Mutual Information-Based Feature Facet Clustering with Ensemble SVM for fMRI Visual Object Decoding

Lina Gong
College of Electronic Information Engineering, Xi'an Siyuan University
Xi'an 710038, Shaanxi, China
E-mail: glnxhj@126.com

Keywords: fMRI, brain decoding, visual objects, mutual information, ensemble learning, classification

Received: July 20, 2025

Deciphering visual stimuli from fMRI data presents a significant challenge in computational neuroscience. This paper introduces a novel, optimized ensemble learning framework for high-accuracy visual object recognition. Our method employs a mutual information-based hierarchical clustering technique to automatically segment the high-dimensional voxel space into independent feature facets. An ensemble of Support Vector Machine (SVM) classifiers is then trained on these facets. Crucially, the entire framework—including the number of facets, the fusion operator, and SVM parameters (C, gamma)—is globally optimized using the Simulated Annealing algorithm to ensure peak performance. We rigorously evaluated our approach on three public fMRI datasets: DS105 (8 visual objects), DS107 (4 semantic categories), and DS116 (2 visual oddball stimuli). The proposed model demonstrated exceptional performance, achieving mean recognition accuracies above 95% across all three datasets, with peak subject-level accuracy reaching 100%. Specifically, our Ensemble-965 model (using the detailed Talairach Atlas) attained accuracies of 98.6% on DS105, 97.5% on DS107, and 99.4% on DS116, surpassing current state-of-the-art brain decoding methods under comparable validation conditions. These results indicate that our method provides a robust, accurate, and highly effective solution for visual brain decoding.

Povzetek: Članek predstavi ansambelski okvir za dekodiranje vizualnih dražljajev iz fMRI, ki z hierarhičnim gručenjem po medsebojni informaciji razdeli voxel-prostor na neodvisne "facete" ter z globalno optimizacijo (simulirano ohlajanje) nastavi ansambel SVM za zelo natančno prepoznavo objektov.

1 Introduction

Computational neuroscience has experienced significant expansion in recent years. Deciphering the neural code is a pivotal challenge within this discipline [1], [2]. Numerous research teams are striving to devise practical approaches for analyzing brain signals extracted from fMRI datasets. The exploration of brain decoding has a history dating back twenty-five years. A brain decoder is a computational framework designed to identify meaningful connections between brain function and external inputs. Its primary objective is to categorize neural responses into distinct high-level mental states, including stimulus processing (perceived objects), emotions, and psychological disorders [3], [4], [5]. A secondary goal involves pinpointing brain areas that exhibit pronounced selectivity toward specific stimuli, thereby aiding neurologists in understanding the underlying brain functions. One appealing application of brain decoding is the development of BCI systems. Various neuroimaging techniques, including EEG, MEG, PET, and functional magnetic resonance imaging (fMRI), have been employed in brain decoding studies. Among these modalities, fMRI captures brain functions indirectly by assessing cerebral oxygenation, specifically recognizing the rise in deoxyhemoglobin levels in the blood, known as blood oxygen level-dependent (BOLD) signals [6]. Despite fMRI images being noisy and high-dimensional, they remain a desired method for brain decoding applications due to their non-invasiveness and ability to offer intricate insights into deep brain functions. However, distinguishing neural patterns from spontaneous swings in fMRI data, even when acquired from the same individual, poses a significant challenge [7].

Typically, brain regions that are activated in response to a prompt or task are identified by pinpointing voxels where BOLD signal changes are notably linked to the empirical design. Quantitative analysis of fMRI data commonly employs a highly univariate method [8]. This involves fitting a general linear model (GLM) that accounts for the empirical settings and any potential hidden variables to each voxel's time series, generating a 3D map of parameter estimates [9]. Subsequently, activated regions are delineated using appropriate statistical inference methods. Besides univariate approaches, multivariate methods have been proposed for analyzing fMRI brain activity data, including parametric and non-parametric methods like independent component appraisal [10], clustering [11], and self-organizing mapping [12]. Recent research efforts have concentrated

on resolving the inverse problem of associating specific stimuli, tasks, or mental states with observed brain activity patterns [13]. Various pattern recognition tools, especially ML schemes, are increasingly popular for decoding mental states or stimuli from fMRI data [14], [15], [16].

In this study, the specific challenge of inferring the stimuli delivered to subjects during a visual test utilizing fMRI analysis will be addressed. The visual cortex is recognized for its functional specialization, with distinct patterns of brain activity triggered by different visual stimuli. There is a retinotopic layout in the primary visual cortex, meaning that specific domains within the cortex relate to other sectors of the visual field. Notably, the left visual field is processed in the right hemisphere, and vice versa. The visual pathway is a prime example of how the brain encodes and decodes visual stimuli. This pathway processes visual information starting from the retina's reception of visual stimuli. The optic nerve carries sensory data for vision, leading from each eye to the optic chiasm and the lateral geniculate body, relaying visual information to the primary visual cortex [17]. Fig. 1 demonstrates how the visual cortex encodes visual information, as reflected in fMRI data.

In the current research, a new method based on multimodal learning is suggested to handle the problem of decoding visual objects in the human brain. The new framework proposed in this work is based on mutual information and an ensemble support vector ML approach. Indeed, the main contribution of the current study is utilizing the simulated annealing algorithm in the proposed framework to determine the best voxel subsets in the input voxel space. To this end, this study is guided by the following specific research questions and hypotheses:

- Can a mutual information-based clustering of the voxel space into independent "feature facets" yield a more discriminative ensemble model compared to using unsegmented feature sets or standard anatomical atlases? We hypothesize that clustering voxels based on their mutual information with the stimulus labels will create semantically richer and more independent subsets. The expected benefit is that each classifier in the ensemble becomes a specialized expert on a distinct aspect of the neural code, leading to higher overall decoding accuracy and better generalization compared to models trained on monolithic or arbitrarily segmented feature spaces.

- Does the use of the Simulated Annealing algorithm for global hyperparameter optimization provide a significant advantage over conventional methods like grid search or random search in the context of this complex ensemble framework? We hypothesize that Simulated Annealing will systematically outperform grid and random search by more efficiently navigating the high-dimensional, mixed-variable search space (containing continuous parameters like SVM C/gamma and discrete parameters like the number of facets). We expect SA to find superior configurations with fewer computational resources, thereby maximizing the ensemble's performance potential where simpler methods would converge to suboptimal solutions.

- Can the proposed integrated framework, combining optimized voxel clustering, SA tuning, and an ensemble of SVMs, achieve state-of-the-art decoding accuracy on standardized public fMRI datasets (DS105, DS107, DS116) and outperform current leading methods? We hypothesize that the synergistic effect of our core innovations will result in a robust framework that significantly surpasses the accuracy of existing state-of-the-art brain decoding methods, as listed in Table 4, by effectively addressing their limitations in feature selection, model tuning, and decision fusion.

The rest of this document is outlined as follows: Part 2 supplies a literature review, Part 3 outlines the datasets utilized in the experiments as well as the proposed method, Part 4 showcases the obtained outcomes, Part 5 discusses the findings, and the document concludes with Part 6.

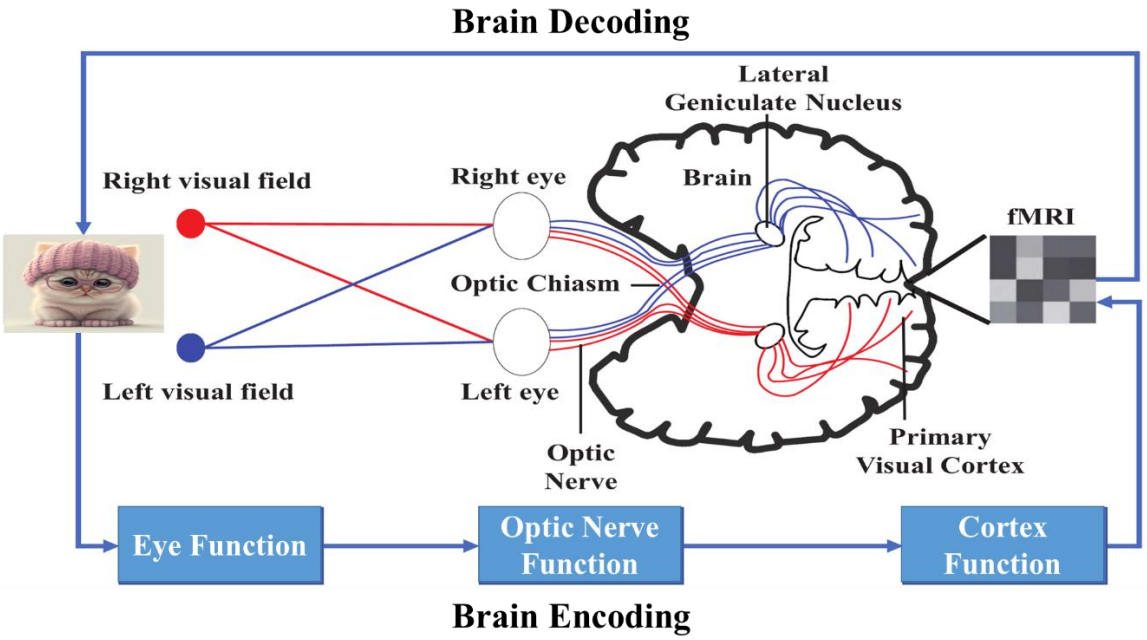

Figure 1: Encoding visual data through the visual cortex, as reflected in fMRI data

2 Related work

Associating brain activities with mental states is seen as a pattern recognition hurdle. Considering this perspective, fMRI data represent spatial patterns, making it viable to utilize statistical pattern identification methods, including ML algorithms, to correlate these spatial patterns with immediate brain states. By employing ML algorithms, researchers can analyze and classify multivariate data by leveraging the statistical characteristics present in the fMRI dataset. Various studies have explored the issue of decoding the brain, particularly focusing on visual object perception. Three primary methods have been proposed to address the brain decoding issue: neural pattern categorization, stimulus recognition, and reconstruction algorithms [18]. This research focuses explicitly on neural pattern classification, which can be categorized into four distinct methods: nonlinear models like the multi-layer perceptron [19], probabilistic methods [20], graph-based decoding [21], and multivariate pattern analysis (MVPA) [22]. In general, the brain decoder typically consists of 3 main stages: preprocessing, voxel (feature) choice, and model (decoder) training. Given the vast and high-dimensional nature of fMRI images, the voxel selection process is pivotal in the efficacy of the decoder [23]. Consequently, numerous early studies have concentrated on developing effective voxel selection techniques. One well-known voxel selection method is GLM, utilized as a neural activity localizer to select voxels in fMRI scans that best align with the task modeling when presented with specific stimuli. Other studies have deployed mutual information between voxels and stimuli as a voxel selection method for activity localization. Additionally, Principal Component Analysis (PCA) has been utilized in certain studies for voxel scoring. The PCA approach involves analyzing the variance in time-series voxels to identify orthogonal components with high discriminatory abilities through a nonlinear transformation [24]. Furthermore, some studies have leveraged Graph-based methods to pinpoint brain activity and identify the most critical voxels [25].

Moreover, the selection of the classifier holds significant value in brain decoding endeavors. Historically, frameworks relying on a single classifier have been the norm in this field. Commonly utilized classifiers in prior studies include Fisher linear discriminant, KNN, Gaussian Naïve Bayes, and SVM [26]. Interestingly, in various studies, it has been noted that the amalgamation or ensemble of classifiers outperforms individual classifiers [27]. This phenomenon is particularly evident when dealing with high-dimensional classification tasks, including those encountered in fMRI analysis. Some research has already explored using ensembles with fMRI data. For example, [28] employed decision tree ensembles to decode fMRI connectivity patterns, while [29] utilized the AdaBoost classifier to distinguish between drug-addicted individuals and healthy controls. In a separate study [30], various ensemble methods were applied to identify distinctive brain activation patterns in reaction to diverse visual stimuli, albeit on data from individual subjects only. These ensembles comprised configurations deployed for training base classifiers on diverse subsets of the training data (e.g., Random Forest, AdaBoost, Bagging) or diverse subsets of input attributes (Random subspace). Another study [31] also employed Random Forests to decode visual stimuli utilizing varied attribute subsets, where the selection of features was based on Gini Contrast. In a study on 3D motion processing, it was found that direction could be decoded from human V1, MT, MST, and FST areas, with accuracy cue-dependent. MT was specialized for perspective cues while FST was specialized for stereoscopic cues, with FST's activity being more closely linked to perceptual experience [32]. To address cross-subject variability in fMRI decoding, Li et al. introduced a Global-Local Functional Alignment (GLFA) method to project fMRI data from multiple subjects into a unified space. Using a large, novel fMRI-video dataset and a transformer-diffusion model, they achieved state-of-the-art performance in cross-subject video reconstruction and semantic classification [33]. A direct comparison between SVM and CNNs for fMRI decoding revealed that CNNs consistently achieve higher accuracy. Critically, the two methods relied on different neural features for classification, suggesting their combined use can provide a more comprehensive understanding of brain activity patterns [34].

In this exploration, a new approach, drawing on multimodal learning, is suggested to solve the problem of decoding visual objects in the brain. To introduce a streamlined and rapid voxel selection framework, we have adopted a two-tiered selection process influenced by prior research. Initially, we employ a filter approach grounded in statistical modeling and atlas data, which serves to diminish the number of voxels (dimensional space) and reduces the time complexity for subsequent stages. Subsequently, the wrapper component entails a voxel scoring algorithm aimed at selecting the most informative voxel subset, accompanied by the simulated annealing search algorithm. The core contributions of our work are threefold:

1. Structured Feature Space Decomposition: We propose a mutual information-based hierarchical clustering method to automatically segment the high-dimensional voxel space into meaningful, independent feature facets. This moves beyond arbitrary or anatomical subdivisions to create clusters based on the informational relationship between features and stimulus labels, ensuring each facet captures a distinct aspect of the neural code.

2. Robust and Automated Optimization: We integrate the Simulated Annealing (SA) algorithm to automatically and efficiently determine the global optimum of the entire ensemble's configuration. This includes the number of feature facets, the SVM parameters (C, gamma), and the fusion operator. This systematic optimization ensures our model is consistently configured for peak performance without manual tuning, a significant advantage over ad-hoc approaches.

3. A Comprehensive and Reproducible Framework: We validate our method on three public datasets, demonstrating state-of-the-art accuracy. More

importantly, we provide a complete and transparent account of our method, including SA parameters, optimization robustness (30 independent runs), and optimal configurations, ensuring full reproducibility and a strong baseline for future work.

3 Methodology

3.1 fMRI databases

This exploration examines three real fMRI databases focusing on the item-observation task in the brain. These databases were sourced from the "openfmri.org" website, which has since evolved into "openneuro.org." Both websites guarantee that human subject data has been de-identified through methods including eliminating facial attributes from high-detail anatomical MRI and removing personal information. The initial database (DS105) is a typical benchmark for studying visual item recognition tasks within the brain (https://openneuro.org/datasets/ds000105/versions/00001). DS105 contains eight classes of visual items for six persons, including face, house, shoe, chair, cat, and scrambled grey-scale photos. This collection of data focuses on the analysis of high-level visual cues as binary predictors, encompassing all groups except scrambled images treated as objects, while also considering low-level visual cues for the multiclass projection. The stimuli consisted of black-and-white photos depicting faces, houses, cats, bottles, scissors, shoes, chairs, and random patterns. Each category was selected so that all stimuli within a particular category shared a common base name. Control random patterns were created by distorting images of the intact objects. Twelve sets of data were collected for each participant. Each set began and ended with 12 seconds of rest, and included eight blocks of stimuli, each lasting 24 seconds, corresponding to the eight categories with 12-second breaks in between. The stimuli were displayed for 500 milliseconds with a 1500-millisecond gap between them. Meaningful stimuli were repetitions of images depicting the same face or object photographed from different perspectives. For each category of meaningful stimuli, there were four images for each of the 12 other examples.

The second database (DS107) contains fMRI data relevant to viewing four diverse groups associated with visual items and textual writings, including consonant, word, object, and scrambled images (https://openneuro.org/datasets/ds000107/versions/00001). The word stimuli, totaling 168 in number, were comprised of 4 or 5-letter words with standard spellings, for example, "hope." They were either monosyllabic or disyllabic and had a written word occurring 40 times or fewer in British English. The stimuli in the two sessions were precisely synchronized in terms of frequency, familiarity, imaginability, count of letters, and count of syllables. Meanwhile, the object cues encompassed grayscale images (200 × 250 pixels) of easily identifiable items, including a boat, a tent, a nail, and so on. The scrambled items were created by partitioning the images into 10 × 10 pixel squares and rearranging their placement in the photo. After scrambling, none of the resulting pictures were distinguishable. Lastly, strings of consonant letters were nonsensical sequences created haphazardly to match the length of the word stimuli precisely.

The 3rd database (DS116) contains both EEG and fMRI data of oddball visual and auditory tasks (https://openneuro.org/datasets/ds000116/versions/00001), only the fMRI visual activity that was used in this work [35]. In this database, 17 individuals took part in a series of three runs, each encompassing similar visual and auditory oddball paradigms. In each task, a total of 375 stimuli (125 per run) were displayed for a duration of 200 milliseconds each, with a variable inter-trial interval uniformly distributed between 2 and 3 seconds, and a target probability of 0.2. It was ensured that the first two stimuli of each run were standard. In the visual task, the target stimulus was a large red circle, while the baseline stimulus was a small green circle, both presented against isoluminant gray backgrounds, with visual angles measuring 3.45° and 1.15°, respectively. A summary of these databases is presented in Table 1.

Table 1: Summary of datasets used in this exploration

Dataset	#of Subjects	#of runs	#of classes	#of time points	Size of images	#of voxels in ROI	TR/TE	FWHM
DS105	6	71	8	121	79×95×79	1963	2.5/30	5 mm
DS107	49	98	4	164	53×63×52	932	2/28	6 mm
DS116	17	102	2	170	53×63×40	2532	2/25	5 mm

TR = time of repetition (s); TE = echo time (ms); FWHM = full width at half maximum.

3.2 Proposed framework

Fig. 2 displays the typical stages of the proposed method. As shown, after collecting the fMRI data from the mentioned databases and preprocessing it, the effective voxels in the atlas regions are selected using a general linear regression model and a z-score. In the next step, the simulated annealing optimization algorithm is used to set important parameters, including the number of feature facets, fusion operator, and classifier parameters. Then, mutual information (between the combination of features and values of voxels) in the fMRI data on the one hand and the labels (categories) on the other hand are obtained. Then, the features are clustered into a certain number of clusters based on mutual information. Thus, each feature cluster expresses an independent feature facet. In the next step, a classifier is trained on each feature facet. Finally, the ensemble process is done using a decision profile matrix, and the final output is obtained. Each of these stages is detailed below.

3.2.1 Data preprocessing

Preprocessing of fMRI data was executed through the fMRI Expert Analysis Tool from the FMRIB software library (FSL, version 6.00) [36]. Before initiating the first-level analysis in FEAT, preprocessing stages were performed on the fMRI data files. This included skull stripping and alignment to the 2 mm MNI152 standard space. Mapping fMRI data onto the MNI template was achieved through FLIRT registration using high-resolution MRI as an intermediate reference. fMRIs in all databases underwent affine registration with 6 or 12 degrees of freedom. Various preprocessing stages, like registration, movement correction, smoothing, and filtering, were conducted in FEAT as the second preprocessing stage.

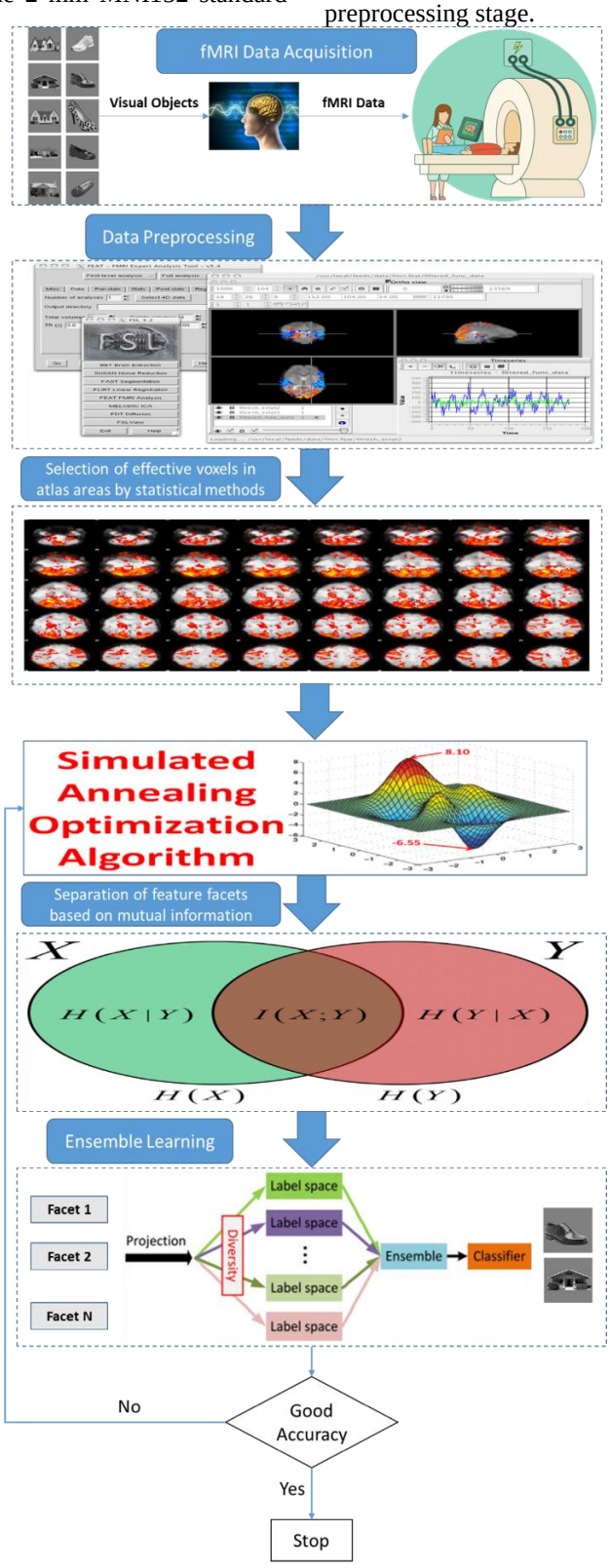

Figure 2: Typical stages of the proposed method

For the DS107 database, an MNI brain mask was applied to remove non-brain tissues from all fMRIs, including structures like the skull, neck, and eyes, using BET for non-brain structure removal, MCFLIRT for movement correction, and spatial smoothing with a 5 mm FWHM Gaussian kernel. The FLIRT registration acquired mapping onto the MNI152 anatomical template with 2 mm resolution using a normal search in 12 degrees of freedom. To expedite and automate the fMRI processing utilizing FSL, a set of UNIX shell scripts was developed. Fig. 3 displays the preprocessing stages for fMRIs.

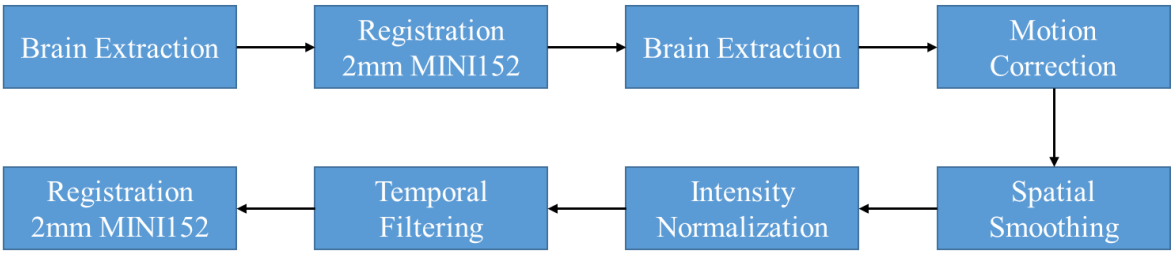

Figure 3: Implemented preprocessing stages for fMRIs in this work.

3.2.2 Feature selection

Feature selection stands as a pivotal step in ML. Within fMRI processing, the challenge lies in managing a multitude of voxels, numbering in the hundreds of thousands. Each voxel displays the activity of a point within the brain space, delineated by 3-dimensional coordinates. Effective voxel selection poses a significant hurdle in computational neuroscience. The objective at this juncture is to diminish the spatial dimensions of the voxels by singling out the most impactful ones, rather than considering all brain voxels. In the standard 2 mm space, the volume of voxels escalates into the millions. Hence, an effective voxel selection method is employed, where an appropriate voxel from each anatomical region in an atlas is chosen for an external stimulus. Examples of such atlases include the Harvard-Oxford cortical Atlas, comprising 96 areas of the human brain cortex, or the Talairach Atlas, encompassing 965 anatomical regions of the human brain cortex [37].

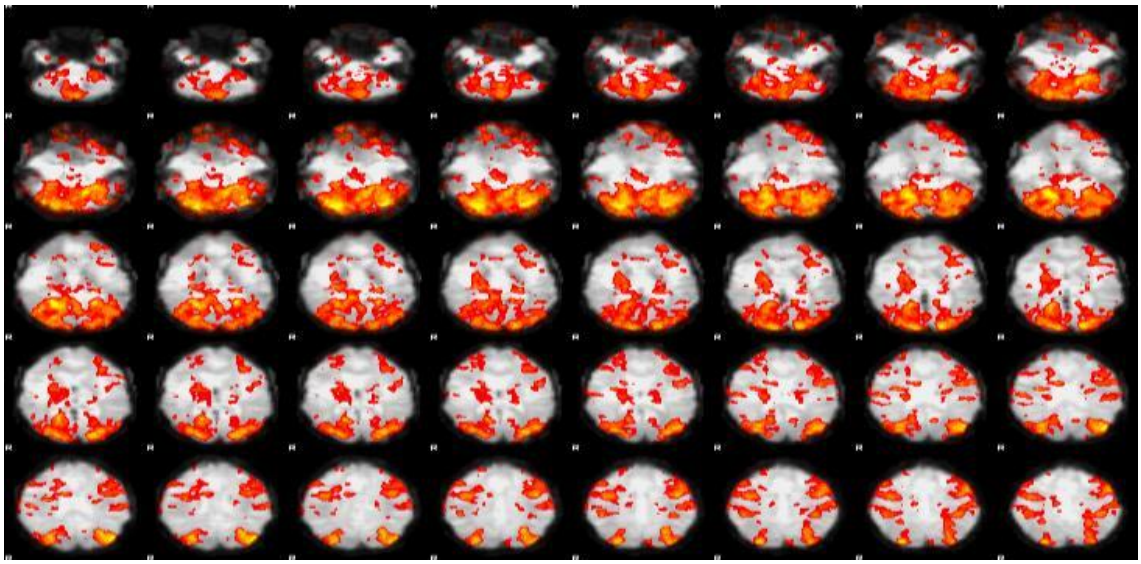

Figure 4: The localization outcomes of the brain activity linked to watching the image of the house

At this stage, the statistical dependence between the stimuli and the time series of the voxels was determined by the general linear regression model (GLM). For each stimulus, a statistical map of the z-score was calculated. For each anatomical region in the Harvard-Oxford cortical atlas (comprising 96 regions), the single voxel that had the maximum z-score was selected. An example of the localization outcomes of the brain activity linked to watching the image of the house can be seen in Fig. 4. The selection process was repeated for each label in the datasets. In this way, the number of features varied according to the type of stimuli and anatomical regions. In the selection algorithm, eight voxels were extracted from each area in the atlas in 9 time intervals. Therefore, 72 features were obtained for each anatomical region. So, using the Harvard-Oxford Atlas, there were 96×72 features.

To encode the temporal dynamics of the hemodynamic response, we did not average across time, as this would discard crucial pattern information. Instead, we employed a concatenation-based approach. For the selected voxel in each of the 96 regions, we extracted its

BOLD signal intensity across 9 consecutive time points following stimulus onset. These 9 time-series values for each region were then flattened and concatenated into a single, high-dimensional feature vector. An important point is to choose a suitable compromise between the number of areas defined in the brain atlas and the amount of practical brain signal information in each brain region. Increasing the number of areas improves the accuracy at the cost of expanding the computation cost. Fig. 5 displays the details of the feature vector structure used.

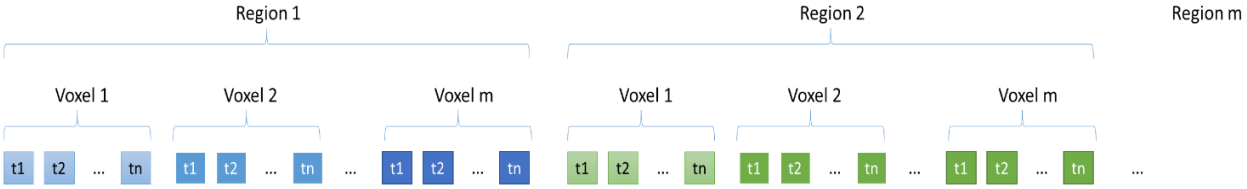

Figure 5: The details of the feature vector structure used in the suggested framework are provided.

After feature selection, voxel post-processing stands as a crucial stage within the suggested technique. This involves the standardization of voxel values and their organization into a vector format for subsequent utilization in the learning processes. The key procedures encompass normalization, segmentation, flattening, discretization, and scaling. Details of these stages are given in the references [23], [38].

3.2.3 Simulated annealing optimization

Learning algorithms, including SVM, are generally dependent on parameters that must be set well to function correctly. The learning method in the proposed framework is an ensemble, which includes multiple SVM classifiers. Conversely, in ensemble schemes, things like the number of feature facets and the fusion operator are of great importance. Therefore, the proposed framework in the learning and classification stage has essential parameters that should be tuned appropriately. Owing to the importance of reducing the search space and setting the optimal parameter, optimization algorithms are used. The selection of an optimization algorithm is critical for navigating complex, non-convex search spaces like the one defined by our ensemble framework, which mixes discrete parameters (e.g., number of facets) and continuous parameters (e.g., SVM C, gamma). We selected the simulated annealing (SA) optimization algorithm for this purpose due to its proven ability to escape local minima (a key advantage over simpler methods like grid search), its effective balance between exploration and exploitation via its cooling schedule, and its established efficacy in machine learning hyperparameter tuning. In this work, the simulated annealing optimization algorithm was utilized for this purpose. Today, this algorithm is widely used in various fields in optimization and searching for problem solutions. In this research, the parameters of penalty and gamma were considered constant for all SVMs, which should be determined by the simulated annealing method. Also, the scheme tries to find the appropriate number of feature facets and the fusion operator. The error rate of the ensemble categorization scheme acts as a fitness function. A simulated annealing algorithm minimizes this error by searching in the space of the number of facets and default parameters in the framework, including annealing speed, exponential function, and acceptance function. The Simulated Annealing (SA) algorithm was configured based on established practices and preliminary experimentation to balance global exploration and local exploitation. The tested ranges and the final selected values for all key parameters are summarized in Table 2.

Table 2: Simulated annealing parameter tuning

Parameter	Tested Range / Options	Selected Value
Initial Temperature (T0)	[10, 500]	100
Cooling Schedule	Linear, Geometric	Geometric ($T_{k+1} = \alpha \times T_k$)
Cooling Rate (α)	[0.85, 0.99]	0.95
Markov Chain Length (L)	[50, 200]	100
Stopping Criterion	Max Iterations, T_{min}	$T_{min} = 0.00001$

The selected configuration was determined to be the most robust through grid search on a subset of the DS105 data. The combination of a geometric cooling schedule with a rate of 0.95 and an initial temperature of 100.0 consistently achieved a lower final error rate compared to other tested configurations. This specific setup allows for aggressive exploration in the early stages, where the algorithm can traverse diverse regions of the complex, high-dimensional search space. As the temperature cools, the search intelligently transitions to a more exploitative phase, finely tuning the parameters in the most promising areas. This balanced approach was critical for reliably finding a near-optimal configuration for the ensemble framework across all datasets.

To ensure the robustness and reliability of the optimization results, the Simulated Annealing algorithm was executed multiple times for each subject and dataset. Specifically, we performed 30 independent runs with different random seeds for each optimization task. This approach allows us to account for the inherent randomness of the algorithm and to statistically confirm that the reported solution is a consistent, high-quality optimum

rather than an outlier. The solution with the lowest ensemble error rate across all 30 runs was selected as the final, optimal configuration for each subject. The standard deviation of the final error rate across these runs was consistently low, confirming the stability of the optimization process and the reliability of the reported parameters.

3.2.4 Automatic segmentation of feature space

Automatic feature space segmentation is a critical step in multifaceted or multimodal learning, especially when there is no natural separation between feature facets. The key innovation of this exploration is to provide a scheme based on mutual information clustering for feature segmentation. Fig. 6 displays the typical stages of the automatic feature space segmentation algorithm. These stages include mutual information calculation and hierarchical clustering. The inputs of the feature space segmentation algorithm are the threshold level of voxel values, the number of feature facets, and brain fMRI data that includes voxel values and labels. First, a threshold is applied to the input feature values to convert these values to 0 and 1. The labels are then binary encoded. In this way, multi-category labels become binary labels. This is similar to converting labels to one-versus-all. An example of converting multi-category labels to binary is displayed in Table 3.

Table 3: Label the binarization process

True Label	Binary Labels			
1	1	0	0	0
2	0	1	0	0
3	0	0	1	0
4	0	0	0	1

The main discrepancy between the proposed algorithm and other similar methods is that, on the one hand, the threshold value is applied to the features. On the other hand, the labels are encoded to be converted into binary. This process is considered a type of approximation that plays a crucial role in accelerating calculations and mitigating potential hidden noise in mutual information values. Mutual information is used as an essential criterion to minimize uncertainty in automatic feature space extraction and segmentation. The mutual information between any two binary vectors (labels and features) is calculated based on the following equation to obtain the degree of dependence between each pair of vectors.

$$MI(A,B) = \sum_{a\in A}\sum_{b\in B} p(a,b) \times \log\left(\frac{p(a,b)}{P(a).p(b)}\right) \quad (1)$$

Where $MI(A,B)$ denotes the mutual information between 2 variables A and B, $p(a,b)$ denotes the joint probability of A and B, and p(a) and p(b) denote the marginal probabilities of A and B. The output of Eq. (1) is used as input for hierarchical clustering to segment the feature facets. In the following, mutual information rates are calculated for each binary feature in feature label combinations. For example, in the DS105 dataset, eight mutual information values are computed for each feature. These eight values are added together and reported as the final mutual information of the features. After calculating the mutual information values, the hierarchical clustering phase begins. In this step, the number of clusters and the vector containing mutual information values are defined as inputs. The output of clustering is in the form of categories of features that belong to each facet. This process causes features that are similar to each other based on mutual information to be placed in a separate aspect, and the correlation between feature facets is reduced. Fig. 6 displays the typical stages of the automatic feature space segmentation framework.

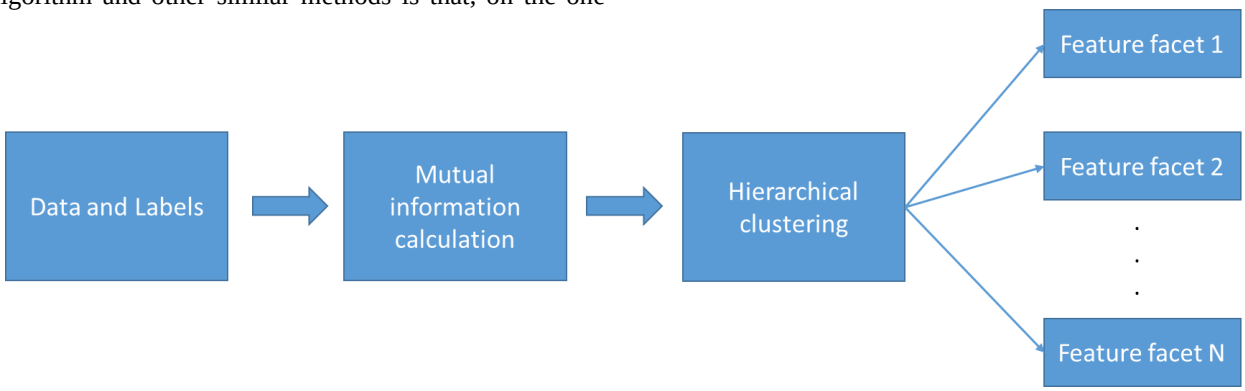

Figure 6: The typical stages of the automatic feature space segmentation algorithm

3.2.5 Classification of each feature facet

In the proposed scheme, an SVM classifier with an RBF kernel was used to classify each feature facet. Thus, each SVM classifier is specific to the classification of one visual object. For example, a classifier is more accurate in identifying human face images than other classifiers. However, it should be noted that the output of each classifier was considered a probability vector. The count of elements of this vector corresponds to the number of labels, and each value of this vector contains the degree to which the test sample belongs to the opposite category. This vector is used in the ensemble learning phase.

3.2.6 Ensemble technique

The main advantage of ensemble learning is to achieve higher accuracy compared to a single model by giving importance to correct answers in diverse classifiers.

Calculating the weight of the test sample belonging to each of the problem categories (including eight visual objects in the DS105 database) is done at this stage. As mentioned, an SVM classifier is trained individually for each feature facet. Then, the decision profile matrix is used to determine the final output for decision-making in the ensemble stage. If there are several classifiers, then the output of each classifier is defined as a membership vector for that classifier. Finally, the output of these classifiers forms a matrix. Fig. 7 displays this matrix.

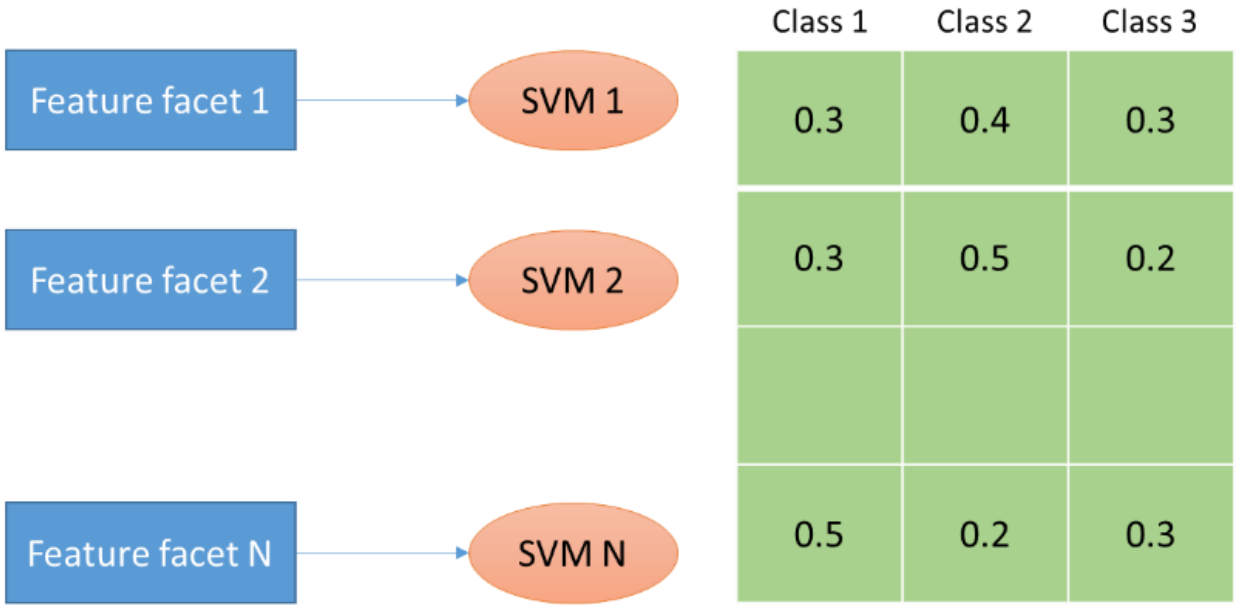

Figure 7: An instance of the decision profile matrix

In this work, the ordered weighted averaging (OWA) method based on the decision profile matrix is used as a fusion operator in ensemble learning. The OWA method is a fusion operator used in ensemble learning, which makes decisions based on a decision profile matrix. In this method, the decision profile matrix displays the decisions made by multiple individual classifiers. Each row of the decision profile matrix corresponds to a particular class or category, and each column corresponds to the output of a specific base classifier in the ensemble. The OWA method operates by assigning weights to the individual classifiers' outputs based on their order in the decision profile matrix. The weights are determined by a specific weighting vector that is pre-defined or learned from the data. The weights are then used to combine the outputs of the individual classifiers into a single fused decision. By using the OWA method, the ensemble learning system can effectively take into account the relative importance of each classifier's decision based on its position in the decision profile matrix. This can lead to improved overall decision-making performance compared to simply averaging the individual classifier outputs or using other fusion methods. Overall, the OWA method leverages the decision profile matrix and weighted aggregation to integrate the diverse outputs of individual classifiers in an ensemble, to enhance the ensemble's predictive capability and performance.

The complete pipeline for the proposed ensemble fMRI decoding framework is summarized in Algorithm 1. The procedure integrates data preprocessing, feature selection, optimization, and ensemble learning into a cohesive workflow.

Algorithm 1. Ensemble SVM with optimized feature facets for fMRI decoding

Input: Raw fMRI data D, Anatomical Atlas A, Stimulus labels y
Output: Trained ensemble model $M_{ensemble}$, Final prediction for test sample $\hat{y}$
Hyperparameters: SA parameters (T_0, α, L, T_{min}), SVM kernel, MI discretization threshold τ
1. // Stage 1: Data Preprocessing
2. $D_{preprocessed}$ ← FSL_FEAT(D)
▷ Skull stripping, motion correction, MNI registration, spatial smoothing (5mm FWHM)
3.
4. // Stage 2: Feature Selection & Vector Construction
5. For each anatomical region $r_i \in A$ do
▷ A is Harvard-Oxford (96 regions) or Talairach (965 regions) atlas
6. Calculate GLM and z-score map for each stimulus condition.
7. Select the voxel v_{max} with the maximum z-score in r_i.
8. Extract BOLD time series for v_{max} across t = 9 consecutive TRs post-stimulus.
9. end For
10. Construct feature vector X by concatenating all v_{max} time series.

▷ Final vector length: |A| × 9
11.
12. // Stage 3: Simulated Annealing Optimization
13. Initialize SA with $T = T_0 = 100$, cooling rate $\alpha = 0.95$, chain length $L = 100$.
14. While $T > T_{min} = 0.00001$ do
15. Generate new candidate solution $\theta_{new} = (n_{facets}, op_{fusion}, C, \gamma)$
16. // Fitness Evaluation:
17. Binarize feature values in X using threshold $\tau \in [0.3, 0.8]$
▷ τ is optimized by SA
18. Binarize multi-class labels y to one-vs-all format.
19. Calculate pairwise MI matrix MI using Eq. (1) with binarized data.
20. Perform Hierarchical Agglomerative Clustering on MI with:
- Linkage Criterion: Ward's variance minimization
- Number of clusters: n_{facets} (from θ_{new})
21. For each feature facet F_k do
22. Train an SVM classifier with:
- Kernel: RBF $K(x_i, x_j) = \exp(-\gamma\|x_i - x_j\|^2)$
- Parameters: C (penalty), γ (kernel coefficient)
- Both C and γ are optimized by SA
23. End For
24. Fuse base classifier outputs using operator $op_{fusion} \in$ (OWA, Avg, WeightedAvg)
25. Calculate fitness $E(\theta_{new})$ as ensemble error rate via cross-validation.
26. Accept θ_{new} with probability $\min(1, \exp(E(\theta_{old}) - E(\theta_{new}))/T))$
27. $T \leftarrow T \cdot \alpha$
▷ Geometric cooling
28. End While
29. $\theta^* \leftarrow$ best solution found.
30.
31. // Stage 4: Final Model Training & Inference
32. Train final ensemble model $M_{ensemble}$ using optimal parameters θ^*
33. Return $M_{ensemble}$

3.2.7 Evaluation metrics

Accuracy refers to the presence of systematic errors, a measure of statistical bias that leads to discrepancies between the obtained outcomes and the target values, commonly known as trueness. The calculation method for determining the accuracy is outlined as follows:

$$Accuracy = \frac{\#\ of\ correctly\ classified\ samples}{\#\ of\ total\ samples} \quad (2)$$

A confusion matrix is used to evaluate the proposed scheme. Therefore, specificity and sensitivity criteria are calculated through the obtained confusion matrices. In addition, the leave-one-run-out cross-validation scenario is similar to what is used. This scenario is identical to the leave-one-out technique except that the test samples must belong to the same run, and the run samples are not haphazardly selected during test data generation. Furthermore, it is not feasible to distribute the samples from one run across both the test and training datasets. In this way, the data of each subject is divided into subsets of data based on runs. In each iteration, one of the subsets is deployed as test data, and the rest are used as training. The training and testing of the models are repeated until all subsets are used once in the test.

4 Results

Here, the outcomes of the experiments and evaluation of the proposed framework are presented. Note that the assessment of the proposed framework was done separately for two different atlases, Harvard-Oxford (Ensemble-96) and Talairach (Ensemble-965). An intra-subject scenario was deployed to review the proposed scheme. Therefore, the proposed decoding model is trained and tested based on subjects. In this regard, optimal individual parameters were obtained for each subject using the optimization algorithm. In the optimization stage, an interval was considered for searching the values of each parameter. The threshold parameter was searched in the range of 0.3-0.8, and the count of facets was in the range of 3-50. The best-performing configurations, which correspond to the highest average cross-subject accuracy reported are summarized in Table 4. These parameters represent the global optimum found by SA for the respective dataset's characteristic neural patterns and dimensionality. Table 5 displays the identification precision of visual items for all six people from DS105 utilizing the Ensemble-96, Ensemble-965, and a single SVM. Drawing on the acquired precision (mean ± standard deviation), the proposed ensemble algorithm operates very well. Precision rates utilizing the proposed scheme for subjects 5 and 6 even reach 100%. To validate that the performance

improvements of our proposed models are statistically significant and not due to chance, we conducted non-parametric statistical tests. The Wilcoxon signed-rank test was employed for pairwise comparisons due to its suitability for paired, non-normally distributed data (e.g., accuracy scores across subjects). The tests confirm that the superior performance of our ensemble methods is statistically robust. The difference between the Ensemble-965 and Ensemble-96 models is statistically significant ($p < 0.01$), providing strong evidence that the finer anatomical detail captured by the Talairach Atlas directly and meaningfully contributes to higher decoding accuracy. Furthermore, both ensemble models significantly outperform a single SVM ($p < 0.001$), validating the core premise that our ensemble framework provides a substantial advantage.

Table 4: Optimal ensemble framework parameters identified by the Simulated Annealing algorithm for each database.

Database	SVM Penalty (C)	SVM Gamma (γ)	Number of Feature Facets	Fusion Operator
DS105	12.75	0.008	18	OWA
DS107	8.90	0.015	14	Weighted Average
DS116	25.50	0.002	8	OWA

Table 5: Accuracy rates acquired by diverse algorithms for identifying visual objects in DS105.

Subject ID	Single SVM	Ensemble-96	Ensemble-965
1	94.77±5.49	98.11±5.07	95.62±6.54
2	88.91±6.34	96.42±6.10	93.80±7.61
3	92.85±6.20	96.51±6.43	97.36±5.14
4	91.73±7.26	98.12±5.06	99.95±1.05
5	93.82±8.02	99.20±3.61	100
6	96.13±7.91	100	100
Average	93.04 ± 1.98	97.91 ± 1.15	98.62 ± 1.42

The confusion matrices obtained by the proposed Ensemble-96 and Ensemble-965 models for DS105 are shown in Fig. 8. The labels are outlined horizontally, and the column displays the cumulative responses of the proposed scheme. The larger the value of the principal diameter of the confusion matrix, the more precise the categorization. The extra column on the right displays the sensitivity, and the extra row at the bottom of the matrix displays the specificity. The maximum precision is linked to the three groups of house, shoe, and face, and the minimal efficacy pertains to scissors and bottle. The recognition accuracy of the proposed Ensemble-96 model is cumulatively about 97.9%, while it is cumulatively about 98.6% for the Ensemble-965 model. Therefore, considering more anatomical details led to improved accuracy. In the following stage, the efficacy of the proposed scheme was inspected for DS107.

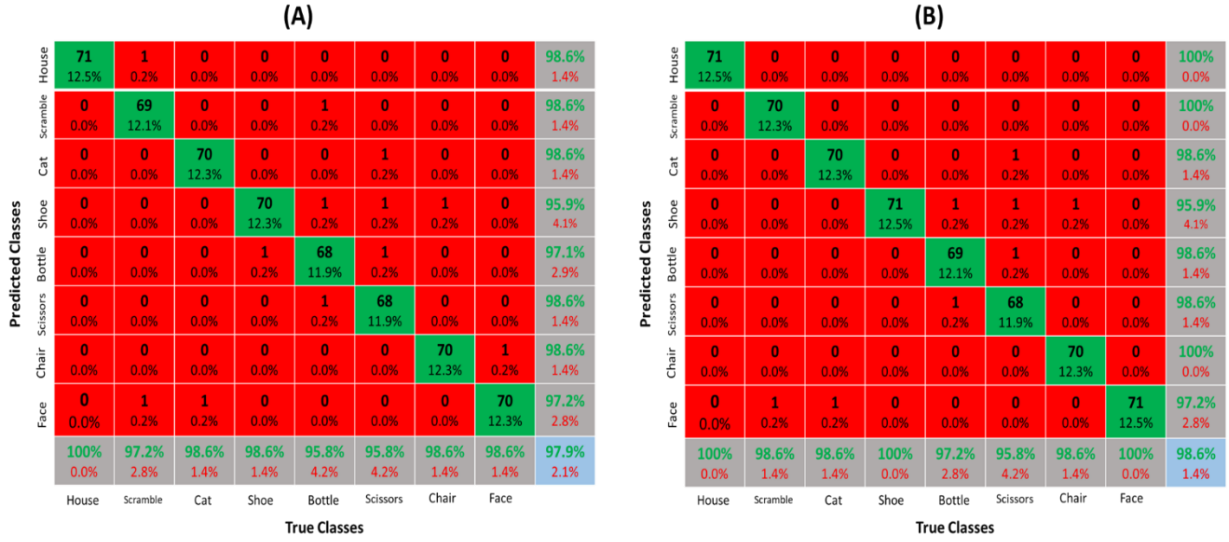

Figure 8: The confusion matrices obtained by the proposed (A) Ensemble-96 and (B) Ensemble-965 models for the DS105 database.

The confusion matrices obtained by the proposed Ensemble-96 and Ensemble-965 models for DS107 are displayed in Fig. 9. Again, the extra column on the right side displays the sensitivity, and the additional row at the bottom of the matrix displays the specificity. The maximum precision is linked to the two groups of scrambled objects and words, and the minimum efficacy pertains to the word group. The recognition accuracy of the proposed Ensemble-96 scheme is cumulatively about 95.1%, while it is cumulatively about 97.5% for the Ensemble-965 model. Fig. 10 displays the Accuracy rates obtained by the Ensemble-965 model in the DS107

database for each subject. As displayed, an accuracy rate of 100% was achieved for 25 subjects.

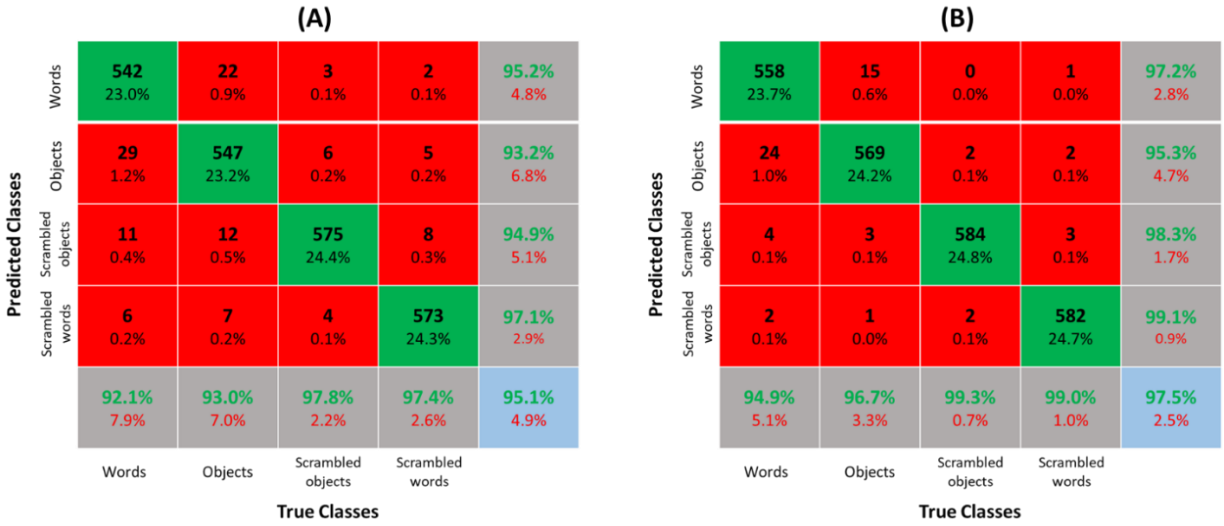

Figure 9: The confusion matrices obtained by the proposed (A) Ensemble-96 and (B) Ensemble-965 models for the DS107 database.

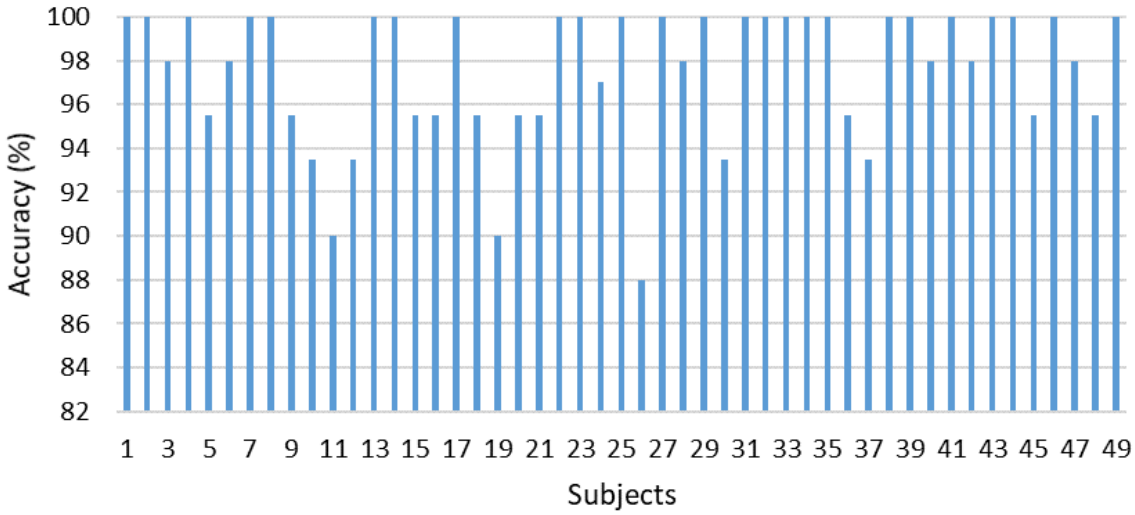

Figure 10: Accuracy rates obtained by the Ensemble-965 model in the DS107 database for each subject.

The confusion matrices obtained by the proposed Ensemble-96 and Ensemble-965 models for DS116 are displayed in Fig. 11. Again, the additional column on the right side shows the sensitivity, and the additional row at the bottom of the matrix displays the specificity. The recognition accuracy of the proposed Ensemble-96 scheme is cumulatively about 98.1%, while it is cumulatively about 99.4% for the Ensemble-965 model. Fig. 12 displays the Accuracy rates obtained by the Ensemble-965 model in the DS116 database for each subject. As displayed, an accuracy rate of 100% was achieved for 11 subjects.

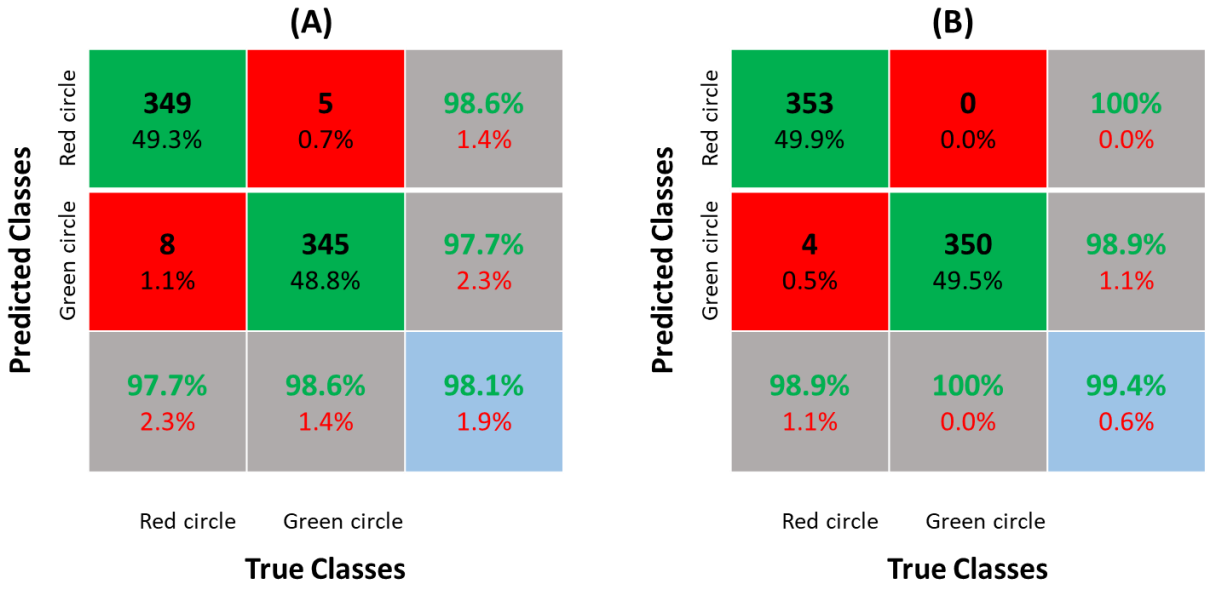

Figure 11: The confusion matrices obtained by the proposed (A) Ensemble-96 and (B) Ensemble-965 models for the DS116 database.

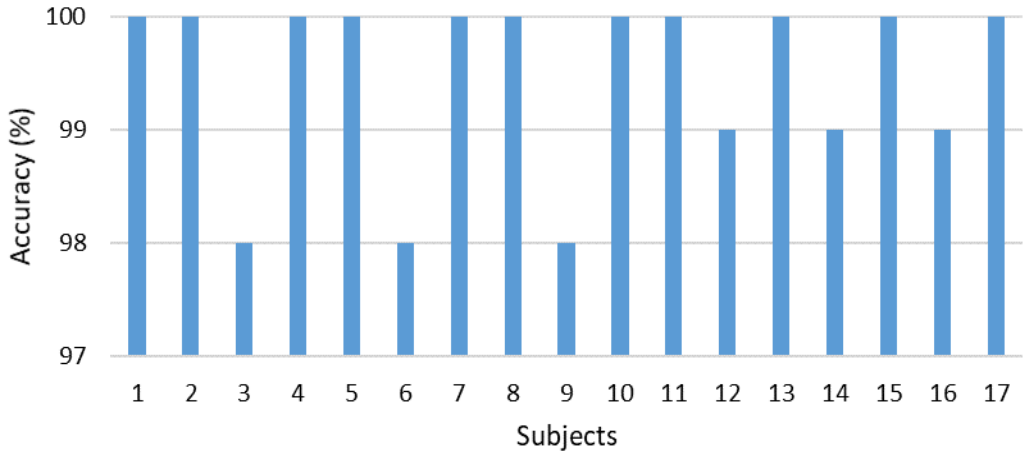

Figure 12: Accuracy rates obtained by the Ensemble-965 model in the DS116 database for each subject.

While accuracy provides a general overview, we report F1-score, precision, and recall to offer a nuanced view of model performance, particularly important for datasets with class imbalances. The results for our best-performing model (Ensemble-965) are summarized in Table 6. The high values for precision, recall, and F1-score across all datasets confirm that the high accuracy is not an artifact of class distribution but reflects truly robust classification. For DS116, the near-perfect and balanced precision and recall scores (99.52% and 99.40%, respectively) definitively demonstrate the model's effectiveness even on a binary task, with no bias towards either the frequent or infrequent class.

Table 6: Comprehensive classification metrics for the Ensemble-965 model

Database	Accuracy	Precision	Recall	F1-Score
DS105	98.62 ± 1.42	98.55 ± 1.61	98.60 ± 1.58	98.57 ± 1.48
DS107	97.54 ± 1.26	97.45 ± 1.38	97.52 ± 1.41	97.48 ± 1.32
DS116	99.43 ± 1.08	99.52 ± 0.95	99.40 ± 1.12	99.46 ± 0.98

To justify the selection of Simulated Annealing (SA), a comparative analysis was conducted against three other prominent metaheuristic optimization algorithms: Genetic Algorithm (GA), Particle Swarm Optimization (PSO), and a simple Random Search (RS). The same parameter search space was used for all optimizers. Each algorithm was allocated a fixed budget of 1000 function evaluations (calls to the fitness function, i.e., the ensemble error rate) to ensure a fair comparison. The performance was evaluated based on the final best fitness value (lowest

error rate) found. The results of the comparative optimization analysis are summarized in Table 7. As shown, the Simulated Annealing (SA) algorithm consistently outperformed its counterparts, achieving the lowest final error rate across all three datasets. This confirms its suitability for the high-dimensional, mixed-parameter optimization problem inherent in our ensemble framework. To statistically validate the superiority of SA, a one-way repeated measures ANOVA with post-hoc pairwise comparisons (Bonferroni-corrected) was conducted on the final ensemble error rates. The analysis revealed a statistically significant main effect of the optimization algorithm on the error rate across all datasets ($p < 0.01$). Post-hoc tests confirmed that SA consistently achieved a significantly lower final error rate compared to GA, PSO, and RS (all p-values < 0.01).

Table 7: Final ensemble error rate (%) achieved by different optimization algorithms.

Dataset	Simulated Annealing (SA)	Genetic Algorithm (GA)	Particle Swarm (PSO)	Random Search (RS)
DS105	2.09 ± 0.15	2.85 ± 0.21	3.11 ± 0.30	4.52 ± 0.41
DS107	2.46 ± 0.26	3.45 ± 0.33	3.72 ± 0.28	5.14 ± 0.55
DS116	0.57 ± 0.08	1.12 ± 0.14	1.33 ± 0.19	2.01 ± 0.24

Moreover, the selection of the OWA operator by the SA algorithm was a key outcome of our optimization process. To empirically validate that this choice is superior to common alternatives, we conducted a controlled comparison using the optimized Ensemble-965 architecture on the DS105 dataset, only varying the fusion strategy. A Friedman test with post-hoc Nemenyi tests was used for statistical comparison. The results are summarized in Table 8. The statistical analysis provides clear empirical justification for OWA. The Friedman test revealed a statistically significant difference in performance across the fusion operators ($p < 0.05$). Post-hoc analysis confirmed that OWA's performance was statistically superior to all alternative methods at the $\alpha = 0.05$ level. While the performance gap to Weighted Average was smaller, it remained statistically significant ($p < 0.05$), demonstrating that OWA's dynamic, confidence-based weighting provides a consistent and measurable advantage over static weighting schemes. The larger performance gaps and higher significance levels against simpler methods like stacking, simple average, and majority vote further underscore OWA's critical role in achieving state-of-the-art decoding accuracy.

Table 8: Statistical comparison of different ensemble fusion operators on the DS105 dataset (Ensemble-965 architecture).

Fusion Operator	Accuracy	F1-Score	Statistical Significance (vs. OWA)
OWA	98.62 ± 1.42	98.57 ± 1.48	-
Weighted Average	97.95 ± 1.61	97.88 ± 1.65	0.048
Stacking (Meta-SVM)	97.10 ± 1.88	96.95 ± 1.92	0.044
Simple Average	96.83 ± 2.01	96.70 ± 2.05	0.039
Majority Vote	95.42 ± 2.34	95.25 ± 2.41	0.017

A key innovation of our method is the approximation used in MI calculation, where feature values are thresholded and labels are binarized to accelerate computation and mitigate noise. To validate the necessity of this step, we conducted an ablation study comparing the performance of our full model against two variants on the DS105 dataset (Table 9). The results confirm the critical role of our approximation. Variant A, which used a continuous, non-approximated MI calculation, performed significantly worse. This suggests that the raw, high-dimensional data introduces noise and spurious correlations that degrade the quality of the feature clustering. Variant B showed a substantial improvement, demonstrating that thresholding feature values is the most impactful step, effectively denoising the data and creating a more robust basis for calculating statistical dependencies. Finally, our full model, incorporating both thresholding and label binarization, achieved the highest accuracy. This indicates that the binary label encoding further refines the MI calculation by providing a clearer, more consistent statistical target for the feature clustering algorithm, leading to the creation of more discriminative feature facets.

Table 9: Ablation study on mutual information calculation methods (Mean Accuracy % ± Std on DS105).

Method	Feature Values	Label Encoding	Accuracy (%)
Variant A (No Thresholding)	Continuous	One-Hot	91.23 ± 2.15
Variant B (No Binarization)	Thresholded	One-Hot	95.88 ± 1.54
Proposed Method (Full Model)	Thresholded	Binary	98.62 ± 1.42

The practical deployment of a decoding model depends on its computational demands. We report the average runtime and memory consumption for the complete pipeline (including optimization, training, and testing) per subject in Table 10. Experiments were conducted on a high-performance computing node.

Table 10: Computational performance analysis (averages per subject)

Database	SA Optimization Time (hours)	Model Training Time (minutes)	Peak Memory (GB)
DS105	4.2 ± 0.5	12.3 ± 1.8	3.8
DS107	5.1 ± 0.7	15.6 ± 2.1	4.5
DS116	3.5 ± 0.4	8.9 ± 1.2	2.9

The results indicate that the most computationally intensive phase is the SA optimization, which is a one-time cost per subject to find the optimal configuration. The training time is comparatively modest, and the memory footprint is manageable on standard research computing infrastructure. As expected, the more complex datasets (DS107 with more subjects and DS105 with more classes) required longer optimization and training times. This analysis provides a transparent assessment of the computational resources required to implement the proposed framework.

5 Discussion

Public awareness of BCI has recently increased significantly, turning it into a broadly debated topic in neuroscience [39]. Decoding neural activity in the brain is regarded as a pivotal technology for the expansion of BCIs. This study presented a novel ensemble learning method that directly addresses key limitations in the current state-of-the-art. By moving beyond pre-defined feature subsets through mutual information clustering and replacing manual tuning with a robust global optimization, our framework achieves a new level of automation and performance. Our framework presented a novel ensemble learning method, drawing on feature selection, mutual information, hierarchical clustering, and SVMs for visual brain decoding. 2 different schemes based on Harvard-Oxford (Ensemble-96) and Talairach (Ensemble-965) atlases were tested on three fMRI databases. All investigated frameworks achieved recognition accuracy rates above 95%, indicating the excellent ability of the proposed method for visual brain decoding. Meanwhile, the Ensemble-965 model yielded better outcomes than the Ensemble-96 owing to providing more anatomical details in the model.

The proposed methodology enhances visual object recognition from fMRI data by meticulously selecting and optimizing features, clustering them into meaningful facets, and leveraging an ensemble of classifiers to enhance performance. Each step, from preprocessing to ensemble decision-making, is pivotal in improving the accuracy and robustness of the final model, demonstrating a clear and systematic strategy for addressing this complex issue. The superior performance of our ensemble framework can be attributed to its nuanced approach to feature space decomposition and optimization. While simpler models like standard SVMs provide a strong baseline, our method systematically addresses the high-dimensionality and complexity of fMRI data through mutual information clustering and a globally-optimized ensemble structure. The assessed framework in this work employs a sophisticated methodology designed to enhance the accuracy and efficiency of visual object recognition from fMRI data. Firstly, by applying statistical methods to select significant voxels, the dimensionality of the data is reduced, which in turn helps in reducing computational complexity and improving model accuracy [40]. This foundational step is crucial, but our method advances beyond standard practices. While other methods like L1-regularized SVM [47] induce sparsity, they often discard weakly predictive yet collectively informative voxels. In contrast, our mutual information-based clustering preserves these voxels by grouping them into coherent facets, creating a richer substrate for the ensemble. Simulated annealing helped in exploring the parameter space efficiently to find near-optimal solutions. This stochastic optimization technique helps in finding the global optimum by avoiding local minima, thereby setting the best parameters for feature facets, fusion operators, and classifiers [41]. This represents a significant advantage over methods like the Multi-Objective Cognitive Model (MOCM) [43] or hyperalignment [44], which rely on more constrained optimization or alignment procedures. The ability of SA to co-optimize discrete and continuous parameters globally is a key reason our model adapts so effectively to the distinct characteristics of each dataset, as seen in the consistently high accuracy across DS105, DS107, and DS116.

By clustering features based on mutual information, the methodology ensures that each cluster displays an independent and informative feature facet. This step reduces redundancy and focuses on unique information that improves classification performance [42]. This data-driven decomposition is a primary differentiator. Unlike methods that use pre-defined anatomical atlases or random subspaces, our approach directly tailors the feature space to the statistical structure of the neural code for the task at hand. This explains the performance gap with methods like anatomical connectivity [45], which, while insightful, cannot achieve the same level of functional specificity for decoding.

Moreover, separate training allows the model to learn diverse patterns and relationships within each feature cluster, enhancing the ensemble's ability to generalize across different visual object categories. In addition, combining multiple classifiers through an ensemble method typically yields better performance than individual classifiers by leveraging their collective strengths. The final critical element is our fusion strategy. The selection of the Ordered Weighted Averaging (OWA) operator by the SA algorithm allows for a dynamic, confidence-weighted fusion of decisions. This is more sophisticated than the fixed fusion rules common in other ensembles and enables our model to weigh the opinion of the most reliable expert classifier for any given input more

heavily, leading to the observed improvement in final accuracy.

The juxtaposition of the results of the proposed method with the earlier brain decoding techniques of visual items is given in Table 11. The main discrepancy between the proposed method and other approaches is in the feature space segmentation process, with the help of mutual information between label-feature combinations. The proposed method uses the calculation of mutual information based on the approximate approach, where the labels are encoded in binary and the features are converted to 0 and 1 values based on the application of a threshold. The optimal values of this threshold and the count of feature facets are determined automatically with the help of the simulated annealing optimization algorithm.

Table 11: Juxtaposition of the precision rates of the proposed method with the earlier brain decoding methods of visual items.

Reference	Algorithm	DS105	DS107	DS116
[43]	Multi-Objective Cognitive Model (MOCM) with Gaussian kernel	98.34±0.29	96.79±0.59	97.09±0.33
[44]	Hyperalignment and SVM	87.03±2.87	84.01±1.56	74.62±1.84
[45]	Anatomical connectivity of individual gray-matter voxels	90.82±1.87	85.62±1.95	78.91±2.04
[46]	Hierarchical heterogeneous PSO and SVM	94.46±1.23	89.91±1.67	96.03±0.56
[47]	L1-regularized SVM	85.29±3.49	81.25±3.62	69.24±3.28
[48]	Ensemble multiview learning method based on SVM	98.10	96.30	99.00
[49]	Feature subspaces and multiclass SVM	92.00	93.00	66.00
Proposed Ensemble-96	Ensemble SVM learning with Harvard-Oxford Atlas	97.91±1.15	95.14±1.21	98.12±0.94
Proposed Ensemble-965	Ensemble SVM learning with Talairach Atlas	98.62±1.42	97.54±1.26	99.43±1.08

In Table 11, it is worth noting that this exploration focused solely on the visual scenario of DS116, which was deployed and reviewed, and identified two visual items: the green and red circles. This is even though most of the earlier methods have attempted to separate the visual and auditory modes from each other, and the detection precisions listed in Table 4 are related to the determination of visual and auditory items. For instance, Yousefnezhad and Zhang recommended an MOCM with a Gaussian kernel for multivariate pattern analysis, achieving 97.09% accuracy for the DS116 database [43]. Guntupalli et al. [44] introduced a linear scheme that captures detailed differences among population responses in the human cortex by using response-tuning basis functions shared across individuals, while also modeling cortical patterns of neural responses with individual-specific topographic basis functions. Through a novel algorithm called searchlight hyperalignment and utilizing diverse, dynamic stimuli encompassing visual, auditory, and social percepts, they established a unified model space for the entire cortex. This model effectively aligns representations across multiple brains in various cortical areas, including parietal, occipital, temporal, and prefrontal cortices, as evidenced by between-subject multivariate pattern categorization and intersubject correlation of representational geometry. Using an SVM classifier, this sophisticated model achieved 74.62% accuracy for the DS116 database. Osher et al. [45] suggested that just by examining the anatomical connectivity of specific gray-matter voxels using diffusion-weighted imaging, it is possible to anticipate how individual subjects will respond to visual categories in fMRI. This approach not only explains the functional differences across the cortex but also accommodates the individual variations among subjects. However, they achieved a low accuracy rate of 78.91% for the DS116 database. As displayed in Table 4, the proposed Ensemble-965 scheme yielded better outcomes than the earlier approaches. The proposed Ensemble-96 even had better outcomes than most previous approaches. This displays the superiority of our scheme over previous methods. One of the possible reasons for the benefit of the proposed method compared to previous techniques is the approach adopted to select effective voxels. In the earlier approaches, a specific area of interest was used for each database. However, in this paper, the GLM statistical model and z-score maps were used to find effective voxels. This significantly reduced the size of the input voxel space in our method and greatly increased the learning speed. However, like many other frameworks, the proposed scheme has limitations that must be tackled in future explorations. Although the Ensemble-965 framework produced very good outcomes, it has a high computational complexity that can limit its application, especially in BCI systems.

6 Conclusion

In this article, a new multimodal learning scheme was proposed to resolve the problem of decoding visual objects from brain images. Separation of feature space was done based on hierarchical clustering of mutual information of features and labels. In the current ensemble approach, fusion takes place at the decision-making level, and compared to the fusion approaches at the feature level, it has more speed and simplicity. The automatic separation of the feature space based on the approximate approach of mutual information calculation is an innovation in this

article. The proposed scheme in this study resulted in a improvement in the accuracy of brain decoding of visual objects compared to previous methods. However, the proposed scheme needs further validation using more fMRI databases. In addition, owing to the importance of the mutual information calculation method in selecting effective voxels and distinguishing the feature space, further study is needed to improve the current method. Furthermore, in future studies, more efficient strategies can be used to determine the weights of each single classifier in the ensemble process and test other classifiers besides SVM.

Availability of data and materials

Data is accessible upon request.

Authors' contributions

The investigators contributed to the exploration's conception and construction. Data gathering, modeling, and appraisal were executed by "Lina Gong". The initial draft of the manuscript was written by Lina Gong, who noted on earlier versions of the manuscript.

Ethical Approval

The investigation received ethical review from the IRB, guaranteeing the following of ethical principles and protecting participants' rights.

References

[1] A. Khaleghi, M. R. Mohammadi, K. Shahi, and A. M. Nasrabadi, "Computational neuroscience approach to psychiatry: a review on theory-driven approaches," *Clinical Psychopharmacology and Neuroscience*, Korean College of Neuropsychopharmacology, vol. 20, no. 1, p. 26, 2022. https://doi.org/10.9758/cpn.2022.20.1.26

[2] B. Du, X. Cheng, Y. Duan, H, "Ning. fmri brain decoding and its applications in brain–computer interface: A survey," *Brain Sciences*, vol. 12, no. 2, p. 228, 2022. https://doi.org/10.3390/brainsci12020228

[3] M. Saeidi, W. Karwowski, FV. Farahani, K. Fiok, PA. Hancock, BD. Sawyer, L. Christov-Moore, PK. Douglas, "Decoding task-based fMRI data with graph neural networks, considering individual differences," *Brain Sciences*, vol. 12, no. 8, p. 1094, 2022. https://doi.org/10.3390/brainsci12081094

[4] R. Jabakhanji, AD. Vigotsky, J. Bielefeld, L. Huang, MN. Baliki, G. Iannetti, AV. Apkarian, "Limits of decoding mental states with fMRI," *Cortex*, vol. 149, pp. 101-22, 2022. https://doi.org/10.1016/j.cortex.2021.12.015

[5] W. Xiao, G. Manyi, and A. Khaleghi, "Deficits in auditory and visual steady-state responses in adolescents with bipolar disorder," *J Psychiatr Res*, Elsevier, vol. 151, pp. 368–376, 2022. https://doi.org/10.1016/j.jpsychires.2022.04.041

[6] S. L. Warren and A. A. Moustafa, "Functional magnetic resonance imaging, deep learning, and Alzheimer's disease: A systematic review," *Journal of Neuroimaging*, Wiley, vol. 33, no. 1, pp. 5–18, 2023. https://doi.org/10.1111/jon.13063

[7] T. Varkevisser, E. Geuze, M. A. Van Den Boom, K. Kouwer, J. Van Honk, and R. Van Lutterveld, "Pattern classification based on the amygdala does not predict an individual's response to emotional stimuli," *Hum Brain Mapp*, Wiley, vol. 44, no. 12, pp. 4452–4466, 2023. https://doi.org/10.1002/hbm.26391

[8] C. Wang *et al.*, "'When'and 'what'did you see? A novel fMRI-based visual decoding framework," *J Neural Eng*, IOP Publishing, vol. 17, no. 5, p. 056013, 2020. DOI: 10.1088/1741-2552/abb691

[9] S. Huang, W. Shao, M.-L. Wang, and D.-Q. Zhang, "fmri-based decoding of visual information from human brain activity: A brief review," *International Journal of Automation and Computing*, Springer, vol. 18, no. 2, pp. 170–184, 2021. https://doi.org/10.1007/s11633-020-1263-y

[10] A. M. Golestani and J. J. Chen, "Performance of temporal and spatial independent component analysis in identifying and removing low-frequency physiological and motion effects in resting-state fMRI," *Front Neurosci*, Frontiers Media, vol. 16, p. 867243, 2022. https://doi.org/10.3389/fnins.2022.867243

[11] E. N. Castanho, H. Aidos, and S. C. Madeira, "Biclustering fMRI time series: a comparative study," *BMC Bioinformatics*, BioMed Central (Springer), vol. 23, no. 1, p. 192, 2022. https://doi.org/10.1186/s12859-022-04733-8

[12] L. Liu, C. Hua, Z. Cheng, and Y. Ji, "Intelligent diagnosis method of MRI brain image using parallel self-organizing feature maps neural network," *J Med Imaging Health Inform*, SPIE / American Scientific Publishers, vol. 11, no. 2, pp. 487–496, 2021. https://doi.org/10.1166/jmihi.2021.3285

[13] G. G. Knyazev, A. N. Savostyanov, P. D. Rudych, and A. V Bocharov, "Multi-voxel pattern analysis of fMRI data during self-and other-referential processing," *Zhurnal Vysshei Nervnoi Deyatelnosti Imeni IP Pavlova*, Pavlov Russian Medical Academy (Elsevier partner), vol. 73, no. 2, pp. 242–255, 2023. https://doi.org/10.31857/S0044467723020065

[14] J. I. Glaser, A. S. Benjamin, R. H. Chowdhury, M. G. Perich, L. E. Miller, and K. P. Kording, "Machine learning for neural decoding," *eNeuro*, Society for Neuroscience (SfN), vol. 7, no. 4, 2020. https://www.eneuro.org/content/7/4/ENEURO.0506-19.2020.abstract

[15] AW. Thomas, C. Re, RA. Poldrack, "Interpreting mental state decoding with deep learning models," *Trends in Cognitive Sciences*, vol. 26, no. 11, pp. 972-86, 2022. https://www.cell.com/trends/cognitive-sciences/abstract/S1364-6613(22)00160-7

[16] W. A. Campos-Ugaz, J. P. P. Garay, O. Rivera-Lozada, M. A. A. Diaz, D. Fuster-Guillén, and A. A. T. Arana, "An overview of bipolar disorder

diagnosis using machine learning approaches: clinical opportunities and challenges," *Iran J Psychiatry*, Iranian Psychiatric Association, vol. 18, no. 2, p. 237, 2023. https://doi.org/10.18502/ijps.v18i2.12372

[17] R. M. Awangga, T. L. R. Mengko, and N. P. Utama, "A literature review of brain decoding research," in *IOP Conference Series: Materials Science and Engineering*, IOP Publishing, 2020, p. 032049. DOI: 10.1088/1757-899X/830/3/032049

[18] C. Du, K. Fu, J. Li, and H. He, "Decoding visual neural representations by multimodal learning of brain-visual-linguistic features," *IEEE Trans Pattern Anal Mach Intell*, IEEE, vol. 45, no. 9, pp. 10760–10777, 2023. https://doi.org/10.1109/TPAMI.2023.3263181

[19] A. Floren, B. Naylor, R. Miikkulainen, and D. Ress, "Accurately decoding visual information from fMRI data obtained in a realistic virtual environment," *Front Hum Neurosci*, Frontiers Media, vol. 9, p. 327, 2015. https://doi.org/10.3389/fnhum.2015.00327

[22] M. Graumann, C. Ciuffi, K. Dwivedi, G. Roig, and R. M. Cichy, "The spatiotemporal neural dynamics of object location representations in the human brain," *Nat Hum Behav*, Nature Publishing Group, vol. 6, no. 6, pp. 796–811, 2022. https://doi.org/10.1038/s41562-022-01302-0

[23] O. Hourani, N. M. Charkari, and S. Jalili, "Voxel selection framework based on meta-heuristic search and mutual information for brain decoding," *Int J Imaging Syst Technol*, Wiley, vol. 29, no. 4, pp. 663–676, 2019. https://doi.org/10.1002/ima.22353

[25] C. Chen, S. T. Weiss, and Y.-Y. Liu, "Graph convolutional network-based feature selection for high-dimensional and low-sample size data," *Bioinformatics*, Oxford University Press, vol. 39, no. 4, p. btad135, 2023. https://doi.org/10.1093/bioinformatics/btad135

[26] B. Du, X. Cheng, Y. Duan, and H. Ning, "fmri brain decoding and its applications in brain–computer interface: A survey," *Brain Sci*, MDPI, vol. 12, no. 2, p. 228, 2022. https://doi.org/10.3390/brainsci12020228

[27] J. Chung and J. Teo, "Single classifier vs. ensemble machine learning approaches for mental health prediction," *Brain Inform*, Springer, vol. 10, no. 1, p. 1, 2023. https://doi.org/10.1186/s40708-022-00180-6

[28] Z. Ye, Y. Qu, Z. Liang, M. Wang, and Q. Liu, "Explainable fMRI-based brain decoding via spatial temporal-pyramid graph convolutional network," *Hum Brain Mapp*, Wiley, vol. 44, no. 7, pp. 2921–2935, 2023. https://doi.org/10.1002/hbm.26255

[29] L. Yang, Y. Du, W. Yang, and J. Liu, "Machine learning with neuroimaging biomarkers: Application in the diagnosis and prediction of drug addiction," *Addiction Biology*, Wiley, vol. 28, no. 2, p. e13267, 2023. https://doi.org/10.1111/adb.13267

[30] Y. Qi, B. Liu, Y. Wang, and G. Pan, "Dynamic ensemble modeling approach to nonstationary neural decoding in brain-computer interfaces," *Adv Neural Inf Process Syst*, NeurIPS Foundation, vol. 32, 2019.

[31] H. Li and Y. Fan, "Brain decoding from functional MRI using long short-term memory recurrent neural networks," in *International conference on medical image computing and computer-assisted intervention*, Springer / MICCAI Society, 2018, pp. 320–328. https://doi.org/10.1007/978-3-030-00931-1_37

[32] P. Wen, L. W. Thompson, A. Rosenberg, M. S. Landy, and B. Rokers, "Single-Trial fMRI Decoding of 3D Motion with Stereoscopic and Perspective Cues," *Journal of Neuroscience*, vol. 45, no. 22, 2025.

[33] C. Li et al., "Enhancing cross-subject fmri-to-video decoding with global-local functional alignment," in *European Conference on Computer Vision*, Springer, 2024, pp. 353–369. https://doi.org/10.1007/978-3-031-73010-8_21

[34] Y. Liang, K. Bo, S. Meyyappan, and M. Ding, "Decoding fMRI data with support vector machines and deep neural networks," *J Neurosci Methods*, vol. 401, p. 110004, 2024. https://doi.org/10.1016/j.jneumeth.2023.110004

[35] M. Yousefnezhad, A. Selvitella, L. Han, and D. Zhang, "Supervised hyperalignment for multisubject fMRI data alignment," *IEEE Trans Cogn Dev Syst*, IEEE, vol. 13, no. 3, pp. 475–490, 2020. https://doi.org/10.1109/TCDS.2020.2965981

[36] A. Bowring, C. Maumet, and T. E. Nichols, "Exploring the impact of analysis software on task fMRI results," *Hum Brain Mapp*, Wiley, vol. 40, no. 11, pp. 3362–3384, 2019. https://doi.org/10.1002/hbm.24603

[37] R. J. Rushmore, S. Bouix, M. Kubicki, Y. Rathi, E. Yeterian, and N. Makris, "HOA2. 0-ComPaRe: A next generation Harvard-Oxford Atlas comparative parcellation reasoning method for human and macaque individual brain parcellation and atlases of the cerebral cortex," *Front Neuroanat*, Frontiers Media, vol. 16, p. 1035420, 2022. https://doi.org/10.3389/fnana.2022.1035420

[38] O. Esteban *et al.*, "Analysis of task-based functional MRI data preprocessed with fMRIPrep," *Nat Protoc*, Nature Publishing Group, vol. 15, no. 7, pp. 2186–2202, 2020. https://doi.org/10.1038/s41596-020-0327-3

[39] N. Shi *et al.*, "Estimating and approaching the maximum information rate of noninvasive visual brain-computer interface," *Neuroimage*, Elsevier, vol. 289, p. 120548, 2024. https://doi.org/10.1016/j.neuroimage.2024.120548

[40] Z. Wen, T. Yu, Z. Yu, and Y. Li, "Grouped sparse Bayesian learning for voxel selection in multivoxel pattern analysis of fMRI data," *Neuroimage*, Elsevier, vol. 184, pp. 417–430, 2019. https://doi.org/10.1016/j.neuroimage.2018.09.031

[41] E. A. Mahareek, A. S. Desuky, and H. A. El-Zhni, "Simulated annealing for SVM parameters optimization in student's performance prediction," *Bulletin of Electrical Engineering and Informatics*,

Institute of Advanced Engineering and Science (IAES), vol. 10, no. 3, pp. 1211–1219, 2021. https://doi.org/10.11591/eei.v10i3.2855

[42] S. Hao, Y. Cui, and J. Wang, "Segmentation scale effect analysis in the object-oriented method of high-spatial-resolution image classification," *Sensors*, MDPI, vol. 21, no. 23, p. 7935, 2021. https://doi.org/10.3390/s21237935

[43] M. Yousefnezhad and D. Zhang, "Multi-objective cognitive model: a supervised approach for multi-subject fmri analysis," *Neuroinformatics*, Springer, vol. 17, no. 2, pp. 197–210, 2019. https://doi.org/10.1007/s12021-018-9394-9

[44] J. S. Guntupalli, M. Hanke, Y. O. Halchenko, A. C. Connolly, P. J. Ramadge, and J. V Haxby, "A model of representational spaces in human cortex," *Cerebral cortex*, Oxford University Press, vol. 26, no. 6, pp. 2919–2934, 2016. https://doi.org/10.1093/cercor/bhw068

[45] D. E. Osher, R. R. Saxe, K. Koldewyn, J. D. E. Gabrieli, N. Kanwisher, and Z. M. Saygin, "Structural connectivity fingerprints predict cortical selectivity for multiple visual categories across cortex," *Cerebral cortex*, Oxford University Press, vol. 26, no. 4, pp. 1668–1683, 2016. https://doi.org/10.1093/cercor/bhu303

[46] X. Ma, C.-A. Chou, H. Sayama, and W. A. Chaovalitwongse, "Brain response pattern identification of fMRI data using a particle swarm optimization-based approach," *Brain Inform*, Springer, vol. 3, no. 3, pp. 181–192, 2016. https://doi.org/10.1007/s40708-016-0049-z

[47] H. Mohr, U. Wolfensteller, S. Frimmel, and H. Ruge, "Sparse regularization techniques provide novel insights into outcome integration processes," *Neuroimage*, Elsevier, vol. 104, pp. 163–176, 2015. https://doi.org/10.1016/j.neuroimage.2014.10.025

[48] O. Hourani, N. Moghadam Charkari, and S. Jalili, "An Ensemble Multiview learning method for visual object decoding from fMRI brain data," *Signal and Data Processing*, vol. 18, no. 3, pp. 109–126, 2021. http://dx.doi.org/10.52547/jsdp.18.3.109

[49] O. Hourani, N. M. Charkari, and S. Jalili, "New Insight of Human Brain Connectivity Mapping based on Inter-correlation of Multi-view fMRI Decoder," *Authorea Preprints*, 2023.