
Electronic Thesis and Dissertation Repository

7-19-2024 11:00 AM

Data Preprocessing and Machine Learning for Intracranial Electroencephalography

Mauricio Cespedes Tenorio, *Western University*

Supervisor: Khan, Ali R., *The University of Western Ontario*

Co-Supervisor: Lau, Jonathan C., *The University of Western Ontario*

A thesis submitted in partial fulfillment of the requirements for the Master of Engineering Science degree in Biomedical Engineering

© Mauricio Cespedes Tenorio 2024

Follow this and additional works at: <https://ir.lib.uwo.ca/etd>



Part of the [Bioelectrical and Neuroengineering Commons](#), [Biomedical Commons](#), [Computational Neuroscience Commons](#), and the [Signal Processing Commons](#)

Recommended Citation

Cespedes Tenorio, Mauricio, "Data Preprocessing and Machine Learning for Intracranial Electroencephalography" (2024). *Electronic Thesis and Dissertation Repository*. 10223. <https://ir.lib.uwo.ca/etd/10223>

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact wlsadmin@uwo.ca.

Abstract

This thesis serves to address the problem of non-standardized preprocessing of intracranial electroencephalography (iEEG) recordings by implementing a software workflow that compiles some of the most common steps followed for the preparation of this type of data. This workflow improves the consistency, replicability, and ease of use of sEEG preprocessing, facilitating the replication and extension of previous studies and the combination of separately preprocessed inter-institutional datasets. Automatic detection of artifacts for iEEG data was also explored as a potential step to include in the preprocessing workflow. Despite training the models with cross-institutional data, poor performance was observed when tested on external datasets, showing the need for more and higher quality cross-institutional datasets to develop truly generalizable models. Future work is needed to include other common preprocessing methods, validate the developed tool with external datasets, and ensure compliance with BIDS standards to establish a standardized preprocessing tool for iEEG recordings.

Keywords: Intracranial Electroencephalography (iEEG), Stereoelectroencephalography (sEEG), BIDS, Data Preprocessing, Artifact Detection, Machine Learning, Epilepsy, Signal Processing.

Summary for Lay Audience

In this thesis, I seek to address a common problem in neuroscience research related to a brain monitoring technology called intracranial electroencephalography (iEEG). This technology, where electrodes are surgically implanted inside of the skull for recordings, is essential for understanding, treating and completing research on epilepsy. Currently, there is no standardized preparation of iEEG data for epilepsy research, which hinders the comparison between studies developed in different institutions and extension of previous findings.

To tackle this issue, I developed a new software workflow that compiles common preprocessing steps for iEEG data. This tool aims for a more consistent preparation of the data between different research settings, making it easier for researchers to replicate and extend findings and to combine data from different institutions. This initial version of the workflow only supports preprocessing of stereoelectroencephalography (SEEG) data, a type of iEEG, chosen due to its common usage in clinical settings compared to other iEEG techniques.

The presented thesis presents a validation of the developed workflow using local datasets. More testing with external datasets is needed to confirm broader usability and generalizability. Additionally, the workflow currently supports only SEEG recordings and limited file formats and preprocessing methods, which require expansion to increase their utility.

I also explored automatic artifact detection for iEEG, which is another fundamental step in iEEG preprocessing with no clear standardized or validated methodology. Although the models tested in this work show promise, they did not perform well enough to be included in the final preprocessing workflow developed. This highlights the need for larger and higher-quality datasets from multiple institutions to develop reliable models that can be standardized in the preprocessing workflow.

Despite these advances, several primary limitations remain to ensure the workflow is universally applicable in epilepsy and neuroscience research. These include the need to perform more extensive testing, increase the number of available methods, improve computational efficiency, and adhere to the BIDS standards. This project not only provides a practical tool for

researchers, but also summarizes the challenges and complexities involved in managing and analyzing iEEG data, highlighting the need for further innovation in this field.

Acknowledgements

First of all, I would like to thank the two people who have made this thesis and this experience possible: Dr. Jonathan Lau and Dr. Ali Khan. Jon, I thank you for all the advice and support you have given me during my master's degree. I appreciate that you have always had an open door for any concerns and looked out for our welfare in the lab. Ali, your expertise and your open door policy have helped me succeed in my master's degree and I am deeply grateful for that. This work would not have been possible without the support of both of you.

To my advisory committee: Dr. Ana Suller-Marti and Dr. Lyle Muller, thank you for your valuable insights and guidance during my master's degree.

To all the members of the lab, who welcomed me here in Canada and always offered me their support and friendship, I will always be grateful for that. I would like to give special recognition to the very good friends I have made during my master's degree: Arun, Mohamed, Alaa, Violet and Keza; I hope to have a lasting friendship with all of you.

Finally, and most importantly, to my girlfriend Elena and my family. Ele, thank you for always being by my side, both in the best and worst of times. To my family, for always supporting me, even from a distance.

List of Acronyms and Abbreviations

AD Artifact detection

AL Anatomical localization of contact electrodes

ANN Artificial Neural Network

ASM Anti-seizure Medications

ASR Artifact Subspace Reconstruction

BIDS Brain Imaging Data Standard

BIP Bipolar

CAR Common Average Reference

CNN Convolutional Neural Network

CV Cross-Validation

db4 Daubechies wavelet of order four

DM Demeaning

DRE Drug-Resistant Epilepsy

DS Downsampling

DWT Discrete Wavelet Transform

ECG Electrocardiography

ECoG Electrocorticography

EDF European Data Format

EEG Electroencephalography

EOG Electrooculography

EZ Epileptogenic Zone

FFT Fast Fourier Transform

FIR Finite Impulse Response

FNUSA St. Anne's University Hospital

GAN Generative Adversarial Network

GRU Gated Recurrent Unit

GUI Graphical user interface

HFOs High Frequency Oscillations

HPF High-Pass Filter

IEDs Interictal Epileptiform Discharges

iEEG Intracranial Electroencephalography

IIR Infinite Impulse Response

ILAE International League Against Epilepsy

JSON JavaScript Object Notation

LDA Linear Discriminant Analysis

LFP Local Field Potential

LR Logistic Regression

LSTM Long Short-Term Memory

LVFA Low Voltage Fast Activity

MEG Magnetoencephalography

ML Machine Learning

MRI Magnetic Resonance Imaging

PLI Power Line Interference

PSD Power Spectrum Density

QC Quality Control

RF Random Forest

RMSE Root Mean Square Error

RNN Recurrent Neural Network

ROI Region of Interest

SEEG Stereoelectroencephalography

SNR Signal-to-Noise Ratio

SOZ Seizure Onset Zone

SVM Support Vector Machine

TLE Temporal Lobe Epilepsy

XGBoost Extreme Gradient Boosting

Contents

Abstract	ii
Summary for Lay Audience	iii
Acknowledgements	v
List of Acronyms and Abbreviations	vi
List of Figures	xv
List of Tables	xxvi
List of Appendices	xxvii
List of Symbols	xxviii
1 Introduction	1
1.1 Overview	1
1.2 Epilepsy definition	2
1.2.1 Definition and overview	2
1.3 Diagnosis and treatment of epilepsy	3
1.3.1 Epilepsy surgery	4
1.4 Intracranial Electroencephalography	6
1.4.1 Basics of iEEG	6
1.4.2 iEEG for identification of SOZ	10

1.4.3	Modern analysis techniques	12
1.5	Brain Imaging Data Structure (BIDS) for iEEG	15
1.6	Preprocessing of iEEG data	17
1.6.1	Overview of preprocessing steps for iEEG	18
	Re-referencing	19
	Drift correction	20
	Power line interference (PLI) filtering	21
	Artifact detection	22
	Downsampling	23
	Channel localization	23
1.6.2	Common preprocessing steps: a brief literature review	25
1.6.3	Existing Solutions and Limitations	26
1.7	Machine learning for iEEG analysis	37
1.7.1	Machine learning overview	37
1.7.2	Machine learning applications for iEEG	40
1.7.3	Automatic artifact detection for preprocessing of iEEG recordings	41
1.8	Thesis Outline	45
1.8.1	Problem statement	45
1.8.2	Objectives	46
2	Methodology	48
2.1	Development and design of a preprocessing workflow	48
2.1.1	Input	50
2.1.2	Epoch extraction	51
2.1.3	Downsampling	51
2.1.4	Drift correction	52
2.1.5	Re-referencing	53
2.1.6	Power Line Interference (PLI) attenuation	53

2.1.7	Channel localization	54
2.1.8	BIDS Validation	58
2.1.9	Outputs	59
2.1.10	Execution of the workflow	60
2.2	Development of an Automatic iEEG Artifact Detector	62
2.2.1	Datasets used and data preprocessing	62
2.2.2	Feature selection	66
2.2.3	Selection of hyperparameters	69
2.2.4	Training and evaluation	71
	Selection of the best hyperparameters	72
	Comparison of similar models	72
	Deep learning vs traditional machine learning evaluation	73
	Clinical insights	73
2.2.5	Metrics used for evaluation	74
3	Results	76
3.1	Validation of the Preprocessing Workflow	76
3.1.1	Validation of each preprocessing step	77
	Epoch extraction	77
	Downsampling	78
	Drift correction	79
	Re-referencing	83
	PLI attenuation	85
	Channel localization	93
3.1.2	Comparison against existing solutions	95
3.2	Validation of the Automatic iEEG Artifact Detector	98
3.2.1	Datasets information	98
3.2.2	Performance evaluation	98

Selection of the best hyperparameters	98
Comparison of similar models	102
Deep learning vs traditional machine learning evaluation	103
3.2.3 Clinical insights	108
4 Conclusions	112
4.1 Thesis Contributions	112
4.2 A more standardized preprocessing of iEEG recordings	113
4.3 Automatic Artifact Detection for iEEG	115
4.4 Remaining challenges, future work & limitations of this study	118
Bibliography	122
A iEEGPrep Report	150
B iEEGPrep snakebids.yaml	156
C iEEGPrep validation	160
C.1 Epoch extraction	160
C.2 Downsampling	161
C.3 Drift correction	162
C.4 Re-referencing	163
C.5 PLI attenuation	164
D Fieldtrip code	167
E Models hyperparameters tested using GridSearchCV	170
F Deep learning models evaluated using cross-validation	171
G Additional results for models comparison	178
G.1 Comparison of similar models	179

G.2 Deep learning vs traditional machine learning evaluation	180
H Copyright Transfers and Reprint Permissions	182
Curriculum Vitae	206

List of Figures

1.1	Graphical representation of the framework to classify epilepsy. Reproduced from [133] with permissions.	3
1.2	a. Examples of conceptual zones that can be used to estimate the EZ. b. Graphical representation of the connections between the epileptogenic network (green) and the healthy tissue network (blue). c. Examples of surgical strategies for epilepsy surgery, including resection of zones (red rectangles) or disconnection of networks (purple dotted line). d. Graphical representation of the conceptual zones that can be estimated using different modalities. Reproduced from [163] with permissions.	5
1.3	Example of physiological high-frequency oscillations (HFOs) recorded using SEEG electrodes for recordings. Data was recorded during wakefulness from an adult patient with left temporal DRE. Top panel shows two bipolar signals after filtering between 0.5 and 100 Hz. Bottom panel shows a zoomed version of the same recording after high-pass filtering of 150 Hz. Reproduced from [49] with permissions.	9
1.4	Examples of interictal epileptiform discharges (IEDs) in SEEG recordings. Figure shows that the recordings from the contacts B1 to B5 have presence of high-amplitude spikes and spike-slow waves. Reproduced from [49] with permissions.	11

1.5	<p>Example of spectral profiles developed by Frauscher <i>et al.</i>. Each panel shows a different statistical distribution of normalized power between 0.5 Hz and 80 Hz, with the red solid line representing the median of the distribution on each frequency point, the red overlay the region between 25 and 75 percentiles, and the dotted red lines showing the maximum and minimum. The color bar shows the percentage of channels that showed a peak in PSD in the specific frequency band in which it is located. Reproduced from [51] with permissions.</p>	14
1.6	<p>Example of a iEEG-BIDS directory for a session with electrical stimulation along with example pictures. First layer shows the different subjects' subfolders and the stimuli directory. (b) Shows the files associated with the recording data, in this case in BrainVision Core format. (c) The information related to the channels in the recording are stored in the <code>_channels.tsv</code> file. (d) The events of importance during the iEEG recording are stored in the <code>_events.tsv</code> file, in this case, showing stimulation events. (e) The <code>_electrodes.tsv</code> file is intended to store information about the electrodes, including their spatial location. (f) The <code>_coordsystem.json</code> file should show the information of the coordinate system used to define the electrodes location. (g) Shows other images that can be stored as part of the directory. Reproduced from [70] with permissions under Creative Commons License.</p>	16
1.7	<p>Electrode contacts localization workflow based on individual or normalized space. Reproduced from [102] with permissions under Creative Commons License.</p>	24

1.8	Diagram showing the different ways to identify the region by iEEG channel. The left side shows the case where a segmentation file is available, in which case it is required to give the iEEG coordinates in the same space as the segmentation. The right side shows the possible paths to follow when a segmentation has not been previously calculated, including the alignment to a common space or the calculation of the segmentation when an MRI file is available. If neither an MRI image nor a plot is available, the only option left is to give the iEEG coordinates in a common space to use a predefined plot in that space. . . .	25
1.9	Visual representation of the classification of articles reviewed in the work done by Mirchi <i>et al.</i> . Reproduced from [104] with permissions under Creative Commons License.	41
2.1	Overview of the preprocessing workflow implemented. A BIDS-iEEG directory is received as the input, which is processed by running different preprocessing steps such as downsampling and re-referencing. Quality-control (QC) steps are also executed, including a validation of the BIDS directory and the generation of a report at the end. This report, along with the clean preprocessed files are the main outputs of the workflow.	50
2.2	Graphical representation of a bipolar (left) and monopolar (right) potential fields in black and the three-dimensional figure used to represent it in the SEEG workflow (orange). Adapted from [146].	57
2.3	Examples for each of the classes present in the FNUSA and Mayo datasets. These correspond to (a) power line noise, (b) muscle artifact, (c) baseline jump artifact, (d) physiological or baseline signal, and (e) pathological signal showing an HFO riding on a spike. Recovered from [112] with permissions under Creative Commons License.	63

2.4	Hilbert transform computed for a bandpass version of the original signal on each of the frequency bands of interest. Envelope computed through the Hilbert transform represents one row in the resulting time-frequency map.	67
3.1	Time plots for the first 60 seconds extracted by iEEGPrep and Fieldtrip from the channel LPIIn02.	78
3.2	Distribution of correlation values with zero lag in time domain (left) and frequency domain (right) between the different pairs of channels present in the extracted epochs using iEEGPrep and Fieldtrip. The amplitude scales for both plots show values very close to $9.99e-1$ (or 0.99), highlighting the strong correlations observed across both time and spectral domains.	79
3.3	Time plots for the first 60 seconds of the downsampled signal from the LPIIn02 channel using iEEGPrep and Fieldtrip.	79
3.4	Distribution of correlation values with zero lag in time domain (left) and frequency domain (right) between the different pairs of channels present in the downsampled epochs using iEEGPrep and Fieldtrip. The amplitude scales for both plots show values very close to $9.99e-1$ (or 0.99), highlighting the strong correlations observed across both time and spectral domains.	80
3.5	Time plots for the first 60 seconds of the detrended signal from the LPIIn02 channel using iEEGPrep and Fieldtrip. It is possible to see that the slow trends previously visible in Figure 3.3 were attenuated after drift correction.	80
3.6	Distribution of correlation values with zero lag in time domain (left) and frequency domain (right) between the different pairs of channels present in the high-pass filtered epochs using iEEGPrep and Fieldtrip. Values obtained show a correlation above 98% for both cases.	81

3.7	Statistical comparison of the first four frequency components between the different pairs of channels present in the high-pass filtered epochs using iEEGPrep and Fieldtrip. Components were calculated using Welch’s method. No statistical significant differences were observed.	82
3.8	Time plots for the first 60 seconds of the detrended signal from the LPIIn02 channel using iEEGPrep and Fieldtrip.	83
3.9	Distribution of amplitude values of the first four frequency components for the channels before (Original, in red) and after drift correction. High-pass filtering was the only method that showed a visual difference for the first two frequency components. Components were calculated using Welch’s method.	83
3.10	Distribution of correlation values with zero lag in time domain (left) and frequency domain (right) between the different pairs of channels present in the re-referenced epochs by iEEGPrep and Fieldtrip. Values obtained show a correlation above 98% for both cases.	84
3.11	Frequency spectrum plots for the first 60 seconds from the channel LPIIn2-3 using iEEGPrep and Fieldtrip. Frequency domain signal was calculated using Welch’s method.	84
3.12	Frequency spectrum plots for the first 60 seconds from the channel LPIIn2-3 using iEEGPrep and Fieldtrip after PLI attenuation using Cleanline algorithm. Frequency domain signal was calculated using Welch’s method.	85
3.13	An example of a signal from the 100 SEEG channels simulated for validation of PLI attenuation methods for the two considered cases. Left column shows the case A, where peaks at the fundamental frequency and its harmonics were induced into the data. Right column shows case B, where more broadband line noise was injected into the simulated data.	86

3.14 Comparison between the Cleanline implementation in MATLAB and Python using signal-to-noise ratio (SNR) and the attenuation in decibels. No statistical significant differences were found between the means of the distributions for both metrics, with p-values above 0.4. A power of 87.1% was achieved with a hypothetical effect size of 0.14 dB for the SNR comparison, while a power of 81.7% was achieved with a hypothetical effect size of 1.1 dB for the attenuation comparison. 87

3.15 Comparison between the Zapline implementation in MATLAB and Python using signal-to-noise ratio (SNR) and the attenuation in decibels. No statistical significant differences were found between the means of the distributions for both metrics, with p-values above 0.4. A power of 85.1% was achieved with a hypothetical effect size of 0.14 dB for the SNR comparison, while a power of 80.9% was achieved with a hypothetical effect size of 1.1 dB for the attenuation comparison. 88

3.16 Comparison between the removePLI implementation in MATLAB and Python using signal-to-noise ratio (SNR) and the attenuation in decibels. No statistical significant differences were found between the means of the distributions for both metrics, with p-values above 0.3. A power of 85.1% was achieved with a hypothetical effect size of 0.14 dB for the SNR comparison, while a power of 81.1% was achieved with a hypothetical effect size of 1.1 dB for the attenuation comparison. 89

3.17 Distribution of SNR on the simulated SEEG channels in case A for the different advanced PLI attenuation methods implemented in iEEGPrep. Zapline showed a significantly higher attenuation compared to the other methods, as indicated by its more negative dB values. No statistically significant difference was found between the Zapline and removePLI methods. 90

3.18	Distribution of attenuation values (dB) on the simulated SEEG channels in case A for the different advanced PLI attenuation methods implemented in iEEG-Prep. Zapline showed a significantly higher SNR compared to the other methods. No statistical significance was found between Cleanline and removePLI methods.	90
3.19	Example of power spectral density curves obtained through Welch’s method for one simulated signal filtered for case A through the different advanced line noise attenuation algorithms implemented in iEEGPrep. Orange and blue signals show simulated signal before and after PLI attenuation respectively.	91
3.20	Power spectral density plot for the channel LPIIn2-3 in the preprocessed EDF file using the different advanced PLI attenuation methods.	92
3.21	Time-frequency map from 0 Hz to 70 Hz of the first 60 seconds of the re-referenced signal with PLI attenuation from the channel LPIIn2-3 using the implemented advanced PLI attenuation methods. Colorbar scale is the same for all plots.	93
3.22	Coronal 2D view of the location of the monopolar channels part of the bipolar channel LAIn9-10 and the virtual midpoint between them. Left frame shows the position of the channel LAIn09 in the gray region (corresponding to ROI <i>ctx-lh-insula</i>). The picture in the middle shows the position of the channel LAIn10 near the intersection between the ROI <i>ctx-lh-insula</i> (gray) and <i>ctx-lh-parstriangularis</i> (red). The right plane shows the virtual midpoint between these two monopolar channels lying in the CSF (blue). The yellow region corresponds to the ROI <i>Left-Cerebral-White-Matter</i>	94
3.23	Examples of instances per class in the FNUSA and Mayo datasets used for training and validation.	99

3.24 Comparison between the tested CNN architectures. Metrics used were accuracy, macro precision, macro recall and macro F1-score. CNN-S, CNN-M and CNN-L refer to the implemented small, medium and large architectures respectively. Box plot shows the interval between the 25% and 75% percentiles, the black line inside the box shows the median and the white dot the mean of each distribution. 101

3.25 Comparison between the tested CNN+LSTM architectures. Metrics used were accuracy, macro precision, macro recall and macro F1-score. CNN+LSTM-S, CNN+LSTM-M and CNN+LSTM-L refer to the implemented small, medium and large architectures respectively. Box plot shows the interval between the 25% and 75% percentiles, the black line inside the box shows the median and the white dot the mean of each distribution. 101

3.26 Comparison between the different traditional machine learning models after hyperparameter tuning. Metrics used were accuracy, macro precision, macro recall and macro F1-score. XGBoost, RF and LR correspond to the XGBoost, random forest and logistic regression classifiers respectively. Box plot shows the interval between the 25% and 75% percentiles while the black line inside the box shows the median and the white dot the mean of each distribution. . . . 102

3.27 Comparison between the different deep learning models after architecture selection. Metrics used were accuracy, macro precision, macro recall and macro F1-score. CNN-Small and CNN+LSTM-Small correspond to the smallest implemented CNN and CNN+LSTM models respectively. Box plot shows the interval between the 25% and 75% percentiles while the black line inside the box shows the median and the white dot the mean of each distribution. 103

3.28 Confusion matrix for the random forest on the external dataset from the University of Florida Neuromedicine Hospital. 105

3.29	Example of one segment of data at the time 0:00:57 (left of the image) classified as noise (class 0) that was not verified as such by the expert reviewer. Middle segment at 0:01:00 shows a clip classified as baseline (class 2) by the model. Annotated channel with the model was LMHc1-2.	109
3.30	Example of two segments of data originally classified as baseline (left) and noise (right) for channel LAM1-2 that were classified as pathological due to LVFA by the expert review.	109
3.31	Signals shown in Figure 3.30 but with an applied high-pass filtered with cut-off frequency at 80 Hz. It shows two segments of data originally classified as baseline (left) and noise (right) for channel LAM1-2 that were classified as pathological due to LVFA by the expert review.	110
C.1	Frequency spectrum plots for the first 60 seconds extracted from the channel LPIn02 by iEEGPrep and Fieldtrip. Frequency domain signal was calculated using Welch's method.	160
C.2	Frequency spectrum plots for the first 60 seconds of the downsampled signal from the channel LPIn02 using iEEGPrep and Fieldtrip. Frequency domain signal was calculated using Welch's method with a window of 3 s and an overlap of 50%.	161
C.3	Time-frequency map of the first 60 seconds of downsampled data from the channel LPIn02 using iEEGPrep and Fieldtrip. Maps were calculated using short-time Fourier Transform (STFT) with a window of 3 s and an overlap of 50%.	161
C.4	Time-frequency map of the first 60 seconds of detrended data from the channel LPIn02 using iEEGPrep and Fieldtrip. Maps were calculated using short-time Fourier Transform (STFT) with a window of 3 s and an overlap of 50%.	162

C.5	Time-frequency map of the first 60 seconds of detrended data from the channel LPIIn02 using the different methods implemented in iEEGPrep. Maps were calculated using short-time Fourier Transform (STFT) with a window of 3 s and an overlap of 50%. Colorbar scale of the three maps is the same. Only high-pass filtering shows a significant reduction of slow frequency oscillations.	162
C.6	Time-frequency map from 0 Hz to 20 Hz of the first 60 seconds of the re-referenced signal from the channels LPIIn02 and LPIIn03 using iEEGPrep and Fieldtrip.	163
C.7	Time-frequency map from 0 Hz to 70 Hz of the first 60 seconds of the re-referenced signal from the channels LPIIn02 and LPIIn03 using iEEGPrep and Fieldtrip. A predominance of line noise can be observed.	163
C.8	Distribution of correlation values with zero lag in time domain (left) and frequency domain (right) between the different pairs of channels present in the re-referenced epochs with PLI attenuation for iEEGPrep and Fieldtrip.	164
C.9	Distribution of time domain RMSE on the hundred simulated sEEG channels in case A for the different advanced PLI attenuation methods implemented in iEEGPrep.	164
C.10	Distribution of frequency domain RMSE on the hundred simulated sEEG channels in case A for the different advanced PLI attenuation methods implemented in iEEGPrep.	165
C.11	Distribution of SNR on the simulated sEEG channels in case B for the different advanced PLI attenuation methods implemented in iEEGPrep.	165
C.12	Example of power spectral density curves obtained through Welch’s method for one simulated signal filtered for case B through the different advanced line noise attenuation algorithms implemented in iEEGPrep. Orange and blue signals show simulated signals before and after PLI attenuation respectively. . . .	166

C.13	Time plot of the first 60 seconds of the re-referenced signal with PLI attenuation from the channels LPIIn02 and LPIIn03 using the implemented advanced PLI attenuation methods.	166
F.1	Small size CNN architecture implemented, referred as <i>CNN-Small</i>	172
F.2	Medium size CNN architecture implemented, referred as <i>CNN-Medium</i>	173
F.3	Large size CNN architecture implemented, referred as <i>CNN-Large</i>	174
F.4	Small size CNN+LSTM architecture implemented, referred as <i>CNN+LSTM-Small</i>	175
F.5	Medium size CNN+LSTM architecture implemented, referred as <i>CNN+LSTM-Medium</i>	176
F.6	Large size CNN+LSTM architecture implemented, referred as <i>CNN+LSTM-Large</i>	177
G.1	Comparison between the different deep learning models after architecture selection per class. CNN-Small and CNN+LSTM-Small correspond to the smallest implemented CNN and CNN+LSTM models respectively.	179
G.2	Examples of correctly classified samples from the <i>noise</i> class.	180
G.3	Examples of correctly and incorrectly classified samples from the <i>baseline</i> or <i>physiology</i> class	180
G.4	Examples of correctly and incorrectly classified samples from the <i>pathology</i> class	181

List of Tables

1.1	Summary of non-review studies analyzed during the literature review. Here, the columns refer to: <i>Re-ref</i> to the montage chosen for re-referencing, <i>Drift</i> to the correction method used for slow-drifts, <i>PLI</i> to the method used to decrease power line interference, <i>AD</i> shows the studies that employed artifact detection (marked with an <i>X</i>), <i>AL</i> shows the studies that performed anatomical localization of the electrodes (marked with an <i>X</i>), and <i>DS</i> shows those studies that performed downsampling to the raw data with an <i>X</i>	26
2.1	Summary of relevant clinical characteristics of the patients from the LHSC dataset used for qualitative evaluation of the developed iEEG classifier.	64
2.2	Summary of the number of clips per category for the three datasets used for training, validation and testing of the proposed iEEG classifier.	65
3.1	Metrics obtained from testing of best deep learning and traditional machine learning models on the external dataset from the University of Florida.	103
3.2	Metrics obtained per class for the best implemented model on the external dataset from the University of Florida.	104

List of Appendices

Appendix A iEEGPrep Report	150
Appendix B iEEGPrep snakebids.yaml	156
Appendix C iEEGPrep validation	160
Appendix D Fieldtrip code	167
Appendix E Models hyperparameters tested using GridSearchCV	170
Appendix F Deep learning models evaluated using cross-validation	171
Appendix G Additional results for models comparison	178
Appendix H Copyright Transfers and Reprint Permissions	182

List of Symbols

Symbol	Description	Unit
n	Harmonic number for line noise contamination (1 for fundamental, 2 for the second harmonic, etc.)	Dimensionless
ω_0	Fundamental frequency for line noise contamination	Hz
$W(f)$	Welch's method result (power spectral density) at a specific frequency	$\mu\text{V}^2/\text{Hz}$

Chapter 1

Introduction

1.1 Overview

This chapter gives a brief overview of different concepts required to understand the developed work. It starts by giving a definition of epilepsy along with its classification according to the International League Against Epilepsy (ILAE). Special focus is given to temporal lobe epilepsy and drug resistant epilepsy, as they are strongly tied to the use of intracranial electroencephalography (iEEG). Presurgical evaluation for epilepsy surgery is later covered, where iEEG is used as part of Phase II of the presurgical investigation.

Intracranial electroencephalography is also covered, describing its basics and how it is used for presurgical evaluation. The last section of the chapter covers an introduction to machine learning and how it can be used to analyze iEEG recordings. At the end of this chapter, an overview of this work is presented, along with its scope and objectives.

1.2 Epilepsy definition

1.2.1 Definition and overview

Epilepsy is defined as a chronic noncommunicable disease of the brain characterized by the presence of recurrent seizures [48, 115]. Epileptic seizures are defined as an uncontrolled excessive electrical activity in the brain, which result in episodes of signs and/or symptoms including involuntary movement and loss of consciousness [115, 42]. Seizures, the primary symptom of epilepsy, can potentially cause brain damage, resulting in significant behavioral, neurological, and cognitive alterations in affected individuals [24]. Furthermore, patients with epilepsy frequently suffer from a range of comorbidities, including but not limited to central nervous diseases such as stroke and dementia, psychiatric disorders such as depression and anxiety, and other non-neurological disorders such as heart disease and hypertension [160, 84].

The classification of epilepsy types plays a crucial role in the clinical assessment and management of patients experiencing seizures. It not only has a direct clinical impact on the diagnosis, treatment and outcome for patients but also is relevant to epilepsy research [133]. The ILAE has defined and redefined this classification of epilepsy since the 1960s [58, 56, 57]. Their most recent definition, released in 2017, presents a multilevel classification, which starts with the type of seizures, followed by the epilepsy types and finally the epilepsy syndromes, as shown in Figure 1.1.

The initial level of epilepsy classification separates seizures into three groups: generalized onset, focal onset, and unknown onset, each specifying the nature and origin of seizure activity [133, 43, 8]. Focal seizures originate in a small defined region of the brain and are further subdivided into five other subgroups according to the latest revision from the ILAE. Generalized seizures affect both hemispheres of the brain [43, 8].

The second level of the framework, shows the different types of epilepsy defined by the ILAE. Generalized epilepsy is characterized by the presence of generalized spikes or seizures on electroencephalography (EEG) recordings. On the other hand, focal epilepsy includes sev-

eral unifocal and multifocal seizures as well as seizures originating in one hemisphere. Furthermore, in their recent revision, the ILAE included a new group of combined generalized and focal epilepsies, which includes patients with both focal and generalized seizures [133].

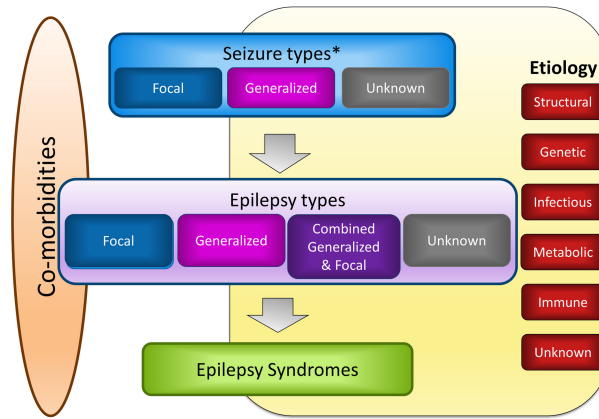


Figure 1.1: Graphical representation of the framework to classify epilepsy. Reproduced from [133] with permissions.

Additional to the definition provided by the ILAE, different forms of focal epilepsy are often named after the lobe where the onset zone (referring to the origin of the seizures) is located as either parietal, frontal, occipital or temporal [132].

The final level of the classification is trying to identify the specific epilepsy syndrome. The ILAE recommends, when possible, the diagnosis for a patient with epilepsy to be performed at all three levels. However, it also acknowledges that it is common to find cases where the epilepsy classification is the final level. According to the ILAE, epilepsy syndromes refer to a group of characteristics that often occur together, including seizure types and imaging features [133]. In 2021, the ILAE provided a more detailed description of these syndromes along with a methodology for their identification [157].

1.3 Diagnosis and treatment of epilepsy

Initial treatment for epilepsy typically involves Anti-seizure Medications (ASM), although these medications do not successfully control or eliminate seizures for all patients. About

one-third of epilepsy patients are diagnosed with drug-resistant epilepsy (DRE), defined by the persistence of seizures despite adequate trials of two different anti-epileptic drugs, according to the ILAE [90, 137]. Similarly, sustained seizure freedom is defined as the freedom of seizures for a minimum of twelve weeks or for a period three times longer than the previous longest seizure-free period, whichever is longer [128]. Approximately, 30 to 50% of patients with temporal lobe epilepsy (TLE) develop resistance to ASM [117].

1.3.1 Epilepsy surgery

Epilepsy surgery is considered the most effective treatment option following the unsuccessful management of seizures after two adequate trials of ASM. [131]. The primary objective of epilepsy surgery is to either completely remove or effectively disconnect the epileptogenic zone (EZ), which is the brain area necessary for initiating seizures [130]. This group of well-selected patients is defined based on the knowledge that certain forms of epilepsy tend to have a poor prognosis with ASM but an excellent surgical outcome [128, 38].

Once a patient is identified as a potential candidate for epilepsy surgery, a presurgical evaluation must be performed to select the ideal candidates for epilepsy surgery. The goal of this process is to evaluate how well the EZ can be *estimated* and if the EZ is possible to disconnect or remove safely [128]. The EZ can only be estimated as it is a hypothetical concept; however, its boundaries can be approximated by using other zones that are believed to be directly related to it. Some of these regions are shown in Subfigure *a* in Figure 1.2, including the seizure onset zone (SOZ) as the brain region where seizure are generated, the irritative zone as the tissue that generates interictal epileptiform discharges (IEDs), and the high-frequencies zone (HFO zone) as the area where HFOs are evoking from [163]. As shown in plot *c* in Figure 1.2, different surgical strategies can be followed to try to disconnect or remove the EZ based on these aforementioned zones.

To define as reliable as possible the EZ, multiple diagnosis modalities are utilized depending on the complexity of each clinical case. The first stage of presurgical evaluation includes

clinical evaluation, long-term video-EEG monitoring (electroencephalography accompanied with video recordings) and magnetic resonance imaging (MRI) with epilepsy protocol and neuropsychological testing [128].

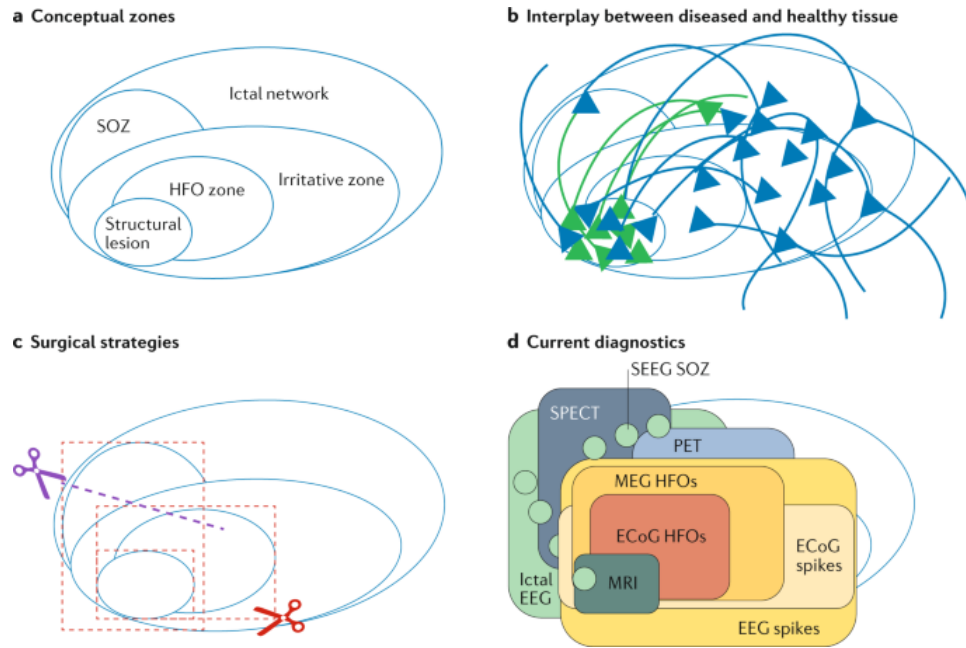


Figure 1.2: **a.** Examples of conceptual zones that can be used to estimate the EZ. **b.** Graphical representation of the connections between the epileptogenic network (green) and the healthy tissue network (blue). **c.** Examples of surgical strategies for epilepsy surgery, including resection of zones (red rectangles) or disconnection of networks (purple dotted line). **d.** Graphical representation of the conceptual zones that can be estimated using different modalities. Reproduced from [163] with permissions.

Frequently other investigations are conducted, including functional studies with different modalities such as MRI (functional MRI or fMRI), SPECT or Wada tests. If a testable hypothesis cannot be generated at this point, phase II of the presurgical investigation including intracranial electroencephalography is directed, where electrodes are placed inside the brain to record brain activity [128, 80]. Intracranial monitoring is often indicated for patients with Temporal Lobe Epilepsy (TLE) that present normal MRI, bilateral mesial temporal sclerosis in the MRI, dual pathology or electroclinical discordance; as well as in extratemporal epilepsy [128].

1.4 Intracranial Electroencephalography

A subset of patients with epilepsy require the implantation of intracranial electrodes for presurgical evaluation after non-invasive modalities fail to accurately delineate the EZ. In this section, we will review the basics of iEEG, along their use in clinical practice for epilepsy and an overview of some research conducted using this modality.

1.4.1 Basics of iEEG

Intracranial EEG refers to the placement of electrodes inside of the skull for the recording of brain electrophysiological activity. Compared to the traditional electroencephalography or scalp EEG, which is recorded from the scalp, iEEG exhibits a signal-to-noise ratio up to 100 times greater than that of conventional scalp EEG, primarily because it bypasses the attenuation effects of the skull and scalp [118, 12]. Due to this, the amplitude of measurements from iEEG are as 10 times higher compared to scalp EEG and they are less affected by external artifacts, such as physiological noise (for example, from the heart) or electro-magnetic noise from the recording room [118, 102].

Intracranial EEG offers highly localized recordings, as electrodes are strategically placed in targeted brain regions to maximize the diagnosis precision. This results in a better localization of the source of the signal compared to non-invasive EEG, where techniques such as source modeling and solving the inverse problem are needed to have an estimation of the source [102, 118]. Simultaneous recording from multiple regions is also possible using iEEG, which allows the analysis of multiple locations at the same time and the interaction between them [102].

Nevertheless, iEEG also has notable drawbacks when compared to scalp EEG, primarily due to its invasive nature and the associated surgical risks [80, 118]. Scalp EEG and other non-invasive modalities such as magnetoencephalography (MEG) also offer a more complete spatial coverage of the brain as electrodes are placed around all the scalp [102]. On the other hand, iEEG is often placed in specific regions of interest, resulting in a sparse spatial sampling

[80]. Furthermore, the limited accessibility of iEEG in clinical setting compared to scalp EEG also presents a challenge [80].

Intracranial electrodes include grid or strip electrodes for cortical surface application, and depth electrodes, which penetrate into the brain to target specific structures [102, 164, 118]. The first type of electrodes are known as subdural electrodes and the modality of iEEG is referred as electrocorticography (ECoG). The second is known as stereoelectroencephalography (SEEG) or stereotactic EEG, which uses semi-rigid shafts or leads with several contacts for recording, usually regularly-spaced [62, 102].

ECoG, compared to SEEG, offers extensive coverage of cortical areas, enabling functional mapping of these regions; however, its sensitivity to neural activities in sulci is lower, and it cannot access deeper brain structures. Furthermore, more potential complications are associated with subdural electrodes compared to SEEG, including intracranial hemorrhage and infections [9].

Clinically, there is a notable shift towards preferring SEEG over subdural electrodes in the last 10 years in North America, while the SEEG approach has been employed in Europe for many decades [2]. This growing preference for SEEG may be attributed to its lower risk profile, reducing complications such as intracranial hemorrhage and infections, as well as recent innovations in imaging and robotics [81]. Another clinical advantage of SEEG is the avoidance of craniotomy, referred to as the temporary removal of a part of the scalp, which is required for ECoG. SEEG implantation is considered a minimally invasive option, as will be detailed later, which also makes the removal of electrodes easier compared to ECoG [80, 2]. Due to this increased importance of SEEG, the work presented on this thesis will mainly focus on this type of iEEG modality.

The origin of electrophysiological activity recorded with intracranial electrodes depends on multiple factors, starting with the type of electrode. To explain this, it is first necessary to introduce the concept of local field potential (LFP). Any transmembrane current (current going from one side of the membrane to the other) in the brain leads to a change in voltage

both inside the cell (referred to as intracellular) as well as outside of it (extracellular) [20]. The superposition of these extracellular deflections at a particular point in the extracellular space is called the local field potential or LFP, and synaptic activity is thought to be primarily responsible for it [20]. The LFP can be recorded with micro or macro electrodes placed in the extracellular space of brain tissue, including SEEG electrodes. However, the extent of the neuronal population captured in an LFP is still unclear. Previous research has shown that several factors could affect the extent of the captured LFP when using intracranial electrodes, including the use of micro or macro electrodes or the recording configuration [94]. Other factors such as the neuron morphology and volume conduction (referring to the capturing of activity from sources far from the recording location) also influence the LFP [94]. On the other hand, ECoG electrodes capture a smooth version of the LFP as they are placed on the surface of the brain. It is believed that these electrodes mostly capture synaptic activity orthogonal to the cortex, however, volume conduction is also present in the recordings [20, 150].

In clinical practice, implantation of SEEG electrodes requires the positioning of the electrodes with respect to the target region, corresponding to those brain regions suspected to be part of the EZ. This positioning is defined using a 3D stereotactic frame (reason for which it is called stereo-EEG). This technique allows the targeting of deep structures through their referencing to a standardized coordinate Talairach atlas [148, 102]. Then, burr holes (small holes in the skull) are used to insert the shafts.

In standard clinical practice, typically 5 to 20 depth electrode shafts are implanted to delineate the EZ. Each shaft comprises 10 to 12 cylindrical contacts, each 2 mm in length and approximately 0.8 mm in diameter, usually spaced at regular intervals [80, 118, 102]. After implantation, iEEG brain signals, along video recording, are captured for an average duration of five to fourteen days depending on the case's complexity. During this period, data are visually analyzed by the clinical team (epileptologist) to try to localize the potential onset of seizures as part of the presurgical evaluation, explained furthermore in the next section [36, 118].

The sampling rate for human iEEG recordings varies across facilities, typically ranging

from 1 kHz and 3 kHz [118]. Research into sampling rates has primarily examined their influence on detecting crucial events in iEEG recordings, suggesting the importance of sampling at least at least at 1 kHz [28, 61]; however, no standardize practice is defined yet.

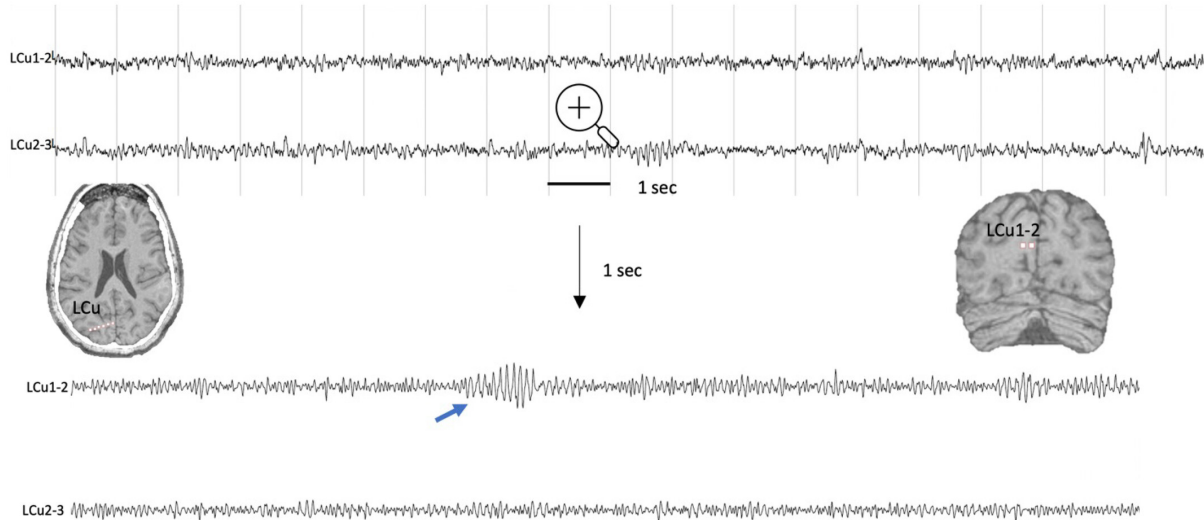


Figure 1.3: Example of physiological high-frequency oscillations (HFOs) recorded using SEEG electrodes for recordings. Data was recorded during wakefulness from an adult patient with left temporal DRE. Top panel shows two bipolar signals after filtering between 0.5 and 100 Hz. Bottom panel shows a zoomed version of the same recording after high-pass filtering of 150 Hz. Reproduced from [49] with permissions.

Recorded iEEG activity can be broadly classified into three distinct phases: ictal, interictal, and post-ictal. The period when seizures occur is referred to as ictal, while the period between the seizures is referred to as interictal and the period just after the seizure is defined as post-ictal [44]. Even though interpretation of iEEG activity is not as standardized compared to scalp EEG, prior efforts have been presented to improve the reading and analysis of these recordings. As for physiological activity, previous investigations have described patterns of physiological activity across regions, but they have been reported to be difficult to see in clinical practice. Other physiological characteristics, such as the presence of physiological high-frequency oscillations (HFOs) in the pre- and postcentral gyri, the mesiotemporal region and the occipital lobe, as shown in Figure 1.3. HFOs consist of the presence of ripples (80–250 Hz) and fast ripples (>250 Hz), and can be seen both in scalp EEG and iEEG [80, 52]. The next subsection will cover more details on pathological patterns that can be seen in iEEG recordings.

1.4.2 iEEG for identification of SOZ

As previously mentioned, iEEG is used in epilepsy as part of the presurgical evaluation to try to determine the location of the EZ. Intracranial recordings are usually associated with the seizure onset zone (SOZ), the region of the brain where seizures are generated [163]. Identification of the SOZ using iEEG is still a field of development and multiple biomarkers have been proposed. In this section, we will discuss a few of the well known biomarkers for delineation of the SOZ.

Specific biomarkers have been found for interictal and ictal periods. Interictal epileptiform discharges (IEDs), also known as interictal spikes, are the most traditional biomarkers in interictal iEEG activity, often present in regions with higher epileptic activity [80, 87]. According to the International Federation of Clinical Neurophysiology, a waveform has to fulfill at least four of the following criteria to be classified as an IED [49]:

- Presence of di-phasic or tri-phasic wave with sharp morphology.
- Different duration compared to the ongoing background activity.
- Asymmetric shape, with a sharp ascending phase and a more slow decaying phase.
- A slow-wave follows the transient.
- The background activity is disrupted by the waveform.
- The distribution of negative and positive potentials suggests it is a spike generated in the brain. Frauscher *et al.* indicate that it is not common for IEDs to have positive and negative polarities when using a referential montage, which helps differentiate IEDs from potential artifacts [49].

Previous studies have demonstrated good concordance between spike density and the SOZ for 56% of patients; however, it has been shown that using spikes for the delineation of the SOZ provides good sensitivity but low specificity [14, 80]. Figure 1.4 shows an example of high-amplitude spikes recorded using SEEG electrodes.

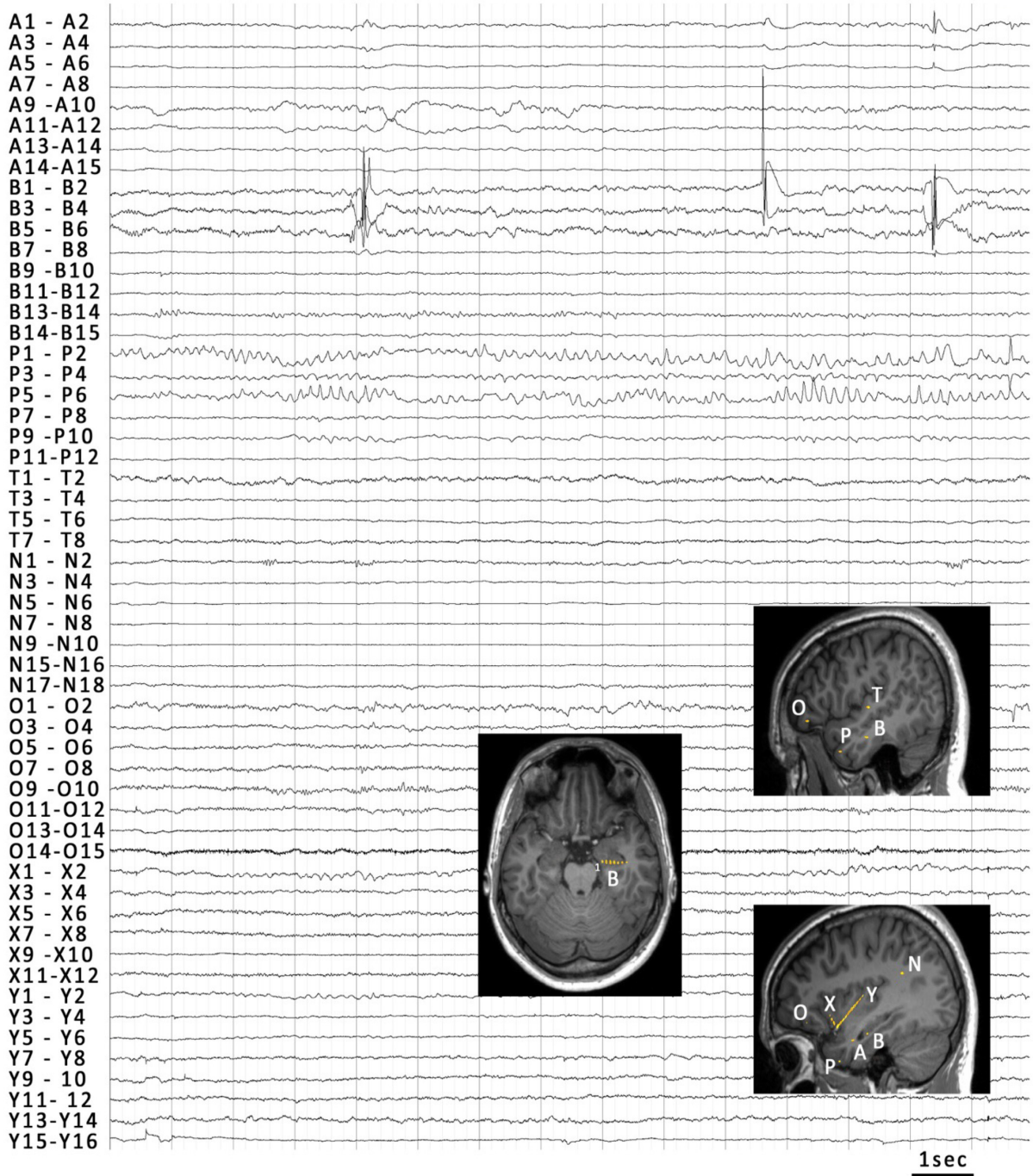


Figure 1.4: Examples of interictal epileptiform discharges (IEDs) in SEEG recordings. Figure shows that the recordings from the contacts B1 to B5 have presence of high-amplitude spikes and spike-slow waves. Reproduced from [49] with permissions.

HFOs are another interictal biomarker of the EZ, considered a very promising option lately. Despite its promises, it has been found that the mapping of the EZ using HFOs is not always

accurate, which could be related to the fact that there exist both pathological and physiological HFOs, and the properties that differentiate them vary across regions [80, 52]. Furthermore, these ripples could be confused with certain non-cerebral artifacts, such as electrode and muscle artifacts, hindering even more their correct identification [49].

Similar biomarkers can also be defined for the ictal period, which remains as the gold standard for delineation of the SOZ [80]. The most common ictal electrographic features are the reduction of background activity, followed by the appearance of high frequency activity and a subsequent slow potential (low-frequency activity) [80, 44]. It has been found that these features can be seen at the onset of seizures for regions inside the SOZ [44]. Other seizure-onset ictal patterns have been described for S EEG, including low-frequency high-amplitude periodic spikes, sharp activity at frequencies lower than 13 Hz and burst of high-amplitude polyspikes [49]. A further description of these is outside of the scope of this thesis as focus was given to interictal activity, as it will be described later.

Other approaches can be used for the clinical delineation of the SOZ using iEEG, including electrical stimulation [80]; however, it is considered to be outside of the scope for this work.

1.4.3 Modern analysis techniques

Despite advancements in iEEG biomarkers for identifying the EZ, epilepsy resective surgery results in seizure freedom in only 50% to 70% of cases, possibly related to the inherent subjectivity of current analysis [153, 39, 15]. As previously mentioned, ictal features remain the gold standard for delineating the EZ and have thus received the most extensive research focus. Recently, interictal features have garnered attention for their potential in delineating the EZ, though they often lack the specificity required for definitive diagnosis, as it is the case of HFOs [80, 52]. Furthermore, current approaches require a large time of hospitalization for monitoring, associated with a high cost and low accessibility [36, 118].

For these reasons, recent efforts have been made towards discovering new biomarkers and enhancing localization of the EZ through the usage of more computer-driven quantified analy-

sis [80]. One notable approach that multiple researchers have followed is the use of normative maps, which define a normal range of variation for specific features for healthy individuals to identify potential abnormalities in patient data [149].

Frauscher *et al.* were pioneers in this area, developing an atlas that characterized normal iEEG activity during periods of resting wakefulness [51]. In this study, the authors analyzed the statistical distribution of normalized Power Spectrum Density (PSD) curves across healthy brain regions using data from intracranial recordings, as exemplified in Figure 1.5. Frequencies between 0.5 Hz and 80 Hz were analyzed and profiles were created using Welch's method for the PSD curves with recordings coming from 106 subjects [51].

A further study by Bernabei *et al.*, expanded the atlas with additional data from 38 subjects [15]. Furthermore, they also investigated the usefulness of these profiles on distinguishing channels from the SOZ and irritative zone from the not affected regions by statistically comparing the PSD distribution of these different groups. Several regions were found to have significant differences; however, many clinically relevant regions for epilepsy did not show any differences, including the hippocampus and the amygdala [15]. Further features were explored in this study, including connectivity between regions, which resulted in a better prediction of the SOZ.

Similar studies have been developed by other authors, where other features such as band-power have been explored, showing potential pathways for better ways to delineate the EZ [149, 156]. Despite their contributions, these studies present several limitations that require careful consideration.

A critical limitation of these studies is that they aim to distinguish healthy brain characteristics using data exclusively from patients with epilepsy, which could compromise the validity of the normative atlas. Therefore, the construction of these profiles of so called normality are based on a very intensive data selection and cleaning of the data to try to identify all of the potential channels contaminated with pathological events, artifacts or that are inside of the clinically-defined SOZ.

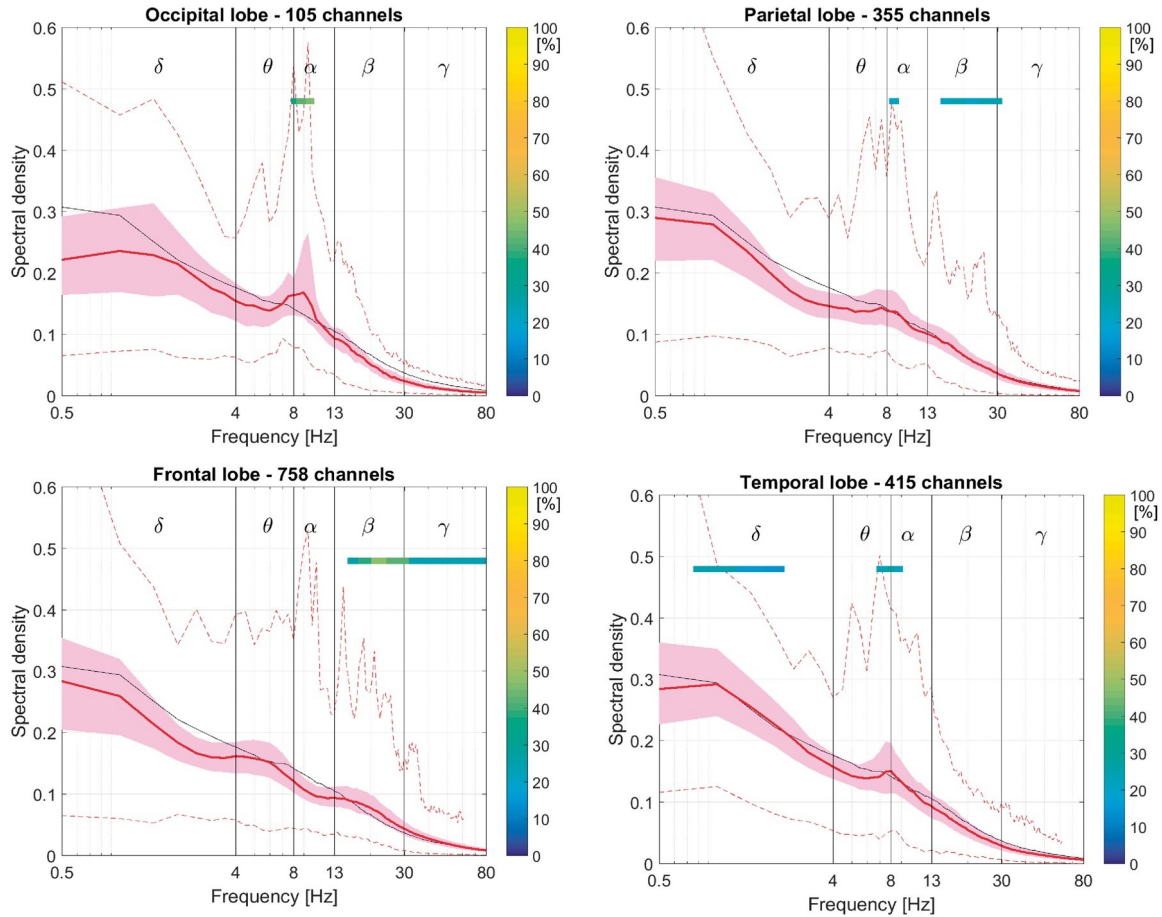


Figure 1.5: Example of spectral profiles developed by Frauscher *et al.*. Each panel shows a different statistical distribution of normalized power between 0.5 Hz and 80 Hz, with the red solid line representing the median of the distribution on each frequency point, the red overlay the region between 25 and 75 percentiles, and the dotted red lines showing the maximum and minimum. The color bar shows the percentage of channels that showed a peak in PSD in the specific frequency band in which it is located. Reproduced from [51] with permissions.

Depending on the study, even further criteria are used, including the time distance of the analyzed event from ictal periods. Most of these cleaning steps are performed manually by a specialist, which is not only time consuming but also injects potential subjectivity into the process. Moreover, these analyses require extensive data samples to ensure robust generalization, highlighting the critical need for their reproducibility. The lack of standardized preprocessing protocols for iEEG data complicates these efforts, interfering with both the replication and extension of research findings.

Automation or semi-automation of some of these cleaning steps, such as the detection

of pathological events or artifacts, presents as a promising avenue to streamline these analyses. Implementing standardized preprocessing methods together with automation could significantly facilitate the execution, replication, and expansion of analyses in iEEG studies

1.5 Brain Imaging Data Structure (BIDS) for iEEG

Standards to structure large datasets are necessary to ensure good sharing practices and facilitate further processing of the data. The Brain Imaging Data Structure (BIDS) standard is a specification designed to organize neuroscience data and metadata with the aim of promoting transparency, reusability and reproducibility [70]. This standard was originally written for Magnetic Resonance Imaging (MRI) data; however, through a community-driven process, it has been extended to other modalities, including MEG and EEG [70, 113, 121].

In 2019, an extension of this specification was published for iEEG data [70]. Figure 1.6 shows an example of a directory compliant with the iEEG-BIDS specification, highlighting the most important parts of the specification.

A summary of the most important parts of the iEEG-BIDS specification are shown below:

- The two original data formats supported to store the iEEG recording data were the European Data Format (EDF) and the BrainVision Core Data Format; however, it has been extended to EEGLAB files, Neurodata Without Borders and Multiscale Electrophysiology Format version 3 (MEF3) according to the official documentation. EDF and BrainVision data formats are still recommended over the others.
- A sidecar JSON file (*_ieeg.json*) has to be specified, showing the metadata from the iEEG recording data, including but not limited to the reference scheme (for example, “bipolar”), sampling frequency and powerline frequency.
- The electrodes localization given by coordinates should be stored in the *_electrodes.tsv* file, usually in 3D space. The term *electrode* is used as a reference to “a single point

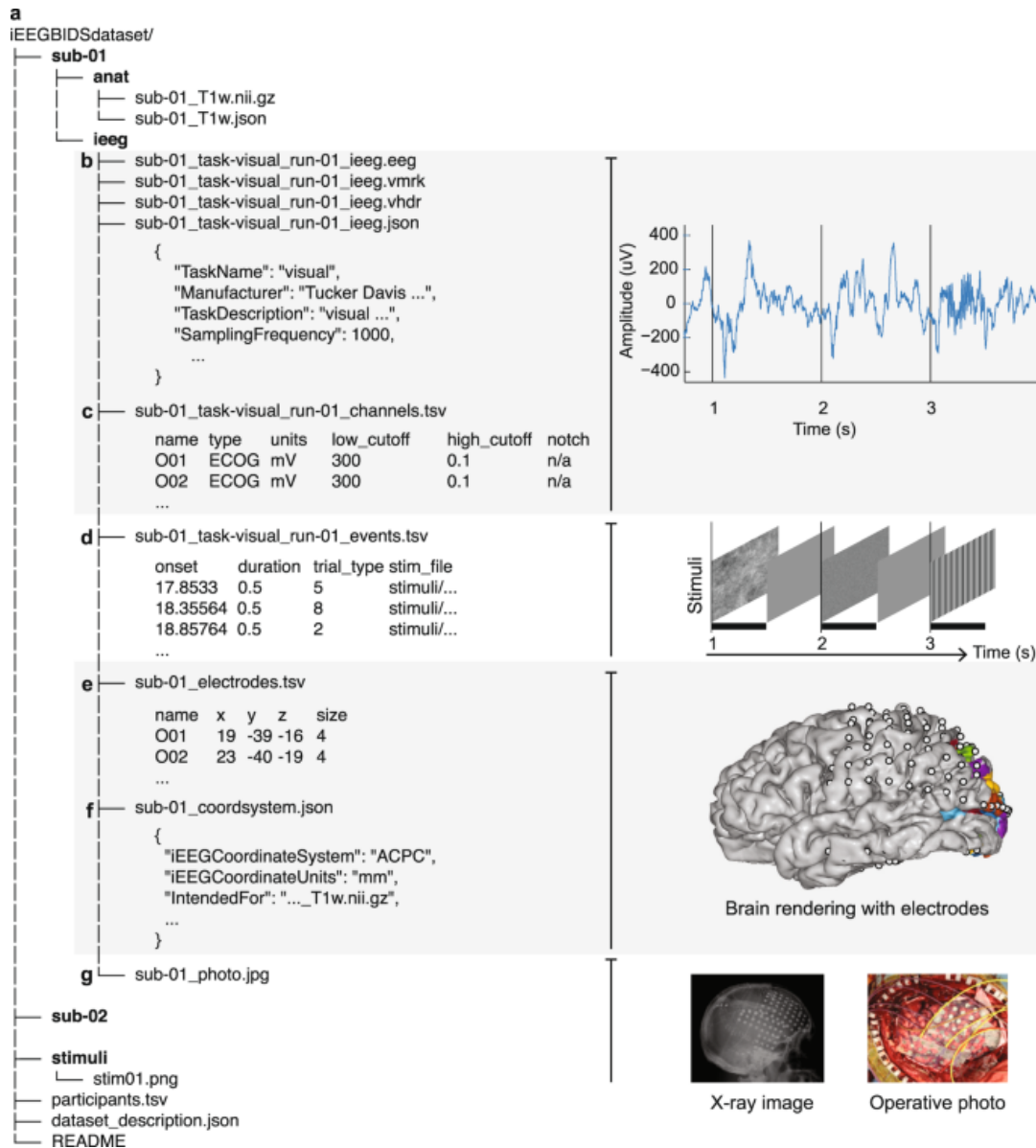


Figure 1.6: Example of a iEEG-BIDS directory for a session with electrical stimulation along with example pictures. First layer shows the different subjects' subfolders and the stimuli directory. **(b)** Shows the files associated with the recording data, in this case in BrainVision Core format. **(c)** The information related to the channels in the recording are stored in the *_channels.tsv* file. **(d)** The events of importance during the iEEG recording are stored in the *_events.tsv* file, in this case, showing stimulation events. **(e)** The *_electrodes.tsv* file is intended to store information about the electrodes, including their spatial location. **(f)** The *_coordsystem.json* file should show the information of the coordinate system used to define the electrodes location. **(g)** Shows other images that can be stored as part of the directory. Reproduced from [70] with permissions under Creative Commons License.

of contact between the acquisition system and the recording site” [70]. The coordinate system specifications used to define the electrode coordinates should be stored in the *_coordsystem.json* file.

- The channels’ information should be stored in the *_channels.tsv* file. A channel represents one time series recorded with the recording system [70]. The information includes but is not limited to the channel name, the type of channel (for example, SEEG and ECoG), sampling frequency and filters cutoff frequencies if any was applied during recording [70].
- Any events of importance during the recording can be stored in the *_events.tsv* file. This could include, for example, electrical stimulation events, indicated by the onset (time at which the event started relative to the beginning of the recording), duration of the event and event label.

For more information, please refer to the original article for the iEEG-BIDS specification and the official documentation [70]. These efforts pave the way for standardizing the preprocessing of iEEG recordings by facilitating the development of workflows based on the BIDS standard.

1.6 Preprocessing of iEEG data

As noted earlier, iEEG recordings exhibit a signal-to-noise ratio approximately ten times greater than standard scalp EEG, which historically suggested minimal artifact presence. Despite this, studies have shown that iEEG recordings still capture electrical potentials generated outside the brain, contradicting earlier assumptions of minimal contamination [102, 101]. Similar to the case of scalp EEG, it has become a common practice to apply preprocessing steps to the data to reduce the effect of these potential artifacts and enhance signals of interest before any further analysis.

It is relevant to highlight that recent trends in scalp EEG analysis have shifted towards minimizing preprocessing, questioning its necessity [33]. Preprocessing is commonly a time-consuming and highly customized task; therefore, the study by Delorme investigated the utility of preprocessing when using the data for further analysis, finding no significant differences for most preprocessing steps when compared to using raw data [33]. Although it is important to consider the results from this study, the author emphasized on how these findings apply for EEG acquired in a very controlled environment and might not extend to other data acquired in different conditions [33, 35]. Similar research has been done for iEEG data by evaluating the results of statistical comparisons and machine learning classification tasks when using clean vs raw data; however, no concrete generalizable results have been found [98, 101].

Despite the fact that preprocessing has become a common practice for iEEG analysis, iEEG analysis traditionally relies on ad hoc workflows [140]. Despite efforts to standardize iEEG data preprocessing, most studies continue to employ custom workflows tailored to specific research needs, incentivized as well by the lack of a software package that integrates common preprocessing steps [140, 101, 98]. The importance of standardization and automation of preprocessing has been recognized for MEG and EEG fields, where there have been different efforts to solve this by the development of software tools like PREP pipeline for EEG [16].

With the recent trend towards machine learning solutions for iEEG, it is fundamental to have standardized preprocessing practices that allow replication of findings and combination of datasets prior analysis [16, 104]. Similarly, for non-machine learning related studies, this is also fundamental to allow for replication or extension of studies.

1.6.1 Overview of preprocessing steps for iEEG

As described in literature, there are certain steps that are commonly used and recommended to preprocess iEEG data. In this section, an overview of each of these steps will be given focusing on their main purpose as part of the preprocessing.

Re-referencing

As described by Mercier *et al.*, recording of iEEG has to be performed using a reference [102]. Electrical potentials recorded by the equipment represents a voltage difference between two points, usually the electrode contact of interest and a reference that can be either external, such as an electrode on the skin, or internal as an electrode contact in the white matter of the brain; this is referred to as online referencing [102]. Common practice involves applying re-referencing post-acquisition, a process known as offline re-referencing.

Intracranial EEG signals are composed not only by local brain activity around the electrode contact but also a mixture of other sources either from other parts of the brain through the volume conduction effect or from outside the brain, as it is the case of line noise activity from the room and recording equipment connected to the electrical grid [102]. Offline re-referencing alleviates some of these components as it reduces power line noise interference and volume conduction effects, increasing the signal-to-noise ratio (SNR) of the signals [102]. There are several re-referencing schemes that can be used, each of them with their own advantages and disadvantages. Here, priority will be given to re-referencing schemes for SEEG recordings as the workflow was initially developed for this type of iEEG electrodes.

Bipolar re-referencing is a standard choice for re-referencing, consisting of the re-reference of each channel to its adjacent contact in the same electrode shaft, performed by the subtraction of the two [98, 102, 103]. This scheme results in a signal composed by the activity near the two electrode contacts and the volume-conducted activity in between [102].

Another popular scheme is common average referencing (CAR). In this scheme first an average signal of all clean channels is computed and then it is subtracted from all channels, aiming to isolate common noise prevalent across all channels [102, 98]. A big difference between CAR and bipolar referencing is that the latest gives more relevance to local activity, while the first one highlights widely distributed features in the signals [102].

Laplacian re-referencing is similar to bipolar re-referencing but instead of using two electrode contacts, each channel is re-referenced against the two adjacent contacts within the same

shaft [102, 98, 97]. This is usually performed as the subtraction of the channel signal with the average signal from the two adjacent contacts [98, 97]. There are other re-referencing schemes used, such as gray-white matter reference (GWR), however, it is outside of the scope of this project to discuss all of them.

Previous studies have shown that Laplacian referencing typically outperforms other re-referencing schemes, followed by bipolar referencing, which has been shown to perform better than CAR when analyzing a multitude of tasks [97, 98, 101]. However, it is important to remember that the selection of a re-referencing scheme should be tailored to the specific objectives of each study, depending on whether the focus is given to local or global neuronal activity [102].

Drift correction

Similar to scalp EEG, iEEG is also affected by slow drifts or low-frequency waves, on top of which the cortical activity is actually found [30]. These drifts are often considered artifacts arising from various sources, including changes in electrode impedance [30]. Slow drifts induce both a noisy trend to the signals, as well as a DC-offset that makes the signal to be centered around zero.

High-pass filtering and detrending are commonly used to attenuate slow drifts in iEEG data. High-pass filtering refers to applying a filter to attenuate activity below a specific cutoff, usually ranging between 0.1 Hz and 2 Hz, which removes both the DC-offset and the trend in the data [30, 102]. Detrending is based on the fit of a low order polynomial (often a linear function) to the data and then the subtraction of them [30, 102]. Demeaning is another alternative, based on the centering of each signal around zero by the subtraction of the mean; however, this only deals with the DC-offset and not the trend in the data. High-pass filtering is often the preferred approach, especially for long-duration data.

Power line interference (PLI) filtering

Electrophysiological recordings, such as scalp EEG and iEEG, are susceptible to electrical noise, commonly known as power line interference (PLI). There are different approaches followed to minimize the effect of PLI, but the most common one is notch filtering, where all of the data between a range of frequencies around the fundamental line noise frequency is filtered. The main disadvantage of this approach is the loss of any other physiological information behind the line noise that could be present in the attenuated frequency band. Waveform distortion is also a known issue when applying notch-filtering [29, 86].

To overcome these limitations, alternative methods such as adaptive filtering, independent component analysis, and spatial filtering techniques have been developed. For this work, three techniques were implemented and evaluated for this work: *removePLI*, *Zapline* and *Cleanline*.

The *removePLI* technique employs a time-based regression approach to mitigate PLI. Briefly, the signal of interest is first band-pass filtered around the fundamental frequency of the power line (50 Hz or 60 Hz). A lattice adaptive notch filter is then used to find the fundamental line frequency on the band-pass filtered data. Once the fundamental frequency is found, the harmonic frequencies are computed recursively and a discrete time-oscillator is used to generate the harmonic sine waves. Finally, a simplified recursive least squares algorithm is used to fit the harmonic sinusoids to the original signal, which should represent the best power line harmonic fits to the original signal. Then, the clean signal would be calculated as the error between the best fit and the original signal [86].

The *Cleanline* algorithm is also based on regression but in frequency domain instead of in time domain and it is implemented in a widely-used toolbox (EEGLAB) [16]. This method transforms the data into frequency domain using the multi-taper transform. Similarly to the previous method, it is assumed that the line noise is represented with sine waves; therefore, a regression in frequency domain is conducted between the amplitude and phase of the sine waves around the frequency of interest (the line frequency) and the multi-taper result. Thompson F-test is utilized to find and discard only those frequencies with sine waves with significant

non-zero [16].

Zapline diverges from regression-based methods, employing a decorrelation strategy to isolate and remove noise. The first step of this algorithm is the separation of the data potentially contaminated by the fundamental line frequency (X_n) and its harmonics from the clean data (X_b), achieved with a square filter. Then, the main idea behind the method is trying to distinguish the part of X_n that is actually contaminated with powerline. To accomplish this a denoising matrix D is computed using the joint decorrelation method [32]. Briefly, a filter is applied X_n to enhance the potential line interference. Then, the main components of the biased X_n are computed, which should represent the line noise contaminated components of X_n . Principal component analysis (PCA) is then used to remove the parts of X_n contaminated with this line noise components. Finally, the clean signal is obtained as a summation of this cleaned X_n and X_b [29].

Artifact detection

Signal-to-noise ratio in intracranial EEG has been found to be around 100 times higher compared to the non-invasive scalp EEG [118, 12]. The recording of electrophysiological activity from inside the brain makes it less susceptible to attenuation caused by the scalp and it also reduces the influence of external noise sources such as movement artifacts [118, 102]. Nevertheless, several studies have found different types of transient artifacts in iEEG recordings, including artifacts caused by eye movement, blinks and muscle movements [116, 12, 78].

As iEEG is typically recorded in brains of patients with epilepsy, electrodes are subject to epileptiform activity. This is not only limited to seizure or ictal activity, but also epileptiform activity often present in interictal periods. When studying the underlying physiological activity from iEEG data, pathological activity is not of interest and it is usually excluded. This is the case for cognitive studies and investigations studying normal activity of the brain using iEEG [15, 51].

In the case of scalp EEG, multiple artifact detection techniques have been proposed and a

few of them are included in commonly used software packages. Some of these include Artifact Subspace Reconstruction (ASR) available in EEGLAB, *ICLabel* based on the identification of noise-related ICA components, also available in EEGLAB; and Autoreject available through the MNE Python package [107, 124, 75]. Nevertheless, none of these tools have been validated for usage in iEEG data. Currently, there are no recommended tools for iEEG artifact detection, possibly making this preprocessing step the hardest to standardize.

Downsampling

Downsampling refers to the offline process of reducing the number of temporal samples in recorded data; in other words, reducing the sampling rate of the signal. This is a common step used in EEG studies and it has been used for iEEG as well [142]. After downsampling, it is important to apply a low pass filter to avoid aliasing on the resampled data, which is a distortion of the signal due to the presence of waves that cannot be sampled properly with the utilized sampling rate [31].

Channel localization

As iEEG is commonly performed along with other imaging modalities, such as computer tomography (CT) and MRI, electrodes positions in the brain can be obtained and aligned with a segmentation of the brain images, which can be used to identify the most relevant brain region recorded by each channel. It is also a common practice to use this localization to discard any signals from electrode contacts located outside of the brain or surrounded mostly by white matter [102].

There are generally two approaches for labeling electrode contacts based on brain segmentation as shown in Figure 1.7: performing a segmentation in the individual space or using an atlas-based approach in a normalized space (also referred to as template). As stated previously, iEEG is usually acquired along with a preoperative MRI and a postoperative CT. The contacts location can be obtained from the latter and the former is the modality commonly used

to get the anatomical segmentation; therefore, an alignment or coregistration between them is required, which is then referred to as *individual space*.

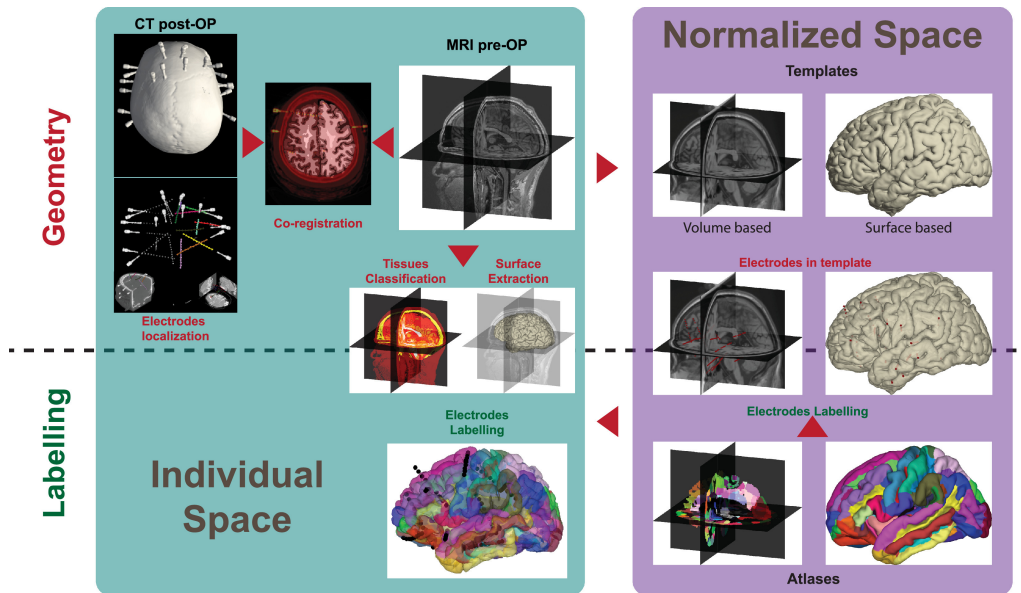


Figure 1.7: Electrode contacts localization workflow based on individual or normalized space. Reproduced from [102] with permissions under Creative Commons License.

As aforementioned, MRI segmentation can be directly performed using tools like SynthSeg [17], representing the individual space approach, or by registering the MRI to a template, such as MNI152 space, and applying an atlas to obtain the brain parcels [102, 46, 45]. These atlases are usually based on average characteristics across subjects, for which it could be less accurate than the individual space parcellation. Figure 1.8 summarizes the workflow that can be followed to obtain the region for each iEEG channel.

In case of using a monopolar reference scheme such as CAR, the region for each electrode can be defined by creating a volumetric mask around the estimated centroid of the contact and then counting the region with the most amount of voxels from this mask [102]. Bipolar case is more complex as the obtained signal represents the activity near the two contacts and in between them. One approach is using a *virtual sensor* in between the two contacts, as an average of their two original positions. However, this can lead to misleading information, especially if the two contacts that form the bipolar channel are located in different brain regions

[102].

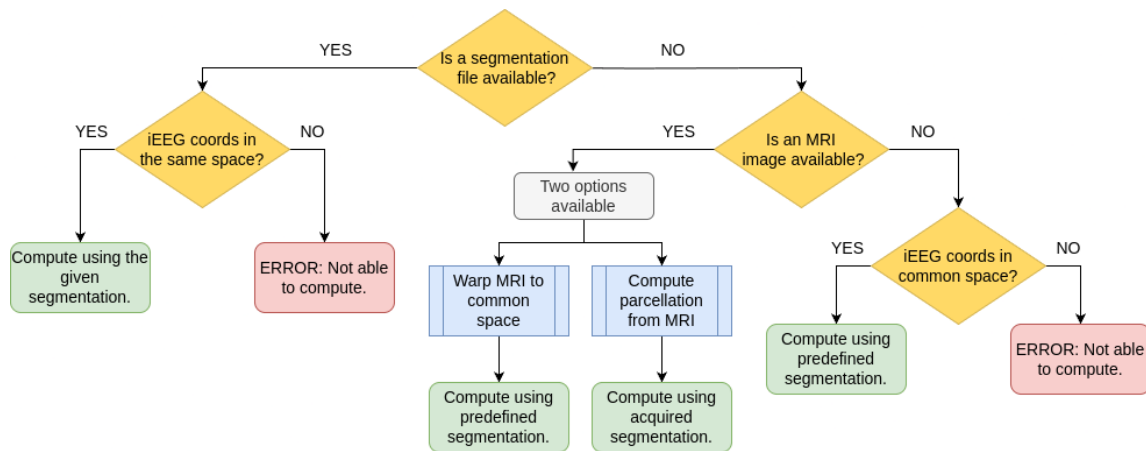


Figure 1.8: Diagram showing the different ways to identify the region by iEEG channel. The left side shows the case where a segmentation file is available, in which case it is required to give the iEEG coordinates in the same space as the segmentation. The right side shows the possible paths to follow when a segmentation has not been previously calculated, including the alignment to a common space or the calculation of the segmentation when an MRI file is available. If neither an MRI image nor a plot is available, the only option left is to give the iEEG coordinates in a common space to use a predefined plot in that space.

1.6.2 Common preprocessing steps: a brief literature review

A non-exhaustive literature review was conducted to assess the usage of the specified preprocessing steps in existing studies. Twenty six studies were reviewed in total, which represents a small selection of studies to span the range of preprocessing methods in previous publications. Table 1.1 shows a summary of the different preprocessing methods applied on each of these studies.

During this review, it was noted that different PLI attenuation methods were used in these studies, including simple notch filtering, adaptive filtering and rejection of channels with high line noise interference. In addition, CAR and bipolar were the two most common re-referencing schemes used in the reviewed publications. Finally, artifact detection was noted to be focused sometimes only in non-cerebral artifacts, but detection of pathological artifacts or both non-cerebral and pathological events was also found during this review.

<i>Paper</i>	<i>Modality</i>	<i>Re-ref</i>	<i>Drift</i>	<i>PLI</i>	<i>AD</i>	<i>AL</i>	<i>DS</i>
Rockhill <i>et al.</i> [129]	SEEG	CAR	None	None	X	X	
Groen <i>et al.</i> [65]	ECoG	CAR	None	Notch	X	X	X
Li <i>et al.</i> [96]	Both	CAR	HPF	Notch			
Bernabei <i>et al.</i> [15]	Both	BIP	HPF	Notch	X	X	X
Frauscher <i>et al.</i> [51]	Both	BIP	DM	Other	X	X	X
David <i>et al.</i> [27]	SEEG	BIP	None	None	X	X	
Kini <i>et al.</i> [88]	ECoG	CAR	None	Notch			
Sumsky & Greenfield [141]	Both	CAR	HPF	Notch			
Taylor <i>et al.</i> [149]	Both	CAR	None	Notch	X	X	X
Wang <i>et al.</i> [156]	Both	CAR	HPF	Notch	X	X	X
Nejedly <i>et al.</i> [111]	SEEG	None	None	Rejection	X		X
Michelmann <i>et al.</i> [103]	SEEG	BIP	HPF	None	X		X
Chen <i>et al.</i> [23]	SEEG	CAR	None	Notch	X	X	X
Kalamangalam <i>et al.</i> [82]	ECoG	None	DM	Notch	X		X
Gomez-Ramirez <i>et al.</i> [67]	Both	None	HPF	Notch			
Das <i>et al.</i> [25]	Both	BIP	None	Notch	X	X	
Jiang <i>et al.</i> [79]	SEEG	BIP	None	Notch	X	X	
Jensen <i>et al.</i> [77]	SEEG	BIP	None	None			
Jedynak <i>et al.</i> [76]	SEEG	BIP	None	None	X		
Gunnarsdottir <i>et al.</i> [66]	SEEG	CAR	HPF	Notch	X	X	
Quitadamo <i>et al.</i> [126]	SEEG	BIP	None	None	X		X
Sun <i>et al.</i> [143]	SEEG	BIP	HPF	Notch	X		X
Sinha <i>et al.</i> [138]	Both	CAR	HPF	Notch	X	X	
Medvedev <i>et al.</i> [100]	Both	None	HPF	Notch	X		
Lai <i>et al.</i> [91]	ECoG	None	None	Notch	X		
Thomschewski <i>et al.</i> [151]	Both	None	HPF	Notch			

Table 1.1: Summary of non-review studies analyzed during the literature review. Here, the columns refer to: *Re-ref* to the montage chosen for re-referencing, *Drift* to the correction method used for slow-drifts, *PLI* to the method used to decrease power line interference, *AD* shows the studies that employed artifact detection (marked with an X), *AL* shows the studies that performed anatomical localization of the electrodes (marked with an X), and *DS* shows those studies that performed downsampling to the raw data with an X.

1.6.3 Existing Solutions and Limitations

Multiple software packages have been developed for the analysis of scalp EEG recordings, which have been recently used for preparation and processing of iEEG data. In this section, we will briefly discuss some of these tools, focusing on their usability for iEEG preprocessing and limitations. Support for each discussed preprocessing step will also be discussed.

- **Magnetoencephalography and Electroencephalography in Python (MNE-Python):**

This is an open-source Python module initially developed for the preparation, processing, and analysis of EEG and MEG data [92]. Support for other functional neuroimaging data has also been added to the package, including SIEEG, ECoG and fNIRS.

According to the official documentation, MNE is capable of reading multiple file types supported by the BIDS-iEEG standard, which will be discussed in the following section, including EDF, BrainVision Core Data Format and EEGLAB files. Extraction of epochs is also supported; however, it requires loading of the full data first, which could be computationally intensive depending on the file size. Support for the described common preprocessing steps:

- *Downsampling:* Resampling is available for both continuous and epoched data, which by default uses the Fast Fourier Transform (FFT) method. This method computes the frequency components of the original signal and truncates the higher components according to the desired new number of samples in the signal. The main limitation of this method is the nature of FFT, which assumes periodicity for the signal, which could not be true for neural signals.
- *Drift correction:* The official documentation recommends high-pass filtering for drift correction. Demeaning and detrending are only supported for epoch objects. The only supported PLI reduction method by MNE-Python is notch filtering, allowing the use of both finite impulse response (FIR) and infinite impulse response (IIR) filters.
- *PLI attenuation:* The only supported PLI reduction method by MNE-Python is notch filtering, allowing the use of both finite impulse response (FIR) and infinite impulse response (IIR) filters.
- *Re-referencing:* Re-referencing in MNE-Python is not designed for SIEEG or ECoG; therefore, there are no predefined montages for re-referencing of iEEG recordings.

CAR can be calculated with this package; however, information about the bad channels to exclude for the average reference calculation must be given. Bipolar re-referencing is available but it requires inputting all the combinations of monopolar channels that will form the new bipolar channels, making the process very manual and time consuming.

- *Channel localization*: MNE-Python only supports localization of the most relevant brain region for each channel through Freesurfer parcellations. Therefore, either the positions of the channels have to be specified in the same space as *fsaverage* (MNI-305 space) or Freesurfer must be executed for the subject of interest before. MNE-Python defines the region for each channel only based on the closest region to each given coordinate. Moreover, when re-referencing to bipolar, the position of the new channel is always set to (0, 0, 0) according to the official documentation, then it is not possible to identify the closest brain region after re-referencing. Other re-referencing schemes such as Laplacian re-reference are not supported.
- *Artifact detection*: This library has availability of multiple methods designed to detect artifacts in MEG/EEG data; however, most of them require external channels to work, such as electrocardiography (ECG) or electrooculography (EOG) channels to detect heart and ocular artifacts. Autoreject is an external package that can be used along with MNE-Python; however, it is optimized for MEG/EEG data and has not been validated for iEEG recordings [75].

Overall, MNE-Python is an open-source package that is optimized for MEG and scalp EEG data, not for iEEG, which makes it difficult to develop a complete preprocessing workflow for iEEG datasets using this tool. Main limitations were found on extraction of epochs of interest for large files, re-referencing, definition of the most relevant brain region per channel and artifact detection. The following limitations were also found for this tool:

- Even though MNE-Python counts with vast documentation, it is still a tool that requires coding expertise to use it to preprocess iEEG datasets, limiting its usability for a group of users.
- No predefined preprocessing pipelines for iEEG has been developed using MNE, which promotes *ad hoc* workflows, hindering reproducibility across research centers.
- MNE has developed a MNE-BIDS-Pipeline for distributed computing systems, where multiple computers can compute in parallel to speed up the process. However, it only supports MEG and EEG datasets, which limits the scalability of the tool for iEEG preprocessing.
- As there is no predefined preprocessing workflow for iEEG data, quality control (QC) of the outputs has to be performed with *ad hoc* scripts, which could potentially limit the QC for experienced users.

- **Brainstorm:**

Brainstorm is an open-source application developed for the processing of brain recordings, including MEG, EEG, fNIRS, ECoG and SIEEG [147]. Brainstorm was designed mostly for usage through a graphical user interface (GUI) to mitigate the need of programming skills for its usage. Within its functions, you can find preprocessing of the data, visualization in time, frequency and time-frequency domain and functional connectivity analysis. Brainstorm is capable of reading multiple iEEG-BIDS file formats like EDF, BrainVision Data Format, Neurodata Without Borders and EEGLAB datasets. Similar to MNE-Python, Brainstorm allows the extraction of epochs based on events; however, no information on how to create a script for this was found.

Support for the described common preprocessing steps:

- *Downsampling:* Resampling can be performed in both the GUI and through the `process_resample()` function. No information about the methods used was

found.

- *Drift correction*: This can be achieved either through the implementation of a high-pass filter. According to the documentation, high-pass filters are FIR filters designed with a Kaiser window, and time and frequency responses can be visualized for QC. Linear detrending and DC-offset removal are also available. Polynomial detrending is not supported.
- *PLI attenuation*: Notch filtering through an IIR filter, for which the frequencies to filter and the notch bandwidth must be given. No other methods were found in the documentation or through the GUI.
- *Re-referencing*: Multiple predefined re-referencing schemes are available for iEEG data, including bipolar montage and common average reference (CAR). Unlike MNE, when applying bipolar re-referencing, the new virtual channel is positioned in the middle of the two contacts. This was found to be the best tool for re-referencing SEEG/ECOG data within the ones investigated.
- *Channel localization*: Region identification for each channel is supported by Brainstorm both by providing the coordinates for each contact in both MNI or subject space. If provided in subject-space, the MRI and parcellation files can be given. Two other options are the estimation of the parcellations from the given MRI using different options such as FastSurfer and FreeSurfer (which require prior manual installation of the tools in the system) and the registration of the MRI to MNI space. If using coordinates in MNI space, default anatomies and parcellation can be used, including Freesurfer templates and the AAL parcellation. Both surface and volume parcellations are supported and the region for each channel is assigned by creating a sphere of a customizable radius around the given coordinate. Therefore, bipolar channels use a sphere located in the midpoint of the two contacts.

The main limitation found for this preprocessing step was the lack of documentation of whether this can be achieved when doing batch processing for multiple

subjects and no options were found to reproduce this when using the pipeline creation tool for the automatic generation of scripts.

- *Artifact detection:* Detection of heart and ocular artifacts is possible when ECG or EOG channels are available. Otherwise, the automatic artifact detection available is based on the identification of events above a certain magnitude threshold for a given frequency band; in more detail, the signals are bandpass filtered using the given frequency range and then it labels as artifacts all the events above the specific threshold.

Brainstorm is designed to be used mainly as a GUI-based application; thus, although a scripting tutorial is provided, there is no formal documentation for each of the functions and the help provided for each of them in the terminal is also very brief. The recommended process, according to the documentation, for scripting is the automatic generation of the scripts through the GUI, without any proper review of the generated code. Based on the information found, the following limitations were found for this tool:

- The transparency of preprocessing workflows developed with Brainstorm is limited to the understanding of the scripts automatically generated, which is difficult due to the limited available documentation.
- The reproducibility of the workflows is also limited both in the case of using GUI, as it is almost impossible to share, or if using automatically generated scripts, since it would be required to understand the code to adapt it to other datasets or needs, which is difficult due to the lack of sufficient documentation.
- The usability of the tool when using the GUI is good but there could be a big learning curve before knowing how to properly use the tool. Furthermore, there are no predefined ways to QC the preprocessing, making the preprocessing process less standardizable.

- No documentation available on how to parallelize Brainstorm processes on distributed computer systems was found, limiting the scalability of the tool.
- Even though Brainstorm can be installed without a MATLAB license, some of its features are only available when installed with MATLAB desktop, as it uses some dependencies from Signal Processing and Parallel Processing, which require the acquisition of a paid license.

- **EEGLAB:**

EEGLAB is another open-source MATLAB toolbox for the processing of electrophysiological signals, including scalp EEG and MEG [34]. Some of the available functionalities include time/frequency analysis, preprocessing and visualization of the data. This tool offers most tutorials and information for usage through their GUI; however, some tutorials are also provided on how to write scripts using EEGLAB functions.

EEGLAB provides information on how to use the software for iEEG data, detailing support for reading multiple iEEG formats (EDF, MEF3 and Neurodata Without Borders). Viewing of recordings and event extraction is also available; however, it references Fieldtrip (discussed below) for further iEEG processing, implying only basic support for iEEG data in EEGLAB.

Support for the described common preprocessing steps:

- *Downsampling*: Resampling of the loaded data can be performed either through a GUI or command. Resampling methods from MATLAB are used by EEGLAB.
- *Drift correction*: For drift correction, EEGLAB suggests using a highpass filter. A FIR filter is used by default and the order of the filter can be customized.
- *PLI attenuation*: Besides common bandpass filtering, the Cleanline algorithm, previously described, can also be used if the corresponding plugin is installed.
- *Re-referencing*: Not available for iEEG data.

- *Channel localization*: Not available for iEEG data.
- *Artifact detection*: Detection of artifacts is available through the Artifact Subspace Reconstruction (ASR) method, in which artifacts are detected by first learning about the properties of a given clean portion of data in a subspace calculated through principal component analysis (PCA) and then identifying these artifacts are segments with deviant statistical properties in this reduced space [19]. ASR, similar to Autoreject, hasn't been validated in iEEG. Detection of periods based on their standard deviation is also available.

Overall, EEGLAB can be used to perform some preprocessing steps that are common between iEEG and scalp EEG. Nevertheless, important features such as re-referencing and channel localization are not available for iEEG data, requiring custom scripts or the use of a combination of tools to build a complete preprocessing workflow, limiting its reproducibility and scalability. Furthermore, proper documentation of the functions is not available, lowering its usability when scripting.

- **Fieldtrip:**

Fieldtrip, similar to Brainstorm and EEGLAB, is another MATLAB toolbox designed for processing and analysis of MEG, EEG and iEEG recordings [114]. Unlike EEGLAB and Brainstorm, Fieldtrip doesn't support usage without the MATLAB desktop application; therefore, a paid license is required for its usage. Another difference from the previously described tools is that Fieldtrip is not GUI based, but all of the documentation and tutorials are based on scripting.

According to the online documentation, several BIDS compliant iEEG file formats are supported, including EDF, EEGLAB file format and BrainVision Data Format. However, the official documentation outlines that the function used to read EDF+ files has been found to be incompatible with some acquisition systems such as EGI 'Net Station'.

Support for the described common preprocessing steps:

- *Downsampling*: All of the preprocessing steps are computed using the preprocessing function `ft_preprocessing`. The parameter *resamplefs* along with the prior function can be used to resample the given data. The function `ft_resampleddata` can also be utilized for the same purpose. Fieldtrip applies an anti-aliasing low pass filter along with the downsampling. Three methods are available (*resample*, *downsample* and *decimate*) but no further information about their differences was found.
- *Drift correction*: High-pass filtering, demeaning and linear detrending are all supported by Fieldtrip. This can be indicated as part of the `ft_preprocessing` configuration or using separate functions. High-pass filter order and type can be customized, supporting both Butterworth IIR and FIR filters.
- *PLI attenuation*: Fieldtrip supports the attenuation of power line noise either using a notch filter or through the function `ft_preproc_dftfilter`. This function supports two methods for PLI attenuation: fitting of sine waves to the specified frequencies (similar to the *removePLI* method previously discussed) or spectrum interpolation for the given frequency band. Alternatively to using this function, the configuration for the `ft_preprocessing` function can also be adjusted accordingly.
- *Re-referencing*: Re-referencing from monopolar to bipolar, laplace and common average reference schemes are supported. To use the automatic grouping offered by Fieldtrip, the channels must be named using an alphanumeric code, where letters represent the group and numbers represent the order of the channel in its group. Channels that are not found to be part of any group are kept as original.
- *Channel localization*: The only information found for channel localization was in the online tutorials, where the region identification per channel can only be performed through the manual identification and labeling of each contact using the acquired postoperative CT image. A parcellation file also has to be given to iden-

tify the regions per channel. No information about any scripting process to achieve this was found.

- *Artifact detection*: The only semi-automatic method for artifact detection supported by Fieldtrip is based on the bandpass filtering of the data, z-transformation of the signals, and further detection of outliers using a given threshold. Predefined functions are given to detect muscle, eyes or heart artifacts; however, these functions are not optimized for iEEG, as Fieldtrip is designed for analysis of different types of electrophysiological recordings.

From all the investigated tools, this is the one that has a better support for scripted preprocessing of iEEG data, as even a previous protocol has been proposed for the standardization of iEEG preprocessing using Fieldtrip [140]. Nevertheless, the following limitations were found for this tool:

- Although the code to preprocess data with Fieldtrip is quite simple, there is no predefined approach to QC the results from each step, which would then require custom scripts to do so, limiting the usability and transparency of the preprocessing.
- Fieldtrip offers support for code execution on distributed computing systems through *qsub*, a tool developed by Fieldtrip developers. However, there is no predefined way to use it in preprocessing workflows built with Fieldtrip, for which coding expertise is required to implement appropriate preprocessing workflows on distributed computing systems. This negatively affects the usability and scalability capacities of the tool for preprocessing workflows.
- Information about how to read complete iEEG directories was not found, with special focus on the BIDS format. This implies that *ad hoc* scripts have to be developed to properly parse large iEEG directories, decreasing the usability of the toolbox. Furthermore, this hinders the transparency of the preprocessing workflows developed as it becomes more challenging to understand and adapt previously shared

preprocessing scripts.

- Although multiple tutorials are offered online, no official documentation on the available functions can be found. Inspecting the code in the Github repository or command line help are then the only ways to get more information about the available functions and parameters when developing a preprocessing workflow, a process in which some users might not feel comfortable.
- As previously mentioned, unlike EEGLAB and Brainstorm, Fieldtrip does not officially support usage on any other platform but MATLAB desktop, constraining its usage to users that can access a license.

Not all the available software tools were included in this review, leaving out some pipelines that could be used for preprocessing. The PREP pipeline is a MATLAB plugin based on EEGLAB for preprocessing of scalp EEG recordings. It doesn't support computation for iEEG data and has been labeled as outdated according to the EEGLAB website. Other pipelines include the Automated Pipeline for Infants Continuous EEG (APICE) and the Harvard Automated Processing Pipeline for Electroencephalography (HAPPE); however, like the Prep Pipeline, they were designed for scalp EEG, not iEEG data [33].

In this section we have described some of the main tools for iEEG analysis and preprocessing. Fieldtrip and Brainstorm stand out as the options that include the most support for iEEG preprocessing; however, poor usability when scripting, lack of predefined QC checkpoints and difficult scalability stand out among their core limitations for iEEG preprocessing. The development of a workflow that improves transparency, reproducibility and scalability then emerges as a promising avenue for improving exchange practices and standardization of iEEG preprocessing.

1.7 Machine learning for iEEG analysis

This section provides an overview of Machine Learning (ML), with focus on its application to iEEG research. The main purpose of this section is to introduce potential machine learning solutions that could help automating the detection of artifacts in iEEG recordings as part of the data preprocessing.

1.7.1 Machine learning overview

Machine learning is the field of study that involves giving computers the ability to learn from data and take decisions without being explicitly programmed to do so. The learning process for machine learning algorithms can be classified between supervised, unsupervised, semi-supervised and reinforced learning [69, 74]. Description of each of them is beyond the scope of this introduction, but the focus will be towards supervised learning, where the algorithm is trained using a desired output so that the machine learns patterns or characteristics to get to this outcome.

Supervised learning can be further classified based on learning tasks or problems to solve between classification and regression. The first refers to a problem where the outcome is a class, for example, defining if a picture shows a dog or a cat; while a regression problem is characterized for predicting a quantitative or continuous outcome, as the prediction of the value of a currency.

Machine learning algorithms can also be classified between traditional/conventional methods and deep learning algorithms. Deep learning refers to the learning process through the transformation of data using successive stacked layers; therefore, the word *deep* in its name. The main idea behind it is that deep learning is able to learn complex and non-linear patterns in the data through transformations to the input data that allow a representation where the task can be achieved more easily.

Most deep learning algorithms are based on Neural Networks or artificial neural networks

(ANNs), which are computational models inspired by the biological neural networks. Over the years, a variety of deep learning algorithms have been developed, such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), generative adversarial networks (GANs), and transformers. Here, we will focus on the first two networks (RNNs and CNNs) as these are the most relevant ones for the presented work.

CNNs are among the most popular deep learning models, inspired by the animal visual cortex, making them particularly effective for computer vision tasks. CNNs operate by identifying specific features within an input image to perform tasks such as image and pattern recognition. It is known that the architecture of CNNs enables a hierarchical feature extraction process: initial layers capture basic elements like edges and colors, which progressively increase in complexity through deeper layers of the network [7].

Convolutional layers are the main component of CNNs, as they allow the network to learn different features present in the input data. In these layers, convolutional operations are performed between a kernel (also referred as filter) and the input, creating *filtered* outputs that highlight relevant features within the data. This represents only the basis of CNNs as there are multiple parts that are not covered here, such as the subsampling process done by the pooling layers and the fully-connected layers used at the end of the network.

Recurrent neural networks (RNNs) are another widely used deep learning architecture, ideal for processing sequential data such as text or audio recordings. The difference of RNNs relies on their ability to process data sequentially while *remembering* or taking in consideration data previously seen, referred to as short-term memory. This is particularly important for tasks where the context is very important, such as text generation and speech recognition. It has also gained popularity in the field of electrophysiology due to the sequential nature of biological signals [5].

Basic RNNs uniquely depend on recurrent connections for the short-term memory. For this reason, simple RNNs suffer from a problem called vanishing gradient, which hinders the learning of long short-term dependencies in the data. Long Short Term Memory networks

(LSTMs) were designed to tackle this problem by the utilization of special memory blocks through the addition of an input gate, an output gate and a forget gate to the architecture of a simple RNN. Gated Recurrent Units (GRU) are a further modification of LSTMs with faster training process due to their lower computational complexity [22, 37].

Outside of deep learning approaches, traditional methods can be found. There are many traditional methods, varying depending on both the type of learning (for example, supervised vs unsupervised) and the tasks (regression or classification). This discussion will be centered on the algorithms employed in this study: logistic regression and tree-based methods.

Logistic regression is a classification algorithm designed to identify a function that distinguishes between classes based on input features [69, 74]. Commonly, a linear boundary is used and the Sigmoid function or logistic is used to calculate the probability of a particular instance to be part of a specific class. In the case of multiclass problems, the softmax function is commonly used for the estimation of probabilities. Logistic regression can be seen as a shallow ANN with a logistic activation function.

Decision trees, known as CART (Classification And Regression Trees), can be applied to both classification and regression tasks. These models partition the feature space using *if-else* decision nodes, effectively categorizing the data. Splitting of the features in classification tasks is usually guided by errors such as the Gini index or entropy metric [69, 74].

Tree-based methods employ ensemble techniques that combine multiple decision trees through either boosting or bagging. Bagging refers to the training of multiple decision trees by randomly sampling the training data with replacement (called *bootstrapping*). The final decision, in the case of classification tasks, is usually taken following the majority vote. Random Forests extend bagging by not only sampling the training data but also sub-sampling features for each tree, which reduces overfitting and improves model accuracy.

Boosting is another ensemble method where trees are grown sequentially, each designed to correct the errors of its predecessor, progressively improving the model's predictive performance. There are different ways to achieve this, including adaptive boosting, where trees are

trained on weighted samples of the dataset depending on the previous tree errors, or gradient boosting, where residual errors from the previous tree are used for training [69, 74].

While each machine learning algorithm has its own set of advantages and limitations, a detailed comparison is beyond the scope of this thesis. However, one important fact to highlight is that the computational cost of deep learning tends to be higher when compared to traditional methods. Therefore, it is important to follow the principle of parsimony—choosing the simplest model capable of performing the desired task—to justify the cost of using deep learning architectures when preferred over simpler approaches [63].

1.7.2 Machine learning applications for iEEG

Machine learning is a tool that has been used to identify complex patterns in biological data, including iEEG recordings. This subsection will focus on a previous review by Mirchi *et al.* about main applications of machine learning for analysis of iEEG datasets [104]. The review specifically targeted articles published before 2020, utilizing data from SEEG or ECoG recordings.

Figure 1.9 shows the classification of the articles performed by the authors, with the majority addressing seizure-related studies, constituting 58% of the total research focus, followed by motor tasks, cognitive tasks and finally sleep staging. For the interest of this work, focus will be given to the first class. For the seizure class, the authors found that most research was focused on seizure prediction or detection, referring to the identification of the ictal/preictal periods [104].

The review also noted that a significant majority of the studies, 63%, employed standard supervised machine learning techniques, followed by 18% using deep learning approaches. Some of the typically used standard algorithms were shallow ANNs and support vector machines (SVM); while for deep learning, RNNs and CNNs were common [104].

Moreover, this article explored some of the most commonly used features as inputs for the models. In the case of seizure detection, time-frequency representations of iEEG data

were commonly utilized; along with other features, such as entropy or kurtosis. For SOZ localization, HFOs and IEDs were used as inputs due to their close relationship to the EZ [73, 50, 1].

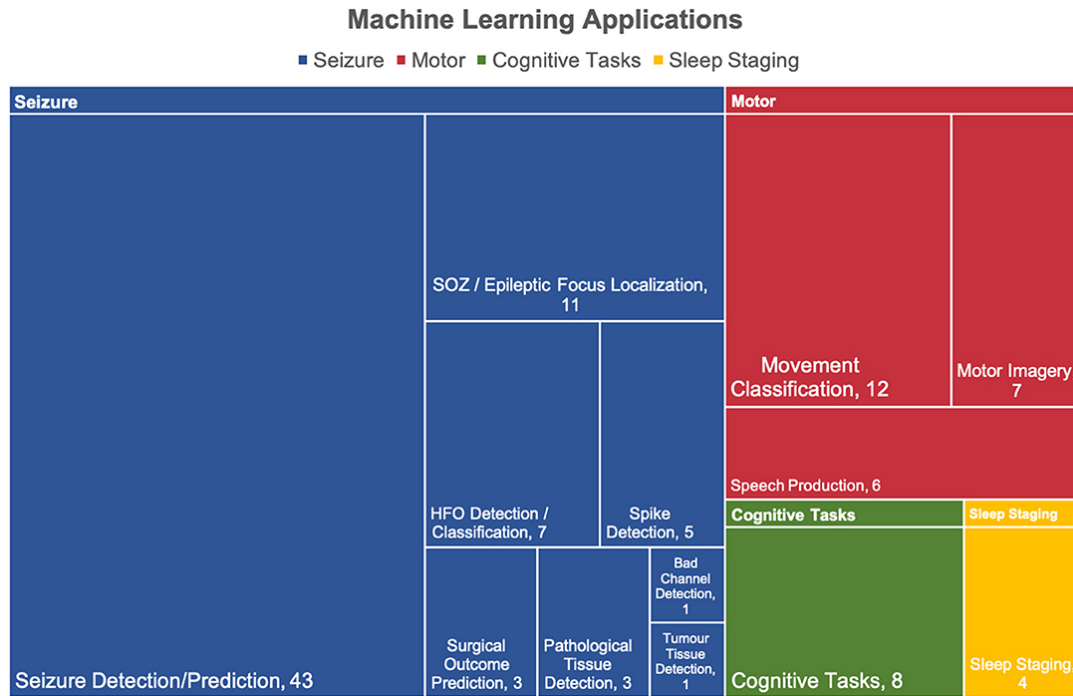


Figure 1.9: Visual representation of the classification of articles reviewed in the work done by Mirchi *et al.*. Reproduced from [104] with permissions under Creative Commons License.

1.7.3 Automatic artifact detection for preprocessing of iEEG recordings

As previously mentioned, iEEG recordings are commonly affected by two types of artifacts, pathological events and non-cerebral artifacts. It was previously believed that iEEG was not affected by non-cerebral artifacts due to its high SNR compared to scalp EEG; however, artifacts such as eye blinks and muscle movement have been now reported during iEEG recording. As iEEG is commonly recorded from brains with epilepsy, pathological events such as spikes and pathological oscillations are also present, which are considered artifacts when analyzing the underlying baseline (non-pathological) activity. Despite the importance of detecting these pathological and noise events in iEEG as part of the preprocessing workflow, there is not a

recommended approach to conduct this artifact identification, hindering its standardization as a preprocessing step for iEEG.

In the literature, there have been several studies on automatic detection of specific interictal pathologies in iEEG. These have mainly focused on either IEDs or HFOs as they are both known biomarkers for epilepsy [73, 50, 1]. IEDs can appear as spikes, sharp waves, spike-and-slow wave complex, sharp-and-slow-wave complex or multiple spike-and-slow wave complexes [47, 1, 83]; while HFOs refers to the presence of ripples (80–250 Hz) or fast ripples (>250 Hz) [80, 52].

Automatic detection of HFOs in scalp EEG and iEEG has been proposed in the past, mostly inspired by the difficulties on visual detection due to the HFOs short duration and low SNR [161, 85]. It has been previously reported that the annotations of HFOs for a 10 minutes recording with 10 channels can take up to 10 hours [161]. Traditional approaches for detection rely on thresholding techniques, where features such as root mean square (RMS) amplitude of a filtered version of the signals or the envelope of the signals obtained through Hilbert transform have been used [91, 158]. More recently, machine learning techniques have been proposed to improve classification of HFOs, which include traditional machine learning techniques such as SVM and LDA, and deep learning techniques such as convolutional neural networks (CNN) or Long Short-Term Memory cells (LSTM), which will be further cover in the next subsection [100, 91].

One of the main challenges related to HFOs detection is the existence of both physiological and pathological HFOs. As mentioned in the first chapter, the properties to differentiate physiological from pathological HFOs have been found to vary across brain regions, which further hinders their classification [80, 52]. Automatic HFOs detectors have also been found to have a low specificity or high false positive rate, mainly due to the misclassification of artifacts, such as false oscillations due to filtering of spikes [134, 105, 151]. Furthermore, inter-rater agreement on classification has been reported between 40% and 80%, which has a direct impact on training and evaluation of detection models [110, 139].

In the case of IEDs, traditional approaches have also been proposed, including template matching and thresholding [59, 127]. Most of these approaches are commonly affected by artifacts with similar morphology such as muscle artifacts or eye blinks, and they fail to capture the high degree of variability in IEDs morphology across subjects [1, 59]. Machine learning techniques have been proposed to solve these limitations, including deep learning approaches like LSTMs and CNNs, and traditional methods such as SVM [1, 59]. However, the lack of cross-institutional large databases for training and validation is currently one of the primary limitations of machine learning solutions [1]. Low inter-rater agreement has also been observed for the identification of IEDs, making the development of standard data sets even more difficult. [13, 10].

Detection of both HFOs and IEDs tend to suffer from low sensitivity due to the misclassification of artifacts as pathological events. There have been previous approaches to try to distinguish artifacts from pathological HFOs and IEDs; however, they lack either evaluation on external datasets or insufficient generalizability reported [127, 162].

Most of these automatic detection algorithms focus on detecting specific types of pathological events due to their clinical importance. For preprocessing of iEEG, pathological events are considered as artifacts in studies focusing on the underlying physiological activity [102]. In these cases, it is mostly relevant to classify the data between physiological or artifacts, which could potentially be pathological or non-cerebral artifacts like eye blinks. This artifact detection task has not been so widely approached in literature, discussed only by two authors based on the conducted literature review [111, 99].

The first study published focuses on the implementation of convolutional neural networks for the classification of interictal segments of three seconds from two institutions between pathological, non-cerebral artifact or baseline activity [111]. Bandpass envelopes were computed using Hilbert transform to construct the input image for the following bands: 20–100 Hz, 80–250 Hz, 200–600 Hz and 500–900 Hz. According to the paper, generalizability to an external dataset was good, with an average F1-score of 0.81 [111]. However, this work presented

two important limitations. First, no open-source model or code was referenced in the publication, which makes it impossible to use or validate the results. Secondly, the sampling rate of the recordings used for training was 5 kHz, significantly higher than the average sampling rates used in clinical practice for iEEG (typically between 1 kHz and 3 kHz [118]), which preclude its usage on external data recorded on a lower sampling rate.

A second study was released afterwards, which included the addition of a third small dataset, used for external testing [99]. A random forest algorithm with customized 68 features extracted from the data was used in this case. This work downsampled the data to 1200 Hz before training since their dataset was recorded at that particular frequency, highlighted the previously mentioned limitation of the first study. However, the classifier performance on the external dataset reported a very low precision for the non-cerebral artifacts class, showing a possible large amount of false positives for this class. Based on the low sensitivity of the baseline class, it can be assumed that this classifier overestimated many baseline clips as artifacts [99].

The first study from Nejedly *et al.* used one of their two datasets for training and validation and the other one for testing [111]. On the other hand, the work from Long *et al.* included a third dataset for testing, reason why the original two datasets from Nejedly *et al.* were only used for training and validation [99]. This hinders a direct comparison between the two algorithms, which is important to define whether deep learning actually provides a significant advantage against traditional machine learning. No comparison with simpler models was given in either of the two papers, fundamental to justify the use of a more complex solution based on the parsimony principle for machine learning.

1.8 Thesis Outline

1.8.1 Problem statement

Preprocessing is fundamental in the field of iEEG, commonly required before inspection and further analysis of the data for research purposes, as shown by the conducted literature review. Preprocessing can be decomposed into multiple steps, including but not limited to: preprocessing of the signals through conventional signal processing methods such as filtering, identification of segments of interests (for example, HFOs, IEDs or segments free of artifacts) and curation of demographic data for each subject.

Typically, these preprocessing steps are carried out in *ad hoc*, standalone, and predominantly manual workflows as there is no standardized approach used [140], which complicates the reproducibility or continuation of research findings. Moreover, implementing preprocessing workflows with existing tools requires substantial coding expertise, constituting a challenge for those with limited software computing experience. These commonly used tools do not integrate quality-control checkpoints to validate the computed preprocessing steps, limiting the transparency of the workflows. The large amount of data recorded from iEEG electrodes makes this process even more difficult, as monitoring over several days with multiple electrodes can result in large files that are difficult to manipulate without the appropriate expertise, highlighting the need for a tool that facilitates data extraction and preparation prior to further analysis. In this thesis, the development of an open source preprocessing workflow for iEEG-BIDS datasets is proposed, trying to improve usability, transparency, reproducibility, scalability and portability compared to existing options.

As previously discussed, artifact detection, including identification of non-cerebral noise and pathological events, has been proposed as an important preprocessing step for iEEG recordings and it was found to be commonly used in iEEG research literature. Nevertheless, a standardized and broadly validated method to achieve this hasn't been defined yet, hindering its incorporation to the proposed preprocessing workflow. In this work, we investigate in detail

this preprocessing step by developing and comparing different automatic artifact detectors for interictal iEEG data in order to evaluate their performance and generalizability, and to gain insight into the main challenges in this field for the standardization of artifact detection for iEEG.

1.8.2 Objectives

The objectives chosen for this thesis work are presented below:

General objective:

- To advance the reproducibility and reliability of epilepsy research through the development of a standardized preprocessing workflow for iEEG recordings and the application of advanced machine learning techniques for the automated detection and classification of artifacts and pathologies in interictal iEEG data.

Specific objectives:

- To design and implement a workflow that integrates essential preprocessing steps for BIDS-iEEG datasets such as filtering and channel re-referencing, aiming for improvements in usability, transparency, reproducibility, scalability and portability compared to existing solutions.
- To evaluate the developed preprocessing workflow by comparing its results with existing tools, aiming for improvements in usability and validation of each preprocessing step.
- To evaluate multiple machine learning algorithms and conduct a comparative analysis of their performance in detecting artifacts and pathological events within interictal iEEG recordings, including a comparison between deep learning and traditional machine learning approaches.
- To investigate the generalizability of machine learning solutions in detecting artifacts and pathological events within interictal iEEG recordings by evaluating the performance of

the models in external data.

- To validate the practical utility of machine learning models by engaging in a qualitative assessment with clinical experts to review their reliability in identifying artifacts and pathological events in continuous interictal iEEG recordings.

The primary goal is to publish all developed code and tools related to the preprocessing workflow and machine learning investigations, aiming to support and facilitate further research and development in the field. However, it must be highlighted that the development and maintenance of a comprehensive software tool takes efforts beyond the scope of this work. This project does not aim to outperform existing machine learning models but rather to explore a variety of features and model types. It emphasizes evaluating the generalizability of these models to enhance our understanding and development of more effective artifact detectors for iEEG.

Chapter 2

Methodology

As the first part of this project, a software workflow to preprocess iEEG datasets was developed. In the next sections, the steps followed for the design and development of this workflow will be covered, detailing all of the different components of this tool.

2.1 Development and design of a preprocessing workflow

The initial step in the development of the preprocessing workflow consisted of a literature review focused towards papers describing recommendations for iEEG preprocessing. This includes mostly review papers covering either the state-of-art of iEEG or the usefulness of specific preprocessing steps [102, 140, 98, 101, 97]. Other relevant papers related to preprocessing for scalp EEG were also reviewed [16, 34]. The most common preprocessing steps found were previously described in the subsection *Overview of preprocessing steps for iEEG* in the Introduction chapter. This section will describe the steps followed for the design and development of the preprocessing workflow.

The proposed SW workflow is based on Python and it was named *iEEGPrep*. Currently, the totality of the workflow can only be executed using SEEG recordings, temporarily excluding its full usage for ECoG data. Re-referencing and channel identification, further described below, are the steps that currently only work for SEEG data. SEEG was chosen as the first type

of iEEG electrodes to support due to their increasing use in clinical practice, as previously discussed. Snakebids was used as the base of the workflow, which is a Python package that extends Snakemake to enable the creation of pipelines for the processing of BIDS datasets [122]. Snakemake is a workflow management system designed to create reproducible and scalable data analyses, as described by their official documentation [108]. The creation of workflows using Snakemake is based on the definition of rules that specify how to generate output files based on the given input files, and these are defined via a human readable Python-based language. These rules can be executed in parallel and are easily scaled to large computing environments such as computing clusters [108]. Snakebids is another Python package that extends Snakemake capabilities by simplifying the creation of workflows designed to process BIDS datasets [122]. Workflows created using Snakemake and Snakebids are reproducible with different BIDS datasets and in different computing environments, particularly useful for the creation of the proposed preprocessing workflow.

A visual overview of the implemented workflow is shown in Figure 2.1. All of the previously reviewed preprocessing steps were considered; however, artifact detection was not included in this first iteration of the tool, as no open-source validated implementation for iEEG data was found. Although epoching or epoch extraction was not discussed in the previous section, it is also a common step used at the beginning of the preprocessing to extract segments of interest for the analysis; for example, a clip during a cognitive task [102]. Therefore, it was included as a first potential step in the workflow.

This pipeline was designed in a flexible way, such that the preprocessing steps to be run can be customized. For example, if desired, the data can go only through downsampling, re-referencing and PLI attenuation, without the need to run the others. However, the order of the steps was fixed; as an example, it is not possible to run re-referencing before downsampling. Besides the preprocessing of the data, a QC component was also added to the workflow, which not only informs the user about the compliance of the input directory with the BIDS standard but also produces outputs for each preprocessing step to ease the manual inspection of the

results without the need to code any further in-house validation, as shown on the right part of Figure 2.1.

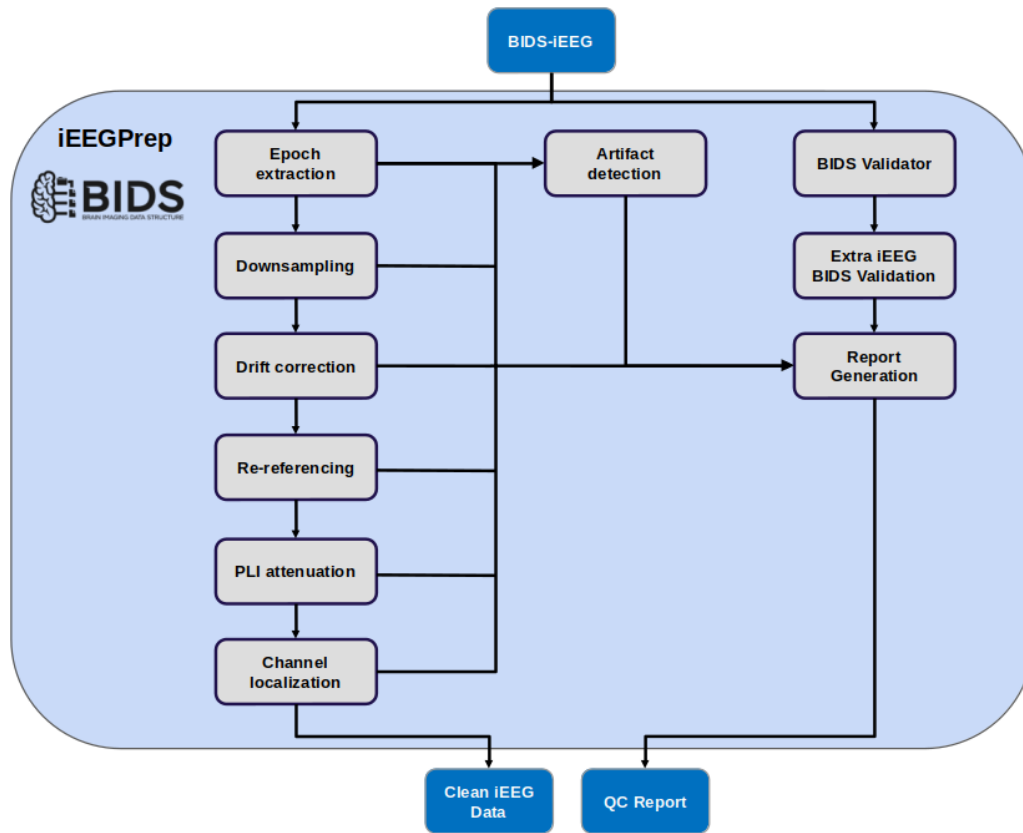


Figure 2.1: Overview of the preprocessing workflow implemented. A BIDS-iEEG directory is received as the input, which is processed by running different preprocessing steps such as downsampling and re-referencing. Quality-control (QC) steps are also executed, including a validation of the BIDS directory and the generation of a report at the end. This report, along with the clean preprocessed files are the main outputs of the workflow.

Next, a description of each of the different parts of the workflow will be presented, detailing the methods employed for each step along with information related to their implementation, like SW packages used.

2.1.1 Input

As aforementioned, the input for the workflow has to be a BIDS-iEEG directory. Currently only EDF format is supported for reading and writing of SEEG data, part of the two recommended

by the BIDS specification. This format was chosen over the BrainVision Core Data Format as the local available SEEG data is stored in EDF, and being able to process these local data was the initial priority for the project.

The input data are parsed using Snakebids, which utilizes Snakemake to allow the directed acyclic graph (DAG) of jobs that need to be executed based on the input and desired output. Snakebids relies on Pybids, another Python package, to parse the BIDS directory as the workflow input [159].

2.1.2 Epoch extraction

The initial step following input parsing is the extraction of epochs or clips of interest for the user from the EDF files. Epoch extraction is configured using two parameters: the start point and the duration of the epoch.

Two alternatives are supported for defining the start point. Users may utilize annotations or events within the EDF file to determine the start point. These events can be easily written as well using packages like *MNE Python* or *Fieldtrip* in MATLAB [64, 92, 114]. Alternatively, the start point can be identified based on channel indices, enabling the extraction of clips starting from a specified time-point for every channel. The duration of the epoch can be specified either by the number of samples to extract or by the event's duration in seconds.

Then, the algorithm searches for all the corresponding events in the given EDF file and generates new EDF files based on the number of events identified. *Pyedfib*, another Python package, is being used as the primary tool for the input/output (I/O) of EDF files throughout the workflow [109].

2.1.3 Downsampling

The implemented downsampling process is a wrapper around MNE implementation with two modifications:

- MNE requires to load the whole data to be able to resample it, being very memory consuming when managing big files, which is not uncommon in SEEG recordings. My wrapper uses the *multiprocessing* Python package to process each of the channels independently in parallel, constrained by the amount of processes available as indicated by the user, minimizing any potential memory issues.
- Before preprocessing, the channels present in the EDF file and the *_channels.tsv* file are compared and only those present in both are preserved. This speeds up the processing by discarding any part of the data that is not relevant.

The parameters used for the MNE method (`mne.io.Raw.resample()`) were kept as default. This method uses a reflection of the signal mirrored on the first and last value of it, followed by zeros, to deal with the potential edge effects of the resampling. By default, a rectangular or Dirichlet window is used to taper the Fourier spectrum before the padding, which intends to alleviate any potential sharp transitions in the signal after resampling (known as ringing).

Downsampling was chosen as the first step as, if executed, it alleviates the computational cost of the following steps, since it reduces the amount of data stored compared to the original file.

2.1.4 Drift correction

Three methods for drift correction were implemented:

- **High-pass filtering:** As high-pass filtering is the most common method for drift correction according to the literature and it is also the recommended option for long-duration data, it was defined as the default method for this preprocessing step. The implementation of high-pass filtering was recovered from the EEG-ASR-Python repository, also based on the Clean Rawdata plug-in for EEGLab in MATLAB.

Briefly, this implementation designs a high-pass filter using the Kaiser method, built based on a given transition band and a desired attenuation. Then, the Scipy function `scipy.signal.filtfilt()` is used to filter the signal, which applies the filter twice (once forward and once backwards) to avoid phase changes to the input data [155]. This function also deals with edge effects by applying padding to the input signal.

- **Demeaning:** As previously described, demeaning just refers to the subtraction of the mean for each channel. Numpy package was used to achieve this [68].
- **Linear detrending:** Although from the reviewed studies, no paper was found to use detrending, it has been described as an alternative for drift correction [102]; therefore, it is also supported in iEEGPrep. Scipy implementation for detrending was utilized through the function `scipy.signal.detrend()`. Currently, there is no supported method for polynomial detrending.

2.1.5 Re-referencing

Although there are multiple re-referencing options, priority was given to the bipolar scheme since it is one of the most commonly used schemes as described in literature. As of now, only re-referencing from monopolar (all signals referenced to the same electrode, either external or internal) to bipolar is supported. As suggested by Mercier *et al.*, the bipolar referencing was computed from the most inner to the most external electrode contact (e1-e2, e2-e3, etc.), equivalent to the methods followed in the common toolbox Brainstorm and Fieldtrip [114, 147]. Reading and writing of the EDF files was done through the Pyedflib library.

2.1.6 Power Line Interference (PLI) attenuation

Four methods are supported for PLI attenuation:

- **Notch filtering:** as it is the most common approach to reduce PLI, notch filtering is supported through a finite impulse response (FIR) filter designed with the window method

through the function `scipy.signal.firwin()`.

- **Cleanline:** The Cleanline algorithm, which was initially developed for EEGLAB in MATLAB, was converted in-house to Python code from the original repository. In this repository, two implementations of the algorithm can be found. The new implementation was the one chosen for the re-implementation in Python.
- **Zapline:** Similarly to Cleanline, the original Zapline algorithm was written in MATLAB and can be used as an independent tool or as a plug-in for EEGLAB [29, 89]. The original implementation of Zapline was then converted to Python code for its usage in iEEGPrep.
- **removePLI:** The algorithm was re-written in Python based on the original source code in Github.

Bipolar re-referencing is known to remove noise with broad spatial distribution, i.e., noise shared amount channels [152, 103]. Therefore, it was decided to first re-reference to remove any potential PLI shared across channels and then run PLI attenuation to correct any other line noise that was not properly attenuated with the re-referencing.

Previous studies have also shown that some PLI attenuation algorithms decrease their performance if high-pass filtering is applied after the algorithm instead of before [16]. As no significant information about the order of re-referencing and drift correction was found, the same order followed in previous scalp EEG pipelines was used [16, 54, 11, 60].

2.1.7 Channel localization

To be able to define the region of the brain where each channel is located, three pieces of data are required:

1. A pre-operative MRI to compute the segmentation of the individual brain. As previously described, this can be performed either through an individual space segmentation or a registration to a common space where an atlas with multiple anatomical labels is defined.

2. A post-operative CT to define the coordinates of each contact electrode. The definition of these coordinates is outside of the scope of the workflow being developed so it was assumed that they were already available, as also required by the iEEG-BIDS standard.
3. Information about the reference scheme used to define the channels since the previous coordinates are usually given per channel.

There are two options that could have been followed to get the estimated brain region for each channel: receive the MRI and offer a way to segment it or receive a precomputed segmentation from the start. The main disadvantage of the first approach is that it would constrain the segmentation to the options supported by the workflow. Another disadvantage is that MRI images are not always available, specially in open SEEG datasets such as the HUP iEEG Epilepsy dataset (ds004100, accessible from OpenNeuro, an online platform for sharing neuroscience BIDS-compliant data).

The main limitation of the second approach is that it forces the user to use another tool beforehand to get the segmentation file. However, there are multiple tools to achieve this, such as SynthSeg or FreeSurfer-volumetric approach [17]. Another limitation is the constraint of the input segmentation file to the supported file formats and also the availability of a table in a supported format that specifies the labels for the segmentation (since the segmentation file only contains integers assigned to each brain region that have to be map to a label usually indicated with a table). This second approach was then chosen for iEEGPrep as it is useful when dealing with datasets with predefined coordinates or those with MRI available.

There are two approaches to define brain segmentation data. Volumetric-based data partition the 3D volume of the brain into different regions and these classification is stored in a voxel level, while surface-based data use meshes to represent the boundaries of each segmented brain structure [95]. Volumetric-based data are often used when the entire 3D brain volume is segmented, specially for subcortical regions since surface-based data are more common for cortical segmentation [95]. Both methods present advantages and disadvantages outside of the scope of this thesis, but it is important to mention as currently iEEGPrep only supports usage

of volumetric-based data, chosen based on the availability of multiple whole-brain atlases and tools for generating volumetric-based segmentation files.

Currently, to allow channel localization, iEEGPrep must receive a NIfTI file with the segmentation and the *_electrodes.tsv* file must have the (x,y,z) coordinates of each electrode contact defined in the same space as the parcellation. In other words, if it is desired to use an atlas for parcellation, the user must guarantee that the coordinates are in the common space used by the atlas; if using an individual subject segmentation, then the coordinates must be in the same space as the resulting segmentation image.

As previously described, only monopolar to bipolar re-referencing is supported; therefore, channel localization was only designed for these two reference schemes. Two options are provided for the localization: volumetric localization or point-based localization.

- *Point-based localization:* For the unipolar case, the coordinate given in the *_electrodes.tsv* file will be used to find the nearest voxel in the NIfTI file containing the parcellation. The region containing this voxel will be assigned to the channel. On the other hand, in case of re-referencing to bipolar, the average coordinate of the two contacts composing the bipolar channel will be used to find the respective voxel.

The disadvantage of this approach is that it could be misleading for monopolar channels if the contact is very close to the boundary of multiple regions. For bipolar channels, as mentioned before, it could not properly represent the region where the captured signals are coming from, especially if the two contacts forming the channel are in different regions.

- *Volumetric localization:* Designed to try to overcome the limitations from point-based localization. Briefly, the idea is to create a volume around the contacts part of the channel that represents the region where the signals are more likely to be coming from; i.e., try to represent the potential field captured by the contacts by using a volume mask.

As shown in Figure 2.2, bipolar scheme captures activity around the two contacts form-

ing the channel and volume-conducted activity in between them [102]. On the other hand, monopolar schemes tend to capture both local and global activity compared to the bipolar reference.

Based on this information, for bipolar channels a cylindrical mask is created around the two contact coordinates, represented as the orange cylinder on Figure 2.2. This mask was generated by first creating a line between the voxels closest to the coordinates of the two electrode contacts in the bipolar channel using the Bresenham's Algorithm and dilating it to have a diameter of 1 mm. The function `skimage.morphology.dilation()` from the Python package `scikit-image` was used for the dilation [154]. The diameter size was chosen only based on the common diameter range for each contact, usually varying between 0.8 to 2 mm [102], as no estimation of the potential field was found.

For the monopolar case, as a simplification of the actual potential field, only the local field around the contact recording was considered. Therefore, a sphere of 1 mm is used around the given coordinate for each contact.

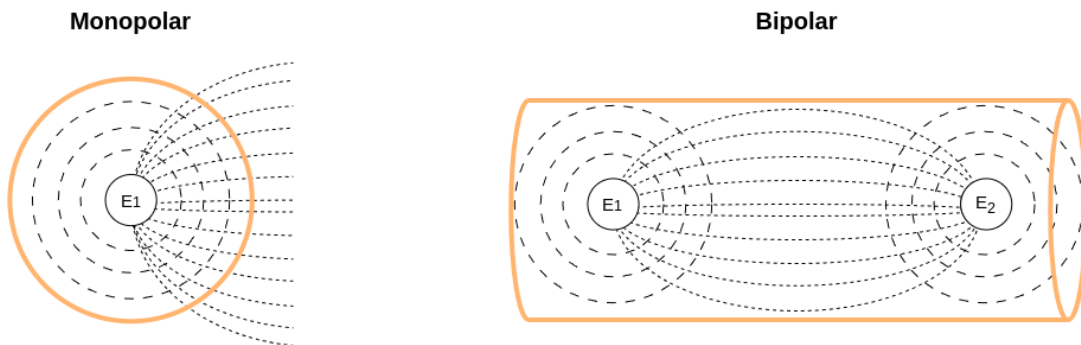


Figure 2.2: Graphical representation of a bipolar (left) and monopolar (right) potential fields in black and the three-dimensional figure used to represent it in the SEEG workflow (orange). Adapted from [146].

Even though this volumetric estimation can represent somehow better the potential field captured by each channel compared to the point-based approach, it is still an estimation and must be interpreted with caution. Special caution has to be taken for monopolar cases, where there is a big influence from the volume-conduction that is not considered

by either the point-based or the volumetric approach.

2.1.8 BIDS Validation

As part of the quality control results produced by iEEGPrep, a validation of the input BIDS directory provided by the user is performed. This validation is composed of two steps. The first is the execution of the official BIDS-Validator tool, which checks if a directory is compliant with the BIDS standard [18]. The flag `--ignoreSubjectConsistency` is used by default to speed up the execution of this step, avoiding checking if any given file for one subject is present in all the other subjects.

It was noted that not all aspects of the iEEG-BIDS standard were checked by the official BIDS-Validator tool. Therefore, the following requirements are checked as a second layer of the validation through in-house scripts using Python:

- All channels present in the *_channels.tsv* file must also be present in the corresponding EDF file.
- The physical unit indicated for each channel in the *_channels.tsv* file must be the same as the one present in the corresponding EDF file.
- The filters applied to each channel must match between the *_channels.tsv* and the EDF files.
- The sampling rate indicated in the *_channels.tsv*, the EDF file and the *_ieeg.json* sidecard must match.
- Electrodes coordinates in the *_electrodes.tsv* file must be numeric.
- All contacts part of the channels in the *_channels.tsv* must be present in the *_electrodes.tsv*.

- The number of channels per type of channel must match between the *_channels.tsv* and the corresponding EDF file. Currently the types checked are: "EEG", "EOG", "ECG", "EMG", "ECOG", "SEEG".
- The length of the file indicated in the sidecar *_ieeg.json* must match the actual length of the recording stored in the EDF file.

2.1.9 Outputs

After all of the steps are executed, their outputs are used to compute specific metrics that are useful to quality-control the workflow; thus, giving the user the opportunity of checking if the obtained results are the desired ones without the need of creating any other scripts. These results are compiled on a HTML document to ease the visualization in any Internet browser. One report is created for every EDF file produced. A description of the QC information for each step in the workflow is provided below:

- *Raw information:* Information about the EDF file being analyzed, indicating the name of the file, measurement date, sampling frequency, number of signals in the file and the duration.
- *BIDS Validation Results:* The official bids-validator outputs are attached in the HTML, along with the results from the in-house iEEG-BIDS validator, showing any potential structural issues in the input dataset.
- *Epoching:* Details about the number of epochs of interest along with their duration is displayed.
- *Downsampling:* The original, target and final sampling rates are shown along with any channels discarded during the step (due to their absence in the *_channels.tsv* file. Size comparison between the original and downsampled files is also detailed.

PSD comparison between input and output channels is displayed as a way to quality-control the downsampled signals. Welch’s method is used to compute the PSD and Plotly, a Python package for plotting, is used to visualize it, allowing the interaction with the figure (for example, zooming, hiding/showing channels, etc.) [72].

- *Drift correction*: First, information about the method used, along with the corresponding parameters and any discarded channels are shown. As QC outputs, a table comparing the original and new mean of each channel is displayed, along with a PSD comparison (same as in the previous step). If a high-pass filter is used, the time and frequency response of the filter is also given.
- *Re-referencing*: If any channels were discarded, they are displayed. A table showing the combination of monopolar channels to form the bipolar equivalents is also displayed.
- *PLI attenuation*: Information about the method used is attached along with a PSD comparison between before and after the step for each channel.
- *Region identification*: The main output for this final step is the amount of channels per region found. This is summarized in a table for the user.

An example of the HTML report generated by the workflow can be found in the Appendix A.

2.1.10 Execution of the workflow

As a large amount of coding is needed to compute each of the preprocessing steps, they were first compiled in a Python package. Poetry was used to manage the dependencies needed for the installation of the package [40].

The workflow can be then executed through the command line with the command:

```
run.py [-h] [--participant_label PARTICIPANT_LABEL
[PARTICIPANT_LABEL ...]] [--run-all] [--epoch] [--downsample]
```



```
[--filter] [--rereference] [--PLI_rej] [--regions -id]
bids_dir output_dir participant [--cores CORES]
```

This command only shows the most important flags for the workflow, please refer to Snakebids official documentation for more information. As shown in the command, the workflow is designed to run only on a participant level as all the outputs are produced in a subject level, no group level analysis is currently being conducted. A description of the parts of the command is given below:

- The fixed arguments are then `bids_dir` and `output_dir`, which should be replaced with the path to the input iEEG-BIDS directory and the output directory.
- The last part (`participant`) is not a flag and should be kept as it indicates Snakebids to run the workflow on a subject level.
- The flag `--participant_label PARTICIPANT_LABEL [PARTICIPANT_LABEL ...]` can be used to indicate the subjects to be preprocessed.
- The flag `--cores CORES` can be used to indicate the amount of cores available for proper scalability. If it is not indicated, Snakemake will automatically estimate the number of cores available in the system.
- The flag `--run-all` can be used to indicate the workflow to run all of the preprocessing steps.
- The other flags can be used to run specific parts of the workflow. For example, if only `--epoch` and `--downsample` are used, then only these two steps will be executed by the app. This gives more flexibility to the workflow. The QC steps are always computed, including the report generation. If none of these flags are given, the workflow will run all of the steps by default.

The parameters for each step can be customized using a YAML configuration file. This allows for a more human-readable customization of the preprocessing compared to common

tools like Fieldtrip and EEGLAB, which require some amount of knowledge of coding, making it not accessible for people without this skill [140]. An example of this configuration file is shown in the Appendix B.

2.2 Development of an Automatic iEEG Artifact Detector

The second part of the work presented on this thesis consisted on the development and testing of an automatic detector of artifacts in EEG recordings. As aforementioned, artifact detection was the only preprocessing step left-out of the initial version of the developed workflow due to the lack of a standardized and validated method for iEEG. In this section, I will cover the development and testing of different machine learning models as solutions for this artifact detection problem, centered on the identification of both pathological and noise events in interictal iEEG recordings. Focus was given to interictal data as it is often the part of the data used for analysis of physiological activity using iEEG recordings [51].

2.2.1 Datasets used and data preprocessing

Three datasets were used for training and evaluation of the model. The first two datasets consisted 193118 and 155182 3-seconds iEEG clips from St. Anne’s University Hospital (FNUSA) and Mayo Clinic respectively [112]. The FNUSA dataset was collected from 14 patients diagnosed with DRE during awake resting state using depth electrodes while the Mayo dataset was acquired during the first night after electrode implantation using depth or ECoG electrodes from 25 patients with DRE; in specific, 1227 of the total 155182 clips came from strip electrodes while the rest from depth electrodes. Referencing montage for these datasets was common-average and both were released with a sampling frequency of 5 kHz.

Both datasets were annotated by three expert reviewers into four categories: muscle and machine artifacts (non-cerebral artifacts), baseline activity, pathophysiological events (for example, HFOs and IEDs) and powerline noise [112]. Figure 2.3 shows examples of clips for

each of these classes. These data were used in the previously discussed paper using CNNs for artifact detection [111].

The third dataset used was released along with the study from Long *et al.*, which implemented an iEEG artifact classifier using random forests [99]. This dataset contained eight minutes of recordings from three subjects while completing a visual task. It was collected at the University of Florida Neuromedicine Hospital using depth electrodes and a sampling rate of 1.2 kHz and annotated by two epileptologists. Data were re-referenced to bipolar montage before release [99]. Annotations were used to extract segments of a duration of 3 seconds to match the clips of the first two datasets, as the model was trained on the epochs of this length.

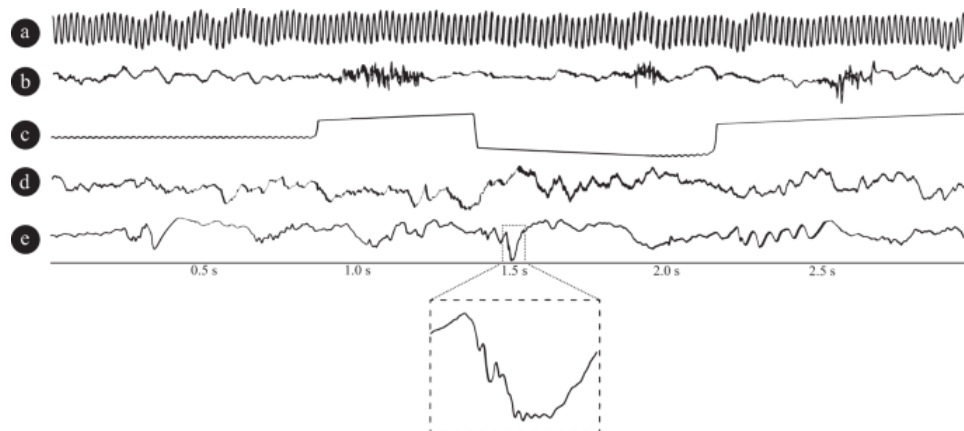


Figure 2.3: Examples for each of the classes present in the FNUSA and Mayo datasets. These correspond to (a) power line noise, (b) muscle artifact, (c) baseline jump artifact, (d) physiological or baseline signal, and (e) pathological signal showing an HFO riding on a spike. Recovered from [112] with permissions under Creative Commons License.

A clinical qualitative assessment of the model was conducted using data from two patients diagnosed with DRE and implanted with depth electrodes as part of the phase II of the presurgical investigation as part of the Western Epilepsy Program at the London Health Science Center (LHSC). Only two patients were evaluated, primarily due to time constraints and availability of clinical fellows for review; however, as the goal of this review was only to explore the behavior of the model and not to actually quantify its performance, this was considered sufficient as a first step to explore the response of the mode. Both patients' recordings were initially sampled at 2048 Hz with an online external reference. It can be appreciated that this sampling

rate is different to the one used in the other datasets; therefore, all datasets were downsampled to 1024 Hz for standardization, as it was required to build the input for the model in the same way as the training data.

A total of four minutes from seven channels were assessed—four from the first subject and three from the second patient from the LHSC database. These four-minute epochs were selected based on previous clinical annotations in the recording files. The annotation ‘*awake trigger*’ was used to extract these epochs as it indicates a period of awake resting state, chosen to match the characteristics of the FNUSA dataset, used for training of the models. Annotations and recordings were also inspected to corroborate that no ictal activity was present in these segments. Table 2.1 shows the most relevant clinical information for these two patients. Based on this clinical information and the preprocessing explained below, signals from channels recording from regions potentially within the SOZ were chosen for qualitative evaluation, as interictal pathological events are more likely to be observed for these channels, which are assumed to be detected by the model.

Subject ID	Sex	Age	Diagnosis			
			Type of epilepsy	Laterality	Lobe localization	Lesional MRI
A	M	51	Focal	Bilateral	Temporal	Yes
B	F	32	Generalized	Bilateral	Temporal/Insular	No

Table 2.1: Summary of relevant clinical characteristics of the patients from the LHSC dataset used for qualitative evaluation of the developed iEEG classifier.

Preprocessing of the three external datasets included downsampling to 1024 Hz, a power of 2, which is optimal for subsequent computational processes. The FNUSA and Mayo datasets were utilized for training, necessitating additional preprocessing to balance the datasets.

Inspection of the FNUSA and Mayo datasets revealed that some subjects had significantly more clips than others, with ratios as high as 21:1. To prevent overfitting, the datasets were modified to balance the number of samples per subject. The median number of samples per subject was calculated, and the number of clips per patient was downsampled to this median, ensuring the class distribution for each individual was maintained. Preserving this distribution

was crucial as the data were used for cross-validation. It is generally advised not to alter the original distribution of validation data to ensure the model’s performance insights are applicable to new datasets with similar distributions. A summary of the total number of clips and their distribution per classification category for all the datasets used during training, validation and testing of the model is presented in Table 2.2.

Classification category	FNUSA	Mayo Clinic	University of Florida Hospital
Non-cerebral artifact	19322	33685	16
Pathology	30630	14969	798
Baseline activity	28712	32214	697
Total	78664	806868	1511

Table 2.2: Summary of the number of clips per category for the three datasets used for training, validation and testing of the proposed iEEG classifier.

Data from LHSC used for qualitative evaluation of the results were preprocessed using the developed tool iEEGPrep. First, the preoperative MRI for each subject was segmented using Synthseg. Preprocessing of the recordings then included: epoch extraction of 4 minutes during awake resting state (marked in the annotations of the EDF files), downsampling to 1024 Hz, drift correction using a high pass filter with transition band from 0.1 Hz to 0.2 Hz, re-referencing to bipolar montage, PLI attenuation using Cleanline and region identification for each channel using the computed parcellation from Sythseg. Channels potentially located inside the SOZ were selected based on this region identification step and the diagnosis for each subject, which included temporal lobe in the SOZ for both of them. For evaluation using the model, the four minutes of continuous data were segmented into 3-second clips with no overlapping between them.

A summary of how each of these datasets was used for this study is presented below:

- The FNUSA and Mayo Clinic datasets were combined and used for training and validation of the evaluated machine learning models. Signals from both of these datasets are referenced to a common average reference (CAR montage), as previously described.
- The third dataset from the University of Florida Neuromedicine Hospital was used for

comparison of the best deep learning and the best traditional machine learning models, focusing on their capability of generalizability to external data. It is important to highlight that these data were published after re-referencing to bipolar montage, which is different to the reference of the training data.

This presents an even bigger challenge for the model as it then implies the evaluation of the generalizability of the model not only to external data but also to other referencing methods.

- The fourth local dataset from LHSC was used for qualitative evaluation of the model when applied to continuous data. As previously mentioned, the main aim of this last step was to explore the behavior of the model and to better understand its performance obtained in the previous step of external testing. To match the dataset used for external testing, a bipolar referencing method was also used for this dataset.

2.2.2 Feature selection

Building on previous studies, my aim was to test both traditional machine learning and deep learning models, requiring the computation of inputs suitable for both approaches. For this reason, inputs for each of the two cases were computed. Following the methodology of Nejedly *et al.*, time-frequency maps were computed using the envelope of bandpass signals derived from the Hilbert transform [111]. Due to the selection of a lower sampling rate for this study, the frequency bands from the previous paper were adjusted to: 20–100 Hz, 80–250 Hz, 200–500 Hz and 450–510 Hz. Specific information on why these bands were chosen was not given in the original paper; however, they are likely to be related to the expected frequencies for ripples and fast ripples for HFOs (80–250 Hz and >250 Hz respectively) [99, 52].

As illustrated in Figure 2.4, the process of computing time-frequency maps using the Hilbert transform involves three key steps. First, the signal is bandpass filtered within each frequency band of interest using a 3rd order Butterworth zero-phase filter, implemented with

the MNE library, similar to how it was previously done [111]. Then, the Hilbert transform is computed on this bandpass signal using the function `scipy.signal.hilbert()` and the envelope can be estimated as the squared absolute value of the result from the function. Finally, the envelope for each frequency band is normalized using the z-score function from Scipy. As shown in Figure 2.4, the resulting image consists of five rows: the top four rows represent the computed results for each frequency band in descending order, and the bottom row displays the z-score version of the original data clip, similar to the methods followed on the study published by Nejedly *et al.* [111].

For the traditional machine learning models implemented in this study, discrete features were computed following methods similar to those described by Long *et al.* [99]. In line with this previous study, Daubechies 4 (db4) wavelets were utilized to decompose the z-scored clips using discrete wavelet transform (DWT) [99].

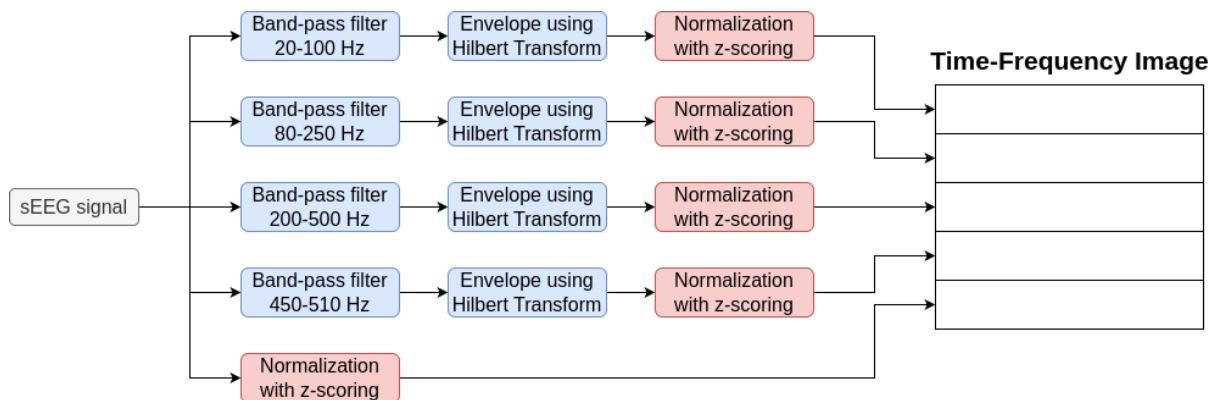


Figure 2.4: Hilbert transform computed for a bandpass version of the original signal on each of the frequency bands of interest. Envelope computed through the Hilbert transform represents one row in the resulting time-frequency map.

Briefly, DWT decomposes a signal through an iterative filtering process. Initially, the original signal is divided into two components using a high-pass and a low-pass filter. The high-pass filtered signal yields the detail coefficients, while the low-pass filtered signal, which retains frequencies up to half the original signal's maximum frequency, is downsampled to half the original sampling rate. The high-pass filtered signal yields the detail coefficients, while the low-pass filtered signal, which retains frequencies up to half the original signal's maximum

frequency, is downsampled to half the original sampling rate [53].

This process of decomposition, including the design of the filters, is executed using wavelets, which are small-duration, wave-like functions. Various types of wavelets exist, such as Morlet, Haar, and Daubechies (db) wavelets. Daubechies wavelets are often a common choice for EEG from their similar spike-wave pattern. Wavelets are subcategorized by their order, which defines the number of samples used and consequently the smoothness of the wavelet [53, 6]. For example, the selected db4 wavelets refer to a Daubechies wavelet of order four. The maximum level of signal decomposition depends on the wavelet, reached when the decomposed signal has fewer samples than the wavelet function.

The function `pywt.wavedec()` from the PyWavelets Python package was used to decompose the iEEG signals using a db4 wavelet [93]. Since no maximum level was specified, the package automatically calculated it, resulting in eight decomposition levels. This function returns the approximation coefficients of the last level of decomposition and the detail coefficients of all levels. The following fourteen features were computed for each array of coefficients:

- 12 statistical features: 5th Percentile, 25th Percentile, 75th Percentile, 95th Percentile, median, mean, standard deviation, variance, root mean square, Kurtosis and skewness. These features help to measure the variability of the data, potentially allowing the identification of events with high variability such as noise or HFOs. Similar features have already been proposed for epilepsy-related applications such as seizure identification [145, 55, 99].
- Shannon entropy, which helps measure the randomness or non-stationarity of a signal and has been previously shown to work as a feature to identify epileptic activity [3, 26, 136, 99].
- Number of times the signal crosses its mean value and the zero amplitude value. These features have been previously proposed for identification of interictal spikes and certain artifacts in scalp EEG, and other scalp EEG applications [125, 4, 144, 99].

Based on the fact that we had eight detail coefficients and one approximation coefficient from the `pywt.wavedec()` results, and that we computed 14 features for each, the input of the discrete models consisted of 126 features.

2.2.3 Selection of hyperparameters

As previously detailed, both deep learning and traditional machine learning models were implemented. Details about the models used are shown below. As it will be explained in the next subsection, k-fold cross-validation (CV) was used to choose the best parameters for each model. The description below shows the models evaluated in this work along with the parameters tested for each model. The complete grid of parameters tested for each model can be found in the Appendix E.

- **Traditional machine learning models:** The best hyperparameters for each model were chosen using grid search cross-validation with 10 folds. Below, the best hyperparameters found are detailed. For the hyperparameters not shown, it can be assumed that they were left as default.

- *Logistic Regression:* This was the simplest model tested, necessary to justify the need of more complex solutions. The function `LogisticRegression()` from `sklearn` was used for this purpose [120]. The hyperparameters tuned were:

- * The inverse regularization strength C , designed to prevent overfitting of the model.
- * Algorithm used to solve the optimization problem (*solver*). Multiple algorithms are available but for this thesis, the Newton's method and SAGA solvers were tested as they are recommended for multinomial problems by `scikit-learn` official documentation.

Furthermore, the L2 penalty (Ridge Regularization) was used.

- *Random Forest*: Inspired by the previous work from Long *et al.*, a random forest was also implemented using the sklearn function `RandomForestClassifier()`.

The hyperparameters tuned were:

- * Number of estimators: referring to the total number of trees in the forest.
- * Maximum depth of each tree in the forest.
- * The maximum numbers of features considered when searching the best split for each tree in the forest (*max_features*).
- * Criterion: the metric used to guide the splits for each tree. Gini impurity and entropy were the metrics considered.

Furthermore, the minimum samples per leaf parameter was set to 0.05.

- *Extreme Gradient Boosting*: Extreme Gradient Boosting was implemented through the function `XGBClassifier()` from the package XGBoost [21], as a tree-based alternative to random forests that can work better for small datasets. The hyperparameters tuned were:

- * Number of estimators: number of boosting rounds.
- * Maximum depth of each tree.
- * The learning rate used by the model.

- **Deep learning models**: For the deep learning (DL) models, no parameters were tuned but at least 3 architectures per model of different complexity were tested and compared (more details about the comparison in the next subsection). Below a brief description of the architectures is provided. Diagrams of each of the architectures implemented can be found in the Appendix F. All of the deep learning models were built and trained using Pytorch’s library [119].

- *CNN*: Each CNN architecture comprised four 2D convolutional layers, followed by data flattening and three linear layers. A skip connection via concatenation,

inspired by DenseNet [71], was implemented between the outputs of the first and third convolutional layers, forming the input for the fourth convolutional layer. This skip connection was introduced to mitigate the rapid overfitting observed during model training. Dropout in the linear layers, 2D max pooling, batch normalization after every convolution were also included to further address the overfitting. The three CNN architectures differed only in the number of filters per convolutional block: the small CNN had half the number of filters of the medium CNN, which in turn had half the number of filters of the large CNN.

- *CNN + LSTM*: The second model combined CNNs and LSTMs to evaluate whether the inclusion of memory cells could better capture temporal dependencies in the data. For this case, six 2D convolutional layers were used. The number of layers was increased to match the model complexity (in terms of parameters) of the previous CNN model, since most linear layers were removed in this configuration. Skip connections through concatenation were implemented between the outputs of the first and third convolutional layers, and between the output of this concatenation and the output of the fifth convolutional block. The output from the CNN was fed into a LSTM, and the LSTM's output was then passed through a linear layer. The same overfitting prevention mechanisms as in the CNN model were used, with an additional dropout rate of 0.5 in the LSTM block.

2.2.4 Training and evaluation

The training and evaluation process comprised four main steps: (1) selecting hyperparameters for each model using cross-validation, (2) comparing similar models based on cross-validation results, (3) comparing the best deep learning model with the best traditional model, and (4) conducting a qualitative clinical evaluation of the best model. Each step is detailed below.

Selection of the best hyperparameters

The hyperparameters tested for each traditional machine learning model implemented (logistic regression, random forest and XGBoosting classifier) are detailed in the Appendix E. Hyperparameter selection was conducted using Grid Search CV, which evaluates different hyperparameter combinations via cross-validation and selects the best combination based on the average scoring across CV folds. The F1-score macro was used as the scoring metric (further discussed below), and group 10-fold cross-validation was the chosen CV method. Unlike standard k-fold CV, group k-fold CV allows separation of subjects between training and validation; i.e., clips coming from a single subject cannot be splitted between the training and the validation sets. A combined version of the curated FNUSA and Mayo datasets was used for both the traditional and DL models in this step.

For the DL models, group 10-fold CV with two repetitions was used; however, Pytorch models are not supported by sklearn's GridSearchCV. As a workaround, data splits for each fold were computed manually, and each architecture was trained and validated per fold. Various metrics were computed for each fold (detailed later in Results), resulting in 20 values per metric for each tested architecture. Then, the different architectures were compared using statistical tests based on the distribution of these values to define the best-performing CNN and CNN+LSTM models. Although cross-validation is not commonly used for deep learning models, it was employed here due to the significant variation observed in results when changing the validation set.

Comparison of similar models

After selection of the best hyperparameters for each traditional model, a comparison between the models was performed to identify the best traditional model. Each algorithm (logistic regression, random forest, and XGBoost classifier) was evaluated using repeated group 10-fold cross-validation with the optimal hyperparameters identified. Five repetitions were used, providing at least 50 samples for each metric computed per model.

Statistical comparisons of the means of the distribution for each metric were conducted for each pair of models. Permutation testing with 100000 resamples and false discovery rate (FDR) correction were used for this purpose, implemented through the functions `scipi.stats.permutation_test` and `fdr_correction()` from `statsmodels` package [135].

A similar comparison was conducted between the two DL models (CNN and CNN+LSTM); however, the models were not retrained with the best hyperparameters as in the previous case. The only hyperparameters tested for the DL models were the number of filters per convolutional block, and they were tested by manually training the architecture for each of the different folds. Therefore, it was not necessary to retrain the best architectures for comparison as these were already previously calculated. The distributions per metric were then compared using the same procedure as in the traditional models.

Deep learning vs traditional machine learning evaluation

Once the best traditional machine learning and DL models were found, we compared them to define whether DL provided any significant improvement compared to non-DL models. These two models were retrained using the whole training dataset (composed by the FNUSA and Mayo datasets) and the third dataset from the University of Florida was used as the external test set.

Clinical insights

As previously mentioned, a qualitative assessment of the final chosen model was conducted in a clinical setting at the University Hospital from the London Health Sciences Centre (LHSC). Four minutes of awake resting data from seven channels in two patients with DRE who underwent iEEG were annotated using the model. These annotations were reviewed by a clinical fellow from the Western Epilepsy Program at LHSC to assess the model's performance. The annotations were presented using the open-source software EDFbrowser, which allows the visualization of the time series in similar way as inspected during clinical evaluation. The clinical

fellow was asked to qualitatively assess her level of agreement with the model outputs and to provide insights into any discrepancies.

2.2.5 Metrics used for evaluation

To compare the different architectures and models, five metrics were used. To explain these metrics, first the following concepts must be understood, which will be explained for a binary classification problem:

- **True positives (TP):** refers to the amount of positive samples that were correctly classified as positive.
- **True negatives (TN):** refers to the amount of negative samples that were correctly classified as negative.
- **False positives (FP):** refers to the amount of negative samples that were incorrectly classified as positive.
- **False negatives (FN):** refers to the amount of positive samples that were incorrectly classified as negative.

Based on these concepts, the metrics used for the comparison of the models and architectures can be found below:

- **Accuracy:** The accuracy is defined as the number samples correctly labeled divided by total number of samples. This metric is not ideal for unbalanced datasets.
- **Macro precision:** In a binary classification, precision is calculated using the formula:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2.1)$$

which can be interpreted as the fraction of samples classified as positives that were actually positives. Macro precision refers as the average precision across classes, so in this

case it would be:

$$\text{Macro-Precision} = \frac{\text{Precision}_{\text{noise}} + \text{Precision}_{\text{phys}} + \text{Precision}_{\text{path}}}{3} \quad (2.2)$$

where $\text{Precision}_{\text{noise}}$, $\text{Precision}_{\text{phys}}$ and $\text{Precision}_{\text{path}}$ correspond to the precision calculated for the noise, physiology or baseline, and pathology classes respectively. Macro precision does not take in consideration the amount of samples per class, which makes it a better option for unbalanced datasets compared to micro precision, where precision is calculated across classes without considering any split.

- **Macro recall:** Recall for a binary classification is defined as:

$$\text{Recall} = \frac{TP}{TP + FN} \quad (2.3)$$

which gives a fraction of how many of all the positive samples were actually classified as positive by the model. Similar to precision, the macro recall was used for the comparisons.

- **Macro F1-score:** The F1-score is a balanced metric between precision and recall and can be calculated using the formula:

$$\text{F1-score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (2.4)$$

The macro version of the F1-score was used for the comparisons done. This was considered as the most important metric as it is the one that better balances the ability of the model to detect positive samples per class while not overestimating any of the classes.

Chapter 3

Results

3.1 Validation of the Preprocessing Workflow

This section will focus on the validation of the designed workflow. Each of the steps was validated sequentially, following the same order as the one used by iEEGPrep, by comparing the outputs of the developed app to the outputs from Fieldtrip for one channel. Fieldtrip was chosen above other options like EEGLAB since the proposed protocol for iEEG analysis written by Stolk *et al.* was developed using this tool, which simplified the scripting process.

For drift-correction and PLI attenuation, comparison between the different supported methods will also be presented. For PLI removal, comparisons between methods using synthetic data will be shown in this chapter as well.

Region identification will be the exception for this validation process as Fieldtrip does not support the localization of channels given the coordinates in advance. Therefore, for this preprocessing step, visual validation using 3D Slicer will be presented.

Finally, the advantages of the developed tool compared to the experience in Fieldtrip will be briefly discussed to highlight some of the advantages of iEEGPrep.

3.1.1 Validation of each preprocessing step

As previously mentioned, the validation of the steps consisted in four main parts:

- All steps will include a visual assessment by comparing iEEGPrep output with Fieldtrip result for one channel. Plots in the time, frequency and time-frequency domain will be shown.
- For all steps, the correlation without lag between each channel preprocessed with iEEGPrep and Fieldtrip was calculated both in time and frequency domain. To calculate the frequency spectrum for each channel, Welch's method was used with a window of 3 s and an overlap of 50%. The distribution of these errors will be shown after each preprocessing step to do a qualitative analysis of these differences.
- For PLI attenuation and drift correction, a statistical comparison between the different methods will be presented. Time, frequency and time-frequency plots will also be shown.
- For PLI attenuation, as the advanced methods were re-implemented from MATLAB code, a validation of each of the algorithms using simulated data was performed. Different metrics were used to compare the two implementations through statistical testing, which will be presented later in this chapter.

Epoch extraction

The epoch from the selected EDF file used for the validation was chosen to be four minutes of data from the event 'awake trigger', present in annotations for the file. In iEEGPrep, this required only changing one line of the configuration file as follows:

```
event: ['awake trigger', 'duration', 240]
```

The obtained epochs from iEEGPrep and Fieldtrip were first confirmed to have the same dimensions. Figure 3.1 shows the first 60 seconds for both epochs for the channel LPIIn02,

which corroborates that iEEGPrep was able to find and extract the desired epoch. This channel will be used for comparisons in the following steps and was chosen since the bipolar channel LPIIn2-3 was noted to have significant power line interference.

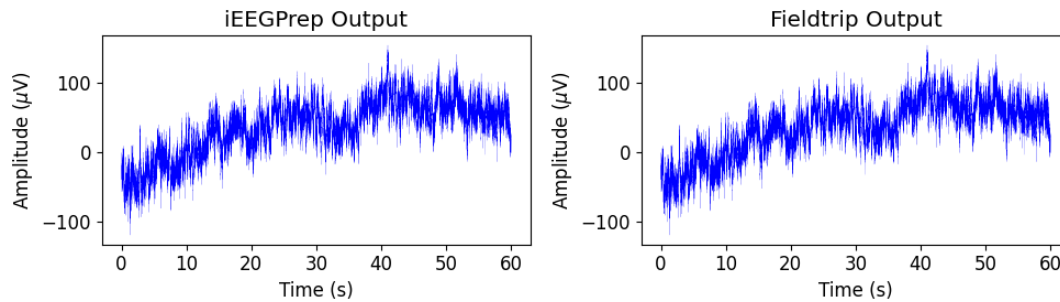


Figure 3.1: Time plots for the first 60 seconds extracted by iEEGPrep and Fieldtrip from the channel LPIIn02.

To do a more quantitative comparison between the outputs, correlation with zero lag in time and frequency domain were calculated for each pair of channels. Zero lag means that the signals are compared using the same time/frequency values; i.e., no lag is added to any of the signals before measuring their similarity using correlation. Therefore, a high correlation with zero lag shows a good similarity between the two signals compared starting from the same time or frequency value. Figure 3.2 shows the distribution of these values for the preprocessed EDF file. It can be seen that both the distribution of correlation values in both the time and frequency domain are around $9.99e-1$, corresponding to a similarity of 99%.

Downsampling

Following the epoch extraction, downsampling from 2048 Hz to 200 Hz was performed, following the downsampling performed in prior iEEG studies [51, 15, 149, 156]. As it can be seen in Figure 3.3, the signals now look less *noisy*; i.e., some high frequency components were removed, as expected from the process. Qualitatively speaking, no significant difference was observed between the signals outputted from iEEGPrep and Fieldtrip, as it can be seen in Figure 3.3. Further qualitative evaluation was performed in the frequency and time-frequency domain, but no differences were observed, as shown in Figures C.2 and C.3 in the Appendix C.

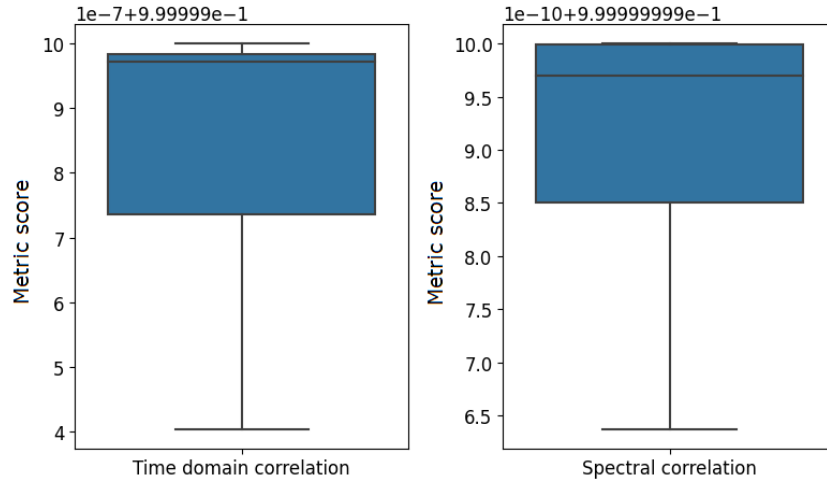


Figure 3.2: Distribution of correlation values with zero lag in time domain (left) and frequency domain (right) between the different pairs of channels present in the extracted epochs using iEEGPrep and Fieldtrip. The amplitude scales for both plots show values very close to $9.99e-1$ (or 0.99), highlighting the strong correlations observed across both time and spectral domains.

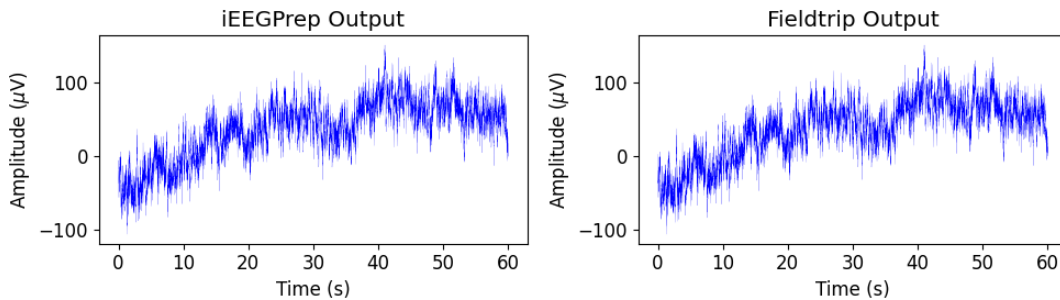


Figure 3.3: Time plots for the first 60 seconds of the downsampled signal from the LPIIn02 channel using iEEGPrep and Fieldtrip.

Similarities between the signals for the whole file were quantified using the correlation value with no lag in both time and frequency domain. As it can be observed from Figure 3.4, both the distribution of values in the time and frequency domain were found to be around $9.99e-1$, showing a high degree of similarity between the signals downsampled using iEEGPrep and Fieldtrip.

Drift correction

As aforementioned, the validation of drift correction was subdivided in two parts.

Comparison against Fieldtrip output using real data

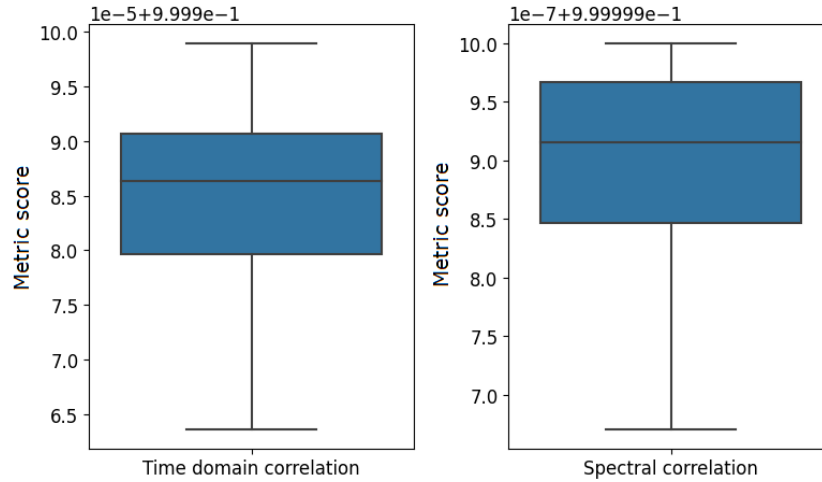


Figure 3.4: Distribution of correlation values with zero lag in time domain (left) and frequency domain (right) between the different pairs of channels present in the downsampled epochs using iEEGPrep and Fieldtrip. The amplitude scales for both plots show values very close to $9.99e-1$ (or 0.99), highlighting the strong correlations observed across both time and spectral domains.

First, drift correction was applied to the resulting data from the previous step (downsampling). The method used for this first part was high-pass filtering, as it is the most common one and the one recommended for long duration data. For the iEEGPrep implementation, the transition band for the FIR filter was chosen to go from 0.1 Hz to 0.2 Hz. In Fieldtrip, the default Butterworth high-pass filter was used with a cut-off frequency of 0.25 Hz.

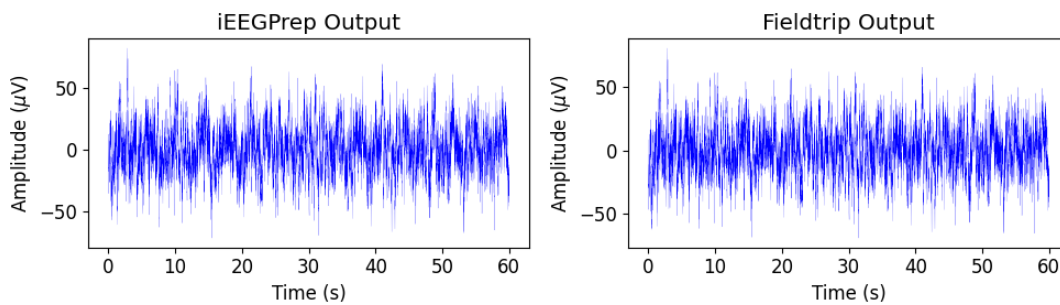


Figure 3.5: Time plots for the first 60 seconds of the detrended signal from the LPI_n02 channel using iEEGPrep and Fieldtrip. It is possible to see that the slow trends previously visible in Figure 3.3 were attenuated after drift correction.

As shown in Figure 3.5, when comparing to the downsampled signal from Figure 3.3, it can be noted that the low frequency trends were attenuated and now the signal is more closed to be

centered at zero in the y axis. One detail that is visible is that the signals do not look completely similar, which differed from the previous preprocessing steps. In the plots from Figure 3.5, this is especially noticeable in the first seconds, as the Fieldtrip output seems to have a peak not visible in the iEEGPrep output. This difference was further confirmed when looking at the correlation values from Figure 3.6, showing lower values compared to the previous steps; nevertheless, a correlation above 98% was observed in the time and frequency domain, still showing a very high degree of similarity between the outputs.

Due to this subtle decrease of the correlation values, statistical comparison of the frequency components was performed, as shown in Figure 3.7. Permutation testing with 100000 resamples and false discovery rate correction was used to check if there was any significant difference between the means of the distributions for each component. Corrected p-values of 0.3396, 0.3243, 0.4694 and 0.4892 were obtained for the 0 Hz, 0.33 Hz, 0.67 Hz and 1 Hz components respectively, showing non-significant differences between the components distributions.

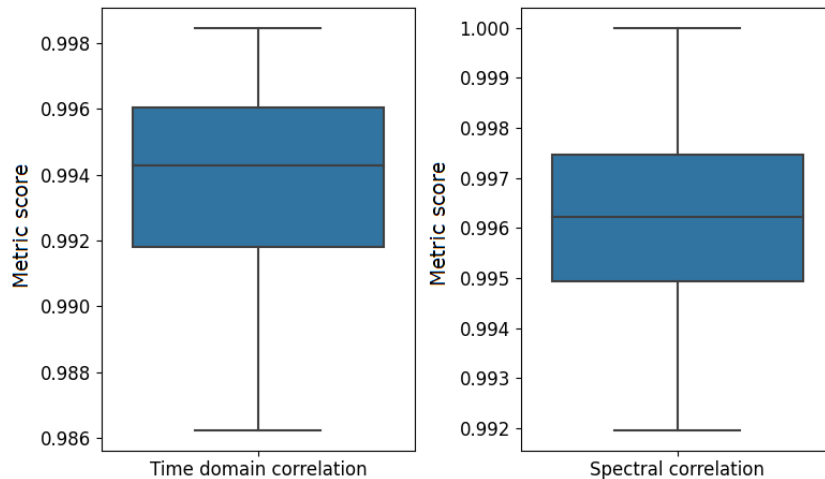


Figure 3.6: Distribution of correlation values with zero lag in time domain (left) and frequency domain (right) between the different pairs of channels present in the high-pass filtered epochs using iEEGPrep and Fieldtrip. Values obtained show a correlation above 98% for both cases.

As expected, from the transition band and cut-off frequencies used, the main differences are seen at the lowest frequencies. For the 1 Hz component, no visible difference is perceived. As shown by Figure 3.7, the only visible variation in the distributions between Python and

MATLAB is observed at 0.33 Hz, which reflects how the Fieldtrip filter is attenuating higher frequencies compared to iEEGPrep implementation, regardless the fact that the Fieldtrip filter was tuned to try to mimic the output given by iEEGPrep; nevertheless, this difference was not found to be significant. These subtle differences can be further observed in the time-frequency map shown in the Appendix C.

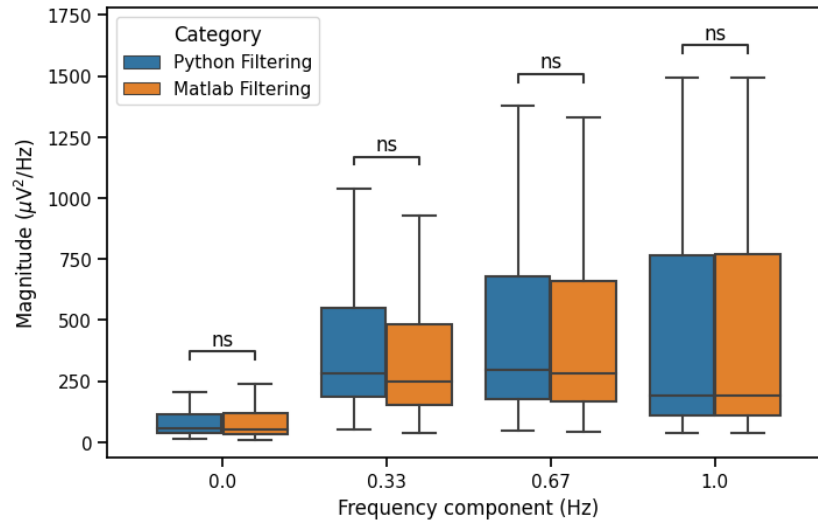


Figure 3.7: Statistical comparison of the first four frequency components between the different pairs of channels present in the high-pass filtered epochs using iEEGPrep and Fieldtrip. Components were calculated using Welch’s method. No statistical significant differences were observed.

Comparison between the different implemented methods

As the second validation for drift correction, a qualitative comparison between the different implemented methods was performed. As shown in Figure 3.8, high-pass filtering was the only method where the original low frequency oscillations observed in the downsampled signal (Figure 3.3) were considerably attenuated.

A visual comparison of the first four frequency components, calculated through Welch’s method, is shown in Figure 3.9. In line with what was seen on the time domain for the LPIIn02 signal, high-pass filtering is the only method with a visible change for the first two components. Linear detrending and demeaning do not seem to have any visible effect in the frequency response, possibly due to a dynamic change of the linear trend and mean of the signal across

time, which can be already perceived from Figure 3.5.

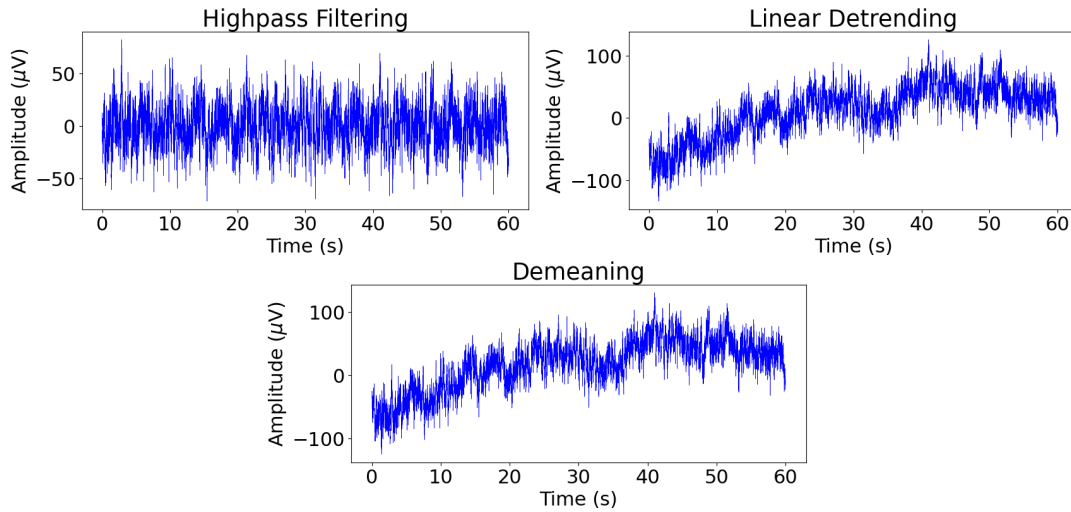


Figure 3.8: Time plots for the first 60 seconds of the detrended signal from the LPIIn02 channel using iEEGPrep and Fieldtrip.

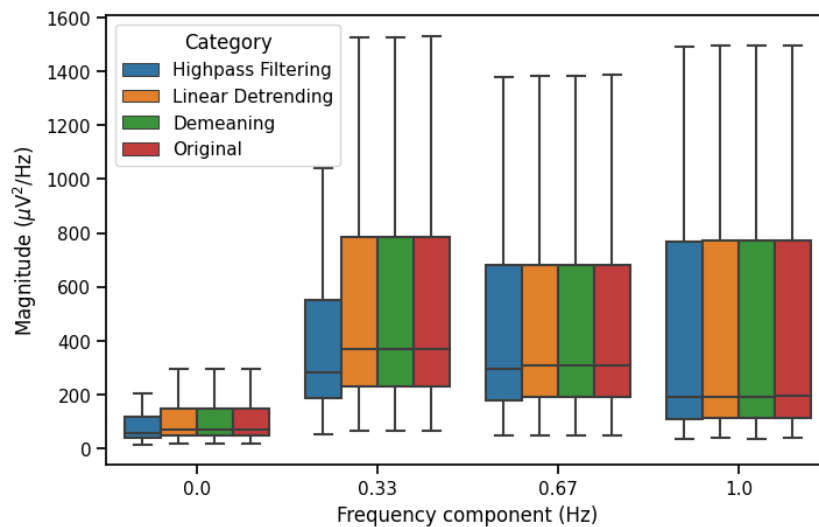


Figure 3.9: Distribution of amplitude values of the first four frequency components for the channels before (Original, in red) and after drift correction. High-pass filtering was the only method that showed a visual difference for the first two frequency components. Components were calculated using Welch's method.

Re-referencing

After applying drift correction to the test EDF file, the channels were re-referenced to a bipolar montage, as this is the only re-reference montage available at this point in iEEGPrep. As

previously mentioned, iEEGPrep and Fieldtrip use the same convention to execute the re-referencing; therefore, no substantial differences were expected.

As shown in Figure 3.10, the correlation values obtained oscillate between 0.98 and 1, similar to the ones found after high-pass filtering. When looking at individual signals, with the example case of the LPIIn2-3 (or LPIIn02-LPIIn03) channel, no evident differences were visually observed, as shown in Figure 3.11 for the frequency domain. Further validation using time-frequency maps did not show visible differences, as shown in the Appendix C.

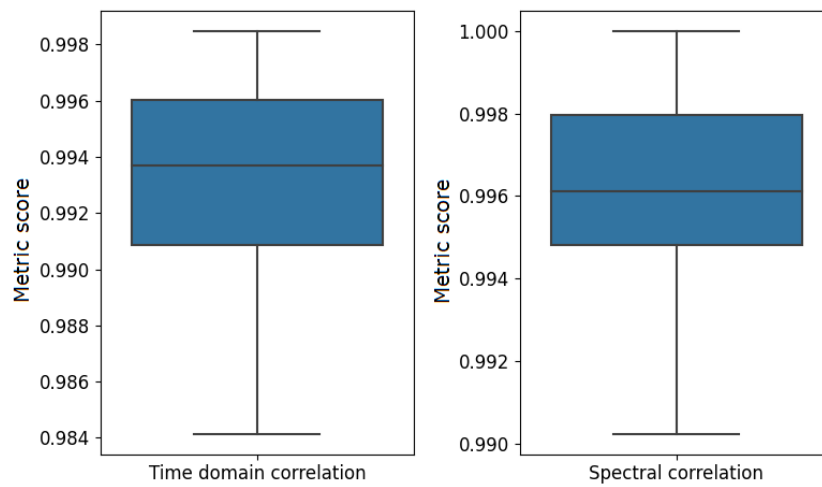


Figure 3.10: Distribution of correlation values with zero lag in time domain (left) and frequency domain (right) between the different pairs of channels present in the re-referenced epochs by iEEGPrep and Fieldtrip. Values obtained show a correlation above 98% for both cases.

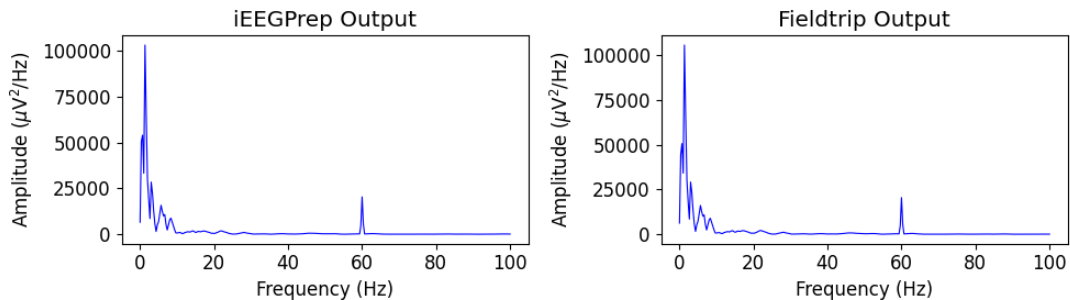


Figure 3.11: Frequency spectrum plots for the first 60 seconds from the channel LPIIn2-3 using iEEGPrep and Fieldtrip. Frequency domain signal was calculated using Welch's method.

PLI attenuation

The validation of this step is composed of three different parts: a comparison between Fieldtrip and iEEGPrep outputs for previously re-referenced data, a validation of each advanced PLI attenuation methods through a comparison between the MATLAB and the in-house Python implementations using simulated data, and finally a comparison between the different implemented advanced methods using simulated data.

Comparison against Fieldtrip output using real data

As briefly mentioned before, channel LPIn2-3 was chosen as the case of study for this report as it showed a significant power line noise contamination, which is visible in Figure 3.11 in the peak around 60 Hz. Cleanline was chosen as the PLI attenuation method for this first comparison, as it is recommended by EEGLAB, one of the most used EEG analysis tools. Figure 3.12 shows how both the MATLAB and Python implementation successfully attenuate the peak around 60 Hz. The same correlation values computed in previous steps were calculated but no difference compared to the distribution from the re-referencing step was seen. The figure showing the correlation values distribution can be found in the Appendix C.

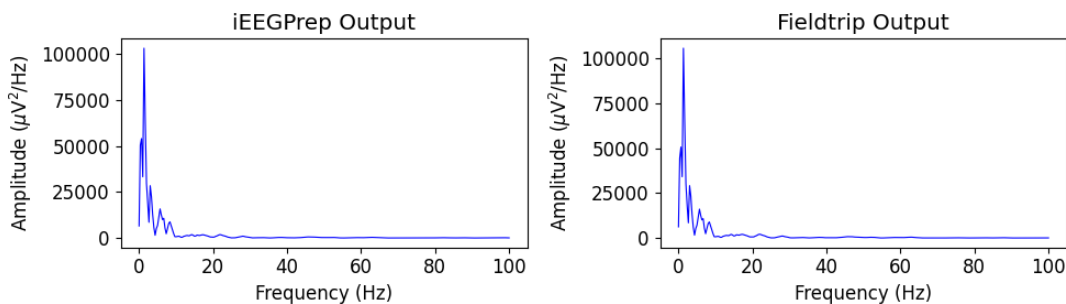


Figure 3.12: Frequency spectrum plots for the first 60 seconds from the channel LPIn2-3 using iEEGPrep and Fieldtrip after PLI attenuation using Cleanline algorithm. Frequency domain signal was calculated using Welch's method.

Comparison against MATLAB's implementation for the used advanced PLI attenuation methods

Three advanced PLI attenuation methods were re-implemented in Python for iEEGPrep: the new implementation of the Cleanline algorithm, Zapline and removePLI. As these al-

gorithms were initially implemented in MATLAB, a comparison between the in-house re-implementation in Python and the original MATLAB was performed to quality control for any possible mistakes in the process. To do so, hundred EEG channels with a sampling rate of 400 Hz were simulated as random pink noise plus frequency components at 60 Hz and its harmonics. An example of these signals is shown in the left column of Figure 3.13, where PLI peaks in the fundamental frequency and two harmonics are visible.

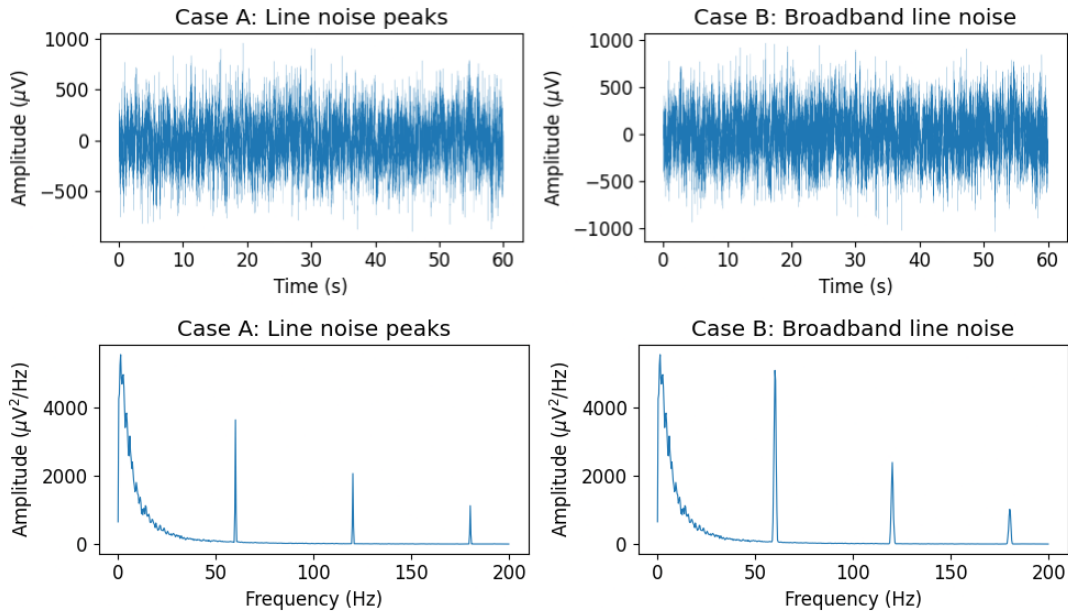


Figure 3.13: An example of a signal from the 100 SEEG channels simulated for validation of PLI attenuation methods for the two considered cases. Left column shows the case A, where peaks at the fundamental frequency and its harmonics were induced into the data. Right column shows case B, where more broadband line noise was injected into the simulated data.

Two metrics were used to compare MATLAB’s original implementation to the Python one for iEEGPrep: SNR and attenuation. Both metrics were computed in decibels. SNR was calculated using the following equation:

$$\text{SNR (dB)} = 10 \cdot \log_{10} \left(\frac{\text{average power of clean signal}}{\text{average power of remainder noise}} \right) \quad (3.1)$$

where the average power was estimated as the mean of the squared signal and the remainder noise was calculated as the subtraction of the filtered signal with the simulated pink noise signal

before the injection of PLI.

$$\text{Attenuation (dB)} = \sum_{n=1}^{n=3} 20 \cdot \log_{10} \left(\frac{BP_{filt, n\omega_0-5 \text{ Hz} \rightarrow n\omega_0+5 \text{ Hz}}}{BP_{noisy, n\omega_0-5 \text{ Hz} \rightarrow n\omega_0+5 \text{ Hz}}} \right) \quad (3.2)$$

where ω_0 represents the fundamental frequency (60 Hz), n the harmonic number (1 for fundamental, 2 for the second harmonic, etc.) and BP the bandpower for a signal between two frequencies. The numerator of the fraction corresponds to the bandpower of the filtered signal between the harmonic frequency minus 5 Hz and plus 5 Hz, and the denominator is the equivalent but for the simulated signal with the PLI. The bandpower from f_1 to f_2 was then calculated as follows:

$$BP_{f_1 \rightarrow f_2} \text{ (dB)} = \int_{f_1}^{f_2} W(f) df \quad (3.3)$$

where $W(f)$ is the value from the Welch's method result at a specific frequency f . The integral was calculated through the trapezoidal rule implemented in Numpy.

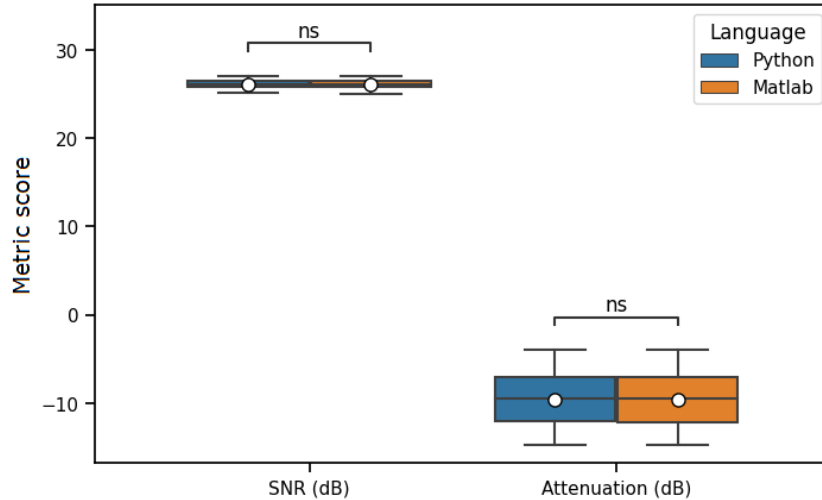


Figure 3.14: Comparison between the Cleanline implementation in MATLAB and Python using signal-to-noise ratio (SNR) and the attenuation in decibels. No statistical significant differences were found between the means of the distributions for both metrics, with p-values above 0.4. A power of 87.1% was achieved with a hypothetical effect size of 0.14 dB for the SNR comparison, while a power of 81.7% was achieved with a hypothetical effect size of 1.1 dB for the attenuation comparison.

Figures 3.14, 3.15 and 3.16 show the distribution of these metrics computed for each chan-

nel using Cleanline, Zapline and removePLI respectively. Statistical significance was computed using permutation testing with 100000 repetitions and FDR correction. No statistically significant differences were found for any of the methods, neither in SNR nor in attenuation values, showing a visually identical distribution for all cases.

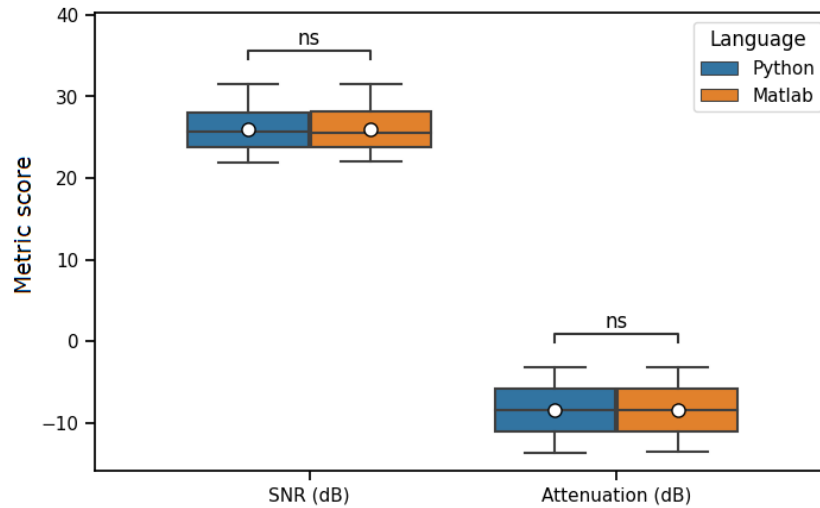


Figure 3.15: Comparison between the Zapline implementation in MATLAB and Python using signal-to-noise ratio (SNR) and the attenuation in decibels. No statistical significant differences were found between the means of the distributions for both metrics, with p-values above 0.4. A power of 85.1% was achieved with a hypothetical effect size of 0.14 dB for the SNR comparison, while a power of 80.9% was achieved with a hypothetical effect size of 1.1 dB for the attenuation comparison.

Furthermore, a power analysis for each of the comparisons was conducted, aiming to determine the likelihood that the test would detect a meaningful difference if one exists. This power analysis was conducted by simulating data under a hypothetical effect size. More specifically, the compared groups were simulated using a normal distribution with the mean of each original group (metrics from MATLAB and Python) and the standard deviation of the combined data from both groups. A hypothetical difference was then added to one of the two groups as the original data did not show any statistically significant difference. This difference was chosen to verify the minimum differences between the means of the two groups that the statistical test is capable of detecting as significant. The significance level was also set to 0.05 in this case and the minimum power considered sufficient was 80%.

In the case of the SNR comparisons, Zapline was the case that showed the highest minimum difference that can be detected with the statistical comparison, with a power of 85.1% when using a hypothetical effect size of 1 dB. On the other hand, the minimum difference that can be detected in the attenuation test to achieve a power above 80% was found to be close to 1.1 dB for all the different methods. Based on these results, the validation was considered satisfactory.

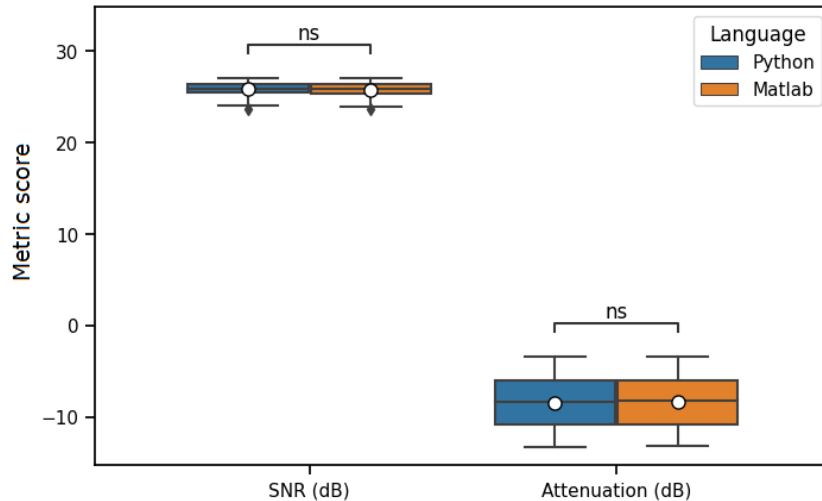


Figure 3.16: Comparison between the removePLI implementation in MATLAB and Python using signal-to-noise ratio (SNR) and the attenuation in decibels. No statistical significant differences were found between the means of the distributions for both metrics, with p-values above 0.3. A power of 85.1% was achieved with a hypothetical effect size of 0.14 dB for the SNR comparison, while a power of 81.1% was achieved with a hypothetical effect size of 1.1 dB for the attenuation comparison.

Comparison between the different implemented methods

As a final step of the PLI attenuation step, a comparison between the different implemented methods was performed. First, two cases of simulated data were evaluated and then the methods were also compared using real data from the test EDF file previously used.

Figure 3.13 shows the two simulated cases, one with sharp peaks in the fundamental and harmonics and the other with more broadband noise. The main used to compare the different methods was the SNR in decibels. Other metrics such as RMSE in time domain and RMSE in frequency domain with the Welch's method results were also considered, results that are shown in the Appendix C.

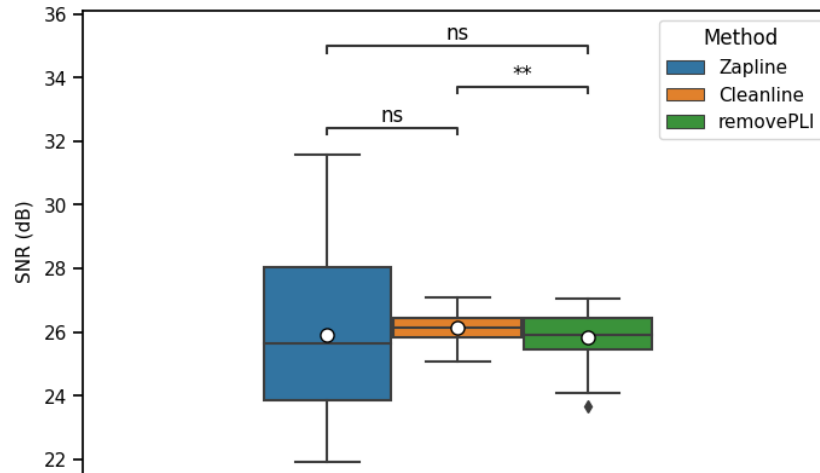


Figure 3.17: Distribution of SNR on the simulated SEEG channels in case A for the different advanced PLI attenuation methods implemented in iEEGPrep. Zapline showed a significantly higher attenuation compared to the other methods, as indicated by its more negative dB values. No statistically significant difference was found between the Zapline and removePLI methods.

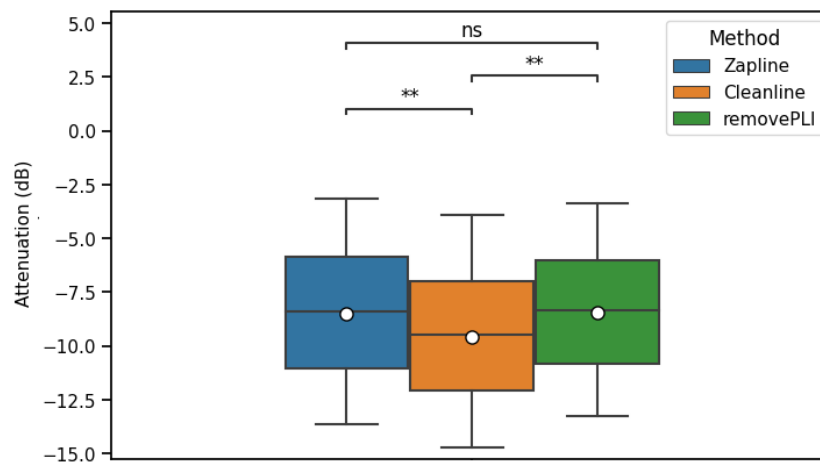


Figure 3.18: Distribution of attenuation values (dB) on the simulated SEEG channels in case A for the different advanced PLI attenuation methods implemented in iEEGPrep. Zapline showed a significantly higher SNR compared to the other methods. No statistical significance was found between Cleanline and removePLI methods.

As shown in Figure 3.17, the Cleanline algorithm showed a significantly better SNR compared to removePLI for case A, with a p-value of $1.14e-3$. No significant difference was found between Cleanline and Zapline when using SNR; however, the attenuation values did show a significant difference between these two methods, with a p-value of $9.375e-3$, as shown in Figure 3.18. Cleanline also showed a better performance when using other metrics such as RMSE,

further described in the Appendix C. No statistically significant differences were observed between Zapline and removePLI for most metrics, with the exception of RMSE in frequency domain. Visually, no major differences were observed in the frequency domain for the three methods, which is exemplified in Figure 3.19. Case B showed more drastic results compared to case A, with Zapline being the only method able to successfully remove the broadband noise; however case A was considered more relevant to this work as it is the standard method for testing PLI. Results for case B can be further seen in the Appendix C.

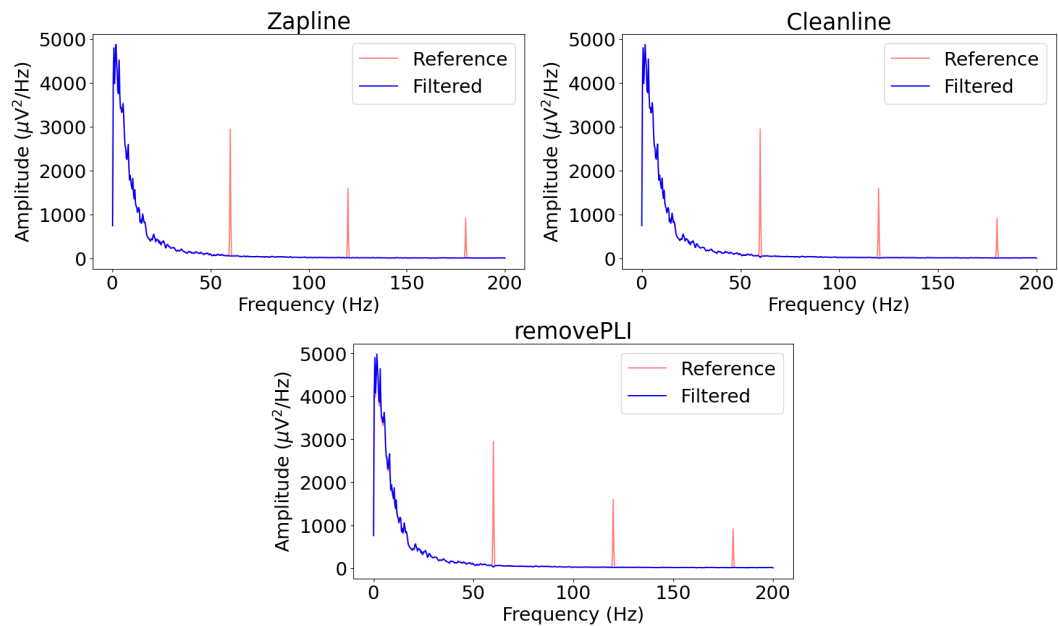


Figure 3.19: Example of power spectral density curves obtained through Welch’s method for one simulated signal filtered for case A through the different advanced line noise attenuation algorithms implemented in iEEGPrep. Orange and blue signals show simulated signal before and after PLI attenuation respectively.

Then, it was intended to verify if the same behavior was observed when using real data. The previously re-referenced EDF file was further preprocessed using iEEGPrep with Cleanline, Zapline and removePLI. The same channel LPIn2-3 was used to qualitatively compare the results. As shown in Figure 3.20, when looking at the average frequency characteristics across time, all of the methods are able to remove the previously observed peak around 60 Hz in Figure 3.11.

As Welch’s method only gives an average frequency characteristic, time-frequency maps

were then plotted to verify that the power line noise was successfully attenuated across time. Before power line noise attenuation, line noise predominated mainly in the first part of the signal, being the component with higher amplitude, which can be seen in the supplemental Figure C.7 in the Appendix C; therefore, the same time frame was plotted after the application of the PLI reduction.

Similar to the results found in case A, when applied to real data, Cleanline was the algorithm that seemed to have a better performance when visually compared, as shown in Figure 3.21. In this case, Zapline seemed to attenuate considerably the noise but it did not remove it completely. The analyzed channel did not show constant line noise contamination across time, as it predominated mainly at the beginning of the signal. Since Zapline is based on the extraction of the line noise component across the whole signal, it is possible that it was not able to remove all of the contamination due to this noise variability across time.

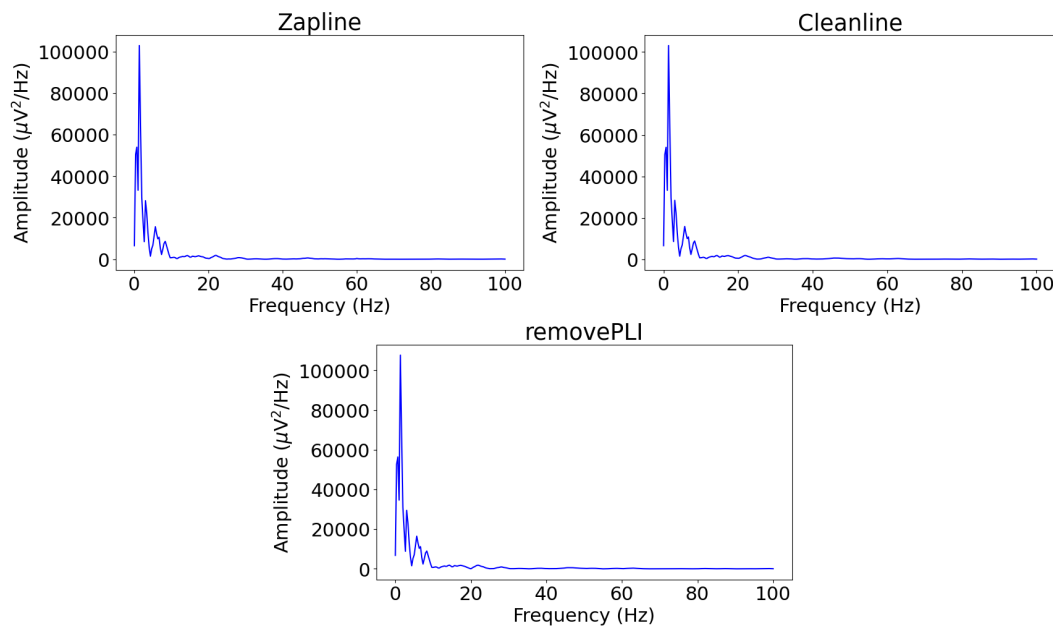


Figure 3.20: Power spectral density plot for the channel LPIIn2-3 in the preprocessed EDF file using the different advanced PLI attenuation methods.

The method removePLI showed a good attenuation with the exception of the beginning of the signal. This is probably related to its removal of the noise through an iterative fitting across time; therefore, the worst fit could be expected to be at the beginning of the signal. This

was also seen when looking at the time plot, where a big artifact is shown at the beginning, compared to Cleanline and Zapline, which can be observed in the supplemental Figure C.13 in the Appendix C.

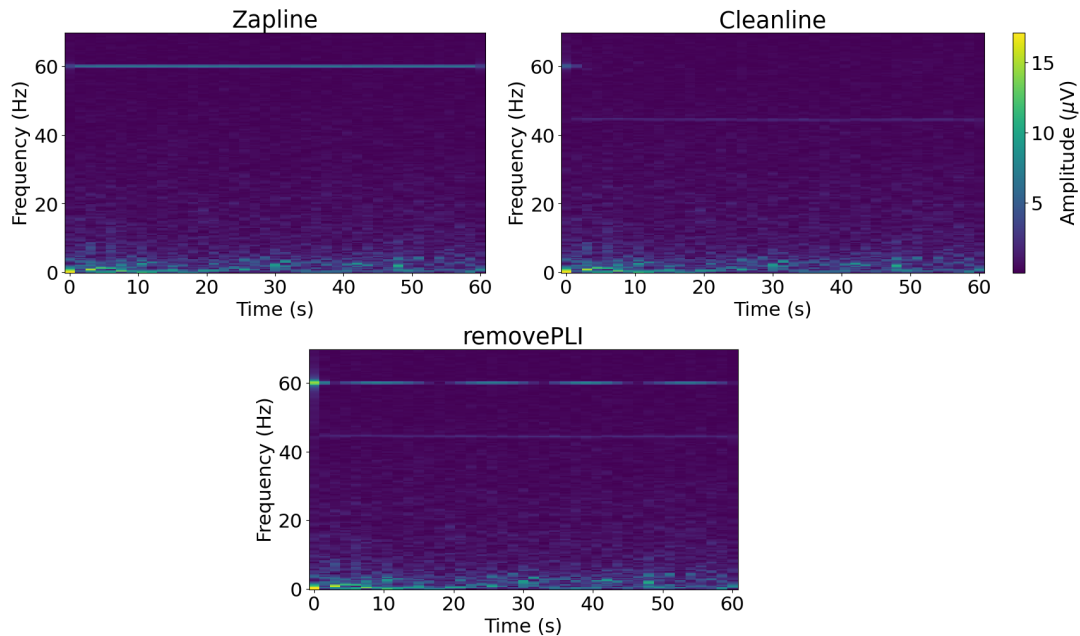


Figure 3.21: Time-frequency map from 0 Hz to 70 Hz of the first 60 seconds of the re-referenced signal with PLI attenuation from the channel LPIn2-3 using the implemented advanced PLI attenuation methods. Colorbar scale is the same for all plots.

From three the performed evaluations, Cleanline showed better performance in one of the simulated cases and also when using in real data. Simulated case with broadband PLI was the only one where Cleanline did not perform well, with Zapline being the only method able to remove the PLI. Further validation using more complex simulated data or large SEEG samples is required to determine the best method for SEEG preprocessing; however, the findings presented here show some of the strengths and limitations of each implemented algorithm.

Channel localization

The last preprocessing step validated was identification of the brain region for each channel. Only a visual validation was performed for this case by looking at the original CT and MRI along with the used parcellation file and electrodes positions. 3D Slicer, an open-source soft-

ware for visualization and processing of medical images and meshes was used for this purpose [41]. The parcellation tested was in the same space as the pre-operative T1w image (considered subject space) and SynthSeg was used to segment the pre-operative MRI, resulting in a parcellation with FreeSurfer’s volumetric regions of interest (ROI) [17].

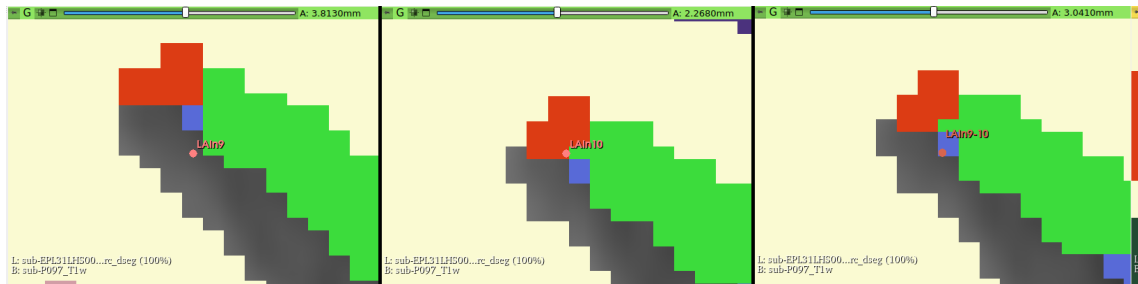


Figure 3.22: Coronal 2D view of the location of the monopolar channels part of the bipolar channel LAIN9-10 and the virtual midpoint between them. Left frame shows the position of the channel LAIN09 in the gray region (corresponding to ROI *ctx-lh-insula*). The picture in the middle shows the position of the channel LAIN10 near the intersection between the ROI *ctx-lh-insula* (gray) and *ctx-lh-parstriangularis* (red). The right plane shows the virtual midpoint between these two monopolar channels lying in the CSF (blue). The yellow region corresponds to the ROI *Left-Cerebral-White-Matter*.

Next, we present one case showing the contribution of having a volume mask for region identification for a bipolar channel instead of just using the midpoint between the two contacts part of the channel. The bipolar channel chosen for this demonstration was LAIN9-10, which is expected to be close to the left anterior insula. This channel was chosen as it resulted in different regions identified when using a volumetric mask compared to the mid point between the contacts.

As shown in Figure 3.22, the ROI were the contact corresponding to LAIN9 is located corresponds to *ctx-lh-insula* (which represents the left insula gray matter) and LAIN10 is located near the intersection between this gray region and the *ctx-lh-parstriangularis* (red ROI, abbreviation for left pars triangularis grey matter). Based on the location of the contacts, it would be expected that the assigned ROI for the bipolar channel LAIN9-10 was *ctx-lh-insula*, as it seems that the region between and around the two contacts part of the channel would mostly correspond to the gray area. However, when looking at the mid coordinate between both con-

tacts in Figure 3.22, we can see that this virtual point ends up lying in a small part of the ROI constituting the cerebrospinal fluid (CSF).

As previously explained, in a bipolar configuration, the captured signal mainly comes from the regions around each contact and in the middle of them from volume conduction. When looking at the different planes in Figure 3.22, the CSF does not seem to have predominance neither around the contact nor in between, but the middle point ended up lying in a small section of the ROI just in between the contacts. Therefore, when running iEEGPrep with point-based localization of the bipolar channel, CSF is assigned to LAIN9-10. On the other hand, when using the volumetric approach, the channel is assigned to the ROI *ctx-lh-insula*, with a 62% of the mask inside it, followed by *CSF* with 17%.

This case shows how using a volumetric mask might give a better intuition of where the signals captured by an SEEG channel might come from. The main result from this step in iEEGPrep just shows the biggest region per channel; nevertheless, to increase the transparency of the results, a supplemental JSON file is given, showing the distribution of the volumetric mask across ROIs per channel, which should facilitate the QC of the results.

3.1.2 Comparison against existing solutions

The process of validation allowed a direct comparison of not only the outputs of iEEGPrep and Fieldtrip but also their usability for the user. The main function used by Fieldtrip for preprocessing is called `ft_preprocessing`, which receives a *structure array* object with the desired configuration, indicating the desired steps to run and some specific parameters for each step, like the high-pass filter frequency. Contrary to iEEGPrep, when using this function Fieldtrip does not allow a quality control of the preprocessing between the different steps and no online information was found with respect to the order of the steps.

To be able to QC each step, several calls of the `ft_preprocessing` had to be implemented in a script, which can also be achieved by using individual functions available for each preprocessing function. Furthermore, additional code had to be used to achieve these QC results;

for example, the computation of a transformation to frequency domain for each channel. The code used for the preprocessing of the file using Fieldtrip can be found in the Appendix D. In iEEGPrep, several QC metrics and plots are already pre-computed and outputted in a HTML report for the user.

Through *snakebids*, iEEGPrep allows for an easy reading of large iEEG-BIDS datasets. On the other hand, Fieldtrip requires the input of the path specific for each file as part of the information of the configuration structure array passed to `ft_preprocessing`, as shown in line 3 from the code in the Appendix D. This would require a *for* loop and the use of another tool such as *glob* to parse the directory.

The re-referencing step was also found to be more rigid compared to my implementation, as the electrodes have to be named using only letters for the electrode name and numbers for the order of the contacts. The developed tool, iEEGPrep, supports parsing of monopolar channels with names with the following formats:

- Text + number. Example: LPIn09
- Text + number + dash + number. Example: LPIn1-9.
- Text + number + text + number. Example: LIn1A-9.

This eases the parsing of monopolar channels for re-referencing compared to Fieldtrip.

Overall, compared to existing tools previously discussed in the Introduction section like Fieldtrip and Brainstorm, iEEGPrep provides improvements in the following points:

- **Usability:** My tool allows an easier parsing of iEEG-BIDS datasets with help of Snakebids, as shown with its comparison to Fieldtrip. In addition, prior tools rely on automatic generation of code or the direct scripting of pipelines, which make it difficult to configure for inexperienced users, especially considering the poor code documentation available. iEEGPrep allows the configuration of the preprocessing steps through a simple human-readable YAML file, easing the usability of my tool. Finally, the independence

of iEEGPrep from paid-based software, like MATLAB, allows its usage to all users independently of their access to private licenses.

- **Scalability:** Main existing solutions available for iEEG preprocessing do not support direct scalability of workflows in distributed computing systems. Fieldtrip was the only tool from the analyzed group that provides documentation for usage in distributed systems using *qsub*; however, no straightforward approach for its usage for preprocessing is available, hindering its implementation for users less experienced in coding. iEEGPrep improves the scalability of the iEEG preprocessing through Snakemake, which allows users to easily scale the workflow based on the available resources through the CLI (like the number of cores or memory available). Future work should focus on better documentation for the users, as it will be later discussed.
- **Transparency and Reproducibility:** Additionally, based on the YAML configuration file used by iEEGPrep, my tool facilitates both the understanding of how the methods are implemented thanks to the availability of descriptions for each method and their corresponding parameters in this file. This improves the transparency of my implementation compared to previous solutions, which do not provide easily accessible documentation of their methods and parameters used, forcing the user to review the code to try to understand the implementations of the methods and their parameters. Furthermore, iEEGPrep improves the transparency and reproducibility of preprocessing applied prior further analysis for published studies or datasets as it only requires the sharing of this single human-readable YAML file. Availability of qc checkpoints in iEEGPrep also improves the easy review of the effect of each preprocessing step on the data, enhancing the transparency of the results, a feature that was not found in any other prior solution.
- **Portability:** The development of iEEGPrep using open-source software and Python as the base language, allows the containerization of the tool and dependencies management, which is considered part of the future work for this thesis.

- **Availability of PLI attenuation methods:** iEEGPrep supports the usage of four different PLI attenuation methods without any additional installation required, which is not available in any other reviewed software tool. This allows users to use their preferred method or easily test results using different methods.

3.2 Validation of the Automatic iEEG Artifact Detector

3.2.1 Datasets information

Initially, we inspected the datasets used for training and validation. Figure 3.23 presents two example clips from each class used for training. During this inspection, many instances from the *noise* and *baseline* classes appeared almost indistinguishable. Even some instances from the baseline class seemed noisier than those in the *noise* group. For example, in the bottom left signal from Figure 3.23, the first part of the signal contains fast oscillations that could be interpreted as noise. No clear visual differences were neither observed when transforming the images to time-frequency maps. On the other hand, the pathological class seemed to be mostly characterized by spikes, making it easier to distinguish from the other two groups.

3.2.2 Performance evaluation

As explained in the methodology, the evaluation of the models consisted in three steps: hyperparameter tuning, selection of the best traditional and deep learning models, and evaluation of the best global model. Next, the results for each step are presented. Finally, the insights gained from the clinical qualitative assessment are shown.

Selection of the best hyperparameters

As aforementioned, hyperparameter tuning for the traditional machine learning models was performed using 10 fold cross validation using macro F1-score as the metrics. Description

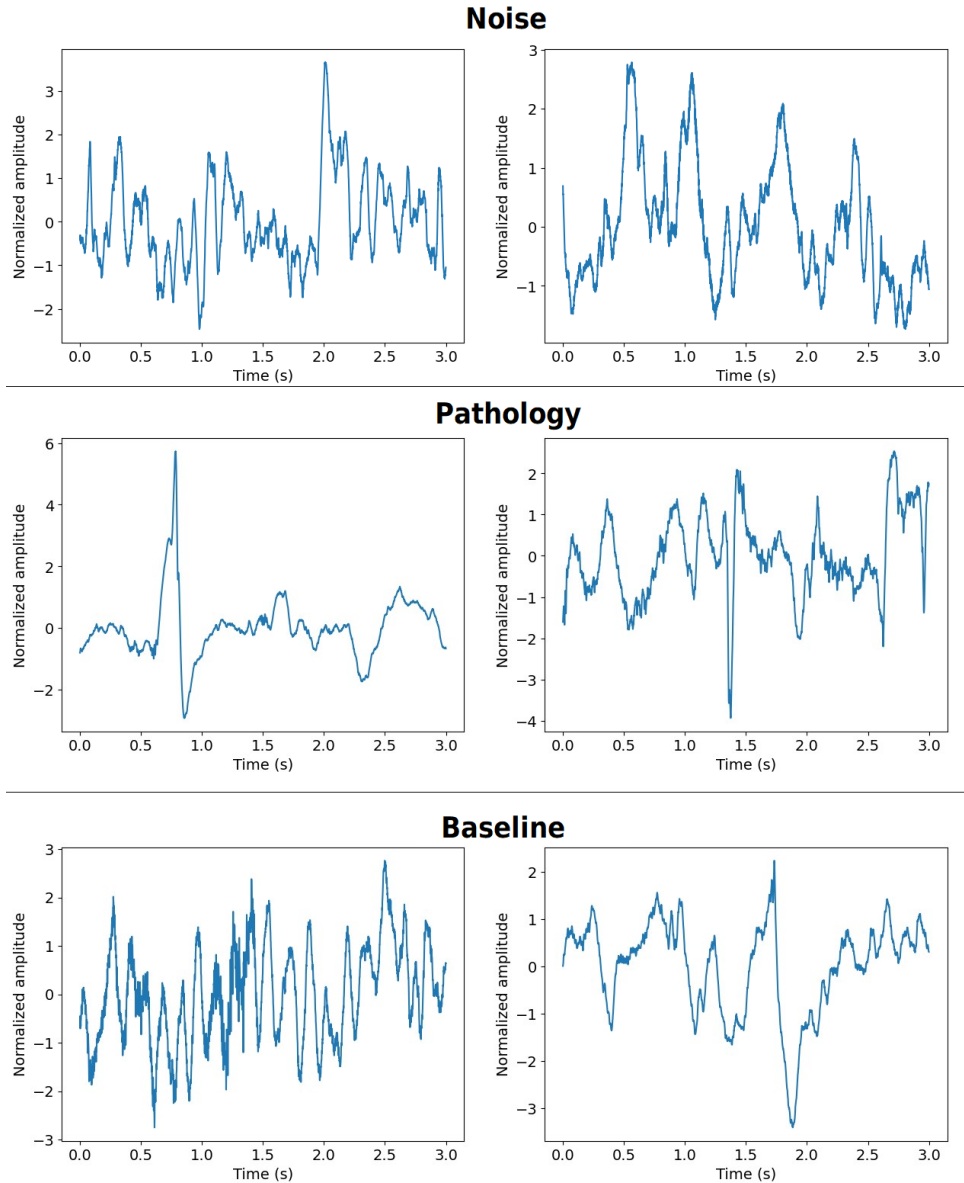


Figure 3.23: Examples of instances per class in the FNUSA and Mayo datasets used for training and validation.

of the evaluated models and their tuned hyperparameters were previously described in the subsection *Selection of best hyperparameters* under the Methodology chapter. Furthermore, the grid of hyperparameters tested for each model can be found in Appendix E. Below, the best hyperparameters found per model are shown. Those hyperparameters not discussed were not included in the grid search and they were kept as default according to the corresponding package used.

- *Logistic Regression:*
 - Inverse of regularization strength C of 1.
 - Solvers: Newton-Conjugate-Gradient method (`newton-cg` with l_2 penalty (Ridge regression)).
 - Multinomial loss.

- *Random Forest:*
 - Criteria function for splits: entropy.
 - A maximum depth per tree of 3.
 - An inspection of 50% of all the features when looking for the best fit.
 - A minimum percentage of 5% of all the samples per leaf.
 - 250 estimators or trees in the forest.

- *Extreme Gradient Boosting:*
 - A learning rate of 0.001.
 - A maximum depth per tree of 4.
 - 1000 estimators or trees used.

In the case of the two deep learning models tested, three architectures were compared for each of them. Figure 3.24 shows the results from the comparison between the three architectures implemented for the CNN model. All the box plots were generated using the Python package Seaborn.

As it can be observed, no statistical differences were obtained between the three architectures. No visible differences between the median and mean were neither observed. Following the parsimony principle, the simplest architecture (CNN-Small) was chosen as the final CNN model. Similarly, no significant statistical differences were obtained between the means of the

metrics distributions for the CNN+LSTM architectures. Then, the simplest model was also chosen as the final one.

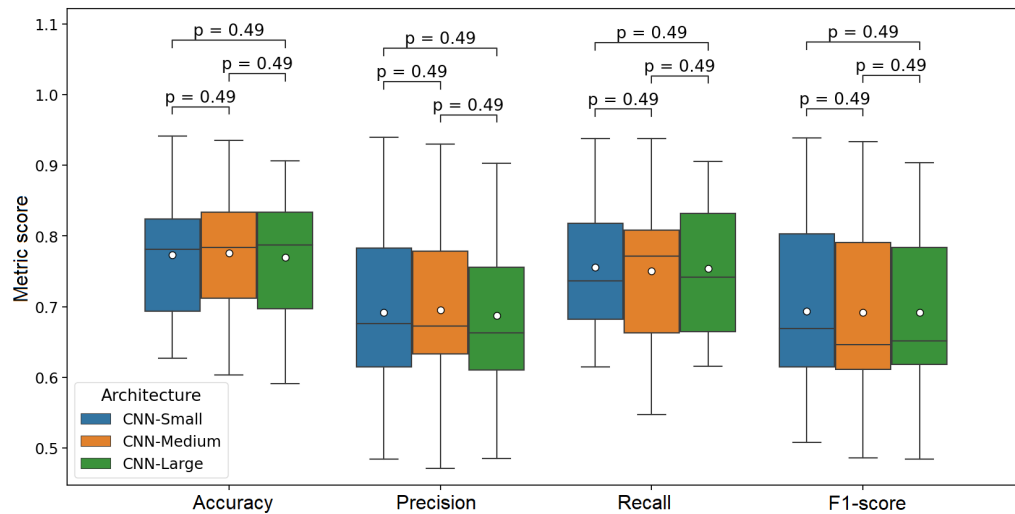


Figure 3.24: Comparison between the tested CNN architectures. Metrics used were accuracy, macro precision, macro recall and macro F1-score. CNN-S, CNN-M and CNN-L refer to the implemented small, medium and large architectures respectively. Box plot shows the interval between the 25% and 75% percentiles, the black line inside the box shows the median and the white dot the mean of each distribution.

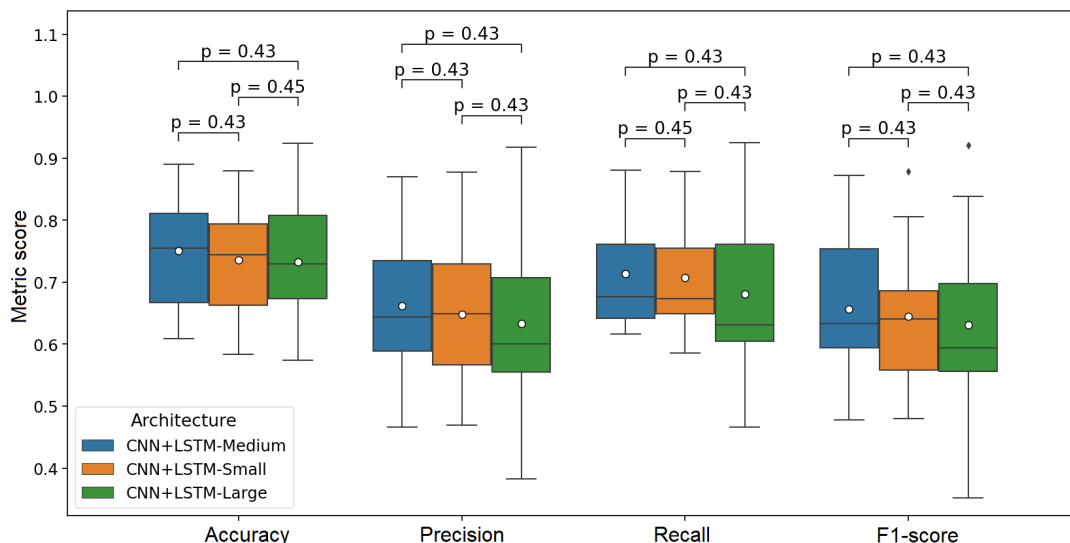


Figure 3.25: Comparison between the tested CNN+LSTM architectures. Metrics used were accuracy, macro precision, macro recall and macro F1-score. CNN+LSTM-S, CNN+LSTM-M and CNN+LSTM-L refer to the implemented small, medium and large architectures respectively. Box plot shows the interval between the 25% and 75% percentiles, the black line inside the box shows the median and the white dot the mean of each distribution.

Comparison of similar models

After selection of the best hyperparameters, a direct comparison between the traditional models was performed using 10 fold cross-validation with five repetitions. Figure 3.26 shows the box plots with the distributions of metrics computer per model. Significant statistical differences were obtained between the logistic regression model and the other two tree-based methods, which justifies the need of complex models. No significant differences were observed between the XGBoost classifier and the random forest; however, the last one was selected as the final traditional model due to its explainability simplicity compared to XGBoosting.

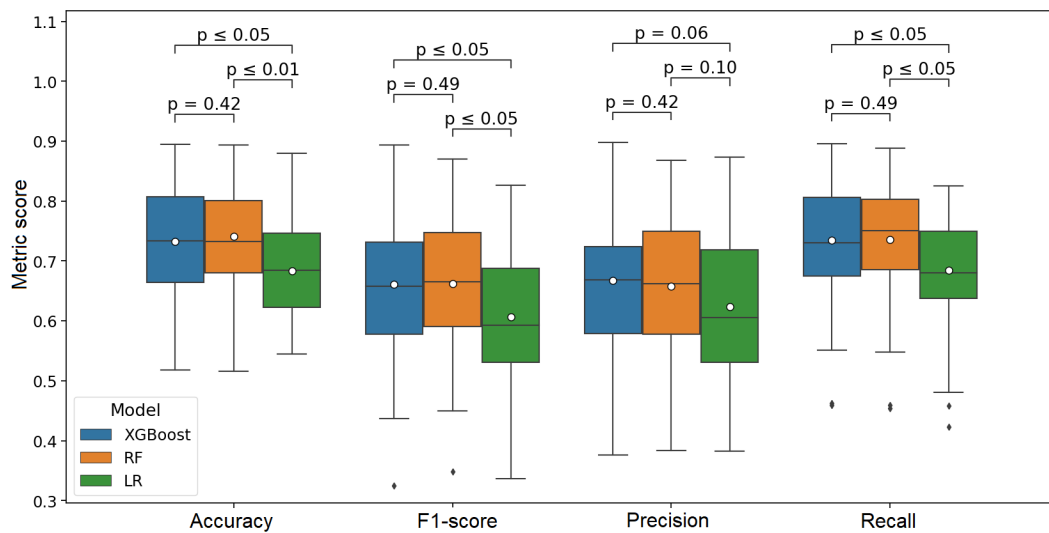


Figure 3.26: Comparison between the different traditional machine learning models after hyperparameter tuning. Metrics used were accuracy, macro precision, macro recall and macro F1-score. XGBoost, RF and LR correspond to the XGBoost, random forest and logistic regression classifiers respectively. Box plot shows the interval between the 25% and 75% percentiles while the black line inside the box shows the median and the white dot the mean of each distribution.

On the other hand, Figure 3.27 shows the distribution per metric for the chosen deep learning models. With p-values around 0.11, no statistically significant differences were obtained for any of these metrics. Nevertheless, the CNN model was selected as the best deep learning model since all of the metrics shows a better performance when only visually comparing the mean and median for each metric, as observed in Figure 3.27; besides the known higher computational complexity of LSTM networks compared to CNNs [123].

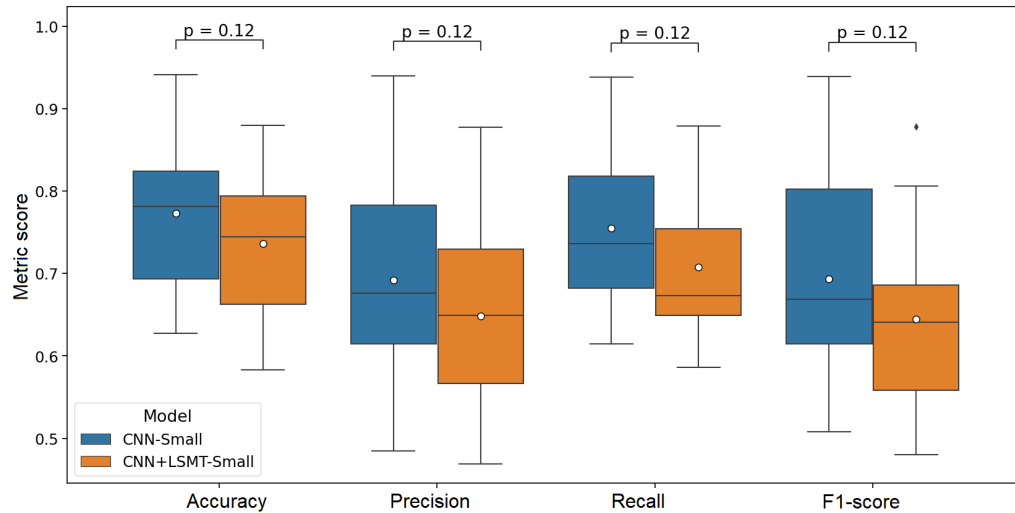


Figure 3.27: Comparison between the different deep learning models after architecture selection. Metrics used were accuracy, macro precision, macro recall and macro F1-score. CNN-Small and CNN+LSTM-Small correspond to the smallest implemented CNN and CNN+LSTM models respectively. Box plot shows the interval between the 25% and 75% percentiles while the black line inside the box shows the median and the white dot the mean of each distribution.

Deep learning vs traditional machine learning evaluation

As aforementioned, to compare the best deep learning model (CNN-Small) against the best traditional machine learning model (random forest), both were retrained using the complete dataset used for cross-validation; i.e., FNUSA and Mayo curated datasets, and then they were tested on the external dataset from the University of Florida. As described in the Methodology chapter, the data from University of Florida used a different referencing method (bipolar) compared to the two datasets used for training (CAR), then this test not only evaluates generalizability of the models to external data but also to a different referencing method. After training, the following metrics were obtained for both models using the dataset from University of Florida:

Model	Accuracy	Macro precision	Macro recall	Macro F1-score
CNN-Small	0.312	0.333	0.317	0.25
Random Forest	0.354	0.477	0.564	0.316

Table 3.1: Metrics obtained from testing of best deep learning and traditional machine learning models on the external dataset from the University of Florida.

As observed, the metrics drastically dropped for both models compared to the CV results. Several factors could explain this discrepancy. One primary reason could be the differences in the dataset labeling processes. For the FNUSA and Mayo datasets, continuous recordings were annotated by three reviewers. On the other hand, the University of Florida dataset was annotated by first detecting potential outlier events in the continuous recording, followed by filtering these events using an IEDs detection algorithm. Therefore, when inspecting the data, it was noted that most events in the University of Florida dataset contained some transient event similar to a spike, which two reviewers classified as pathological events, noise, or baseline activity. This difference in the annotation methods could be one of the reasons for such a change in the metrics.

To try to explain these results, the metrics for each class were computed for the best model, the random forest, along with the confusion matrix for better explainability, which are shown in Table 3.2 and Figure 3.28 respectively. The results showed how the noise class was significantly overestimated, particularly for the baseline class, where even more samples were misclassified as noise than as baseline. Some other samples from the pathology class were also misclassified as noise. This caused the precision for the noise class to drop drastically to 0.02 while actually detecting all the true positives. Thus, it is evident that the model is drastically overestimating the noise class, similar to the results previously found [99].

Class	Precision	Recall	F1-score	Number of samples
Noise	0.02	1.00	0.05	16
Pathology	0.99	0.37	0.54	798
Baseline	0.42	0.32	0.36	697
Macro average	0.48	0.35	0.36	1511

Table 3.2: Metrics obtained per class for the best implemented model on the external dataset from the University of Florida.

For the pathology class, the model showed a very high precision but low recall, contrary to the previous case. The confusion matrix shows that only three baseline clips were misclassified as pathological. However, many pathological events were misdiagnosed as baseline or noise, thereby reducing the recall. Finally, for baseline, neither a good precision or recall were

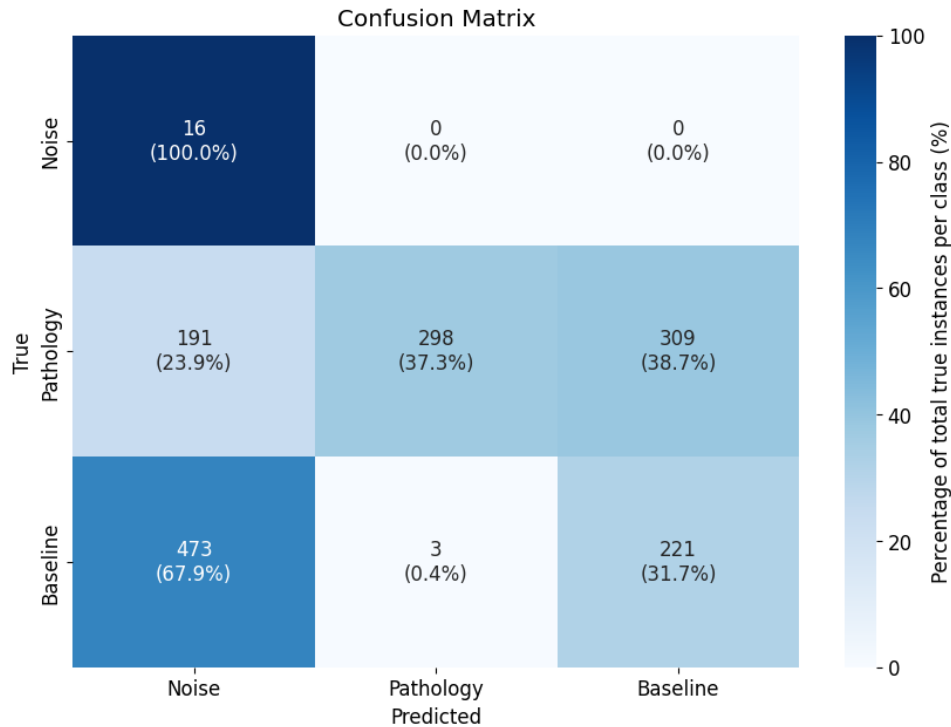


Figure 3.28: Confusion matrix for the random forest on the external dataset from the University of Florida Neuromedicine Hospital.

obtained since many pathological events were misclassified as baseline and many true baseline events were misdiagnosed as noise.

To gain a better understanding of the results, both correctly and incorrectly classified samples from each class were examined. For true pathological events, correctly classified samples tended to exhibit more prominent spikes compared to incorrectly classified ones. For events misclassified as noise, these samples indeed appeared noisier compared to those correctly classified as either pathology or baseline, with the model displaying high confidence in some cases, such as the top left event in Figure G.4. Additional examples are provided in the Appendix G.

As mentioned before,, one possible cause of the mismatch between the model's output and the reviewer criteria could be the differing annotation processes used. High disagreement rates between reviewers when annotating HFOs and IEDs, reported to be as high as 60%, could also contribute to the high error rate in the pathological class [110, 139, 13, 10]. Additionally, the different referencing method used in the dataset from the University of Florida compared to the

training data could also have an impact on these results.

The random forest evaluated in this work, which resulted as the best traditional machine learning model, has some resemblance to the one presented by Long *et al.* in their published work [99]. In addition, the datasets and their split for training and testing was very similar. Therefore, it is important to compare my results from the ones presented in this previous work. Before diving into the results, despite the similarities of the models, it is important to acknowledge some of the main differences in the training methods between this thesis and the work from Long *et al.*:

- **Dataset split:** In both this thesis and the previous work from Long *et al.*, the FNUSA and Mayo Clinic datasets were combined for training of the model while the data from the University of Florida were used as an external testing set. However, in this published work, the authors splitted the combined dataset into training and testing once again, which resulted in two testing sets: an internal set (from the same institutions used for training) and an external set (the one from the University of Florida), which differs from the approach followed in this thesis. In addition, Long *et al.* did not perform a similar curation of the data set to balance the number of samples per subject followed in this thesis for the training data set [99].
- **Input selection:** The input for the traditional models in this thesis, including the random forest, was uniquely based on features calculated from DWT components, while Long *et al.* calculated features for other frequency bands besides the ones obtained from DWT. Features used also differ a little between the two works, as some tailored features were used by Long *et al.* [99].
- **Probability threshold moving (PTM):** The paper from Long *et al.* described the tuning of the threshold for each class based on the validation data, which was not performed in this thesis, where the default method of assigning the class with the highest probability was used [99].

When talking about results, Long *et al.* showed better generalizability results compared to the ones presented on Table 3.2, with F1-scores of 0.148, 0.795 and 0.539 for the artifact, pathology and baselines classes respectively. However, it was not expected to replicate the results from this work due to the following reasons:

- Features that cannot be directly obtained from the iEEG data were also considered in their work, which have a high importance for their model including anatomical location of the channel, whether a channel is inside or outside of the soz, and the state of consciousness during the recording (awake or sleep). While these features could play an important role in the classification of interictal iEEG signals, I wanted to test models using features uniquely coming from the recordings, as these other features are not always available and it requires further data preparation before the usage of the model.
- In the work from Long *et al.*, even though the signals from the test data used bipolar referencing, some features from this test data were calculated using the monopolar CAR version of the signals instead. The authors justify this by describing how some morphological characteristics of the signals might change after re-referencing to bipolar, which might negatively affect the model performance. The released dataset only contains the bipolar signals, which make it impossible to use any monopolar information in this thesis. Therefore, even though the authors of this previous research claim to have a model that generalized to CAR and bipolar, the test results are based on the original signals using CAR before they were combined into the bipolar channels. On the other hand, I tested my models using only the bipolar signals to check their generalizability to another referencing method, which can also justify the decrease of the performance compared to the work from Long *et al.*
- Finally, this previous paper applies prior probability adjustment (PPA) to modify the probabilities for each class as outputted by the model considering the differences in the class distributions between the dataset used for training and test dataset. Nevertheless,

applying PPA on a dataset and then reporting the model performance on this same data might not be ideal as the model's outputs were modified to adjust the test set, which might bias the results presented. No similar methods were applied in this thesis.

As it will be discussed in the next chapter, future work is required to evaluate the direct effect of re-referencing differences in the performance of artifact detectors.

3.2.3 Clinical insights

As a final step in the evaluation of the best model (the random forest), multiple channels from two subjects from the LHSC were annotated with the model's outputs and converted to EDF format for inspection by a clinical fellow at the University Hospital, as previously explained. The recordings were then loaded into EDFBrowser and visualized using an amplitude scale between $30 \mu\text{V}/\text{mm}$ and $70 \mu\text{V}/\text{mm}$, and a time scale of $30 \text{ mm}/\text{s}$, closely matching the usual configuration for reviewing iEEG data in the hospital. Adjacent channels were also visualized during the review to provide the reviewer with more spatial context. The main observations gathered from this review process are presented below:

- All reviewed recordings contained a significant number of samples labeled as noise, but the reviewer could not identify any noisy patterns in the data. To provide a visual comparison, some manually identified noisy events from other recordings were shown. These events were clearly visible, typically having a very high amplitude compared to the baseline (e.g., eye blinks) or a high degree of high-frequency noise with an amplitude higher than the baseline activity. Therefore, it is likely, as expected from the previously obtained metrics on the University of Florida dataset, that the model is overestimating the number of noisy segments and classifying even clips with an unnoticeable amount of noise as noisy. Figure 3.29 shows one of these clips that the model labeled as noise, but the reviewers did not agree. As observed in the image, no clear differences are seen between the segment labeled as baseline and the one classified as noise.

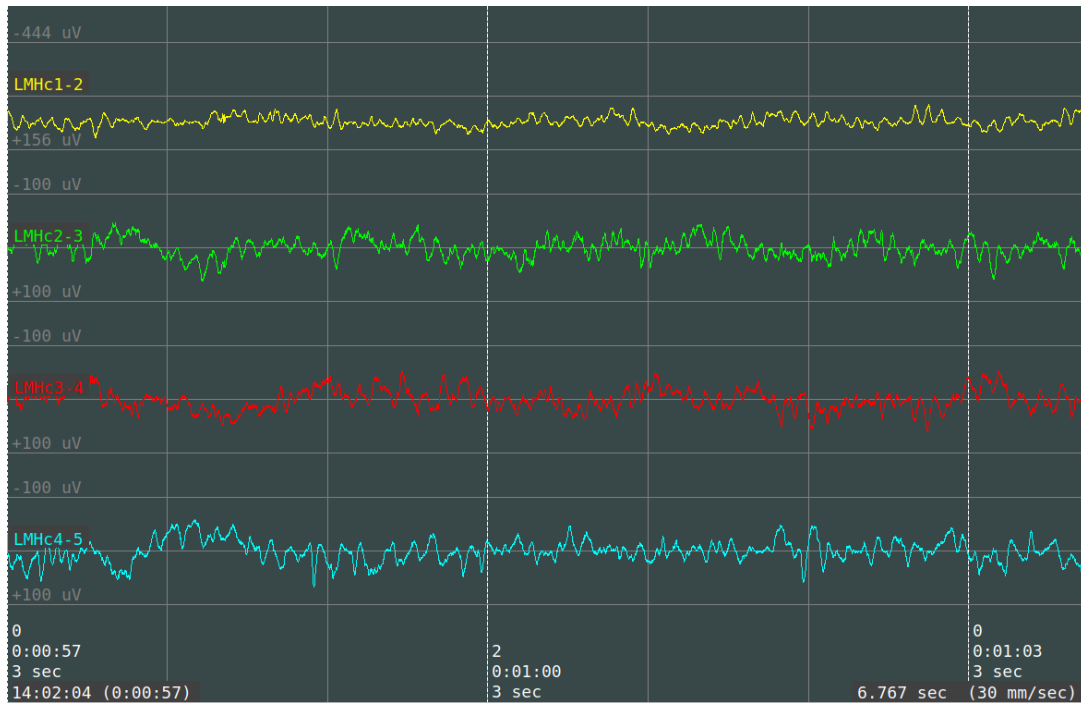


Figure 3.29: Example of one segment of data at the time 0:00:57 (left of the image) classified as noise (class 0) that was not verified as such by the expert reviewer. Middle segment at 0:01:00 shows a clip classified as baseline (class 2) by the model. Annotated channel with the model was LMHc1-2.

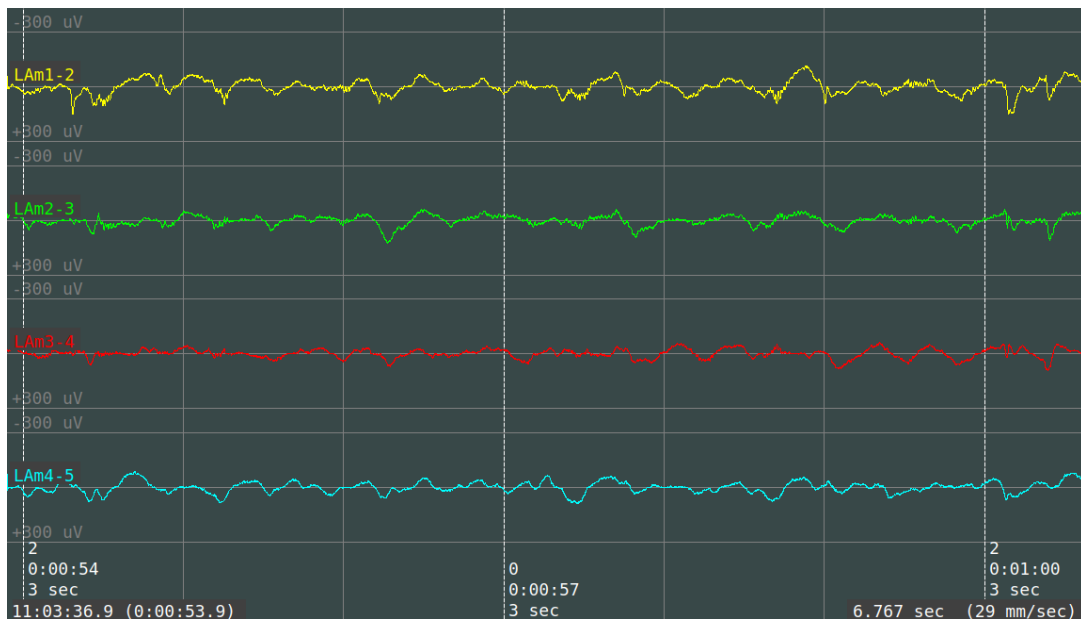


Figure 3.30: Example of two segments of data originally classified as baseline (left) and noise (right) for channel LAM1-2 that were classified as pathological due to LVFA by the expert review.

- The reviewer validated the pathological events corresponding to interictal spikes; however, many clips containing only HFOs were not identified as pathological by the model. When reviewing the data, the reviewer annotated channels exhibiting low voltage fast activity (LVFA) as pathological, which the model missed in most cases. Afterwards, the signal was high-pass filtered using a cutoff frequency of 80 Hz to verify if these oscillations had frequencies in the range of HFOs, which was confirmed. Figure 3.30 illustrates one of these cases where the model annotated two clips as baseline and noise, while the reviewer indicated that they contained pathological activity through LVFA. These signals were then high-pass filtered, clearly revealing the oscillations in the HFOs frequency range for both the annotated channel LAm1-2 as for LAm2-3, as shown in Figure 3.31.

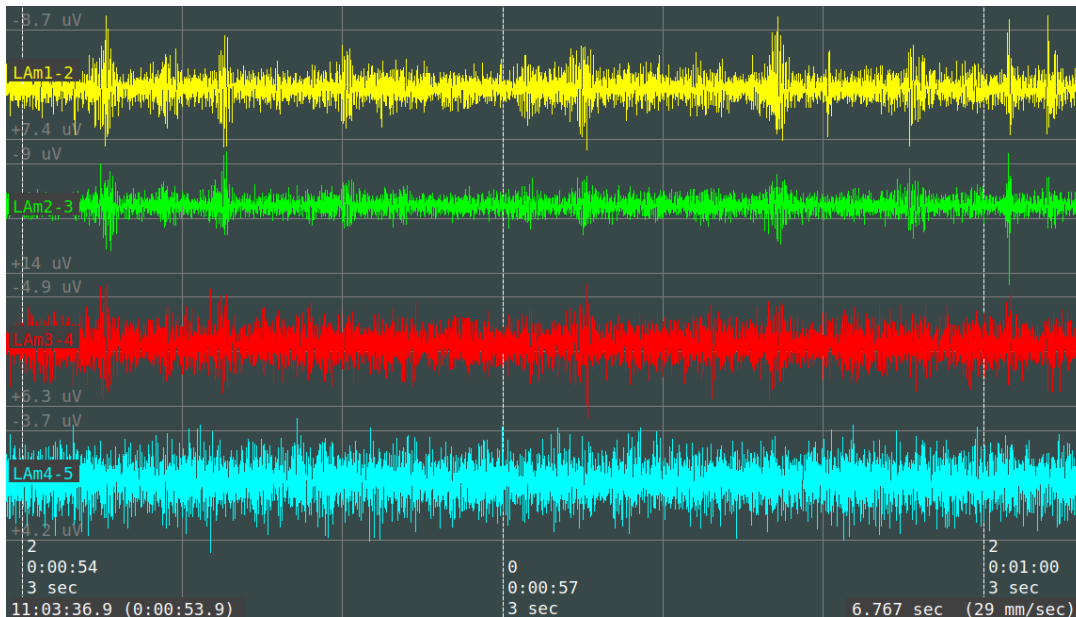


Figure 3.31: Signals shown in Figure 3.30 but with an applied high-pass filtered with cut-off frequency at 80 Hz. It shows two segments of data originally classified as baseline (left) and noise (right) for channel LAM1-2 that were classified as pathological due to LVFA by the expert review.

This could indicate that the training data lacked sufficient pathological events with only HFOs and no spikes. As previously mentioned, this was noticed when originally inspecting the training datasets, as most pathological events seemed to have interictal spikes, as

exemplified in Figure 2.3.

Chapter 4

Conclusions

4.1 Thesis Contributions

This thesis serves to address the challenge of non-standardized preprocessing of iEEG data. The main contribution of this work is the exhaustive research conducted to compile various useful preprocessing steps for iEEG. To this end, a comprehensive literature review of common preprocessing steps and an evaluation of how different methods applied to scalp EEG could be adapted to iEEG were conducted. Such research concludes in the development of a software workflow design that allows for more standardized preprocessing of scalp EEG datasets. This process facilitates a more consistent, transparent, reproducible, and user-friendly approach to data preparation, which is crucial for improving reproducibility and extension of results across different research environments.

The developed workflow integrates different preprocessing steps—such as line noise attenuation and re-referencing—that ensure datasets are prepared consistently before subsequent analysis. The workflow also includes user-friendly interfaces to assist researchers and clinicians in navigating through the preprocessing stages efficiently, regardless of the user’s expertise level. The configuration of the complete pipeline through a single human-readable YAML file also simplifies the sharing of the preprocessing parameters between research centers, bene-

fitting the reproduction and extension of previous studies and the seamless integration between preprocessed inter-institutional datasets.

Artifact detection for iEEG has not been standardized and validated; however, it is still a fundamental part of preprocessing not included in iEEGPrep. Multicenter released iEEG datasets were used to compare different deep learning and traditional machine learning models for artifact detection in interictal iEEG data, resulting in a random forest using features from discrete wavelet transform coefficients as the best model found. Even though the best models shows a good performance on the cross-validation results, testing on a external dataset and qualitative assessment in a clinical setting, showed poor generalizability of the model on recordings from institutions not used for training; possibly justified by inter-institutional differences between the recordings and their annotation, poor inter-rater agreement previously reported and insufficient training data to capture inter-subject and inter-institutional variability. The analysis conducted not only proposes potential models for artifact detection in interictal iEEG data, but, more importantly, it highlights some of the main remaining challenges to develop good machine learning models for this task.

4.2 A more standardized preprocessing of iEEG recordings

In this work, an initial version of a software workflow for the preprocessing of iEEG-BIDS datasets was presented. Validation against the commonly used tool Fieldtrip showed that each of the steps work as expected. Main differences were found due to variations between the drift correction implemented in iEEGPrep and Fieldtrip; however, they were found to be not statistically significant. Nevertheless, the observed differences highlight the importance of creating a tool that allows a more standardized preprocessing of iEEG datasets to minimize these potential variations between studies and preprocessed datasets. The implemented workflow, iEEGPrep, tries to achieve this by concatenating all of the configuration information into one human-readable YAML file, which can be easily shared by research groups after its usage.

Different methods implemented for drift correction and PLI attenuation were also compared. High-pass filtering showed the best performance on reducing both the mean and low frequency oscillations in iEEG data, which was expected based on previous recommendations. For PLI attenuation, Zapline showed a significantly higher performance when using simulated data; however, Cleanline showed a better performance when visually comparing the algorithms using real data. Further evaluation using large iEEG datasets should be performed to have a solid conclusion of the best algorithm for this type of data; nevertheless, the analysis showed the strengths and limitations of each of these methods. Zapline showed a better attenuation when the line noise is relatively stable across time, while Cleanline and removePLI seemed to deal better with dynamic PLI. The method removePLI presented more artifacts at the beginning of the signals, possibly caused through its time-based sequential fitting of noise. Cleanline showed to be the most robust algorithm as it worked well both for simulated and real signals.

Direct comparison against Fieldtrip also highlighted the strengths of the developed tool, mainly towards usability and scalability. My workflow iEEGPrep allows for a configuration of the parameters through a human-readable YAML file, which should be easier to a user with no coding experience compared to building a script for preprocessing using Fieldtrip. Furthermore, Snakemake makes distribution of computation resources, either when using local or distributed systems, easier for the user compared to the configuration needed in Fieldtrip through scripting.

A comparison against other previous solutions was also previously discussed, showing advantages for iEEGPrep in usability, reproducibility, scalability, portability, transparency and availability of PLI methods. Easy configuration through a human-readable YAML file boosts the usability and transparency of the tool, while also making it easy to share between studies and research settings. The use of Snakemake and Snakebids facilitates the scalability of the tool and parsing of iEEG-BIDS directories. iEEGPrep also offers a human-readable report that shows the effect of each preprocessing step on the dataset; enhancing the transparency of the results. Finally, iEEGPrep is developed completely with open-source tools, which removes the

need of a paid license compared to these other MATLAB-based toolboxes.

Even though several benefits of iEEGPrep were presented and discussed here, further work is required before public release of this workflow. This includes but it is not limited to:

- Reorganization of the outputs in a BIDS compliant approach.
- Support for other re-referencing techniques such as Laplacian and CAR, and support of ECoG preprocessing.
- Proper software packaging and testing of the tool.
- Testing with external large BIDS-iEEG datasets.

To my knowledge, this is the first automated toolbox for preprocessing of iEEG data. This tool aims the standardization of iEEG data preparation prior to experimental or clinical analysis, enhancing the reproducibility, accessibility, and quality control, in iEEG related studies.

4.3 Automatic Artifact Detection for iEEG

Artifact detection has been described as an important preprocessing step for iEEG recordings, particularly applicable to interictal data, where there is a combination of the baseline activity with pathological events and non-cerebral artifacts. Nevertheless, no standardized and validated approach has been defined for interictal iEEG data, which led to the temporary exclusion of this step from iEEGPrep. To tackle this limitation, this work also presented the training and evaluation of multiple traditional machine learning and deep learning models for classifying interictal clips of iEEG recordings. For traditional models, the tested simple model was found to perform significantly worse than the more complex ones, justifying the need for more sophisticated solutions to target this problem. On the other hand, it was found that the addition of LSTM cells to CNN architectures did not result in a significant performance increase, perhaps driven by the very small duration of the clips used for this classification.

The best traditional model (a random forest) exhibited a better performance compared to the best deep learning model. This highlights the importance of comparing traditional algorithms when implementing deep learning solutions, as deep learning models do not always achieve a better performance. In spite of the good validation metrics of the models and the use of two institutions for training, the performance of the best models on external datasets was found to be poor, reflecting on a low generalizability of the solution. Potential reasons for the poor performance of both models on external data include the lack of sufficient inter-institutional data for training and the different referencing method used in the test dataset compared to the training data. Other factors, such as non-standardized selection and labeling of clips and inter-rater variability, could also contribute. Overestimation of the noisy clips and misidentification of pathological events were found to be the most significant causes of the low macro metrics.

Clinical insights obtained from qualitative evaluation of the random forest's output confirmed the overestimation of the noise class by the model. The clinical reviewer was unable to corroborate any of the clips classified by the algorithm as noisy. Furthermore, misidentification of epochs contaminated with HFOs but no spikes was also recorded.

In this work we showed that the introduction of an non-cerebral artifact class to detection of interictal pathologies resulted in a high specificity but low recall for pathologies identification, contrary to the low specificity and high recall seen in previous HFOs and IEDs detectors. This shows that the addition of this extra class can result in an increase of the specificity for these detectors; however, further work is needed to achieve a better balance between specificity and recall. Based on the analysis presented, the following recommendations are given for future developments of interictal artifact detectors:

- Better sharing practices are needed to improve the development of interictal artifact detectors. Inter-institutional data are required to build robust machine learning classifiers. Furthermore, efforts are needed to reduce inter-rater disagreement in labeling interictal pathologies, as this directly impacts model performance. This could involve having a group of experts review the same recordings to reach a consensus on the labels. It is also

recommended to release future datasets as continuous recordings with annotations rather than as epochs of data, since publishing clips limits the algorithms to the chosen epoch length and impedes validation on continuous data. Formats like EDF can be used to store the recordings along with the corresponding labels.

- Non-cerebral artifacts labeling should focus on large artifacts that can be confirmed in a clinical setting, with special attention to those that can be misclassified as pathologies when using traditional IEDs and HFOs detectors.
- Further research is needed to determine the best methods for detecting artifacts in iEEG recordings, including the investigation of epoch lengths used as inputs for the models and the normalization techniques used prior to classification. In this study, since the dataset was released in clips, each segment was z-scored separately. This could remove some important features, such as the commonly observed higher amplitude of non-cerebral artifacts compared to baseline activity. Normalization of the data is fundamental for iEEG classification tasks, but other approaches, such as normalizing larger epochs of data, could help the model learn other important time-dependent features like amplitude variation.
- Future implementation of artifact detection could focus on two step approaches, similar to those used for HFOs and IEDs detection. In these approaches, potential events are first identified with a non-machine learning algorithm (for example, a simple thresholding approach), and then machine learning is only used to classify these events. To achieve this, more detailed annotations of the datasets would also be useful, where not only segments of data are annotated as pathological or non-cerebral artifacts but also specify the type of pathology or artifact. This could be similar to how it is done for other datasets like heart arrhythmia datasets, where specific details of the type of arrhythmia are provided [106].
- Addition of spatial information to the inputs could also be beneficial to increase the performance of the developed models. During data assessment with the clinical reviewer, it

was noted that adjacent channels were often loaded, providing information such as the propagation of interictal spikes. Incorporating this data into the model inputs could potentially help the algorithms differentiate between pathological events and non-cerebral artifacts.

Future work in this area should be focused on increasing the quantity and quality of available data before the testing of new algorithms. Inter-institutional cooperation is required to generate datasets that can generalize better. Finally, closer collaboration with the clinical environment is encouraged to better understand how to design the datasets, which could improve agreement between automatic detectors and clinical expert criteria.

4.4 Remaining challenges, future work & limitations of this study

While this thesis made significant contributions towards standardizing preprocessing of iEEG datasets through the development of a software workflow and the investigation of automatic interictal artifact detection, several challenges and limitations remain that must be acknowledged and addressed in future work. Currently, this author is focusing on the following work, which are considered limitations of the thesis:

1. External testing of the workflow

The developed preprocessing workflow has been tested only with local BIDS data. Further testing on available external datasets is necessary to confirm its usability across different research settings.

2. BIDS compliance

Although iEEGPrep receives and validates iEEG-BIDS datasets, the outputs of the model are not yet BIDS compliant. Additional work is needed to ensure both inputs and outputs adhere to BIDS standards, qualifying this tool as a BIDS app.

3. **Limited supported methods**

Several preprocessing steps in iEEGPrep currently support only specific methods or inputs, such as EDF files and re-referencing from monopolar to bipolar schemes. Expanding support to include other BIDS recording formats, like BrainVision Core Data Format, and preprocessing settings, such as common average referencing, is essential for enhancing the workflow's usability.

4. **Scalability and computational resources**

The current version of the pipeline was tested only on the computing clusters of the Digital Research Alliance of Canada, which possess substantial computing resources. It is necessary to test the workflow in environments with fewer computational resources to ensure scalability. Additionally, comparing the computational costs of this workflow with common tools like EEGLAB or Fieldtrip is also necessary.

5. **Better software practices**

To release this workflow as a publicly available tool, improved software practices are needed. These include comprehensive software documentation, tool packaging, and the implementation of continuous integration/continuous development (CI/CD) practices.

Some other limitations and future work currently outside of my scope include:

1. **User dependency**

Although designed for user-friendliness, the effectiveness of the software workflow still relies on the user's ability to correctly set configuration parameters and interpret outputs. Variability in user expertise and interaction with the software may lead to inconsistent data processing results.

2. **Impact of referencing changes on automatic artifact detection**

In this work, to compare the best deep learning solution against the best traditional machine learning model, their performance was evaluated by assessing their generalizability

to an external dataset. However, for this test dataset, a different re-referencing approach (bipolar referencing) was used compared to the training data, which used CAR. This could have had a negative impact on the model performance, as the morphology of the signals might vary depending on the chosen referencing method. This is considered as a limitation of this study, as it depends on the open-datasets found. Future work is needed to assess the direct impact of changes in referencing between training and test data on the model generalizability.

3. Model generalizability

Although it was not the primary goal, the best models for interictal iEEG artifact detection demonstrated poor generalizability in both external and local datasets. Enhancing the collection and annotation of large, inter-institutional datasets is needed. Furthermore, labeling extensive clinical recording segments using machine learning alone is computationally intensive. A two-step approach—initially identifying potential events with less resource-intensive methods, followed by machine learning-based labeling—should be explored to improve the feasibility of automatic artifact detectors.

4. Data sharing practices

As previously mentioned, larger inter-institutional datasets are needed to improve the generalizability of the artifact detectors for interictal iEEG recordings, annotated using consistent approaches. Furthermore, efforts are needed to improve the inter-rater agreement before training of the machine learning models to reduce the potential rater biases in the algorithms.

5. Double-step artifact detection

As a possible future avenue in the field of artifact detection in iEEG, it is worth investigating a two-step approach, where events of interest are first detected using a simple algorithm, such as thresholding techniques, and then machine learning is used to determine whether these events are artifacts or not. This This addresses one of the limitations of

this work which is the segmentation of continuous ieeg recordings into non-overlapping 3-second clips. This could have a direct influence on model performance if events of interest, such as spikes, are incorrectly split between two clips. Furthermore, if the model is trained with the events of interest always at a specific position within the window (e.g., by building the clips around these events), it could help increase the model performance, as it could focus on specific parts of the image instead of also addressing the task of identifying when the event is occurring.

In addition, it would reduce the computational cost compared to running a machine learning model on large continuous datasets and could also allow models to focus only on classifying these events rather than also on identifying them in continuous data.

Bibliography

- [1] Bahman Abdi-Sargezeh, Sepehr Shirani, Saeid Sanei, Clive Cheong Took, Oana Geman, Gonzalo Alarcon, and Antonio Valentin. A review of signal processing and machine learning techniques for interictal epileptiform discharge detection. *Computers in Biology and Medicine*, 168:107782, January 2024.

- [2] Hussam Abou-Al-Shaar, Andrea A. Brock, Bornali Kundu, Dario J. Englot, and John D. Rolston. Increased nationwide use of stereoencephalography for intracranial epilepsy electroencephalography recordings. *Journal of Clinical Neuroscience*, 53:132–134, July 2018.

- [3] U. Rajendra Acharya, H. Fujita, Vidya K. Sudarshan, Shreya Bhat, and Joel E.W. Koh. Application of entropies for automated diagnosis of epilepsy using EEG signals: A review. *Knowledge-Based Systems*, 88:85–96, November 2015.

- [4] Mashaël Aldayel and Abeer Al-Nafjan. A comprehensive exploration of machine learning techniques for EEG-based anxiety detection. *PeerJ Computer Science*, 10:e1829, January 2024.

- [5] M. Alfaro-Ponce and I. Chairez. Continuous and recurrent pattern dynamic neural networks recognition of electrophysiological signals. *Biomedical Signal Processing and Control*, 57:101783, March 2020.

- [6] Ibrahim Aliyu and Chang Gyoon Lim. Selection of optimal wavelet features for epileptic EEG signal classification with LSTM. *Neural Computing and Applications*, 35(2):1077–1097, January 2023.
- [7] Neena Aloysius and M. Geetha. A review on deep convolutional neural networks. In *2017 International Conference on Communication and Signal Processing (ICCSP)*, pages 0588–0592, Chennai, April 2017. IEEE.
- [8] Haleema Anwar, Qudsia Umaira Khan, Natasha Nadeem, Iqra Pervaiz, Muhammad Ali, and Fatima Fayyaz Cheema. Epileptic seizures. *Discoveries*, 8(2):e110, June 2020.
- [9] Ravindra Arya, Francesco T. Mangano, Paul S. Horn, Katherine D. Holland, Douglas F. Rose, and Tracy A. Glauser. Adverse events related to extraoperative invasive EEG monitoring with subdural grid electrodes: A systematic review and meta-analysis. *Epilepsia*, 54(5):828–839, May 2013.
- [10] Elham Bagheri, Justin Dauwels, Brian C. Dean, Chad G. Waters, M. Brandon Westover, and Jonathan J. Halford. Interictal epileptiform discharge characteristics underlying expert interrater agreement. *Clinical Neurophysiology*, 128(10):1994–2005, October 2017.
- [11] N.W. Bailey, M. Biabani, A.T. Hill, A. Miljevic, N.C. Rogasch, B. McQueen, O.W. Murphy, and P.B. Fitzgerald. Introducing RELAX: An automated pre-processing pipeline for cleaning EEG data - Part 1: Algorithm and application to oscillations. *Clinical Neurophysiology*, 149:178–201, May 2023.
- [12] Tonio Ball, Markus Kern, Isabella Mutschler, Ad Aertsen, and Andreas Schulze-Bonhage. Signal quality of simultaneously recorded invasive and non-invasive EEG. *NeuroImage*, 46(3):708–716, July 2009.

- [13] Daniel T. Barkmeier, Aashit K. Shah, Danny Flanagan, Marie D. Atkinson, Rajeev Agarwal, Darren R. Fuerst, Kouros Jafari-Khouzani, and Jeffrey A. Loeb. High inter-reviewer variability of spike detection on intracranial EEG addressed by an automated multi-channel algorithm. *Clinical Neurophysiology*, 123(6):1088–1095, June 2012.
- [14] Fabrice Bartolomei, Agnes Trébuchon, Francesca Bonini, Isabelle Lambert, Martine Gavaret, Marmaduke Woodman, Bernard Giusiano, Fabrice Wendling, and Christian Bénar. What is the concordance between the seizure onset zone and the irritative zone? A SEEG quantified study. *Clinical Neurophysiology*, 127(2):1157–1162, February 2016.
- [15] John M Bernabei, Nishant Sinha, T Campbell Arnold, Erin Conrad, Ian Ong, Akash R Pattnaik, Joel M Stein, Russell T Shinohara, Timothy H Lucas, Dani S Bassett, Kathryn A Davis, and Brian Litt. Normative intracranial EEG maps epileptogenic tissues in focal epilepsy. *Brain*, 145(6):1949–1961, June 2022.
- [16] Nima Bigdely-Shamlo, Tim Mullen, Christian Kothe, Kyung-Min Su, and Kay A. Robbins. The PREP pipeline: standardized preprocessing for large-scale EEG analysis. *Frontiers in Neuroinformatics*, 9, June 2015.
- [17] Benjamin Billot, Douglas N. Greve, Oula Puonti, Axel Thielscher, Koen Van Leemput, Bruce Fischl, Adrian V. Dalca, and Juan Eugenio Iglesias. SynthSeg: Segmentation of brain MRI scans of any contrast and resolution without retraining. *Medical Image Analysis*, 86:102789, May 2023.
- [18] Ross Blair, Zack Michael, Krzysztof J. Gorgolewski, Nell Hardcastle, Teal Hobson-Lowther, David Nishikawa, Suyash Bhogawar, Stefan Appelhoff, Mainak Jas, Brian Grass, Christopher J. Markiewicz, Chris Holdgraf, Alexander Jones, Rohan Goyal, Robert Oostenveld, Gregory Noack, William Triplett, Mikaël Naveau, Matthew Zito, Joke Durnez, Nicolas Traut, Parul Sethi, Yaroslav O. Halchenko, Dimitri Papadopoulos, Joseph Wexler, Taylor Salo, Adam Thomas, Michael Hanke, Soichi Hayashi, David E.

- Warren, Dianne Patterson, Duncan Macleod, Eduard Ort, Franklin Feingold, Horea Christian, Jakub Kaczmarzyk, Russell Poldrack, Travis Riddle, and Wazeer Zulfikar. bids-validator, March 2022.
- [19] Sarah Blum, Nadine S. J. Jacobsen, Martin G. Bleichner, and Stefan Debener. A Riemannian Modification of Artifact Subspace Reconstruction for EEG Artifact Handling. *Frontiers in Human Neuroscience*, 13:141, April 2019.
- [20] György Buzsáki, Costas A. Anastassiou, and Christof Koch. The origin of extracellular fields and currents — EEG, ECoG, LFP and spikes. *Nature Reviews Neuroscience*, 13(6):407–420, June 2012.
- [21] Tianqi Chen and Carlos Guestrin. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 785–794, San Francisco California USA, August 2016. ACM.
- [22] Yuexing Chen and Jiarun Li. Recurrent Neural Networks algorithms and applications. In *2021 2nd International Conference on Big Data & Artificial Intelligence & Software Engineering (ICBASE)*, pages 38–43, Zhuhai, China, September 2021. IEEE.
- [23] Yvonne Y. Chen, Lyndsey Aponik-Gremillion, Eleonora Bartoli, Daniel Yoshor, Sameer A. Sheth, and Brett L. Foster. Stability of ripple events during task engagement in human hippocampus. *Cell Reports*, 35(13):109304, June 2021.
- [24] CMH Lahore Medical College & Institute of Dentistry, Lahore, Pakistan, Haleema Anwar, Qudsia Umaira Khan, CMH Lahore Medical College & Institute of Dentistry, Lahore, Pakistan, Natasha Nadeem, CMH Lahore Medical College & Institute of Dentistry, Lahore, Pakistan, Iqra Pervaiz, CMH Lahore Medical College & Institute of Dentistry, Lahore, Pakistan, Muhammad Ali, CMH Lahore Medical College & Institute of Dentistry, Lahore, Pakistan, Fatima Fayyaz Cheema, and CMH Lahore Medical College &

- Institute of Dentistry, Lahore, Pakistan. Epileptic seizures. *Discoveries*, 8(2):e110, June 2020.
- [25] Anup Das, Carlo De Los Angeles, and Vinod Menon. Electrophysiological foundations of the human default-mode network revealed by intracranial-EEG recordings during resting-state and cognition. *NeuroImage*, 250:118927, April 2022.
- [26] Mahshid Dastgoshadeh and Zahra Rabiei. Detection of epileptic seizures through EEG signals using entropy features and ensemble learning. *Frontiers in Human Neuroscience*, 16:1084061, February 2023.
- [27] Olivier David, Anne-Sophie Job, Luca De Palma, Dominique Hoffmann, Lorella Minotti, and Philippe Kahane. Probabilistic functional tractography of the human cortex. *NeuroImage*, 80:307–317, October 2013.
- [28] Kathryn A. Davis, Seth P. Devries, Abba Krieger, Temenuzhka Mihaylova, Daniela Minecan, Brian Litt, Joost B. Wagenaar, and William C. Stacey. The effect of increased intracranial EEG sampling rates in clinical practice. *Clinical Neurophysiology*, 129(2):360–367, February 2018.
- [29] Alain De Cheveigné. ZapLine: A simple and effective method to remove power line artifacts. *NeuroImage*, 207:116356, February 2020.
- [30] Alain De Cheveigné and Dorothée Arzounian. Robust detrending, rereferencing, outlier detection, and inpainting for multichannel data. *NeuroImage*, 172:903–912, May 2018.
- [31] Alain De Cheveigné and Israel Nelken. Filters: When, Why, and How (Not) to Use Them. *Neuron*, 102(2):280–293, April 2019.
- [32] Alain De Cheveigné and Lucas C. Parra. Joint decorrelation, a versatile tool for multichannel data analysis. *NeuroImage*, 98:487–505, September 2014.

- [33] Arnaud Delorme. EEG is better left alone. *Scientific Reports*, 13(1):2372, February 2023.
- [34] Arnaud Delorme and Scott Makeig. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1):9–21, March 2004.
- [35] Arnaud Delorme and Jeffery A. Martin. Automated Data Cleaning for the Muse EEG. In *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 1–5, Houston, TX, USA, December 2021. IEEE.
- [36] Matteo Demuru, Dorien Van Blooijis, Willemiek Zweiphenning, Dora Hermes, Frans Leijten, Maeike Zijlmans, and on behalf of the RESPECT group. A Practical Workflow for Organizing Clinical Intraoperative and Long-term iEEG Data in BIDS. *Neuroinformatics*, 20(3):727–736, July 2022.
- [37] Zahra Ebrahimi, Mohammad Loni, Masoud Daneshtalab, and Arash Gharehbaghi. A review on deep learning methods for ECG arrhythmia classification. *Expert Systems with Applications: X*, 7:100033, September 2020.
- [38] Jerome Engel. Surgery for Seizures. *New England Journal of Medicine*, 334(10):647–653, March 1996.
- [39] Dario J. Englot and Edward F. Chang. Rates and predictors of seizure freedom in resective epilepsy surgery: an update. *Neurosurgical Review*, 37(3):389–405, July 2014.
- [40] Sébastien Eustace and The Poetry contributors. Poetry: Python packaging and dependency management made easy.
- [41] Andriy Fedorov, Reinhard Beichel, Jayashree Kalpathy-Cramer, Julien Finet, Jean-Christophe Fillion-Robin, Sonia Pujol, Christian Bauer, Dominique Jennings, Fiona Fennessy, Milan Sonka, John Buatti, Stephen Aylward, James V. Miller, Steve Pieper,

- and Ron Kikinis. 3D Slicer as an image computing platform for the Quantitative Imaging Network. *Magnetic Resonance Imaging*, 30(9):1323–1341, November 2012.
- [42] Robert S. Fisher, Carlos Acevedo, Alexis Arzimanoglou, Alicia Bogacz, J. Helen Cross, Christian E. Elger, Jerome Engel, Lars Forsgren, Jacqueline A. French, Mike Glynn, Dale C. Hesdorffer, B.I. Lee, Gary W. Mathern, Solomon L. Moshé, Emilio Perucca, Ingrid E. Scheffer, Torbjörn Tomson, Masako Watanabe, and Samuel Wiebe. ILAE Official Report: A practical clinical definition of epilepsy. *Epilepsia*, 55(4):475–482, April 2014.
- [43] Robert S. Fisher, J. Helen Cross, Jacqueline A. French, Norimichi Higurashi, Edouard Hirsch, Floor E. Jansen, Lieven Lagae, Solomon L. Moshé, Jukka Peltola, Eliane Roulet Perez, Ingrid E. Scheffer, and Sameer M. Zuberi. Operational classification of seizure types by the International League Against Epilepsy: Position Paper of the ILAE Commission for Classification and Terminology. *Epilepsia*, 58(4):522–530, April 2017.
- [44] Robert S. Fisher, Helen E. Scharfman, and Marco deCurtis. How Can We Identify Ictal and Interictal Abnormal Activity? In Helen E. Scharfman and Paul S. Buckmaster, editors, *Issues in Clinical Epileptology: A View from the Bench*, volume 813, pages 3–23. Springer Netherlands, Dordrecht, 2014.
- [45] Vladimir Fonov, Alan C. Evans, Kelly Botteron, C. Robert Almli, Robert C. McKinstry, and D. Louis Collins. Unbiased average age-appropriate atlases for pediatric studies. *NeuroImage*, 54(1):313–327, January 2011.
- [46] Vs Fonov, Ac Evans, Rc McKinstry, Cr Almli, and Dl Collins. Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *NeuroImage*, 47:S102, July 2009.
- [47] Amal Fouad, Hamed Azizollahi, Jean-Eudes Le Douget, François-Xavier Lejeune, Mario Valderrama, Liliana Mayor, Vincent Navarro, and Michel Le Van Quyen. Inter-

ictal epileptiform discharges show distinct spatiotemporal and morphological patterns across wake and sleep. *Brain Communications*, 4(5):fcac183, September 2022.

- [48] Epilepsy Foundation. What is epilepsy? <https://www.epilepsy.com/what-is-epilepsy>. Accessed: 2024-01-07.
- [49] B. Frauscher, D. Mansilla, C. Abdallah, A. Astner-Rohracher, S. Beniczky, M. Brazdil, V. Gnatkovsky, J. Jacobs, G. Kalamangalam, P. Perucca, P. Ryvlin, S. Schuele, J. Tao, Y. Wang, M. Zijlmans, and A. McGonigal. Learn how to interpret and use intracranial EEG findings. *Epileptic Disorders*, 26(1):1–59, February 2024.
- [50] Birgit Frauscher, Fabrice Bartolomei, Katsuhiko Kobayashi, Jan Cimbalnik, Maryse A. Van ‘T Klooster, Stefan Rampp, Hiroshi Otsubo, Yvonne Höller, Joyce Y. Wu, Eishi Asano, Jerome Engel, Philippe Kahane, Julia Jacobs, and Jean Gotman. High-frequency oscillations: The state of clinical research. *Epilepsia*, 58(8):1316–1329, August 2017.
- [51] Birgit Frauscher, Nicolas Von Ellenrieder, Rina Zelmann, Irena Doležalová, Lorella Minotti, André Olivier, Jeffery Hall, Dominique Hoffmann, Dang Khoa Nguyen, Philippe Kahane, François Dubeau, and Jean Gotman. Atlas of the normal intracranial electroencephalogram: neurophysiological awake activity in different cortical areas. *Brain*, 141(4):1130–1144, April 2018.
- [52] Birgit Frauscher, Nicolás Von Ellenrieder, Rina Zelmann, Christine Rogers, Dang Khoa Nguyen, Philippe Kahane, François Dubeau, and Jean Gotman. High-Frequency Oscillations in the Normal Human Brain. *Annals of Neurology*, 84(3):374–385, September 2018.
- [53] Paula Karenina De Macedo Machado Freire, Celso Augusto Guimarães Santos, and Gustavo Barbosa Lima Da Silva. Analysis of the use of discrete wavelet transforms

- coupled with ANN for short-term streamflow forecasting. *Applied Soft Computing*, 80:494–505, July 2019.
- [54] Laurel J. Gabard-Durnam, Adriana S. Mendez Leal, Carol L. Wilkinson, and April R. Levin. The Harvard Automated Processing Pipeline for Electroencephalography (HAPPE): Standardized Processing Software for Developmental and High-Artifact Data. *Frontiers in Neuroscience*, 12:97, February 2018.
- [55] Tapan Gandhi, Bijay Ketan Panigrahi, and Sneha Anand. A comparative study of wavelet families for EEG signal classification. *Neurocomputing*, 74(17):3051–3057, October 2011.
- [56] H. Gastaut. Clinical and Electroencephalographical Classification of Epileptic Seizures. *Epilepsia*, 11(1):102–112, March 1970.
- [57] H. Gastaut, Wf Caveness, H. Landolt, Aml Dehaas, Fl Mcnaughton, O. Magnus, Jk Merlis, Da Pond, J. Radermecker, and Ws Vanleeuwen. A Proposed International Classification of Epileptic Seizures. *EPILEPSIA*, 5(3):297–306, 1964. Web of Science ID: WOS:A1964WY86600007.
- [58] Henri Gastaut and Benjamin G. Zifkin. Classification of The Epilepsies. In *Drugs for the Control of Epilepsy*. CRC Press, 1992.
- [59] David Geng, Ayham Alkhachroum, Manuel A Melo Bicchi, Jonathan R Jagid, Iahn Cajigas, and Zhe Sage Chen. Deep learning for robust detection of interictal epileptiform discharges. *Journal of Neural Engineering*, 18(5):056015, October 2021.
- [60] Cristina Gil Ávila, Felix S. Bott, Laura Tiemann, Vanessa D. Hohn, Elisabeth S. May, Moritz M. Nickel, Paul Theo Zebhauser, Joachim Gross, and Markus Ploner. DISCOVER-EEG: an open, fully automated EEG pipeline for biomarker discovery in clinical neuroscience. *Scientific Data*, 10(1):613, September 2023.

- [61] Stephen V. Gliske, Zachary T. Irwin, Cynthia Chestek, and William C. Stacey. Effect of sampling rate and filter settings on High Frequency Oscillation detections. *Clinical Neurophysiology*, 127(9):3042–3050, September 2016.
- [62] Jorge Gonzalez-Martinez, Juan Bulacio, Andreas Alexopoulos, Lara Jehi, William Bingaman, and Imad Najm. Stereoelectroencephalography in the “difficult to localize” refractory focal epilepsy: Early experience from a North American epilepsy center. *Epilepsia*, 54(2):323–330, February 2013.
- [63] Marco Gori, Alessandro Betti, and Stefano Melacci. Learning principles. In *Machine Learning*, pages 53–111. Elsevier, 2024.
- [64] Alexandre Gramfort. MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, 7, 2013.
- [65] Iris I. A. Groen, Giovanni Piantoni, Stephanie Montenegro, Adeen Flinker, Sasha Devore, Orrin Devinsky, Werner Doyle, Patricia Dugan, Daniel Friedman, Nick F. Ramsey, Natalia Petridou, and Jonathan Winawer. Temporal Dynamics of Neural Responses in Human Visual Cortex. *The Journal of Neuroscience*, 42(40):7562–7580, October 2022.
- [66] Kristin M Gunnarsdottir, Adam Li, Rachel J Smith, Joon-Yi Kang, Anna Korzeniewska, Nathan E Crone, Adam G Rouse, Jennifer J Cheng, Michael J Kinsman, Patrick Landazuri, Utku Uysal, Carol M Ulloa, Nathaniel Cameron, Iahn Cajigas, Jonathan Jagid, Andres Kanner, Turki Elarjani, Manuel Melo Bicchi, Sara Inati, Kareem A Zaghoul, Varina L Boerwinkle, Sarah Wyckoff, Niravkumar Barot, Jorge Gonzalez-Martinez, and Sridevi V Sarma. Source-sink connectivity: a novel interictal EEG marker for seizure localization. *Brain*, 145(11):3901–3915, November 2022.
- [67] Jaime Gómez-Ramírez, Shelagh Freedman, Diego Mateos, José Luis Pérez Velázquez, and Taufik A. Valiante. Exploring the alpha desynchronization hypothesis in resting state

- networks with intracranial electroencephalography and wiring cost estimates. *Scientific Reports*, 7(1):15670, November 2017.
- [68] Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. Array programming with NumPy. *Nature*, 585(7825):357–362, September 2020.
- [69] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. Springer Series in Statistics. Springer New York, New York, NY, 2009.
- [70] Christopher Holdgraf, Stefan Appelhoff, Stephan Bickel, Kristofer Bouchard, Sasha D’Ambrosio, Olivier David, Orrin Devinsky, Benjamin Dichter, Adeen Flinker, Brett L. Foster, Krzysztof J. Gorgolewski, Iris Groen, David Groppe, Aysegul Gunduz, Liberty Hamilton, Christopher J. Honey, Mainak Jas, Robert Knight, Jean-Philippe Lachaux, Jonathan C. Lau, Christopher Lee-Messer, Brian N. Lundstrom, Kai J. Miller, Jeffrey G. Ojemann, Robert Oostenveld, Natalia Petridou, Gio Piantoni, Andrea Pigorini, Nader Pouratian, Nick F. Ramsey, Arjen Stolk, Nicole C. Swann, François Tadel, Bradley Voytek, Brian A. Wandell, Jonathan Winawer, Kirstie Whitaker, Lyuba Zehl, and Dora Hermes. iEEG-BIDS, extending the Brain Imaging Data Structure specification to human intracranial electrophysiology. *Scientific Data*, 6(1):102, June 2019.
- [71] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q. Weinberger. Densely Connected Convolutional Networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2261–2269, Honolulu, HI, July 2017. IEEE.
- [72] Plotly Technologies Inc. Collaborative data science, 2015.

- [73] Julia Jacobs, Maeike Zijlmans, Rina Zelman, Claude-Édouard Chatillon, Jeffrey Hall, André Olivier, François Dubeau, and Jean Gotman. High-frequency electroencephalographic oscillations correlate with outcome of epilepsy surgery. *Annals of Neurology*, 67(2):209–220, February 2010.
- [74] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An Introduction to Statistical Learning: with Applications in R*. Springer Texts in Statistics. Springer US, New York, NY, 2021.
- [75] Mainak Jas, Denis A. Engemann, Yousra Bekhti, Federico Raimondo, and Alexandre Gramfort. Autoreject: Automated artifact rejection for MEG and EEG data. *NeuroImage*, 159:417–429, October 2017.
- [76] Maciej Jedynak, Anthony Boyer, Manuel Mercier, Blandine Chanteloup-Forêt, Manik Bhattacharjee, Philippe Kahane, and Olivier David. SEEG electrode shaft affects amplitude and latency of potentials evoked with single pulse electrical stimulation. *Journal of Neuroscience Methods*, 403:110035, March 2024.
- [77] Michael A. Jensen, Harvey Huang, Gabriela Ojeda Valencia, Bryan T. Klassen, Max A. Van Den Boom, Timothy J. Kaufmann, Gerwin Schalk, Peter Brunner, Gregory A. Worrell, Dora Hermes, and Kai J. Miller. A motor association area in the depths of the central sulcus. *Nature Neuroscience*, 26(7):1165–1169, July 2023.
- [78] Karim Jerbi, Samson Freyermuth, Sarang Dalal, Philippe Kahane, Olivier Bertrand, Alain Berthoz, and Jean-Philippe Lachaux. Saccade Related Gamma-Band Activity in Intracerebral EEG: Dissociating Neural from Ocular Muscle Activity. *Brain Topography*, 22(1):18–23, June 2009.
- [79] Haiteng Jiang, Vasileios Kokkinos, Shuai Ye, Alexandra Urban, Anto Bagić, Mark Richardson, and Bin He. Interictal SEEG Resting-State Connectivity Localizes the

- Seizure Onset Zone and Predicts Seizure Outcome. *Advanced Science*, 9(18):2200887, June 2022.
- [80] Barbara C. Jobst, Fabrice Bartolomei, Beate Diehl, Birgit Frauscher, Philippe Kahane, Lorella Minotti, Ashwini Sharan, Nastasia Tardy, Gregory Worrell, and Jean Gotman. Intracranial EEG in the 21st Century. *Epilepsy Currents*, 20(4):180–188, July 2020.
- [81] Holger Joswig, Jonathan C Lau, Mahmoud Abdallat, Andrew G Parrent, Keith W MacDougall, Richard S McLachlan, Jorge G Burneo, and David A Steven. Stereoelectroencephalography Versus Subdural Strip Electrode Implantations: Feasibility, Complications, and Outcomes in 500 Intracranial Monitoring Cases for Drug-Resistant Epilepsy. *Neurosurgery*, 87(1):E23–E30, July 2020.
- [82] Giridhar P. Kalamangalam and Mircea I. Chelaru. Functional Connectivity in Dorsolateral Frontal Cortex: Intracranial Electroencephalogram Study. *Brain Connectivity*, 11(10):850–864, December 2021.
- [83] Nick Kane, Jayant Acharya, Sandor Beniczky, Luis Caboclo, Simon Finnigan, Peter W. Kaplan, Hiroshi Shibasaki, Ronit Pressler, and Michel J.A.M. Van Putten. A revised glossary of terms most commonly used by clinical electroencephalographers and updated proposal for the report format of the EEG findings. Revision 2017. *Clinical Neurophysiology Practice*, 2:170–185, 2017.
- [84] Mark R Keezer, Sanjay M Sisodiya, and Josemir W Sander. Comorbidities of epilepsy: current concepts and future perspectives. *The Lancet Neurology*, 15(1):106–115, January 2016.
- [85] Karolin Kerber, Matthias Dümpelmann, Björn Schelter, Pierre Le Van, Rudolf Korinthenberg, Andreas Schulze-Bonhage, and Julia Jacobs. Differentiation of specific ripple patterns helps to identify epileptogenic areas for surgical procedures. *Clinical Neurophysiology*, 125(7):1339–1345, July 2014.

- [86] Mohammad Reza Keshtkaran and Zhi Yang. A fast, robust algorithm for power line interference cancellation in neural recording. *Journal of Neural Engineering*, 11(2):026017, April 2014.
- [87] Hui Ming Khoo, Nicolás Von Ellenrieder, Natalja Zazubovits, Daniel He, François Dubeau, and Jean Gotman. The spike onset zone: The region where epileptic spikes start and from where they propagate. *Neurology*, 91(7), August 2018.
- [88] Lohith G Kini, John M Bernabei, Fadi Mikhail, Peter Hadar, Preya Shah, Ankit N Khambhati, Kelly Oechsel, Ryan Archer, Jacqueline Boccanfuso, Erin Conrad, Russell T Shinohara, Joel M Stein, Sandhitsu Das, Ammar Kheder, Timothy H Lucas, Kathryn A Davis, Danielle S Bassett, and Brian Litt. Virtual resection predicts surgical outcome for drug-resistant epilepsy. *Brain*, 142(12):3892–3905, December 2019.
- [89] Marius Klug and Niels A. Kloosterman. Zapline-plus: A Zapline extension for automatic and adaptive removal of frequency-specific noise artifacts in M/EEG. *Human Brain Mapping*, 43(9):2743–2758, June 2022.
- [90] Patrick Kwan, Alexis Arzimanoglou, Anne T. Berg, Martin J. Brodie, W. Allen Hauser, Gary Mathern, Solomon L. Moshé, Emilio Perucca, Samuel Wiebe, and Jacqueline French. Definition of drug resistant epilepsy: Consensus proposal by the ad hoc Task Force of the ILAE Commission on Therapeutic Strategies. *Epilepsia*, 51(6):1069–1077, June 2010.
- [91] Dakun Lai, Xinyue Zhang, Kefei Ma, Zichu Chen, Wenjing Chen, Heng Zhang, Han Yuan, and Lei Ding. Automated Detection of High Frequency Oscillations in Intracranial EEG Using the Combination of Short-Time Energy and Convolutional Neural Networks. *IEEE Access*, 7:82501–82511, 2019.

- [92] Eric Larson, Alexandre Gramfort, Denis A Engemann, Jaakko Leppakangas, Christian Brodbeck, Mainak Jas, Teon Brooks, Jona Sassenhagen, Martin Luessi, Daniel McCloy, Jean-Remi King, Richard Höchenberger, Roman Goj, Guillaume Favelier, Clemens Brunner, Marijn van Vliet, Mark Wronkiewicz, Chris Holdgraf, Alex Rockhill, Joan Massich, Yousra Bekhti, Mathieu Scheltienne, Stefan Appelhoff, Alan Leggitt, Andrew Dykstra, Rob Luke, Romain Trachel, Lorenzo De Santis, Asish Panda, Mikołaj Magnuski, Britta Westner, Martin Billinger, Dan G Wakeman, Daniel Strohmeier, Hari Bharadwaj, Tal Linzen, Alexandre Barachant, Emily Ruzich, Christopher J Bailey, Adam Li, Clément Moutard, Luke Bloy, Fede Raimondo, Jussi Nurminen, Jair Montoya, Marmaduke Woodman, Ingo Lee, Martin Schulz, Nick Foti, Cathy Nangini, José C García Alanis, Scott Huberty, Olaf Hauk, Ross Maddox, Dimitri Papadopoulos Orfanos, Roan LaPlante, Ashley Drew, Christoph Dinh, Guillaume Dumas, Johann Benerradi, Thomas Hartmann, Eduard Ort, Paul Pasler, Stefan Reppinger, Alexander Rudiuk, Ana Radanovic, Brad Buran, Mathurin Massias, Matti Hämäläinen, Praveen Sripad, Valerii Chirkov, Christopher Mullins, Félix Raimundo, Phillip Alday, Ram Pari, Simon Kornblith, Yaroslav Halchenko, Yu-Han Luo, Johannes Kasper, Keith Doelling, Mads Jensen, Tanay Gahlot, Adonay Nunes, Alexandre Gramfort, Dirk Güttlin, kjs, Alejandro Weinstein, Camilo Lamus, Catalina María Galván, Cristóbal Moënné-Locco, Dmitrii Altukhov, Erica Peterson, Erkka Heinila, Jevri Hanna, Jon Houck, Michiru Kaneda, Natalie Klein, Paul Roujansky, Simon Kern, Antti Rantala, Burkhard Maess, Christian O'Reilly, Henrich Kolkhorst, Hubert Banville, Jack Zhang, Jacob Woessner, Kostiantyn Maksymenko, Maggie Clarke, Matteo Anelli, Nikolai Chapochnikov, Pierre-Antoine Bannier, Saket Choudhary, Carina Forster, Cora Kim, Felix Klotzsche, Fu-Te Wong, Ivana Kojcic, Jesper Duemose Nielsen, Kaisu Lankinen, Kambiz Tabavi, Louis Thibault, Moritz Gerster, Nathalie Gayraud, Nick Ward, Santeri Ruuskanen, Ana Radanovic, Andrew Quinn, Antoine Gauthier, Basile Pinsard, Dominik Welke, Dominik Welke, Emily Stephen, Erik Hornberger, Evan Hathaway, Evgenii Kalenkovich, Fahimeh Mamashli,

Giorgio Marinato, Hafeza Anevar, Jan Sosulski, Jeff Stout, Joshua Calder-Travis, Larry Eisenman, Lorenz Esch, Marian Dovgialo, Nicolas Barascud, Nicolas Legrand, Rotem Falach, Samuel Deslauriers-Gauthier, Silvia Cotroneo, Steve Matindi, Steven Bierer, Victor Férat, Victoria Peterson, Zvi Baratz, Alessandro Tonin, Alexander Kovrig, Annalisa Pascarella, Apoorva Karekal, Carlos de la Torre, Chetan Gohil, Christina Zhao, Dominik Krzemiński, Dominique Makowski, Ezequiel Mikulan, Gennadiy Belonosov, George O'Neill, Hamza Abdelhedi, Jean-Baptiste Schiratti, Jen Evans, John Veillette, Jordan Drew, Joshua Teves, Judy D Zhu, Kristijan Armeni, Kyle Mathewson, Laura Gwilliams, Lenny Varghese, Lukas Gemein, Lukas Hecker, Lx37, Mats van Es, Matt Boggess, Matthias Eberlein, Mohamed Sherif, Nataliia Kozhemiako, Naveen Srinivasan, Niklas Wilming, Oleh Kozynets, Pierre Ablin, Quentin Bertrand, Reza Shoorangiz, Rodrigo Hübner, Sara Sommariva, Sena Er, Sheraz Khan, Sophie Herbst, Sumalyo Datta, Theodore Papadopoulo, Thomas Jochmann, Thomas Samuel Binns, Timon Merk, Tod Flak, Tom Dupré la Tour, Tristan Stenner, Tziona NessAiver, akshay0724, sviter, Aaron Earle-Richardson, Abram Hindle, Achilleas Koutsou, Adeline Fecker, Adina Wagner, Alex Ciok, Andy Gilbert, Aniket Pradhan, Anna Padee, Anne-Sophie Dubarry, Anton Nikolas Waniek, Archit Singhal, Ariel Rokem, Arne Pelzer, Austin Hurst, Ben Beasley, Bruno Nicenboim, Carlos de la Torre, Christian Clauss, Christian Mista, Chun-Hui Li, Claire Braboszcz, Daniel Carlström Schad, Daniel Hasegan, Daniel Tse, Darin Erat Sleiter, David Haslacher, David Sabbagh, Demetres Kostas, Desislava Petkova, Dinara Issagaliyeva, Diptyajit Das, Dominik Wetzel, Eberhard Eich, Elizabeth DuPre, Ellen Lau, Emanuele Olivetti, Enrico Varano, Enzo Altamiranda, Eric Brayet, Etienne de Montalivet, Evgeny Goldstein, Federico Zamberlan, Florin Pop, Frederik D Weber, Gansheng Tan, Geoff Brookshire, George O'Neill, Giulio, Gonzalo Reina, Hamid Maymandi, Hermann Sonntag, Hongjiang Ye, Hyonyoung Shin, Hüseyin Orkun Elmas, Ilias Machairas, Ivan Skelin, Ivan Zubarev, Jakub Kaczmarzyk, Jan Zerfowski, Jasper J F van den Bosch, Jeroen Van Der Donckt, Johan van der Meer,

Johannes Niediek, Josh Koen, Joshua J Bear, Juergen Dammers, Julia Guiomar Niso Galán, Julius Welzel, Katarina Slama, Katrin Leinweber, Laetitia Grabot, Lau Møller Andersen, Leonardo S Barbosa, Liberty Hamilton, Lorenzo Alfine, Lukáš Hejtmánek, Maksym Balatsko, Manfred Kitzbichler, Manoj Kumar, Manorama Kadwani, Manu Sutela, Marcin Koculak, Mark Alexander Henney, Martin Schulz, Martin van Harmelen, MartinBaBer, Matt Courtemanche, Matt Tucker, Matteo Visconti di Oleggio Castello, Matthias Dold, Matti Toivonen, Maureen Shader, Mauricio Cespedes, Michael Krause, Milan Rybář, Mingjian He, Mohammad Daneshzand, Nicolas Gensollen, Nicole Proulx, Niels Focke, Nikolas Chalas, Omer Shubi, Pablo Mainar, Padma Sundaram, Pedro Silva, Peter J Molfese, Proloy Das, Qian Chu, Quanliang Li, Quentin Barthélemy, Rahul Nadkarni, Ramiro Gatti, Ramonapariciog Apariciogarcia, Rasmus Aagaard, Reza Nasri, Richard Koehler, Riessarius Stargardsky, Robert Oostenveld, Robert Seymour, Robin Tibor Schirrmester, Ryan Law, Sagun Pai, Sam Perry, Samuel Louviot, Sawradip Saha, Sebastiaan Mathot, Sebastian Major, Sebastien Treguer, Sebastián Castaño, Senwen Deng, Sergey Antopolskiy, Simeon Wong, Simeon Wong, Simon-Shlomo Poil, Sondre Foslien, Sourav Singh, Stanislas Chambon, Steven Bethard, Steven M Gutstein, Svea Marie Meyer, T Wang, Thomas Donoghue, Thomas Moreau, Thomas Radman, Timothy Gates, Tom Ma, Tom Stone, Tommy Clausner, Toomas Erik Anijärv, Xiaokai Xia, Yiping Zuo, Zhi Zhang, buildqa, and luzpaz. MNE-Python, January 2024.

- [93] Gregory Lee, Ralf Gommers, Filip Waselewski, Kai Wohlfahrt, and Aaron O’Leary. PyWavelets: A Python package for wavelet analysis. *Journal of Open Source Software*, 4(36):1237, April 2019.
- [94] Scott F. Lempka and Cameron C. McIntyre. Theoretical Analysis of the Local Field Potential in Deep Brain Stimulation Applications. *PLoS ONE*, 8(3):e59839, March 2013.
- [95] John D. Lewis, Gleb Bezgin, Vladimir S. Fonov, D. Louis Collins, and Alan C. Evans. A sub+cortical fMRI-based surface parcellation. *Human Brain Mapping*, 43(2):616–632,

February 2022.

- [96] Adam Li, Chester Huynh, Zachary Fitzgerald, Iahn Cajigas, Damian Brusko, Jonathan Jagid, Angel O. Claudio, Andres M. Kanner, Jennifer Hopp, Stephanie Chen, Jennifer Haagensen, Emily Johnson, William Anderson, Nathan Crone, Sara Inati, Kareem A. Zaghloul, Juan Bulacio, Jorge Gonzalez-Martinez, and Sridevi V. Sarma. Neural fragility as an EEG marker of the seizure onset zone. *Nature Neuroscience*, 24(10):1465–1474, October 2021.
- [97] Guangye Li, Shize Jiang, Sivylla E. Paraskevopoulou, Meng Wang, Yang Xu, Zehan Wu, Liang Chen, Dingguo Zhang, and Gerwin Schalk. Optimal referencing for stereo-electroencephalographic (SEEG) recordings. *NeuroImage*, 183:327–335, December 2018.
- [98] Shengjie Liu, Guangye Li, Shize Jiang, Xiaolong Wu, Jie Hu, Dingguo Zhang, and Liang Chen. Investigating Data Cleaning Methods to Improve Performance of Brain–Computer Interfaces Based on Stereo-Electroencephalography. *Frontiers in Neuroscience*, 15:725384, October 2021.
- [99] Sarah Long, Maria Bruzzone, Sotiris Mitropanopoulos, Giridhar Kalamangalam, and Aysegul Gunduz. Identification and classification of pathology and artifacts for human intracranial cognitive research. *NeuroImage*, 270:119961, April 2023.
- [100] A. V. Medvedev, G. I. Agoureeva, and A. M. Murro. A Long Short-Term Memory neural network for the detection of epileptiform spikes and high frequency oscillations. *Scientific Reports*, 9(1):19374, December 2019.
- [101] Steven L. Meisler, Michael J. Kahana, and Youssef Ezzyat. Does data cleaning improve brain state classification? *Journal of Neuroscience Methods*, 328:108421, December 2019.

- [102] Manuel R. Mercier, Anne-Sophie Dubarry, François Tadel, Pietro Avanzini, Nikolai Axmacher, Dillan Cellier, Maria Del Vecchio, Liberty S. Hamilton, Dora Hermes, Michael J. Kahana, Robert T. Knight, Anais Llorens, Pierre Megevand, Lucia Melloni, Kai J. Miller, Vitória Piai, Aina Puce, Nick F Ramsey, Caspar M. Schwiedrzik, Sydney E. Smith, Arjen Stolk, Nicole C. Swann, Mariska J Vansteensel, Bradley Voytek, Liang Wang, Jean-Philippe Lachaux, and Robert Oostenveld. Advances in human intracranial electroencephalography research, guidelines and good practices. *NeuroImage*, 260:119438, October 2022.
- [103] Sebastian Michelmann, Matthias S. Treder, Benjamin Griffiths, Casper Kerrén, Frédéric Roux, Maria Wimber, David Rollings, Vijay Sawlani, Ramesh Chelvarajah, Stephanie Gollwitzer, Gernot Kreiselmeier, Hajo Hamer, Howard Bowman, Bernhard Staesina, and Simon Hanslmayr. Data-driven re-referencing of intracranial EEG based on independent component analysis (ICA). *Journal of Neuroscience Methods*, 307:125–137, September 2018.
- [104] Nykan Mirchi, Nebras M. Warsi, Frederick Zhang, Simeon M. Wong, Hrishikesh Suresh, Karim Mithani, Lauren Erdman, and George M. Ibrahim. Decoding Intracranial EEG With Machine Learning: A Systematic Review. *Frontiers in Human Neuroscience*, 16:913777, June 2022.
- [105] Tonmoy Monsoor, Yipeng Zhang, Atsuro Daida, Shingo Oana, Qiuqing Lu, Shaun A. Hussain, Aria Fallah, Raman Sankar, Richard J. Staba, William Speier, Vwani Roychowdhury, and Hiroki Nariai. Optimizing detection and deep learning-based classification of pathological high-frequency oscillations in epilepsy. *Clinical Neurophysiology*, 154:129–140, October 2023.
- [106] G.B. Moody and R.G. Mark. The impact of the MIT-BIH Arrhythmia Database. *IEEE Engineering in Medicine and Biology Magazine*, 20(3):45–50, June 2001.

- [107] Tim R. Mullen, Christian A. E. Kothe, Yu Mike Chi, Alejandro Ojeda, Trevor Kerth, Scott Makeig, Tzyy-Ping Jung, and Gert Cauwenberghs. Real-time neuroimaging and cognitive monitoring using wearable dry EEG. *IEEE Transactions on Biomedical Engineering*, 62(11):2553–2567, November 2015.
- [108] Felix Mölder, Kim Philipp Jablonski, Brice Letcher, Michael B. Hall, Christopher H. Tomkins-Tinch, Vanessa Sochat, Jan Forster, Soohyun Lee, Sven O. Twardziok, Alexander Kanitz, Andreas Wilm, Manuel Holtgrewe, Sven Rahmann, Sven Nahnsen, and Johannes Köster. Sustainable data analysis with Snakemake. *F1000Research*, 10:33, April 2021.
- [109] Holger Nahrstaedt, Simon Kern, Luca Cerina, Stefan Appelhoff, David T.H. Kao, Craig Franklin, Shannon Clarke, Jukka Zitting, David Ojeda, Chadwick Boulay, Ogi Moore, and Daniil Kosachev. holgern/pyedflib: v0.1.23, February 2020.
- [110] Hiroki Nariai, Joyce Y. Wu, Danilo Bernardo, Aria Fallah, Raman Sankar, and Shaun A. Hussain. Interrater reliability in visual identification of interictal high-frequency oscillations on electrocorticography and scalp EEG. *Epilepsia Open*, 3(S2):127–132, December 2018.
- [111] Petr Nejedly, Jan Cimbalnik, Petr Klimes, Filip Plesinger, Josef Halamek, Vaclav Kremen, Ivo Viscor, Benjamin H. Brinkmann, Martin Pail, Milan Brazdil, Gregory Worrell, and Pavel Jurak. Intracerebral EEG Artifact Identification Using Convolutional Neural Networks. *Neuroinformatics*, 17(2):225–234, April 2019.
- [112] Petr Nejedly, Vaclav Kremen, Vladimir Sladky, Jan Cimbalnik, Petr Klimes, Filip Plesinger, Filip Mivalt, Vojtech Travnicek, Ivo Viscor, Martin Pail, Josef Halamek, Benjamin H. Brinkmann, Milan Brazdil, Pavel Jurak, and Gregory Worrell. Multicenter intracranial EEG dataset for classification of graphoelements and artifactual signals. *Scientific Data*, 7(1):179, June 2020.

- [113] Guiomar Niso, Krzysztof J. Gorgolewski, Elizabeth Bock, Teon L. Brooks, Guillaume Flandin, Alexandre Gramfort, Richard N. Henson, Mainak Jas, Vladimir Litvak, Jeremy T. Moreau, Robert Oostenveld, Jan-Mathijs Schoffelen, Francois Tadel, Joseph Wexler, and Sylvain Baillet. MEG-BIDS, the brain imaging data structure extended to magnetoencephalography. *Scientific Data*, 5(1):180110, June 2018.
- [114] Robert Oostenveld, Pascal Fries, Eric Maris, and Jan-Mathijs Schoffelen. FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Computational Intelligence and Neuroscience*, 2011:1–9, 2011.
- [115] World Health Organization. Epilepsy. <https://www.who.int/news-room/fact-sheets/detail/epilepsy>. Accessed: 2024-01-07.
- [116] Hiroshi Otsubo, Ayako Ochi, Katsumi Imai, Tomoyuki Akiyama, Ayataka Fujimoto, Cristina Go, Peter Dirks, and Elizabeth J. Donner. High-frequency oscillations of ictal muscle activity and epileptogenic discharges on intracranial EEG in a temporal lobe epilepsy patient. *Clinical Neurophysiology*, 119(4):862–868, April 2008.
- [117] Yulia S. Panina, Elena E. Timechko, Anna A. Usoltseva, Kristina D. Yakovleva, Elena A. Kantimirova, and Diana V. Dmitrenko. Biomarkers of Drug Resistance in Temporal Lobe Epilepsy in Adults. *Metabolites*, 13(1):83, January 2023.
- [118] Josef Parvizi and Sabine Kastner. Promises and limitations of human intracranial electroencephalography. *Nature Neuroscience*, 21(4):474–483, April 2018.
- [119] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in pytorch. 2017.
- [120] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau,

- M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.
- [121] Cyril R. Pernet, Stefan Appelhoff, Krzysztof J. Gorgolewski, Guillaume Flandin, Christophe Phillips, Arnaud Delorme, and Robert Oostenveld. EEG-BIDS, an extension to the brain imaging data structure for electroencephalography. *Scientific Data*, 6(1):103, June 2019.
- [122] Peter Van Dyken, Ali Khan, Tristan Kuehn, Jason Kai, Remi Gau, Alaa Taha, and Mohamed Yousif. `khanlab/snakebids: 0.11.2`, February 2024.
- [123] Huyen Trang Phan, Ngoc Thanh Nguyen, and Dosam Hwang. Convolutional attention neural network over graph structures for improving the performance of aspect-level sentiment analysis. *Information Sciences*, 589:416–439, April 2022.
- [124] Luca Pion-Tonachini, Ken Kreutz-Delgado, and Scott Makeig. The ICLabel dataset of electroencephalographic (EEG) independent component (IC) features. *Data in Brief*, 25:104101, August 2019.
- [125] Jan Pyrzowski, Jean-Eudes Le Douget, Amal Fouad, Mariusz Siemiński, Joanna Jędrzejczak, and Michel Le Van Quyen. Zero-crossing patterns reveal subtle epileptiform discharges in the scalp EEG. *Scientific Reports*, 11(1):4128, February 2021.
- [126] Lucia Rita Quitadamo, Roberto Mai, Francesca Gozzo, Veronica Pelliccia, Francesco Cardinale, and Stefano Seri. Kurtosis-Based Detection of Intracranial High-Frequency Oscillations for the Identification of the Seizure Onset Zone. *International Journal of Neural Systems*, 28(07):1850001, September 2018.
- [127] Robert J. Quon, Stephen Meisenhelter, Edward J. Camp, Markus E. Testorf, Yinchun Song, Qingyuan Song, George W. Culler, Payam Moein, and Barbara C. Jobst. AiED: Artificial intelligence for the detection of intracranial interictal epileptiform discharges. *Clinical Neurophysiology*, 133:1–8, January 2022.

- [128] Chaturbhuj Rathore and Kurupath Radhakrishnan. Concept of epilepsy surgery and presurgical evaluation. *Epileptic Disorders*, 17(1):19–31, March 2015.
- [129] Alexander P Rockhill, Alessandra Mantovani, Brittany Stedelin, Caleb S Nerison, Ahmed M Raslan, and Nicole C Swann. Stereo-EEG recordings extend known distributions of canonical movement-related oscillations. *Journal of Neural Engineering*, 20(1):016007, February 2023.
- [130] F. Rosenow. Presurgical evaluation of epilepsy. *Brain*, 124(9):1683–1700, September 2001.
- [131] Philippe Ryvlin, J Helen Cross, and Sylvain Rheims. Epilepsy surgery in children and adults. *The Lancet Neurology*, 13(11):1114–1126, November 2014.
- [132] Shah T Sarmast, Abba Musa Abdullahi, and Nusrat Jahan. Current Classification of Seizures and Epilepsies: Scope, Limitations and Recommendations for Future Action. *Cureus*, September 2020.
- [133] Ingrid E. Scheffer, Samuel Berkovic, Giuseppe Capovilla, Mary B. Connolly, Jacqueline French, Laura Guilhoto, Edouard Hirsch, Satish Jain, Gary W. Mathern, Solomon L. Moshé, Douglas R. Nordli, Emilio Perucca, Torbjörn Tomson, Samuel Wiebe, Yue-Hua Zhang, and Sameer M. Zuberi. Ilae classification of the epilepsies: Position paper of the Commission for Classification and Terminology. *Epilepsia*, 58(4):512–521, April 2017.
- [134] Nicolina Sciaraffa, Manousos A. Klados, Gianluca Borghini, Gianluca Di Flumeri, Fabio Babiloni, and Pietro Aricò. Double-Step Machine Learning Based Procedure for HFOs Detection and Classification. *Brain Sciences*, 10(4):220, April 2020.
- [135] Skipper Seabold and Josef Perktold. statsmodels: Econometric and statistical modeling with python. In *9th Python in Science Conference*, 2010.

- [136] Rajeev Sharma, Ram Pachori, and U. Acharya. Application of Entropy Measures on Intrinsic Mode Functions for the Automated Identification of Focal Electroencephalogram Signals. *Entropy*, 17(2):669–691, February 2015.
- [137] S. D. Shorvon and D. M. G. Goodridge. Longitudinal cohort studies of the prognosis of epilepsy: contribution of the National General Practice Study of Epilepsy and other studies. *Brain*, 136(11):3497–3510, November 2013.
- [138] Nishant Sinha, John S. Duncan, Beate Diehl, Fahmida A. Chowdhury, Jane De Tisi, Anna Miserocchi, Andrew William McEvoy, Kathryn A. Davis, Sjoerd B. Vos, Gavin P. Winston, Yujiang Wang, and Peter Neal Taylor. Intracranial EEG Structure-Function Coupling and Seizure Outcomes After Epilepsy Surgery. *Neurology*, 101(13), September 2023.
- [139] Aaron M. Spring, Daniel J. Pittman, Yahya Aghakhani, Jeffrey Jirsch, Neelan Pillay, Luis E. Bello-Espinosa, Colin Josephson, and Paolo Federico. Interrater reliability of visually evaluated high frequency oscillations. *Clinical Neurophysiology*, 128(3):433–441, March 2017.
- [140] Arjen Stolk, Sandon Griffin, Roemer Van Der Meij, Callum Dewar, Ignacio Saez, Jack J. Lin, Giovanni Piantoni, Jan-Mathijs Schoffelen, Robert T. Knight, and Robert Oostenveld. Integrated analysis of anatomical and electrophysiological human intracranial data. *Nature Protocols*, 13(7):1699–1723, July 2018.
- [141] Stefan Sumsky and L. John Greenfield. Network analysis of preictal iEEG reveals changes in network structure preceding seizure onset. *Scientific Reports*, 12(1):12526, July 2022.
- [142] Congzhong Sun and Chaozhou Mou. Survey on the research direction of EEG-based signal processing. *Frontiers in Neuroscience*, 17:1203059, July 2023.

- [143] Yulin Sun, Weipeng Jin, Xiaopeng Si, Xingjian Zhang, Jiale Cao, Le Wang, Shaoya Yin, and Dong Ming. Continuous Seizure Detection Based on Transformer and Long-Term iEEG. *IEEE Journal of Biomedical and Health Informatics*, 26(11):5418–5427, November 2022.
- [144] Dwi Sunaryono, Joko Siswantoro, Riyanarto Sarno, Rahardian Indarto Susilo, and Shoffi Izza Sabilla. Epilepsy Detection using Combination DWT and Convolutional Neural Networks Based on Electroencephalogram. In *2023 International Seminar on Intelligent Technology and Its Applications (ISITIA)*, pages 411–416, Surabaya, Indonesia, July 2023. IEEE.
- [145] Piyush Swami, Tapan K. Gandhi, Bijaya K. Panigrahi, Manjari Tripathi, and Snehanand. A novel robust diagnostic model to detect seizures in electroencephalography. *Expert Systems with Applications*, 56:116–130, September 2016.
- [146] Andrea Szelényi, Lorenzo Bello, Hugues Duffau, Enrica Fava, Guenther C. Feigl, Miroslav Galanda, Georg Neuloh, Francesco Signorelli, and Francesco Sala. Intraoperative electrical stimulation in awake craniotomy: methodological aspects of current practice. *Neurosurgical Focus*, 28(2):E7, February 2010.
- [147] François Tadel, Sylvain Baillet, John C. Mosher, Dimitrios Pantazis, and Richard M. Leahy. Brainstorm: A User-Friendly Application for MEG/EEG Analysis. *Computational Intelligence and Neuroscience*, 2011:1–13, 2011.
- [148] Jean Talairach, Pierre Tournoux, and Mark Rayport. *Co-planar stereotaxic atlas of the human brain : 3-dimensional proportional system : an approach to cerebral imaging*. Thieme, Stuttgart, 1 edition, 1988.
- [149] Peter N Taylor, Christoforos A Papasavvas, Thomas W Owen, Gabrielle M Schroeder, Frances E Hutchings, Fahmida A Chowdhury, Beate Diehl, John S Duncan, Andrew W McEvoy, Anna Miserocchi, Jane De Tisi, Sjoerd B Vos, Matthew C Walker, and Yujiang

- Wang. Normative brain mapping of interictal intracranial EEG to localize epileptogenic tissue. *Brain*, 145(3):939–949, April 2022.
- [150] Brandon J. Thio, Aman S. Aberra, Grace E. Dessert, and Warren M. Grill. Ideal current dipoles are appropriate source representations for simulating neurons for intracranial recordings. *Clinical Neurophysiology*, 145:26–35, January 2023.
- [151] Aljoscha Thomschewski, Nathalie Gerner, Patrick B. Langthaler, Eugen Trinkka, Arne C. Bathke, Jürgen Fell, and Yvonne Höller. Automatic vs. Manual Detection of High Frequency Oscillations in Intracranial Recordings From the Human Temporal Lobe. *Frontiers in Neurology*, 11:563577, October 2020.
- [152] Amy Trongnetrpunya, Bijurika Nandi, Daesung Kang, Bernat Kocsis, Charles E. Schroeder, and Mingzhou Ding. Assessing Granger Causality in Electrophysiological Data: Removing the Adverse Effects of Common Signals via Bipolar Derivations. *Frontiers in Systems Neuroscience*, 9, January 2016.
- [153] José F. Téllez-Zenteno, Raj Dhar, and Samuel Wiebe. Long-term seizure outcomes following epilepsy surgery: a systematic review and meta-analysis. *Brain*, 128(5):1188–1198, May 2005.
- [154] Stéfan Van Der Walt, Johannes L. Schönberger, Juan Nunez-Iglesias, François Boulogne, Joshua D. Warner, Neil Yager, Emmanuelle Gouillart, and Tony Yu. scikit-image: image processing in Python. *PeerJ*, 2:e453, June 2014.
- [155] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C. J. Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert

- Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, and Paul van Mulbregt. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3):261–272, March 2020.
- [156] Yujiang Wang, Gabrielle M. Schroeder, Jonathan J. Horsley, Mariella Panagiotopoulou, Fahmida A. Chowdhury, Beate Diehl, John S. Duncan, Andrew W. McEvoy, Anna Misrocchi, Jane De Tisi, and Peter N. Taylor. Temporal stability of intracranial electroencephalographic abnormality maps for localizing epileptogenic tissue. *Epilepsia*, 64(8):2070–2080, August 2023.
- [157] Elaine C. Wirrell, Rima Nabbout, Ingrid E. Scheffer, Taoufik Alsaadi, Alicia Bogacz, Jacqueline A. French, Edouard Hirsch, Satish Jain, Sunao Kaneko, Kate Riney, Pauline Samia, O. Carter Snead, Ernest Somerville, Nicola Specchio, Eugen Trinka, Sameer M. Zuberi, Simona Balestrini, Samuel Wiebe, J. Helen Cross, Emilio Perucca, Solomon L. Moshé, and Paolo Tinuper. Methodology for classification and definition of epilepsy syndromes with list of syndromes: Report of the ILAE Task Force on Nosology and Definitions. *Epilepsia*, 63(6):1333–1348, June 2022.
- [158] Simeon M. Wong, Olivia N. Arski, Adriana M. Workewych, Elizabeth Donner, Ayako Ochi, Hiroshi Otsubo, O. Carter Snead, and George M. Ibrahim. Detection of high-frequency oscillations in electroencephalography: A scoping review and an adaptable open-source framework. *Seizure*, 84:23–33, January 2021.
- [159] Tal Yarkoni, Christopher Markiewicz, Alejandro De La Vega, Krzysztof Gorgolewski, Taylor Salo, Yaroslav Halchenko, Quinten McNamara, Krista DeStasio, Jean-Baptiste Poline, Dmitry Petrov, Valérie Hayot-Sasson, Dylan Nielson, Johan Carlin, Gregory Kiar, Kirstie Whitaker, Elizabeth DuPre, Adina Wagner, Lee Tirrell, Mainak Jas, Michael Hanke, Russell Poldrack, Oscar Esteban, Stefan Appelhoff, Chris Holdgraf, Isla Staden, Bertrand Thirion, Dave Kleinschmidt, John Lee, Matteo Di Castello, Michael

- Notter, and Ross Blair. PyBIDS: Python tools for BIDS datasets. *Journal of Open Source Software*, 4(40):1294, August 2019.
- [160] Alan W.C. Yuen, Mark R. Keezer, and Josemir W. Sander. Epilepsy is a neurological and a systemic disorder. *Epilepsy & Behavior*, 78:57–61, January 2018.
- [161] Rina Zelman, Maeike Zijlmans, Julia Jacobs, Claude-E. Châtillon, and Jean Gotman. Improving the identification of High Frequency Oscillations. *Clinical Neurophysiology*, 120(8):1457–1464, August 2009.
- [162] Yipeng Zhang, Qiuqing Lu, Tonmoy Monsoor, Shaun A. Hussain, Joe X. Qiao, Noriko Salamon, Aria Fallah, Myung Shin Sim, Eishi Asano, Raman Sankar, Richard J. Staba, Jerome Engel, William Speier, Vwani Roychowdhury, and Hiroki Nariai. Refining epileptogenic high-frequency oscillations using deep learning: a reverse engineering approach. *Brain Communications*, 4(1):fcab267, January 2022.
- [163] Maeike Zijlmans, Willemieke Zweiphenning, and Nicole Van Klink. Changing concepts in presurgical assessment for epilepsy surgery. *Nature Reviews Neurology*, 15(10):594–606, October 2019.
- [164] Dominik Zumsteg and Heinz Gregor Wieser. Presurgical Evaluation: Current Role of Invasive EEG. *Epilepsia*, 41(s3), March 2000.

Appendix A

iEEGPrep Report

In the next pages, a printed version of the HTML report generated by iEEGPrep as part of the quality-control part of the workflow is presented. This PDF was generated by printing to PDF the original HTML, which can be found on my personal website. The Latin Filler visible in the report is a temporary filling that will be replaced with information about each section later on the development of this tool.

sEEGPrep Report

RAW INFORMATION

BIDS VALIDATION
RESULTS

EPOCHING

DOWNSAMPLING

DETRENDING

REREFERENCING

PLI

REGIONS ID

Raw Information

Et quis adipiscing ipsum exercitation est amet adipiscing anim laboris. Aute et laborum quis ex ad officia deserunt sit laborum esse. Consequat do officia reprehenderit magna fugiat velit duis labore proident qui et labore laborum elit. Quis dolor id amet non aliquip Lorem dolor quis aliqua non nisi nisi.

Info

Filename	bids/sub-EPL31LHS0026/ses-V02SE06/ieeg/sub-EPL31LHS0026_ses-V02SE06_task-full_rec-regionID_run-01_clip-01_ieeg.edf
Measurement Date	2022-10-30 12:32:23
Sampling frequency (Hz)	200.0
Signals in file	80
Duration	00:04:00.000000 (HH:MM:SS)

BIDS Validation Results

BIDS Validator Results

Suspendisse potenti. Quisque blandit urna vitae maximus tempor. Nunc dictum et elit luctus scelerisque. Pellentesque sodales placerat ante eu vulputate. Interdum et malesuada fames ac ante ipsum primis in faucibus. Maecenas nunc erat, accumsan et scelerisque eget, laoreet eget orci. Donec porta at justo vel ultricies. Integer vel libero sit amet justo rhoncus tempor vitae semper magna. Nunc at augue facilisis, viverra ante id, malesuada orci.

bids-validator@1.14.0

1: [ERR] The compulsory file /dataset_description.json is missing. See Section 03 (Modality agnostic files) of the BIDS specification. (code: 57 - DATASET_DESCRIPTION_JSON_MISSING)

Please visit https://neurostars.org/search?q=DATASET_DESCRIPTION_JSON_MISSING for existing conversations about this issue.

1: [WARN] Tabular file contains custom columns not described in a data dictionary (code: 82 - CUSTOM_COLUMN_WITHOUT_DESCRIPTION)

```
./sub-EPL31LHS0026/ses-V02SE06/ieeg/sub-EPL31LHS0026_ses-V02SE06_task-full_run-01_events.tsv
Evidence: Columns: time_abs, time_rel, event not defined, please define in: /events.json, /task-full_events.json, /run-01_events.json, /task-full_run-01_events.json, /sub-EPL31LHS0026/sub-EPL31LHS0026_events.json, /sub-EPL31LHS0026/sub-EPL31LHS0026_task-full_run-01_events.json, /sub-EPL31LHS0026/sub-EPL31LHS0026_task-full_run-01_events.json, /sub-EPL31LHS0026/sub-EPL31LHS0026_task-full_run-01_events.json, /sub-EPL31LHS0026/sub-EPL31LHS0026_ses-V02SE06/sub-EPL31LHS0026_ses-V02SE06_task-full_run-01_events.json, /sub-EPL31LHS0026/sub-EPL31LHS0026_ses-V02SE06/sub-EPL31LHS0026_ses-V02SE06_task-full_run-01_events.json, /sub-EPL31LHS0026/sub-EPL31LHS0026_ses-V02SE06/ieeg/sub-EPL31LHS0026_ses-V02SE06_events.json, /sub-EPL31LHS0026/sub-EPL31LHS0026_ses-V02SE06/ieeg/sub-EPL31LHS0026_ses-V02SE06_task-full_run-01_events.json, /sub-EPL31LHS0026/sub-EPL31LHS0026_ses-V02SE06/ieeg/sub-EPL31LHS0026_ses-V02SE06_task-full_run-01_events.json, /sub-EPL31LHS0026/sub-EPL31LHS0026_ses-V02SE07/ieeg/sub-EPL31LHS0026_ses-V02SE07_task-full_run-01_events.tsv
Evidence: Columns: time_abs, time_rel, event not defined, please define in: /events.json, /task-full_events.json, /run-01_events.json, /task-full_run-01_events.json, /sub-EPL31LHS0026/sub-EPL31LHS0026_events.json, /sub-EPL31LHS0026/sub-EPL31LHS0026_task-full_run-01_events.json, /sub-EPL31LHS0026/sub-EPL31LHS0026_task-full_run-01_events.json, /sub-EPL31LHS0026/sub-EPL31LHS0026_task-full_run-01_events.json, /sub-EPL31LHS0026/sub-EPL31LHS0026_ses-V02SE07/sub-EPL31LHS0026_ses-V02SE07_task-full_events.json, /sub-EPL31LHS0026/sub-EPL31LHS0026_ses-V02SE07/sub-EPL31LHS0026_ses-V02SE07_run-01_events.json, /sub-EPL31LHS0026/sub-EPL31LHS0026_ses-V02SE07/ieeg/sub-EPL31LHS0026_ses-V02SE07_task-full_events.json, /sub-EPL31LHS0026/sub-EPL31LHS0026_ses-V02SE07/ieeg/sub-EPL31LHS0026_ses-V02SE07_events.json, /sub-EPL31LHS0026/sub-EPL31LHS0026_ses-V02SE07/ieeg/sub-EPL31LHS0026_ses-V02SE07_task-full_run-01_events.json, /sub-EPL31LHS0026/sub-EPL31LHS0026_ses-V02SE07/ieeg/sub-EPL31LHS0026_ses-V02SE07_task-full_run-01_events.json, /sub-EPL31LHS0026/sub-EPL31LHS0026_ses-V02SE08/ieeg/sub-EPL31LHS0026_ses-V02SE08_task-full_run-01_events.tsv
Evidence: Columns: time_abs, time_rel, event not defined, please define in: /events.json, /task-full_events.json, /run-01_events.json, /task-full_run-01_events.json, /sub-EPL31LHS0026/sub-EPL31LHS0026_events.json, /sub-EPL31LHS0026/sub-EPL31LHS0026_task-full_run-01_events.json, /sub-EPL31LHS0026/sub-EPL31LHS0026_task-full_run-01_events.json, /sub-EPL31LHS0026/sub-EPL31LHS0026_ses-V02SE08/sub-EPL31LHS0026_ses-V02SE08_events.json, /sub-EPL31LHS0026/sub-EPL31LHS0026_ses-V02SE08/sub-EPL31LHS0026_ses-V02SE08_task-full_events.json, /sub-EPL31LHS0026/sub-EPL31LHS0026_ses-V02SE08/sub-EPL31LHS0026_ses-V02SE08_run-01_events.json, /sub-EPL31LHS0026/sub-EPL31LHS0026_ses-V02SE08/sub-EPL31LHS0026_ses-V02SE08_task-full_run-01_events.json, /sub-EPL31LHS0026/sub-EPL31LHS0026_ses-V02SE08/ieeg/sub-EPL31LHS0026_ses-V02SE08_task-full_events.json, /sub-EPL31LHS0026/sub-EPL31LHS0026_ses-V02SE08/ieeg/sub-EPL31LHS0026_ses-V02SE08_run-01_events.json, /sub-EPL31LHS0026/sub-EPL31LHS0026_ses-V02SE08/ieeg/sub-EPL31LHS0026_ses-V02SE08_task-full_run-01_events.json, /sub-EPL31LHS0026/sub-EPL31LHS0026_ses-V02SE09/ieeg/sub-EPL31LHS0026_ses-V02SE09_task-full_run-01_events.tsv
Evidence: Columns: time_abs, time_rel, event not defined, please define in: /events.json, /task-full_events.json, /run-01_events.json, /task-full_run-01_events.json, /sub-EPL31LHS0026/sub-EPL31LHS0026_events.json, /sub-EPL31LHS0026/sub-EPL31LHS0026_task-full_events.json, /sub-EPL31LHS0026/sub-EPL31LHS0026_task-full_run-01_events.json, /sub-EPL31LHS0026/sub-EPL31LHS0026_task-full_run-01_events.json, /sub-EPL31LHS0026/sub-EPL31LHS0026_ses-V02SE09/sub-EPL31LHS0026_ses-V02SE09_events.json, /sub-EPL31LHS0026/sub-EPL31LHS0026_ses-V02SE09_task-full_events.json, /sub-EPL31LHS0026/sub-EPL31LHS0026_ses-V02SE09/sub-EPL31LHS0026_ses-V02SE09_run-01_events.json, /sub-EPL31LHS0026/sub-EPL31LHS0026_ses-V02SE09/ieeg/sub-EPL31LHS0026_ses-V02SE09_events.json, /sub-EPL31LHS0026/sub-EPL31LHS0026_ses-V02SE09/ieeg/sub-EPL31LHS0026_ses-V02SE09_task-full_events.json, /sub-EPL31LHS0026/sub-EPL31LHS0026_ses-V02SE09/ieeg/sub-EPL31LHS0026_ses-V02SE09_run-01_events.json, /sub-EPL31LHS0026/sub-EPL31LHS0026_ses-V02SE09/ieeg/sub-EPL31LHS0026_ses-V02SE09_task-full_run-01_events.json
```

Please visit https://neurostars.org/search?q=CUSTOM_COLUMN_WITHOUT_DESCRIPTION for existing conversations about this issue.

2: [WARN] The recommended file /README is missing. See Section 03 (Modality agnostic files) of the BIDS specification. (code: 101 - README_FILE_MISSING)

Please visit https://neurostars.org/search?q=README_FILE_MISSING for existing conversations about this issue.

Summary:

20 Files, 367.92GB
1 - Subject
4 - Sessions

Available Tasks:

full

Available Modalities:

ieEG

If you have any questions, please post on <https://neurostars.org/tags/bids>.



EEG/iEEG BIDS Validator Report

Suspendisse potenti. Quisque blandit urna vitae maximus tempor. Nunc dictum et elit luctus scelerisque. Pellentesque sodales placerat ante eu vulputate. Interdum et malesuada fames ac ante ipsum primis in faucibus. Maecenas nunc erat, accumsan et scelerisque eget, laoreet eget orci. Donec porta at justo vel ultricies. Integer vel libero sit amet justo rhoncus tempor vitae semper magna. Nunc at augue facilisis, viverra ante id, malesuada orci.

- Info:** Test PASSED. All channels present on the channels.tsv file were found in the corresponding EDF file.
- Info:** Test PASSED. Units in the channels.tsv file match the units in the EDF file.
- Info:** Test PASSED. Filters defined in the channels.tsv match with the defined filters in the EDF file.
- Info:** Test PASSED. Sampling frequency in the files channels.tsv, EDF and json file match.
- Info:** Test PASSED. All electrodes positions in the electrodes.tsv file are numeric, as required by the BIDS standard.
- Info:** Assuming unipolar reference as the parameter 'iEEGReference' is not set to 'bipolar' in the json file.
- Info:** Test PASSED. All the contacts forming the channels in the channels.tsv and EDF files are present in the electrodes.tsv.
- Info:** Test PASSED. Number of channels specified in the json file match the number of channels in the channels.tsv and EDF files.

Warning: Length specified in the edf file (86387 seconds) doesn't seem to match the length indicated in the JSON file (23.997).

Epoch rule report

Events found

Suspendisse potenti. Quisque blandit urna vitae maximus tempor. Nunc dictum et elit luctus scelerisque. Pellentesque sodales placerat ante eu vulputate. Interdum et malesuada fames ac ante ipsum primis in faucibus. Maecenas nunc erat, accumsan et scelerisque eget, laoreet eget orci. Donec porta at justo vel ultricies. Integer vel libero sit amet justo rhoncus tempor vitae semper magna. Nunc at augue facilisis, viverra ante id, malesuada orci.

Epoch start event	Relative start time	Duration	Clip number
awake trigger	05:02:08.458008	00:04:00.000000	clip-01

Downsampling rule report

Sampling rates

Suspendisse potenti. Quisque blandit urna vitae maximus tempor. Nunc dictum et elit luctus scelerisque. Pellentesque sodales placerat ante eu vulputate. Interdum et malesuada fames ac ante ipsum primis in faucibus. Maecenas nunc erat, accumsan et scelerisque eget, laoreet eget orci. Donec porta at justo vel ultricies. Integer vel libero sit amet justo rhoncus tempor vitae semper magna. Nunc at augue facilisis, viverra ante id, malesuada orci.

Original sampling rate	Target sampling rate	New sampling rate	Discarded channels
2048.0	200	200.0	['Patient Event', 'C149', 'C150', 'C151', 'C152', 'C153', 'C154', 'C155', 'C156', 'C157', 'C158', 'C159', 'C160', 'C161', 'C162', 'C163', 'C164', 'C165', 'C166', 'C167', 'C168', 'C169', 'C170', 'C171', 'C172', 'C173', 'C174', 'C175', 'C176', 'C177', 'C178', 'C179', 'C180', 'C181', 'C182', 'C183', 'C184', 'C185', 'C186', 'C187', 'C188', 'C189', 'C190', 'C191', 'C192', 'C193', 'C194', 'C195', 'C196', 'C197', 'C198', 'C199', 'C200', 'C201', 'C202', 'C203', 'C204', 'C205', 'C206', 'C207', 'C208', 'C209', 'C210', 'C211', 'C212', 'C213', 'C214', 'C215', 'C216', 'C217', 'C218', 'C219', 'C220', 'C221', 'C222', 'C223', 'C224', 'C225', 'C226', 'C227', 'C228', 'C229', 'C230', 'C231', 'C232', 'C233', 'C234', 'C235', 'C236', 'C237', 'C238', 'C239', 'C240', 'C241', 'C242', 'C243', 'C244', 'C245', 'C246', 'C247', 'C248', 'C249', 'C250', 'C251', 'C252', 'C253', 'C254', 'C255', 'C256', 'DC1', 'DC2', 'DC3', 'DC4', 'DC5', 'DC6', 'DC7', 'DC8', 'DC9', 'DC10', 'DC11', 'DC12', 'DC13', 'DC14', 'DC15', 'DC16', 'TRIG', 'OSAT', 'PR', 'Pleth']

Size comparison

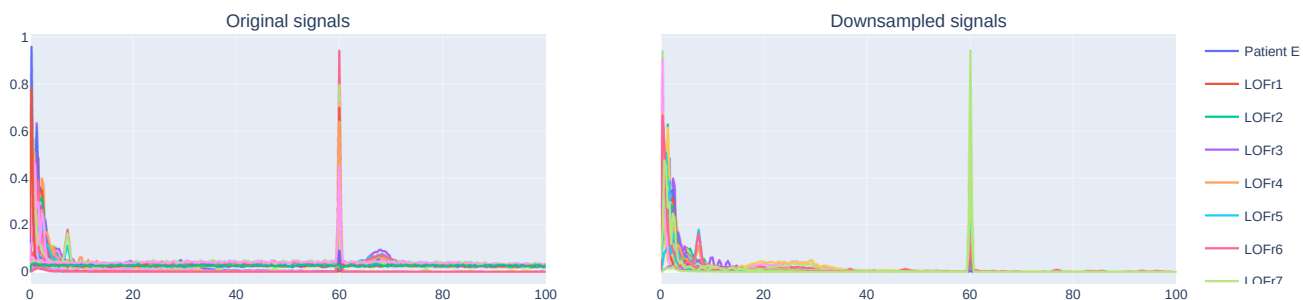
Suspendisse potenti. Quisque blandit urna vitae maximus tempor. Nunc dictum et elit luctus scelerisque. Pellentesque sodales placerat ante eu vulputate. Interdum et malesuada fames ac ante ipsum primis in faucibus. Maecenas nunc erat, accumsan et scelerisque eget, laoreet eget orci. Donec porta at justo vel ultricies. Integer vel libero sit amet justo rhoncus tempor vitae semper magna. Nunc at augue facilisis, viverra ante id, malesuada orci.

Original size (MB)	Downsampled size (MB)	Compression rate
273.25	14.27	19.14

PSD QC

Suspendisse potenti. Quisque blandit urna vitae maximus tempor. Nunc dictum et elit luctus scelerisque. Pellentesque sodales placerat ante eu vulputate. Interdum et malesuada fames ac ante ipsum primis in faucibus. Maecenas nunc erat, accumsan et scelerisque eget, laoreet eget orci. Donec porta at justo vel ultricies. Integer vel libero sit amet justo rhoncus tempor vitae semper magna. Nunc at augue facilisis, viverra ante id, malesuada orci.

Normalized PSD



Detrending rule report

General information

Suspendisse potenti. Quisque blandit urna vitae maximus tempor. Nunc dictum et elit luctus scelerisque. Pellentesque sodales placerat ante eu vulputate. Interdum et malesuada fames ac ante ipsum primis in faucibus. Maecenas nunc erat, accumsan et scelerisque eget, laoreet eget orci. Donec porta at justo vel ultricies. Integer vel libero sit amet justo rhoncus tempor vitae semper magna. Nunc at augue facilisis, viverra ante id, malesuada orci.

- **Method used:** HighPass
- **Discarded channels:** []
- **Transition band (Hz):** [0.1, 0.2]

Comparison of channels mean

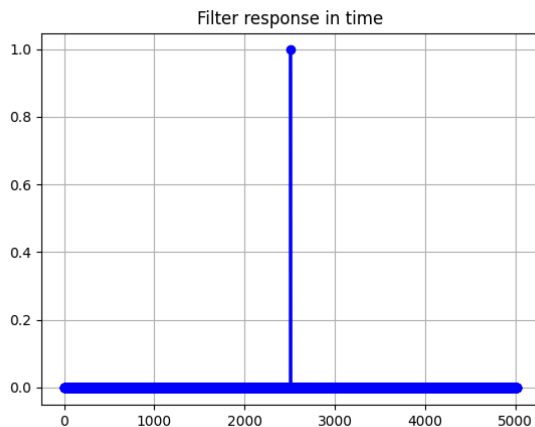
Suspendisse potenti. Quisque blandit urna vitae maximus tempor. Nunc dictum et elit luctus scelerisque. Pellentesque sodales placerat ante eu vulputate. Interdum et malesuada fames ac ante ipsum primis in faucibus. Maecenas nunc erat, accumsan et scelerisque eget, laoreet eget orci. Donec porta at justo vel ultricies. Integer vel libero sit amet justo rhoncus tempor vitae semper magna. Nunc at augue facilisis, viverra ante id, malesuada orci.

Channel	Original mean	New mean
LOFr1	23.048317	-0.002896
LOFr2	47.710714	0.015531
LOFr3	33.249204	-0.027254
LOFr4	10.846365	0.008861
LOFr5	10.147426	-0.105171
LOFr6	17.601156	0.000834
LOFr7	10.147426	-0.105171

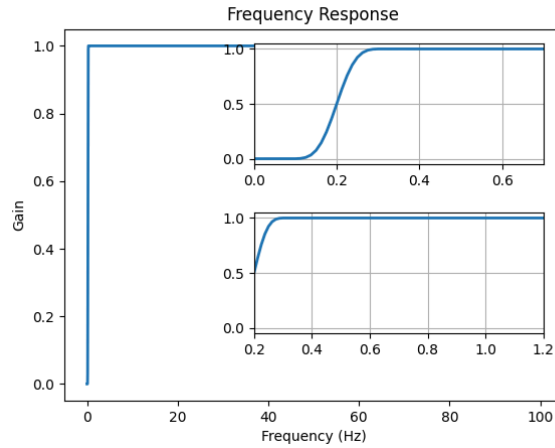
Highpass filter information

Suspendisse potenti. Quisque blandit urna vitae maximus tempor.

Filter time response



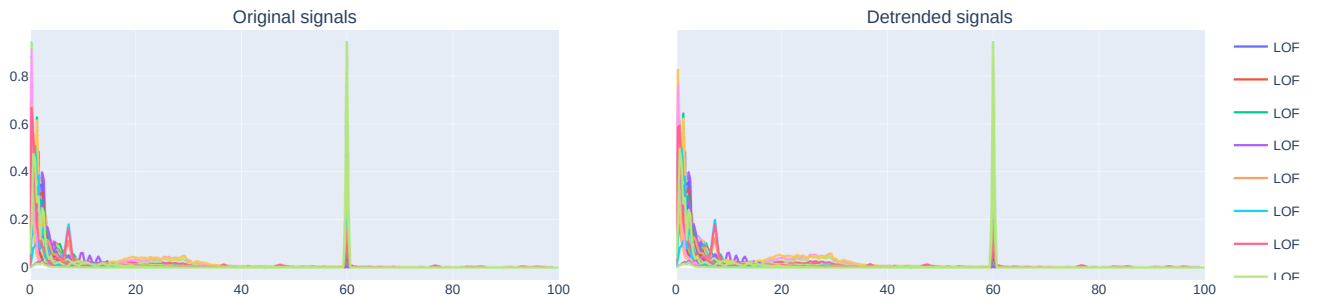
Filter frequency response



PSD QC

Suspendisse potenti. Quisque blandit urna vitae maximus tempor. Nunc dictum et elit luctus scelerisque. Pellentesque sodales placerat ante eu vulputate. Interdum et malesuada fames ac ante ipsum primis in faucibus. Maecenas nunc erat, accumsan et scelerisque eget, laoreet eget orci. Donec porta at justo vel ultricies. Integer vel libero sit amet justo rhoncus tempor vitae semper magna. Nunc at augue facilisis, viverra ante id, malesuada orci.

Normalized PSD



Rereferencing rule report

General information

Suspendisse potenti. Quisque blandit urna vitae maximus tempor. Nunc dictum et elit luctus scelerisque. Pellentesque sodales placerat ante eu vulputate. Interdum et malesuada fames ac ante ipsum primis in faucibus. Maecenas nunc erat, accumsan et scelerisque eget, laoreet eget orci. Donec porta at justo vel ultricies. Integer vel libero sit amet justo rhoncus tempor vitae semper magna. Nunc at augue facilisis, viverra ante id, malesuada orci.

- Discarded channels: []

Bipolar channels found

Suspendisse potenti. Quisque blandit urna vitae maximus tempor. Nunc dictum et elit luctus scelerisque. Pellentesque sodales placerat ante eu vulputate. Interdum et malesuada fames ac ante ipsum primis in faucibus. Maecenas nunc erat, accumsan et scelerisque eget, laoreet eget orci. Donec porta at justo vel ultricies. Integer vel libero sit amet justo rhoncus tempor vitae semper magna. Nunc at augue facilisis, viverra ante id, malesuada orci.

Bipolar channel	Unipolar channels
LOFr1-2	(LOFr1, LOFr2)
LOFr2-3	(LOFr2, LOFr3)
LOFr3-4	(LOFr3, LOFr4)
LOFr4-5	(LOFr4, LOFr5)
LOFr5-6	(LOFr5, LOFr6)
LOFr6-7	(LOFr6, LOFr7)

PLI rule report

General information

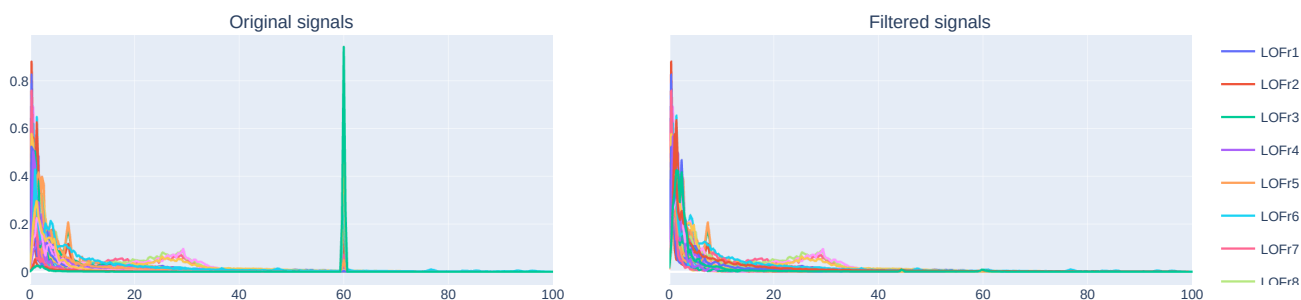
Suspendisse potenti. Quisque blandit urna vitae maximus tempor. Nunc dictum et elit luctus scelerisque. Pellentesque sodales placerat ante eu vulputate. Interdum et malesuada fames ac ante ipsum primis in faucibus. Maecenas nunc erat, accumsan et scelerisque eget, laoreet eget orci. Donec porta at justo vel ultricies. Integer vel libero sit amet justo rhoncus tempor vitae semper magna. Nunc at augue facilisis, viverra ante id, malesuada orci.

- **Method:** Cleanline
- **Line frequency (Hz):** 60
- **Bandwidth (Hz):** 8

PSD QC

Suspendisse potenti. Quisque blandit urna vitae maximus tempor. Nunc dictum et elit luctus scelerisque. Pellentesque sodales placerat ante eu vulputate. Interdum et malesuada fames ac ante ipsum primis in faucibus. Maecenas nunc erat, accumsan et scelerisque eget, laoreet eget orci. Donec porta at justo vel ultricies. Integer vel libero sit amet justo rhoncus tempor vitae semper magna. Nunc at augue facilisis, viverra ante id, malesuada orci.

Normalized PSD



Region Identification rule report

General information

Suspendisse potenti. Quisque blandit urna vitae maximus tempor. Nunc dictum et elit luctus scelerisque. Pellentesque sodales placerat ante eu vulputate. Interdum et malesuada fames ac ante ipsum primis in faucibus. Maecenas nunc erat, accumsan et scelerisque eget, laoreet eget orci. Donec porta at justo vel ultricies. Integer vel libero sit amet justo rhoncus tempor vitae semper magna. Nunc at augue facilisis, viverra ante id, malesuada orci.

- **Discarded channels:** []

Regions report

Suspendisse potenti. Quisque blandit urna vitae maximus tempor. Nunc dictum et elit luctus scelerisque. Pellentesque sodales placerat ante eu vulputate. Interdum et malesuada fames ac ante ipsum primis in faucibus. Maecenas nunc erat, accumsan et scelerisque eget, laoreet eget orci. Donec porta at justo vel ultricies. Integer vel libero sit amet justo rhoncus tempor vitae semper magna. Nunc at augue facilisis, viverra ante id, malesuada orci.

Region	Number of channels	Channels
Left-Cerebral-White-Matter	21	[LOFr1-2\LOFr2-3\LOFr7-8\LPCg2-3\LPCg3-4\LPCg4-5\LPCg5-6\n\LPCg6-7\LPCg7-8\LPCg8-9\LPCg9-10\LAm5-6\LAHc5-6\LAHc6-7\n\LAHc7-8\LAHc8-9\LAHc9-10\LPHc4-5\LPHc5-6\LPHc6-7\LPHc7-8]
ctx-lh-lateralorbitofrontal	3	[LOFr3-4\LOFr4-5\LOFr5-6]
CSF	12	[LOFr6-7\LOFr9-10\LTepo5-6\LTepo6-7\LTepo7-8\LPIn3-4\LPIn4-5\n\LPIn5-6\LPIn7-8\LPIn8-9\LPIn9-10\RAm8-9]
ctx-lh-parsorbitalis	1	[LOFr8-9]
ctx-lh-posteriorcingulate	1	[LPCa1-2]

Appendix B

iEEGPrep snakebids.yaml

Below, the customizable part of the configuration YAML file used with iEEGPrep workflow is shown. Note that this doesn't represent the complete file as it was only intended to show the part of it that can be customized by the user.

In particular, two parts are important. First the `pybids_inputs` can be modified to detect the different files needed by the pipeline. This follows Snakebids standard, so please refer to the original documentation for more information. After the indicator *Pipeline parameters*, the parameters for the different preprocessing steps are given along with an explanation for their customization.

```
 bids_dir: '/path/to/bids_dir'  
output_dir: '/path/to/output_dir'  
  
# Resources  
local_scratch: '/tmp'  
local_scratch_env: 'SLURM_TMPDIR'  
  
#enable printing debug statements during parsing -- disable if generating dag  
↪ visualization  
debug: False  
  
derivatives: False #will search in bids/derivatives if True; can also be path(s) to  
↪ derivatives datasets  
  
#list of analysis levels in the bids app  
analysis_levels: &analysis_levels  
- participant
```


#mapping from analysis_level to set of target rules or files

targets_by_analysis_level:

participant:

- '' # if "", then the first rule is run

#this configures the pybids grabber - create an entry for each type of input you want

↪ to grab

indexed by name of input

dictionary for each input is passed directly to pybids get()

pybids_inputs:

edf:

filters:

datatype: 'ieeg'

suffix: 'ieeg'

extension: '.edf'

wildcards:

- subject

- session

- task

- run

events:

filters:

suffix: 'events'

extension: '.tsv'

wildcards:

- subject

- session

- task

- run

channels:

filters:

suffix: 'channels'

extension: '.tsv'

wildcards:

- subject

- session

- task

- run

electrodes:

filters:

suffix: 'electrodes'

extension: '.tsv'

wildcards:

- subject

- session

json:

filters:

suffix: 'ieeg'

```

    extension: '.json'
wildcards:
  - subject
  - session
  - task
  - run

parc:
  filters:
    suffix: 'orig'
    extension: '.mgz'
  wildcards:
    - subject
    - session
    - acquisition
    - task
    - run
  custom_path: /path/to/sub-{subject}_desc-synthsegcortparc_dseg.nii.gz

# ----- Pipeline parameters-----
# Event:[label or index, 'duration' or 'n_samples', amount as seconds or int]
event: ['awake trigger', 'duration', 240] # 4 min 'awake trigger'

# 1. Parameters for downsampling step:
# (1.a) Target sampling rate
# Default: 200 Hz
target_srate: 1024

# 2. Parameters for detrending step:
# (2.a) Detrend method
# Options: 'LinearDetrend','HighPass' (default), 'Demmean'
detrend_method: 'HighPass'

# (2.b) Transition band of high pass filter used in detrending step (if selected)
highpass: [0.1, 0.2]

# 3. Parameters for PLI attenuation
# (3.a) Power line interference removal method
# Options: 'Cleanline', 'Zapline', 'NotchFilter' (default), 'PLIremoval'
methodPLI: 'Cleanline'

# (3.b) Electrical grid fundamental frequency
# Default: 60 Hz
lineFreq: 60

# (3.c) Bandwidth for Notch filter (TODO) or Cleanline
# Note: in 'Cleanline', the bandwidth represents the frequency range
# (around the theoretical fundamental and harmonic frequencies) in which
# to search for possible peaks associated with line noise.
# Default: 4 (ex: if lineFreq=60, the range would go from 58 to 62)
bandwidth: 8

# (3.d) Number of harmonics to remove when using Notch Filter

```

```
# Default: 1 (only removes fundamental frequency)
n_harmonics: 1

# 4. Parameters for region identification step
# (4.1) Boolean to discard signals from electrodes located in white matter or
# 'unknown' region
# Default: True (discards the signals)
discard_wm_un: True
reference_edf: 'bipolar' # 'unipolar' or 'bipolar'
vol_version: True

# 4. Parameter for regions identification and rereferencing step
# (4.1) Column names in tsv file with electrodes information
# Format: [{type of channel}, {channel label}, {x position}, {y position}, {z
↪ position}, {group}]
# Default: ['type', 'label', 'x', 'y', 'z', 'group'] or None (both options are equivalent)
tsv_cols: ['type', 'label', 'x', 'y', 'z', 'orig_group']

# Color table required to assign labels for each value in the parcellation file
colortable: /path/to/colortable.tsv

# QC
bids_validator_flags: '--ignoreSubjectConsistency'
```

Appendix C

iEEGPrep validation

Below, extra validation Figures for the validation of each of the preprocessing steps are shown. This is an extension to the results presented in Chapter 2.

C.1 Epoch extraction

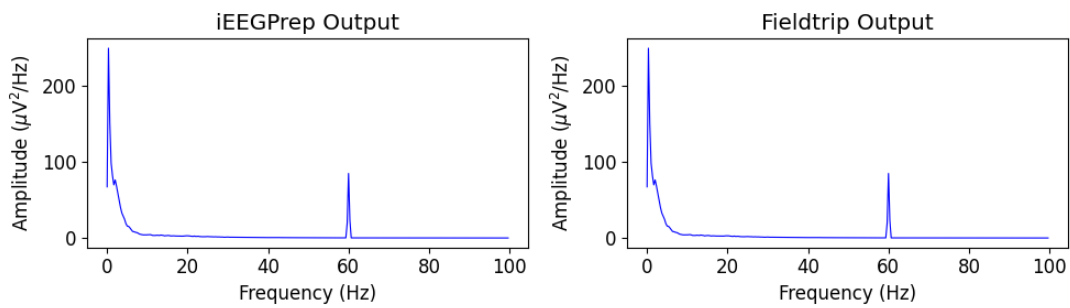


Figure C.1: Frequency spectrum plots for the first 60 seconds extracted from the channel LPIIn02 by iEEGPrep and Fieldtrip. Frequency domain signal was calculated using Welch's method.

C.2 Downsampling

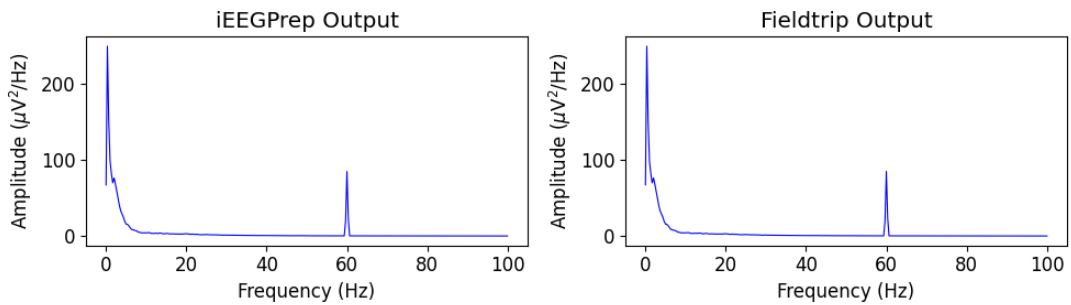


Figure C.2: Frequency spectrum plots for the first 60 seconds of the downsampled signal from the channel LPIIn02 using iEEGPrep and Fieldtrip. Frequency domain signal was calculated using Welch's method with a window of 3 s and an overlap of 50%.

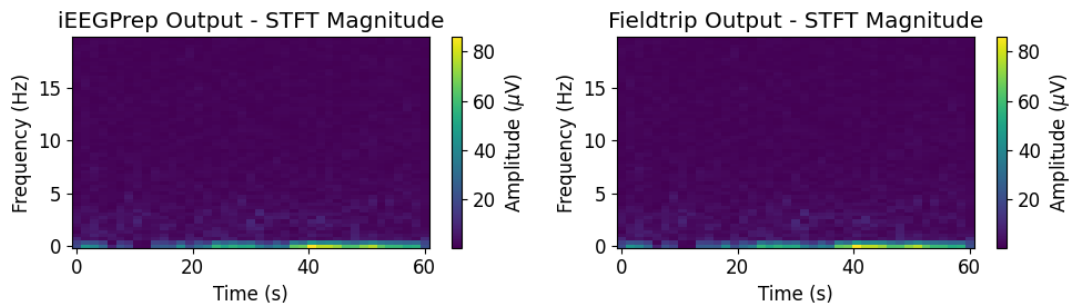


Figure C.3: Time-frequency map of the first 60 seconds of downsampled data from the channel LPIIn02 using iEEGPrep and Fieldtrip. Maps were calculated using short-time Fourier Transform (STFT) with a window of 3 s and an overlap of 50%.

C.3 Drift correction

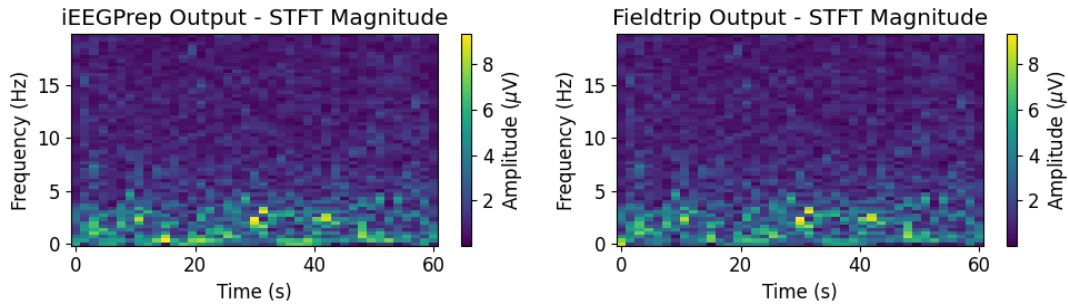


Figure C.4: Time-frequency map of the first 60 seconds of detrended data from the channel LPI_n02 using iEEGPrep and Fieldtrip. Maps were calculated using short-time Fourier Transform (STFT) with a window of 3 s and an overlap of 50%.

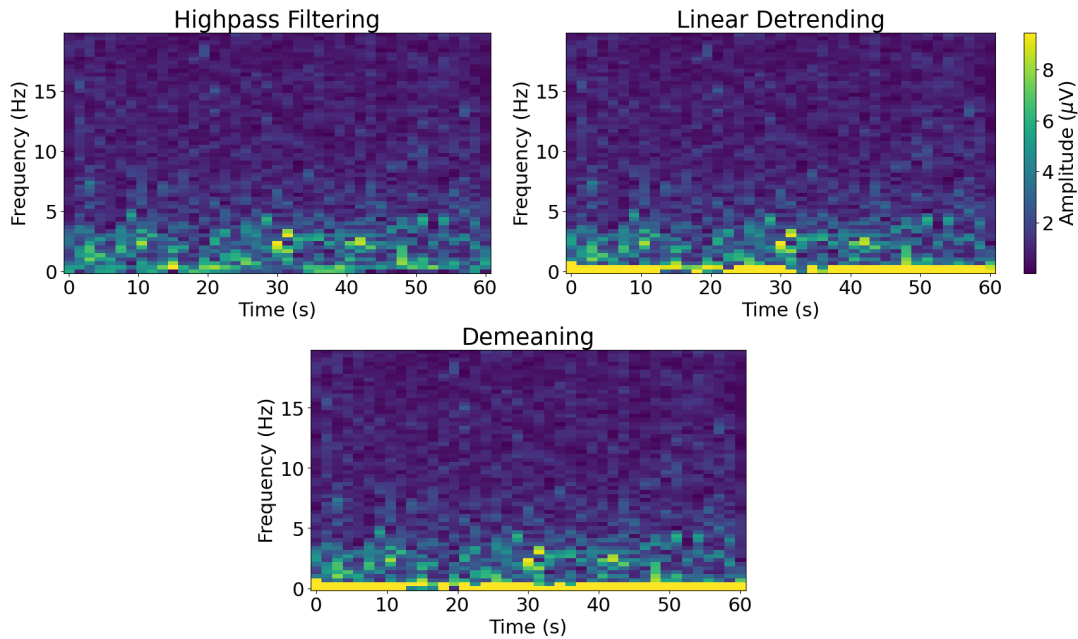


Figure C.5: Time-frequency map of the first 60 seconds of detrended data from the channel LPI_n02 using the different methods implemented in iEEGPrep. Maps were calculated using short-time Fourier Transform (STFT) with a window of 3 s and an overlap of 50%. Colorbar scale of the three maps is the same. Only high-pass filtering shows a significant reduction of slow frequency oscillations.

C.4 Re-referencing

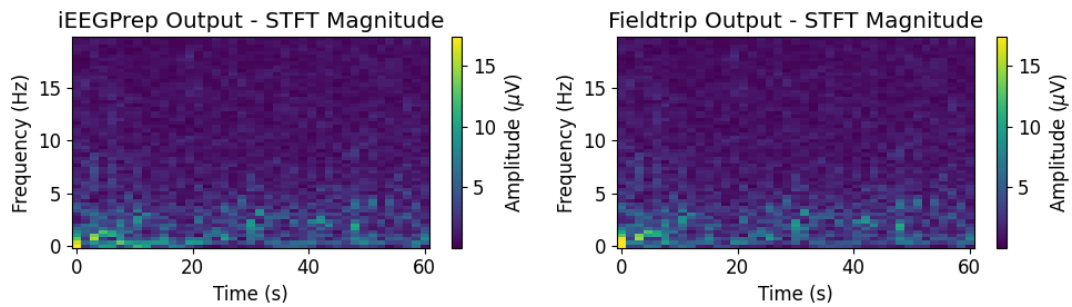


Figure C.6: Time-frequency map from 0 Hz to 20 Hz of the first 60 seconds of the re-referenced signal from the channels LPIIn02 and LPIIn03 using iEEGPrep and Fieldtrip.

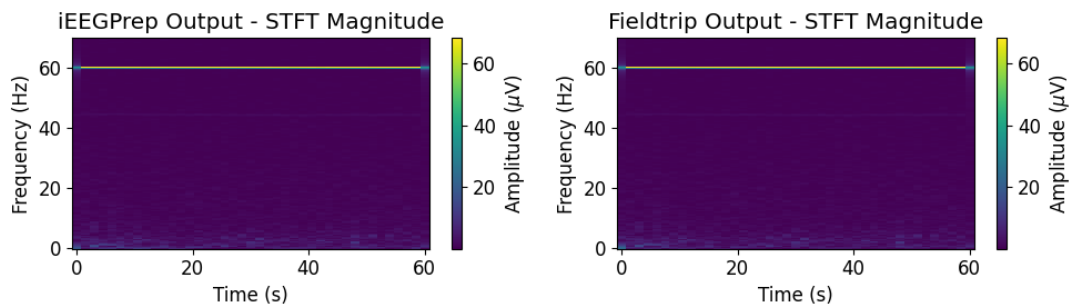


Figure C.7: Time-frequency map from 0 Hz to 70 Hz of the first 60 seconds of the re-referenced signal from the channels LPIIn02 and LPIIn03 using iEEGPrep and Fieldtrip. A predominance of line noise can be observed.

C.5 PLI attenuation

Comparison against Fieldtrip output using real data

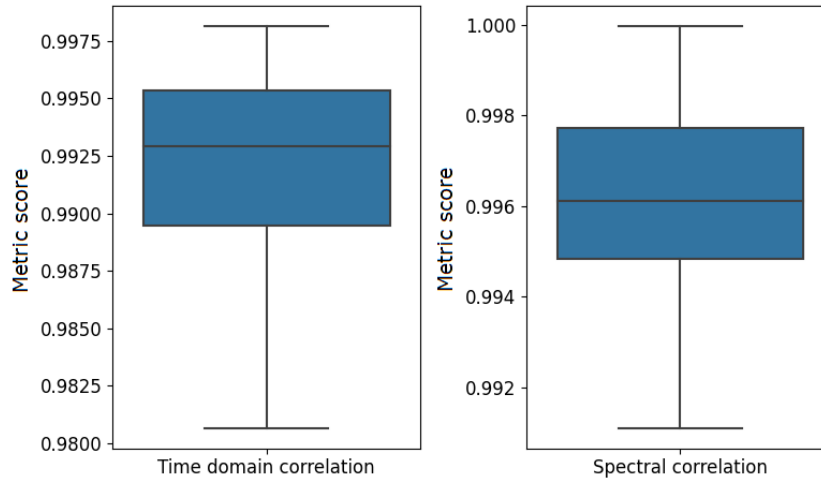


Figure C.8: Distribution of correlation values with zero lag in time domain (left) and frequency domain (right) between the different pairs of channels present in the re-referenced epochs with PLI attenuation for iEEGPrep and Fieldtrip.

Comparison between the different implemented methods

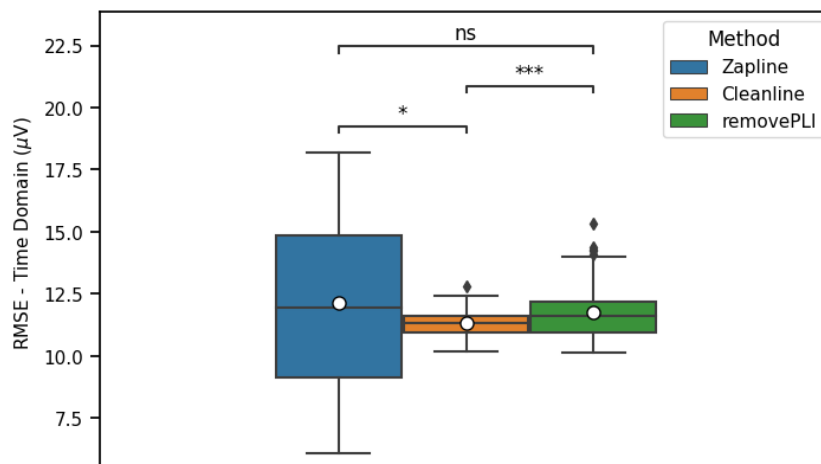


Figure C.9: Distribution of time domain RMSE on the hundred simulated sEEG channels in case A for the different advanced PLI attenuation methods implemented in iEEGPrep.

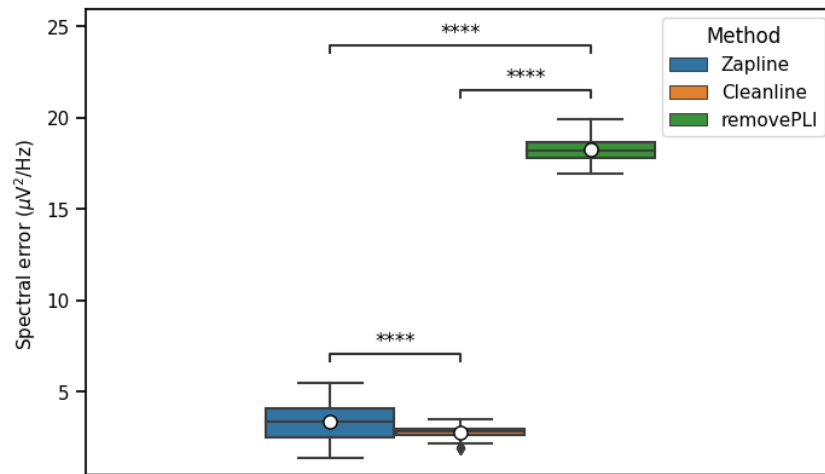


Figure C.10: Distribution of frequency domain RMSE on the hundred simulated sEEG channels in case A for the different advanced PLI attenuation methods implemented in iEEGPrep.

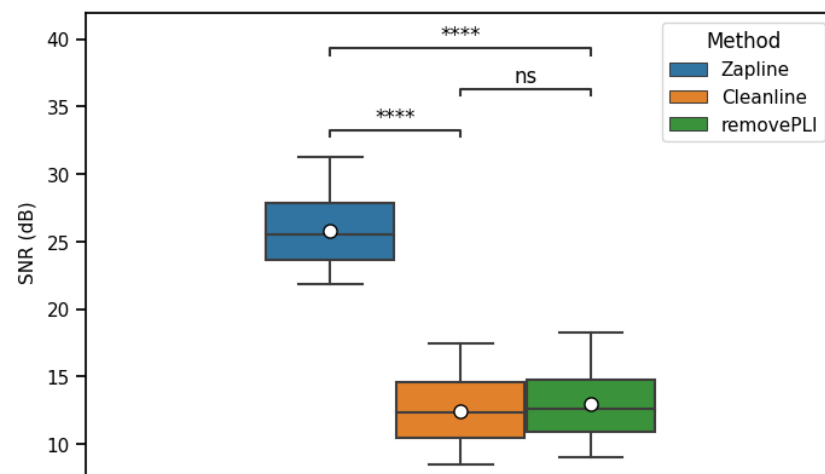


Figure C.11: Distribution of SNR on the simulated sEEG channels in case B for the different advanced PLI attenuation methods implemented in iEEGPrep.

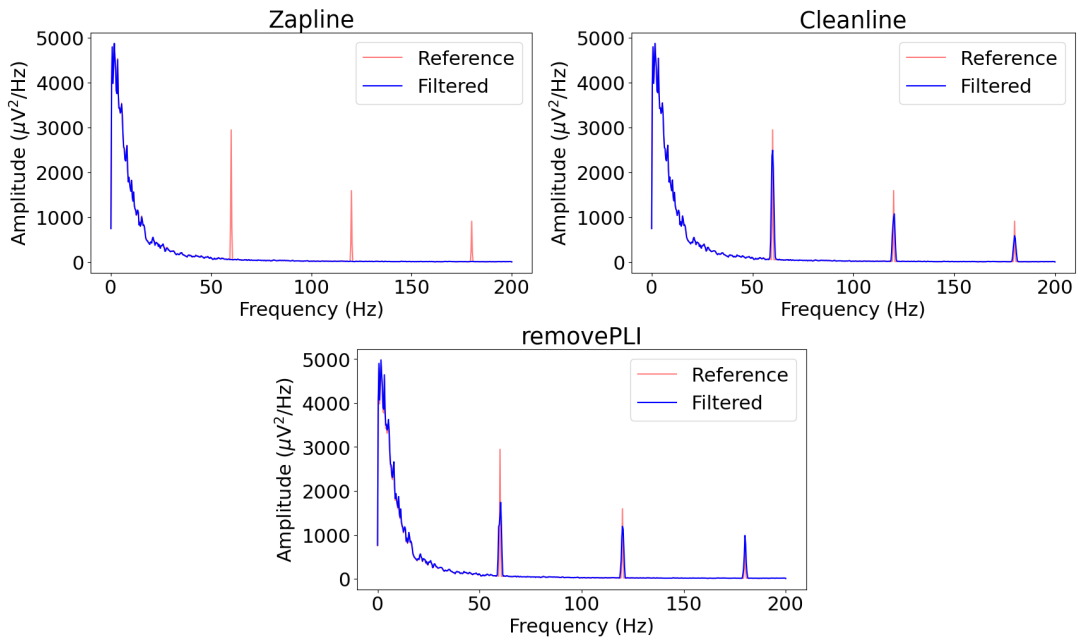


Figure C.12: Example of power spectral density curves obtained through Welch's method for one simulated signal filtered for case B through the different advanced line noise attenuation algorithms implemented in iEEGP_{REP}. Orange and blue signals show simulated signals before and after PLI attenuation respectively.

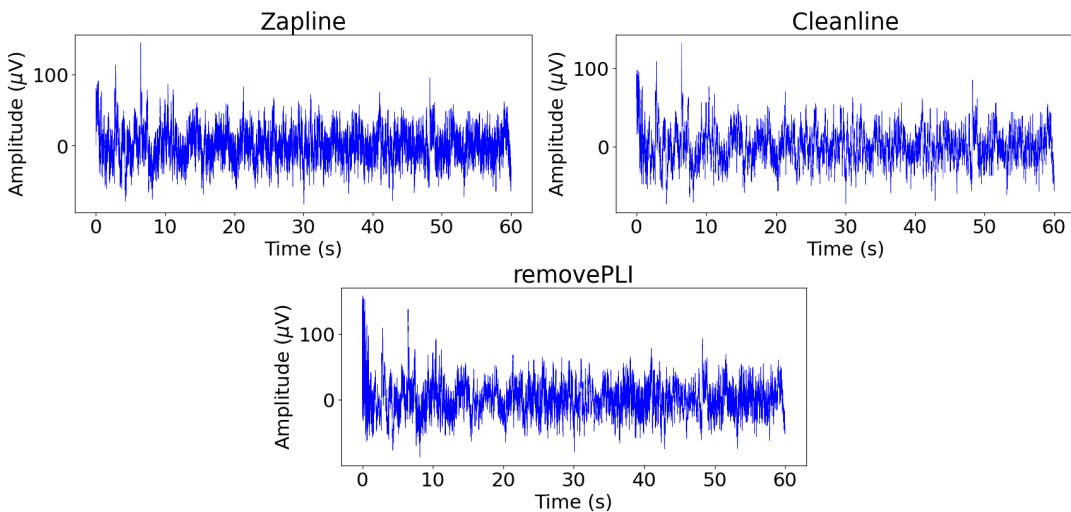


Figure C.13: Time plot of the first 60 seconds of the re-referenced signal with PLI attenuation from the channels LPI_n02 and LPI_n03 using the implemented advanced PLI attenuation methods.

Appendix D

Fieldtrip code

Below, the code used for preprocessing of one EDF file with Fieldtrip is found. This was used as part of the validation for sEEGPrep discussed in Chapter 2.

```
1 clear; clc;
2
3 filename = '/bids_dir/sub-EPL31LHS0026/ses-V02SE06/ieeg/...
4 sub-EPL31LHS0026_ses-V02SE06_task-full_run-01_ieeg.edf';
5
6 %% Get the epoch
7 % define trials
8 cfg                = [];
9 cfg.dataset        = filename;
10 cfg.trialfun       = 'trialfun_annotation';
11 cfg = ft_definetrial(cfg);
12 cfg                = ft_definetrial(cfg);
13 % read the data
14 cfg.continuous     = 'yes';
15 % cfg.channel       = {'LPIn2', 'LPIn3'}; %idx=73,74
16 data               = ft_preprocessing(cfg);
17 idx = 73;
18
19 % Save data
20 signals = data.trial(1);
21 signals = signals{1};
22 save('epoching.mat', "signals");
23
24 labels = data.label;
25 save('unipolar_labels_matlab.mat', "labels");
26
27 %% Downsample
28 % the default functionality in ft_resampleddata applies a firfs
29 % anti-aliasing filter that has its cutoff at the new Nyquist frequency
30 cfg = [];
31 cfg.resamplefs = 200;
```

```

32 data = ft_resampleddata(cfg, data);
33
34 % Save data
35 signals = data.trial(1);
36 signals = signals{1};
37 save('Downsampling.mat', "signals");
38
39 %% Drift correction using high-pass
40
41 cfg = [];
42 cfg.demean = 'yes';
43 cfg.baselinewindow = 'all';
44 cfg.padding = 4*60;
45 cfg.padtype = 'data';
46 cfg.hpfilter = 'yes';
47 cfg.hpfreq = 0.25;
48 cfg.hpfiltord = 3;
49 data = ft_preprocessing(cfg, data);
50
51 % Save data
52 signals = data.trial(1);
53 signals = signals{1};
54 save('Highpass.mat', "signals");
55
56
57 %% Re-reference
58
59 depths = {'LOFr*', 'LPCg*', 'LAm*', 'LAHc*', 'LPHc*', 'LTePo*', 'LAIIn*', 'LPIIn*', ...
60          'ROFr*', 'RPCg*', 'RAm*', 'RAHc*', 'RPHc*', 'RAIIn*', 'RPIIn*'};
61 for d = 1:numel(depths)
62     cfg = [];
63     cfg.channel = ft_channelselection(depths{d}, data.label);
64     cfg.reref = 'yes';
65     cfg.refchannel = 'all';
66     cfg.refmethod = 'bipolar';
67     reref_depths{d} = ft_preprocessing(cfg, data);
68 end
69
70 % Append together
71 cfg = [];
72 data = ft_appenddata(cfg, reref_depths{:});
73
74 % Save data
75 signals = data.trial(1);
76 signals = signals{1};
77 save('Reref.mat', "signals");
78
79 labels = data.label;
80 save('bipolar_labels_matlab.mat', "labels");
81
82 %% Run cleanline
83 data = signals;
84 signals = [];
85 signals.data = data;

```

```
86 signals.srate = 200;
87 lineNoiseIn = struct('lineNoiseMethod', 'clean', ...
88     'lineNoiseChannels', 1:size(data,1),...
89     'Fs', 200, ...
90     'lineFrequencies', [60],...
91     'p', 0.01, ...
92     'fScanBandWidth', 8, ...
93     'taperBandWidth', 2, ...
94     'taperWindowSize', 4, ...
95     'taperWindowStep', 1, ...
96     'tau', 100, ...
97     'pad', 2, ...
98     'fPassBand', [0 100], ...
99     'maximumIterations', 10);
100 [signal, lineNoiseOut] = cleanLineNoise(signals, lineNoiseIn);
101 signals = signal.data;
102 save('cleanline.mat', "signals");
```

Appendix E

Models hyperparameters tested using GridSearchCV

Below, we show the parameters tested for each of the implemented traditional machine learning models. The parameters were tested using the grid search CV with the function `GridSearchCV()` from `sklearn` package. 10-fold CV was used for this purpose.

- *Logistic Regression:*

```
1 param_grid = {'logistic__C': [0.001, 0.01, 0.1, 1, 10, 100],
2               "logistic__penalty":["l2"],
3               'logistic__solver': ['saga', 'newton-cg'],
4               'logistic__multi_class': ['multinomial']}
```

- *Random Forest:*

```
1 param_grid = dict({'n_estimators': [250, 500, 750, 1000],
2                   'max_depth': [3, 5, 7],
3                   'max_features' : [0.5, 0.75],
4                   'min_samples_leaf': [0.05],
5                   'criterion': ['gini', 'entropy']
6                   })
```

- *Extreme Gradient Boosting:*

```
1 param_grid = dict({'n_estimators': [500, 750, 1000],
2                   'max_depth': [2, 3, 4],
3                   'learning_rate' : [0.001, 0.01]
4                   })
```

Appendix F

Deep learning models evaluated using cross-validation

Next, the deep learning architectures tested during selection of the best parameters can be found. These Figures were produced with the Python package `torchview` and later edited.

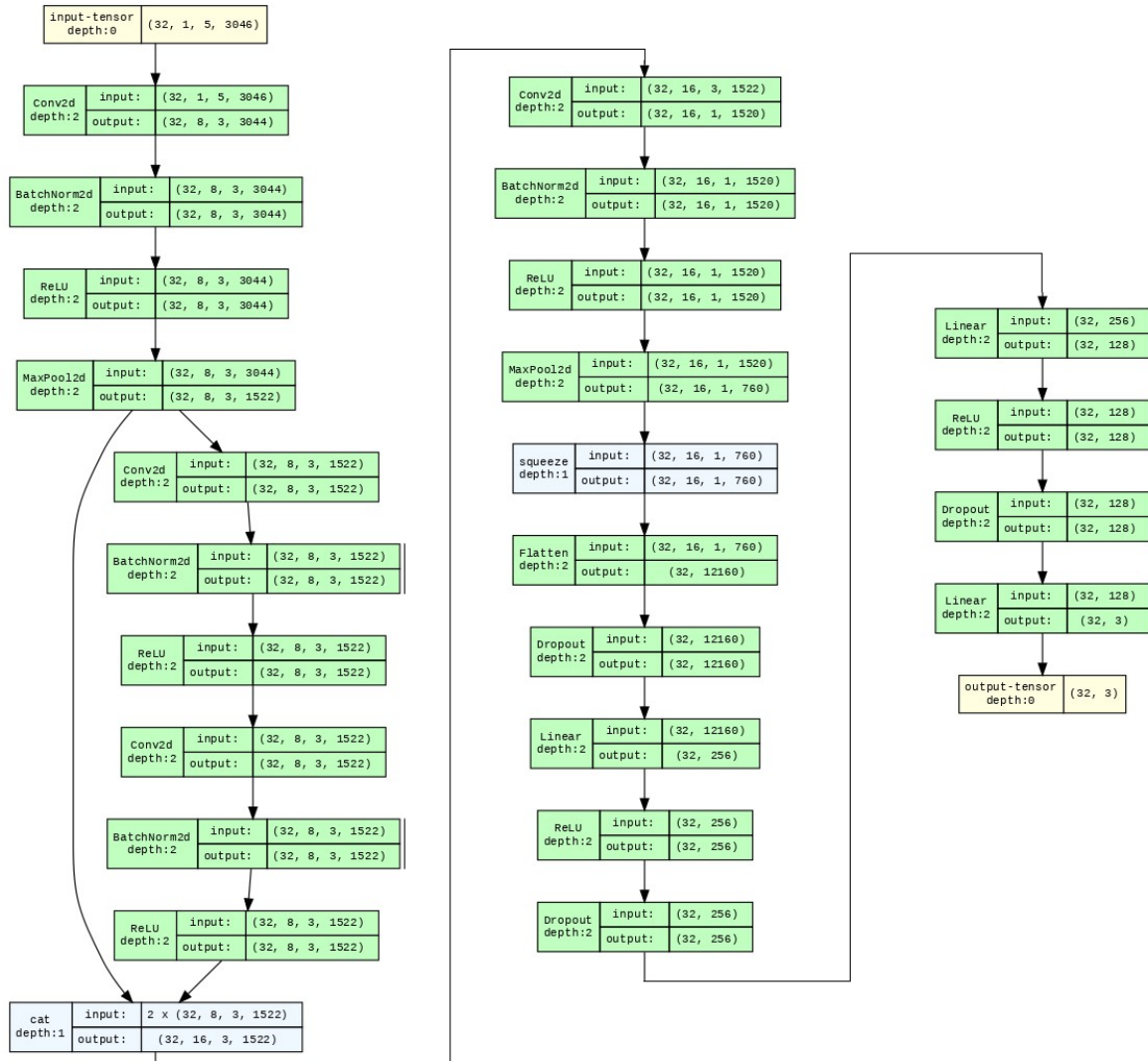


Figure F.1: Small size CNN architecture implemented, referred as *CNN-Small*.

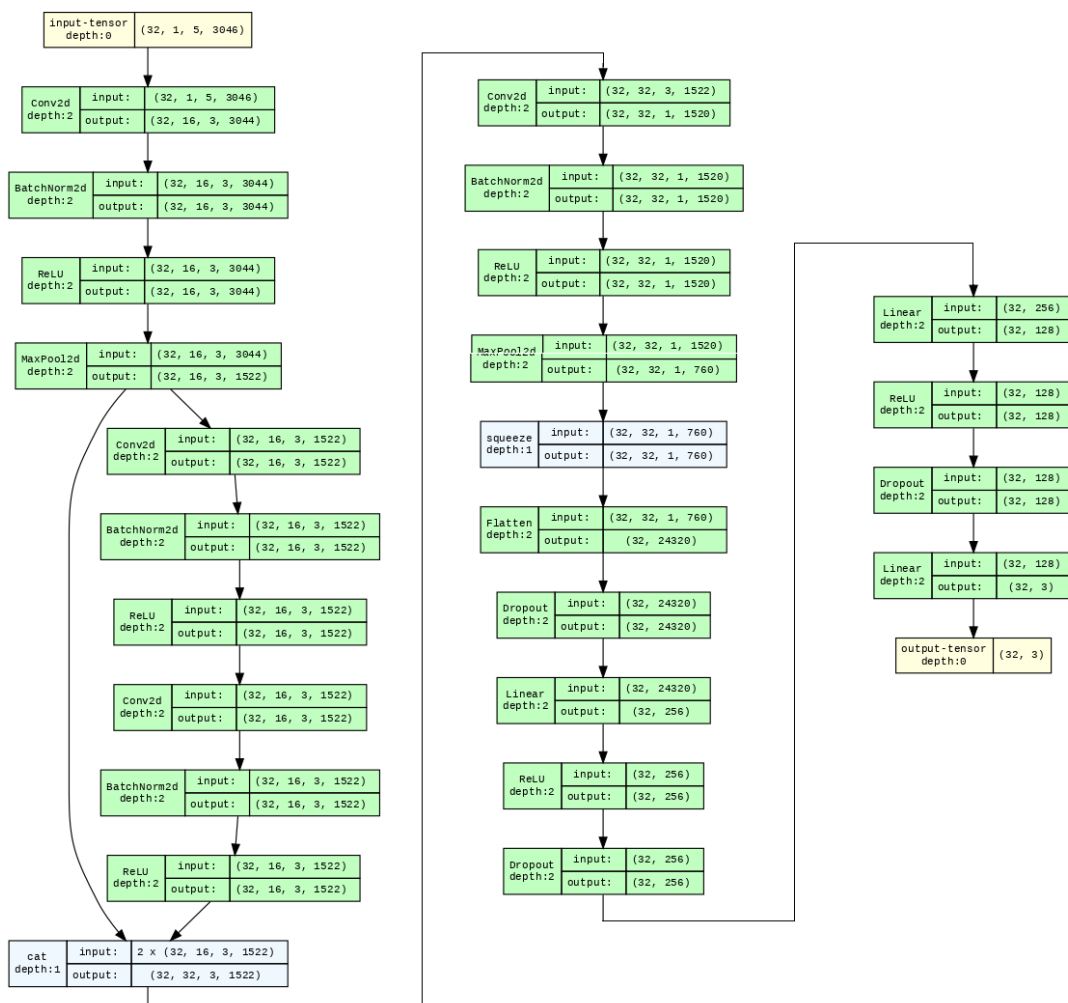


Figure F.2: Medium size CNN architecture implemented, referred as *CNN-Medium*.

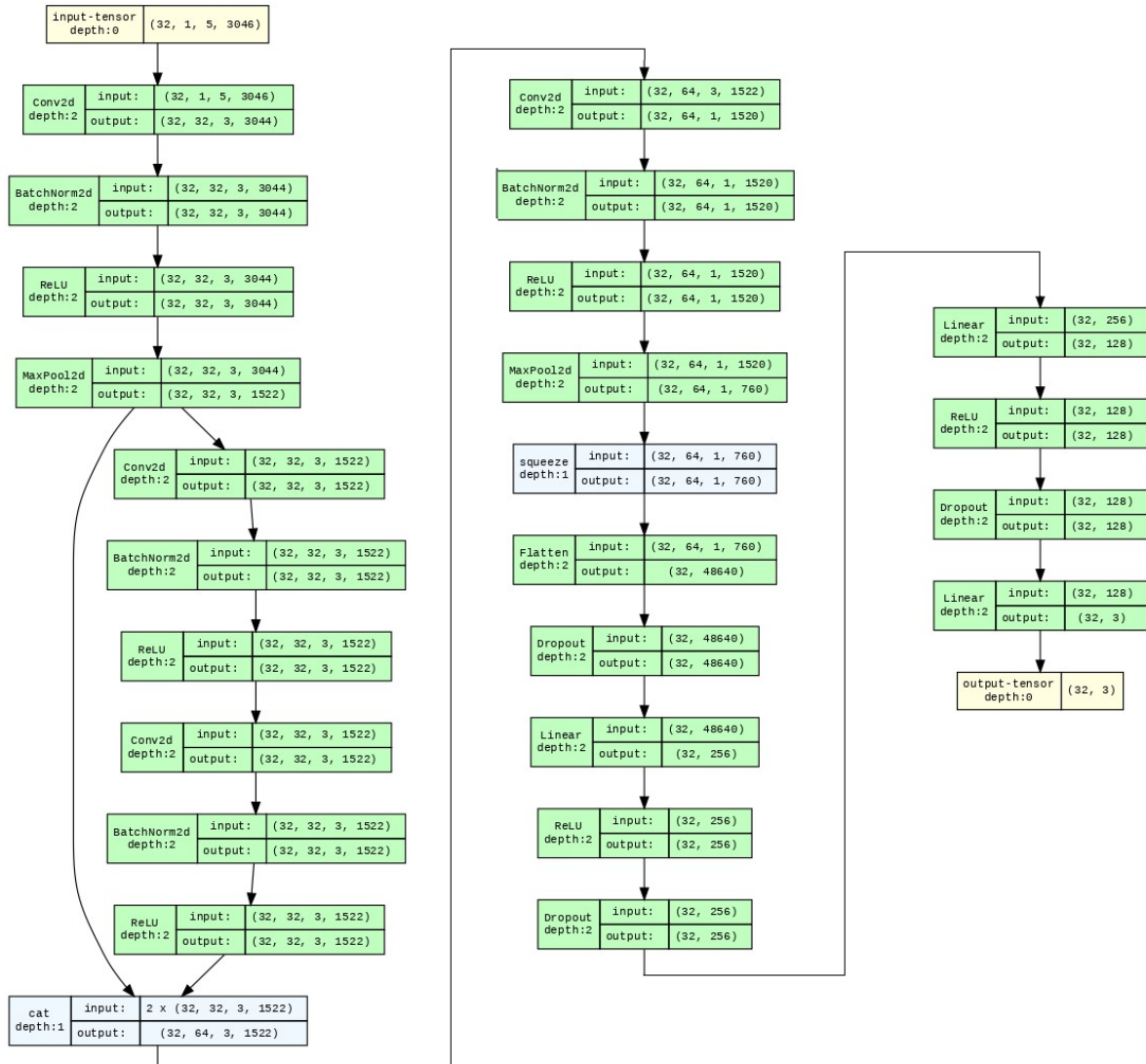


Figure F.3: Large size CNN architecture implemented, referred as *CNN-Large*.

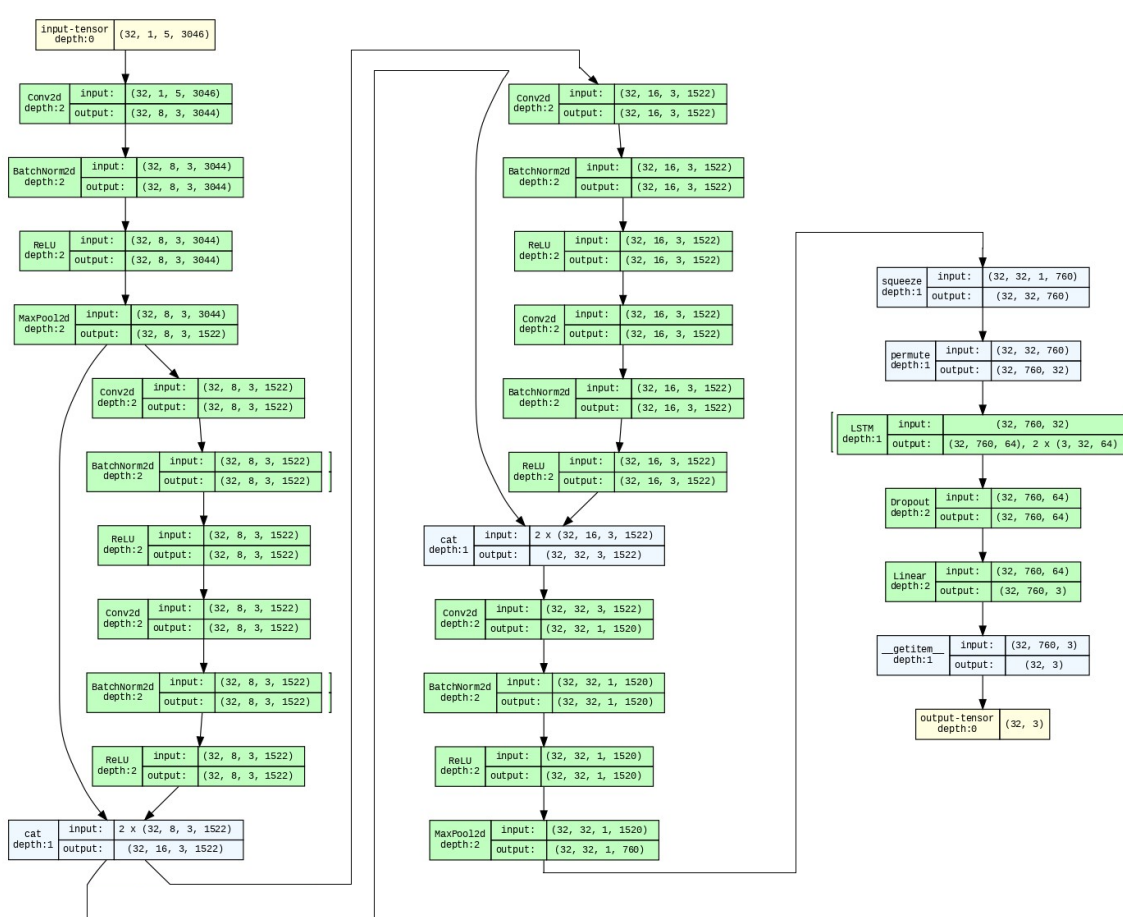


Figure F.4: Small size CNN+LSTM architecture implemented, referred as *CNN+LSTM-Small*.

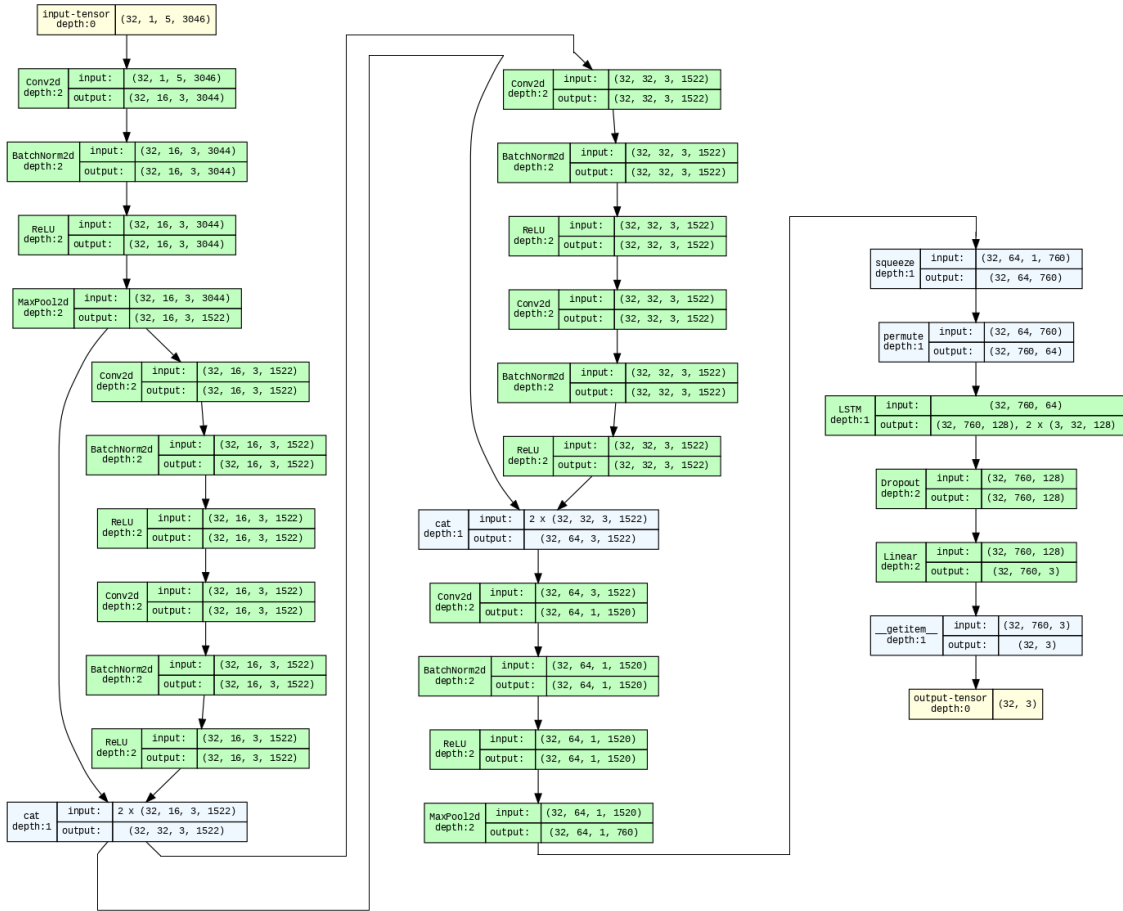


Figure F.5: Medium size CNN+LSTM architecture implemented, referred as *CNN+LSTM-Medium*.

Appendix G

Additional results for models comparison

In the next pages, additional Figures to the results presented in Chapter 2 are shown. The same section names from the Results section from Chapter 2 were used to facilitate the correspondence between the two parts.

G.1 Comparison of similar models

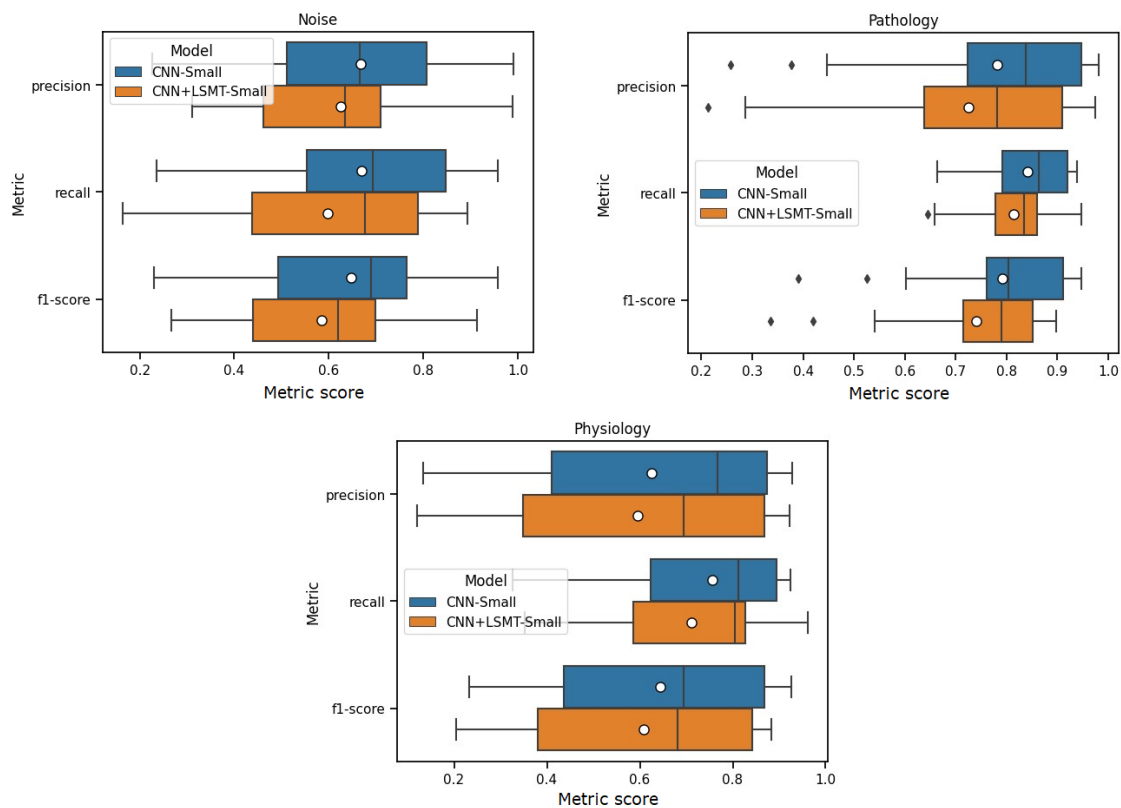


Figure G.1: Comparison between the different deep learning models after architecture selection per class. CNN-Small and CNN+LSTM-Small correspond to the smallest implemented CNN and CNN+LSTM models respectively.

G.2 Deep learning vs traditional machine learning evaluation

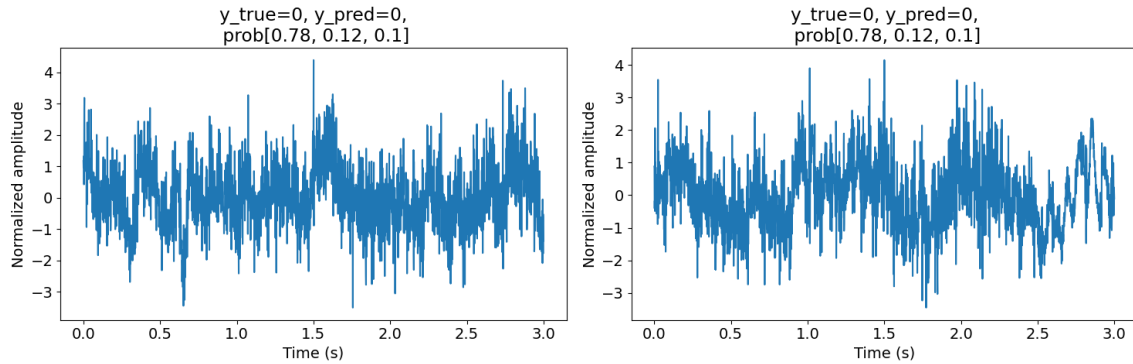
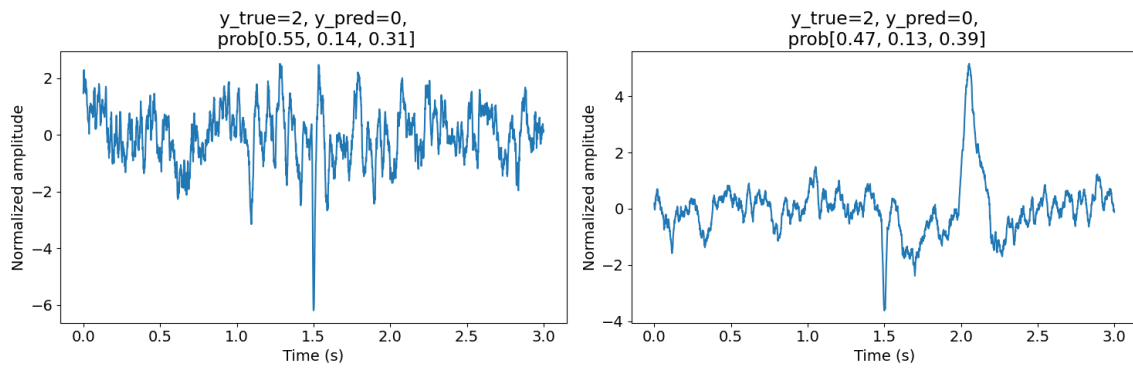


Figure G.2: Examples of correctly classified samples from the *noise* class.

Incorrect prediction



Correct prediction

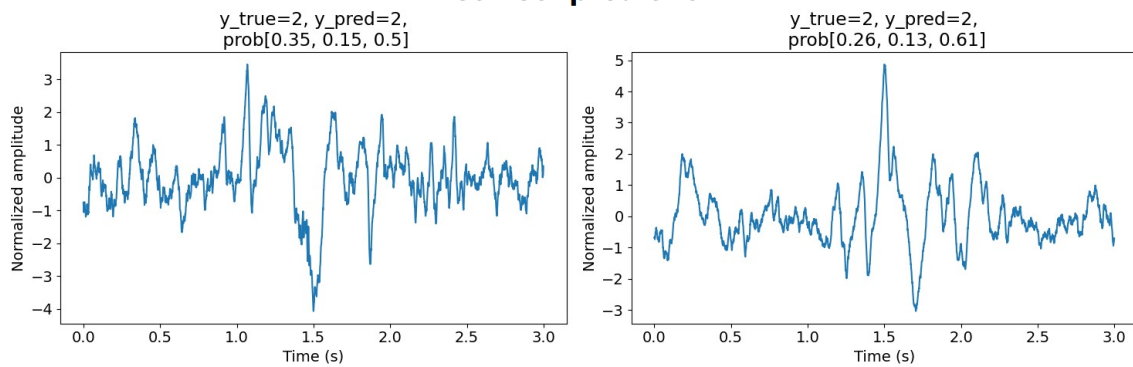
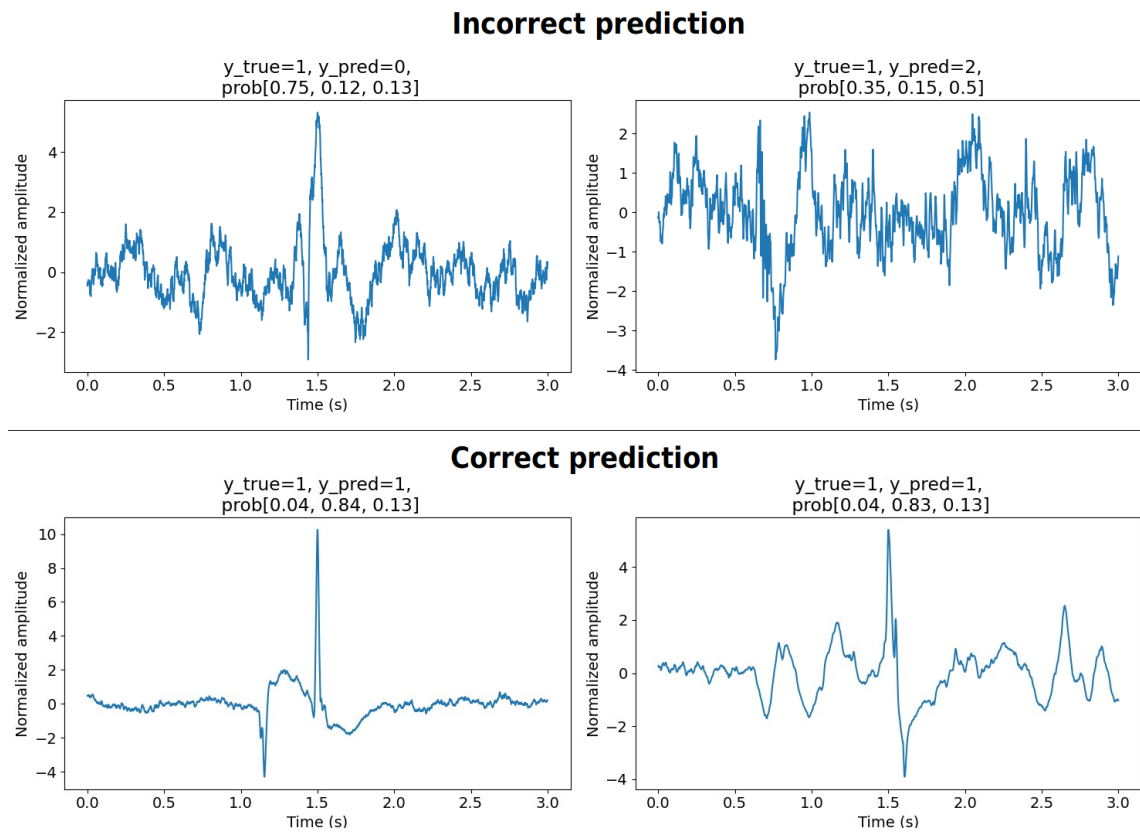


Figure G.3: Examples of correctly and incorrectly classified samples from the *baseline* or *physiology* class

Figure G.4: Examples of correctly and incorrectly classified samples from the *pathology* class

Appendix H

Copyright Transfers and Reprint Permissions

In the next pages, the permissions to reproduce Figures 1.1, 1.2 and 1.5 are found.

JOHN WILEY AND SONS LICENSE
TERMS AND CONDITIONS

May 20, 2024

This Agreement between Mr. Mauricio Cespedes Tenorio ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

License Number	5793210066470
License date	May 20, 2024
Licensed Content Publisher	John Wiley and Sons
Licensed Content Publication	Epilepsia
Licensed Content Title	ILAE classification of the epilepsies: Position paper of the ILAE Commission for Classification and Terminology
Licensed Content Author	Sameer M. Zuberi, Yue-Hua Zhang, Samuel Wiebe, et al
Licensed Content Date	Mar 8, 2017
Licensed Content Volume	58
Licensed Content Issue	4
Licensed Content Pages	10
Type of use	Dissertation/Thesis
Requestor type	University/Academic
Format	Electronic

Portion	Figure/table
Number of figures/tables	1
Will you be translating?	No
Title of new work	Data preparation and machine learning for intracranial electroencephalography
Institution name	University of Western Ontario
Expected presentation date	Jul 2024
Portions	Figure 1
	Mr. Mauricio Cespedes Tenorio
Requestor Location	
	Attn: Western University
Publisher Tax ID	EU826007151
Total	0.00 USD

Terms and Conditions

TERMS AND CONDITIONS

This copyrighted material is owned by or exclusively licensed to John Wiley & Sons, Inc. or one of its group companies (each a "Wiley Company") or handled on behalf of a society with which a Wiley Company has exclusive publishing rights in relation to a particular work (collectively "WILEY"). By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the billing and payment terms and conditions established by the Copyright Clearance Center Inc., ("CCC's Billing and Payment terms and conditions"), at the time that you opened your RightsLink account (these are available at any time at <http://myaccount.copyright.com>).

Terms and Conditions

- The materials you have requested permission to reproduce or reuse (the "Wiley Materials") are protected by copyright.
- You are hereby granted a personal, non-exclusive, non-sub licensable (on a stand-alone basis), non-transferable, worldwide, limited license to reproduce the Wiley Materials for the purpose specified in the licensing process. This license, **and any CONTENT (PDF or image file) purchased as part of your order**, is for a one-time use only and limited to any maximum distribution number specified in the license. The first instance of republication or reuse granted by this license must be completed within two years of the date of the grant of this license (although copies prepared before the end date may be distributed thereafter). The Wiley Materials shall not be used in any other manner or for any other purpose, beyond what is granted in the license. Permission is granted subject to an appropriate acknowledgement given to the author, title of the material/book/journal and the publisher. You shall also duplicate the copyright notice that appears in the Wiley publication in your use of the Wiley Material. Permission is also granted on the understanding that nowhere in the text is a previously published source acknowledged for all or part of this Wiley Material. Any third party content is expressly excluded from this permission.
- With respect to the Wiley Materials, all rights are reserved. Except as expressly granted by the terms of the license, no part of the Wiley Materials may be copied, modified, adapted (except for minor reformatting required by the new Publication), translated, reproduced, transferred or distributed, in any form or by any means, and no derivative works may be made based on the Wiley Materials without the prior permission of the respective copyright owner. **For STM Signatory Publishers clearing permission under the terms of the [STM Permissions Guidelines](#) only, the terms of the license are extended to include subsequent editions and for editions in other languages, provided such editions are for the work as a whole in situ and does not involve the separate exploitation of the permitted figures or extracts**, You may not alter, remove or suppress in any manner any copyright, trademark or other notices displayed by the Wiley Materials. You may not license, rent, sell, loan, lease, pledge, offer as security, transfer or assign the Wiley Materials on a stand-alone basis, or any of the rights granted to you hereunder to any other person.
- The Wiley Materials and all of the intellectual property rights therein shall at all times remain the exclusive property of John Wiley & Sons Inc, the Wiley Companies, or their respective licensors, and your interest therein is only that of having possession of and the right to reproduce the Wiley Materials pursuant to Section 2 herein during the continuance of this Agreement. You agree that you own no right, title or interest in or to the Wiley Materials or any of the intellectual property rights therein. You shall have no rights hereunder other than the license as provided for above in Section 2. No right, license or interest to any trademark, trade name, service mark or other branding ("Marks") of WILEY or its licensors is granted hereunder, and you agree that you shall not assert any such right, license or interest with respect thereto
- NEITHER WILEY NOR ITS LICENSORS MAKES ANY WARRANTY OR REPRESENTATION OF ANY KIND TO YOU OR ANY THIRD PARTY, EXPRESS, IMPLIED OR STATUTORY, WITH RESPECT TO THE MATERIALS OR THE ACCURACY OF ANY INFORMATION CONTAINED IN THE MATERIALS, INCLUDING, WITHOUT LIMITATION, ANY IMPLIED WARRANTY OF MERCHANTABILITY, ACCURACY, SATISFACTORY QUALITY, FITNESS FOR A PARTICULAR PURPOSE, USABILITY, INTEGRATION OR NON-INFRINGEMENT AND ALL SUCH WARRANTIES ARE HEREBY EXCLUDED BY WILEY AND ITS LICENSORS AND WAIVED BY YOU.

- WILEY shall have the right to terminate this Agreement immediately upon breach of this Agreement by you.
- You shall indemnify, defend and hold harmless WILEY, its Licensors and their respective directors, officers, agents and employees, from and against any actual or threatened claims, demands, causes of action or proceedings arising from any breach of this Agreement by you.
- IN NO EVENT SHALL WILEY OR ITS LICENSORS BE LIABLE TO YOU OR ANY OTHER PARTY OR ANY OTHER PERSON OR ENTITY FOR ANY SPECIAL, CONSEQUENTIAL, INCIDENTAL, INDIRECT, EXEMPLARY OR PUNITIVE DAMAGES, HOWEVER CAUSED, ARISING OUT OF OR IN CONNECTION WITH THE DOWNLOADING, PROVISIONING, VIEWING OR USE OF THE MATERIALS REGARDLESS OF THE FORM OF ACTION, WHETHER FOR BREACH OF CONTRACT, BREACH OF WARRANTY, TORT, NEGLIGENCE, INFRINGEMENT OR OTHERWISE (INCLUDING, WITHOUT LIMITATION, DAMAGES BASED ON LOSS OF PROFITS, DATA, FILES, USE, BUSINESS OPPORTUNITY OR CLAIMS OF THIRD PARTIES), AND WHETHER OR NOT THE PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. THIS LIMITATION SHALL APPLY NOTWITHSTANDING ANY FAILURE OF ESSENTIAL PURPOSE OF ANY LIMITED REMEDY PROVIDED HEREIN.
- Should any provision of this Agreement be held by a court of competent jurisdiction to be illegal, invalid, or unenforceable, that provision shall be deemed amended to achieve as nearly as possible the same economic effect as the original provision, and the legality, validity and enforceability of the remaining provisions of this Agreement shall not be affected or impaired thereby.
- The failure of either party to enforce any term or condition of this Agreement shall not constitute a waiver of either party's right to enforce each and every term and condition of this Agreement. No breach under this agreement shall be deemed waived or excused by either party unless such waiver or consent is in writing signed by the party granting such waiver or consent. The waiver by or consent of a party to a breach of any provision of this Agreement shall not operate or be construed as a waiver of or consent to any other or subsequent breach by such other party.
- This Agreement may not be assigned (including by operation of law or otherwise) by you without WILEY's prior written consent.
- Any fee required for this permission shall be non-refundable after thirty (30) days from receipt by the CCC.
- These terms and conditions together with CCC's Billing and Payment terms and conditions (which are incorporated herein) form the entire agreement between you and WILEY concerning this licensing transaction and (in the absence of fraud) supersedes all prior agreements and representations of the parties, oral or written. This Agreement may not be amended except in writing signed by both parties. This Agreement shall be binding upon and inure to the benefit of the parties' successors, legal representatives, and authorized assigns.
- In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall prevail.
- WILEY expressly reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

- This Agreement will be void if the Type of Use, Format, Circulation, or Requestor Type was misrepresented during the licensing process.
- This Agreement shall be governed by and construed in accordance with the laws of the State of New York, USA, without regards to such state's conflict of law rules. Any legal action, suit or proceeding arising out of or relating to these Terms and Conditions or the breach thereof shall be instituted in a court of competent jurisdiction in New York County in the State of New York in the United States of America and each party hereby consents and submits to the personal jurisdiction of such court, waives any objection to venue in such court and consents to service of process by registered or certified mail, return receipt requested, at the last known address of such party.

WILEY OPEN ACCESS TERMS AND CONDITIONS

Wiley Publishes Open Access Articles in fully Open Access Journals and in Subscription journals offering Online Open. Although most of the fully Open Access journals publish open access articles under the terms of the Creative Commons Attribution (CC BY) License only, the subscription journals and a few of the Open Access Journals offer a choice of Creative Commons Licenses. The license type is clearly identified on the article.

The Creative Commons Attribution License

The [Creative Commons Attribution License \(CC-BY\)](#) allows users to copy, distribute and transmit an article, adapt the article and make commercial use of the article. The CC-BY license permits commercial and non-

Creative Commons Attribution Non-Commercial License

The [Creative Commons Attribution Non-Commercial \(CC-BY-NC\) License](#) permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.(see below)

Creative Commons Attribution-Non-Commercial-NoDerivs License

The [Creative Commons Attribution Non-Commercial-NoDerivs License \(CC-BY-NC-ND\)](#) permits use, distribution and reproduction in any medium, provided the original work is properly cited, is not used for commercial purposes and no modifications or adaptations are made. (see below)

Use by commercial "for-profit" organizations

Use of Wiley Open Access articles for commercial, promotional, or marketing purposes requires further explicit permission from Wiley and will be subject to a fee.

Further details can be found on Wiley Online Library
<http://olabout.wiley.com/WileyCDA/Section/id-410895.html>

Other Terms and Conditions:

v1.10 Last updated September 2015

Questions? customercare@copyright.com.



SPRINGER NATURE LICENSE
TERMS AND CONDITIONS

May 20, 2024

This Agreement between Mr. Mauricio Cespedes Tenorio ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	5793210564080
License date	May 20, 2024
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Nature Reviews Neurology
Licensed Content Title	Changing concepts in presurgical assessment for epilepsy surgery
Licensed Content Author	Maeike Zijlmans et al
Licensed Content Date	Jul 24, 2019
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Would you like a high resolution image with your order?	no
Will you be translating?	no

Circulation/distribution	1 - 29
Author of this Springer Nature content	no
Title of new work	Data preparation and machine learning for intracranial electroencephalography
Institution name	University of Western Ontario
Expected presentation date	Jul 2024
Portions	Figure 1
	Mr. Mauricio Cespedes Tenorio
Requestor Location	
	Attn: Western University
Total	0.00 USD

Terms and Conditions

Springer Nature Customer Service Centre GmbH Terms and Conditions

The following terms and conditions ("Terms and Conditions") together with the terms specified in your [RightsLink] constitute the License ("License") between you as Licensee and Springer Nature Customer Service Centre GmbH as Licensor. By clicking 'accept' and completing the transaction for your use of the material ("Licensed Material"), you confirm your acceptance of and obligation to be bound by these Terms and Conditions.

1. Grant and Scope of License

1. 1. The Licensor grants you a personal, non-exclusive, non-transferable, non-sublicensable, revocable, world-wide License to reproduce, distribute, communicate to the public, make available, broadcast, electronically transmit or create derivative works using the Licensed Material for the purpose(s) specified in your RightsLink Licence Details only. Licenses are granted for the specific use requested in the order and for no other use, subject to these Terms and Conditions. You acknowledge and agree that the rights granted to you under this License do not include the right to modify, edit, translate, include in collective works, or create derivative works of the Licensed Material in whole or in part unless expressly stated in your RightsLink Licence Details. You may use the Licensed Material only as permitted under this Agreement and will not reproduce, distribute, display, perform, or otherwise use or exploit any Licensed Material in any way, in whole or in part, except as expressly permitted by this License.

1. 2. You may only use the Licensed Content in the manner and to the extent permitted by these Terms and Conditions, by your RightsLink Licence Details and by any applicable laws.

1. 3. A separate license may be required for any additional use of the Licensed Material, e.g. where a license has been purchased for print use only, separate permission must be obtained for electronic re-use. Similarly, a License is only valid in the language selected and does not apply for editions in other languages unless additional translation rights have been granted separately in the License.

1. 4. Any content within the Licensed Material that is owned by third parties is expressly excluded from the License.

1. 5. Rights for additional reuses such as custom editions, computer/mobile applications, film or TV reuses and/or any other derivative rights requests require additional permission and may be subject to an additional fee. Please apply to journalpermissions@springernature.com or bookpermissions@springernature.com for these rights.

2. Reservation of Rights

Licensor reserves all rights not expressly granted to you under this License. You acknowledge and agree that nothing in this License limits or restricts Licensor's rights in or use of the Licensed Material in any way. Neither this License, nor any act, omission, or statement by Licensor or you, conveys any ownership right to you in any Licensed Material, or to any element or portion thereof. As between Licensor and you, Licensor owns and retains all right, title, and interest in and to the Licensed Material subject to the license granted in Section 1.1. Your permission to use the Licensed Material is expressly conditioned on you not impairing Licensor's or the applicable copyright owner's rights in the Licensed Material in any way.

3. Restrictions on use

3. 1. Minor editing privileges are allowed for adaptations for stylistic purposes or formatting purposes provided such alterations do not alter the original meaning or intention of the Licensed Material and the new figure(s) are still accurate and representative of the Licensed Material. Any other changes including but not limited to, cropping, adapting, and/or omitting material that affect the meaning, intention or moral rights of the author(s) are strictly prohibited.

3. 2. You must not use any Licensed Material as part of any design or trademark.

3. 3. Licensed Material may be used in Open Access Publications (OAP), but any such reuse must include a clear acknowledgment of this permission visible at the same time as the figures/tables/illustration or abstract and which must indicate that the Licensed Material is not part of the governing OA license but has been reproduced with permission. This may be indicated according to any standard referencing system but must include at a minimum 'Book/Journal title, Author, Journal Name (if applicable), Volume (if applicable), Publisher, Year, reproduced with permission from SNCSC'.

4. STM Permission Guidelines

4. 1. An alternative scope of license may apply to signatories of the STM Permissions Guidelines ("STM PG") as amended from time to time and made available at <https://www.stm-assoc.org/intellectual-property/permissions/permissions-guidelines/>.

4. 2. For content reuse requests that qualify for permission under the STM PG, and which may be updated from time to time, the STM PG supersede the terms and

conditions contained in this License.

4. 3. If a License has been granted under the STM PG, but the STM PG no longer apply at the time of publication, further permission must be sought from the Rightsholder. Contact journalpermissions@springernature.com or bookpermissions@springernature.com for these rights.

5. Duration of License

5. 1. Unless otherwise indicated on your License, a License is valid from the date of purchase ("License Date") until the end of the relevant period in the below table:

Reuse in a medical communications project	Reuse up to distribution or time period indicated in License
Reuse in a dissertation/thesis	Lifetime of thesis
Reuse in a journal/magazine	Lifetime of journal/magazine
Reuse in a book/textbook	Lifetime of edition
Reuse on a website	1 year unless otherwise specified in the License
Reuse in a presentation/slide kit/poster	Lifetime of presentation/slide kit/poster. Note: publication whether electronic or in print of presentation/slide kit/poster may require further permission.
Reuse in conference proceedings	Lifetime of conference proceedings
Reuse in an annual report	Lifetime of annual report
Reuse in training/CME materials	Reuse up to distribution or time period indicated in License
Reuse in newsmedia	Lifetime of newsmedia
Reuse in coursepack/classroom materials	Reuse up to distribution and/or time period indicated in license

6. Acknowledgement

6. 1. The Licensor's permission must be acknowledged next to the Licensed Material in print. In electronic form, this acknowledgement must be visible at the same time as the figures/tables/illustrations or abstract and must be hyperlinked to the journal/book's homepage.

6. 2. Acknowledgement may be provided according to any standard referencing system and at a minimum should include "Author, Article/Book Title, Journal name/Book imprint, volume, page number, year, Springer Nature".

7. Reuse in a dissertation or thesis

7. 1. Where 'reuse in a dissertation/thesis' has been selected, the following terms apply: Print rights of the Version of Record are provided for; electronic rights for use only on institutional repository as defined by the Sherpa guideline (www.sherpa.ac.uk/romeo/) and only up to what is required by the awarding institution.

7. 2. For theses published under an ISBN or ISSN, separate permission is required. Please contact journalpermissions@springernature.com or bookpermissions@springernature.com for these rights.

7. 3. Authors must properly cite the published manuscript in their thesis according to current citation standards and include the following acknowledgement: *'Reproduced with permission from Springer Nature'*.

8. License Fee

You must pay the fee set forth in the License Agreement (the "License Fees"). All amounts payable by you under this License are exclusive of any sales, use, withholding, value added or similar taxes, government fees or levies or other assessments. Collection and/or remittance of such taxes to the relevant tax authority shall be the responsibility of the party who has the legal obligation to do so.

9. Warranty

9. 1. The Licensor warrants that it has, to the best of its knowledge, the rights to license reuse of the Licensed Material. **You are solely responsible for ensuring that the material you wish to license is original to the Licensor and does not carry the copyright of another entity or third party (as credited in the published version).** If the credit line on any part of the Licensed Material indicates that it was reprinted or adapted with permission from another source, then you should seek additional permission from that source to reuse the material.

9. 2. EXCEPT FOR THE EXPRESS WARRANTY STATED HEREIN AND TO THE EXTENT PERMITTED BY APPLICABLE LAW, LICENSOR PROVIDES THE LICENSED MATERIAL "AS IS" AND MAKES NO OTHER REPRESENTATION OR WARRANTY. LICENSOR EXPRESSLY DISCLAIMS ANY LIABILITY FOR ANY CLAIM ARISING FROM OR OUT OF THE CONTENT, INCLUDING BUT NOT LIMITED TO ANY ERRORS, INACCURACIES, OMISSIONS, OR DEFECTS CONTAINED THEREIN, AND ANY IMPLIED OR EXPRESS WARRANTY AS TO MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. IN NO EVENT SHALL LICENSOR BE LIABLE TO YOU OR ANY OTHER PARTY OR ANY OTHER PERSON OR FOR ANY SPECIAL, CONSEQUENTIAL, INCIDENTAL, INDIRECT, PUNITIVE, OR EXEMPLARY DAMAGES, HOWEVER CAUSED, ARISING OUT OF OR IN CONNECTION WITH THE DOWNLOADING, VIEWING OR USE OF THE LICENSED MATERIAL REGARDLESS OF THE FORM OF ACTION, WHETHER FOR BREACH OF CONTRACT, BREACH OF WARRANTY, TORT, NEGLIGENCE, INFRINGEMENT OR OTHERWISE (INCLUDING, WITHOUT LIMITATION, DAMAGES BASED ON LOSS OF PROFITS, DATA, FILES, USE, BUSINESS OPPORTUNITY OR CLAIMS OF THIRD PARTIES), AND WHETHER OR NOT THE PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. THIS LIMITATION APPLIES NOTWITHSTANDING ANY FAILURE OF ESSENTIAL PURPOSE OF ANY LIMITED REMEDY PROVIDED HEREIN.

10. Termination and Cancellation

10. 1. The License and all rights granted hereunder will continue until the end of the applicable period shown in Clause 5.1 above. Thereafter, this license will be terminated and all rights granted hereunder will cease.

10. 2. Licensor reserves the right to terminate the License in the event that payment is not received in full or if you breach the terms of this License.

11. General

11. 1. The License and the rights and obligations of the parties hereto shall be construed, interpreted and determined in accordance with the laws of the Federal Republic of Germany without reference to the stipulations of the CISG (United

Nations Convention on Contracts for the International Sale of Goods) or to Germany's choice-of-law principle.

11. 2. The parties acknowledge and agree that any controversies and disputes arising out of this License shall be decided exclusively by the courts of or having jurisdiction for Heidelberg, Germany, as far as legally permissible.

11. 3. This License is solely for Licensor's and Licensee's benefit. It is not for the benefit of any other person or entity.

Questions? For questions on Copyright Clearance Center accounts or website issues please contact springernaturesupport@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777. For questions on Springer Nature licensing please visit <https://www.springernature.com/gp/partners/rights-permissions-third-party-distribution>

Other Conditions:

Version 1.4 - Dec 2022

Questions? customercare@copyright.com.

OXFORD UNIVERSITY PRESS LICENSE
TERMS AND CONDITIONS

May 20, 2024

This Agreement between Mr. Mauricio Cespedes Tenorio ("You") and Oxford University Press ("Oxford University Press") consists of your license details and the terms and conditions provided by Oxford University Press and Copyright Clearance Center.

License Number 5793211283572

License date May 20, 2024

Licensed content publisher Oxford University Press

Licensed content publication Brain

Licensed content title Atlas of the normal intracranial electroencephalogram: neurophysiological awake activity in different cortical areas

Licensed content author Frauscher, Birgit; von Ellenrieder, Nicolas

Licensed content date Mar 1, 2018

Type of Use Thesis/Dissertation

Institution name

Title of your work Data preparation and machine learning for intracranial electroencephalography

Publisher of your work University of Western Ontario

Expected publication date Jul 2024

Permissions cost 0.00 USD

Value added tax 0.00 USD

Total 0.00 USD

Title of new work Data preparation and machine learning for intracranial electroencephalography

Institution name University of Western Ontario

Expected presentation date Jul 2024

Portions Figure 4

Mr. Mauricio Cespedes Tenorio

Requestor Location

Attn: Western University

Publisher Tax ID GB125506730

Total 0.00 USD

Terms and Conditions

**STANDARD TERMS AND CONDITIONS FOR REPRODUCTION OF
MATERIAL FROM AN OXFORD UNIVERSITY PRESS JOURNAL**

1. Use of the material is restricted to the type of use specified in your order details.
2. This permission covers the use of the material in the **English** language in the following territory: *world*. If you have requested additional permission to translate this material, the terms and conditions of this reuse will be set out in clause 12.
3. This permission is limited to the particular use authorized in (1) above and does not allow you to sanction its use elsewhere in any other format other than specified above, nor does it apply to quotations, images, artistic works etc that have been reproduced from other sources which may be part of the material to be used.
4. No alteration, omission or addition is made to the material without our written consent. Permission **must** be re-cleared with Oxford University Press if/when you decide to reprint.

5. The following credit line appears wherever the material is used: author, title, journal, year, volume, issue number, pagination, by permission of Oxford University Press or the sponsoring society if the journal is a society journal. Where a journal is being published on behalf of a learned society, the details of that society **must** be included in the credit line.
6. For the reproduction of a full article from an Oxford University Press journal for whatever purpose, the corresponding author of the material concerned should be informed of the proposed use. Contact details for the corresponding authors of all Oxford University Press journal contact can be found alongside either the abstract or full text of the article concerned, accessible from www.oxfordjournals.org Should there be a problem clearing these rights, please contact journals.permissions@oup.com
7. If the credit line or acknowledgement in our publication indicates that any of the figures, images or photos was reproduced, drawn or modified from an earlier source it will be necessary for you to clear this permission with the original publisher as well. If this permission has not been obtained, please note that this material cannot be included in your publication/photocopies.
8. While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by Oxford University Press or by Copyright Clearance Center (CCC)) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and Oxford University Press reserves the right to take any and all action to protect its copyright in the materials.
9. This license is personal to you and may not be sublicensed, assigned or transferred by you to any other person without Oxford University Press's written permission.
10. Oxford University Press reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.
11. You hereby indemnify and agree to hold harmless Oxford University Press and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.
12. Unless otherwise specified, inclusion under a Creative Commons license or any other Open Access license allowing onward reuse is prohibited.
13. Where permission to translate has been agreed, the credit line and any copyright/disclaimer notices provided by OUP shall be included on the same page as the Material, translated into the language of the new work, except for trademark names:

Translated and reproduced by permission of Oxford University Press on behalf of the <Society Name>. Translation Disclaimer: OUP and the <Society Name> are not responsible or in any way liable for the accuracy of the translation. The Licensee is solely responsible for the translation in this publication/reprint.
14. Other Terms and Conditions:

Questions? customer@copyright.com.





This is a License Agreement between Mauricio Cespedes Tenorio ("User") and Copyright Clearance Center, Inc. ("CCC") on behalf of the Rightsholder identified in the order details below. The license consists of the order details, the Marketplace Permissions General Terms and Conditions below, and any Rightsholder Terms and Conditions which are included below.

All payments must be made in full to CCC in accordance with the Marketplace Permissions General Terms and Conditions below.

Order Date	21-Jun-2024	Type of Use	Republish in a thesis/dissertation
Order License ID	1496964-1	Publisher Portion	John Wiley Image/photo/illustration
ISSN	1950-6945		

LICENSED CONTENT

Publication Title	Epileptic disorders	Publication Type	e-Journal
Article Title	Learn how to interpret and use intracranial EEG findings	Start Page	1
Date	01/01/1999	End Page	59
Language	English	Issue	1
Country	United States of America	Volume	26
Rightsholder	John Wiley & Sons - Books	URL	http://www.epilepticdisorders.com

REQUEST DETAILS

Portion Type	Image/photo/illustration	Distribution	Worldwide
Number of Images / Photos / Illustrations	2	Translation	Original language of publication
Format (select all that apply)	Electronic	Copies for the Disabled?	No
Who Will Republish the Content?	Academic institution	Minor Editing Privileges?	No
Duration of Use	Life of current edition	Incidental Promotional Use?	No
Lifetime Unit Quantity	Up to 499	Currency	USD
Rights Requested	Main product		

NEW WORK DETAILS

Title	Data Preparation and Machine Learning for Intracranial Electroencephalography	Institution Name	University of Western Ontario
Instructor Name	Mauricio Cespedes Tenorio	Expected Presentation Date	2024-07-19

ADDITIONAL DETAILS

The Requesting Person / Organization to Appear on the License	Mauricio Cespedes Tenorio
--	---------------------------

REQUESTED CONTENT DETAILS

Title, Description or Numeric Reference of the Portion(s)	Figure 5 and Figure 7	Title of the Article / Chapter the Portion Is From	Learn how to interpret and use intracranial EEG findings
Editor of Portion(s)	Frauscher, B.; Mansilla, D.; Abdallah, C.; Astner-Rohracher, A.; Beniczky, S.; Brazdil, M.; Gnatkovsky, V.; Jacobs, J.; Kalamangalam, G.; Perucca, P.; Ryvlin, P.; Schuele, S.; Tao, J.; Wang, Y.; Zijlmans, M.; McGonigal, A.	Author of Portion(s)	Frauscher, B.; Mansilla, D.; Abdallah, C.; Astner-Rohracher, A.; Beniczky, S.; Brazdil, M.; Gnatkovsky, V.; Jacobs, J.; Kalamangalam, G.; Perucca, P.; Ryvlin, P.; Schuele, S.; Tao, J.; Wang, Y.; Zijlmans, M.; McGonigal, A.
Volume / Edition	26	Publication Date of Portion	2024-02-23
Page or Page Range of Portion	1-59		

RIGHTSHOLDER TERMS AND CONDITIONS

No right, license or interest to any trademark, trade name, service mark or other branding ("Marks") of WILEY or its licensors is granted hereunder, and you agree that you shall not assert any such right, license or interest with respect thereto. You may not alter, remove or suppress in any manner any copyright, trademark or other notices displayed by the Wiley material. This Agreement will be void if the Type of Use, Format, Circulation, or Requestor Type was misrepresented during the licensing process. In no instance may the total amount of Wiley Materials used in any Main Product, Compilation or Collective work comprise more than 5% (if figures/tables) or 15% (if full articles/chapters) of the (entirety of the) Main Product, Compilation or Collective Work. Some titles may be available under an Open Access license. It is the Licensors' responsibility to identify the type of Open Access license on which the requested material was published, and comply fully with the terms of that license for the type of use specified. Further details can be found on Wiley Online Library <http://olabout.wiley.com/WileyCDA/Section/id-410895.html>.

Marketplace Permissions General Terms and Conditions

The following terms and conditions ("General Terms"), together with any applicable Publisher Terms and Conditions, govern User's use of Works pursuant to the Licenses granted by Copyright Clearance Center, Inc. ("CCC") on behalf of the applicable Rightsholders of such Works through CCC's applicable Marketplace transactional licensing services (each, a "Service").

1) **Definitions.** For purposes of these General Terms, the following definitions apply:

"License" is the licensed use the User obtains via the Marketplace platform in a particular licensing transaction, as set forth in the Order Confirmation.

"Order Confirmation" is the confirmation CCC provides to the User at the conclusion of each Marketplace transaction. "Order Confirmation Terms" are additional terms set forth on specific Order Confirmations not set forth in the General Terms that can include terms applicable to a particular CCC transactional licensing service and/or any Rightsholder-specific terms.

"Rightsholder(s)" are the holders of copyright rights in the Works for which a User obtains licenses via the Marketplace platform, which are displayed on specific Order Confirmations.

"Terms" means the terms and conditions set forth in these General Terms and any additional Order Confirmation Terms collectively.

"User" or "you" is the person or entity making the use granted under the relevant License. Where the person accepting the Terms on behalf of a User is a freelancer or other third party who the User authorized to accept the General Terms on the User's behalf, such person shall be deemed jointly a User for purposes of such Terms.

"Work(s)" are the copyright protected works described in relevant Order Confirmations.

2) **Description of Service.** CCC's Marketplace enables Users to obtain Licenses to use one or more Works in accordance with all relevant Terms. CCC grants Licenses as an agent on behalf of the copyright rightsholder identified in the relevant Order Confirmation.

3) **Applicability of Terms.** The Terms govern User's use of Works in connection with the relevant License. In the event of any conflict between General Terms and Order Confirmation Terms, the latter shall govern. User acknowledges that Rightsholders

have complete discretion whether to grant any permission, and whether to place any limitations on any grant, and that CCC has no right to supersede or to modify any such discretionary act by a Rightsholder.

4) **Representations; Acceptance.** By using the Service, User represents and warrants that User has been duly authorized by the User to accept, and hereby does accept, all Terms.

5) **Scope of License; Limitations and Obligations.** All Works and all rights therein, including copyright rights, remain the sole and exclusive property of the Rightsholder. The License provides only those rights expressly set forth in the terms and conveys no other rights in any Works

6) **General Payment Terms.** User may pay at time of checkout by credit card or choose to be invoiced. If the User chooses to be invoiced, the User shall: (i) remit payments in the manner identified on specific invoices, (ii) unless otherwise specifically stated in an Order Confirmation or separate written agreement, Users shall remit payments upon receipt of the relevant invoice from CCC, either by delivery or notification of availability of the invoice via the Marketplace platform, and (iii) if the User does not pay the invoice within 30 days of receipt, the User may incur a service charge of 1.5% per month or the maximum rate allowed by applicable law, whichever is less. While User may exercise the rights in the License immediately upon receiving the Order Confirmation, the License is automatically revoked and is null and void, as if it had never been issued, if CCC does not receive complete payment on a timely basis.

7) **General Limits on Use.** Unless otherwise provided in the Order Confirmation, any grant of rights to User (i) involves only the rights set forth in the Terms and does not include subsequent or additional uses, (ii) is non-exclusive and non-transferable, and (iii) is subject to any and all limitations and restrictions (such as, but not limited to, limitations on duration of use or circulation) included in the Terms. Upon completion of the licensed use as set forth in the Order Confirmation, User shall either secure a new permission for further use of the Work(s) or immediately cease any new use of the Work(s) and shall render inaccessible (such as by deleting or by removing or severing links or other locators) any further copies of the Work. User may only make alterations to the Work if and as expressly set forth in the Order Confirmation. No Work may be used in any way that is unlawful, including without limitation if such use would violate applicable sanctions laws or regulations, would be defamatory, violate the rights of third parties (including such third parties' rights of copyright, privacy, publicity, or other tangible or intangible property), or is otherwise illegal, sexually explicit, or obscene. In addition, User may not conjoin a Work with any other material that may result in damage to the reputation of the Rightsholder. Any unlawful use will render any licenses hereunder null and void. User agrees to inform CCC if it becomes aware of any infringement of any rights in a Work and to cooperate with any reasonable request of CCC or the Rightsholder in connection therewith.

8) **Third Party Materials.** In the event that the material for which a License is sought includes third party materials (such as photographs, illustrations, graphs, inserts and similar materials) that are identified in such material as having been used by permission (or a similar indicator), User is responsible for identifying, and seeking separate licenses (under this Service, if available, or otherwise) for any of such third party materials; without a separate license, User may not use such third party materials via the License.

9) **Copyright Notice.** Use of proper copyright notice for a Work is required as a condition of any License granted under the Service. Unless otherwise provided in the Order Confirmation, a proper copyright notice will read substantially as follows: "Used with permission of [Rightsholder's name], from [Work's title, author, volume, edition number and year of copyright]; permission conveyed through Copyright Clearance Center, Inc." Such notice must be provided in a reasonably legible font size and must be placed either on a cover page or in another location that any person, upon gaining access to the material which is the subject of a permission, shall see, or in the case of republication Licenses, immediately adjacent to the Work as used (for example, as part of a by-line or footnote) or in the place where substantially all other credits or notices for the new work containing the republished Work are located. Failure to include the required notice results in loss to the Rightsholder and CCC, and the User shall be liable to pay liquidated damages for each such failure equal to twice the use fee specified in the Order Confirmation, in addition to the use fee itself and any other fees and charges specified.

10) **Indemnity.** User hereby indemnifies and agrees to defend the Rightsholder and CCC, and their respective employees and directors, against all claims, liability, damages, costs, and expenses, including legal fees and expenses, arising out of any use of a Work beyond the scope of the rights granted herein and in the Order Confirmation, or any use of a Work which has been altered in any unauthorized way by User, including claims of defamation or infringement of rights of copyright, publicity, privacy, or other tangible or intangible property.

11) **Limitation of Liability.** UNDER NO CIRCUMSTANCES WILL CCC OR THE RIGHTSHOLDER BE LIABLE FOR ANY DIRECT, INDIRECT, CONSEQUENTIAL, OR INCIDENTAL DAMAGES (INCLUDING WITHOUT LIMITATION DAMAGES FOR LOSS OF BUSINESS PROFITS OR INFORMATION, OR FOR BUSINESS INTERRUPTION) ARISING OUT OF THE USE OR INABILITY TO USE A WORK, EVEN IF ONE OR BOTH OF THEM HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. In any event, the total liability of the Rightsholder and CCC (including their respective employees and directors) shall not exceed the total amount actually paid by User for the relevant License. User assumes full liability for the actions and omissions of its principals, employees, agents, affiliates, successors, and assigns.

12) **Limited Warranties.** THE WORK(S) AND RIGHT(S) ARE PROVIDED "AS IS." CCC HAS THE RIGHT TO GRANT TO USER THE RIGHTS GRANTED IN THE ORDER CONFIRMATION DOCUMENT. CCC AND THE RIGHTSHOLDER DISCLAIM ALL OTHER WARRANTIES RELATING TO THE WORK(S) AND RIGHT(S), EITHER EXPRESS OR IMPLIED, INCLUDING WITHOUT LIMITATION IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. ADDITIONAL RIGHTS MAY BE REQUIRED TO USE ILLUSTRATIONS, GRAPHS, PHOTOGRAPHS, ABSTRACTS, INSERTS, OR OTHER PORTIONS OF THE WORK (AS OPPOSED TO THE ENTIRE WORK) IN A MANNER CONTEMPLATED BY USER; USER UNDERSTANDS AND AGREES THAT NEITHER CCC NOR THE RIGHTSHOLDER MAY HAVE SUCH ADDITIONAL RIGHTS TO GRANT.

13) **Effect of Breach.** Any failure by User to pay any amount when due, or any use by User of a Work beyond the scope of the License set forth in the Order Confirmation and/or the Terms, shall be a material breach of such License. Any breach not cured within 10 days of written notice thereof shall result in immediate termination of such License without further notice. Any unauthorized (but licensable) use of a Work that is terminated immediately upon notice thereof may be liquidated by payment of the Rightsholder's ordinary license price therefor; any unauthorized (and unlicensable) use that is not terminated immediately for any reason (including, for example, because materials containing the Work cannot reasonably be recalled) will be subject to all remedies available at law or in equity, but in no event to a payment of less than three times the Rightsholder's ordinary license price for the most closely analogous licensable use plus Rightsholder's and/or CCC's costs and expenses incurred in collecting such payment.

14) **Additional Terms for Specific Products and Services.** If a User is making one of the uses described in this Section 14, the additional terms and conditions apply:

a) *Print Uses of Academic Course Content and Materials (photocopies for academic coursepacks or classroom handouts).* For photocopies for academic coursepacks or classroom handouts the following additional terms apply:

i) The copies and anthologies created under this License may be made and assembled by faculty members individually or at their request by on-campus bookstores or copy centers, or by off-campus copy shops and other similar entities.

ii) No License granted shall in any way: (i) include any right by User to create a substantively non-identical copy of the Work or to edit or in any other way modify the Work (except by means of deleting material immediately preceding or following the entire portion of the Work copied) (ii) permit "publishing ventures" where any particular anthology would be systematically marketed at multiple institutions.

iii) Subject to any Publisher Terms (and notwithstanding any apparent contradiction in the Order Confirmation arising from data provided by User), any use authorized under the academic pay-per-use service is limited as follows:

A) any License granted shall apply to only one class (bearing a unique identifier as assigned by the institution, and thereby including all sections or other subparts of the class) at one institution;

B) use is limited to not more than 25% of the text of a book or of the items in a published collection of essays, poems or articles;

C) use is limited to no more than the greater of (a) 25% of the text of an issue of a journal or other periodical or (b) two articles from such an issue;

D) no User may sell or distribute any particular anthology, whether photocopied or electronic, at more than one institution of learning;

E) in the case of a photocopy permission, no materials may be entered into electronic memory by User except in order to produce an identical copy of a Work before or during the academic term (or analogous period) as to which any particular permission is granted. In the event that User shall choose to retain materials that are the subject of a photocopy permission in electronic memory for purposes of producing identical copies more than one day after such retention (but still within the scope of any permission granted), User must notify CCC of such fact in the applicable permission request and such retention shall constitute one copy actually sold for purposes of calculating permission fees due; and

F) any permission granted shall expire at the end of the class. No permission granted shall in any way include any right by User to create a substantively non-identical copy of the Work or to edit or in any other way modify the Work (except by means of deleting material immediately preceding or following the entire portion of the Work copied).

iv) **Books and Records; Right to Audit.** As to each permission granted under the academic pay-per-use Service, User shall maintain for at least four full calendar years books and records sufficient for CCC to determine the numbers of copies made by User under such permission. CCC and any representatives it may designate shall have the right to audit such books and records at any time during User's ordinary business hours, upon two days' prior notice. If any such audit shall determine that User shall have underpaid for, or underreported, any photocopies sold or by three percent (3%) or more, then User shall bear all the costs of any such audit; otherwise, CCC shall bear the costs of any

such audit. Any amount determined by such audit to have been underpaid by User shall immediately be paid to CCC by User, together with interest thereon at the rate of 10% per annum from the date such amount was originally due. The provisions of this paragraph shall survive the termination of this License for any reason.

b) **Digital Pay-Per-Uses of Academic Course Content and Materials (e-coursepacks, electronic reserves, learning management systems, academic institution intranets).** For uses in e-coursepacks, posts in electronic reserves, posts in learning management systems, or posts on academic institution intranets, the following additional terms apply:

i) The pay-per-uses subject to this Section 14(b) include:

A) **Posting e-reserves, course management systems, e-coursepacks for text-based content**, which grants authorizations to import requested material in electronic format, and allows electronic access to this material to members of a designated college or university class, under the direction of an instructor designated by the college or university, accessible only under appropriate electronic controls (e.g., password);

B) **Posting e-reserves, course management systems, e-coursepacks for material consisting of photographs or other still images not embedded in text**, which grants not only the authorizations described in Section 14(b)(i)(A) above, but also the following authorization: to include the requested material in course materials for use consistent with Section 14(b)(i)(A) above, including any necessary resizing, reformatting or modification of the resolution of such requested material (provided that such modification does not alter the underlying editorial content or meaning of the requested material, and provided that the resulting modified content is used solely within the scope of, and in a manner consistent with, the particular authorization described in the Order Confirmation and the Terms), but not including any other form of manipulation, alteration or editing of the requested material;

C) **Posting e-reserves, course management systems, e-coursepacks or other academic distribution for audiovisual content**, which grants not only the authorizations described in Section 14(b)(i)(A) above, but also the following authorizations: (i) to include the requested material in course materials for use consistent with Section 14(b)(i)(A) above; (ii) to display and perform the requested material to such members of such class in the physical classroom or remotely by means of streaming media or other video formats; and (iii) to “clip” or reformat the requested material for purposes of time or content management or ease of delivery, provided that such “clipping” or reformatting does not alter the underlying editorial content or meaning of the requested material and that the resulting material is used solely within the scope of, and in a manner consistent with, the particular authorization described in the Order Confirmation and the Terms. Unless expressly set forth in the relevant Order Confirmation, the License does not authorize any other form of manipulation, alteration or editing of the requested material.

ii) Unless expressly set forth in the relevant Order Confirmation, no License granted shall in any way: (i) include any right by User to create a substantively non-identical copy of the Work or to edit or in any other way modify the Work (except by means of deleting material immediately preceding or following the entire portion of the Work copied or, in the case of Works subject to Sections 14(b)(1)(B) or (C) above, as described in such Sections) (ii) permit “publishing ventures” where any particular course materials would be systematically marketed at multiple institutions.

iii) Subject to any further limitations determined in the Rightsholder Terms (and notwithstanding any apparent contradiction in the Order Confirmation arising from data provided by User), any use authorized under the electronic course content pay-per-use service is limited as follows:

A) any License granted shall apply to only one class (bearing a unique identifier as assigned by the institution, and thereby including all sections or other subparts of the class) at one institution;

B) use is limited to not more than 25% of the text of a book or of the items in a published collection of essays, poems or articles;

C) use is limited to not more than the greater of (a) 25% of the text of an issue of a journal or other periodical or (b) two articles from such an issue;

D) no User may sell or distribute any particular materials, whether photocopied or electronic, at more than one institution of learning;

E) electronic access to material which is the subject of an electronic-use permission must be limited by means of electronic password, student identification or other control permitting access solely to students and instructors in the class;

F) User must ensure (through use of an electronic cover page or other appropriate means) that any person, upon gaining electronic access to the material, which is the subject of a permission, shall see:

- o a proper copyright notice, identifying the Rightsholder in whose name CCC has granted permission,
- o a statement to the effect that such copy was made pursuant to permission,
- o a statement identifying the class to which the material applies and notifying the reader that the material has been made available electronically solely for use in the class, and
- o a statement to the effect that the material may not be further distributed to any person outside the class, whether by copying or by transmission and whether electronically or in paper form, and User must also ensure that such cover page or other means will print out in the event that the person accessing the material chooses to print out the material or any part thereof.

G) any permission granted shall expire at the end of the class and, absent some other form of authorization, User is thereupon required to delete the applicable material from any electronic storage or to block electronic access to the applicable material.

iv) Uses of separate portions of a Work, even if they are to be included in the same course material or the same university or college class, require separate permissions under the electronic course content pay-per-use Service. Unless otherwise provided in the Order Confirmation, any grant of rights to User is limited to use completed no later than the end of the academic term (or analogous period) as to which any particular permission is granted.

v) Books and Records; Right to Audit. As to each permission granted under the electronic course content Service, User shall maintain for at least four full calendar years books and records sufficient for CCC to determine the numbers of copies made by User under such permission. CCC and any representatives it may designate shall have the right to audit such books and records at any time during User's ordinary business hours, upon two days' prior notice. If any such audit shall determine that User shall have underpaid for, or underreported, any electronic copies used by three percent (3%) or more, then User shall bear all the costs of any such audit; otherwise, CCC shall bear the costs of any such audit. Any amount determined by such audit to have been underpaid by User shall immediately be paid to CCC by User, together with interest thereon at the rate of 10% per annum from the date such amount was originally due. The provisions of this paragraph shall survive the termination of this license for any reason.

c) ***Pay-Per-Use Permissions for Certain Reproductions (Academic photocopies for library reserves and interlibrary loan reporting) (Non-academic internal/external business uses and commercial document delivery)***. The License expressly excludes the uses listed in Section (c)(i)-(v) below (which must be subject to separate license from the applicable Rightsholder) for: academic photocopies for library reserves and interlibrary loan reporting; and non-academic internal/external business uses and commercial document delivery.

- i) electronic storage of any reproduction (whether in plain-text, PDF, or any other format) other than on a transitory basis;
- ii) the input of Works or reproductions thereof into any computerized database;
- iii) reproduction of an entire Work (cover-to-cover copying) except where the Work is a single article;
- iv) reproduction for resale to anyone other than a specific customer of User;
- v) republication in any different form. Please obtain authorizations for these uses through other CCC services or directly from the rightsholder.

Any license granted is further limited as set forth in any restrictions included in the Order Confirmation and/or in these Terms.

d) ***Electronic Reproductions in Online Environments (Non-Academic-email, intranet, internet and extranet)***. For "electronic reproductions", which generally includes e-mail use (including instant messaging or other electronic transmission to a defined group of recipients) or posting on an intranet, extranet or Intranet site (including any display or performance incidental thereto), the following additional terms apply:

- i) Unless otherwise set forth in the Order Confirmation, the License is limited to use completed within 30 days for any use on the Internet, 60 days for any use on an intranet or extranet and one year for any other use, all as measured from the "republication date" as identified in the Order Confirmation, if any, and otherwise from the date of the Order Confirmation.
- ii) User may not make or permit any alterations to the Work, unless expressly set forth in the Order Confirmation (after request by User and approval by Rightsholder); provided, however, that a Work consisting of photographs or other still images not embedded in text may, if necessary, be resized, reformatted or have its resolution modified

without additional express permission, and a Work consisting of audiovisual content may, if necessary, be “clipped” or reformatted for purposes of time or content management or ease of delivery (provided that any such resizing, reformatting, resolution modification or “clipping” does not alter the underlying editorial content or meaning of the Work used, and that the resulting material is used solely within the scope of, and in a manner consistent with, the particular License described in the Order Confirmation and the Terms.

15) **Miscellaneous.**

a) User acknowledges that CCC may, from time to time, make changes or additions to the Service or to the Terms, and that Rightsholder may make changes or additions to the Rightsholder Terms. Such updated Terms will replace the prior terms and conditions in the order workflow and shall be effective as to any subsequent Licenses but shall not apply to Licenses already granted and paid for under a prior set of terms.

b) Use of User-related information collected through the Service is governed by CCC’s privacy policy, available online at www.copyright.com/about/privacy-policy/.

c) The License is personal to User. Therefore, User may not assign or transfer to any other person (whether a natural person or an organization of any kind) the License or any rights granted thereunder; provided, however, that, where applicable, User may assign such License in its entirety on written notice to CCC in the event of a transfer of all or substantially all of User’s rights in any new material which includes the Work(s) licensed under this Service.

d) No amendment or waiver of any Terms is binding unless set forth in writing and signed by the appropriate parties, including, where applicable, the Rightsholder. The Rightsholder and CCC hereby object to any terms contained in any writing prepared by or on behalf of the User or its principals, employees, agents or affiliates and purporting to govern or otherwise relate to the License described in the Order Confirmation, which terms are in any way inconsistent with any Terms set forth in the Order Confirmation, and/or in CCC’s standard operating procedures, whether such writing is prepared prior to, simultaneously with or subsequent to the Order Confirmation, and whether such writing appears on a copy of the Order Confirmation or in a separate instrument.

e) The License described in the Order Confirmation shall be governed by and construed under the law of the State of New York, USA, without regard to the principles thereof of conflicts of law. Any case, controversy, suit, action, or proceeding arising out of, in connection with, or related to such License shall be brought, at CCC’s sole discretion, in any federal or state court located in the County of New York, State of New York, USA, or in any federal or state court whose geographical jurisdiction covers the location of the Rightsholder set forth in the Order Confirmation. The parties expressly submit to the personal jurisdiction and venue of each such federal or state court.

Last updated October 2022

Curriculum Vitae

Name: Mauricio Cespedes Tenorio

Post-Secondary Education and Degrees: University of Costa Rica
San Jose, Costa Rica
2017 - 2021 B.S. in Electrical and Computer Engineering

University of Western Ontario
London, ON
2022 - 2023 MESc in Biomedical Engineering

Honours and Awards: Excellence Scholarship
2017-2020

Related Work Experience: Teaching Assistant
The University of Western Ontario
2023-2024

Publications and Conferences

Conference proceedings and research abstracts

- M. Cespedes *et al.*, "Implementation and Evaluation of an Interictal Intracranial EEG Classifier", *Clinical Neurological Sciences Research Day*, Western University, London, Ontario, Oct. 2024.
- M. Cespedes *et al.*, "iEEGPrep: towards standard preprocessing for the analysis of intracranial EEG recordings", *Canadian League Against Epilepsy Annual Scientific Meeting*, Oct. 2023.

- M. Cespedes *et al.*, "iEEGPrep: towards standard preprocessing for the analysis of intracranial EEG recordings", *Clinical Neurological Sciences Research Day*, Western University, London, Ontario, 2023.
- M. Cespedes Tenorio and D. S. Dumani, "Modeling thermometry image perturbations during photoacoustic imaging-guided photothermal therapy," *2021 IEEE UFFC Latin America Ultrasonics Symposium (LAUS)*, Oct. 2021.
- M. Cespedes Tenorio and D. S. Dumani, "Multivariable fuzzy logic controlled photothermal therapy," *IFAC-PapersOnLine*, vol. 54, no. 15, pp. 400–405, 2021.