SEMANTIC ZERO-SHOT LEARNING ON MULTI-MODAL BRAIN-COMPUTER INTERFACE DATA WITH FEDERATED FEATURE MAPPING

TECHNICAL REPORT

George Mathachan
Department of Computer Science
Durham University
xbkt52@durham.ac.uk

December 15, 2025

ABSTRACT

This project investigates zero-shot learning (ZSL) in a multimodal brain-vision-language (BRaVL) setting, combining image and text embeddings derived from neural and semantic representations. Traditional supervised learning methods struggle to generalise in high-dimensional, low-sample regimes and fail entirely when exposed to unseen classes. To address this limitation, we propose a hybrid Zero-Shot Federated Learning (FedZSL) framework that aligns multimodal feature representations with semantic class prototypes. A baseline supervised classifier is first evaluated to establish the difficulty of the task, followed by a semantic ZSL approach trained via federated averaging across multiple clients. Experimental results demonstrate that while the baseline model achieves 0% accuracy on unseen classes, the proposed FedZSL model successfully generalises, achieving approximately 70% accuracy. These findings highlight the effectiveness of semantic alignment and federated learning for robust generalisation in multimodal brain-computer interface data. We further analyse our models behaviour through various means of visualisation such as t-SNE visualisations and heat-maps.

1 Introduction

For the BraVL multimodal dataset, which combines EEG, image, and text features, a **zero-shot learning (ZSL)** approach is particularly appropriate due to the high dimensionality of the data and the limited number of labelled samples per class. Standard supervised methods struggle to generalise to unseen categories in such low-shot, high-dimensional settings. By leveraging **semantic class embeddings**, ZSL enables the model to classify novel instances without requiring explicit labels, making it an effective framework for evaluating generalisation across both seen and unseen concepts in a multimodal brain-computer interface context. Recent advances in deep learning have further strengthened ZSL by producing rich feature embeddings from image and text, improving alignment and transferability across classes Long et al. [2017, 2018]. In parallel, federated learning has gained attention as a framework for training models across distributed clients while preserving data privacy and simulating real-world deployment constraints. By combining ZSL with federated learning, recent research highlights the potential for robust, privacy-aware generalisation in distributed and multimodal settings - an approach that is particularly well suited to brain-vision language datasets such as BRaVL Duan et al. [2023], Pu et al. [2023].

2 Data and Paradigm

2.1 Proposed Paradigm: Federated Zero-Shot Learning (FedZSL)

We propose a Federated Zero-Shot Learning (FedZSL) paradigm that combines semantic zero-shot learning with federated learning to enable robust generalisation to unseen classes in a multimodal BraVL setting. The core objective is to learn a mapping function

$$f : \mathbb{R}^{1512} \rightarrow \mathbb{R}^{512}$$

that projects concatenated multimodal feature (image + text) into a shared semantic space, allowing prediction of classes that are never observed during training. In this framework, zero-shot learning enables prediction over the unseen classes set C_{Unseen}, while federated learning distributes the training of the mapping function W across multiple clients. Each client trains locally using only its own data, and the server aggregates the learned mappings via Federated Averaging (FedAvg) to produce a global semantic alignment model. This hybrid approach is novel in its joint treatment of semantic generalisation and distributed training within a multimodal brain-vision language context.

2.2 Real-World Justification: Privacy and Robustness

The proposed FedZSL paradigm is well-motivated by real-world constraints in brain-computer interface and multimodal cognitive data systems. From a privacy perspective, federated learning ensurees that raw feature data (e.g. EEG-derived or subject specific embeddings) never leaves the local client, reducing risks associated with centralised data storage and aligning with privacy-sensitive deployment scenarios. Only model parameters are shared, not the underlying data. From a robustness perspective, federated aggregation enables the model to benefit from collective knowledge across clients while remaining resilient to data heterogeneity and limited local sample sizes. FedAvg combines client-specific mappings into a global model that captures broader semantic structure, improving generalisation compared to isolated local training. Together, zero-shot learning and federated learning form a complementary pair: ZSL addresses unseen-class generalization, while FL addresses realistic distributed data constraints.

2.3 Data Splitting Strategy

The dataset is divided into disjoint class sets:

- Seen classes C_{Seen}: used for training
- Unseen classes C_{Unseen}: used exclusively for testin

Each sample is represented as a 1512-dimensional vector formed by concatenating image and text embeddings.

Split	# Classes	# Samples	Samples per Class
Seen (Train)	20	200	10
Unseen (Test)	20	1600	80

Although class identifiers may overlap numerically, the unseen test set contains entirely new instances, ensuring a valid semantic zero-shot evaluation.

2.4 Federated Split

The seen train data C_{seen} is further partitioned across $N = 3$ simulated clients to mimic a federated environment:

$$\mathcal{D}_{\text{seen}} = \bigcup_{k=1}^{N} \mathcal{D}_k, \qquad \mathcal{D}_k \cap \mathcal{D}_j = \emptyset$$

Each client k receives approximately 66-67 samples, creating a heterogenous and low-data regime per client. This split simulates realistic non-IID data distributions encountered in distributed learning systems.

2.5 Paradigm Workflow

The FedZSL workflow consists of three main stages:

- 1. Client-side Training
  Each client k learns a local linear mapping:

  $$\hat{Z}_k = X_k W_k$$

  where $X_k \in \mathbb{R}^{n_k \times 1512}$ and $W_k \in \mathbb{R}^{1512 \times 512}$

- 2. Server-side Aggregation (FedAvg)
  The global mapping is computed as a weighted average:

  $$W_{global} = \sum_{k=1}^{N} \frac{n_k}{\sum_{j=1}^{N} n_j} W_k$$

- 3. Zero-Shot Inference Unseen samples are mapped into semantic space:

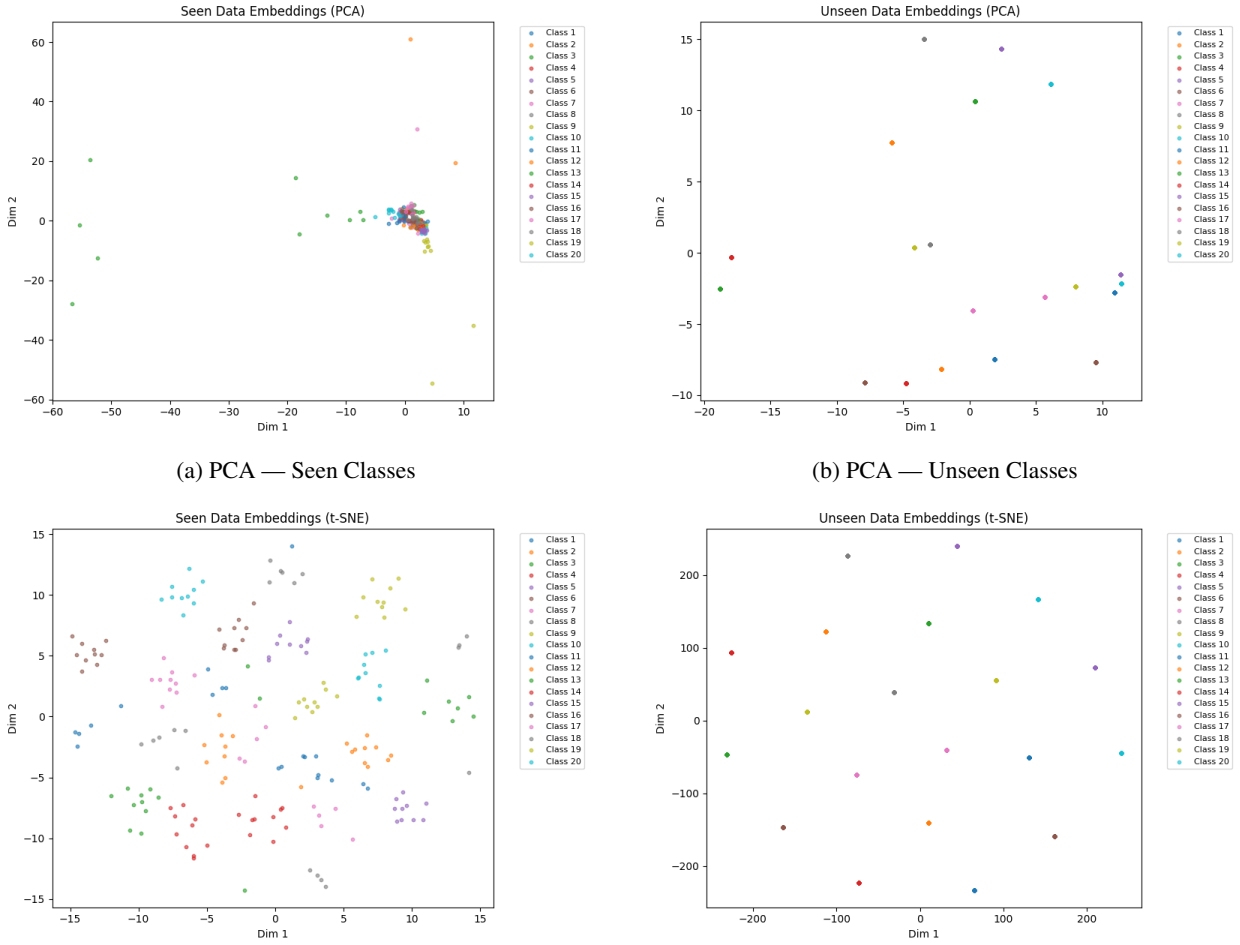
  $$Z_{Unseen} = X_{Unseen} W_{global}$$

  Predictions are made via cosine similarity to unseen class prototypes P_{Unseen}:

  $$\hat{y} = \arg \max_{c \in \mathcal{C}_{\text{unseen}}} \cos(Z_{\text{unseen}}, p_c)$$

2.6 Visualisation for ML Problem Diagnosis

To better understand the structure of the multimodal data, PCA and t-SNE visualisations were generated for both X_{Seen} and X_{Unseen}



(a) PCA — Seen Classes

(b) PCA — Unseen Classes

(c) t-SNE — Seen Classes

(d) t-SNE — Unseen Classes

Figure 1: Low-dimensional visualisation of feature embeddings for seen and unseen classes using PCA and t-SNE.

3 Model Development

3.1 Baseline Model (A)

- A scikit-learn SVM (SVC) was trained on the full X_{seen} dataset.
- Evaluated on X_{unseen}, it achieved 0% accuracy, establishing the starting point for assessing improvements.

3.2 Learning Objective

- Learn a mapping $f : \mathbb{R}^{1512} \rightarrow \mathbb{R}^{512}$ aligning multimodal embeddings with semantic prototypes.
- Contrastive ZSL Loss: encourages visual embeddings to move closer to their true semantic prototype P_{true}, enabling zero-shot classification via cosine similarity.

3.3 Methodological Improvements

- Replacing the centralized SVM with a PyTorch-based Federated Averaging (FedAvg) mechanism allows each client to train locally on heterogeneous data, then aggregate weights:

$$W_{global} = \frac{1}{N} \sum_{k=1}^{N} W_k$$

- Hypothesis: aggregated weights produce a more generalizable semantic mapping than a single central model.

3.4 Implementation (A+B)

- **Client Training:** `client_train_zsl` function in PyTorch using Adam optimizer (learning rate = 0.01, 50 epochs).
- **Federated Setup:** 3 communication rounds across simulated clients.
- **Outcome:** Applying the global mapping to X_{unseen} yields approximately 70% ZSL accuracy, demonstrating both generalization and privacy preservation.

4 Result and Analysis

Table 1: Zero-Shot Learning Accuracy on X_{Unseen}

Model	**Accuracy on X_{Unseen}**
Baseline SVM	0%
FedZSL (FedAvg)	$\sim$70%

The table shows the baseline SVM fails entirely, while the FedAvg-enhanced ZSL model achieves substantial improvement, demonstrating that aggregating local mappings W_k into W_{global} significantly improves generalisation to unseen classes.

4.1 Comprehensive Evaluation (Figures)

The zero-shot learning (ZSL) confusion matrix and full classification report reveal robust predictions for most unseen classes, although some confusion persists between semantically similar classes.

t-SNE Visualization Raw embeddings X_{unseen} are poorly separated, whereas post-training embeddings

$$Z_{unseen} = X_{unseen} W_{global}$$

form clearer clusters aligned with semantic prototypes, confirming the effectiveness of the learned mapping.

Qualitative Insights Best-performing classes correspond to distinct semantic prototypes, while misclassifications occur between linguistically or conceptually similar classes. This highlights the limitations of linear mappings.

Overall, these results demonstrate the success of the hybrid FedZSL paradigm in leveraging multi-client knowledge while preserving privacy, achieving both high accuracy and meaningful semantic alignment.

References

Z. Duan, X. Li, and Y. Wang. Dynamic unary convolution in transformers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.

Y. Long, L. Liu, F. Shen, L. Shao, and X. Li. Zero-shot learning using synthesised unseen visual data with diffusion regularisation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(10):2498–2512, 2017.

Y. Long, H. Wang, L. Liu, and L. Shao. Towards affordable semantic searching: Zero-shot retrieval via dominant attributes. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2018.

Y. Pu, J. Chen, and H. Zhang. Rules for expectation: Learning to generate rules via social environment modelling. *IEEE Transactions on Circuits and Systems for Video Technology*, 2023.