

# Preprocessing Pipelines for EEG

Sherly Maria,<sup>1, a)</sup> and Chandra J,<sup>2, b)</sup>

<sup>1</sup>*Research Scholar  
Department of Computer Science  
CHRIST (Deemed to be University)*

<sup>2</sup>*Associate Professor  
Department of Computer Science  
CHRIST (Deemed to be University)*

a)[sherly.maria@res.christuniversity.in](mailto:sherly.maria@res.christuniversity.in)

b)[chandra.j@christuniversity.in](mailto:chandra.j@christuniversity.in)

(Use the Microsoft Word template style: *Author Email*) or (Use Times New Roman Font: 10 pt, Italic, Centered)  
**Electroencephalogram (EEG) signals collected, present a lot of challenges in order to process the data. Usually, the signals collected contain a lot of artifacts and noises. To address this issue and to make the preprocessing method easier and automated. EEG is widely used to record brain signals and activity for clinical and research purposes. EEG signals are the best way to understand brain signals compared to other methods because of how accurate it is. However, it comes with certain setbacks like being highly sensitive to noise and susceptible to artifacts. Hence developing a pre-processing method ensures a smooth understanding of the signals. These pre-processing methods include filtering and noise removal techniques. Section 1 includes the pre-processing pipelines that have been popularly used by researchers during this study. Section 2 consists of the results and comparisons of various pipelines and our understanding of what is more effective.**

## Keywords

EEG, Stress, Pipelines, Signals, Noise, Artifacts

## INTRODUCTION

Electroencephalography (EEG) is a non-invasive method to capture brain activity with utility over all fields including medical and research. During the acquisition of data, the signal could be noisy and contain artifacts due to either electrical interference, incorrect placement of electrodes, or due to the participant's eye and muscle movements or clenching of the jaw. These artifacts could mess up the readings if left uncorrected and therefore it is necessary to get them removed for the purpose of data analysis. The methods to remove these artifacts could be filtering, noise removal, or correction of data. Standardized, automated processing pipelines provide a lot of advantages that would help in the uniform removal of artifacts and efficiency with larger data. Flexible pipelines also provide us with the advantage of swapping methods for the pipeline according to the requirement.

### Existing Pre-processing Pipelines

1. HAPPE(Harvard Automated Processing Pipeline for Electroencephalography)

- a. EEG recordings that have larger sample sizes and are of high density are usually very difficult to pre-process due to how unstable they could be. HAPPE focuses on the most common challenges that the researchers face like recordings being short or high levels of noise and artifacts. HAPPE[2] uses a combination of Q-ICA and ICA with MARA-based rejection components mainly for the conditions like high artifacts and smaller recording sizes. This processing pipeline could perform analysis for smaller datasets while preserving a great proportion of underlying signals. A few limitations like ICA decomposition requirements would not be met due to the data length. However, this could be approached by using PCA instead of ICA on channel-level data. Please note that using PCA could increase the number of artifacts, however, this is a minor setback. HAPPE is also freely available and is a stand-alone software.
2. MADE(Maryland Analysis of Development EEG)
    - a. MADE[1] was developed with the intention to standardize the EEG pre-processing method. This pipeline was mainly targeting paediatric data and was meant to be susceptible to the noises and artifacts of collecting data from paediatric populations. MADE is mainly event-related and provides a report at the end of the entire process describing the features of the processed data to help understand the quality of the data. It contains custom-written scripts and can be changed according to the requirements. MADE software is also freely available in GitHub.
3. PREP(Standardized Pre-processing for Large-Scale EEG Analysis)
    - a. Standardizing EEG processing usually involves referencing and high pass filtering. The main advantage of PREP [3] is to have an early-stage identification of interpolation of channels. This is used to keep a record of channels where the data could go bad. PREP is known to keep the complete record of the algorithm and interpolation to make sure that the EEG characteristics like noise detection. PREP also doesn't use a high pass filter downstream to maximize the usefulness of this data across applications. This is done by careful application of noise-removal and bad channel detection. Therefore, bringing in uniform statistical behaviour.
4. NEAR (An artifact removal pipeline for human new-born EEG data)
    - a. NEAR's procedure to remove artifacts is mainly focused on new-born and infant EEG data which is extremely sensitive. This makes it difficult to remove noise and is a much-needed pipeline for artifacts removal for kids. NEAR [4] is also experimenting on the frequency tagging paradigm using SSVEP and using ERP measures. The main approach of NEAR is to use time-frequency analysis or desynchronization in specific frequency ranges. NEAR is also experimentally known to perform better than MADE for paediatric data. NEAR software is also publicly available at GitHub.
5. GROUP-LEVEL EEG-PROCESSING PIPELINE FOR FLEXIBLE SINGLE TRIAL-BASED ANALYSES INCLUDING LINEAR MIXED MODELS
    - a. This method uses a single trial-based analysis of EEG and behavioural data. This method is mainly based on LMM analysis of ERP data. This pipeline is also open to experimenting and can

be modified according to one's requirements. Parts of the pipeline can be broken down and swapped for other approaches.

6. ADJUST (An Automatic EEG artifact detector based on the joint use of spatial and temporal features)
  - a. ADJUST [6] is an automatic algorithm that identifies artifacts' independent components by combining stereotyped artifacts. Mainly temporal and spatial features that are specific. These features were mainly optimized to capture blinks and discontinuities on feature selection. After a lot of experiments, it showed that ADJUST's classification of ICA matches a manual one and even the removal of noisy features by ADJUST was much better. This mainly also proved that ADJUST was a fast and efficient and automatic way to use ICA for artifact removal.
  
7. FASTER (Fully Automated Statistical Thresholding for EEG artifact Rejection)
  - a. FASTER [7] is an artifact rejection method for various aspects of the data using EEG time series and ICA. FASTER has been compared to the SCADS method and is known to have a 60% sensitivity for the detection of noisy channels and eye movements. FASTER is also known to aggregate the ERP across subject datasets and helps in detecting outlier datasets.

## Proposed Pipeline

The main aim of this pipeline is to correct the artifacts that involve a few channels and to let go of contaminated and noisy channels. The bad channels would be converted into epochs. Post this, filtering methods would be applied and Wavelet analysis and ICA would be performed to refine the data. All these steps would be flexible but as noted multiple times in previously mentioned pipelines, ICA is one of the most preferred methods.

The first step would be to choose the dataset. We decided to go with the DEAP dataset available as an open-source dataset for emotion detection. The second step would be artifact removal. This could begin with removing the bad electrodes and finding the ones with the lesser amplitude. The difference in amplitudes could say a lot about the intensity of noise. A larger value of the amplitude would implicate more noisy data and thus be eliminated. Among the most recent trends of artifact removal, using rejection cycles would make the job easier. A rejection matrix is then used to understand bad samples and channels.

Post this filtering would need to be performed, using a low-pass filter with the frequency ranging between 50-40 Hertz. Low pass filtering is one of the most effective ways to remove line noise.

However, extensive filtering (above 0.1 Hz) can introduce critical distortions in the data (8). While alternatives to high-pass filtering exist, such as local detrending methods (9), we have not observed better performance of these methods. It is preferred to apply high pass filters when the data is continuous, while low pass could be applied at any point.

## Independent Component Analysis

ICA is mostly used with EEG data in order to remove physiological noises like eye blinks and jaw clenching. ICA is known to be efficient and one of the best methods because it removes very specific artifacts from the data without tampering with EEG data.

ICA is very commonly used with adult data and not very widespread in the use with paediatric data like NEAR and MADE. It's quite easy to collect the data in Adults as they are attentive and are susceptible to lesser noise as compared to kids[10]. These factors would contribute to a low success rate of separation of neural and non-neural sources. A good ICA decomposition requires several considerations. In order to obtain a reliable separation in ICs, the data must be high-pass filtered at least at 1 Hz and should not contain high amplitude noise (e.g., motion artifacts) [11]. However, high-pass data filtering may not be suitable for many EEG analyses (e.g., ERPs are distorted, and slow waves may be lost)[12]. Hence it is recommended to have a combination of the high pass and low pass filter depending on the data flow.

## Conclusion

Despite being a widespread technology, the pre-processing pipeline has no standardized procedure because of the lack of researchers in this field[13]. We have combined the downsides of various pipelines like MADE, PREP, ADJUST, NEAR and thoroughly understood the drawbacks of each method making this more optimized for Adult data. Apart from bringing flexibility and adaptive algorithms, this pipeline can also be used for detecting different segmentation methods[14 and 15]. The outcome of this research is to help propose a suitable method to be able to automate the stress detection technique and to analyse the best suitable method according to the participant's requirement.

## REFERENCES

1. M.K. Islam, A. Rastegarnia, Z. Yang "Methods for artifact detection and removal from scalp EEG: a review", *Neurophysiol. Clin./Clin. Neurophysiol.*, 46 (4) (2016), pp. 287-305, [10.1016/j.neucli.2016.07.002](https://doi.org/10.1016/j.neucli.2016.07.002)
2. L.J. Gabard-Durnam, A.S. Mendez Leal, C.L. Wilkinson, A.R. Levin  
The Harvard Automated Processing Pipeline for Electroencephalography (HAPPE): standardized processing software for developmental and high-artifact data. *Front. Neurosci.*, 12 (2018), p. 12, [10.3389/fnins.2018.00097](https://doi.org/10.3389/fnins.2018.00097)
3. R. Debnath, G.A. Buzzell, S. Morales, M.E. Bowers, S.C. Leach, N.A. Fox  
The Maryland analysis of developmental EEG (MADE) pipeline  
*Psychophysiology*, 57 (6) (2020), Article e13580,
4. Velu Prabhakar, Kumaravelab, Elisabetta Farellaa, Eugenio Pariseb, Marco Buiattib, "NEAR: An artifact removal pipeline for human newborn EEG data", <https://doi.org/10.1016/j.dcn.2022.101068>
5. [Bigdely-Shamlo et al., 2015](#)N. Bigdely-Shamlo, T. Mullen, C. Kothe, K.-M. Su, K.A. Robbins, "The PREP pipeline: standardized preprocessing for large-scale EEG analysis"  
*Front. Neuroinform.*, 9 (June) (2015), pp. 1-20, [10.3389/fninf.2015.00016](https://doi.org/10.3389/fninf.2015.00016)
6. A. Mognon, J. Jovicich, L. Bruzzone, M. Buiatti "ADJUST: an automatic EEG artifact detector based on the joint use of spatial and temporal features"  
*Psychophysiology*, 48 (2) (2011), pp. 229-240, [10.1111/j.1469-8986.2010.01061.x](https://doi.org/10.1111/j.1469-8986.2010.01061.x)
7. H. Nolan, R. Whelan, R.B. Reilly, "FASTER: fully automated statistical thresholding for EEG artifact rejection"*J. Neurosci. Methods*, 192 (1) (2010), pp. 152-162, [10.1016/j.jneumeth.2010.07.015](https://doi.org/10.1016/j.jneumeth.2010.07.015)
8. A. De Cheveigné, L.C. Parra"Joint decorrelation, a versatile tool for multichannel data analysis"*NeuroImage*, 98 (2014), pp. 487-505, [10.1016/j.neuroimage.2014.05.068](https://doi.org/10.1016/j.neuroimage.2014.05.068)
9. A. de Cheveigné, D. Arzounian, "Robust detrending, rereferencing, outlier detection, and inpainting for multichannel data", *NeuroImage*, 172 (2018), pp. 903-912, [10.1016/j.neuroimage.2018.01.035](https://doi.org/10.1016/j.neuroimage.2018.01.035).

10. M. Eisermann, A. Kaminska, M.-L. Moutard, C. Soufflet, P. Plouin, “Normal EEG in childhood: from neonates to adolescents”, *Neurophysiol. Clin./Clin. Neurophysiol.*, 43 (1) (2013), pp. 35-65, [10.1016/j.neucli.2012.09.091](https://doi.org/10.1016/j.neucli.2012.09.091)
11. Winkler, I., Debener, S., Müller, K., Tangermann, M., 2015. On the influence of high-pass filtering on ICA-based artifact reduction in EEG-ERP. In: Proceedings of the 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 4101–5.
12. M. Eisermann, A. Kaminska, M.-L. Moutard, C. Soufflet, P. Plouin” Normal EEG in childhood: from neonates to adolescents”, *Neurophysiol. Clin./Clin. Neurophysiol.*, 43 (1) (2013), pp. 35-65, [10.1016/j.neucli.2012.09.091](https://doi.org/10.1016/j.neucli.2012.09.091)
13. E. Kushnerenko, R. Ceponienė, P. Balan, V. Fellman, R. Näätänen, “Maturation of the auditory change detection response in infants: a longitudinal ERP study” *Neuroreport*, 13 (15) (2002), pp. 3-8
14. P.J. Marshall, Y. Bar-Haim, N.A. Fox, “Development of the EEG from 5 months to 4 years of age”, *Clin. Neurophysiol.*, 113 (8) (2002), pp. 1199-1208
15. C.A. Nelson, C.S. Monk, “The use of event-related potentials in the study of cognitive development”, *Handbook of Developmental Cognitive Neuroscience*, MIT Press