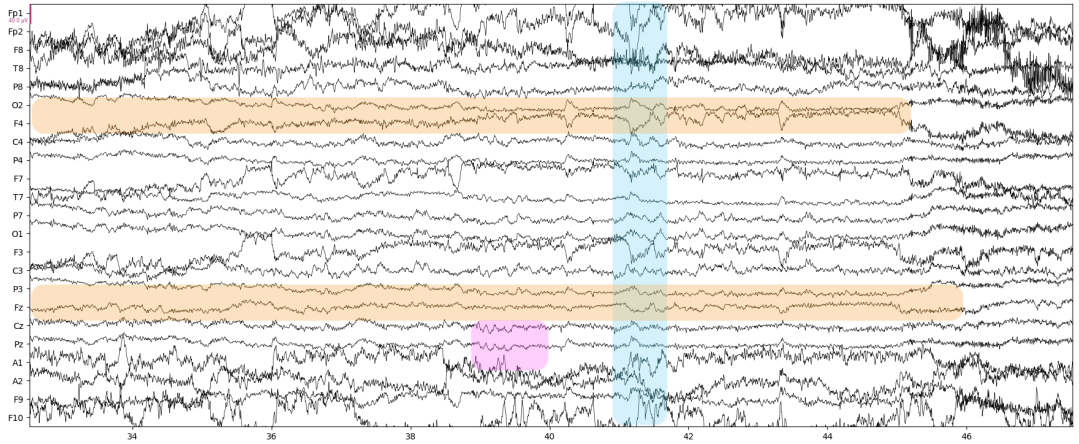

Figure A.7: fragment of recording preprocessed with wICA, orange boxes are muscle artifacts, blue box is blink artifact and pink box is epileptic spike train

A.5. Methods of optimization of the selected concept pipeline

The following optimization steps are applied to the IClabel preprocessing pipeline to ensure a robust performance.

1. The number of ICA components is maximized to one less than the number of good channels in the recording to improve the ability of the analysis to effectively identify and separate all the different types of artifacts and neural signals.

2. The data to fit the ICA on is expanded by not downsampling to 250 Hz and by using not only the resting state of the protocol but all the stages of the recording where tests are performed. Because the accurate estimation of all components requires a sufficient number of samples, approximately given by $N^2 * k$, where N is the number of components and k is a constant typically set to 20. In this study (128 channel recording), 330000 samples are needed, which is required by 330 s of recording with a 1000 Hz sample rate.

3. The data to fit the ICA is optimized by rejecting segments with extreme values above 150 μV and removing the notch filter for line noise.

A.6. Results of the validation of the developed preprocessing pipeline

In table A.4 the signal preservation and artifact removal metrics for the optimized preprocessing pipeline are shown. The optimized preprocessing pipeline shows a lower signal preservation and lower artifact removal compared to the IC-label concept pipeline. The potential to detect the Berger effect is illustrated

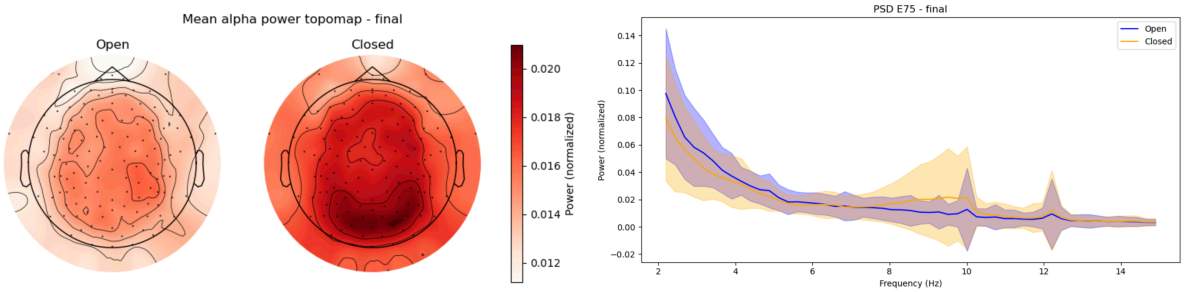

Figure A.8: Topoplot and power spectrum as a mean over all participants for eyes open and eyes closed resting state conditions for the optimized preprocessing pipeline

in Figure A.8. The results of the Bayesian t-test are presented in Table A.5. The preprocessing pipeline can detect the Berger effect, and the power to do so is higher than that for the IC-label concept pipeline.

mean ± SD	SER	ARR	correlation	mutual information
All data	2.4 ± 3.5	17.1 ± 14.1	0.56 ± 0.21	0.56 ± 0.26

Table A.4: All the cleaning performance metrics for the optimized preprocessing pipeline. Abbreviations: SD (standard deviation), SER (Signal Error Ratio), ARR (Artifact Residue Ratio)

pipelines	BF10	cohens d	p-value
final	3.257e+04	0.613902	4.440602e-07

Table A.5: Results of the Bayesian t-test for the difference in alpha power between eyes open and closed for the final preprocessing pipelines. Abbriviations: BF10 (bayesian factor)

The stability of improvement (figure A.9) is less than the IClabel concept pipeline, and although some files increase significantly in quality, others decrease in quality. Lastly, the same segment as for the concept pipelines is shown in figure A.10. The muscle artifacts are better removed than in the IClabel concept pipeline and the epileptic spike train remains intact. However, there are still some large artifacts, especially eye movements.

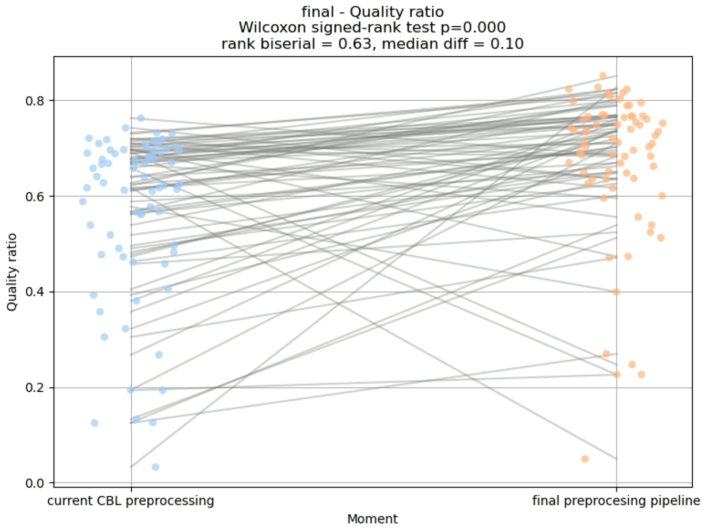

Figure A.9: Paired scatter plot of the optimized preprocessing pipeline to show the difference before and after preprocessing

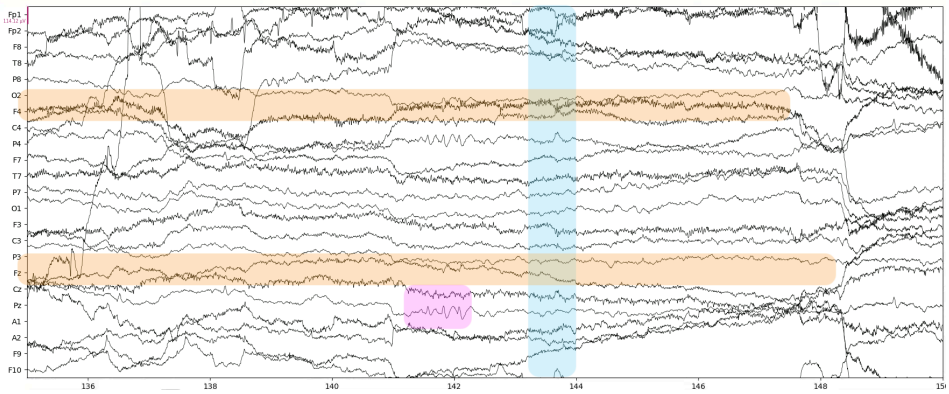

Figure A.10: Fragment of recording preprocessed with the optimized IC-label preprocessing pipeline, orange boxes are muscle artifacts, blue box is blink artifact and pink box is epileptic spike train

A.7. Discussion of the performance of the developed preprocessing pipeline

The developed preprocessing pipeline performs with sufficient quality to support further EEG analyses, provided that its inherent limitations are acknowledged. Further optimization is required to develop a broadly applicable standard preprocessing pipeline. However, the resulting preprocessing pipeline and systematic development approach, including the performance metrics used, provide a solid foundation for a standardized, reliable, automated preprocessing pipeline and quality assessment for pediatric EEG. Suggestions for future optimizations are elaborated in Appendix C.

The sufficient quality of performance from the preprocessing pipeline is based on quantitative quality metrics that indicate an overall improvement in data quality with adequate signal preservation and effective artifact removal. However, the inconsistent performance of the pipeline across recordings limits its broader applicability. This variability may stem from methodological choices, such as applying ICA decomposition from concatenated data from all task conditions rather than solely from resting-state data. While including all conditions increases the data quantity for component estimation, it may reduce the representativeness of the components to resting-state recordings. This shows the trade-off between the number and quality of the extracted components. Interestingly, despite lower SER and ARR scores than the original ICLabel concept pipeline, the optimized pipeline performed better in detecting the Berger effect. This potentially indicates that the SER and ARR metrics are more sensitive to performance inconsistency than functional test outcomes, such as the Bayesian t-test.

A fundamental limitation in developing and validating the preprocessing pipeline is the absence of a universally accepted “gold standard” for clean EEG signals. The evaluation of artifact removal and signal preservation relies on the assumption that EEG segments can be classified as either ”clean” or ”artifactual.” However, in practice, clean segments may still contain low-level noise, and artifact-labeled segments may include relevant neural signals. This limits the interpretability of performance metrics, such as the SER and ARR, as these may reflect a mixture of artifact and signal alterations rather than isolated effects.

This uncertainty poses a dilemma regarding the optimal outcome of preprocessing. In some cases, minimally processed EEG may be preferable, especially in clinical contexts, as it ensures that all neural signals are preserved, even at the cost of retaining noise or artifacts. However, well-known artifacts, such as muscle activity or eye movements, are likely to interfere with the analysis when they are not properly removed. This creates an inherent trade-off between signal preservation and artifact reduction, which may differ depending on the situation for which the preprocessing pipeline should be adaptable.

Lastly, the pipeline was implemented in Python using MNE, whereas most existing pediatric pipelines are developed in MATLAB environments. As a result, direct comparisons with prior work are limited, not only because of differing algorithmic implementations but also because shared evaluation frameworks are lacking. All these limitations reflect the reasons for the development of the present pipeline and emphasize the urgent need for ongoing advancements in pediatric EEG preprocessing. Continued refinement and optimization of pediatric EEG preprocessing frameworks are essential for improving data and analysis quality and enhancing reproducibility. Advancing standardized and reliable preprocessing

methods will facilitate more robust research outcomes and better support clinical impact.

B

Assessing cleaning performance

In this appendix, all used cleaning performance metrics of the preprocessing pipeline are explained in further detail. The functions to assess the data quality, signal preservation and artifact removal metrics are available from the GitHub repository (https://github.com/Ilse2001/code-repository-msc-thesis-eeg-fc) in the Quality assessment file.

B.1. Data quality

Data quality is defined as the extent to which the EEG signal is free from external contamination. The quantitative signal quality assessment for large-scale continuous scalp EEG (Zhao et al., 2023) is used to calculate the data quality, because it is based on four robust and well-founded criteria and is fully automated, making it suitable for big data applications. The MATLAB algorithm is translated to Python.

The processing flow of the quality assessment is illustrated in figure B.1. First, the EEG recording is divided into segments of one second in one channel, therefore it is important that the data does not contain any slow drifts, so the data is high pass filtered with a frequency of 1 Hz before providing it to the quality assessment. Each segment is assessed for artifacts using four criteria:

1. Constant/flat signal: median absolute deviation (MAD) below 10^{-4} μV
2. High amplitude: amplitude above 150 μV or z-score MAD above 3
3. High frequency noise: noise to signal ratio (NSR) above 0.5 or z-score above 5. (For the NSR, the method of Cristian Kothe was used, which assumes the power above 40 Hz as noise and below 40 Hz as the signal (Bigdely-Shamlo et al., 2015))
4. Low correlation: maximum correlation below 0.6 (The 98 percentile Pearson correlation was used to exclude high correlations due to noise).

The evaluation of these four criteria results in four Boolean masks (channels × time segments) with 1 if an artifact is detected and zero for clean segments. These four matrices are combined into a final matrix by taking the maximum of the four matrices for all segments. Thus, if one of the four criteria identifies an artifact, the segment is called artifactual. Using this final matrix of bad segments, the overall quality ratio can be assessed as the ratio of clean segments to the total number of segments. The code is uploaded to GitHub under the name “Functions Quality Assessment.” The main function can be called as quality = assess_quality(raw), where raw is an MNE-Python object of type mne.io.BaseRaw containing continuous EEG data, and the output is a DataFrame containing the criteria matrices and the overall quality ratio.

This automated quality assessment resulting in the quality ratio is validated using the quality scores of the neurologists to ensure a good performance. This validation is fully explained in Appendix D

B.2. Signal preservation

The signal to error ratio (SER) is used to quantify neural signal preservation and is implemented as described in Bailey et al., 2023. The SER is calculated per channel using Equation B.1, where the

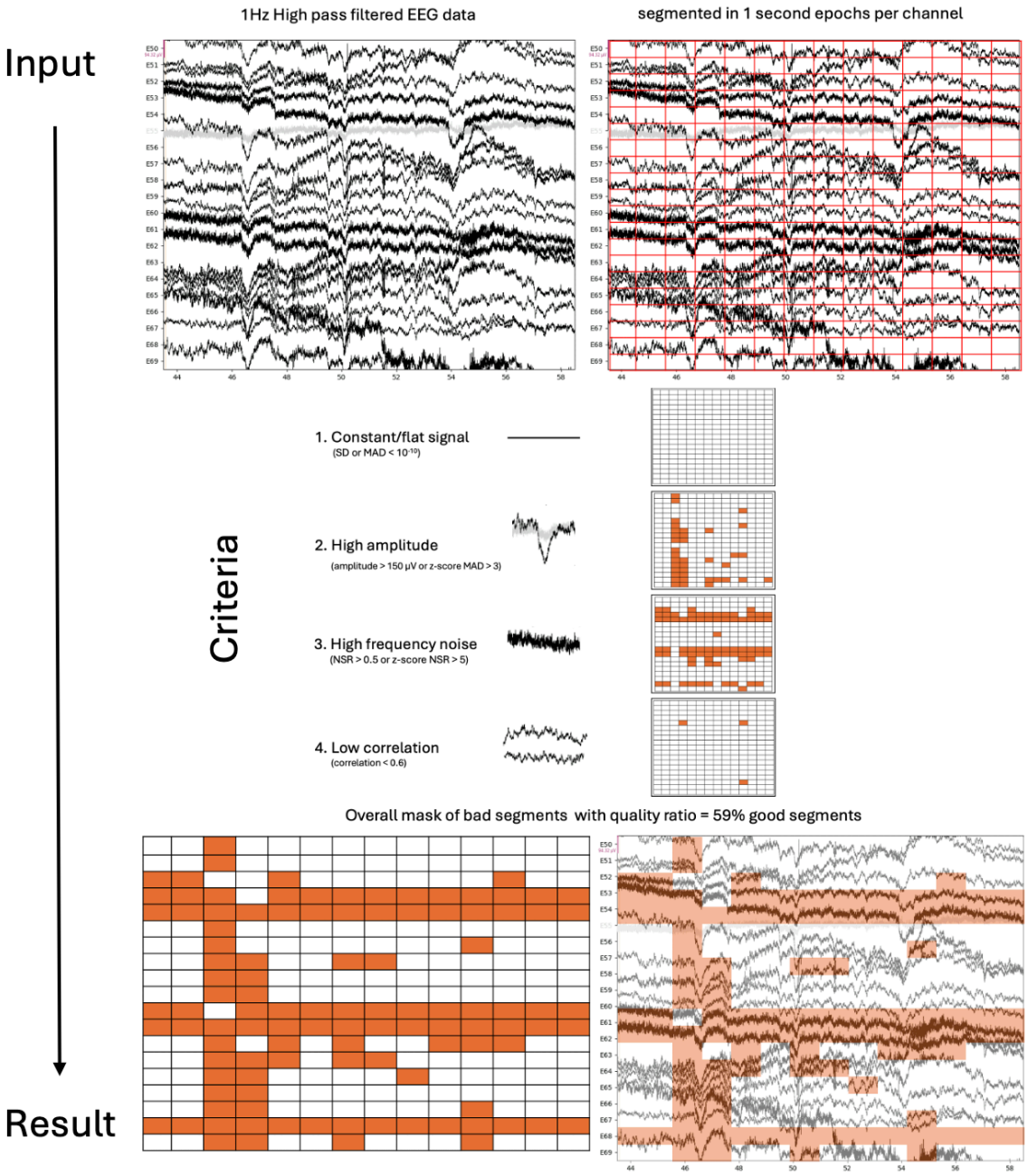

Figure B.1: Flow diagram of the automated EEG quality assessment method. First, 1 Hz high pass filtered EEG data is segmented into one-second epochs per channel. Each segment is then evaluated on four criteria: flat signal, high amplitude, high-frequency noise, and low correlation. Boolean masks are created for each criterion, where segments containing artifacts are marked. These masks are combined using a logical maximum operation, resulting in a final mask representing all artifactual segments. From this, a quality score is calculated as the ratio of clean segments to the total number of segments.

power of the original signal is divided by the signal removed by preprocessing. The total SER is then calculated by taking the weighted average of all the individual SER values of the channels. The weighting is based on the mean power of the channel, with the idea that channels with more power contain more information and are therefore more important. The SER is calculated exclusively on the clean segments of the signal as identified by the data quality assessment. A high SER indicates good cleaning performance because the power of the removed signal should be small, as the clean segments do not contain large artifacts.

$$SER_i = 10 \log_{10} \frac{E\left\{\left(\text{original signal}_i\right)^2\right\}}{E\left\{\left(\text{removed signal}_i\right)^2\right\}} (\text{ for clean segments }) \tag{B.1}$$

In addition to the SER, the correlation of the signal before and after preprocessing is added as a signal preservation measure to determine whether the information in the signal remains mostly intact. This is an addition to the SER, as the SER only provides information on the power ratio and not on the informational content. The Pearson correlation coefficient is calculated with the Numpy corrcoef function (Harris et al., 2020).

B.3. Artifact removal

The artifact to residue ratio is used to quantify artifact removal and is implemented as described in Bailey et al., 2023. The ARR is calculated per channel using Equation B.2, where the power of the removed signal is divided by the signal that remains after preprocessing. The total ARR is then calculated by taking the weighted average of all the individual ARR values of the channels. The weighting is based on the mean power of the signal that was removed from the channel, with the idea that channels where most power is removed contain more artifacts and are therefore more important for assessing artifact removal. The ARR is calculated exclusively on the artifact segments of the signal identified by the data quality assessment. A high ARR indicates good cleaning performance because the power of the removed signal should be large, as the artifact segments contain overwhelmingly large artifacts compared to the neurological signal.

$$ARR_i = 10 \log_{10} \frac{E\left\{\left(\text{removed signal}_i\right)^2\right\}}{E\left\{\left(\text{cleaned signal}_i\right)^2\right\}} (\text{ for segments with artifacts }) \tag{B.2}$$

In addition to the ARR, the mutual information between the removed and cleaned signals is added as a measure of artifact removal. The mutual information provides an indication of whether the noise is correctly separated from the neurological signal, as these are assumed to be two independent signals. If there is mutual information left, this either means that there is still artifact left in the cleaned signal or that there is neurological information removed with the artifacts. This is an addition to the ARR, as the ARR only provides information on the power ratio and not on the informational content. The mutual information is estimated using the mutual information regression feature from scikit-learn (Pedregosa et al., 2011).

B.4. Strength of Berger effect

The reduction in occipital alpha power following eye opening is known as the Berger effect, which is a robust and well-documented neural response Barry et al., 2007; Berger, 1929. This effect is used as a benchmark to evaluate whether the neural signal is preserved and distinguishable after preprocessing. Alpha power is measured in resting state with eyes open and closed separately by calculating the mean power in the individual alpha band of the normalized PSD, which was defined as the individual alpha peak frequency (iAPF) ± 2 Hz. The iAPF was defined as the highest peak between 6 and 12 Hz in the power spectrum of the eyes closed resting state. To assess the strength of the Berger effect, a Bayesian t-test is performed on the root mean square power difference in the alpha band between the eyes open and eyes closed resting state conditions. A large effect size indicates that relevant neural signals are retained, which is crucial for subsequent experimental analyses Clayson et al., 2021.

B.5. Stability

The stability of the preprocessing performance is important to ensure that all types of EEG files are well improved with the preprocessing pipeline and that the preprocessing pipeline is not overfitted to a specific EEG file. Therefore, the quality before and after preprocessing was measured for all preprocessing pipelines, and the difference is analyzed using the Wilcoxon t-test. Additionally, Cohen's d is calculated to show the stability of the difference, and the median deviation indicates the effect size to compare between pipelines.

B.6. Visual inspection

Visual inspection is included as a complementary step to evaluate the performance of the preprocessing pipelines. Although objective metrics provide valuable and standardized information regarding signal preservation and artifact removal, they can not account for all exceptions or data-specific nuances. Therefore, manual inspection remains essential to assess whether artifacts are effectively removed and whether meaningful neural signals are preserved. Special attention is given to features such as epileptic discharges, which may resemble artifacts in their temporal and spectral characteristics but originate from neural sources. If such signals are removed, it may indicate that other, less visually apparent neural signals are also lost. Visual inspection thus helps protect against overcorrection and supports the interpretation of objective metrics by ensuring that preprocessing does not compromise the neurophysiological validity of the data.

C

Suggestions for future improvements of the developed preprocessing pipeline

A key step toward improvement is enhancing the generalizability of the pipeline. This could be achieved by systematically examining which recordings respond well and which do not, and whether consistent factors underlie these differences (e.g., signal quality, artifact profile, or task conditions). Based on these findings, the pipeline can be more precisely tailored to address the shortcomings observed in less responsive recordings, an effect that was not present in the concept pipelines.

In addition, the combination of preprocessing steps appears to be a promising strategy. This approach has already been used in the literature and is supported by the complementary strengths observed in this study. For instance, muscle artifacts were dominant in many recordings, which made the MWF less effective as a standalone method. However, the ICLabel-based pipeline removed these artifacts. Conversely, eye movement artifacts remained partially present in the ICLabel pipeline but could potentially be addressed using the MWF. The limitations of individual methods can be mitigated by strategically combining preprocessing techniques, such as MWF and ICA. A detailed inspection of ICA components and their classification, especially in cases where artifact removal remains suboptimal, may help guide further refinement.

Lastly, another important aspect concerning the optimization of preprocessing is the difference between resting-state recording and EEG recording during cognitive tasks. To maximize the number of ICA components, the pipeline was fitted on at least 330 seconds of recording. As the resting-state recordings alone were shorter, data from task-related segments were added to reach this threshold value. However, because only resting-state data were included in the final analysis, this may have affected the pipeline's performance, specifically for resting-state EEG. A separate ICA fitting on only resting-state data, or alternatively reducing the number of components, might improve artifact removal without compromising the integrity of the resting-state signal. Systematically investigating preprocessing performance on resting-state versus task segments could thus be a valuable next step in refining the pipeline. For future research, it might be beneficial to make the preprocessing pipeline modular to adapt to the type of data that is analyzed.

In addition to further optimizing the current pediatric pipeline, an alternative approach could be to explore mature Python-based preprocessing pipelines developed for adult EEG, such as PyLINE Anijärv and Campbell, 2024. Although PyLINE shows promise as a standardized and modular tool, it currently lacks extensive validation, particularly in pediatric populations. Investigating its applicability and necessary adaptations for pediatric EEG could provide a valuable complementary strategy for robust preprocessing.

D

Quality assessment - validation

D.1. Methods

A total of 71 EEG resting state recordings were evaluated to validate the automated quality assessment, resulting in a quality ratio of good segments (Q-score). Each recording was assessed using the automated pipeline and independently by experienced neurologists who provided visual quality ratings (N-scores). The N-score was used as a reference measure in this validation because visual inspection remains the clinical standard for evaluating EEG quality. Although inherently subjective, such assessments are grounded in extensive clinical experience and are widely accepted in clinical and research settings.

To determine the validity of the Q-score, its relationship with the N-score was assessed using the Spearman rank correlation. This non-parametric method is chosen because the neurologists' scores are ordinal and do not satisfy the assumptions of parametric tests as the difference between the ratings of the neurologists is not guaranteed to be equal for every scale.

D.2. Results

Figure D.1 shows the relationship between the quality rating of the neurologists (N-scores) and the automatically computed quality scores (Q-scores) in a scatter plot. A positive correlation was observed between the N-scores and Q-scores ($\rho = 0.243$, $p = 0.04139$). An example of a 15 s recording assessed using the automated pipeline is shown in Figure D.2.

D.3. Discussion

The moderate correlation between the neurologists' visual ratings (N-scores) and the automated quality scores (Q-scores) indicates that the automated assessment sufficiently reflects the expert judgment. This supports the use of the automated score as a practical and objective tool for evaluating EEG data quality, particularly in the context of comparing preprocessing pipelines.

However, a potential limitation of this comparison lies in the difference in viewing environments. Neurologists assessed the EEG recordings using a clinical viewer that applies automatic settings that may be influenced by excessive artifacts, possibly leading to an underestimation of data quality. In contrast, the automated scores were generated on a research server, where a dedicated EEG viewer allowed for more detailed inspection.

In conclusion, despite the differences in viewing environment and the inherent subjectivity of visual scoring, the observed correspondence supports the utility of the automated score as a reliable indicator of EEG data quality.

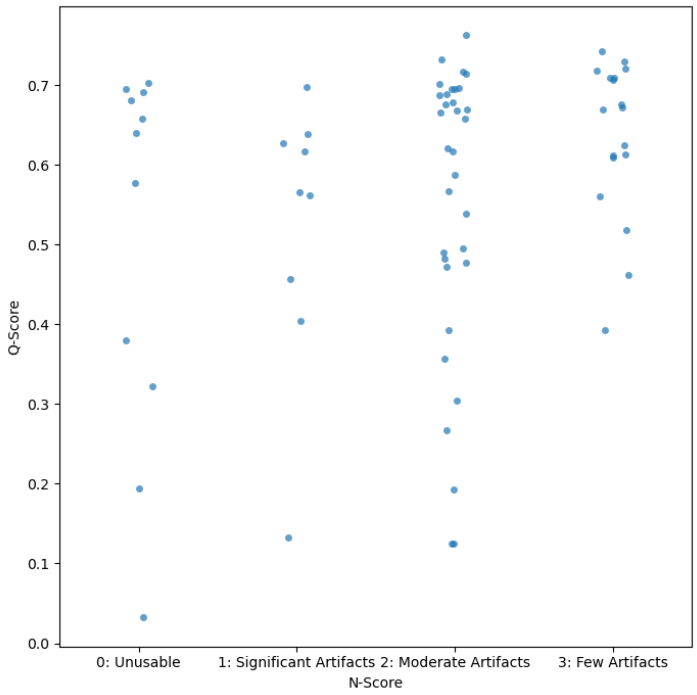

Figure D.1: Relationship between neurologists' quality ratings and the automatically computed quality score. The Scatter plot shows individual data points, with higher neurologist ratings (N-scores) generally corresponding to higher automated quality scores (Q-scores).

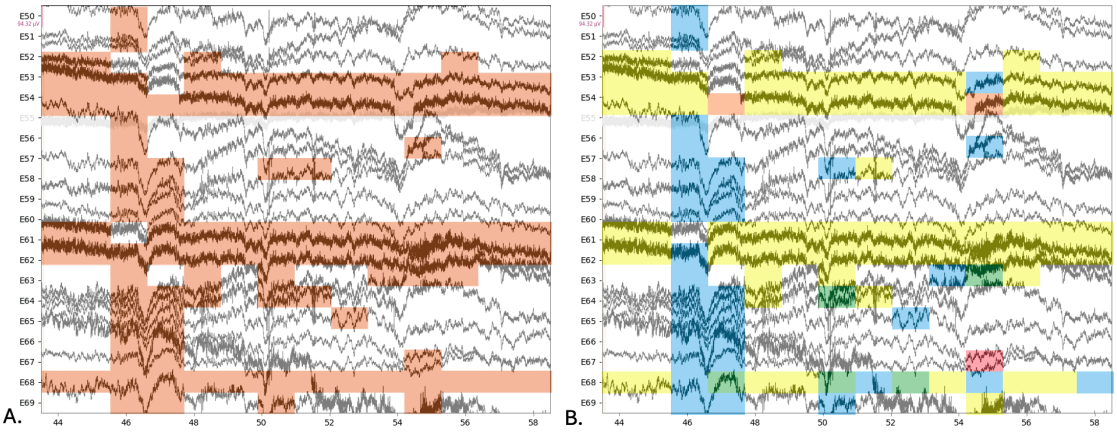

Figure D.2: Example of 15 seconds from a recording assessed with the automated quality pipeline. (A) Overall mask of the bad windows. (B) Individual masks per criteria: yellow - HF noise, blue - extreme amplitude, red - low correlation.