Scalp EEG Prep

v0.1.0 eeg scalp-eeg electrophysiology epilepsy neurology neuroscience preprocessing quality-control artifact-removal bids

The first step of the scalp EEG workflows.

Repository	`https://github.com/pennsieve-apps/eeg-prep`	License	Apache-2.0
Authors	Pennsieve (University of Pennsylvania)	Generated	2026-11-13 21:33 UTC
Sample run	`pipeline:b5154b53-747e-4a3b-bc79-06d88d7aec35:64bc9e9467aa`	Run time	2 min 18 s

This document shows what **Scalp EEG Prep** produces. Pennsieve ran version v0.1.0 once, on the test inputs the app declares for itself, and collected what it wrote. It is not the app's documentation and not its output: the outputs are the files listed under Results. Read it to decide whether the app does what you need before you run it on your own data.

Parts marked AI-WRITTEN were written by a language model (us.anthropic.claude-opus-5-5) from the files and records cited in this document; every number in them was checked against that material. Everything else, including every figure, table and value, is copied from the run's output files and its record. No figure was drawn or altered; images are downscaled copies.

IN THIS DOCUMENT

1 At a glance
2 About the app
3 Test inputs
4 Figures
5 Results
6 Run it yourself
7 How to cite
8 Files in this document

1 · At a glance

The Scalp EEG Prep app reads clinical scalp EEG in common formats with MNE-Python, keeps EEG channels under their 10-20 names, detects bad channels with pyprep, filters, re-references (average by default, with bipolar-only recordings kept bipolar), removes eye, muscle and heart artifacts with ICA and ICLabel, marks bad segments and reads seizure labels. This sample run used the app's own test data, written by its test-data generator. From that input the app produced a BIDS-EEG derivative with one preprocessed EDF+ file and its sidecar files for each of two recordings, sub-01_ses-01_task-routine_run-01 and sub-chbx01_task-eeg_run-01. The run also produced a separate quality-control output with a per-channel QC table, a summary file and figures. In this run, sub-01 was average-referenced and sub-chbx01 stayed bipolar.

AI-WRITTEN

RAN ON
1 input · 9 files · 3.3 MB

PRODUCED
2 outputs · 22 files · 4.9 MB

SHOWN HERE
3 figures

RUN TIME
2 min 18 s

2 · About the app

The first step of the scalp EEG workflows. Reads clinical scalp EEG (BIDS-EEG, EDF/EDF+, BDF, BrainVision, EEGLAB, FIF, Nihon Kohden, Natus) with MNE-Python, keeps the EEG channels under their 10-20 names, finds bad channels (pyprep), filters, re-references (average by default; bipolar-only recordings stay bipolar), removes eye, muscle and heart artifacts with ICA + ICLabel (optional ASR), marks bad segments, reads seizure labels, and writes one BIDS-EEG derivative (EDF+) plus a QC table and figures.

Takes

Input	What	Formats
eeg	Scalp EEG recordings: a BIDS-EEG dataset (events.tsv seizure rows, HED-SCORE `eventType`, channels.tsv, participants.tsv are read), or loose EDF/EDF+/BDF, BrainVision (.vhdr), EEGLAB (.set), FIF, Nihon Kohden (.EEG) or Natus/Nicolet (.e) files; CHB-MIT `-summary.txt` files next to the EDFs and an optional labels.csv. Folders and symlinks are followed.	`application/edf, text/csv, text/tab-separated-values, application/json, application/octet-stream`

Produces

Output	What	Formats
eeg	A BIDS-EEG derivative: dataset_description.json, participants.tsv, recordings.csv, provenance.json, and per recording <stem>_desc-preproc_{eeg.edf (EDF+, µV), eeg.json, channels.tsv, events.tsv, electrodes.tsv, coordsystem.json}.	`application/edf, text/csv, text/tab-separated-values, application/json`
qc	eeg-prep/qc.csv (channels kept/bad/dropped with reasons, ICA components with their ICLabel class, bad-segment totals), one QC figure per recording, an overview figure, summary.json and provenance.json.	`text/csv, image/png, application/json`

3 · Test inputs

What the app ran on for this sample: the test data its app.yml declares, built for this run and discarded afterwards (app commit `da0a2c452706`).

eeg 9 files · 3.3 MB

Written by the app's own test-data generator (python -m tests.make_fixture "$REVIEW_INPUT_DIR").

In the app's input directory as `eeg/`.

4 · Figures

Figures marked (test) show the app's generated test data, not real measurements.

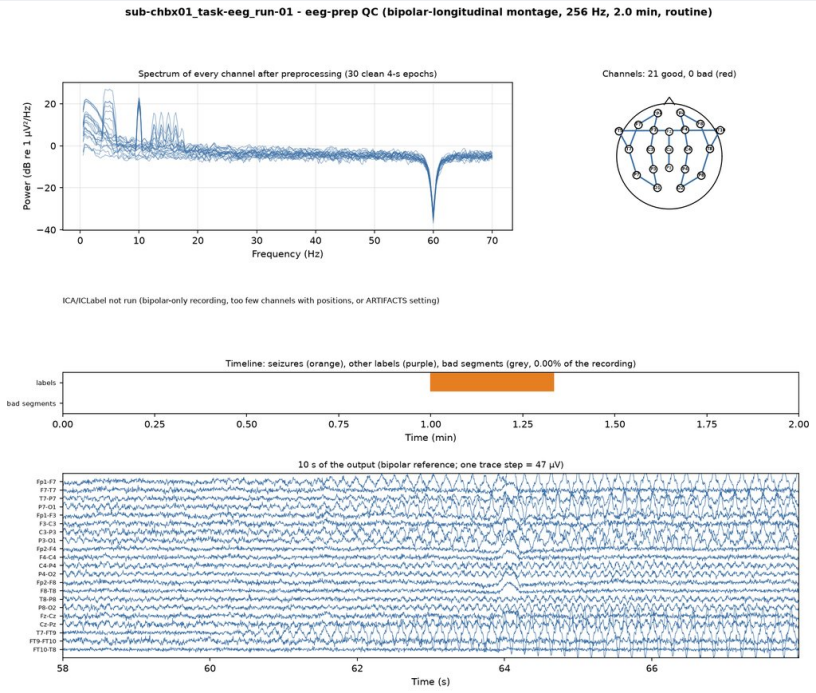


Figure 1 (test). Quality-control summary produced by the eeg-prep app (version 0.1.0) for recording sub-chbx01_task-eeg_run-01. It shows the per-channel spectrum after preprocessing, channel status on a head map, a timeline of labels and bad segments, a short excerpt of the output traces, and a note that ICA/ICLabel was not run for this recording. AI-WRITTEN

qc · qc/eeg-prep/sub-chbx01_task-eeg_run-01_qc.png R14

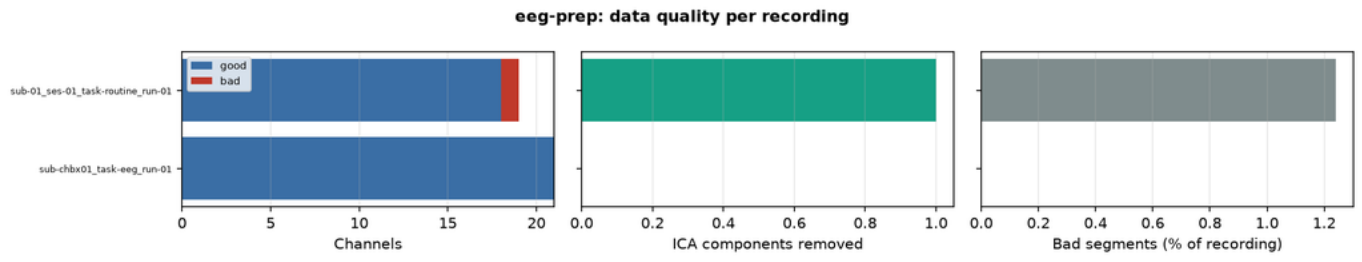


Figure 2 (test). Overview of data quality per recording from the eeg-prep app, with panels for good and bad channels, ICA components removed, and bad segments as a percentage of each recording, for sub-01_ses-01_task-routine_run-01 and sub-chbx01_task-eeg_run-01. AI-WRITTEN

qc · qc/eeg-prep/overview_qc.png R15

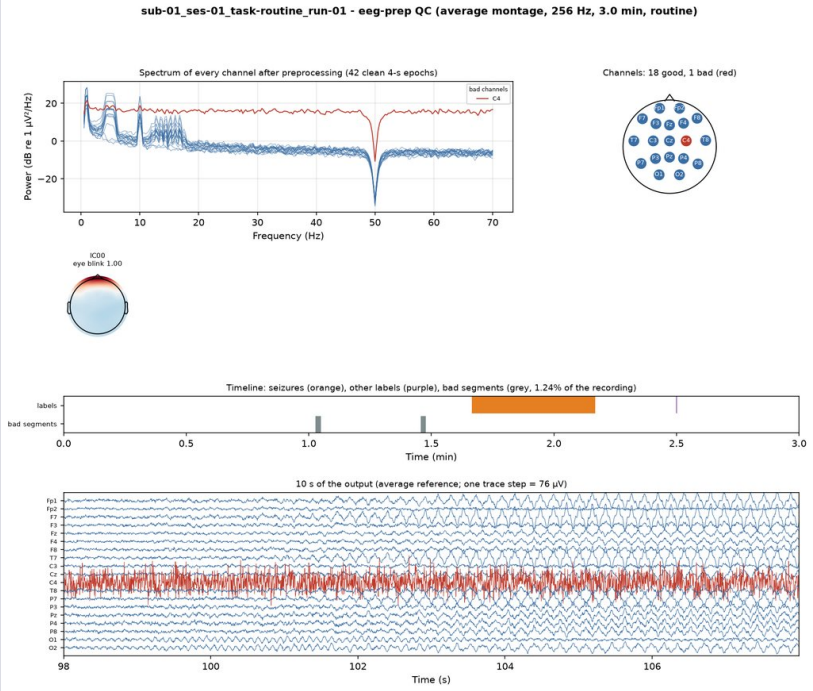


Figure 3 (test). Quality-control summary produced by the eeg-prep app (version 0.1.0, average reference, ICLabel artifact removal) for recording sub-01_ses-01_task-routine_run-01. It shows the per-channel spectrum after preprocessing with the bad channel highlighted, channel status on a head map, the removed ICA component, a timeline of labels and bad segments, and a short excerpt of the output traces. AI-WRITTEN

qc · qc/eeg-prep/sub-01_ses-01_task-routine_run-01_qc.png R16

5 · Results

What the app wrote, by output port, as delivered in `Scalp-EEG-Prep-v0.1.0-sample.zip`.

> **eeg** A BIDS-EEG derivative: dataset_description.json, participants.tsv, recordings.csv, provenance.json, and per recording <stem>_desc-preproc_{eeg.edf (EDF+, µV), eeg.json, channels.tsv, events.tsv, electrodes.tsv, coordsystem.json}.

Test EEG data of 9 files written by the app's own test-data generator (python -m tests.make_fixture).

AI-WRITTEN

Files

Total	16 files · 2.9 MB
.edf	2 · 2.9 MB
.json	6 · 10.0 KB
.tsv	7 · 5.5 KB
.csv	1 · 581 B

Recorded by the app R1

App	eeg-prep 0.1.0
Ran	2026-11-13 18:45 UTC
Peak memory	315.4 MB
Tools	python 3.12.16 · mne 1.13.2 · mne-bids 0.20.0 · pyprep 0.10.0 · mne-icalabel 0.10.0 · onnxruntime 1.31.1 · meegkit 0.2.0 · python-picard 0.8.2 · numpy 2.5.4 · scipy 1.18.1 · matplotlib 3.11.2

`eeg/sub-01/ses-01/eeg/sub-01_ses-01_task-routine_run-01_desc-preproc_coordsystem.json` R9

EEGCoordinateSystem	Other

EEGCoordinateUnits	m
EEGCoordinateSystemDescription	Template positions, not measured: MNE-Python's 10-05 montage 'standard_1005' (named 'colin27_1005' since MNE 1.13) in MNE head coordinates (x towards the right preauricular point, y towards the nasio…

eeg/sub-chbx01/eeg/sub-chbx01_task-eeg_run-01_desc-preproc_coordsystem.json R10

EEGCoordinateSystem	Other
EEGCoordinateUnits	m
EEGCoordinateSystemDescription	Template positions, not measured: MNE-Python's 10-05 montage 'standard_1005' (named 'colin27_1005' since MNE 1.13) in MNE head coordinates (x towards the right preauricular point, y towards the nasio…

eeg/dataset_description.json R11

Name	eeg-prep derivative
BIDSVersion	1.10.0
DatasetType	derivative

eeg/sub-chbx01/eeg/sub-chbx01_task-eeg_run-01_desc-preproc_eeg.json R12

TaskName	eeg
SamplingFrequency	256
EEGReference	bipolar: each channel is the first electrode minus the second (see contacts in channels.tsv)
PowerLineFrequency	60
RecordingDuration	120
RecordingType	continuous
EEGChannelCount	21
ECGChannelCount	0
EMGChannelCount	0
EOGChannelCount	0
MiscChannelCount	0
TriggerChannelCount	0
EEGPlacementScheme	10-20
SourceSoftwareFilters	n/a
Sources	eeg/chbx/chbx01_01.edf

eeg/sub-01/ses-01/eeg/sub-01_ses-01_task-routine_run-01_desc-preproc_eeg.json R13

TaskName	routine
SamplingFrequency	256
EEGReference	average of the good EEG channels (computed by eeg-prep)
PowerLineFrequency	50
RecordingDuration	180
RecordingType	continuous
EEGChannelCount	19
ECGChannelCount	0
EMGChannelCount	0
EOGChannelCount	0
MiscChannelCount	0
TriggerChannelCount	0
EEGPlacementScheme	10-20

SourceSoftwareFilters	n/a
Sources	eeg/bids/sub-01/ses-01/eeg/sub-01_ses-01_task-routine_run-01_eeg.edf

eeg/participants.tsv R4

2 of 2 rows.

participant_id	age	sex
sub-01	34	f
sub-chbx01		

eeg/sub-chbx01/eeg/sub-chbx01_task-eeg_run-01_desc-preproc_events.tsv R6 one row

onset	60
duration	20
trial_type	seizure
source	annotation
channels	n/a
hed	n/a

eeg/sub-01/ses-01/eeg/sub-01_ses-01_task-routine_run-01_desc-preproc_events.tsv R7

4 of 4 rows.

onset	duration	trial_type	source	channels	hed
61.711	1.176	bad_segment	eeg-prep	n/a	n/a
87.5	1.059	bad_segment	eeg-prep	n/a	n/a
100	30	seizure	annotation	T7,P7	sz_foc_ia
150	n/a	lights_off	labels.csv	n/a	n/a

1 more table(s) were read but are not shown: eeg/recordings.csv R8.

> **qc** eeg-prep/qc.csv (channels kept/bad/dropped with reasons, ICA components with their ICLabel class, bad-segment totals), one QC figure per recording, an overview figure, summary.json and provenance.json.

qc.csv lists per-recording channel status with original names and reasons: for sub-01, EKG and Photic channels were dropped by type, and C4 was marked bad for deviation (robust amplitude z > 5) and high-frequency noise. The output also holds per-recording QC figures, an overview figure of channels, removed ICA components and bad segments, and a provenance file recording parameters such as HIGHPASS 0.5, REFERENCE average, ARTIFACTS iclabel and ICLABEL_THRESHOLD 0.8. AI-WRITTEN

Files

Total	6 files · 1.9 MB
.png	3 · 1.9 MB
.json	2 · 8.3 KB
.csv	1 · 5.7 KB

Recorded by the app R2

App	eeg-prep 0.1.0
Ran	2026-11-13 18:45 UTC
Peak memory	315.4 MB
Tools	python 3.12.16 · mne 1.13.2 · mne-bids 0.20.0 · pyprep 0.10.0 · mne-icalabel 0.10.0 · onnxruntime 1.31.1 · meegkit 0.2.0 · python-picard 0.8.2 · numpy 2.5.4 · scipy 1.18.1 · matplotlib 3.11.2

qc/eeg-prep/summary.json R3

app	eeg-prep

version	0.1.0

`qc/eeg-prep/qc.csv` R5

10 of 65 rows, 8 of 11 columns.

recording_id	item	name	original_name	type	status	reason	iclabel_label
sub-01_ses-01_task-routine_run-01	dropped	EKG	EKG	ecg	dropped	type ECG	
sub-01_ses-01_task-routine_run-01	dropped	Photic	Photic	trigger	dropped	type TRIG	
sub-01_ses-01_task-routine_run-01	channel	Fp1	EEG Fp1-REF	EEG	good		
sub-01_ses-01_task-routine_run-01	channel	Fp2	EEG Fp2-REF	EEG	good		
sub-01_ses-01_task-routine_run-01	channel	F7	EEG F7-REF	EEG	good		
sub-01_ses-01_task-routine_run-01	channel	F3	EEG F3-REF	EEG	good		
sub-01_ses-01_task-routine_run-01	channel	Fz	EEG Fz-REF	EEG	good		
sub-01_ses-01_task-routine_run-01	channel	F4	EEG F4-REF	EEG	good		
sub-01_ses-01_task-routine_run-01	channel	F8	EEG F8-REF	EEG	good		
sub-01_ses-01_task-routine_run-01	channel	T7	EEG T3-REF	EEG	good		

6 · Run it yourself

On Pennsieve

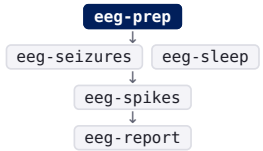
1. Open **Scalp EEG Prep** in the Pennsieve App Store and choose **Try this app**, or add it as a step in a workflow.
2. Give it **eeg** (application/edf, text/csv, text/tab-separated-values) from one of your datasets. The sample's generated inputs are small synthetic files the app's repository can regenerate (`tests/` in the repository).
3. Pick a compute node and start the run.
4. Collect the results from the run's page: as a download, or into a dataset folder when the app is a workflow step.

Suggested workflows

Public workflows on Pennsieve that run Scalp EEG Prep as a step; open one and run it on your data.

Scalp EEG Seizure Detection & Monitoring OFFICIAL

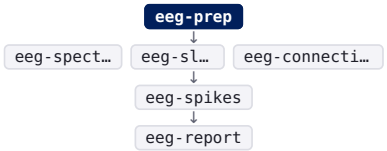
Scalp EEG: seizure candidates in SzCORE format, scored; spikes by sleep stage;



5 steps · **Open on Pennsieve ↗**

Scalp EEG Interictal Biomarkers OFFICIAL

Scalp EEG: spectral biomarkers, PDR, spikes by sleep stage, sleep, connectivity,



6 steps · **Open on Pennsieve ↗**

Locally, as a container

Every Pennsieve app is a container that reads `INPUT_DIR`, writes `OUTPUT_DIR` and takes its parameters from the environment, so it runs anywhere Docker does. Build it from the release's source and run it on a folder laid out as in section 3 (one subfolder per input port):

```
git clone https://github.com/pennsieve-apps/eeg-prep eeg-prep
cd eeg-prep && git checkout v0.1.0
docker build -t eeg-prep:v0.1.0 .

mkdir -p input output        # your inputs under ./input
docker run --rm \
  -v "$PWD/input:/input:ro" -v "$PWD/output:/output" \
  -e INPUT_DIR=/input -e OUTPUT_DIR=/output \
  eeg-prep:v0.1.0
```

The app declares 4 GB of memory; give Docker at least that. The results land in `./output`, the same files as section 5.

7 · How to cite

Results made with this app rest on three things. Please cite all three.

1. **THE METHOD**
   The app's citation file names no single work it implements; cite the works it builds on, below.
2. **THIS RELEASE**
   Pennsieve (University of Pennsylvania) (2026). eeg-prep (version v0.1.0). Pennsieve Discover. doi:10.26275/c0uo-bidi
3. **PENNSIEVE**
   Name the platform as a resource, with its Research Resource Identifier RRID:SCR_021677.

FOR YOUR METHODS SECTION

Analyses were run with Scalp EEG Prep v0.1.0 (doi:10.26275/c0uo-bidi) on Pennsieve (RRID:SCR_021677).

Also builds on

1. Alexandre Gramfort, Martin Luessi, Eric Larson, Denis A. Engemann, Daniel Strohmeier, Christian Brodbeck, et al. (2013). MEG and EEG data analysis with MNE-Python. Frontiers in Neuroscience. doi:10.3389/fnins.2013.00267
2. Stefan Appelhoff, Matthew Sanderson, Teon Brooks, Marijn van Vliet, Romain Quentin, Chris Holdgraf, et al. (2019). MNE-BIDS: Organizing electrophysiological data into the BIDS format and facilitating their analysis. Journal of Open Source Software. doi:10.21105/joss.01896
3. Cyril R. Pernet, Stefan Appelhoff, Krzysztof J. Gorgolewski, Guillaume Flandin, Christophe Phillips, Arnaud Delorme, et al. (2019). EEG-BIDS, an extension to the brain imaging data structure for electroencephalography. Scientific Data. doi:10.1038/s41597-019-0104-8
4. Nima Bigdely-Shamlo, Tim Mullen, Christian Kothe, Kyung-Min Su, Kay A. Robbins (2015). The PREP pipeline: standardized preprocessing for large-scale EEG analysis. Frontiers in Neuroinformatics. doi:10.3389/fninf.2015.00016
5. Stefan Appelhoff, Austin J. Hurst, Aamna Lawrence, Adam Li, Yorguin José Mantilla Ramos, Christian O'Reilly, et al. (2026). PyPREP: A Python implementation of the preprocessing pipeline (PREP) for EEG data. doi:10.5281/zenodo.6363575
6. Adam Li, Jacob Feitelberg, Anand Prakash Saini, Richard Höchenberger, Mathieu Scheltienne (2022). MNE-ICALabel: Automatically annotating ICA components with ICLabel in Python. Journal of Open Source Software. doi:10.21105/joss.04484
7. Luca Pion-Tonachini, Ken Kreutz-Delgado, Scott Makeig (2019). ICLabel: An automated electroencephalographic independent component classifier, dataset, and website. NeuroImage. doi:10.1016/j.neuroimage.2019.05.026
8. Te-Won Lee, Mark Girolami, Terrence J. Sejnowski (1999). Independent Component Analysis Using an Extended Infomax Algorithm for Mixed Subgaussian and Supergaussian Sources. Neural Computation. doi:10.1162/089976699300016719
9. Pierre Ablin, Jean-Francois Cardoso, Alexandre Gramfort (2018). Faster Independent Component Analysis by Preconditioning With Hessian Approximations. IEEE Transactions on Signal Processing. doi:10.1109/tsp.2018.2844203

10. Tim R. Mullen, Christian A. E. Kothe, Yu Mike Chi, Alejandro Ojeda, Trevor Kerth, Scott Makeig, et al. (2015). Real-time neuroimaging and cognitive monitoring using wearable dry EEG. IEEE Transactions on Biomedical Engineering. doi:10.1109/tbme.2015.2481482
11. Chi-Yuan Chang, Sheng-Hsiou Hsu, Luca Pion-Tonachini, Tzyy-Ping Jung (2020). Evaluation of Artifact Subspace Reconstruction for Automatic Artifact Components Removal in Multi-Channel EEG Recordings. IEEE Transactions on Biomedical Engineering. doi:10.1109/tbme.2019.2930186
12. Nicolas Barascud, Pedro Alcocer, Paulroujansky, Juan Jesús Torre Tresols, Matthijs Pals, Romain Quentin, et al. (2021). nbara/python-meegkit: v0.1.1. doi:10.5281/zenodo.5643659
13. Jonathan Dan, Una Pale, Alireza Amirshahi, William Cappelletti, Thorir Mar Ingolfsson, Xiaying Wang, et al. (2024). SzCORE : Seizure Community Open-Source Research Evaluation framework for the validation of electroencephalography -based automated seizure detection algorithms. Epilepsia. doi:10.1111/epi.18113
14. Paolo Detti (2020). Siena Scalp EEG Database. doi:10.13026/5d4a-j060
15. Paolo Detti, Giampaolo Vatti, Garazi Zabalo Manrique de Lara (2020). EEG Synchronization Analysis for Seizure Prediction: A Study on Data of Noninvasive Recordings. Processes. doi:10.3390/pr8070846
16. Jonathan Dan, Paolo Detti (2024). BIDS Siena Scalp EEG Database. doi:10.5281/zenodo.10640762
17. Ali Shoeb (2010). CHB-MIT Scalp EEG Database. doi:10.13026/c2k01r
18. Ary L. Goldberger, Luis A. N. Amaral, Leon Glass, Jeffrey M. Hausdorff, Plamen Ch. Ivanov, Roger G. Mark, et al. (2000). PhysioBank, PhysioToolkit, and PhysioNet. Circulation. doi:10.1161/01.cir.101.23.e215
19. Dorottya Cserpan, Ece Boran, Richard Rosch, San Pietro Lo Biundo, Georgia Ramantani, Johannes Sarnthein (2021). Dataset of EEG recordings of pediatric patients with epilepsy based on the 10-20 system. doi:10.18112/openneuro.ds003555.v1.0.1
20. Dorottya Cserpan, Ece Boran, Santo Pietro Lo Biundo, Richard Rosch, Johannes Sarnthein, Georgia Ramantani (2021). Scalp high-frequency oscillation rates are higher in younger children. Brain Communications. doi:10.1093/braincomms/fcab052
21. Ali Hossam Shoeb (2009). Application of machine learning to epileptic seizure onset detection and treatment. https://dspace.mit.edu/handle/1721.1/54669
22. The Siesta Group. edfio. https://github.com/the-siesta-group/edfio

From the app's citation file (CITATION.cff) and the release's record on Pennsieve Discover.

8 · Files in this document

Every figure, table and value above is copied from one of these files in `Scalp-EEG-Prep-v0.1.0-sample.zip`, the sample run's download.

Ref	File	Size
R1	`eeg/provenance.json`	5.7 KB
R2	`qc/eeg-prep/provenance.json`	5.7 KB
R3	`qc/eeg-prep/summary.json`	2.6 KB
R4	`eeg/participants.tsv`	48 B
R5	`qc/eeg-prep/qc.csv`	5.7 KB
R6	`eeg/sub-chbx01/eeg/sub-chbx01_task-eeg_run-01_desc-preproc_events.tsv`	79 B
R7	`eeg/sub-01/ses-01/eeg/sub-01_ses-01_task-routine_run-01_desc-preproc_events.tsv`	208 B
R8	`eeg/recordings.csv`	581 B
R9	`eeg/sub-01/ses-01/eeg/sub-01_ses-01_task-routine_run-01_desc-preproc_coordsystem.json`	313 B
R10	`eeg/sub-chbx01/eeg/sub-chbx01_task-eeg_run-01_desc-preproc_coordsystem.json`	313 B
R11	`eeg/dataset_description.json`	514 B
R12	`eeg/sub-chbx01/eeg/sub-chbx01_task-eeg_run-01_desc-preproc_eeg.json`	1.5 KB
R13	`eeg/sub-01/ses-01/eeg/sub-01_ses-01_task-routine_run-01_desc-preproc_eeg.json`	1.7 KB
R14	`qc/eeg-prep/sub-chbx01_task-eeg_run-01_qc.png`	1.0 MB
R15	`qc/eeg-prep/overview_qc.png`	38.8 KB
R16	`qc/eeg-prep/sub-01_ses-01_task-routine_run-01_qc.png`	877.3 KB

About this document

- Generated 2026-11-13 21:33 UTC by Pennsieve's curation-service for Scalp EEG Prep v0.1.0, sample run pipeline:b5154b53-747e-4a3b-bc79-06d88d7aec35:64bc9e9467aa, app commit da0a2c452706.
- Material: the files the app wrote in the sample run (13 read, up to 256.0 KB each, headers only for large data files), the fixtures step's record of the test inputs, the app's app.yml and CITATION.cff, and the run's record.
- Model: us.anthropic.claude-opus-5-5. It wrote the captions and prose marked `AI-WRITTEN`; sentences with a number not found in the material were removed.
- A release report is generated for each release of the app; this one describes v0.1.0. It is a derived document, not part of the app's published release.