# Brain-Computer Interface with AI Integration

Pavitha Dissanayake
*Department of Computer Science and Engineering*
*University of Moratuwa*
pavitha.21@cse.mrt.ac.lk

Dr. Uthayasanker Thayasivam
*Department of Computer Science and Engineering*
*University of Moratuwa*
ruthaya@cse.mrt.ac.lk

***Abstract*—Brain–Computer Interfaces (BCIs) translate neural activity into actionable commands, but decoding meaningful patterns from EEG signals remains challenging due to low signal-to-noise ratios and inter-subject variability. This study investigates EEG-based ERP classification using the N170 dataset from the EEG-ExPy benchmark. Classical covariance-based pipelines, including XdawnCov + TS and ERPCov + TS, achieved strong performance with mean AUCs above 0.69, demonstrating effective feature extraction from single-trial EEG. A deep residual CNN (DeepResNetEEG) successfully learned temporal–spatial patterns during training, reaching high training accuracy, but exhibited overfitting with test performance near chance. These preliminary results highlight the promise of deep architectures for EEG representation learning while emphasizing the need for improved generalization. Future work will focus on integrating regularization, data augmentation, and transformer-based architectures, as well as extending experiments to additional ERP paradigms such as P300. Overall, this work provides a reproducible framework for advancing AI-enhanced EEG-based BCI research.**

***Index Terms*—Brain-Computer Interface, EEG, Deep Learning, Neural Networks, AI Integration**

## I. Introduction

Brain–Computer Interfaces (BCIs) enable direct communication between the human brain and external devices by translating neural activity into executable commands. They have gained increasing attention in recent years for their potential applications in neurorehabilitation, assistive technologies, and human–machine interaction. Among various neuroimaging techniques, Electroencephalography (EEG) remains the most widely used in BCI systems due to its non-invasive nature, affordability, and high temporal resolution. However, decoding meaningful patterns from EEG signals remains a significant challenge because of low signal-to-noise ratio, non-stationarity, and high inter-subject variability.

Traditional EEG-based BCI approaches have relied on handcrafted feature extraction methods such as Common Spatial Patterns (CSP) and bandpower analysis, followed by classical machine learning models like Support Vector Machines (SVM) or Linear Discriminant Analysis (LDA). While effective to an extent, these methods struggle to generalize across different sessions and subjects. In contrast, recent advances in deep learning and hybrid architectures,combining convolutional and recurrent models,have shown promising improvements in automatically learning spatio-temporal features from raw EEG signals.

Despite these advances, many existing studies face limited reproducibility and inconsistent benchmarking, making it difficult to compare model performance fairly across datasets and methodologies. To address this, NeuroTechX introduced the EEG-ExPy benchmark, a standardized framework for EEG-based experimentation, offering reproducible pipelines and diverse motor imagery datasets. Leveraging this benchmark allows researchers to fairly assess algorithmic improvements while ensuring transparency and replicability.

This research project explores the integration of advanced AI models within the EEG-ExPy framework to enhance EEG-based motor imagery classification. Specifically, the aim is to design and evaluate deep learning architectures that improve signal representation, optimize generalization across subjects, and advance the reproducibility of BCI research. Preliminary experiments focus on baseline replication and model adaptation using convolutional-recurrent hybrids, setting the stage for future integration of attention-based or transformer modules.

## II. Related Work

Recent advances in EEG-based Brain–Computer Interfaces (BCIs) have centered on improving the decoding of neural responses such as event-related potentials (ERPs) and steady-state visual evoked potentials (SSVEPs). Traditional classifiers,such as Linear Discriminant Analysis (LDA), Support Vector Machines (SVM), and Task-Related Component Analysis (TRCA),have shown strong within-subject accuracy but often fail to generalize across subjects or paradigms. To address these limitations, deep learning and hybrid architectures have been increasingly adopted to model the complex spatio-temporal structure of EEG signals.

N170-related research primarily targets face and visual stimulus recognition tasks. Early ERP studies using handcrafted spectral features were constrained by low signal-to-noise ratios and limited trial-level discrimination. More recent methods leverage lightweight convolutional and recurrent networks capable of learning discriminative temporal–spatial features directly from raw EEG. Liu et al. [11] proposed a compact CNN architecture achieving substantial accuracy improvements across visual paradigms, while Zhang et al. [17] demonstrated that spatial principal component analysis can enhance SVM-based EEG decoding performance. These studies highlight the benefits of feature learning but also underscore the need for generalizable models that can adapt across paradigms and subjects.

In P300 classification, several approaches aim to improve both accuracy and robustness in detecting target responses. Classical frameworks such as stepwise linear discriminant analysis and Bayesian classifiers have been gradually replaced by multi-scale convolutional and hybrid recurrent models. Borra et al. [3] introduced MS-EEGNet, a lightweight multi-scale CNN that achieved superior decoding accuracy while maintaining computational efficiency. Afrah et al. [1] extended this with an unsupervised CLSTM autoencoder for feature extraction, achieving improved cross-session stability. Furthermore, domain adaptation and data fusion approaches,such as Du and Li et al. [7] and Ermaganbet et al. [8] ,enable inter-subject transfer learning using centralized multi-person data and dual-input CNNs. Self-supervised approaches like SpellerSSL [9] further enhance P300 speller robustness through representation learning from unlabeled EEG signals.

SSVEP-based BCIs present another challenge: decoding periodic responses from visual flickers in real time. Canonical correlation analysis and TRCA remain strong baselines, but recent advances show that combining them with deep neural networks leads to significant gains. Transformer-based models such as SSVEPformer [4] and time–frequency fusion networks like SSVEP-TFFNet [5] outperform traditional methods by learning global temporal dependencies and sub-band frequency interactions. Similarly, Dong et al. [6] used an xLSTM with spatial attention to enhance SSVEP spelling performance, underscoring the benefit of attention mechanisms in decoding periodic neural dynamics.

While these studies demonstrate the growing power of deep learning in EEG classification, several gaps remain. Most models are dataset- or paradigm-specific, limiting cross-paradigm adaptability. Others achieve strong accuracy but lack reproducibility due to non-standardized preprocessing and evaluation pipelines. The EEG-ExPy benchmark addresses this limitation by offering reproducible open-source notebooks for N170, P300, and SSVEP paradigms, enabling consistent evaluation across algorithms [2], [10]. However, few studies have systematically explored integrating advanced AI models within this standardized framework.

In this context, this research builds upon prior work by incorporating deep hybrid architectures within the EEG-ExPy benchmark to improve classification accuracy, reproducibility, and generalization. By focusing on reproducible pipelines, lightweight architectures, and potential transformer integration, the aim is to advance the interpretability and real-world applicability of EEG-based BCIs.

## III. Methodology

This study investigates EEG-based brain–computer interface (BCI) decoding using publicly available datasets from the EEG-ExPy benchmark, focusing on event-related potentials (ERPs) such as N170 and P300. The objective is to enhance single-trial classification and cross-subject generalization through deep learning models while maintaining comparability with established baseline methods.

### A. Data and Preprocessing

The datasets include multi-channel EEG recordings (32–64 channels, 250–512 Hz sampling rate) collected during visual cognitive tasks involving face/non-face or target/non-target stimuli. Each trial is time-locked to stimulus onset and epoched from 200 ms to +800 ms. Preprocessing steps include 1–30 Hz bandpass filtering, baseline correction, and artifact rejection via Independent Component Analysis (ICA) when necessary. The data are standardized across channels to reduce inter-subject amplitude variability.

### B. Baseline Models

To establish reference performance, traditional EEG classifiers such as Logistic Regression, Linear Discriminant Analysis (LDA), and pyRiemann's Minimum Distance to Mean (MDM) are implemented. These models operate on vectorized or covariance-based EEG features and serve as strong baselines for ERP classification.

### C. Deep Learning Models

Advanced architectures are then developed to capture spatial-temporal dependencies within EEG signals. A Convolutional Neural Network (CNN) is employed for localized pattern extraction across time and electrode dimensions. Additional experiments include Multilayer Perceptrons (MLPNNs) for compact representations and hybrid CNN-RNN or transformer-based architectures for modeling sequential dependencies across channels and time. Each model outputs class probabilities corresponding to the cognitive condition (e.g., target vs. non-target).

### D. Experimental Design

Experiments are conducted in both within-subject and cross-subject paradigms to evaluate generalization performance. The primary evaluation metrics are Accuracy (ACC) and Area Under the ROC Curve (AUC), supplemented by F1-score and precision/recall where appropriate. Model training uses the PyTorch framework on Python 3.10+, with GPU acceleration for deep learning models.

Implementation Plan

The workflow comprises four phases:

- (1) Data preprocessing and quality verification
- (2) Model implementation and validation
- (3) Experimental evaluation and metric computation
- (4) Result analysis and reporting

This pipeline ensures reproducibility, comparability with EEG-ExPy benchmarks, and a structured foundation for extending toward transformer-based BCI models in future work.

## IV. Experiments and Results

### A. Experimental Setup

Experiments were conducted using the EEG-ExPy N170 dataset, which contains multi-channel event-related potentials (ERPs) recorded during visual stimulus recognition tasks. Each trial consisted of face and non-face stimuli, with epochs time-locked to stimulus onset (200 ms to +800 ms) and band-limited to 1–30 Hz. Following preprocessing and artifact rejection, data were standardized and split using Stratified Shuffle Split cross-validation (10 folds, 25% test split) to ensure class balance and robustness of evaluation.

The comparison included both classical and deep learning approaches. Baseline models implemented included:

- **Vect + LR**: Vectorized EEG features with logistic regression
- **Vect + RegLDA**: Regularized LDA with automatic shrinkage
- **ERPCov + TS**: ERP covariance features with tangent-space logistic regression
- **ERPCov + MDM**: Minimum-distance-to-mean classifier on ERP covariance
- **XdawnCov + TS / MDM**: Covariance-based pipelines using Xdawn spatial filtering

These represent state-of-the-art ERP classification methods widely used for benchmarking in BCI literature.

For deep learning, a Deep Residual CNN (DeepResNetEEG) model was developed. The network accepts EEG inputs in the shape (channels × samples × 1) and employs multiple convolutional and residual blocks to capture temporal–spatial dependencies. It includes batch normalization, ReLU activations, and skip connections to stabilize learning, followed by global average pooling and a dense softmax classifier. The model was trained with the Adam optimizer (learning rate = 0.001), batch size = 8, and 50 epochs. Implementation was performed in TensorFlow/Keras, wrapped with SciKeras for sklearn compatibility.

All experiments were executed on Python 3.11 with PyTorch, MNE, and pyRiemann frameworks, and utilized GPU acceleration for deep models. Evaluation metrics included Area Under the ROC Curve (AUC) as the primary metric, complemented by accuracy for the deep learning model to monitor overfitting behavior.

### B. Results

A summary of AUC performance across all methods is presented below:

Classical pipelines based on covariance features (particularly XdawnCov + TS and ERPCov + TS) achieved the highest mean AUC scores, indicating strong discrimination between face and non-face trials. These methods benefit from tangent-space mapping, which effectively linearizes covariance representations of ERP patterns.

In contrast, the DeepResNetEEG model displayed clear overfitting tendencies. While training accuracy consistently reached the high 90 % range, test accuracy plateaued around 50 %, with a corresponding mean AUC of 0.507. This indicates that the CNN successfully captures dataset-specific temporal–spatial structures but struggles to generalize across subjects or sessions. The limited dataset size and high inter-subject variability likely contribute to this gap, as deep architectures typically require larger and more diverse samples for robust generalization.

TABLE I
PERFORMANCE COMPARISON OF DIFFERENT METHODS

| **Method** | **Mean** | **Std** | **Max** |
|---|---|---|---|
| DeepResNetEEG | 0.507 | 0.018 | 0.525 |
| ERPCov + MDM | 0.683 | 0.015 | 0.698 |
| ERPCov + TS | 0.697 | 0.029 | 0.726 |
| Vect + LR | 0.650 | 0.022 | 0.672 |
| Vect + RegLDA | 0.663 | 0.027 | 0.690 |
| XdawnCov + MDM | 0.656 | 0.016 | 0.672 |
| XdawnCov + TS | 0.700 | 0.029 | 0.729 |

### C. Discussion

Despite its limited test performance, the preliminary deep learning results are encouraging, as the strong training convergence confirms that EEG temporal–spatial features are being effectively encoded. This establishes a foundation for improving generalization through several strategies:

- Regularization and data augmentation (e.g., dropout, noise injection, channel dropout) to reduce overfitting
- Cross-subject normalization and domain adaptation to handle inter-participant variability
- Transfer learning or pretraining using related EEG datasets to leverage shared cognitive representations
- Incorporation of transformer-based attention mechanisms to better capture long-range dependencies across time and electrodes

Future work will extend experiments to the P300 paradigm and assess transformer-based architectures against the current baselines. Overall, these preliminary results demonstrate that while classical covariance-based models remain strong performers for ERP classification, deep residual networks hold potential once appropriately regularized and scaled for EEG data.

## V. Conclusion and Future Work

This study presents preliminary investigations into EEG-based brain–computer interface decoding using the EEG-ExPy N170 dataset. Classical covariance-based pipelines, particularly XdawnCov + TS and ERPCov + TS, demonstrated strong performance in discriminating face versus non-face stimuli, achieving mean AUC values above 0.69. In contrast, a DeepResNetEEG model successfully captured temporal–spatial patterns within the training set, but exhibited significant overfitting, resulting in near-chance test performance. These findings indicate that while deep learning architectures can encode meaningful EEG representations, robust cross-subject generalization remains a key challenge.

The primary limitations observed so far include overfitting of deep models, limited dataset size, and inter-subject variability, which collectively constrain the test performance of high-capacity architectures. Additionally, the current experiments are restricted to the N170 paradigm; evaluation on other ERP types such as P300 and SSVEP is required to assess model versatility and multi-paradigm generalization.

Future work will focus on improving model generalization and expanding the experimental scope. Planned extensions include:

- Integration of regularization techniques (dropout, early stopping) and data augmentation strategies to mitigate overfitting
- Exploration of transformer-based architectures to enhance temporal feature extraction across EEG channels
- Extension of experiments to P300 decoding, enabling evaluation of model robustness across different ERP paradigms
- Investigation of domain adaptation and transfer learning to improve cross-subject performance

By addressing these challenges, the study aims to develop lightweight, generalizable deep learning models for EEG classification, contributing both practical insights and a reproducible framework for future BCI research using publicly available EEG datasets.

## References


[1] R. Afrah, Z. Amini, and R. Kafieh. An unsupervised feature extraction method based on clstm-ae for accurate p300 classification. *Journal of Biomedical Physics & Engineering*, 2024.

[2] Bruno Aristimunha, Raphael Y. de Camargo, Walter H. Lopez Pinaya, Sylvain Chevallier, Alexandre Gramfort, and Cedric Rommel. Evaluating the structure of cognitive tasks with transfer learning. arXiv preprint, 2023.

[3] D. Borra et al. Ms-eegnet: A lightweight multi-scale convolutional neural network for p300 decoding. *Journal of Neural Engineering*, 2021.

[4] J. Chen, Y. Zhang, P. Peng, et al. A transformer-based deep neural network model for ssvep classification (ssvepformer). arXiv preprint, 2022.

[5] Y. Dai, Z. Chen, et al. A time-frequency feature fusion-based deep learning network for ssvep frequency recognition (ssvep-tffnet). *Frontiers in Neuroscience*, 2025.

[6] W. Dong, C. Xu, et al. Enhanced ssvep bionic spelling via xlstm-based deep learning and spatial attention. *Biomimetics*, 2025.

[7] P. Du, P. Li, et al. Single-trial p300 classification algorithm based on centralized multi-person data fusion cnn. *Frontiers in Neuroscience*, 2023.

[8] Z. Ermaganbet, A. Mussabayeva, et al. Subject-independent p300 speller classification using time-frequency representation and double input cnn with feature concatenation. In *IEEE DSP Conference*, 2023.

[9] J. Hong, G. Mackellar, and S. Ghane. Spellerssl: Self-supervised learning with p300 aggregation for speller bcis. arXiv preprint, 2025.

[10] Roman Kessler, Alexander Enge, and Michael A. Skeide. How eeg preprocessing shapes decoding performance. *Communications Biology*, 2025.

[11] Yi Liu, Zach Quince, Steven Goh, Shoryu Teragawa, and Tobias Low. A lightweight deep learning model for eeg classification across visual stimuli. University of Southern Queensland preprint, 2023.

[12] V. Marochko et al. Integrated gradients for enhanced interpretation of p3b-erp classifiers trained with eeg-superlets in traditional and virtual environments. In *CEUR Workshop*, 2025.

[13] J. A. O'Reilly et al. Blind source separation of event-related potentials using a recurrent neural network. *bioRxiv*, 2024.

[14] J. A. O'Reilly, J. Wehrman, et al. Neural correlates of face perception modeled with a convolutional recurrent neural network. *bioRxiv*, 2023.

[15] Y. Ravipati, N. Pouratian, et al. Evaluating deep learning performance for p300 neural signal classification. In *AMIA Annual Symposium Proceedings*, 2024.

[16] Siyuan Zang, Xiaojun Ding, Meihong Wu, and Changle Zhou. An eeg classification-based method for single-trial n170 latency detection and estimation. *Computational and Mathematical Methods in Medicine*, 2022.

[17] G. Zhang et al. Assessing the effectiveness of spatial pca on svm-based decoding of eeg data. *NeuroImage*, 2023.