A Appendix: Supplemental materials

A.1 Hierarchical scheme

In this paper, we took advantage of simple feedforward hierarchical filtering of perceived images to use for joint learning with EEG features. We observed that heuristically, the simple four layers of filtering without feedback work best on the limited benchmarking datasets. The four hierarchical filters are as follows (and visualized through Fig. A1 to Fig. A3):

- **V1 (Edge detection)**: Sobel filtering captures the role of V1 in detecting simple edges and orientation. The Sobel filter computes the gradient magnitude of an image to detect edges. The gradient magnitude is given by:

  $$G = \sqrt{(S_x^2 + S_y^2)}$$

  where:
  - S_x is the gradient in the x direction.
  - S_y is the gradient in the y direction.

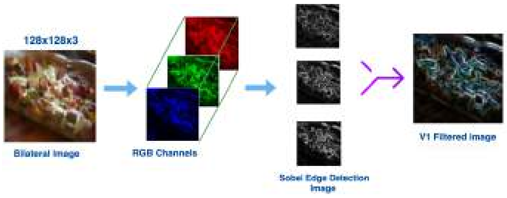

Figure A1: V1 filtering process

- **V2 (Texture and contour detection)**: Local Binary Patterns (LBPs) and contour detection simulate V2's role in recognizing textures, contours, and boundary details, processing the information from the prior layers. The LBP operator describes the local texture of an image by comparing each pixel with its neighbors. Contours represent the boundaries and edges in an image, being one of the most important features required to identify objects and shapes.

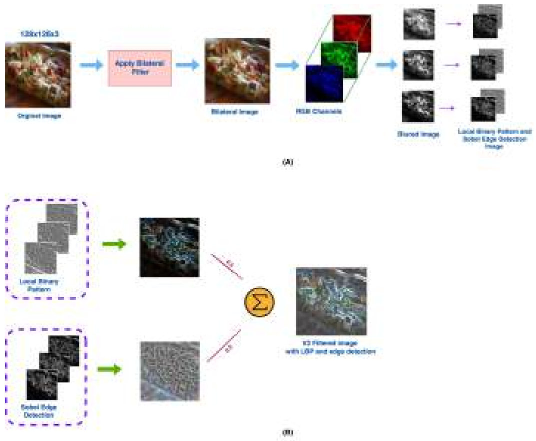

Figure A2: V2 filtering process

- **V3 (Motion and color processing)**: Motion detection and color saturation filtering reflect V3's role in processing dynamic information and complex color features, processing the information of the perceived sight. The HSV (Hue, Saturation, Value) color model separates color information into three channels. The saturation channel is extracted as:

  $$S = \text{Saturation}(H, S, V)$$

where:

- H is the Hue channel.
- S is the Saturation channel.
- V is the Value channel.

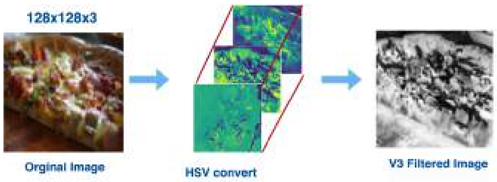

Figure A3: V3 filtering process

- **V4 (Curvature and shape detection)**: V4 is characterized as a mid-tier cortical area in the ventral visual pathway, positioned between earlier visual areas like V1/V2 and higher-level areas in the inferotemporal cortex. Hough circles and shape detection simulate V4's involvement in higher-level shape and form recognition, which is crucial for object identification. The outcomes of V4 were shown heuristically to be redundant in terms of the jointly learned feature space of the original image. Therefore, we used V1-V3, added to the original image, for feature extraction from the image.

A.2 MANOVA Analysis

The results of the MANOVA test can be seen in Table B1. The significance of the difference is presented as F Value and probabilities. All 4 tests yield high indication of significance. The Wilks' Lambda for the group is nearly zero. This indicates that nearly all the variability in the EEG features can be explained by group differences, suggesting a very strong separation between the models.
The Pillai's Trace is also high (2.9960), which confirms that a substantial amount of variance in the features is explained by the group effect. Hotelling-Lawley and Roy's Greatest Root are very high. This supports the idea that the groups are distinct.

Test Statistic	Value	Num DF	Den DF	F Value	Pr > F
Intercept					
Wilks' Lambda	0.0016 ↓	128	4477	22338.16 ↑	0.00001 ↓
Pillai's Trace	0.9984 ↑	128	4477	22338.16 ↑	0.00001 ↓
Hotelling-Lawley	638.6608 ↑	128	4477	22338.16 ↑	0.00001 ↓
Roy's Root	638.6608 ↑	128	4477	22338.16 ↑	0.00001 ↓
V1-V3 and original					
Wilks' Lambda	0.00001 ↓	384	13430.15	27048.62 ↑	0.00001 ↓
Pillai's Trace	2.9960 ↑	384	13437	26053.44 ↑	0.00001 ↓
Hotelling-Lawley	2404.6657 ↑	384	13078.52	28027.41 ↑	0.00001 ↓
Roy's Root	1073.1941 ↑	128	4479	37553.41 ↑	0.00001 ↓

Table B1: Multivariate Linear Model Results for EEG Features

A.3 ThoughtViz embedding space

The ThoughtViz dataset was analyzed to examine the embedding space generated by t-SNE (t-Distributed Stochastic Neighbor Embedding), a widely used dimensionality reduction technique. The t-SNE algorithm maps high-dimensional EEG features to a lower-dimensional space (in this case, two dimensions) while preserving the local structure of the data.

The resulting visualization, shown in Figure A4, provides insights into how well the models differentiate the EEG data. Each point in the plot represents a sample in the dataset, and the spatial arrangement reflects the similarity between samples: points that are closer together indicate higher similarity, while points farther apart signify greater dissimilarity.

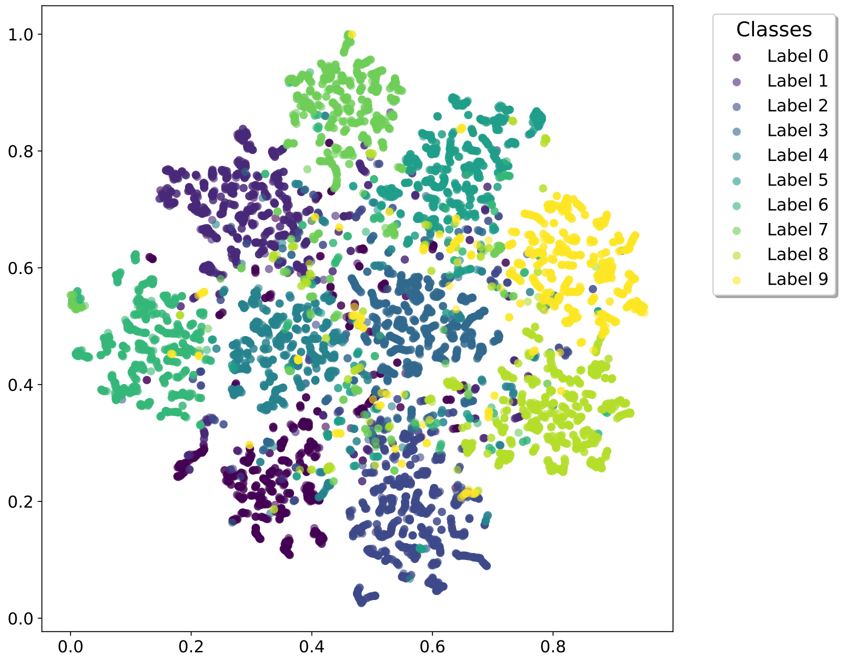

Figure A4: The t-SNE plot for ThoughtViz dataset. The figure illustrates the first two dimensions of the t-SNE map (horizontal as the first dimension). The clustering indicates clear separation of EEG feature groups, highlighting the efficacy of the proposed hierarchical scheme.