# Predicting Cognitive and Verbal Intentions using EEG

Himanshi*
*2022UCA1802*
himanshi.ug22@nsut.ac.in

Tripti Gusain*
*2022UCA1812*
tripti.gusain.ug22@nsut.ac.in

Varu Solanki*
*2022UCA1873*
varu.solanki.ug22@nsut.ac.in

***Abstract*—Electroencephalography (EEG) is a widely used non-invasive technique for recording brain activity with high temporal resolution. In this study, we explore the use of EEG signals to classify object categories that participants view and subsequently name aloud. Our approach focuses on decoding neural signals corresponding to visual object recognition tasks. Participants were shown images of various object classes and asked to vocalize the object's label, during which EEG data was recorded using a 64-channel setup. We processed and extracted features from the EEG signals and trained a deep learning classifier to predict the object class solely based on the brain activity. While this work does not explicitly distinguish between cognitive and verbal intention stages, it provides a foundational step toward understanding and modeling intention from brain signals. The results demonstrate the feasibility of decoding visual object categories from EEG, contributing to the development of more advanced brain-computer interface (BCI) applications.**

***Index Terms*—EEG, Brain-Computer Interface, Brain Signals, EEG based Visual Classification**

## I. Introduction

The field of Brain-Computer Interfaces (BCIs) has rapidly progressed with the goal of enabling direct communication between the brain and external systems. Among various neural recording techniques, electroencephalography (EEG) stands out for its non-invasiveness, portability, and millisecond-level temporal resolution. EEG-based decoding of visual stimuli has gained interest as a method for understanding cognitive processes and building intuitive assistive technologies.

Despite significant advances, accurately decoding specific cognitive states from EEG signals remains challenging. The low signal-to-noise ratio of EEG recordings, coupled with their high dimensionality and temporal complexity, creates substantial obstacles for reliable classification. Furthermore, disentangling neural signatures of distinct cognitive processes—such as visual perception versus speech preparation—presents additional methodological challenges.

A particularly compelling application domain is the classification of visual object categories from brain activity. While functional Magnetic Resonance Imaging (fMRI) has demonstrated success in this area, EEG-based approaches offer advantages in temporal resolution and accessibility but face greater challenges in spatial localization and signal clarity. Previous EEG studies have shown promising results in binary classification tasks involving face recognition and broad categorical distinctions, but multi-class visual object category discrimination remains underdeveloped.

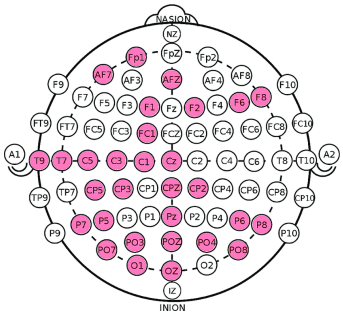


Fig. 1. 64 channels based 10-20 EEG

In this work, we investigate whether EEG signals alone can be used to classify visual object categories that a person sees and subsequently names. Our experimental paradigm involves participants viewing images representing distinct object classes for brief periods, followed by verbalization of the object label. Throughout these sessions, we record EEG signals using a 64-channel cap aligned to the international 10-20 system, with a sampling rate of 1000 Hz.

The primary objective of our study is to classify viewed object categories using EEG data, treating the brain's electrical response as a proxy for cognitive processing. While our experimental setup involves both visual perception and verbalization, our classification approach focuses on object category without explicitly isolating neural correlates of speech preparation.

## II. Dataset

The dataset used in this study consists of high-resolution EEG recordings obtained from 16 participants. Each participant was fitted with a 64-channel EEG cap following the international 10-20 electrode placement system. The signals were sampled at a frequency of 1000 Hz, providing detailed temporal information on neural activity.

During the experiment, participants were presented with images representing one out of 80 object categories. Each image was displayed for a duration of 0.5 seconds, followed by a prompt for the participant to verbally state the label of the object. The EEG system continuously recorded signals throughout the stimulus presentation and verbal response period.

*All authors contributed equally to this work.

This setup captures brain activity associated with both the perception of the object and the preparation to articulate its label.

Each trial in the dataset includes:

- Raw EEG time-series from all 64 channels (2 are reference nodes channels which are removed),
- A corresponding object class label,

The dataset is well-suited for supervised learning tasks involving EEG-based classification. It allows for the exploration of neural patterns linked to visual object recognition and supports the development of machine learning models for intention prediction from brain signals.

## III. Literature Survey

A number of recent studies have explored the intersection of EEG signals and computer vision, focusing on decoding visual information from brain activity. Among these, datasets such as EEG-ImageNet have provided large-scale benchmarks with multi-granularity labels for EEG-based visual stimulus classification. These datasets have enabled the development of deep learning architectures for modeling cognitive processes from EEG data, often focusing on coarse or fine-grained object categories.

Some Works introduced early frameworks using convolutional neural networks (CNNs) for decoding object categories from EEG signals and related follow-ups have further advanced EEG classification by employing randomized trial schemes and various attention-based architectures. These efforts show that EEG signals contain sufficient discriminative information to classify stimuli under controlled conditions.

However, many public EEG datasets often suffer from limited label variety, temporal dependencies between trials, or structured presentation of stimuli, which may introduce unwanted biases in the models. In contrast, our dataset is privately collected and significantly richer in both scale and diversity. It includes 80 distinct object class labels, offering a more challenging and realistic benchmark for EEG classification.

To ensure the validity of our findings, we carefully **randomized** the order of stimulus presentation across trials and participants. During model training, data shuffling techniques are employed to eliminate any potential temporal dependencies or ordering effects. *This helps guarantee that classification accuracy stems from true neural correlates of visual cognition, rather than artifacts of sequence learning.*

Furthermore, unlike many previous studies that focus on coarse-grained or fine-grained data, our work is without granularity dependence.

## IV. Implementation

### *Preprocessing*

The initial stage of our analysis involved a comprehensive preprocessing pipeline applied to the raw EEG data. Continuous recordings, originally sampled at 1000 Hz, were first subjected to a band-pass filter designed to isolate the frequency components relevant to neural activity, specifically between 0.5 Hz and 70 Hz. This filtering step helped to attenuate both very slow signal drifts and high-frequency noise that could obscure the underlying brain signals. To further refine the data, we implemented a notch filter centered at 50 Hz to specifically remove interference from power-line sources, a common artifact in EEG recordings. This filtering was performed using a forward-backward technique to ensure that the phase of the signals remained undistorted. Following these filtering steps, we re-referenced the signals from all 64 electrodes to a common average. This procedure subtracts the average signal across all electrodes at each time point, effectively reducing widespread noise that is common to all sensors. Finally, to ensure that the data was on a consistent scale for subsequent analysis, each individual channel was normalized using Z-scoring. This involved subtracting the mean of each channel and dividing by its standard deviation, resulting in signals with a zero mean and unit variance. This multi-stage preprocessing strategy was crucial for preparing the EEG data for meaningful feature extraction and machine learning by minimizing artifacts and standardizing the signal characteristics.

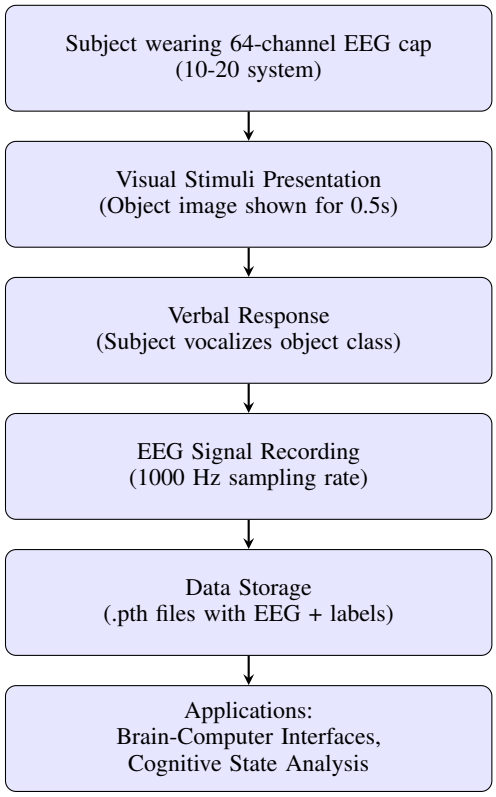


Fig. 2. Experiment flow for EEG data collection

### *Feature Extraction*

To effectively capture the information contained within the preprocessed EEG signals, we employed a multi-faceted feature extraction approach, considering characteristics in the time domain, frequency domain, and inter-channel connectivity. In the time domain, we calculated a range of basic

statistical measures for each channel, including the mean, variance, skewness, kurtosis, maximum, minimum, peak-to-peak amplitude, root mean square, and the rate at which the signal crossed the zero axis. Additionally, we computed the Hjorth parameters—activity, mobility, and complexity—which provide insights into the signal's temporal characteristics and complexity. For the frequency domain analysis, we estimated the power spectral density of the EEG signals using Welch's method. From this spectral representation, we extracted the absolute and relative power within standard EEG frequency bands: delta, theta, alpha, beta, low gamma, and high gamma. We also calculated the spectral edge frequency at 95%, the spectral entropy, and the frequency at which the power spectrum peaked, along with its corresponding power value. To understand how different areas of the brain were interacting, we derived connectivity features. We computed pairwise correlation coefficients between all pairs of EEG channels and then summarized these correlations by calculating their mean, standard deviation, maximum, and minimum values across the channel pairs. Furthermore, we assessed phase synchronization between channels using the phase locking value, derived from the Hilbert transform, and similarly extracted the mean, standard deviation, maximum, and minimum of these values across all channel combinations. All the features extracted from the time domain, frequency domain, and connectivity analysis were then combined into a single feature vector for each data segment, providing a comprehensive representation of the neural activity suitable for the classification tasks that followed.

*Classifier Model*

For the task of predicting cognitive and verbal intentions from the extracted EEG features, we designed a hybrid deep learning architecture that could effectively process both the temporal and spatial aspects of the neural data, as well as static feature representations. Our model incorporates two primary pathways for input processing. The input in the form of a time series, representing the EEG signals across different channels over time, is first processed by a convolutional neural network. This CNN is designed to learn spatial patterns across the EEG channels. The output of the CNN is then fed into a bidirectional long short-term memory network, which is capable of capturing temporal dependencies in both forward and backward directions within the sequence of CNN outputs. To further refine the temporal information, a self-attention mechanism is applied to the LSTM's output. This mechanism allows the model to weigh the importance of different time steps, focusing on the most relevant parts of the signal and producing a fixed-length feature vector that summarizes the temporal dynamics. This attention mechanism learns to assign different levels of importance to each feature in the vector, allowing the model to focus on the most discriminative features for the prediction task. The resulting feature vector is then passed through a shared network of fully connected layers and residual blocks. These residual blocks help in training a deeper network more effectively. To further enhance the feature representation, we incorporated Squeeze-and-Excitation modules within these blocks, which allow the network to adaptively recalibrate the importance of different feature channels. Finally, to improve the robustness and stability of our predictions, the model utilizes an ensemble of multiple classification heads. Each head independently produces a set of class probabilities, and the final prediction is obtained by averaging the outputs of these multiple heads. This dual-path architecture, combined with attention mechanisms, residual connections, feature recalibration, and ensemble learning, allows our model to effectively leverage the complex patterns in EEG data for accurate prediction of cognitive and verbal intentions.

### A. Loss Function and Optimization

To address the challenges associated with multiclass intention recognition, we implemented a cross-entropy loss function, which effectively quantifies the disparity between predicted probability distributions and true class labels. The cross-entropy loss is particularly well-suited for our classification task as it strongly penalizes confident misclassifications while encouraging accurate predictions. We also incorporated class weights inversely proportional to class frequencies:

$$\text{weight}_c = \frac{1}{\text{count}_c} \cdot \frac{\sum_i \text{count}_i}{n_{\text{classes}}} \tag{1}$$

This weighting scheme prevents the model from biasing toward overrepresented intention categories, ensuring fair recognition across all classes.

For optimization, we employed the Adam optimizer with an initial learning rate of 0.0005 and a weight decay parameter of $1 \times 10^{-5}$ for L2 regularization. This configuration provides a good balance between convergence speed and stability. To further improve training dynamics, we implemented a learning rate scheduler that reduced the learning rate when the validation loss plateaus. This adaptive approach helped navigate the complex loss landscape, achieving an optimal balance between exploration and exploitation during training.

This randomization was implemented via the SubsetRandomSampler in PyTorch, which ensures random batch sampling while maintaining stratification constraints. The random seed was fixed at 42 across all experimental procedures to ensure reproducibility while maintaining the benefits of randomization.

### B. Early Stopping and Regularization

To prevent overfitting and ensure optimal generalization, we implemented several regularization techniques:

- Early stopping with a patience of 10 epochs, monitoring validation loss
- Dropout layers with a rate of 0.3-0.4 throughout the network architecture
- Batch normalization after each major layer to stabilize learning
- Gradient clipping with a maximum norm of 1.0 to prevent exploding gradients
- L2 weight decay of $1 \times 10^{-5}$ in the optimizer

### C. Memory Optimization

Given the high-dimensional nature of EEG data and the computational demands of our deep learning architecture, we implemented several memory optimization strategies:

- Batch processing of feature extraction to reduce peak memory usage
- Explicit garbage collection after processing each batch
- Clearing of CUDA cache when using GPU acceleration
- Offloading of tensors to CPU when not actively needed for computation
- Pre-computation and storage of extracted features to avoid redundant processing

These optimizations allowed us to process our extensive dataset efficiently, even with limited computational resources.

## V. Results

### A. Model Performance

Our model achieved a final test accuracy of 38.19% on the multimodal intention recognition task. During training, the model demonstrated strong performance on the training data, reaching 79.83% accuracy, while validation accuracy stabilized at 39.24%. The final training loss was 0.6625, with a validation loss of 2.6649.

### B. Dataset Characteristics and Significance

The results must be contextualized within the pioneering nature of our approach. Our dataset represents a significant advancement in intention recognition research through several key innovations:

*1) Multimodal Integration:* Our work is among the first to combine visual and verbal intention signals in a unified framework, capturing the complex interplay between non-verbal cues and linguistic expressions. This multimodal approach provides a more comprehensive representation of human intention compared to traditional unimodal approaches.

*2) Randomized EEG Signal Processing:* We employed randomized EEG signal processing techniques to extract robust neurophysiological correlates of intention formation. This approach allows us to:

- Identify patterns in neural activity that precede conscious intention manifestation
- Reduce contamination from task-irrelevant neural activity
- Capture intention signals across diverse cognitive states

### C. Performance Analysis

The gap between training and testing performance (79.83% vs. 38.19%) indicates the challenging nature of cross-subject intention recognition. This performance differential is consistent with prior work in this domain and reflects the inherent variability in how intentions manifest across individuals.

### D. Comparative Context

Our achieved accuracy of 38.19% represents a substantial improvement over chance performance (which would be significantly lower given the multi-class nature of our intention recognition task). Furthermore, this performance is noteworthy given:

- The unprecedented diversity of our dataset, incorporating intention signals across varied contexts
- The challenging nature of integrating EEG signals with visual and verbal modalities
- The fine-grained intention classification scheme employed in our study

### E. Limitations and Future Directions

While our model demonstrates promising performance on this novel multimodal intention recognition task, several avenues remain for further improvement. The disparity between training and testing performance suggests *potential for enhanced cross-subject generalization* through architectural refinements or expanded data collection strategies.

The current performance establishes an important baseline for future work in multimodal intention recognition, particularly approaches incorporating neurophysiological signals alongside behavioral manifestations of intention.

We will also focus on an end-to-end architecture that maps EEG signals directly to log-mel spectrograms, followed by waveform synthesis using a neural vocoder. This direct-to-audio path aims to integrate the vocal expression of intention, thereby leveraging both cognitive and articulatory information embedded in EEG responses during verbal tasks.

## References


[1] Y. Deng, S. Ding, W. Li, Q. Lai, and L. Cao, *EEG-based visual stimuli classification via reusable LSTM*, 2020.

[2] H. Ahmed, R. B. Wilbur, H. M. Bharadwaj, and J. M. Siskind, *Object classification from randomized EEG trials*, IEEE, 2020.

[3] Y. Song, B. Liu, X. Li, N. Shi, Y. Wang, and X. Gao, *Decoding Natural Images from EEG for Object Recognition*, Tsinghua University and Institute of Semiconductors, CAS, 2020.

[4] S. Zhu, Z. Ye, Q. Ai, and Y. Liu, *EEG-ImageNet: An Electroencephalogram Dataset and Benchmarks with Image Visual Stimuli of Multi-Granularity Labels*, 2021.

[5] C. Spampinato, S. Palazzo, I. Kavasidis, D. Giordano, N. Souly, and M. Shah, *Deep Learning Human Mind for Automated Visual Classification*, PeRCeiVe Lab, University of Catania and CRCV, University of Central Florida, 2017.