

Review

Not peer-reviewed version

---

# How to Process Infant Electroencephalography (Eeg) Data: A Critical Review of 10 Preprocessing Pipelines on Challenges, Strengths, and Future Directions

---

[Efthymios Papatzikis](#) \*

Posted Date: 28 April 2025

doi: 10.20944/preprints202504.2375.v1

Keywords: Infant EEG; EEG Preprocessing Pipelines; EEG Processing; EEG Pipelines Evaluation; Electroencephalography Analysis



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Review

# How to Process Infant Electroencephalography (EEG) Data: A Critical Review of 10 Preprocessing Pipelines on Challenges, Strengths, and Future Directions

Efthymios Papatzikis<sup>1,2</sup>

<sup>1</sup> Department of Early Childhood Education and Care, Oslo Metropolitan University, Oslo, Norway; efp331@mail.harvard.edu

<sup>2</sup> Advanced Health Intelligence and Brain-Inspired Technologies (ADEPT) Research Group, Oslo Metropolitan University, Oslo, Norway

**Abstract:** Infant electroencephalography (EEG) is an essential tool for understanding early brain development, yet it presents unique challenges such as frequent movement artifacts, low signal-to-noise ratio (SNR), and rapid developmental changes. In recent years, specialized preprocessing pipelines have been developed to address these issues and improve the accuracy of infant EEG data analysis. In alignment with this evolving research landscape, this critical review performs a comparative analysis of ten prominent preprocessing pipelines – APICE, ADJUST, GADS, BEAPP, HAPPE+ER, HAPPE, HAPPILEE, MADE, NEAR, and the Modular pipeline – focusing on five key criteria: artifact handling, automation and scalability, flexibility and adaptability, developmental sensitivity, and validation against empirical data. The review employs a multi-dimensional analysis framework developed using a reverse design approach, where the analysis criteria emerged organically through an in-depth study of the pipelines' specific methods and applications. The comparative analysis involved benchmarking each pipeline against the five criteria using a structured system of qualitative descriptors translated into numerical indices, enabling in the end a more robust comparative overview of their strengths and limitations. Results of the critical review reveal that specific pipelines excel in automation and flexibility, making them suitable for large-scale or multi-site studies, while others demonstrate exceptional developmental sensitivity, specifically addressing the needs of neonatal EEG data. Overall, no single pipeline was found to be universally superior, but each contributes uniquely depending on specific research requirements. Future directions include enhancing automation, improving developmental sensitivity, and promoting empirical validation to ensure reproducibility and adaptability across diverse datasets. This critical analysis and review provides a foundation for researchers and clinicians to make informed decisions about preprocessing strategies in infant EEG studies, advancing the field of developmental neuroscience.

**Keywords:** infant EEG; EEG preprocessing pipelines; EEG processing; EEG pipelines evaluation; electroencephalography analysis

---

## 1. Introduction

Infant electroencephalography (EEG) has emerged as an indispensable tool in understanding early brain development, providing researchers with insights into neural processes from birth through the first years of life. However, the technical challenges inherent in recording EEG from infants – such as frequent movement, low signal-to-noise ratio (SNR), and rapid developmental changes – have necessitated the development of specialized preprocessing pipelines. More specifically, frequent infant movement during naturalistic settings introduces significant motion artifacts, which can significantly alter EEG spectral signatures and complicate data interpretation (Georgieva et al., 2020). These motion artifacts, including jaw and limb movements, generate distortions that primarily affect the beta frequency band, particularly over peripheral electrode sites.

Furthermore, achieving a high signal-to-noise ratio (SNR) in infant EEG recordings presents unique challenges. Conventional signal analysis often fails to provide optimal data quality in infants due to anatomical and behavioral differences compared to adults. For instance, studies have demonstrated that using weighted averaging and soft thresholding techniques can effectively improve SNR, particularly in high-risk newborns, by reducing the impact of noise and artifacts present in the EEG signals (Papatzikis et al., 2024; Falkner & Abedi, 2010). Additionally, spectral analysis approaches, particularly using Fast Fourier Transformation, have also been shown to be useful for quantifying EEG signals, especially in the context of reducing artifacts and improving reliability in preterm infants (Victor et al., 2005).

Most importantly, the rapid developmental changes in sleep patterns and EEG signal complexity during the first weeks of life add another great layer of complexity to infant EEG analysis. The maturation of sleep states is associated with evolving EEG characteristics, such as changes in entropy and spectral power (Wielek et al., 2019). Therefore, these dynamics necessitate sophisticated adaptive techniques, including for example machine learning classifiers, to effectively capture and interpret these changing neural signatures over time.

In this context, numerous preprocessing pipelines have been recently introduced to improve the quality of infant EEG data, by automating artifact rejection, and streamlining data processing. These pipelines use advanced techniques to enhance, at the preprocessing level of the signal, data reliability, and ensure that meaningful neural signals can be extracted from noisy EEG recordings. In research, this means that studies on infant brain development can be conducted with higher data accuracy and reliability, leading to more robust conclusions, while clinically, these pipelines may facilitate better monitoring of infants' neural health by minimizing the impact of artifacts and noise, thus enabling earlier identification of neurodevelopmental issues and more precise assessments.

However, despite the advancements in pipeline development, there remains a significant gap in the literature regarding a comprehensive approach and review of these tools in terms of their profiles, effectiveness, adaptability, and suitability for various infant populations. Most existing literature focuses on the technical aspects of specific pipelines – either individually or in small groups – but there is limited overall critical analysis across different approaches. As the varying methodologies and target applications of these pipelines make it challenging for researchers and clinicians to determine which pipeline is most suitable for their specific needs, a critical review of these infant EEG signal preprocessing pipelines is essential to synthesize current knowledge, identify gaps, and provide guidance on best practices for the field.

In this critical review, I discuss ten key infant EEG pipelines capable of analyzing infant EEG as presented in recent literature, examining their strengths, limitations, and potential impact on the field. These pipelines – APICE (Flo et al., 2022), ADJUST (Leach et al., 2020), GADS (Stevenson et al., 2014), BEAPP (Levin et al., 2018), HAPPE+ER (Monachino et al., 2022), HAPPE (Gabard-Durnam et al., 2018), HAPPILEE (Lopez et al., 2022), MADE (Debnath et al., 2020), NEAR (Kumaravel et al., 2022), and the modular pipeline (Coelli et al., 2024; called Modular from this point on) – are approached through a critical lens to highlight their contributions and areas for improvement.

### *1.1. Overview of the Infant EEG Pipelines*

One prominent approach to improving infant EEG preprocessing involves the automation of artifact detection and rejection, as exemplified by the Automated Pipeline for Infants Continuous EEG (APICE) developed by Flo et al. (2022) and the adapted algorithm Adjusted-ADJUST by Leach et al. (2020). On the one hand, APICE is a MATLAB-based pipeline that integrates with EEGLAB to efficiently preprocess continuous EEG data from infants, and it is specifically designed to handle the unique challenges of this special population, including high artifact loads and variability in signal quality. The pipeline employs adaptive thresholds and iterative artifact rejection techniques that target both physiological (e.g., heartbeats, muscle activity) and non-physiological artifacts (e.g., line noise). Notably, APICE uses a combination of Independent Component Analysis (ICA), Denoising Source Separation (DSS), and transient artifact correction, making it highly versatile for different

types of infant EEG data. The preprocessing is done on continuous data, which improves data recovery and ensures greater flexibility in downstream analysis. On the other hand, the Adjusted-ADJUST algorithm by Leach et al. (2020) also provides significant advancements tailored for infant EEG data, specifically addressing the limitations of traditional adult-focused artifact removal methods. Adjusted-ADJUST modifies the original ADJUST algorithm by optimizing the detection of ocular and muscle artifacts, which are more frequent and less predictable in infants compared to adults, while it enhances artifact classification by integrating spatial and temporal features specific to pediatric data, such as modifying the parameters used for detecting blinks and horizontal eye movements to better accommodate the increased noise and variability found in this population. Additionally, adjusted-ADJUST incorporates features like automated classification through ICLabel and checks for alpha peak preservation, ensuring that neural activity is retained while reducing false artifact detection.

Tailored specifically for neonatal EEG, the Generalized Artifact Detection System (GADS) by Stevenson et al. (2014) and the Neonatal EEG Artifact Removal (NEAR) pipeline by Kumaravel et al. (2022) employ specialized techniques to handle the high artifact load characteristic of recordings from newborns. GADS utilizes a two-stage system based on Support Vector Machines (SVMs) to distinguish between major and minor artifacts. It extracts 14 features from EEG epochs at multiple time scales – ranging from 2 to 32 seconds – such as mean amplitude, frequency-band energies, and the Hurst exponent, to effectively classify artifacts. This multi-scale approach provides context for both short and long-duration artifacts, enabling the system to perform well in detecting different artifact types, from muscle movements to electrode disconnections, while its machine learning-based approach ensures robust detection even with high inter- and intra-patient variability typical in neonatal EEG data. NEAR, in contrast, employs a different strategy tailored for newborn-specific artifacts by using a combination of the Local Outlier Factor (LOF) and Artifact Subspace Reconstruction (ASR). LOF is a density-based algorithm that effectively identifies bad channels by analyzing their local density compared to neighboring channels, making it suitable for the noisy, inconsistent data commonly found in neonatal recordings. Following bad channel identification, NEAR applies ASR to remove transient high-amplitude artifacts while preserving the underlying neural signals. This two-step approach ensures that both persistent and transient artifacts are addressed effectively, making NEAR particularly adept at handling non-stereotyped artifacts like head and arm movements.

Building upon the foundation laid by Gabard-Durnam et al. (2018) with the original HAPPE pipeline, subsequent developments like HAPPE+ER (Monachino et al., 2022) and HAPPILEE (Lopez et al., 2022) have further refined and specialized preprocessing techniques for infant EEG through this 'sequel' of preprocessing systems. More specifically, HAPPE automates artifact rejection and is optimized for datasets with low electrode counts, which is a common limitation in infant EEG studies due to the sensitivity of infant scalps. It integrates multiple preprocessing steps, including automatic channel rejection, de-trending, and ICA, making it a comprehensive solution for initial artifact handling. HAPPE+ER expands on these capabilities by specifically optimizing for event-related potential (ERP) analyses, while including a series of automated steps such as event marker detection, epoch segmentation, and enhanced artifact removal to handle the complexities associated with ERP paradigms. Additionally, HAPPE+ER incorporates wavelet-based denoising and supports automated generation of ERP quality reports, allowing researchers to evaluate both data quality and pipeline effectiveness consistently across developmental datasets. This is particularly beneficial for studies examining cognitive development, where ERP responses to stimuli provide crucial insights. HAPPILEE, on the other hand, is designed specifically for low-density EEG setups, which are often necessary in pediatric or field-based research contexts where high-density electrode caps may be impractical. This pipeline processes both resting-state and task-based EEG recordings and integrates artifact rejection methods optimized for fewer electrodes, while it uses modified bad channel detection techniques, such as power spectrum analysis and correlation-based evaluations, to identify and manage noisy channels effectively. HAPPILEE integrates seamlessly with HAPPE+ER



for ERP analysis, providing a tailored solution for managing lower density data without compromising preprocessing rigor, making it ideal for portable or community-based research in young children.

For large-scale and longitudinal studies, platforms like the Batch Electroencephalography Automated Processing Platform (BEAPP) by Levin et al. (2018) and the modular EEG preprocessing pipeline by Coelli et al. (2024) provide flexible and efficient solutions, each with unique features tailored to specific needs. BEAPP operates on a modular framework that integrates several freely available software tools, such as EEGLAB, FieldTrip, and the PREP pipeline, to streamline preprocessing in a batch mode. It offers users the ability to manage EEG files from multiple acquisition setups, perform re-referencing, minimize artifacts, and conduct spectral analyses through customizable settings, while including a graphical user interface (GUI), which facilitates the selection of modules and parameters, ensuring reproducibility and allowing non-programmers to effectively utilize complex EEG processing tools. BEAPP seems to be particularly useful in multisite and longitudinal studies due to its capability to handle EEG datasets collected under diverse conditions, making it ideal for scaling up infant EEG research with consistent preprocessing quality. The Modular pipeline by Coelli et al. (2024), on the other hand, emphasizes the flexibility of method selection at each preprocessing step. It includes a comparative framework that allows researchers to select from different ICA methods (e.g., Extended Infomax and Second Order Blind Identification (SOBI)) and re-referencing strategies (e.g., Common Average Reference (CAR), robust-CAR, Reference Electrode Standardization Technique (REST), and Reference Electrode Standardization Interpolation Technique (RESIT)). This modular structure provides a framework for objectively assessing the impact of each preprocessing choice on signal quality using quantitative indicators such as Mutual Information Reduction (MIR) and automatic Independent Component (IC) classification metrics. Such flexibility is particularly beneficial in the context of infant EEG, where artifact types and signal quality can vary widely, requiring an adaptable approach that can optimize preprocessing for each specific dataset.

Finally, addressing the rapid developmental changes in neural activity across early developmental stages, the Maryland Analysis of Developmental EEG (MADE) pipeline by Debnath et al. (2020) offers specialized preprocessing tailored for developmental EEG research. MADE employs a combination of automated artifact rejection methods, such as ASR and ICA, to address the high variability and artifact prevalence in infant EEG. Additionally, it includes a feature-based selection of components to retain neural signals pertinent to developmental research, ensuring that neural activity related to maturation profiles in different stages is preserved. Most importantly, the pipeline also integrates developmental-specific filtering and re-referencing methods designed to minimize the loss of crucial developmental signals often masked by standard adult-focused preprocessing approaches. This focus on retaining the integrity of age-related EEG features makes MADE a robust tool for providing more accurate insights into the evolving brain activity of infants, thereby enhancing the understanding of early brain development.

## 2. Methods

Building upon this initial summative overview, a comparative study and analysis of the selected infant EEG preprocessing pipelines seemed to be crucial to understand their effectiveness, weaknesses and overall applicability in research. As a result, I decided to employ a multi-dimensional benchmarking framework focusing on five core criteria that appear to be particularly relevant to the nature of infant EEG data – that is, artifact handling, automation and scalability, flexibility and adaptability, developmental sensitivity, and validation against empirical data.

These criteria were selected following a reverse design approach, which emerged as I engaged deeply with the content and methodological processes of the 10 selected pipelines. In the absence of any other relevant framework to rely on, and rather than loosely predefining my analysis criteria at the outset, I allowed the criteria to crystallize organically through a comprehensive study of the pipelines themselves, involving a thorough review of the specific techniques, goals, and practical

implementations within each pipeline, reflecting relative performance based on available evidence and documented use cases from their creators/authors. I found for this reverse design approach to be particularly effective, as the pipelines revealed which criteria were most relevant and necessary for robust critical analysis, aligning it in result closely with real-world applications, and ensuring that the chosen criteria are both empirically grounded and reflective of the current state of the field, as well as leading to a more nuanced and context-sensitive review.

*Artifact Handling:* One of the most critical aspects of benchmarking these pipelines seemed to be how effectively they manage artifacts, given the susceptibility of infant EEG to various sources of interference, such as movement and muscle activity which usually result in a much lower signal-to-noise ratio (SNR) than in the adult EEG. Infant EEG is notoriously prone to this wide array of artifacts, and in this context, effective artifact handling requires the use of advanced techniques such as Independent Component Analysis (ICA), adaptive filtering, or even machine learning algorithms tailored specifically to the high level of noise present in infant EEG recordings. Pipelines that strike a balance between optimizing artifact rejection while preserving neural signal integrity are particularly important.

*Automation and Scalability:* Equally essential seemed to be the pipeline's ability to automate processes (for example, the pipeline's reliance on user input for artifact rejection) and scale efficiently (for example, the pipeline's ability to perform batch analyses across datasets, particularly in the context of longitudinal studies). The capacity for automation – that is, minimizing manual intervention and thus reducing human error – seems to directly impacting the consistency and reliability of EEG processing, while scalability, on the other hand, addresses how well the pipeline manages large, often complex datasets involving either many hours of continuous EEG recordings, or recordings with a lot of different measurement conditions. Pipelines that require minimal manual adjustments while supporting high-throughput data processing seem to be far more practical for research conducted in both clinical and experimental settings.

*Flexibility and Adaptability:* As already mentioned, infant EEG data vary significantly between subjects due to developmental stages, scalp physiology, and recording environments. As a result, the ability of a pipeline to be customized to suit different datasets and research questions becomes crucial. In this sense, 'adaptability' includes the pipeline's capability to adjust filtering settings, integrate various preprocessing methods, and incorporate user-defined parameters, while extending to different EEG systems, including both high- and low-density electrode arrays. Pipelines that offer modularity, may allow researchers to selectively implement processing steps that are more tailored to their specific datasets, and especially in the infant EEG research context, may be better suited to handle the present diversity.

*Developmental sensitivity:* Infant brain activity evolves rapidly in the early stages of life, particularly in the context of prematurity (Benders et al., 2015; Tataranno et al., 2018). Consequently, it is crucial for preprocessing pipelines to account for these developmental changes to ensure the accurate representation of neural signals. To achieve this, pipelines must effectively distinguish between developmental variations and meaningful neural activity, preserving age-specific features of brain activity rather than misclassifying them as artifacts. Pipelines that incorporate developmental considerations – such as optimizing ICA parameters or applying age-specific thresholds for artifact rejection – are especially valuable in the context of developmental neuroscience research.

*Validation against empirical data:* The ability to generate clean, analyzable data that supports reliable conclusions about brain function in the infant EEG context is the true test of a pipeline's utility. As a result, pipelines that have been validated across multiple datasets, research labs, and study designs should be considered more robust. Moreover, those that report quantitative metrics such as sensitivity and specificity in artifact detection, or provide comparisons with manual preprocessing, should be deemed more reliable. Openly available code and clear documentation further enhance a pipeline's credibility, as transparency is vital for ensuring reproducibility and enabling scrutiny by the broader scientific community (Thimbleby, 2024; Rupprecht et al., 2020).

Based on these five criteria, I developed qualitative descriptors to systematically benchmark each pipeline's capabilities. These qualitative descriptors were translated into numerical indices (tables 1-5) to ensure a uniform, quantifiable profile for each criterion. Throughout the review process – which involved carefully reviewing each pipeline's documentation and using the qualitative descriptors to identify strengths and weaknesses – I employed this system to systematically quantify the capabilities of each pipeline across the five criteria, ultimately producing a structured comparative analysis.

Table 1. ‘Artifact Handling’ descriptors and numerical scale.

Artifact Handling	
This criterion benchmarks how well the pipeline deals with artifacts, such as muscle movements or electrical noise, which can corrupt EEG signals.	
Index	Artifact Handling Description
1	Minimal artifact handling, limited to manual inspection or rejection. No automated techniques used.
2	Basic artifact removal using standard filters (e.g., high-pass/low-pass) with minimal automation.
3	Uses ICA or similar standard techniques to address common artifacts but lacks optimization for specific issues in infant EEG (e.g., eye or muscle artifacts).
4	Employs multiple automated techniques, such as ICA along with source separation or adaptive filtering, optimized for infants. Targets both common and specific types of artifacts.
5	Utilizes advanced methods such as machine learning or sophisticated noise models to specifically target different artifact types in infant EEG, including head movement, muscle noise, and environmental noise. Includes validation of artifact removal to ensure neural signal preservation.

Table 2. ‘Automation and Scalability’ descriptors and numerical scale.

Automation and Scalability	
This criterion benchmarks the degree of automation in the pipeline and how easily it can be scaled to handle multiple datasets or large-scale projects.	
Index	Automation and Scalability Description
1	No automation; manual intervention required for every step. Poor scalability; designed for single dataset use.
2	Some automation (e.g., predefined script runs); however, most steps still require manual adjustments. Limited batch processing ability.
3	Moderate automation; key steps are automated, but user intervention is needed for artifact rejection or threshold adjustments. Limited but functional batch processing for smaller datasets.
4	Highly automated with most preprocessing steps handled without user input. Batch processing available for larger datasets, allowing for moderate scalability.

5	Fully automated pipeline with end-to-end processing capabilities. Strong batch processing and scalability across multiple datasets, supporting high-throughput EEG studies. Suitable for longitudinal, multi-site research.
---	---

Table 3. ‘Flexibility and Adaptability’ descriptors and numerical scale.

Flexibility and Adaptability	
This criterion benchmarks how flexible and adaptable the pipeline is to different experimental settings, data types, and research questions.	
Index	Flexibility and Adaptability Description
1	Very rigid; designed for one specific dataset or configuration without customization options. No flexibility in parameter tuning.
2	Limited flexibility; some parameters can be adjusted, but the pipeline is not easily adaptable to different EEG systems or experimental designs.
3	Moderately flexible; configurable parameters are available, allowing some adaptation to different datasets and experimental needs. Can accommodate different electrode counts and settings.
4	High flexibility; pipeline can be adapted to multiple EEG acquisition systems and experimental paradigms. Many configurable parameters are available for artifact rejection, filtering, etc. Suitable for a range of infant EEG studies.
5	Extremely flexible; modular approach with customizable processing steps, allowing for tailored analysis for different research questions, environments, and EEG systems. Supports both high-density and low-density electrode arrays. Can easily accommodate novel study designs.

Table 4. ‘Developmental Sensitivity’ descriptors and numerical scale.

Developmental Sensitivity	
This criterion benchmarks how well the pipeline accounts for developmental changes in infants, such as age-related adjustments to EEG characteristics.	
Index	Developmental Sensitivity Description
1	No specific considerations for age or developmental stage. Uses parameters optimized for adults or generalized settings.
2	Minimal consideration for development; uses fixed parameters that are marginally suitable for infants. Limited adaptation to age-specific features.
3	Moderate developmental sensitivity; some age-specific parameters are used, and developmental considerations are made, but not consistently throughout the pipeline.
4	High developmental sensitivity; uses age-specific thresholds and considerations for different stages of infant development. Integrates specialized preprocessing based on age or developmental stage.



5	Fully optimized for developmental changes; extensive use of age-appropriate algorithms, thresholds, and metrics throughout the pipeline. Includes adjustments for different infant age groups (e.g., preterm vs. full-term), and integrates empirical knowledge of infant brain development. Validation shows strong performance across different developmental stages.
---	---

Table 5. ‘Validation against Empirical Data’ descriptors and numerical scale.

Validation Against Empirical Data	
This criterion benchmarks how well the pipeline has been validated, particularly with empirical data, to ensure reliability and reproducibility of its outputs.	
Index	Validation Against Empirical Data Description
1	No validation against empirical data. No reported quantitative metrics on pipeline performance.
2	Limited validation; tested on a small dataset with minimal empirical evidence. No cross-validation or use of independent datasets.
3	Moderate validation; tested on a larger sample size, but primarily within a single dataset. Some quantitative metrics reported, such as sensitivity or specificity.
4	Good validation; cross-validated on multiple independent datasets or different environments. Empirical metrics, such as accuracy, sensitivity, and specificity, are well-reported.
5	Extensive validation with empirical data from multiple sources and across different populations. Includes comparisons with manual preprocessing and quantitative performance metrics. Demonstrates high reliability and reproducibility in various research settings. Openly accessible results or data are available for verification by the broader research community.

3. Results

3.1. Individual Analysis

*APICE*: *APICE* demonstrates strong versatility in handling high artifact loads with iterative artifact detection and adaptive thresholds using multiple algorithms, such as DSS and ICA to target both physiological and non-physiological artifacts. The use of adaptive thresholds, which are based on the distribution of voltage values for each subject, allows *APICE* to detect and correct artifacts in a targeted manner, especially in high-artifact datasets like those collected from infants. This subject-level adaptability enhances the robustness of artifact detection across different populations without the need for main adjustments. The pipeline also performs artifact correction on continuous data, incorporating adaptive correction methods such as automated drift removal and targeted denoising, significantly enhancing data recovery before epoching. This approach is particularly effective in retaining signal integrity by reducing the impact of non-stereotyped artifacts over long recording periods. *APICE* excelled in artifact handling (index of 5), automation and scalability (index of 5), and flexibility (index of 5). However, although artifact handling received a descriptor index of 5, it is important to note that the effectiveness of ICA was limited in neonate data, which could temper the final choice of descriptor slightly when considering this specific subset. Developmental sensitivity

received an index of 4, indicating room for improvement in specific optimizations for neonatal EEG – as mentioned above –, particularly due to the high inter-trial variability and developmental changes affecting artifact decomposition techniques like ICA and DSS. Finally, validation against empirical data received an index of 4, reflecting its strong but not exhaustive validation efforts. This descriptor index takes into account specific limitations, such as the marginal effect of ICA or DSS improvements across the different datasets evaluate in the literature.

*Adjusted-ADJUST*: The Adjusted-ADJUST pipeline is notable for enhancing artifact detection using modified ICA and additional algorithms tailored specifically to infant EEG data collected with geodesic nets. It employs modifications aimed at handling the unique properties of infant data, such as adjustments for ocular artifact detection and modifications for blink and horizontal eye movement detection. Specifically, the ocular artifact detection algorithms were adjusted by changing the spatial feature parameters and removing the temporal measure for more robustness against noise, which is often higher in infant datasets. These features improve artifact classification for developmental EEG and make the pipeline more robust against increased noise levels. The authors of the algorithm compared the adjusted pipeline to the original-ADJUST and ICLabel algorithms in terms of three performance measures: classification agreement with expert coders, number of trials retained after artifact removal, and reliability of the EEG signal after preprocessing. Adjusted-ADJUST demonstrated higher classification agreement scores with expert coders compared to the original-ADJUST and ICLabel (i.e., the adjusted-ADJUST algorithm achieved an agreement score of 80.78% compared to 57.83% for original-ADJUST and 73.46% for ICLabel). Based on these characteristics, Adjusted-ADJUST received in my analysis an index of 4 for artifact handling due to its reliance on manual parameter adjustments, which however introduces variability and reduces full automation effectiveness. Flexibility and adaptability were indexed at 5, thanks to the pipeline's customization options and compatibility with external EEG tools like EEGLAB. The pipeline also allows users to modify parameters such as z-score thresholds for blink detection, enabling greater adaptability for different datasets. Developmental sensitivity was indexed at 4 when compared to other pipelines, as it includes specific adjustments for ocular artifacts, with the need of further refinements and validations to handle the physiological and behavioral variabilities across all developmental stages effectively, such as the differences in blink rates, skull density, and movement artifacts. Finally, validation against empirical data received an index of 3, reflecting the need for more extensive empirical validation across multiple diverse infant populations, including different developmental stages, acquisition setups, and experimental paradigms. Compared to other pipelines, such as APICE, NEAR, and HAPPE, which have been validated across a variety of datasets and paradigms, Adjusted-ADJUST lacks validation across multiple setups and does not seem to produce comprehensive quality metrics at each preprocessing stage.

*GADS*: The GADS pipeline utilizes a two-stage SVM-based approach to distinguish between major and minor artifacts in neonatal EEG, achieving high accuracy in artifact detection, especially for major artifacts with a median AUC of 1.00 (IQR: 0.95-1.00). The first stage of the SVM-based approach specifically employs 14 different features extracted from EEG epochs, including mean amplitude, frequency-band energies, Hurst exponent, and other measures, across different time scales (2s, 4s, 16s, and 32s epochs) to classify the artifacts effectively. In this approach, detection is relatively straightforward for major artifacts, due to significant differences in amplitude and frequency between background EEG and artifact, whereas the detection of minor artifacts is more challenging. The median AUC for minor artifacts presented to be 0.89 (IQR: 0.83-0.95), indicating substantial but variable accuracy. However, GADS is less consistent in detecting subtle low-amplitude artifacts, often failing to identify minor fluctuations that could still significantly impact the overall signal quality. For this quite comprehensive yet focused approach, GADS received an index of 4 for artifact handling, while its reliance on pre-set features, which are not optimized for different neonatal conditions, limits the pipeline's generalizability across diverse datasets, achieving an index of 3 for flexibility and adaptability. The pipeline received an index of 4 for automation and scalability, as it effectively automates artifact detection but lacks the flexibility of fully modular pipelines like

APICE or BEAPP. Unlike APICE, which offers adaptive thresholds and multiple algorithms for artifact detection, or BEAPP, which provides extensive batch processing capabilities across diverse acquisition systems, GADS relies on a more fixed approach, limiting its adaptability across different datasets. Developmental sensitivity also received an index of 4, as it addresses neonatal-specific EEG characteristics but lacks specific adaptations for older children, while the validation against empirical data received an index of 3 as it was performed specifically in a NICU setting, using leave-one-out cross-validation on a cohort of 51 neonates, which limits broader applicability.

**NEAR:** NEAR takes an innovative approach by employing LOF for bad channel detection and ASR for transient artifact removal, specifically calibrated for newborn EEG. LOF is a density-based algorithm that identifies bad channels by comparing the local density of data points to those of its neighbors, which is effective for detecting channels with atypical noise levels. To achieve optimal results, LOF is calibrated using the F1 Score metric (a measure that combines precision and recall to provide a balanced evaluation of classification performance), allowing NEAR to effectively distinguish between good and bad channels, even in challenging newborn datasets. ASR, on the other hand, is used to remove transient, high-amplitude artifacts, supporting two ASR processing modes: Correction (ASRC) and Removal (ASRR). The ASRR mode is generally preferred for newborn EEG, as it has been found to remove artifacts more effectively while preserving the underlying neural activity (e.g., ANOVA results: ASRR showed significant effects ( $F(1,13) = 5.13$ ,  $P = 0.041$ ) compared to ASRC ( $F(1,13) = 1.68$ ,  $P = 0.22$ ); power spectrum peak recovery was higher in ASRR; ERP effects were clearer with ASRR). In this context, NEAR excelled in artifact handling (index of 5) due to its novel use of LOF for channel detection and calibrated ASR for artifact removal. Developmental sensitivity was indexed at 4, reflecting its precise targeting of newborn EEG and tailored parameters for neonatal characteristics, although not incorporating parameters for prematurity, while automation and scalability were indexed at 4, featuring batch processing and adaptive calibration, though the requirement for manual adjustments for data for different experimental setups or different artifact profiles limited scalability. For this latter reason, flexibility and adaptability also received an index of 4. Finally, validation against empirical data received an index of 4, with successful validation across multiple neonatal datasets using metrics like the F1 Score. I did not assign an index of 5 in this case due to limitations such as the lack of validation across all developmental stages and other experimental contexts, which would be necessary to establish broader generalizability.

**HAPPE:** The Harvard Automated Processing Pipeline for EEG (HAPPE) utilizes Wavelet-enhanced Independent Component Analysis (W-ICA) and the Multiple Artifact Rejection Algorithm (MARA) to effectively separate artifacts and perform automated component rejection. Due to these advanced techniques that effectively reduce artifacts while preserving the EEG signal it achieved an index of 4 for artifact handling. However, it did not receive the highest index for this criterion because of challenges in handling certain complex artifacts and limitations in achieving full automation, necessitating manual intervention at times. Speaking about automation and scalability, these were indexed at 4, as the reliance on MATLAB limits scalability to broader environments. By the same token, although HAPPE includes batch processing capabilities, which enhances scalability for large datasets, the same dependency on MATLAB may pose challenges for integrating with other software platforms commonly used in research environments, justifying once more the assigned index for this specific criterion. In the same line of reasoning, flexibility and adaptability were indexed at 3, as customization options are limited due to the pipeline's primary optimization for MATLAB, restricting adaptation to different experimental setups, electrode configurations, or cross-platform use. However, the integration with BEAPP partially addresses these limitations by providing a user-friendly interface that supports diverse channel layouts and enhances flexibility in large-scale studies. Developmental sensitivity was also indexed at 3, indicating that more focused calibration for neonatal EEG characteristics is needed. While HAPPE includes some developmental adjustments, it lacks the granularity and optimizations essential for neonatal EEG, especially compared to specialized pipelines like NEAR, which are calibrated for neonatal-specific data. Finally, validation against

empirical data received an index of 4, supported by extensive testing on a large developmental dataset, including EEG data from infants and young children, demonstrating its effectiveness in high-artifact conditions. However, further validation across multiple contexts, such as different developmental stages and experimental paradigms, would confirm its robustness and generalizability for diverse developmental conditions.

*HAPPE+ER*: HAPPE+ER is an extension of HAPPE, specifically optimized for event-related potential (ERP) analyses. As an extension, it inherits both the strengths and limitations of the original HAPPE framework. For this reason, HAPPE+ER showcased slightly better capabilities from HAPPE, with indices of 4 for artifact handling, automation and scalability, and validation against empirical data. Flexibility and adaptability were indexed with a 3, while developmental sensitivity with a 4. In this extension of HAPPE, the improvements for ERP analyses include the addition of automated event marker detection, enhanced epoch segmentation, and wavelet-based denoising, which collectively enhance artifact handling and data quality especially in ERP settings. Additionally, developmental sensitivity was improved by incorporating event-specific preprocessing techniques that better account for the characteristics of ERP signals in infants, though optimizations for other developmental stages are still lacking. Despite these enhancements, limitations in customization and broader validation persist as in the HAPPE pipeline.

*HAPPILEE*: HAPPILEE is also designed as an extension of the HAPPE pipeline, but for low-density EEG setups. This pipeline extension was indexed at a high level in artifact handling (index of 4) due to the effective artifact correction methods tailored for low-density configurations, while automation, thanks to several automated processing steps that simplify data handling, was also indexed with a 4; though some complex artifacts still require manual intervention, and full automation across different research contexts remains a challenge. Flexibility and adaptability were indexed with a 4, reflecting the pipeline's focus on low-density configurations, which nevertheless limits broader applicability. Developmental sensitivity was indexed with a 3, as it lacks broader calibration for different developmental stages. The pipeline's current methods for bad channel detection, while functional for low-density setups, are not fully optimized for varying channel densities and often fail to account for the unique characteristics of neonatal and infant EEG, leading to less effective identification of bad channels – after all this is a specialized version and should be treated as such. Moreover, the thresholding system for segment rejection is not adequately adaptive, making it less effective for data with high artifact variability typically seen in infant populations. This results in either over-rejection or retention of noisy data, which impacts data quality and limits the pipeline's utility in diverse developmental research settings when also considered along its low-density set-up inclination. Finally, validation against empirical data was indexed at 4, reflecting robust validation efforts with multiple neonatal datasets, although the focus on specific experimental contexts limits generalizability as in the previous cases of HAPPE+ER and HAPPE.

*BEAPP*: BEAPP integrates several freely available preprocessing tools, such as EEGLAB, FieldTrip, and the PREP pipeline, into a highly flexible framework designed to support multisite and longitudinal EEG studies. BEAPP stands out for its ability to seamlessly handle data from multiple acquisition setups and its emphasis on automation, making it particularly suited for large-scale, multi-site studies requiring consistent and reproducible workflows. It provides extensive capabilities for artifact minimization, such as muscle artifacts, eye movement artifacts, and line noise, and re-referencing across various EEG systems, all of which are facilitated by an accessible graphical user interface (GUI). The GUI makes it particularly accessible for non-programmers, enabling a broader range of users to effectively manage preprocessing steps by offering intuitive parameter setting, visual feedback on processing progress, and easy-to-navigate options for customizing analysis workflows. BEAPP includes advanced artifact removal tools, such as automated line noise removal and ICA, which help to maintain data quality even in large and heterogeneous datasets. For all this, BEAPP was indexed at 4 for artifact handling due to its use of sophisticated tools like Cleanline, ICA, and MARA, which effectively handle common EEG artifacts. However, manual adjustments for filtering parameters, such as determining appropriate cutoff frequencies or adjusting notch filter



settings, may require user expertise to optimize data quality and avoid unintended distortions, which can be challenging without a solid understanding of signal processing. It goes without saying that automation and scalability were indexed with a 5, as BEAPP supports extensive batch processing – all within the user-friendly GUI –, while flexibility and adaptability also received an index of 5, owing to its modular framework, which allows seamless integration with multiple external tools, making it highly customizable for different EEG systems and datasets. Developmental sensitivity received an index of 4, as BEAPP incorporates features from pipelines like HAPPE, which provides effective solutions for handling infant EEG, such as targeted artifact rejection tailored to high-artifact infant data and enhanced preprocessing steps to accommodate shorter attention spans and high variability in data quality. However, it inherits certain limitations from these pipelines, too, such as the need for more extensive validation across different brain maturation groups and experimental contexts. Finally, validation against empirical data received an index of 4, reflecting successful validation efforts compared to other established pipelines and strong data retention rates. So far, BEAPP has been validated on datasets involving pediatric EEG and typical adult EEG in controlled lab environments, such as data from the Infant Sibling Project (ISP; for the pediatric EEG cohort), which included different sampling rates and acquisition setups like 64-channel and 128-channel Geodesic Sensor Nets. However, further validation on more diverse datasets, such as clinical populations, multi-site international datasets, and datasets with different EEG acquisition systems, remains pending and would definitely help towards its generalizability.

**MODULAR:** The Modular EEG preprocessing pipeline provides extensive customization options through the selection of different ICA methods, such as SOBI and Extended Infomax, and re-referencing strategies like CAR, REST, and RESIT. SOBI is known for its effectiveness in separating sources based on second-order statistics, while Extended Infomax is an advanced version of the Infomax algorithm that can handle both sub-Gaussian and super-Gaussian sources. CAR helps in reducing common noise by averaging signals from all electrodes, REST aims to reconstruct the signal as if referenced to a neutral point, and RESIT provides a refined approach to minimize residual artifacts through interpolation. Because of these collective characteristics, Modular received a high index of 5 for flexibility and adaptability, reflecting its extensive customization options. On the other hand, artifact handling was indexed with a 4, as there is some inconsistency in managing subtle or low-amplitude artifacts, such as eye blinks or muscle twitches, which often require manual refinement and adjustment to ensure data quality across different experimental contexts. Automation and scalability were also indexed with a 4 due to the inclusion of batch processing, although the need for manual selection of artifact components and re-referencing methods somewhat diminishes the ability towards full automation. For instance, manual intervention is often required to accurately identify and reject noisy components that automated methods may misclassify, as evidenced in the relevant literature (Coelli et al., 2024), which found that applying ICA led to a higher percentage of components classified as 'Other' due to misclassifications, requiring human correction to improve the quality of signal processing. Developmental sensitivity received a 3 as Modular does not incorporate age-specific adaptations for developmental EEG datasets, such as specialized filtering techniques or reference options tailored to the unique characteristics of differentiating infant EEG signals. Finally, validation against empirical data received an index of 4, based on quantitative measures such as Mutual Information Reduction (MIR) and empirical tests across adult datasets. MIR provides insights into the independence of signal components, but as suggested by the creators and authors, it may be beneficial to also consider other validation metrics like Dipolarity or Signal-to-Noise Ratio (SNR) for a more comprehensive assessment of signal quality and physiological plausibility. Additionally, broader validation in developmental populations is particularly important, as already mentioned a few times, due to the unique neurophysiological characteristics and variability in EEG signals at different stages of development.

**MADE:** MADE is specifically tailored for developmental EEG, utilizing ASR and ICA to address the high variability and artifact prevalence typical in infant EEG. It also incorporates feature-based component selection and the FASTER EEGLAB plugin for bad channel removal and interpolation,



which ensures that critical developmental signals are retained, thereby enhancing its applicability for understanding early brain maturation. Compared to other commonly used methods, such as traditional threshold-based artifact rejection, MADE's feature-based component selection offers a more targeted and efficient approach, minimizing data loss while retaining essential neural signals. MADE employs an automatic classification of independent components (ICs) using the adjusted-ADJUST method, enhancing artifact rejection by avoiding subjective biases and targeting features that are better suited for developmental data, and ensuring more effective artifact removal and better preservation of valuable data. However, the pipeline has several limitations, including its reliance on MATLAB, which restricts accessibility to users without appropriate software licenses and training. The computational demands of ICA, especially for high-density EEG, can also be a bottleneck, requiring significant processing time and hardware resources. Additionally, the fixed parameters used for artifact rejection may not be optimal for all datasets, reducing the flexibility needed to adapt to different recording conditions or developmental age groups, which could limit its broader applicability across varied research contexts. For these characteristics, MADE received specific indices for each criterion as follows: Artifact handling was indexed with a 4 due to its effective use of ICA and automated artifact correction techniques to address high artifact contamination. Automation and scalability were also assigned a 4, with MADE being fully automated, incorporating batch processing, and supporting data from various hardware systems, though computational demands and reliance on MATLAB limit accessibility. Flexibility and adaptability were indexed with a 3, as it is heavily focused on low-density EEG setups, while the reliance on the EEGLAB environment also means that users need familiarity with this specific software, which further limits its broader applicability without significant modification. Developmental sensitivity was assigned a 4, with MADE effectively addressing challenges unique to infant EEG, such as high artifact levels and shorter recording lengths, though further customization for different developmental stages could improve its utility. Finally, validation against empirical data was indexed at 4 based on successful validation with large developmental datasets, including EEG recordings from infants (12 months old), young children (3-6 years old), and late adolescents (16 years old). The infant data was collected using a 128-channel EGI system, the childhood data with a 64-channel BioSemi Active 2 system, and the late adolescent data with a 64-channel EGI system, demonstrating effective artifact removal and high data retention. The validation involved comparing the percentage of retained trials and the quality of processed EEG signals across different preprocessing pipelines, highlighting MADE's superior performance in retaining clean data while minimizing artifact influence. However, the specific index of 4 also reflects certain limitations in the validation process, including for example the challenges in maintaining consistency in artifact rejection across datasets with varying noise profiles. Additionally, the empirical validation lacked a comprehensive assessment of inter-rater reliability in identifying artifacts, which could affect consistency in real-world applications.

### 3.2. Comparative Overview

A summative presentation of the ten selected infant EEG preprocessing pipelines provides a comparative overview of their respective strengths and areas for improvement across all five criteria. The individual criteria indices assigned, are summarized in the Summative Indices Table (table 6) below, which is complemented by a Radar Chart (Figure 1) that visually represents the profiles of the different pipelines, enabling a quick comparative analysis, and a Heatmap (Figure 2) that highlights variations in capabilities levels across the criteria for each pipeline.

Table 6. Summative indices for all 10 EEG Pipelines.

Pipeline	Artifact Handling	Automation and Scalability	Flexibility and Adaptability	Developmental Sensitivity	Validation Against Empirical Data
APICE	5	5	5	4	4
Adjusted-ADJUST	4	4	5	4	3
GADS	4	4	3	4	3
NEAR	5	4	4	4	4
HAPPE	4	4	3	3	4
HAPPE+ER	4	4	3	4	4
HAPPILEE	4	4	4	3	4
BEAPP	4	5	5	4	4
MODULAR	4	4	5	3	4
MADE	4	4	3	4	4

Starting with the radar chart, APICE clearly stands out in terms of artifact handling, automation and scalability, and flexibility and adaptability, all indexed at the highest level of 5. This demonstrates that APICE is well-suited for handling high artifact loads while maintaining flexibility and ease of use. BEAPP was similarly indexed highly in these categories, specifically excelling in automation and scalability, making it particularly useful for large-scale studies involving multiple data acquisition systems. The radar chart also reveals that HAPPE and its extensions (HAPPE+ER and HAPPILEE) lag behind in some areas, particularly regarding flexibility and adaptability, indicating potential challenges in accommodating various experimental setups. The heatmap provides additional insights by visually indicating which criteria each pipeline excels at and where they may fall short. For instance, pipelines such as NEAR and APICE receive high numerical indices for developmental sensitivity, showcasing their ability to account for the rapid changes in infant brain development. However, it is also apparent that GADS and HAPPE struggle with flexibility and adaptability, showcasing relatively low indices, which suggests these pipelines may require more manual adjustments or might not be easily customized for different research contexts.

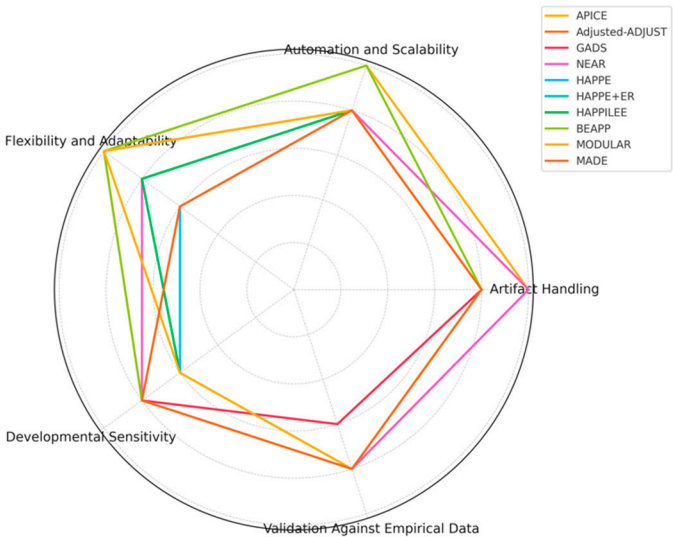
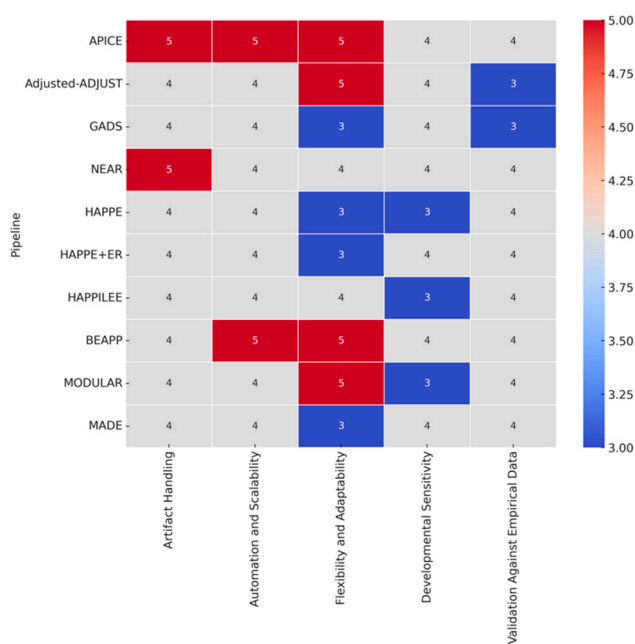


Figure 1. Radar chart of evaluation results for the 10 pipelines.



**Figure 2.** Heatmap showing variations in performance across the five criteria for all pipelines.

Analyzing the correlation matrix (Figure 3) reveals interesting relationships between the criteria. For instance, the moderate positive correlation ( $r = .38$ ) between artifact handling and automation and scalability suggests that pipelines that excel in effective artifact management are also more likely to automate preprocessing tasks efficiently. This relationship is particularly noticeable for APICE and BEAPP, which perform well across both criteria. A significant strong positive correlation ( $r = .56$ ) exists between automation and scalability and flexibility and adaptability, implying that the pipelines that are scalable are also generally more adaptable to different research conditions. This is well illustrated by the performance of the MODULAR pipeline, which allows users to select from different preprocessing methods, thereby providing flexibility while being able to scale effectively. However, the data also indicates some notable negative or negligible relationships. For instance, there is no significant correlation ( $r = .00$ ) between flexibility and adaptability and developmental sensitivity, suggesting that a pipeline’s ability to be flexible and customized does not necessarily imply that it will be developmentally sensitive. Similarly, the correlation between Developmental Sensitivity and Validation Against Empirical Data is negative ( $r = -.33$ ), indicating that the more developmentally sensitive pipelines might be less validated against diverse empirical datasets. This outcome may stem from the fact that specialized developmental considerations often target specific age ranges or population subsets, potentially reducing the generalizability of the validation efforts.

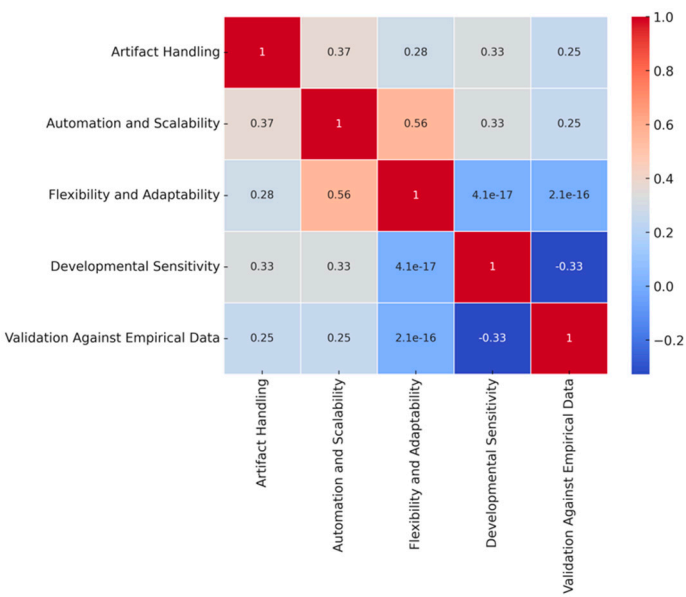


Figure 3. Correlation Matrix with relationships between criteria.

4. Discussion

The comparative overview presented above highlights the diverse strengths and limitations among the reviewed infant EEG preprocessing pipelines, demonstrating that – at this point – pipeline selection should be informed by the specific goals, contexts, and constraints of individual research or clinical applications. No single pipeline was found universally superior across all evaluated criteria; instead, each pipeline excels within particular dimensions, underlining the importance of aligning methodological choices with the research question, developmental stage, and logistical needs.

In light of this, the importance of robust preprocessing cannot be overstated. Clinically, pipelines, such as APICE and NEAR, that excel in handling infant-specific artifacts, may offer significant potential for application in NICUs. Effective artifact handling in these contexts allows clinicians to make more precise neurological assessments, facilitating in result early diagnosis and intervention for infants at risk for neurodevelopmental disorders. Furthermore, the modularity and flexibility offered by pipelines like MODULAR, BEAPP and adjusted-ADJUST may provide clinicians with adaptable frameworks that can be tailored to different clinical environments, diverse EEG hardware configurations, and varied patient populations, accelerating the translation of research-based EEG analysis techniques into practical clinical tools.

From a research perspective, the insights provided by this comparative analysis can significantly inform methodological decisions and enhance the quality of EEG-based developmental studies. In this framework, researchers can strategically choose pipelines based on their study-specific demands, such as the necessity for analyzing differential or specialized datasets provided by HAPPE, HAPPE+ER and HAPPILEE or the developmental specificity offered by pipelines like MADE and GADS. Informed pipeline selection enables researchers to optimize data quality, minimize biases introduced by inadequate preprocessing, and most importantly ensure methodological rigor. Therefore, recognizing the strengths and weaknesses of each pipeline promotes methodological transparency, aiding in the clear reporting of preprocessing decisions while also facilitating more accurate interpretations of developmental phenomena, thereby significantly advancing the field of developmental neuroscience.

5. Limitations

Despite these promising findings, several limitations should be carefully considered in the context of this comprehensive review. First, while developmental sensitivity is critically important in infant EEG analysis, this study revealed inconsistencies and gaps in how effectively pipelines

incorporate developmental considerations. Pipelines such as APICE, NEAR, and MADE for example demonstrate sensitivity to developmental changes; however, consistent age-specific optimizations across multiple stages, from preterm neonates to toddlers, remain somewhat limited across most pipelines. As infant EEG exhibits rapid and substantial developmental transformations, the lack of extensive developmental tailoring across age groups could inadvertently introduce systematic biases into the data interpretation, potentially obscuring important developmental trajectories.

Second, although validation against empirical data emerged as a key criterion in the pipeline analysis, this review identified variability in the depth and breadth of validation studies conducted. Significantly, all pipelines relied on validation within restricted experimental paradigms or specific age groups. As a result, the limited cross-validation across diverse populations and experimental conditions restricts the generalizability and robustness of pipeline recommendations. As such, more extensive and systematic validation efforts – particularly involving independent datasets collected from multiple research and clinical settings – are essential to bolster pipeline reliability and facilitate broader acceptance in both the research and clinical communities.

Finally, issues related to computational efficiency, proprietary software dependency (e.g., reliance on MATLAB), and user expertise requirements seem to be common limitations across pipelines, potentially hindering widespread adoption. For instance, computationally intensive techniques like ICA, DSS, or other advanced machine learning algorithms, while highly effective, may impose constraints on computational resources, affecting scalability, particularly in resource-limited clinical or field-based research contexts, posing practical challenges, particularly for researchers or clinicians operating in settings with limited technological resources or where rapid data analysis turnaround is required. Additionally, dependence on specialized software or platforms may restrict pipeline accessibility, necessitating specific technical skills or costly software licenses, thus limiting broader applicability and adoption in diverse research and clinical environments.

## 6. Conclusion

Moving forward, researchers and clinicians should consider integrating the findings of this critical review into their methodological decision-making by explicitly matching pipeline strengths to their specific needs. Although significant progress has been indeed made in advancing methodologies for artifact rejection, automation, and flexibility, several challenges remain since the field remains still fragmented as infant EEG is inherently complex, influenced by high levels of movement artifacts, low signal-to-noise ratios, and rapid developmental changes that necessitate tailored approaches for effective analysis.

Most importantly, this analysis underscores, in practical terms, the need for collaborative, community-driven efforts towards pipeline standardization and further validation. Encouraging transparency through open-source software, clearly documented codebases, and shared validation datasets would facilitate community-wide replication, independent validation, and ultimately foster methodological advancements. Such collaborative efforts would enhance the reproducibility and reliability of infant EEG research findings, supporting robust discoveries in developmental neuroscience and the clinical translation of these findings to enhance infant neurodevelopmental outcomes.

## References

- Benders, M. J., Palmu, K., Menache, C., Borradori-Tolsa, C., Lazeyras, F., Sizonenko, S., ... & Hüppi, P. S. (2015). Early brain activity relates to subsequent brain growth in premature infants. *Cerebral Cortex*, 25(9), 3014-3024.
- Coelli, S., Calcagno, A., Cassani, C. M., Temporiti, F., Reali, P., Gatti, R., ... & Bianchi, A. M. (2024). Selecting methods for a modular EEG pre-processing pipeline: An objective comparison. *Biomedical Signal Processing and Control*, 90, 105830. <https://doi.org/10.1016/j.bspc.2023.105830>
- 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, e13580. <https://doi.org/10.1111/psyp.13580>
- Falkner, T., & Abedi, A. (2010). Improving the signal to noise ratio of event-related EEG signals in high-risk newborns. *25th Biennial Symposium on Communications*.



- Fló, A., Gennari, G., Benjamin, L., & Dehaene-Lambertz, G. (2022). Automated pipeline for infants continuous EEG (APICE): A flexible pipeline for developmental cognitive studies. *Developmental Cognitive Neuroscience*, 54, 101077. <https://doi.org/10.1016/j.dcn.2022.101077>
- 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. *Frontiers in Neuroscience*, 12, 97. <https://doi.org/10.3389/fnins.2018.00097>
- Georgieva, S., Lester, S., Noreika, V., Yilmaz, M. N., Wass, S., & Leong, V. (2020). Toward the understanding of topographical and spectral signatures of infant movement artifacts in naturalistic EEG. *Frontiers in Neuroscience*, 14, 352. <https://doi.org/10.3389/fnins.2020.00352>
- Kumaravel, V. P., Farella, E., Parise, E., & Buiatti, M. (2022). NEAR: An artifact removal pipeline for human newborn EEG data. *Developmental Cognitive Neuroscience*, 54, 101068. <https://doi.org/10.1016/j.dcn.2022.101068>
- Leach, S. C., Morales, S., Bowers, M. E., Buzzell, G. A., Debnath, R., Beall, D., & Fox, N. A. (2020). Adjusting ADJUST: Optimizing the ADJUST algorithm for pediatric data using geodesic nets. *Psychophysiology*, 57(8), e13566. <https://doi.org/10.1111/psyp.13566>
- Levin, A. R., Méndez Leal, A. S., Gabard-Durnam, L. J., & O'Leary, H. M. (2018). BEAPP: The Batch Electroencephalography Automated Processing Platform. *Frontiers in Neuroscience*, 12, 513. <https://doi.org/10.3389/fnins.2018.00513>
- Lopez, K. L., Monachino, A. D., Morales, S., Leach, S. C., Bowers, M. E., & Gabard-Durnam, L. J. (2022). HAPPILEE: HAPPE in Low Electrode Electroencephalography, a standardized pre-processing software for lower density recordings. *NeuroImage*, 260, 119390. <https://doi.org/10.1016/j.neuroimage.2022.119390>
- Monachino, A. D., Lopez, K. L., Pierce, L. J., & Gabard-Durnam, L. J. (2022). The HAPPE plus Event-Related (HAPPE+ER) software: A standardized preprocessing pipeline for event-related potential analyses. *Developmental Cognitive Neuroscience*, 57, 101140. <https://doi.org/10.1016/j.dcn.2022.101140>
- Papatzikis, E., Dimitropoulos, K., Tataropoulou, K., Kyrtsoudi, M., Pasoudi, E., O'Toole, J. M., & Nika, A. (2024). The father's singing voice may impact premature infants' brain more than their mother's: A study protocol and preliminary data on a singing and EEG randomized controlled trial (RCT) based on the fundamental frequency of voice and kinship parameters. *medRxiv*. <https://doi.org/10.1101/2024.09.29.24314570>
- Rupprecht, L., Davis, J. C., Arnold, C., Gur, Y., & Bhagwat, D. (2020). Improving reproducibility of data science pipelines through transparent provenance capture. *Proceedings of the VLDB Endowment*, 13(12), 3354-3368.
- Stevenson, N. J., O'Toole, J. M., Korotchikova, I., & Boylan, G. B. (n.d.). Artefact detection in neonatal EEG. Irish Centre for Fetal and Neonatal Translational Research, University College Cork.
- Tataranno, M. L., Claessens, N. H., Moeskops, P., Toet, M. C., Kersbergen, K. J., Buonocore, G., ... & Benders, M. J. (2018). Changes in brain morphology and microstructure in relation to early brain activity in extremely preterm infants. *Pediatric research*, 83(4), 834-842.
- Thimbleby, H. (2024). Improving science that uses code. *The Computer Journal*, 67(4), 1381-1404.
- van der Reijden, C. S., Mens, L. H. M., & Snik, A. F. M. (2005). EEG derivations providing auditory steady-state responses with high signal-to-noise ratios in infants. *Ear & Hearing*, 26(3), 299-309.
- Victor, S., Appleton, R. E., Beirne, M., Marson, A. G., & Weindling, A. M. (2005). Spectral analysis of electroencephalography in premature newborn infants: Normal ranges. *Pediatric Research*, 57(3), 336-341. <https://doi.org/10.1203/01.PDR.0000153868.77623.43>
- Wielek, T., Del Giudice, R., Lang, A., Wisłowska, M., Ott, P., & Schabus, M. (2019). On the development of sleep states in the first weeks of life. *PLOS ONE*, 14(10), e0224521. <https://doi.org/10.1371/journal.pone.0224521>

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.