Inside Out ML: Emotion Classification from fMRI Using Parcellation and Projection Techniques

CS 6140 Machine Learning Spring 2025 Research Project

Ananda Anissa Vaughn Francis
Khoury College of Computer Sciences
Northeastern University
Boston, MA, USA
francis.ana@northeastern.edu

***Abstract*—This study explores feature extraction techniques for analyzing high-dimensional, auto-correlated data in multi-class classification tasks using functional magnetic resonance imaging (fMRI).Extending beyond the scope of CS 6140 Machine Learning, this investigation aims to establish a protocol for applying machine learning and deep learning methods to fMRI data by identifying meaningful brain parcels and data projections for emotion classification.**

***Keywords*—functional magnetic resonance imaging (fMRI), multi-class classification, feature extraction, emotion recognition**

I. Introduction

Recent years, researchers have investigated the ability of deep learning and machine learning models to analyze neuroimages. Before applying a model to a dataset, due to the high-dimensionality and density of neuroimaging data, practitioners often reduce dimensions using brain parcellation masks. These masks average voxel values over specific brain parcels, to focus analysis by regions of interest (ROI), rather than by voxel, which can go up to a 1 billion [9].

Given this development, studies have begun to explore best practices for choosing brain parcellations technique for feature extraction. The inspiration for our investigation came from a study that analyzed a teleological approach [10], recommending brain parcellation based on the aim of the study. To expand upon this work, this study investigated coefficients of a traditional machine learning model to identify important embeddings and regions of interest for emotion classification.

This paper bridges machine learning methodology with neuroscience application by evaluating how both statistical and domain-specific feature extraction techniques affect multi-class classification of emotional states from fMRI data. Our study results will recommend interdisciplinary researchers specific ROI and embeddings to focus on when studying emotion and mood in neurodegenerative, neurodevelopmental and psychiatric disorders.

Identify applicable funding agency here. If none, delete this.

II. Data

The dataset comprises 270 fMRI scans collected from 30 participants by Northeastern University's Affective and Brain Sciences Lab. Each subject was shown emotionally evocative images designed to elicit anger, fear, or disgust. This was followed by a priming word that was either congruent (same emotion as the image), incongruent (different emotion), or neutral (emotionally neutral).

Each subject completed all combinations of emotional state and priming condition (3 emotions × 3 priming types = 9 scans per subject), resulting in 270 scans in total. Each scan is stored as a .nii.gz file and labeled with subject ID, emotion state, and priming condition. The classification targets are emotional state and priming condition, while subject ID, used to model the hierarchical structure, introduces auto-correlation across samples from the same participant.

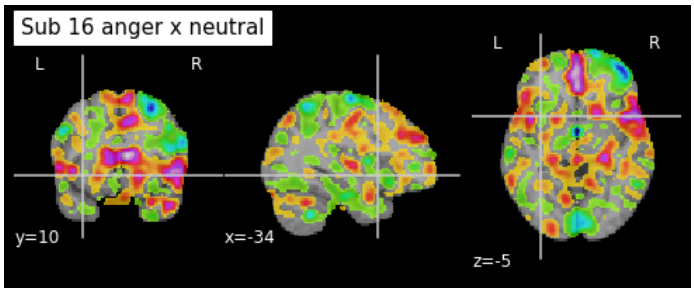

Fig. 1. Functional Magnetic Resonance Imaging (fMRI) scan of Subject 16 under neutral priming for angry emotion state

The dataset's size poses unique challenges for classification. The input data is 4D, with dimensions 270 × 91 × 109 × 91, resulting in 243 million total values (902,629 per scan). This is too large to feed directly into a classifier. Therefore, dimensionality reduction and brain parcellation techniques will be applied to reducing each scan for downstream modeling. In addition, due to size the dataset is stored in a Hugging Face dataset and loaded locally. In the future, this setup will be modified to support streaming, avoiding the need to store all 270 .nii.gz files locally.

III. Implementation

Initially, all algorithms were attempted from scratch, aligning with the course's emphasis on foundational understanding. However, given time constraints in project deadlines and computational runtime, focus switched to scikit-learn library implementations. Future efforts will involve revisiting these scratch implementations (included in the codebase) to identify how improved code structure and methodology might achieve comparable or superior results to open-source alternatives. To balance this shift, the project scope was broadened to dive deeper into the parcellation and projection techniques' nature and their real-world applicability in emotion analysis.

A. *Brain Parcellations*

The nature of neuroimaging data is not a tabular structure, but rather spatial and temporal dependencies to show where and when brain activity occurs. Each data point represents a specific voxel activation value which for fMRI can go up to 1 million features per image. Aside from high dimensionality, these raw images often contain noisy signals that are not related to blood oxygen level dependent (BOLD) signals, which reflect brain activity.

Brain parcellation uses predetermined masker objects or atlas maps to segment the high-dimensional voxel data into groups that represent specific regions of the brain. These regions of interest (ROI) are non-overlapping, and produce a structured signal (an average activation value for each ROI) that can be used as a feature and minimize noise [4]. All atlases were accessed using the nilearn library, which provides a standardized interface to widely-used functional neuroimaging templates. Each .nii.gz file was first loaded using Nibabel converted to voxel-level arrays, and transformed into 1D feature vectors using the NiftiLabelsMasker class in Nilearn. The corresponding atlas maps were retrieved via `nilearn.datasets.fetch_atlas_[ATLAS_NAME]`, and stored in a dictionary along with their labels for each ROI [4].

The atlas maps were chosen for 2 principles: granularity and determinism. Some atlas maps produced feature counts that exceeded the number of samples. These were ruled out. The second consideration was the atlas type. There are two types of atlases used for brain parcellation: probabilistic and deterministic. A deterministic atlas offers a definitive, non-overlapping segmentation of the brain. Each brain region in a deterministic atlas is represented as a distinct 3D image, where individual voxels are assigned to only one specific ROI [4]. For probabilistic atlases, the output data is represented by 4D images where each of the 3D components where individual values reflect the probability a voxel is related to that region [4]. Only deterministic parcellation maps were used to make interpretability seamless with clear feature separation.

1) ***Harvard Oxford:*** *:* The Harvard-Oxford atlases are probabilistic anatomical maps created from structural MRI data by the Harvard Center for Morphometric Analysis [4]. They segment both cortical and subcortical regions and are widely used in structural parcellation. The Cortical Atlas (48 ROIs) encompasses bilateral cortical structures at a threshold of 0 with a resolution of 1 mm. The Left Cortical Atlas (96 ROIs) targets cortical regions in the left hemisphere at the same threshold and resolution. The Subcortical Atlas (21 ROIs) includes key subcortical structures such as the thalamus, hippocampus, and amygdala, also at 1 mm resolution and 0 threshold.

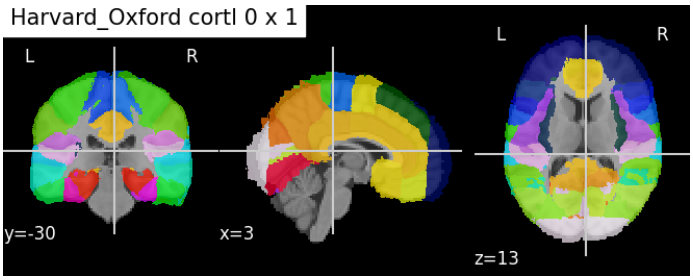

Fig. 2. Deterministic atlas map of Harvard Oxford Left Cortical Atlas at 3D coordinatates (3, -30, 13)

2) ***Talairach:*** : The Talairach atlas minimizes individual anatomical variations through a hierarchical system. Each map organizes anatomical labels using "volume and sub volumetric components," providing different levels of segmentation: hemispheres, lobes, gyri, tissues, and Brodmann Areas [7]. The Brodmann Areas variant segments the brain into 71 regions of interest (ROIs) based on Brodmann's cytoarchitectonic map [E]. The Gyrus Level variant provides 55 ROIs, segmenting the brain at the level of its cortical gyri. Due to these segmentations low-dimensionality, we combined Hemi-Lobe-Tissue feature set, totaling 22 ROIs: Hemisphere (7 ROIs), Lobe (12 ROIs), and Tissue type (3 ROIs).

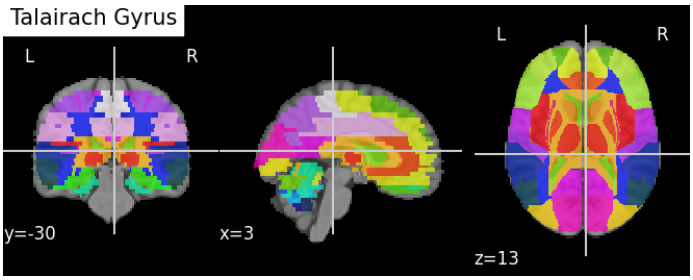

Fig. 3. Deterministic atlas map of Talaraich Gyrus at 3D coordinatates (3, -30, 13)

3) ***Schaefer:*** *:* The Schaefer atlas has 100 regions of interest (ROIs) within the cerebral cortex, organized into 17 large-scale networks, all at 1mm resolution. This atlas, derived from resting-state fMRI (rs-fMRI) data, uses a novel gradient-weighted Markov Random Field (gwMRF) model. This integrated approach detects abrupt transitions in functional connectivity patterns (local gradient) and clusters similar functional connectivity patterns regardless of spatial proximity (global similarity) to create homogeneous ROI [4].

4) ***AAL SPM12*** *:* The Automated Anatomical Labeling (AAL) atlas provides an automated anatomical parcellation

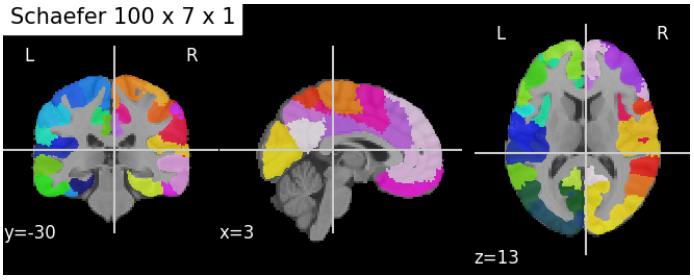

Fig. 4. Deterministic atlas map of Schaefer atlas (100 ROI, 17 yeo networks) at 3D coordinatates (3, -30, 13)

of a single-subject, spatially normalized T1-weighted brain volume from the Montreal Neurological Institute (MNI) [4]. AAL parcellates voxel data by assigning labels to specific points identified by coordinates, calculating the proportion of activated voxels within defined anatomical volumes of interest (AVOI) intersected by a sphere centered at given coordinates and determining the percentage of voxels in an activated cluster that fall within specific AVOIs [6].

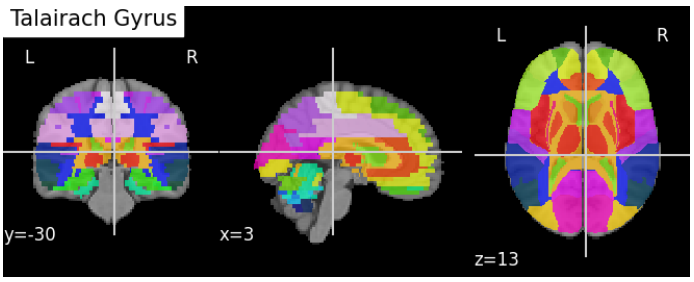

Fig. 5. Deterministic atlas map of Automated Anatomaical Labeling atlas at 3D coordinatates (3, -30, 13)

5) ***Juelich 0 x 1****:* The Juelich cytoarchitectonic atlas (juelich 0 x 1) creates brain maps based on cellular architecture with threshold of 0 and resolution mm of 1. These maps are derived from peer-reviewed probability maps that outline cortical and subcortical areas [4]. To create a parcellation, these individual probability maps are integrated into a maximum probability map. This is done by assigning each voxel to the cytoarchitectonic region with the highest probability, ensuring a complete and anatomically informed segmentation[5].

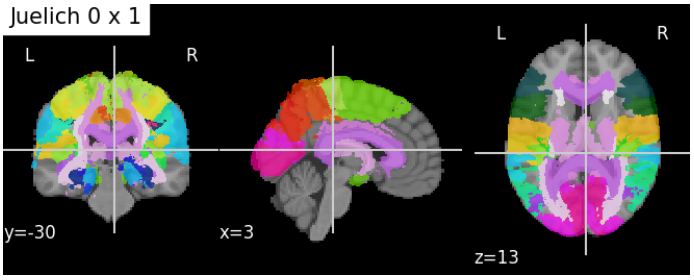

Fig. 6. Deterministic atlas map of Juelich cytoarchitectonic atlas at 3D coordinatates (3, -30, 13)

B. Geometric Projections

Before applying any projection/manifold techniques, the raw 3D neuroimaging data was flattened into a 1D vector. This 1D data was then standardized using StandardScaler() from scikit-learn, ensuring a mean of 0 and a standard deviation of 1.

Unlike brain parcellation maps, which inherently define a fixed number of features based on predefined brain regions, statistical dimensionality reduction methods require determining an optimal number of components. For methods like Principal Component Analysis (PCA), explained variance was used, while nonlinear manifold methods relied on cost functions [2].

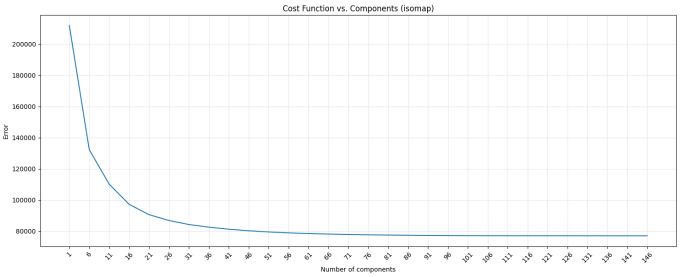

Fig. 7. Line plot showing reconstruction error rate of change as Isomap component features increases wiht plateau at approximately 50 components.

To find this optimal component count, the cost functions iteratively calculated the loss or error for various component numbers. For computational efficiency, error was evaluated for every fifth component count. These error values were visualized on a line plot to identify an ”elbow point”, where the loss stops decreasing significantly. Future work could automate this by calculating the derivative to find when the rate of change plateaus near zero.

The maximum feature count was set to 269 for iteration. A key constraint for dimensionality reduction, especially PCA states if the number of data samples (N) is less than the number of input features (D), the linear subspace dimensionality is at most N-1 [3]. In our case, with 270 data samples, the maximum threshold for features was therefore 269.

Modified Locally Linear Embedding, Autoencoder and Sammon Mapping were initially included in the investigation but were removed due to the computational inefficiency to run them (kernel died every iteration).

1) ***t-Distributed Stochastic Neighbor Embedding****:* T-distributed Stochastic Neighbor Embedding (t-SNE) is a data visualization technique that reduces high-dimensional data into a two or three-dimensional map. It works by converting data point similarities into probabilities, then minimizing the difference between these probabilities in both the original and lower-dimensional spaces [3]. T-SNE excels at revealing local data structure and clusters, making it useful for visualizing complex relationships. However, it’s computationally intensive, its results can vary due to a non-convex cost function, and it doesn’t explicitly preserve global data structure [11]. t-SNE is not typically used for feature extraction as its founders are

uncertain how it scales beyond 3D embedding [11]. TSNE() was imported from sci-kit learn's manifold package and initialized with 3 components, an auto learning rate, perplexity of 3.

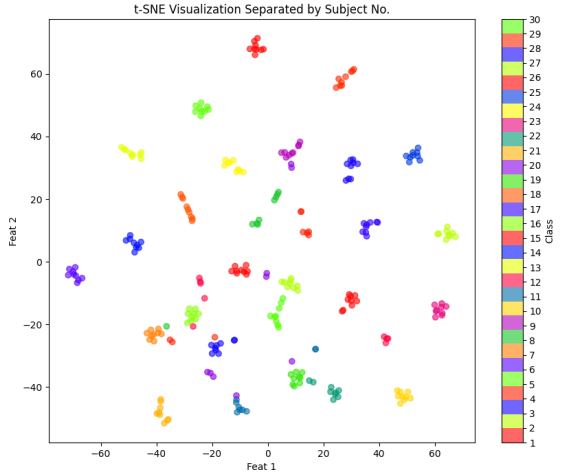

Fig. 8. 2D scatter plot for t-SNE embeddings, coloring the points based on their subject ID. Note: Autocorrelation in clustering

*2) **Hessian Eigen Mapping**:* Hessian Eigen Mapping (HLLE), uses a Hessian-based quadratic form to recover local linear structure: proposing a solution to LLE's regularization problem. A key constraint for HLLE is that the number of neighbors must be greater than $n \cdot \frac{n+3}{2}$, where n is the number of components. For our dataset of 270 instances, a scratch implementation of the quadratic formula revealed that the maximum components for HLLE is 21, requiring at least 262 neighbors (269 was used for this experiment, which can reduce variance of the estimation, and due to its proximity to 262, will likely not introduce bias). This method can be accessed in scikit-learn's LocallyLinearEmbedding by setting method='hessian'.

*3) **Principal Component Analysis**:* Principal Component Analysis (PCA), also known as the Karhunen Loeve transform [3] decomposes a dataset into orthogonal components (linear basis vector) that capture maximum variance and minimize the average reconstruction error. Scikit-learn implements PCA as a transformer in the decomposition package. Best practice dictates reducing the number of components to the minimum required to explain 99% of the total variance. In this specific case, 226 components were found to meet this criterion, as observed from the cumulative explained variance curve.

*4) **Isomap**:* Isomap, or Isometric Feature Mapping, uses MDS to reduce data dimensionality while preserving geodesic distances. Rather than use Euclidean distance metrics, Isomap measures distances along the curved "manifold" (theoretical data geometry where high-dimensional data truly lies) [3]. The process involves defining neighborhoods for each data point, constructing a graph where neighboring points are

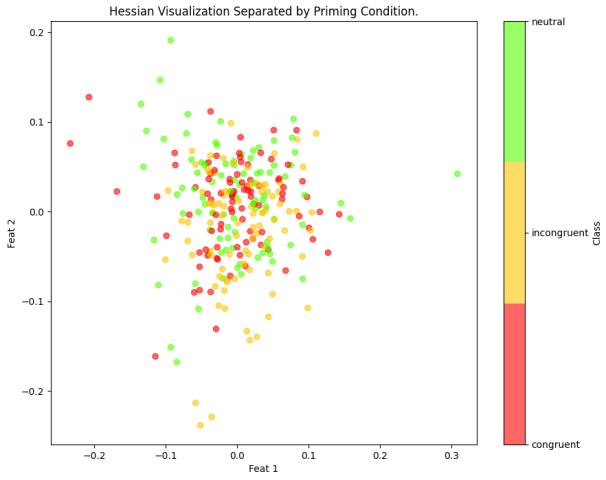

Fig. 9. 2D scatter plot for Hessian Eigenmapping embeddings, coloring the points based on their priming class.

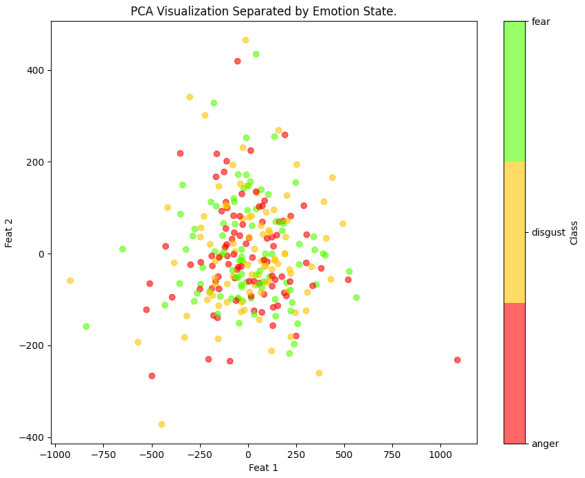

Fig. 10. 2D scatter plot for PCA embeddings, coloring the points based on their emotion class.

linked by their Euclidean distance, and then approximating geodesic distances as the shortest path along this graph [11]. Multidimensional Scaling (MDS) is applied to this geodesic distance matrix to generate the lower-dimensional projection. When the number of components surpassed 150, the distance matrix generated was not positive semi-definite (PSD) and the optimal component count was 51.

*5) **Multidimensional Scaling**:* Multidimensional Scaling (MDS) is a linear technique that aims to preserve the pairwise distances (or dissimilarities) from the original high-dimensional space as closely as possible. It works by finding the eigenvectors of the distance matrix. When the input dissimilarities are Euclidean distances, MDS yields results equivalent to PCA [3]. Although MDS can produce similar results to

Fig. 11. 2D scatter plot for Isomap embeddings, coloring the points based on their priming class.

PCA when using Euclidean dissimilarities, it was included in the analysis to explore potential variations. Additionally, its optimal component count was set to 50, determined by observing the elbow point in its cost function.

Fig. 12. 2D scatter plot for MDS embeddings, coloring the points based on their emotion class.

C. *Exploratory Data Analysis*

To better understand how various parcellation and projection methods perform feature extraction, this investigation delved into three key statistical aspects: multicollinearity, homogeneity of variance of features across target variables, and autocorrelation given that each subject provided multiple samples. Beyond this robust statistical protocol, each feature extraction method was also visually analyzed to provide further context.

*1) **Multicollinearity**:* Multicollinearity occurs when independent variables are highly correlated with each other. This inflates the standard errors of parameter estimates, making it difficult to determine the individual impact of each variable and potentially leading to confusing or misleading statistical results [1]. In this investigation, multicollinearity was assessed with heatmaps and Variance Inflation Factor (VIF) calculations. For methods producing fewer than 30 input features (specifically, "Harvard Oxford sub 0 x 1", "Talairach Hemi x Lobe x Tissue", "t-SNE", and "Hessian"), correlation matrices were visualized to identify strong correlations. To quantify multicollinearity, for each feature extraction method, the VIF value was calculated for all features, and sequentially dropped until the highest VIF until all remaining VIFs were below 5. The percentage of features dropped were stored and visualized in a bar plot for comparison. All collinear features remained as part of the analysis and future work aims to explore model performance with their removal.

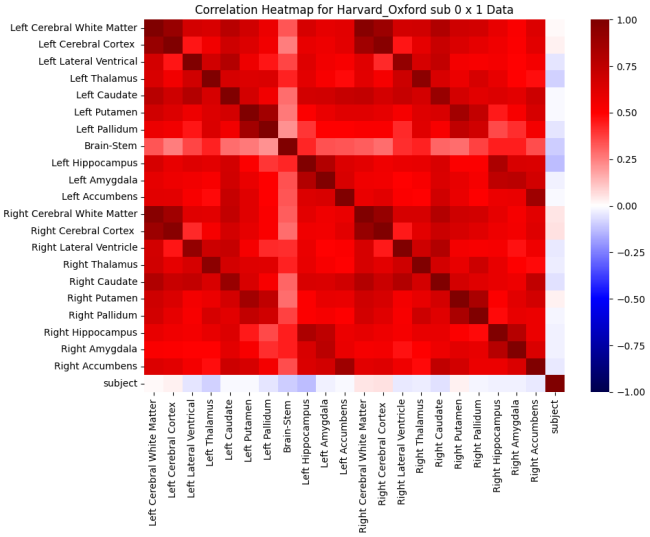

Fig. 13. Heatmap of Harvard Oxford Subcortical atlas displaying multicollinearity (dark red squares). Correlations almost solely positive (aside from subject ID)

*2) **Homogeneity of Variance**:* Homogeneity of variance (or homoscedasticity) states that the variability of observations should be roughly equal across all groups or levels of the independent variables. Heteroscedasticity can degrade model estimates and make it harder to detect true effects [1]. In this investigation, homogeneity of variance was assessed using the Levene test. This was applied to each feature from each feature extraction technique, examining its variance across different target variables: emotion and priming condition. The p values were stored and averaged out across each feature extraction method. For any methods with p values on average closer to alpha=0.05, multi-panel violin plots were used to further investigate the change in variance.

*3) **Autocorrelation**:* Autocorrelation occurs when observations within a dataset are not independent, meaning one observation can predict another. This violates a core assumption of

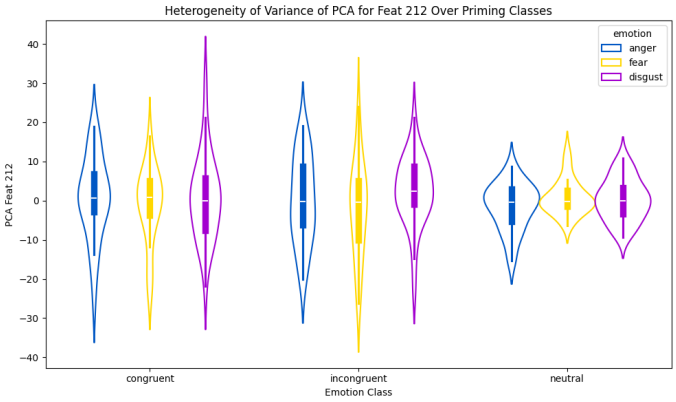

Fig. 14. Multi-panel violin plots for Principal Component 212, displaying heteroscedasticity across priming condition classes (Levene test $p-value \leq alpha-0.05$

most statistical techniques and can lead to incorrect analysis [1]. This investigation assesses autocorrelation by identifying which brain parcellation and dimensionality reduction methods produced features that were overly dependent on the subject number. For each feature extraction method all the correlation coefficients of all the features to the subject number was calculated and an average correlation was computed. These average correlations were then visualized in bar plots to clearly show which methods exhibited the highest degree of autocorrelation with the subject identifier, indicating potential issues for subsequent analyses.

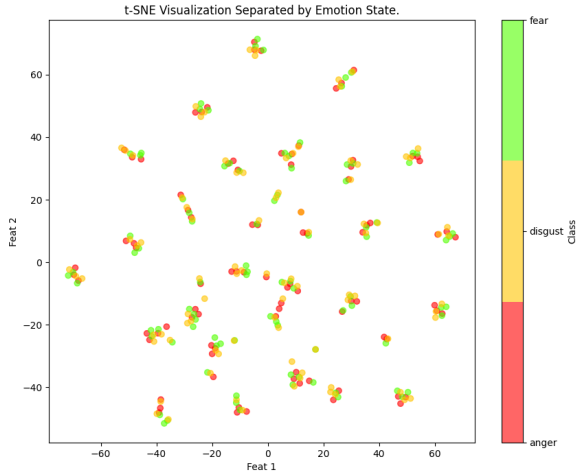

Fig. 15. 2D scatter plot for t-SNE embeddings, coloring the points based on their emotion state, yet clusters are non-separable. Geometric projection is prioritizing separability for subject ID not emotion class. Evidence of autocorrelation (refer to Figure 7).

D. *Multinomial Classification*

1) ***Hierarchical K-Fold Cross Validation****:* To ensure robust model training and prevent data leakage from autocorrelated samples (i.e., multiple data points from the same subject), Hierarchical K-Fold Cross Validation was employed. This approach guarantees that no subject's data appears in both the training and testing splits, thus providing a more reliable assessment of model generalization.

2) ***Logistic Regression****:* Logistic regression was used to quantify how a one-unit change in independent variables (features) affects the probability of a specific outcome. This method is particularly suitable for multiclass classification tasks, where the dependent variable represents discrete categories. For our logistic regression model, the sci-kit learn package LogisticRegression was initialized with penalty set to l2 regularization, dual set to False, max iteration set to 5000, and multi-class set to multinomial to handle our target variables. To optimize runtime across iterations, especially to mitigate kernel interruptions during training warm start is set to True. While time constraints prevented extensive hyperparameter tuning, these settings were chosen for practical efficiency. Model coefficients, reflecting feature importance, were captured for each fold, averaged, and stored for analysis.

3) ***Evaluation****:* The evaluation protocol was designed for comprehensive performance assessment and error analysis. Every misclassified instance was recorded to facilitate a detailed examination of residuals and understand specific failure modes. True positives and negatives were calculated per fold, aggregated, and then used to generate averaged confusion matrices for both training and testing splits for each feature extraction method. These were saved as visual heatmaps. Detailed classification reports, including recall, F1-score, accuracy, and precision were stored for the final fold. And finally, ROC AUC scores, aggregated across all folds, were computed and stored to provide an overall measure of model discriminative power. These evaluation steps, implemented through the HierarchicalKFoldCV class and supporting functions, ensure a thorough and robust assessment of the models trained on various parcellation and dimensionality reduction techniques.

4) ***Reproducibility****:* The trained model from each cross-validation fold was saved as a .pkl file to enable future reference and validation. All feature-extracted datasets, along with performance metrics across folds and configurations, were systematically stored to ensure reproducibility of the analysis.

IV. Results

A. *Multicollinearity*

For multicollinearity, none of the features produced by geometric projections had initial VIF values higher than 5. This is likely due to the fact that to be considered a linear basis vector, none of the other vectors can be expressed as a linear combination of the others. In geometric projection methods, the resulting components are specifically constructed to be orthogonal or linearly independent. This inherent design reduces multicollinearity, which is why the Variance Inflation Factor (VIF) values remain low and no features were collinear.

In contrast, the brain parcellation techniques demonstrated significantly higher multicollinearity, which is expected given

the nature of the brain and deterministic atlas maps. Voxel values are grouped into one ROI but likely have functional and structural interactions with other parcels outside the designated region. As a result, signal redundancy and overlap between ROIs emerge, particularly in anatomically or functionally similar areas.

When examining the heatmap, this is confirmed by strong, positive correlations between mirroring structures across hemispheres (e.g., Right Cerebral White Matter and Left Cerebral White Matter). These patterns reflect the brain's bilateral symmetry and interconnectivity, underscoring the challenges of applying linear modeling assumptions to parcellated brain data without first addressing collinearity. Researchers interested in brain parcellation techniques should be mindful of the statistical assumptions of the models being used.

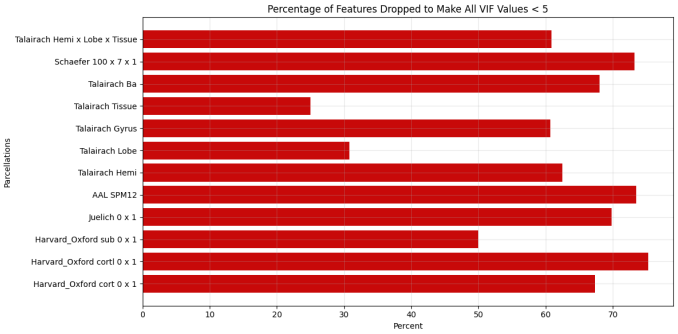

Fig. 16. Bar plot of percentages of features for brain parcellation techniques that initial VIF values were greater than 5

B. *Homogeneity of Variance*

For both parcellation and projection techniques, there was, on average, more heterogeneity of input features across priming classes than emotion classes. Among the parcellation methods, only Schaefer and Juelich exhibited any heterogeneity across emotion classes, and even then, it was minimal: less than 2% of features. In contrast, all parcellation techniques showed heterogeneity across priming classes. The Juelich atlas had the lowest at 12%, while the Talairach Hemisphere method showed the highest, with 42% of features demonstrating class-based variance. For projection techniques, PCA consistently exhibited the most heteroscedasticity across both target classes: 2.7% of features varied across emotion classes, and a significant 69% varied across priming classes. Isomap followed closely, with 1.95% and 40.5% feature heterogeneity for emotion and priming classes, respectively. MDS and Hessian Eigenmapping also showed some heteroscedasticity for priming classes, though to a lesser extent. Notably, t-SNE exhibited no heteroscedasticity across either class type, suggesting it produces more class-invariant embeddings under the parameters used.

These findings highlight a key distinction in the representational structure of each target class. Emotion categories appear to elicit more globally consistent patterns, while priming conditions introduce more localized or class-specific variability: especially pronounced in projection-based features. This may reflect underlying differences in how priming manipulations influence distributed neural responses versus emotion-induced states, which tend to be more stable across regions.

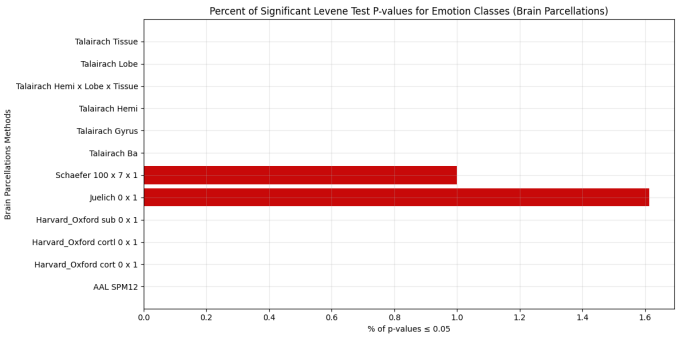

Fig. 17. Bar plot of percentage of p values from Levene test greater than .05 for each brain parcellation technique across emotion classes

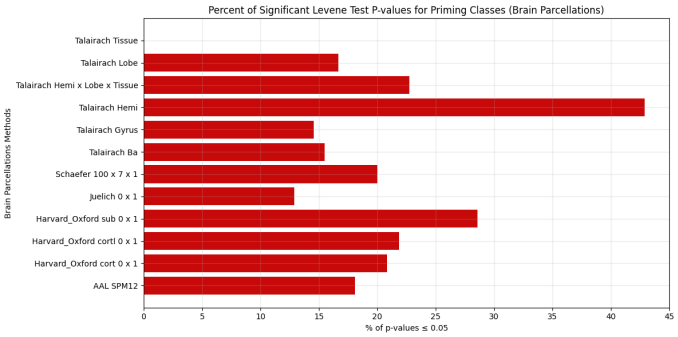

Fig. 18. Bar plot of percentage of p values from Levene test greater than .05 for each brain parcellation technique across priming classes

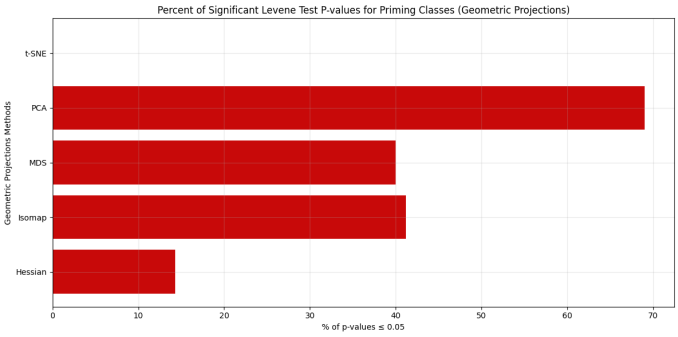

Fig. 19. Bar plot of percentage of p values from Levene test greater than .05 for each geometric projection technique across priming classes

C. *Autocorrelation*

As expected, t-SNE produced the highest number of features with a strong positive correlation to subject number. This likely limits its utility in predicting target classes, as it appears to overfit or align too closely with subject-specific variance. PCA, Hessian Eigenmapping, and Isomap also exhibited some

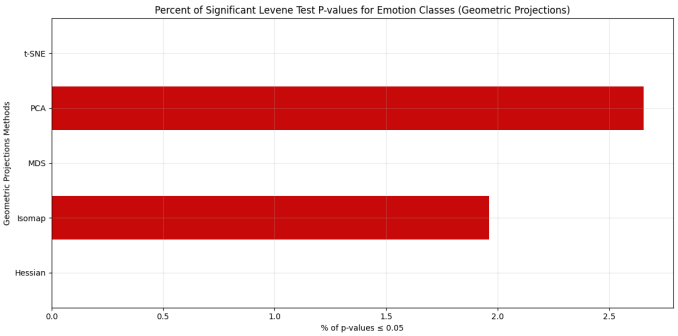

Fig. 20. Bar plot of percentage of p values from Levene test greater than .05 for each geometric projection technique across emotion classes

features with strong positive correlation to subject number, though to a lesser extent. Interestingly, Hessian Eigenmapping had the highest number of features with strong negative correlations to subject number. This was followed by Schaefer, Isomap, AAL SPM12, Harvard-Oxford Left Cortical, Talairach Gyrus, and Talairach Brodmann Areas (no particular order).

This pattern is likely due to the nature of geometric projection methods, which operate in high-dimensional space where many neighboring data points are autocorrelated. As a result, these methods may inadvertently capture subject-specific variance. In contrast, parcellation-based methods do not explicitly model the global data structure across subjects, and therefore exhibit fewer features strongly correlated with subject number.

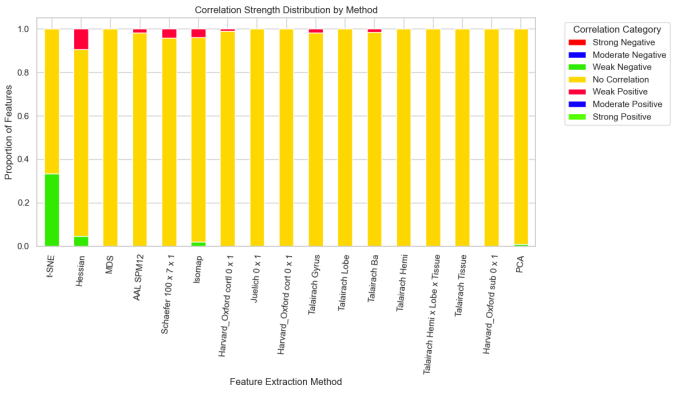

Fig. 21. Stack bar plot, representing buckets percentages of features with each type of relationship to the subject ID

D. *Multinomial Logistic Regression*

*1) **Largest Feature Coefficients By Class**:* For anger, Brodmann area 11, the Left Lingual Gyrus, and the Middle Frontal Gyrus showed the most prominent coefficients, with both positive and negative correlations indicating complex neural activations. Fear was significantly associated with the Right Central Opercular Cortex, Brodmann area 11, and the Pulvinar, while disgust exhibited strong coefficients in the GM Visual cortex V2 BA18 and the GM Secondary somatosensory cortex / Parietal operculum OP2. These findings, derived from multiple atlas methods including Talairach, Harvard Oxford, Juelich, and AAL SPM12, underscore the distributed nature of emotional processing within the brain.

Regarding priming conditions, the Middle Temporal Gyrus emerged as a highly influential feature for neutral priming, demonstrating the largest positive coefficients across Talairach and Harvard Oxford atlases. Other significant areas for neutral priming included Brodmann area 39 and the Temporal Lobe. Conversely, congruent priming showed notable negative coefficients in the Middle Temporal Gyrus and Posterior Cingulate, while incongruent priming was strongly associated with negative coefficients in the Temporal Lobe and Lateral Occipital Cortex, inferior division. These results highlight the temporal lobe's critical role in processing contextual cues related to priming.

TABLE I
TOP 5 EMOTION COEFFICIENTS BY FEATURE

Index	Feature	Method
0	Brodmann area 11	Talairach Ba
1	Left Lingual Gyrus	Harvard_Oxf
2	Middle Frontal Gyrus	Talairach Gy
3	b'7Networks_LH_Vis_8'	Schaefer 100
4	Left Occipital Fusiform Gyrus	Harvard_Oxf
5	Right Central Opercular Cortex	Harvard_Oxf
6	Brodmann area 11	Talairach Ba
7	Pulvinar	Talairach Ba
8	Calcarine_L	AAL SPM12
9	Brodmann area 43	Talairach Ba
10	GM Visual cortex V2 BA18	Juelich 0 x 1
11	GM Secondary somatosensory cortex / Parietal operculum OP2	Juelich 0 x 1
12	Right Central Opercular Cortex	Harvard_Oxf
13	Vermis_9	AAL SPM12
14	Caudate_R	AAL SPM12

TABLE II
TOP 5 PRIMING COEFFICIENTS BY FEATURE

Index	Feature	Method
0	Middle Temporal Gyrus	Talairach Gyrus
1	Middle Temporal Gyrus, temporooccipital part	Harvard_Oxford cort 0 x 1
2	Brodmann area 39	Talairach Ba
3	Temporal Lobe	Talairach Hemi x Lobe x Tiss
4	GM Inferior parietal lobule PGp	Juelich 0 x 1
5	Middle Temporal Gyrus	Talairach Gyrus
6	Feat 7	Hessian
7	Posterior Cingulate	Talairach Gyrus
8	Feat 4	Hessian
9	GM Broca's area BA44	Juelich 0 x 1
10	Temporal Lobe	Talairach Hemi x Lobe x Tiss
11	Lateral Occipital Cortex, inferior division	Harvard_Oxford cort 0 x 1
12	GM Inferior parietal lobule Pga	Juelich 0 x 1
13	Lingual Gyrus	Talairach Gyrus
14	Limbic Lobe	Talairach Hemi x Lobe x Tiss

*2) **Largest Test ROC AUC Scores By Target Variable**:* For Emotion Classification, test scores consistently ranged from 0.4805 (Isomap) to 0.5738 (Juelich 0 x 1), suggesting limited discriminative power in distinguishing individual emotional states. Methods such as Harvard Oxford cort 0 x 1 (0.5606), Talairach Ba (0.5365), and Schaefer 100 x 7 x 1 (0.5394) achieved some of the higher test AUCs for emotion. The overfit

ratios, which represent the ratio of test to train AUC, varied from 0.5295 (AAL SPM12) to 0.7959 (Hessian), indicating a notable drop in performance from training to testing and highlighting the inherent complexity and variability in distinguishing individual emotional states based on brain activity alone.

The bias-variance tradeoff seemed to be better balanced for Priming Classification, where the scores demonstrated considerably higher performance and robustness. Key methods such as PCA (0.8553) and Schaefer 100 x 7 x 1 (0.8539) consistently yielded the highest test AUCs, suggesting strong discriminative capabilities for priming conditions. Other methods like Harvard Oxford cort 0 x 1 (0.8134), Juelich 0 x 1 (0.8070), and Talairach Gyrus (0.8082) also showed strong performance. The over ratios, generally above 0.8 (ranging from 0.6703 for Isomap to 0.9519 for t-SNE), indicate effective generalization from training to unseen priming data for many methods.

3) ***Largest F1 Scores By Class****:* F1 scores further corroborated these findings, showing that priming conditions were classified with greater accuracy than individual emotions. Neutral priming achieved the highest F1 scores, notably with Schaefer 100 x 7 x 1 (0.7692), followed by congruent and incongruent priming. In contrast, F1 scores for specific emotions like anger (0.3962, 0.4800), disgust (0.3077, 0.3617), and fear (0.4048, 0.5051) were considerably lower, suggesting that these emotions are less distinctly represented in the feature space for reliable classification. PCA and Schaefer 100 x 7 x 1 generally performed well across priming F1 scores.

TABLE III
F1 SCORES BY METHOD

Index	Method	Label	F1 Score
0	PCA	congruent	0.715789
1	Talairach Gyrus	congruent	0.649123
2	Schaefer 100 x 7 x 1	congruent	0.627451
3	Harvard_Oxford cort 0 x 1	congruent	0.616667
4	Juelich 0 x 1	congruent	0.597938
5	Schaefer 100 x 7 x 1	incongruent	0.688889
6	PCA	incongruent	0.678899
7	Juelich 0 x 1	incongruent	0.659091
8	Harvard_Oxford cortl 0 x 1	incongruent	0.629213
9	Talairach Hemi x Lobe x Tissue	incongruent	0.607143
10	Schaefer 100 x 7 x 1	neutral	0.769231
11	Talairach Gyrus	neutral	0.736842
12	Juelich 0 x 1	neutral	0.705882
13	Harvard_Oxford cort 0 x 1	neutral	0.693333
14	Hessian	neutral	0.666667
15	Talairach Ba	anger	0.48
16	Juelich 0 x 1	anger	0.463158
17	AAL SPM12	anger	0.453782
18	MDS	anger	0.412698
19	PCA	anger	0.396226
20	Harvard_Oxford sub 0 x 1	disgust	0.361702
21	Schaefer 100 x 7 x 1	disgust	0.337079
22	Talairach Hemi x Lobe x Tissue	disgust	0.318841
23	Juelich 0 x 1	disgust	0.315789
24	t-SNE	disgust	0.307692
25	Juelich 0 x 1	fear	0.505051
26	Talairach Hemi x Lobe x Tissue	fear	0.432
27	PCA	fear	0.426966
28	Harvard_Oxford cortl 0 x 1	fear	0.422222
29	Harvard_Oxford sub 0 x 1	fear	0.404762

4) ***Chi Square Associations in Residuals by Method****:* Chi-square association tests on residuals confirmed significant dependencies between observed emotion/priming states and predicted labels across nearly all methods, evidenced by very low p-values. This statistical significance demonstrates that HierarchicalKFoldCV may not have been as helpful in preventing autocorrelation as attempted. There were significant associations between subject and predicted label, and subject and priming, underscoring the importance of IID in Logistic Regression multinomial classification tasks.

V. CONCLUSION

This study successfully tested the robustness of various feature extraction techniques for analyzing fMRI data, revealing distinct neural regions associated with specific emotions and priming conditions. While our logistic regression model exhibited strong classification performance for priming conditions, consistently achieving high test AUC scores and F1 scores, the reliable differentiation of individual emotional states proved more challenging, highlighting the intricate nature of emotion processing. Moving forward, future work will focus on exploring more advanced classification models like SVMs for multilabel tasks, investigating brain parcellation by race to address demographic considerations, employing sophisticated feature selection methods beyond model coefficients, and analyzing the stability of model parameters across folds to ensure robust and generalizable findings.

ACKNOWLEDGMENT

Thank you Professor Ahmad for a great semester and special thank you to Parnian for the constant support !

REFERENCES

[1] A. F. Zuur, E. N. Ieno, and C. S. Elphick, "A protocol for data exploration to avoid common statistical problems," *Methods in Ecology and Evolution*, vol. 1, no. 1, pp. 3–14, Nov. 2009. doi:10.1111/j.2041-210x.2009.00001.x

[2] "2.2. Manifold Learning," *scikit*, [Online]. Available: https://scikit-learn.org/stable/modules/manifold.html (accessed Jun. 16, 2025).

[3] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. Cambridge, MA: MIT Press, 2021.

[4] "API references," *Nilearn*, [Online]. Available: https://nilearn.github.io/dev/modules/ (accessed Jun. 16, 2025).

[5] B. T. Thomas Yeo *et al.*, "The organization of the human cerebral cortex estimated by intrinsic functional connectivity," *Journal of Neurophysiology*, vol. 106, no. 3, pp. 1125–1165, Sep. 2011. doi:10.1152/jn.00338.2011

[6] A. Schaefer *et al.*, "Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI," *Cerebral Cortex*, vol. 28, no. 9, pp. 3095–3114, Jul. 2017. doi:10.1093/cercor/bhx179

[7] O. Guy-Evans, "Brodmann areas of The brain: Anatomy and functions," *Simply Psychology*, [Online]. Available: https://www.simplypsychology.org/brodmann-areas.html (accessed Jun. 16, 2025).

[8] L. Zhang, M. Wang, M. Liu, and D. Zhang, "A survey on Deep Learning for neuroimaging-based Brain Disorder Analysis," *Frontiers in Neuroscience*, vol. 14, Oct. 2020. doi:10.3389/fnins.2020.00779

[9] P. Moghimi *et al.*, "Evaluation of functional MRI-based human brain parcellation: A Review," *Journal of Neurophysiology*, vol. 128, no. 1, pp. 197–217, Jul. 2022. doi:10.1152/jn.00411.2021

[10] L. van der Maaten and G. Hinton, "Visualizing Data using t-SNE," *Journal of Machine Learning Research*, vol. 9, no. 2579, Nov. 2008.