ADVANCEMENTS IN MACHINE LEARNING PARADIGMS: BRIDGING TRADITIONAL ML, ZERO-SHOT LEARNING, MULTI-AGENT RULE GENERATION, AND DEEP LEARNING

A PREPRINT

ABSTRACT

This report explores the evolution of machine learning (ML) from traditional methods to advanced paradigms that push the boundaries of artificial intelligence capabilities. Traditional ML techniques, while foundational, often face limitations in scalability, generalisation, and reliance on annotated data. Advanced paradigms such as zero-shot learning (ZSL) and multi-agent rule generation extend ML's reach by enabling models to generalise to unseen classes and autonomously generate complex decision rules within dynamic environments. Deep learning (DL) further enhances these paradigms through hierarchical representation learning, supporting more adaptable and expressive models.

1 Introduction

Machine learning has historically operated within the confines of the traditional paradigm, where models learn predictive patterns from sufficiently large and representative labelled datasets. Supervised and unsupervised methods excel when task distributions remain stable and all classes or behaviours are observed during training. Yet modern AI problems—from semantic retrieval and autonomous agents to dynamic game environments—often violate these assumptions. Systems must recognise unseen categories, reason from sparse information, or adapt to changing rules in complex multi-agent environments. Such demands expose the rigidity of conventional ML pipelines and motivate the development of advanced machine learning paradigms. One major advancement is zero-shot learning (ZSL), which enables a model to make predictions for classes without any visual training examples. [1] demonstrated a key breakthrough by synthesising unseen visual features using semantic attributes and diffusion regularisation, allowing ZSL to be reframed as a conventional supervised task on generated data. This synthesis-based approach highlights a shift in ML: instead of learning only from available data, models can imagine or simulate data for unseen classes. A related progression is zero-shot retrieval, which seeks to retrieve instances based on a query described only by a set of dominant or semantic attributes. The Zero-Shot Retrieval via Dominant Attributes framework [2] shows how models can infer latent instance-level attributes from coarse semantic embeddings, enabling flexible instance retrieval even in large, unlabelled databases. These approaches collectively bypass the limitations of dataset coverage and reduce reliance on costly human annotations. Beyond single-model reasoning, ML continues to evolve toward multi-agent and rule-generating systems.[3] introduced a paradigm where rules governing an environment are generated, evaluated, and evolved through multi-agent reinforcement learning, game-theoretic modelling, and generative architectures. By treating rules themselves as learnable entities, this framework extends ML far beyond fixed-task settings, enabling agents to reason about, adapt to, and even design the governing dynamics of their environment. This represents a profound expansion of ML's capabilities—from predicting outcomes within rules to learning the rules themselves. Parallel to these conceptual advancements, deep learning continues to push the boundaries of representational learning. Deep architectures such as transformers provide strong global reasoning abilities but often lack local inductive biases. Duan et al. (2023) address this gap by proposing Dynamic Unary Convolution, which enhances transformers with adaptive, task-aware convolutional operations. By capturing local and mid-level features more effectively, such models improve performance in dense vision tasks and highlight how deep learning architectures evolve to support the flexibility required by advanced ML paradigms. Together, these developments show that advanced ML lies at the intersection of semantic reasoning, generative modelling, multi-agent dynamics, and adaptive deep architectures. They bridge traditional ML with modern DL-based systems, enabling new forms of generalisation and more robust real-world applicability.

2 Data and Paradigm

2.1 Dataset Overview and Exploration

This work uses the *ThingsEEG-Text* dataset, which contains 561-dimensional EEG feature vectors recorded while participants viewed object images drawn from the THINGS database. Experiments focus on a single subject (S10) using a predefined category-level split comprising 1,654 seen and 200 unseen object categories, each with 10 trials, resulting in 16,540 seen and 16,000 unseen samples. The dataset is high-dimensional, noisy EEG signals and an extreme low-shot regime, with only 10 trials per category. Under these conditions, normal supervised classification is highly prone to overfitting and poor generalisation. Further analysis indicates weak natural alignment between EEG representations and visual semantic structure, motivating a formulation that prioritises semantic generalisation over direct label prediction.

2.2 Data Splitting Strategy and Paradigm Design

To address these challenges, the task is formulated as a *category-level zero-shot learning* problem in which object categories are partitioned into disjoint seen and unseen sets. Formally, let Y_{seen} and Y_{unseen} denote the sets of seen and unseen categories, respectively, with $Y_{\text{seen}} \cap Y_{\text{unseen}} = \varnothing$. Given EEG samples $\mathbf{x}i \in \mathbb{R}^{561}$, all data associated with categories in Yseen are used to learn a mapping $f : \mathbb{R}^{561} \to \mathbb{R}^{d}$ into a semantic embedding space, while samples from Y_{unseen} are reserved exclusively for zero-shot evaluation. Within the seen categories, samples are randomly split into training and validation sets using an 80/20 ratio at the sample level, preserving the same category set in both splits. Crucially, no information from unseen categories is used during training or model selection to make sure it is a strict zero-shot evaluation setting.

Table 1: Zero-Shot Data Splitting Strategy (ThingsEEG-Text)

Split	Category Type	Number of Categories	Samples per Category	Total EEG Samples
Training	Seen categories	1,654	∼8	13,232
Validation	Seen categories	1,654	∼2	3,308
Test (Zero-Shot)	Unseen categories	200	80	16,000

2.3 Paradigm Motivation, Workflow, and Real-World Value

The adopted paradigm frames EEG decoding as a zero-shot semantic classification problem, in which models learn to map EEG signals into a semantic embedding space rather than directly predicting class labels. This formulation enables inference over object categories that are never observed during training, addressing the practical limitation that collecting labelled EEG data for every class is costly, time-consuming, and subject to fatigue. Figure 1 provides an overview of the zero-shot workflow and its real-world brain–computer interface application. Exploratory data visualisation using cross-modal similarity heatmaps (Figure 2) reveals weak natural alignment between EEG and image representations, highlighting a fundamental semantic mismatch that motivates the need for multimodal learning. By integrating image-based semantic embeddings, the paradigm enables semantic transfer across modalities and subjects, improving robustness and scalability beyond subject- or class-specific memorisation. From a real-world perspective, this paradigm reduces annotation cost by avoiding exhaustive EEG data collection while supporting generalisation to unseen concepts, making it well suited for brain–computer interface systems operating under limited data availability and evolving task requirements.

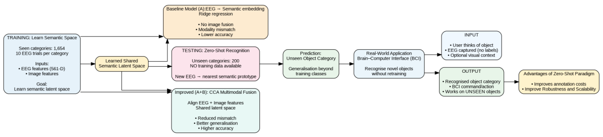
TRAINING: Learn Semantic Space
Seen categories: 1,654
10 EEG trials per category
Inputs:
• EEG features (561-D)
• Image features
Goal:
Learn semantic latent space
Learned Shared Semantic Latent Space
Baseline Model (A):EEG → Semantic embedding
Ridge regression
• No image fusion
• Modality mismatch
• Lower accuracy
TESTING: Zero-Shot Recognition
Unseen categories: 200
NO training data available
New EEG → nearest semantic prototype
Improved (A+B): CCA Multimodal Fusion
Align EEG + Image features
Shared latent space
• Reduced mismatch
• Better generalisation
• Higher accuracy
Prediction:
Unseen Object Category
Generalisation beyond training classes
Real-World Application
Brain–Computer Interface (BCI)
Recognise novel objects without retraining
INPUT
• User thinks of object
• EEG captured (no labels)
• Optional visual context
OUTPUT
• Recognised object category
• BCI command/action
• Works on UNSEEN objects
Advantages of Zero-Shot Paradigm
• Improves annotation costs
• Improve Robustness and Scalability

Figure 1: Overview of zero-shot paradigm workflow and real world application

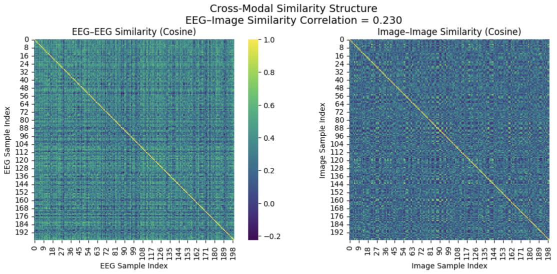

Figure 2: Cross-modal similarity heatmap illustrating within-modality semantic structure.

3 Model Development

3.1 Learning Objective and Model Specification

Under the zero-shot learning paradigm defined previously, this section specifies the concrete learning objective and model components. Let $\mathcal{D}_{\text{seen}} = \{(\mathbf{x}_i^{\text{EEG}}, \mathbf{x}_i^{\text{img}}, \mathbf{x}_i^{\text{txt}}, y_i)\}_{i=1}^{N_s}$ denote the training dataset with EEG signals, image features, text embeddings, and labels from seen categories. The zero-shot test set $\mathcal{D}_{\text{unseen}} = \{(\mathbf{x}_j^{\text{EEG}}, \mathbf{x}_j^{\text{img}}, y_j)\}_{j=1}^{N_u}$ contains disjoint unseen categories. The task is to learn a mapping $f_\theta : \mathbb{R}^{d_{\text{EEG}}} \to \mathbb{R}^{d_s}$ that projects EEG into semantic space. Classification compares predicted embeddings to class prototypes using accuracy, precision, recall, and macro F1-score (evaluated on unseen categories only), explicitly measuring semantic generalisation rather than memorisation.

3.2 Baseline Model (A): Ridge Regression with Image Prototypes

Learning objective. The baseline model learns a linear mapping from EEG features to semantic space via ridge regression. EEG features are first standardised and reduced via PCA:

$$\tilde{\mathbf{x}}^{\text{EEG}} = \text{PCA}(\text{Standardise}(\mathbf{x}^{\text{EEG}})). \tag{1}$$

A linear ridge regression model maps reduced features to semantic embeddings by minimising

$$\mathcal{L}(\mathbf{W}) = \sum_i \left\|\mathbf{W}\tilde{\mathbf{x}}_i^{\text{EEG}} - \mathbf{x}_i^{\text{txt}}\right\|_2^2 + \lambda\|\mathbf{W}\|_2^2. \tag{2}$$

Ridge regression is selected for its closed-form solution, numerical stability in high-dimensional settings, and inherent regularisation through the ℓ_2 penalty $\lambda\|\mathbf{W}\|_2^2$, which provides robustness under limited per-class supervision.

Zero-shot inference. At test time, class prototypes are computed from image embeddings for unseen categories (preserving the zero-shot constraint):

$$\mathbf{p}_c = \frac{1}{|c|}\sum_{y_i=c} \mathbf{x}_i^{\text{img}}. \tag{3}$$

Given a test EEG sample, ridge regression predicts an embedding $\hat{\mathbf{z}} = \mathbf{W}\tilde{\mathbf{x}}^{\text{EEG}}$. Classification is performed via nearest-prototype matching using cosine similarity:

$$\hat{y} = \arg\max_c \cos(\hat{\mathbf{z}}, \mathbf{p}_c). \tag{4}$$

This baseline establishes a reference point for semantic transfer using EEG alone, without cross-modal alignment.

3.3 Methodological Improvement (B): CCA-Based Multimodal Fusion

Motivation and theoretical hypothesis. Data exploration revealed weak alignment between EEG and image modalities (cross-modal correlation $r \approx 0.23$). This suggests that EEG features encode neural responses in a space fundamentally misaligned with visual-semantic structure. Standard supervised approaches and direct semantic regression both suffer from this modality gap: EEG representations, while sensitive to object categories, do not naturally project into visual semantic space. We hypothesise that learning a shared latent space via Canonical Correlation Analysis (CCA) addresses this fundamental misalignment. CCA discovers linear projections that maximise cross-modal correlation, creating a representation space where EEG and image features are semantically aligned. Theoretically, in this aligned space, EEG embeddings should exhibit higher cosine similarity to true class prototypes derived from images, improving zero-shot accuracy beyond the baseline.

Improved model: CCA-based fusion. The improved model implements multimodal alignment through CCA. Both modalities are independently preprocessed: EEG and image features are standardised and reduced via PCA to the same dimensionality:

$$\tilde{\mathbf{x}}^{\text{EEG}} = \text{PCA}_{d_p}(\text{Standardise}(\mathbf{x}^{\text{EEG}})), \quad \tilde{\mathbf{x}}^{\text{img}} = \text{PCA}_{d_p}(\text{Standardise}(\mathbf{x}^{\text{img}})). \tag{5}$$

CCA then learns linear projection matrices $\mathbf{U}$ and $\mathbf{V}$ that maximise canonical correlation between the two modalities:

$$\max_{\mathbf{U},\mathbf{V}} \text{ corr}\left(\mathbf{U}^\top\tilde{\mathbf{X}}^{\text{EEG}}, \mathbf{V}^\top\tilde{\mathbf{X}}^{\text{img}}\right). \tag{6}$$

The resulting projections define a shared latent space where d_c canonical variates (with $d_c \leq d_p$) explain the maximum correlation between modalities. This space represents a compromise representation capturing shared semantics while suppressing modality-specific noise.

Zero-shot inference in aligned space. At test time, EEG samples from unseen categories are projected into the learned CCA space:

$$\mathbf{z}_{\text{test}}^{\text{EEG}} = \mathbf{U}^\top \tilde{\mathbf{x}}_{\text{test}}^{\text{EEG}}. \tag{7}$$

Image-based prototypes for unseen classes are similarly projected:

$$\mathbf{p}_c^{\text{CCA}} = \frac{1}{|c|} \sum_{y_i = c} \mathbf{V}^\top \tilde{\mathbf{x}}_i^{\text{img}}. \tag{8}$$

Final predictions are obtained via cosine similarity in the aligned space:

$$\hat{y} = \arg\max_c \cos(\mathbf{z}_{\text{test}}^{\text{EEG}}, \mathbf{p}_c^{\text{CCA}}). \tag{9}$$

Critically, both Model A and Model A+B use identical image-derived prototypes; the methodological improvement stems entirely from CCA-based alignment. This design isolates the contribution of multimodal fusion to zero-shot performance.

3.4 Training, Hyperparameter Selection, and Regularisation

Model capacity is controlled via PCA dimensionality $d_p \in \{50, 100\}$ and CCA dimensionality $d_c \in \{20, 40, 60\}$ (with $d_c \leq d_p$). Grid search on the validation split selects hyperparameters using macro F1-score to reduce class imbalance sensitivity. Best hyperparameters are used for final test-set evaluation, avoiding information leakage.

Both PCA and CCA admit closed-form solutions; convergence is guaranteed by linearity. Overfitting is mitigated through dimensionality reduction suppressing noise, CCA's correlation-maximising objective suppressing modality-specific variation, and constrained latent dimensionality enforcing a bottleneck. Validation performance plateaus when dimensionality exceeds the tuned range, confirming regularisation prevents overfitting.

3.5 Summary

This section compares baseline Model A (ridge regression) with improved Model A+B (CCA-aligned multimodal fusion), both using identical image-derived prototypes. By modifying only the alignment mechanism, this isolates the contribution of cross-modal fusion to zero-shot generalisation.

4 Result Analysis

The ablation results in Table 2 demonstrate a clear performance gap between the EEG-only baseline and the proposed CCA-based multimodal model, highlighting the limitations of semantic prototype matching without explicit cross-modal alignment. The baseline achieves near-random accuracy and macro F1-score, confirming that EEG representations alone provide insufficient semantic structure for reliable zero-shot generalisation to unseen categories. Introducing CCA-based fusion consistently improves accuracy, precision, recall, and macro F1-score by aligning EEG features with image-derived semantic prototypes in a shared latent space, thereby enabling more discriminative zero-shot inference. Performance varies systematically with representation dimensionality: low PCA dimensions underperform due to limited representational capacity, while overly large CCA dimensions degrade performance, suggesting overfitting and noise amplification. The optimal configuration (PCA = 100, CCA = 20) reflects a balance between expressive power and regularisation, aligning with the theoretical expectation that effective zero-shot learning depends on robust semantic alignment rather than model complexity. The substantially higher final test performance relative to validation results further indicates that the learned alignment generalises beyond model selection, supporting the robustness and reliability of the proposed approach of CCA-based multimodal zero-shot learning framework.

Table 2: Ablation Study Zero-shot EEG decoding performance for the baseline model (A) and the CCA-based multimodal model (A+B) under different PCA and CCA dimensionalities.

Model	PCA dim	CCA dim	Accuracy	Precision	Recall	Macro F1
Baseline (A)	100	–	0.0111	0.0026	0.0111	0.0024
A+B (CCA)	50	20	0.0091	0.0070	0.0102	0.0074
A+B (CCA)	50	40	0.0070	0.0067	0.0071	0.0059
A+B (CCA)	100	20	0.0553	0.0509	0.0695	0.0514
A+B (CCA)	100	40	0.0426	0.0419	0.0527	0.0410
A+B (CCA)	100	60	0.0287	0.0281	0.0370	0.0275
A+B (CCA), Final Test	100	20	**0.0811**	**0.0830**	**0.0811**	**0.0794**

Figures 3 and 4 compare zero-shot classification behaviour on 20 unseen classes for the EEG-only baseline model (A) and the proposed CCA-based multimodal model (A+B). The baseline confusion matrix exhibits highly concentrated predictions on a small subset of classes with little diagonal structure, indicating poor semantic discrimination and confirming the theoretical limitation that EEG-only representations lack sufficient structure for zero-shot generalisation. In contrast, the A+B model displays a markedly stronger diagonal pattern with more evenly distributed correct predictions, demonstrating that aligning EEG features with image-derived semantic representations in a shared latent space improves class separability for unseen categories. Residual off-diagonal errors reflect the inherent difficulty of zero-shot EEG decoding and imperfect cross-modal alignment rather than random behaviour. The qualitative differences between the two confusion matrices closely mirror the quantitative improvements observed in accuracy, precision, recall, and macro-F1 in the ablation study, supporting the robustness and reliability of the results. Together, these findings validate the proposed hypothesis that zero-shot EEG decoding performance is primarily limited by semantic misalignment, and that CCA-based multimodal fusion effectively addresses this limitation.

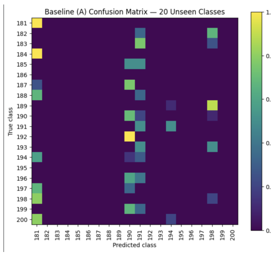

Figure 3: Baseline model (A): confusion matrix for 20 unseen categories.

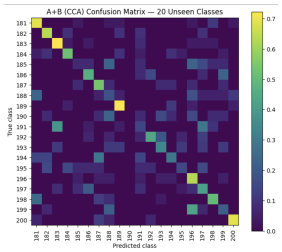

Figure 4: Improved model (A+B): CCA-based multimodal fusion for 20 unseen categories.

Figure 5 presents a box plot–based failure analysis of the proposed CCA-based multimodal model (A+B) by comparing cosine similarity to the true semantic prototype for correctly and incorrectly classified EEG samples in the shared latent space. The box plot shows that correct predictions are associated with consistently higher similarity values and a tighter interquartile range, indicating strong cross-modal alignment between EEG representations and their corresponding image-derived semantic prototypes. In contrast, incorrectly classified samples exhibit a lower median similarity and substantially greater variance, including negative values, reflecting weak or incorrect semantic alignment. This highlights a key limitation of the proposed model which is although CCA improves alignment on average, it does not guarantee robust semantic alignment for all EEG samples, particularly in the presence of noisy or ambiguous neural signals. Overall, the separation between success and failure cases provides causal insight into model behaviour, demonstrating that residual zero-shot classification errors arise from imperfect multimodal alignment rather than random noise or deficiencies in the similarity metric itself.

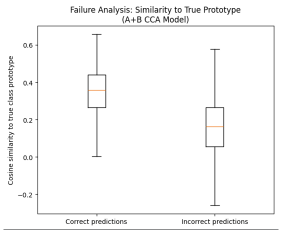

Figure 5: Failure analysis of the CCA-based model (A+B)

References

[1] Long, Y., Liu, L., Shen, F., Shao, L., & Li, X. (2017). Zero-shot learning using synthesised unseen visual data with diffusion regularisation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*

[2] Long, Y., Liu, L., Shen, Y., & Shao, L. (2018). Towards Affordable Semantic Searching: Zero-Shot Retrieval via Dominant Attributes. *Proceedings of the AAAI Conference on Artificial Intelligence*, 32(1).

[3] Pu, J., Duan, H., Zhao, J., & Long, Y. (2024). Rules for Expectation: Learning to Generate Rules via Social Environment Modeling. *IEEE Transactions on Circuits and Systems for Video Technology*.