



SOFTWARE ENGINEERING PROJECT

**Prestimulus EEG-Based Behavioral
Prediction in the HBN-EEG Dataset**

BY

**Nantawat Suksirisunt
Naytitorn Chaovirachot**

**DEPARTMENT OF COMPUTER ENGINEERING
FACULTY OF ENGINEERING
KASETSART UNIVERSITY**

Academic Year 2565

Prestimulus EEG-Based Behavioral Prediction in the HBN-EEG Dataset

BY

**Nantawat Suksirisunt
Naytitorn Chaovirachot**

**This Project Submitted in Partial Fulfillment of the
Requirement for Bachelor Degree of Engineering
(Software Engineering)**

**Department of Computer Engineering, Faculty of
Engineering KASERTSART UNIVERSITY
Academic Year 2565**

Approved By:

Advisor **Date**
(Asst. Prof.Dr. Thanawin Rakthanmanon)

Co-Advisor **Date**
(Assoc. Prof. Dr. Theerawit Wilaiprasitporn)

Head of Department **Date**
(Assoc. Prof.Dr. Punpiti Piamsa-Nga)

Abstract

Electroencephalography (EEG) research commonly relies on interactive notebook environments for data inspection and experimentation. While flexible, these workflows introduce hidden execution states, environment dependency, and limited reproducibility, making collaboration and systematic experimentation difficult. Furthermore, large-scale EEG datasets require significant computational resources that are not always available on user machines.

This project develops a baseline pipeline for prestimulus EEG-based behavioral prediction on the Healthy Brain Network EEG (HBN-EEG) dataset. We focus on the Contrast Change Detection (CCD) task and use the 2-second inter-trial interval (ITI) segment as a constrained prestimulus window for modeling.

The proposed baseline extracts steady-state and spectral features from CCD-ITI epochs after standard preprocessing (e.g., bandpass filtering, artifact mitigation, and normalization). We train lightweight models to predict reaction time (RT) and trial-level accuracy as regression targets, and evaluate performance using regression metrics (e.g., MAE/MSE).

To support reproducible experimentation, the system includes a lightweight web interface for configuring preprocessing and running analyses without relying on a full notebook environment or heavy local machine learning installations.

Acknowledgement

Put your acknowledgement paragraph here.

Nantawat Suksirisunt
Naytitorn Chaovirachot

Table of Contents

Content	Page
Abstract	i
Acknowledgement	ii
List of Tables	v
List of Figures	vi
Chapter 1 Introduction	1
1.1 Background	1
1.2 Problem Statement	1
1.3 Project Aim	1
1.4 Terminology	2
Chapter 2 Literature Review and Related Work	3
2.1 Challenge and Dataset Context	3
2.2 HBN-EEG Resources	3
2.3 Positioning of This Work	4
Chapter 3 Objectives and Method Overview	5
3.1 Research Objectives	5
3.2 Dataset and Tasks	5
3.3 Preprocessing and Features	5
Chapter 4 Software Architecture Design	6
4.1 Domain Model	6
4.2 Training and Evaluation Flow	6

4.3 Method Notes	7
Chapter 5 AI Component Design	8
5.1 Business Context and AI Integration	8
5.2 Goal Hierarchy	9
5.3 Task Requirements Analysis Using AI Canvas	10
5.3.1 AI Task Requirements	10
5.3.2 AI Canvas Summary	10
5.3.3 Innovation	11
5.4 User Experience Design with AI	12
5.4.1 Plot Mode	12
5.4.2 Grid Plot Mode	13
5.4.3 Data Mode	13
5.4.4 AI Interaction Mode	13
5.5 Deployment Strategy	24
5.5.1 Deployment Plan	24
5.5.2 Proof of Concept	25
5.6 Reflection and Future Development	25
Chapter 6 Deliverables and Evaluation	27
6.1 Artifacts	27
6.2 Evaluation Protocol	27
Chapter 7 Conclusion and Discussion	28
7.1 Reflection	28
7.2 Challenges	28
7.3 Future Work	28
Appendix A: Example	33
Appendix B: About L^AT_EX	35

List of Tables

Page

List of Figures

	Page
4.1 System architecture overview	7
5.1 Self-supervised learning for semantic feature extraction from EEG signals.	8
5.2 React Notebook Interface – Model visualization and execution	14
5.3 Epoch visualization	15
5.4 Evoked response	15
5.5 Evoked grid view	16
5.6 Evoked joint plot	16
5.7 Evoked topomap	17
5.8 Evoked per condition	18
5.9 Frequency spectrum	19
5.10 Power spectral density grid	19
5.11 Signal-to-noise ratio spectrum	20
5.12 Signal-to-noise ratio grid	20
5.13 Sensor layout	21
5.14 Time-domain signal	22
5.15 Training Interface – Dataset selection, SSL configuration, and training log	22
5.16 Prediction Interface – Uploading input data and viewing output class with confidence	23
5.17 Evaluation Interface – Upload trained model and view performance metrics and curve	23
5.18 Comparison Interface – Visual comparison between SSL and Supervised models	24

Chapter 1

Introduction

1.1 Background

This project aligns with the EEG Foundation Challenge (EEG2025), which calls for models that generalize across tasks and subjects using large-scale, high-density EEG. The Healthy Brain Network EEG (HBN-EEG) dataset provides 128-channel recordings across six tasks, with FAIR, BIDS, and HED annotations that enable reproducible analyses. The competition emphasizes zero-shot transfer to new tasks and subjects, and prediction of latent psychopathology factors.¹

1.2 Problem Statement

Under the challenge's constraints, the test phase exposes only limited CCD segments (e.g., ITI), making behavior prediction difficult without full event-locked trials. We aim to establish a constrained baseline using the 2-second CCD ITI to predict reaction time (RT) and accuracy.

1.3 Project Aim

Our goal is to build a principled baseline framework that:

- Verifies SSVEP and P300 presence with a reproducible preprocessing pipeline (SNR spectra and evoked responses),
- Predicts CCD RT and accuracy using features derived strictly from ITI segments,

¹See [1, 2, 3].

- Establishes a clear, reproducible baseline under the competition constraints.

1.4 Terminology

- **SSVEP:** Steady-State Visual Evoked Potential, frequency-locked EEG response to periodic stimulation.
- **P300:** Positive ERP component around 300 ms after salient events, associated with attention and context updating.
- **ITI:** Inter-Trial Interval, baseline period between trials; in CCD, flickering gratings persist during ITI.
- **BIDS/HED:** Data and event annotation standards supporting consistent, analysis-ready datasets.

Chapter 2

Literature Review and Related Work

2.1 Challenge and Dataset Context

The EEG Foundation Challenge proposes two tracks: (1) zero-shot decoding across new tasks and subjects, and (2) prediction of psychopathology factors from EEG. It leverages an unprecedented high-density, multi-task HBN-EEG dataset formatted in BIDS with HED annotations.¹

2.2 HBN-EEG Resources

- **HBN-EEG FAIR Implementation (2024):** Presents analysis-ready EEG with integrated behavioral events and HED annotations, enabling reproducible analyses.²
- **Transdiagnostic Resource (2017):** Describes the HBN biobank's multimodal, large-scale and community-sampled design to support dimensional (transdiagnostic) research.³
- **EEG + Eye Tracking Dataset (2017):** Provides active and passive paradigms, including steady-state and contrast decision tasks, supporting developmental brain investigations.⁴

¹[1, 2].

²[2].

³[3].

⁴[4].

2.3 Positioning of This Work

Our study focuses on a constrained but practical scenario: prestimulus behavior prediction (RT, accuracy) from CCD ITI-only segments in HBN-EEG. This targets the challenge setting of limited test data while keeping the methodology simple and reproducible using FAIR/BIDS/HED resources.

Chapter 3

Objectives and Method Overview

3.1 Research Objectives

- **Baseline (CCD-ITI only):** Predict reaction time (RT) and accuracy as regression targets using features derived strictly from the 2-second CCD ITI.
- **Verification:** Confirm presence of SSVEP and P300 components via SNR spectra and evoked response plots.

3.2 Dataset and Tasks

HBN-EEG provides six tasks: Resting State, Surround Suppression (SuS), Movie Watching, Sequence Learning, Contrast Change Detection (CCD), and Symbol Search. CCD includes periodic stimulation during ITI, yielding steady-state signals suitable for baseline modeling under limited test segments.¹

3.3 Preprocessing and Features

- **Preprocessing:** Bandpass filtering, epoching (CCD ITI: –2 to 0 s), artifact mitigation, and normalization.
- **Verification:** Compute SNR spectra at stimulation frequencies; plot evoked responses to confirm SSVEP/P300.
- **Baseline Features:** ITI-derived steady-state features (spectral power, harmonics, topographies) for RT/accuracy prediction.

¹[2, 4].

Chapter 4

Software Architecture Design

4.1 Domain Model

The architecture reflects a research workflow tailored to the EEG2025 challenge: baseline prediction using CCD-ITI segments and cross-task transfer from SuS. Components:

- **DataLoader:** Ingests BIDS/HED-formatted HBN-EEG, parses events and metadata for SuS and CCD.¹
- **PreprocessingPipeline:** Bandpass filters, epochs (SuS, CCD-ITI), artifact mitigation, normalization.
- **BaselineModel:** Maps ITI-derived steady-state features to RT and accuracy.
- **TransferTrainer:** Pretrains on SuS steady-state features; fine-tunes on CCD-ITI for behavior prediction.
- **Evaluator:** Computes metrics (MAE for RT, accuracy/F1), produces SNR/ERP verification plots.
- **Visualizer:** Summarizes metrics and renders spectra/evoked responses.

4.2 Training and Evaluation Flow

Scenario: Baseline and Transfer Workflow

1. Load SuS and CCD-ITI segments; preprocess and verify SSVEP/P300.

¹[2].

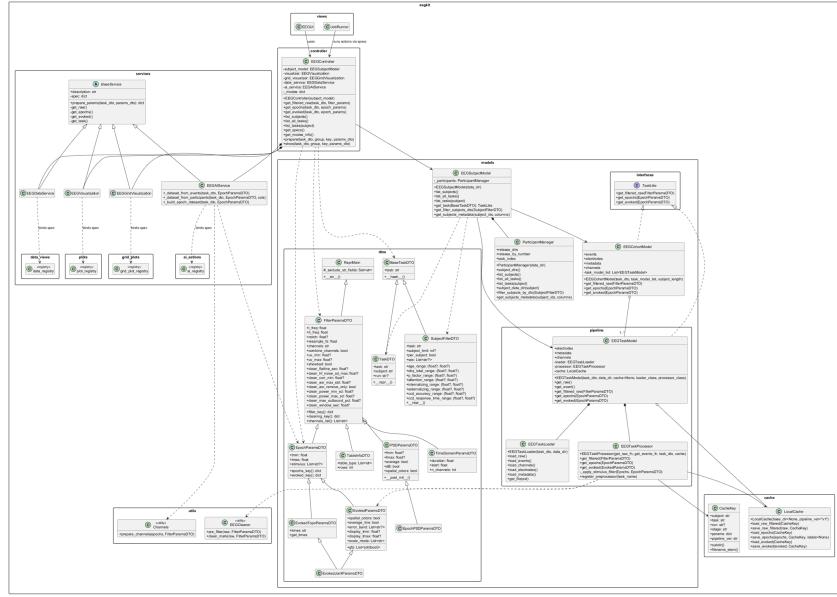


Figure 4.1: System architecture overview

2. Train baseline on CCD-ITI to predict RT/accuracy.
3. Pretrain on SuS steady-state; fine-tune on CCD-ITI for transfer.
4. Evaluate metrics and visualize spectra/evoked responses.

4.3 Method Notes

- **Features:** Spectral power at stimulation frequencies and harmonics; topographic distributions.
- **Metrics:** RT regression (MAE/MSE); accuracy classification (Accuracy/F1).
- **Plots:** SNR spectra, evoked response overlays for SuS and CCD.

Chapter 5

AI Component Design

5.1 Business Context and AI Integration

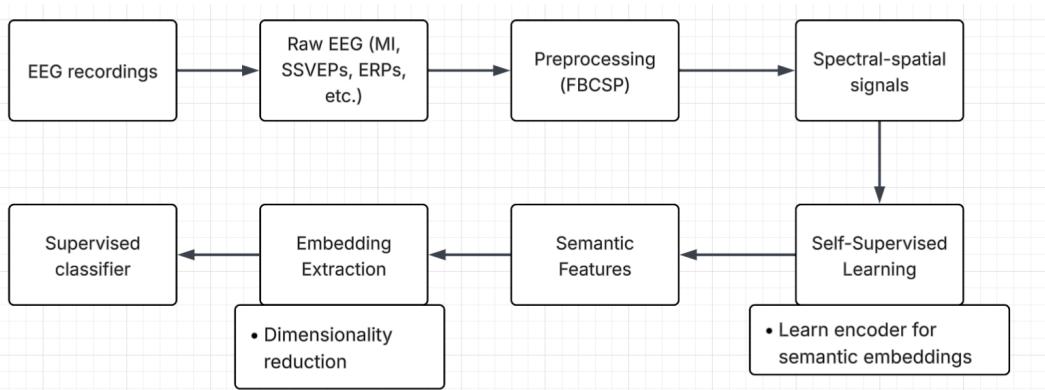


Figure 5.1: Self-supervised learning for semantic feature extraction from EEG signals.

Diagram Explanation: The system is designed as an experimental EEG research platform integrating signal visualization, dataset inspection, and machine learning experimentation into a unified workflow. The frontend interface allows users to explore EEG signals, construct datasets, and execute AI models, while the backend handles preprocessing, model execution, and evaluation.

The workflow proceeds as:

signal inspection → dataset verification → model training → evaluation → comparison

This ensures that machine learning results are directly connected to verified signal characteristics rather than blindly training on raw data.

Why is AI suitable for this problem?

EEG analysis involves high-dimensional temporal signals with complex spatial relationships across electrodes. Manual feature engineering is difficult because:

- patterns vary across subjects
- noise and artifacts are common
- signal distributions change across recording sessions

Machine learning models can learn latent representations that capture discriminative neural patterns more effectively than handcrafted features.

Is the problem large, complex, or always changing?

Yes.

EEG data presents three major challenges:

- Large — continuous multi-channel time series
- Complex — spatial-temporal correlations
- Non-stationary — signals differ between subjects and sessions

Because of this variability, deterministic rule-based systems are unsuitable.

Can we accept an answer that's not 100% perfect?

Yes. This is a research system.

The system is intended for research experimentation rather than safety-critical decision making. Model predictions are used as analytical indicators, not medical diagnoses. Therefore probabilistic predictions with measurable accuracy are acceptable.

5.2 Goal Hierarchy

Organizational Goal: Enable efficient experimentation in EEG machine learning research

System Goal: Provide an interactive platform that links signal inspection with machine learning experimentation

User Goal: Explore EEG signals, construct datasets, train models, and compare results without manual scripting

AI Model Goal: Learn representations that generalize across subjects and improve classification performance

Success Metrics:

- classification accuracy improvement over baseline
- consistent performance on unseen subjects
- reduced experiment setup time
- reproducible experiment workflow

5.3 Task Requirements Analysis Using AI Canvas

5.3.1 AI Task Requirements

- **Requirements (REQ):** Learn discriminative EEG representations for motor-imagery classification using standardized datasets.
- **Specifications (SPEC):** Evaluate whether learned representations improve classification accuracy and generalization compared to baseline approaches.
- **Environment (ENV):** Evaluated on multi-subject EEG datasets with different recording conditions. (e.g., noise, subject variation).

5.3.2 AI Canvas Summary

- **Input:** Preprocessed EEG segments
- **Output:** Predicted motor-imagery class labels
- **Success Criteria:** Model performs better than random guessing and generalizes to unseen subjects. (Ideally >80% accuracy)

5.3.3 Innovation

The system integrates visualization-driven dataset validation with model training inside a single interactive interface, reducing mismatch between inspected signals and training data.

5.4 User Experience Design with AI

The platform follows a research-oriented workflow where users first inspect signals before executing machine learning models. To support this process, the interface uses a **mode-action** structure.

Interface Overview: The interface contains four primary operational modes:

- **Plot Mode:** — detailed single-view signal inspection
- **Grid Plot Mode:** — comparative multi-condition visualization
- **Data Mode:** — structured data inspection and preparation
- **AI Mode:** — machine learning experimentation

Each mode exposes only relevant actions.

Visualization and Inspection Modes

5.4.1 Plot Mode

Provides detailed EEG inspection including sensor layout, time-domain plots, frequency plots, epoch plots, evoked responses, and SNR analysis.

Available actions include:

- sensor layout visualization
- time-domain signal plots
- frequency-domain plots
- epoch visualization
- evoked response plots
- evoked topography plots
- SNR spectrum analysis

5.4.2 Grid Plot Mode

Allows comparison across conditions using PSD, SNR, and evoked grids.

Available actions include:

- PSD Grid
- SNR Grid
- Evoked Grid

5.4.3 Data Mode

Displays structured EEG data tables to confirm dataset composition.

Available actions include:

- EEG sample tables
- epoch tables
- metadata

5.4.4 AI Interaction Mode

AI Mode enables machine learning experimentation using the prepared EEG data.

Available actions include:

- Training
- Prediction
- Evaluation
- Model Comparison

System Behavior: Upon user interaction, the backend automatically executes preprocessing, training, evaluation, and result generation. Logs, graphs, and final metrics are dynamically updated and available for export.

Feedback Loop: inspect → adjust → train → evaluate → compare → refine

Researchers can rapidly modify dataset selections and rerun experiments without rewriting scripts or managing computing environments. This significantly shortens the experimental cycle compared to notebook-based pipelines.

Interface Screenshots:

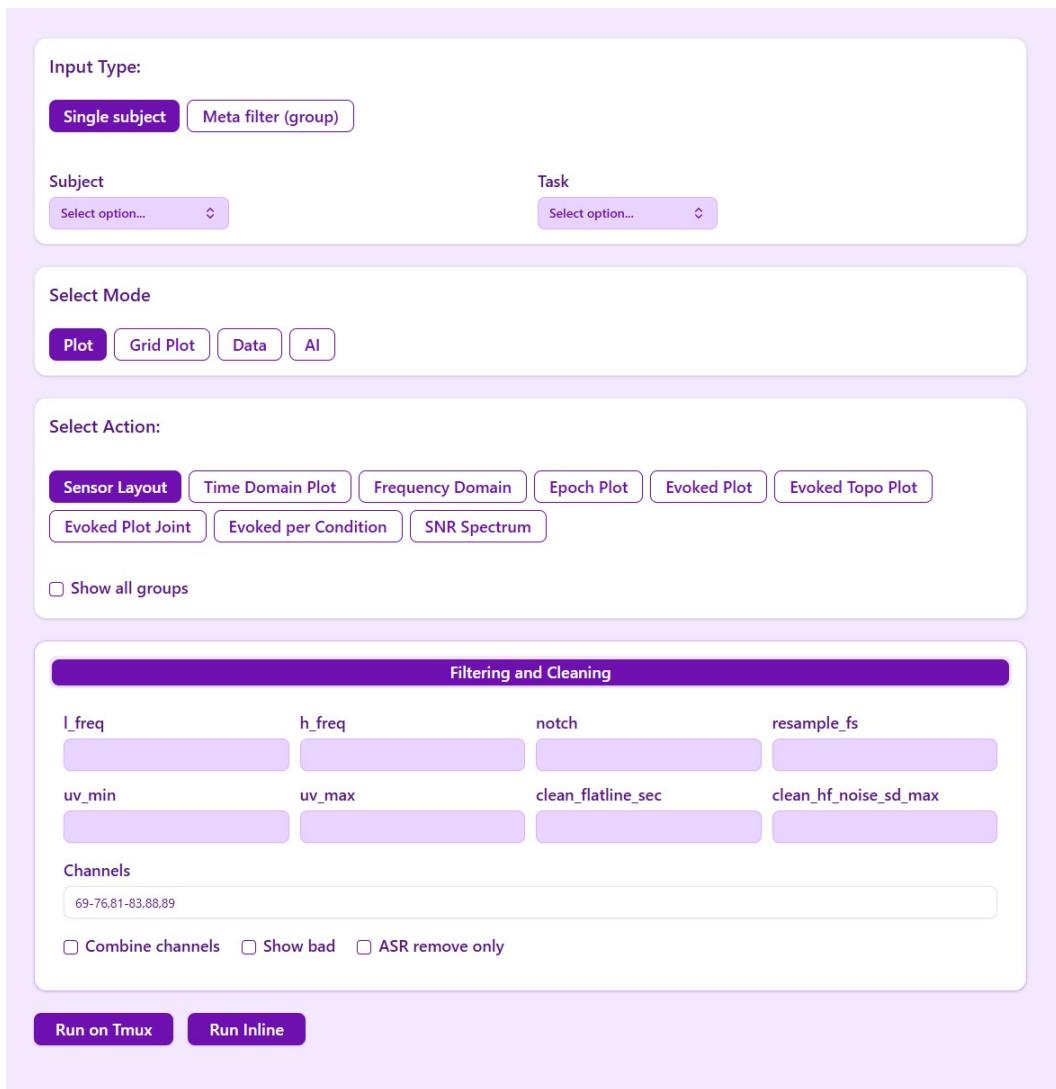


Figure 5.2: React Notebook Interface – Model visualization and execution

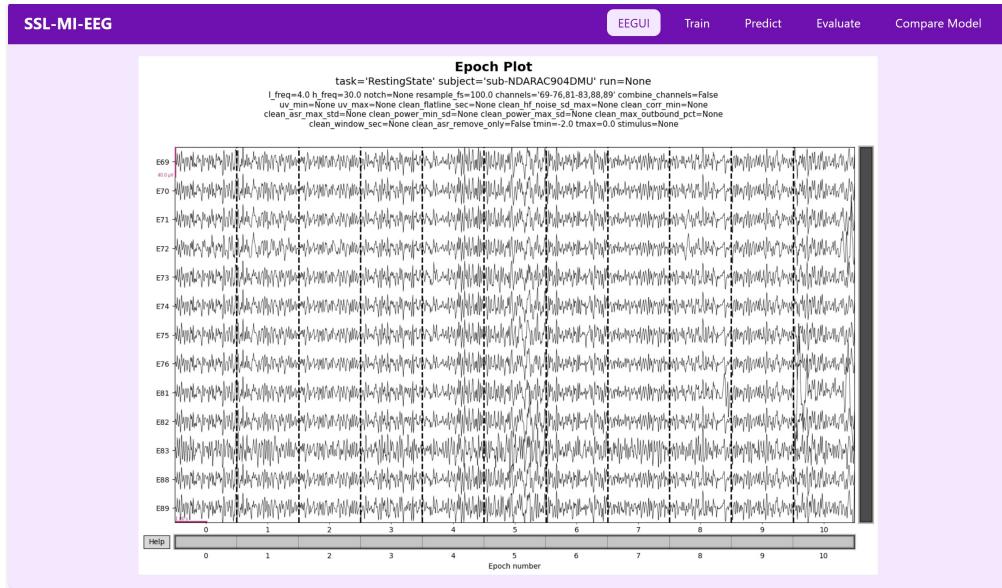


Figure 5.3: Epoch visualization

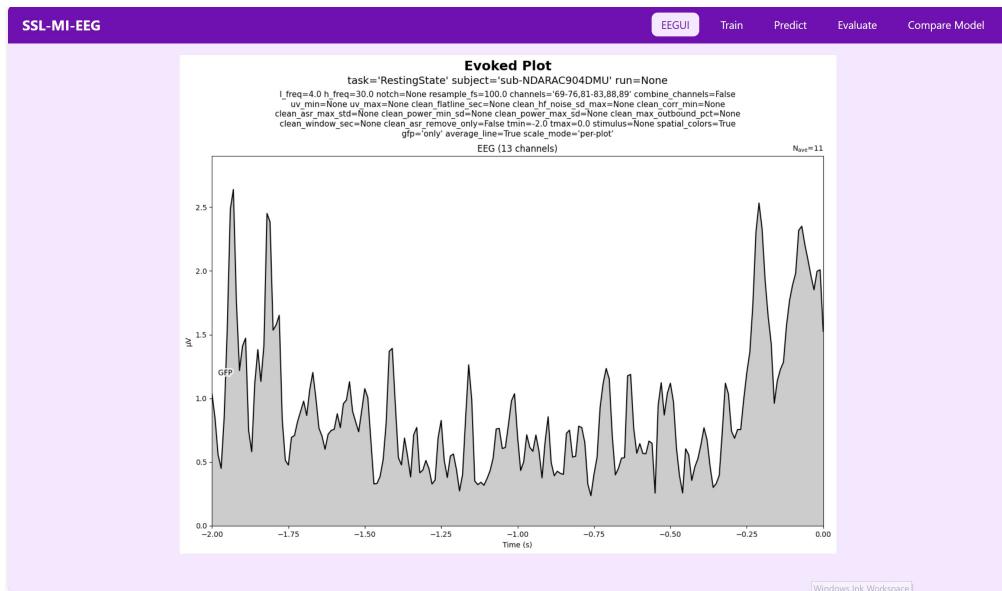


Figure 5.4: Evoked response

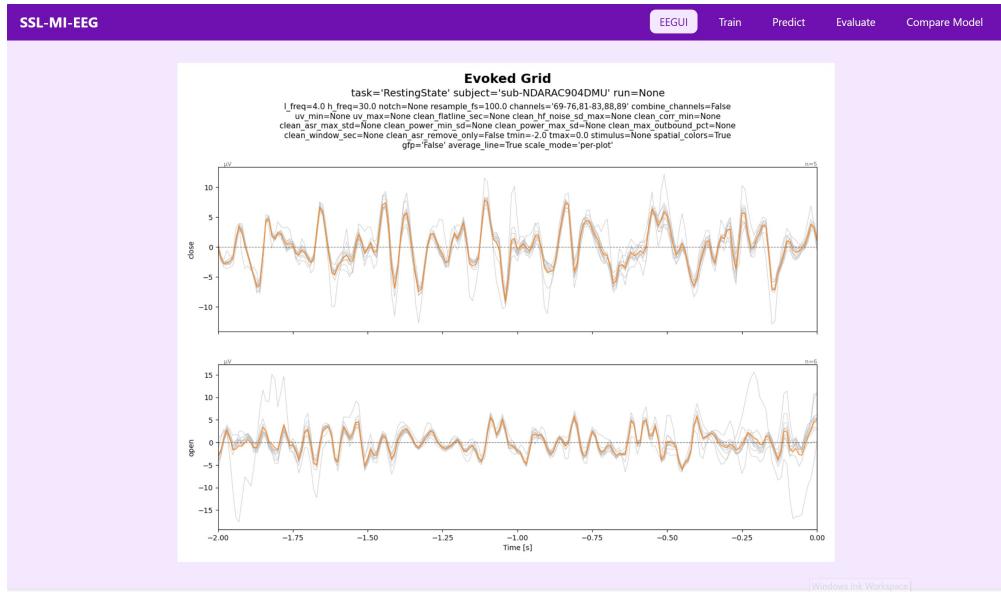


Figure 5.5: Evoked grid view

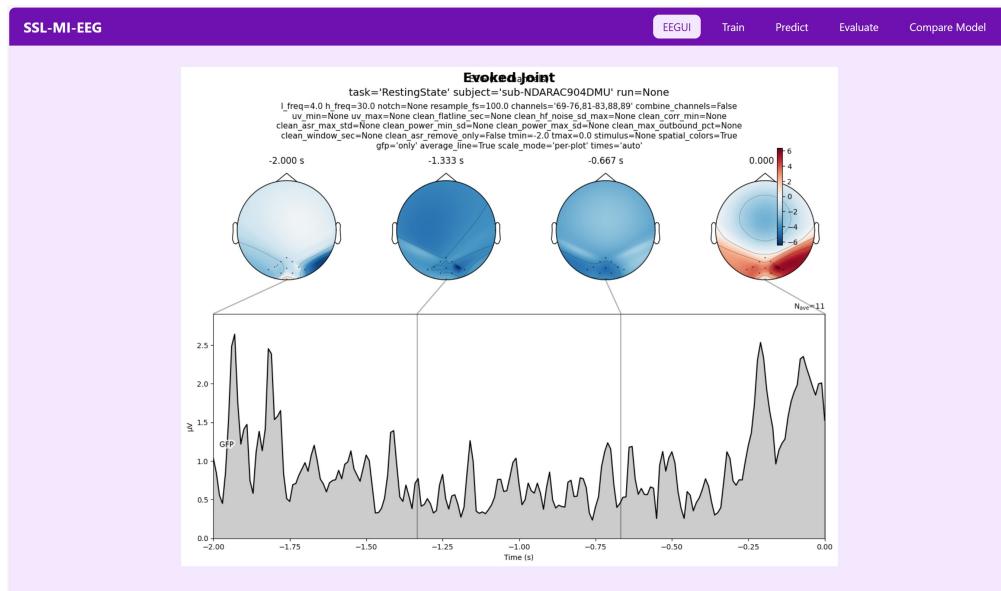


Figure 5.6: Evoked joint plot

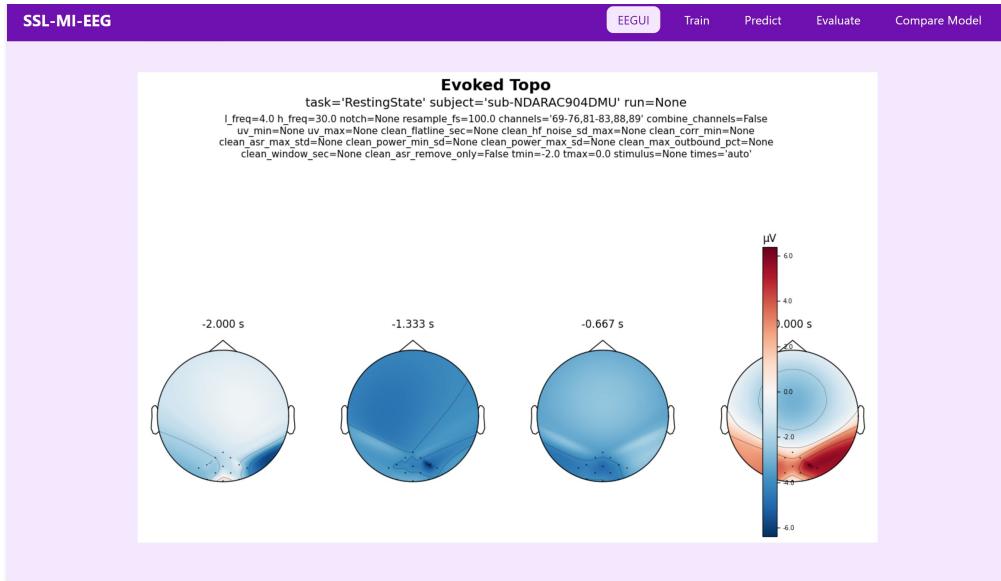


Figure 5.7: Evoked topomap

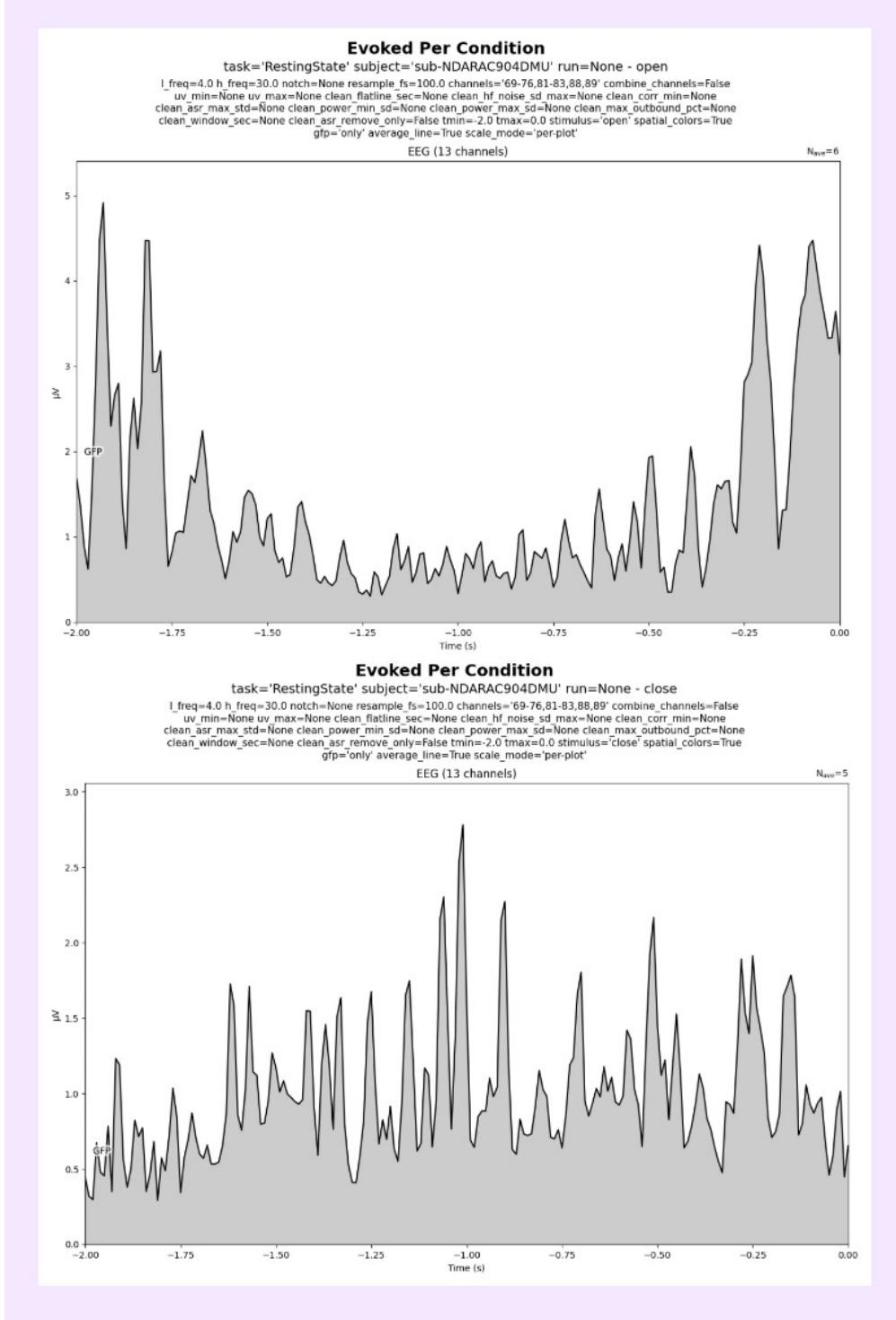


Figure 5.8: Evoked per condition

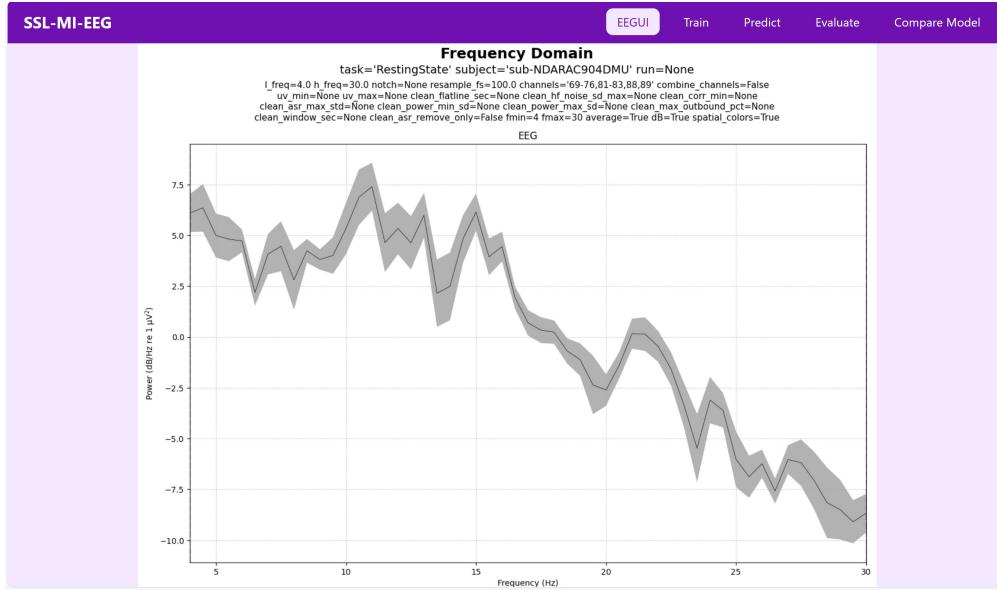


Figure 5.9: Frequency spectrum



Figure 5.10: Power spectral density grid



Figure 5.11: Signal-to-noise ratio spectrum

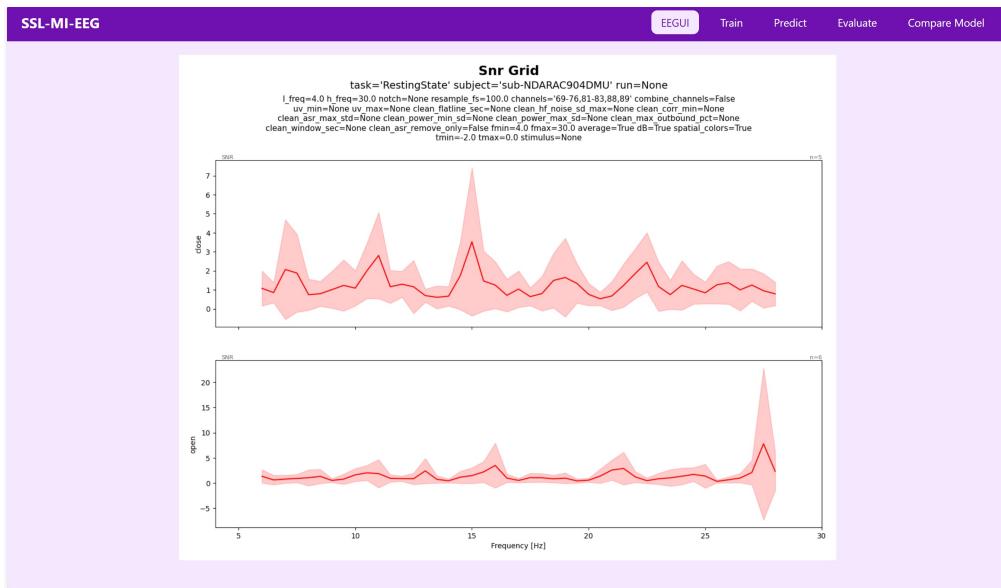


Figure 5.12: Signal-to-noise ratio grid

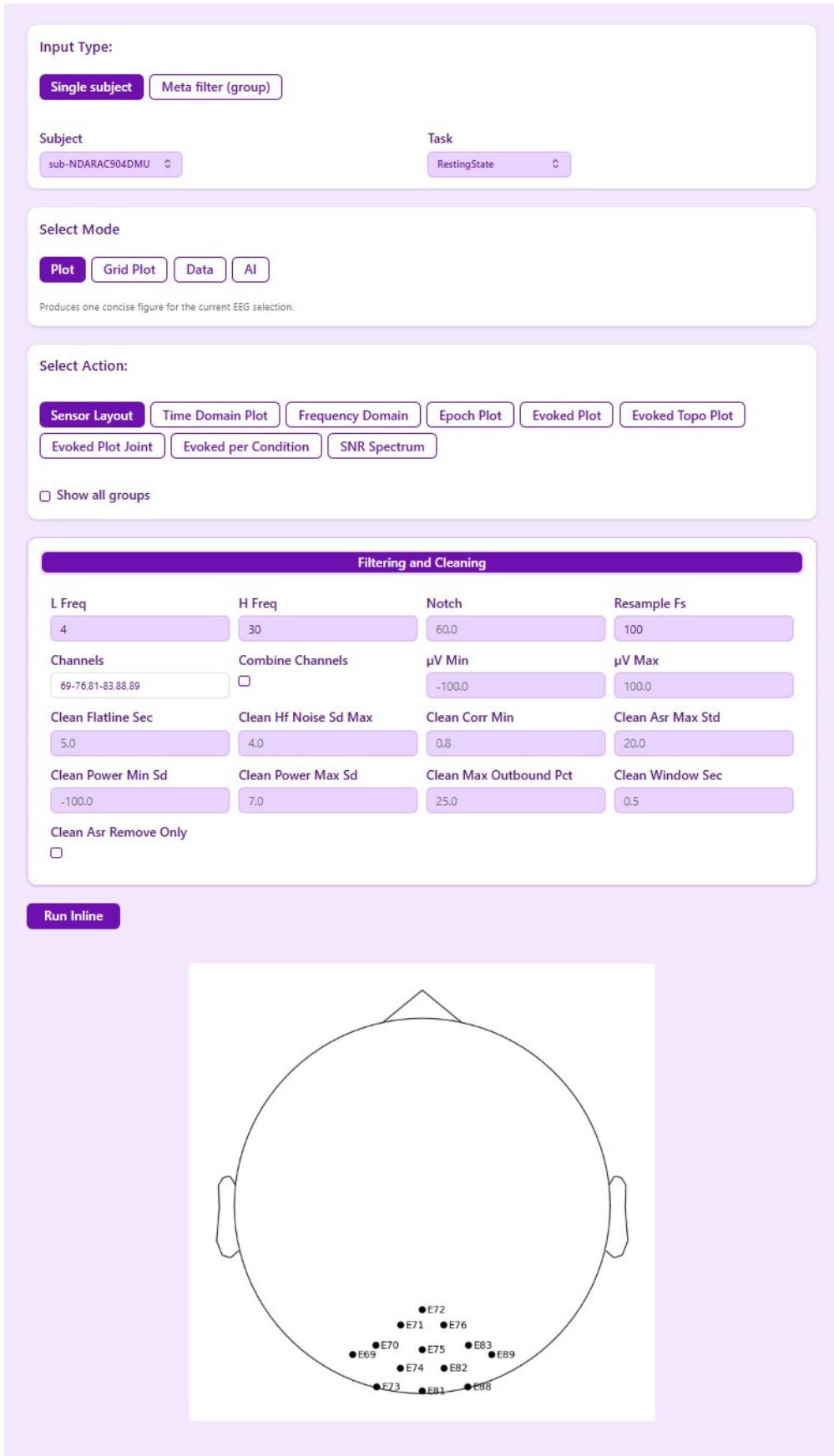


Figure 5.13: Sensor layout

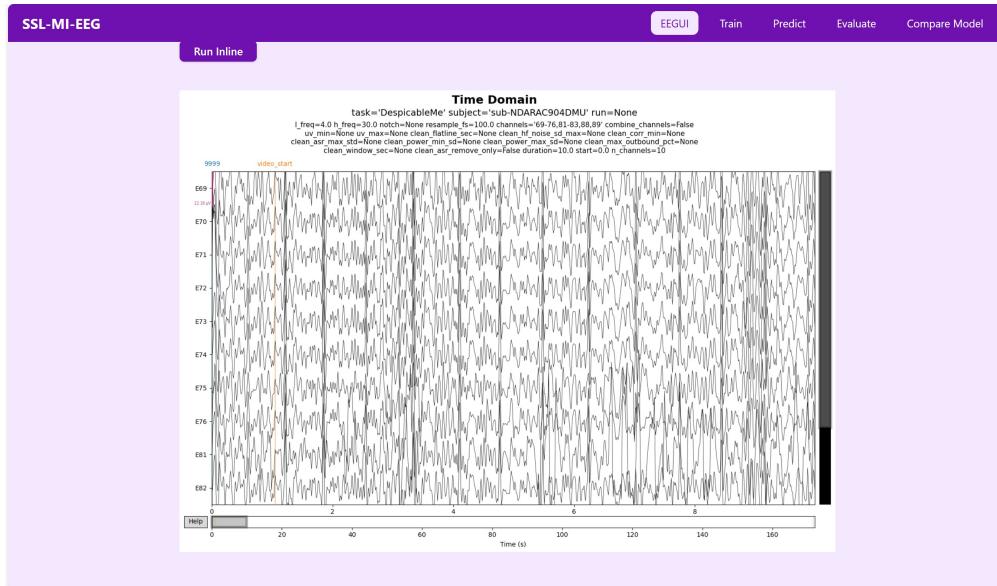


Figure 5.14: Time-domain signal

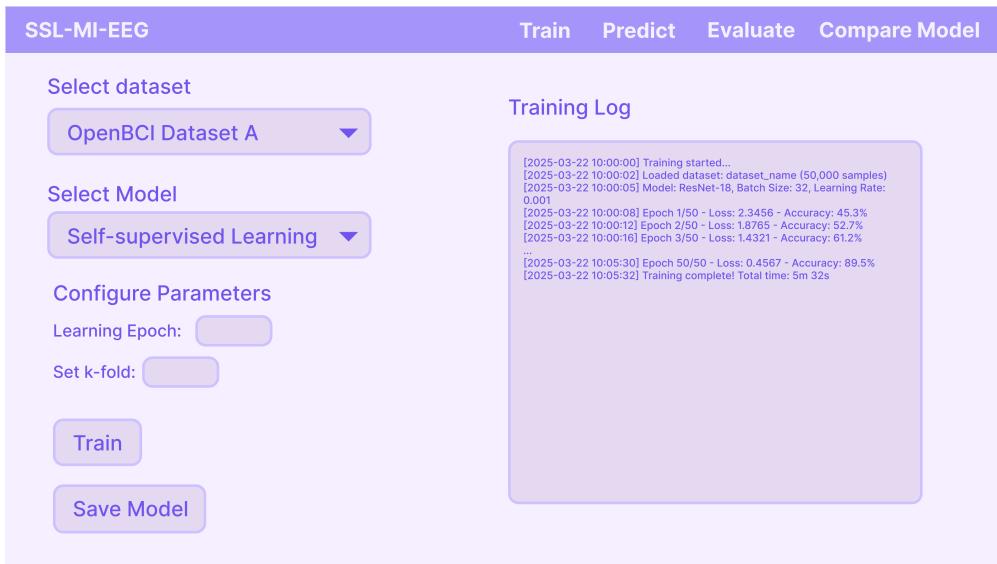


Figure 5.15: Training Interface – Dataset selection, SSL configuration, and training log

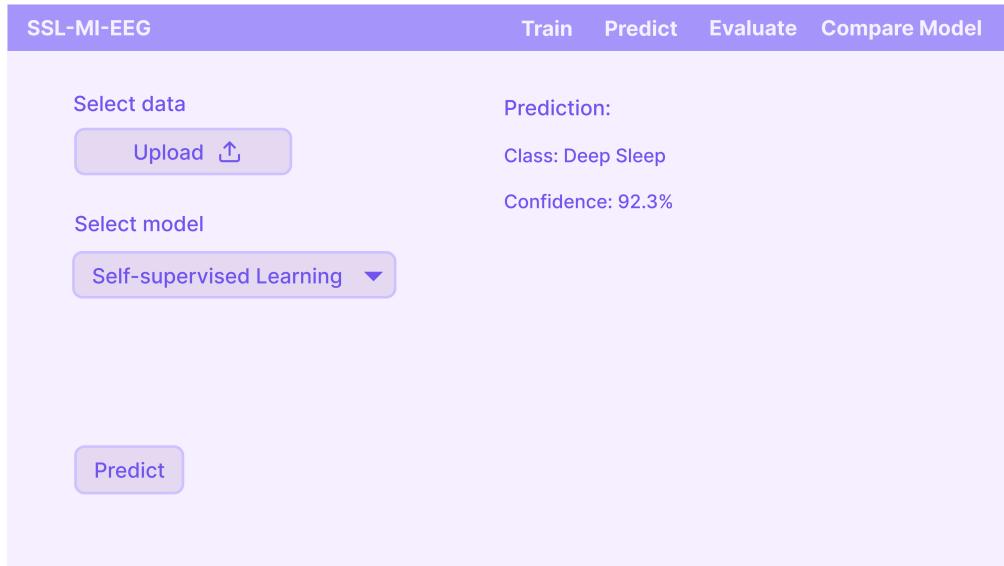


Figure 5.16: Prediction Interface – Uploading input data and viewing output class with confidence



Figure 5.17: Evaluation Interface – Upload trained model and view performance metrics and curve

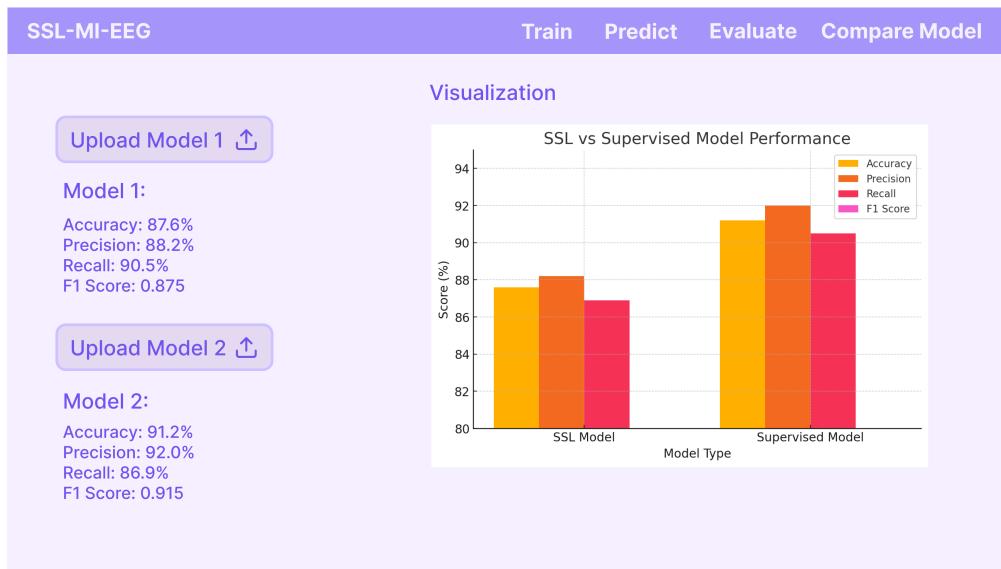


Figure 5.18: Comparison Interface – Visual comparison between SSL and Supervised models

5.5 Deployment Strategy

5.5.1 Deployment Plan

The AI-based application is deployed as a web-based system in a research environment. The trained model is hosted on a backend server and accessed through a browser interface.

The system consists of:

- Frontend — interactive research interface
- Backend — REST API service
- Model — EEG classification network

Technologies used:

- React — interactive frontend framework
- FastAPI — model service endpoints
- TensorFlow — neural network execution

- scientific Python libraries — preprocessing and evaluation

The frontend web system replaces manual notebook execution and allows controlled interaction with the EEG pipeline, and the backend exposes endpoints for training, prediction, and evaluation, allowing modular integration and testing.

The core model used in our system is based on **MixNet-BCI**, an open-source EEG classification framework developed by VISTEC. This model is integrated into our workflow for signal preprocessing, feature extraction, and classification of motor imagery tasks.

5.5.2 Proof of Concept

The model pipeline was validated using benchmark EEG datasets under two conditions:

- **Subject-Dependent Setting:** Model is trained and tested on individual subjects.
- **Subject-Independent Setting:** Model is trained on a group of subjects and tested on unseen individuals.

The system successfully supports end-to-end workflow:
visualization → dataset selection → training → inference → evaluation

Results can be accessed both through the interface and API endpoints.

5.6 Reflection and Future Development

Lessons Learned

- SSL can extract high-quality embeddings from unlabeled EEG data that generalize to MI classification tasks.
- Multi-paradigm training (e.g., SSVEP + ERP) provides richer features than MI alone.
- Integrated workflows reduce experimentation overhead

Challenges

- subject variability affects generalization
- interpreting learned representations remains difficult
- noisy EEG signals impact performance

Future Work

- transformer-based temporal modeling
- real-time prediction interface
- automated hyperparameter search
- live experiment integration

Chapter 6

Deliverables and Evaluation

6.1 Artifacts

- **Project repository:** Contains preprocessing scripts, baseline modeling, and transfer training code. [Link: to be added]
- **Figures:** SNR spectra and evoked response plots verifying SSVEP/P300 for SuS and CCD; updated class diagram.
- **Logs:** Preprocessing and training logs for reproducibility.

6.2 Evaluation Protocol

- **Baseline (CCD-ITI only):** Train on available CCD-ITI segments; report RT MAE/MSE and accuracy/F1.
- **Transfer (SuS → CCD-ITI):** Pretrain on SuS; fine-tune on CCD-ITI; compare metrics against baseline.
- **Verification:** Visualize SNR spectra and evoked responses to confirm signal components.

Chapter 7

Conclusion and Discussion

7.1 Reflection

Summarize the key outcomes and contributions of the project.

7.2 Challenges

List the main technical and non-technical challenges encountered.

- Challenge 1: ...
- Challenge 2: ...

7.3 Future Work

Outline planned next steps, improvements, and open questions.

- Next step 1: ...
- Next step 2: ...

Reference

Bibliography

- [1] B. Aristimunha, D. Truong, P. Guetschel, S. Y. Shirazi, I. Guyon, A. R. Franco, M. P. Milham, A. Dotan, S. Makeig, A. Gramfort, J.-R. King, M.-C. Corsi, P. A. Valdés-Sosa, A. Majumdar, A. Evans, T. J. Sejnowski, O. Shriki, S. Chevallier, and A. Delorme, “Eeg foundation challenge: From cross-task to cross-subject eeg decoding,” NeurIPS 2025 Competition Proposal (preprint), 2025.
- [2] S. Y. Shirazi, A. Franco, M. S. Hoffmann, N. B. Esper, D. Truong, A. Delorme, M. P. Milham, and S. Makeig, “Hbn-eeg: The fair implementation of the healthy brain network (hbn) electroencephalography dataset,” bioRxiv, 2024.
- [3] L. M. Alexander, J. Escalera, L. Ai, C. Andreotti, K. Febre, A. Mangone, N. Vega-Potler, N. Langer, A. Alexander, M. Kovacs, S. Litke, B. O’Hagan, J. Andersen, B. Bronstein, A. Bui, M. Bushey, H. Butler, V. Castagna, N. Camacho, E. Chan, D. Citera, J. Clucas, S. Cohen, S. Dufek, M. Eaves, B. Fradera, J. Gardner, N. Grant-Villegas, G. Green, C. Gregory, E. Hart, S. Harris, M. Horton, D. Kahn, K. Kabotyanski, B. Karmel, S. P. Kelly, K. Kleinman, B. Koo, E. Kramer, E. Lennon, C. Lord, G. Mantello, A. Margolis, K. R. Merikangas, J. Milham, G. Minniti, R. Neuhaus, A. Levine, Y. Osman, L. C. Parra, K. R. Pugh, A. Racanello, A. Restrepo, T. Saltzman, B. Septimus, R. Tobe, R. Waltz, A. Williams, A. Yeo, F. X. Castellanos, A. Klein, T. Paus, B. L. Leventhal, R. C. Craddock, H. S. Koplewicz, and M. P. Milham, “An open resource for transdiagnostic research in pediatric mental health and learning disorders,” *Scientific Data*, vol. 4, no. 170181, 2017.
- [4] N. Langer, E. J. Ho, L. M. Alexander, H. Y. Xu, R. K. Jozanovic, S. Henin, A. Petroni, S. Cohen, E. T. Marcelle, L. C. Parra, M. P. Milham, and S. P. Kelly, “A resource for assessing information processing

in the developing brain using eeg and eye tracking,” *Scientific Data*, vol. 4, no. 170040, 2017.

- [5] Overleaf, “Learn latex in 30 minutes,” https://www.overleaf.com/learn/latex/Learn_LaTeX_in_30_minutes.

Appendix A

Appendix A: Example

<TIP: Put additional or supplementary information/data/figures in appendices. />

Appendix B

Appendix B: About \LaTeX

\LaTeX (stylized as \LaTeX) is a software system for typesetting documents. \LaTeX markup describes the content and layout of the document, as opposed to the formatted text found in WYSIWYG word processors like Google Docs, LibreOffice Writer, and Microsoft Word. The writer uses markup tagging conventions to define the general structure of a document, to stylize text throughout a document (such as bold and italics), and to add citations and cross-references.

\LaTeX is widely used in academia for the communication and publication of scientific documents and technical note-taking in many fields, owing partially to its support for complex mathematical notation. It also has a prominent role in the preparation and publication of books and articles that contain complex multilingual materials, such as Arabic and Greek.

Overleaf has also provided a 30-minute guide on how you can get started on using \LaTeX . [5]