



Decoding fMRI data with support vector machines and deep neural networks

Yun Liang^a, Ke Bo^b, Sreenivasan Meyyappan^c, Mingzhou Ding^{a,*}

^a J. Crayton Pruitt Family Department of Biomedical Engineering, University of Florida, Gainesville, FL, USA

^b The Cognitive and Affective Neuroscience Lab, Dartmouth College, Hanover, NH, USA

^c Center for Mind and Brain, University of California, Davis, CA, USA

ARTICLE INFO

Keywords:

fMRI
Multivariate pattern analysis
Support vector machine
Convolutional neural network
Spatial attention
Emotion processing

ABSTRACT

Background: Multivoxel pattern analysis (MVPA) examines fMRI activation patterns associated with different cognitive conditions. Support vector machines (SVMs) are the predominant method in MVPA. While SVM is intuitive and easy to apply, it is mainly suitable for analyzing data that are linearly separable. Convolutional neural networks (CNNs) are known to have the ability to approximate nonlinear relationships. Applications of CNN to fMRI data are beginning to appear with increasing frequency, but our understanding of the similarities and differences between CNN models and SVM models is limited.

New method: We compared the two methods when they are applied to the same datasets. Two datasets were considered: (1) fMRI data collected from participants during a cued visual spatial attention task and (2) fMRI data collected from participants viewing natural images containing varying degrees of affective content.

Results: We found that (1) both SVM and CNN are able to achieve above-chance decoding accuracies for attention control and emotion processing in both the primary visual cortex and the whole brain, (2) the CNN decoding accuracies are consistently higher than that of the SVM, (3) the SVM and CNN decoding accuracies are generally not correlated, and (4) the heatmaps derived from SVM and CNN are not significantly overlapping.

Comparison with existing methods: By comparing SVM and CNN we pointed out the similarities and differences between the two methods.

Conclusions: SVM and CNN rely on different neural features for classification. Applying both to the same data may yield a more comprehensive understanding of neuroimaging data.

1. Introduction

Functional magnetic resonance imaging (fMRI) exploits blood-oxygen-level-dependent (BOLD) signals to map neural activities (Mahmoudi et al., 2012). Conventional analyses of fMRI data using methods such as the general linear model (GLM) to compare BOLD activities evoked by different cognitive conditions at the single voxel level or at the level of a region of interest (ROI) to gain insights into the neural basis of cognitive functions. These conventional analyses, referred to as the univariate approach here, have yielded much of our current understanding of the neural underpinnings of human cognition. Like any methods, the univariate approach to fMRI data analysis has limitations. For example, for a voxel to be considered activated by a cognitive event, it has to be consistently activated in a population of participants, which is often difficult to achieve given the idiosyncratic nature of voxel-level

neural responses to subtle experimental manipulations and can thus lead to failure to detect the presence of neural signals; see, e.g., (Haxby et al., 2011). At the ROI level, univariate analysis entails averaging BOLD responses across voxels, which disregards the information contained in the heterogeneous response patterns, again reducing the sensitivity of our ability to detect the presence of neural signals that exhibit differential neural response patterns but not overall changes in ROI-level response magnitude.

Multivoxel pattern analysis (MVPA), which can be performed at the single subject level and takes into account the spatial variations of the BOLD activity across voxels, overcomes the limitations of the univariate approach and have become the main approach for providing information that complements the univariate approach (Haynes, 2015; Lewis-Peacock and Norman, 2014; Mitchell et al., 2003; Norman et al., 2006; Tagliazucchi and Laufs, 2014; Tagliazucchi et al., 2012).

* Corresponding author.

E-mail address: mding@bme.ufl.edu (M. Ding).

<https://doi.org/10.1016/j.jneumeth.2023.110004>

Received 4 July 2023; Received in revised form 21 October 2023; Accepted 27 October 2023

Available online 31 October 2023

0165-0270/© 2023 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Currently, the support vector machine (SVM) (Sain, 1996) is the most widely used MVPA method for analyzing fMRI data (Bonnici et al., 2013; LaConte et al., 2005; Song et al., 2011). When decoding between different cognitive conditions using SVM, a classifier is trained on training data and tested on testing data, and above-chance level decoding accuracy is taken as evidence of the involvement of the ROI in the cognitive function being tested. The MVPA literature has grown significantly in recent years (Baucom et al., 2012; Kotz et al., 2013; Said et al., 2010; Sitaram et al., 2011). Novel insights not possible with traditional univariate methods have emerged in all fields of cognitive neuroscience. For example, we used the MVPA approach to decode neural responses to natural images containing varying degrees of emotional content and reveal the existence of affect-specific neural representations in the retinotopic visual cortex (Bo et al., 2021). In addition to decoding accuracy, the heatmaps (or weight maps) are another SVM construct that can be derived from the classifier, which provide information on the contribution of different voxels to the classifier performance. Despite their clear importance for developing cognitive neural models, to date, heatmaps are far less utilized than decoding accuracy. We have used SVM heatmaps to uncover the microstructures underlying the control of spatial vs feature attentional in the dorsal attention network (DAN) (Rajan et al., 2021). Despite SVM's successes, the linear nature of the method is both a strength and a weakness. Being linear, SVM is intuitive to understand and computationally efficient, which is a strength. On the other hand, being linear, SVM is effective only when the data is linearly separable. Recent studies suggest that the mapping between neural activity and cognitive functions may be nonlinear (Birn et al., 2001) and as such linear SVM may not be able to effectively characterize the underlying brain patterns (Farahani et al., 2019). Another known weakness of SVM is its limited ability for handling high-dimensional data; the need for expert feature selection to reduce data dimensionality may bias the results (Vieira et al., 2017).

The advent of AI-inspired methods such as deep neural networks (DNNs) has the potential to overcome SVM's limitations and provide information that complements the SVM. DNNs, especially the Convolutional Neural Networks (CNNs), have emerged as a technique for analyzing multivariate neuroimaging data and yielded insights not possible with other methods (Abrol et al., 2021; Sarraf et al., 2016). Similar to SVMs, when used to decode different experimental conditions, a CNN model is trained on training data and tested on testing data. Both decoding accuracy and heatmaps can be derived from the trained CNN models. Our recent application of CNN models to fMRI data recorded from patients suffering from trigeminal neuralgia revealed novel insights into the generation and perception of TN pain that were not possible with other methods (Liang et al., 2023).

Given the prevalence of SVM and the emerging significance of CNN in MVPA analysis of neuroimaging data, we consider it both necessary and timely to compare CNNs and SVMs by applying them to the same datasets and reveal the similarities and differences between the two methods. For this purpose, two well-characterized datasets were considered: (1) fMRI data recorded from participants performing a cued spatial attentional task (attention dataset) and (2) fMRI data recorded from participants viewing natural images containing varying levels of affective content (emotion dataset). For both datasets, the number of participants is twenty ($N = 20$), which is typical of many fMRI studies of cognitive neuroscience problems (Desmond and Glover, 2002). Recognizing that the data from a single subject is not enough to train a robust CNN model, we have adopted an across-subject approach, in which fMRI data from 19 subjects were pooled together to train a CNN model which was then tested on the 20th subject; repeating the process 20 times yielded robust decoding accuracies and heatmaps. The SVM decoding followed the same across-subject approach to facilitate comparison. The main quantities of interest are the decoding accuracy and the heatmap. Two hypotheses were tested: (1) CNN will achieve higher decoding accuracy than SVM and (2) the heatmaps from the two methods differ,

namely, CNN and SVM rely on different neural information for decoding and classification.

2. Materials and methods

Two fMRI datasets were used to compare SVM and CNN decoding. To facilitate the comparison, recording, preprocessing, and analyses were carried out identically for both datasets. We note that these datasets have been used in previous publications to address different questions (Bo et al., 2021; Meyyappan et al., 2021; Rajan et al., 2021).

2.1. The attention dataset

The experimental protocol was approved by the Institutional Review Board of the University of Florida. Twenty right-handed healthy volunteers (15 men and 5 women; mean age: 24.65 ± 2.87 years) with normal or corrected-to-normal vision and no history of neurological or psychological disorders provided written informed consent and participated in the study. This dataset has been used in prior publications to address different questions (Meyyappan et al., 2021).

The participants performed a cued visual spatial/feature attention experiment while simultaneous EEG-fMRI were recorded (only the fMRI data were considered here). The start of each trial was signaled by an auditory cue, which directed the participants to direct their attention to either a spatial location (left or right) or a color (red or green). Following a delay period, varied randomly from 3000 to 6600 ms, two colored rectangles (red or green) were presented for a duration of 200 ms, with one in each of the two peripheral locations. The subject's task was to report the orientation of the rectangle (target) appearing at the cued location or having the cued color. For feature (color) trials, the two rectangles displayed were always of the opposite color; for spatial trials, the two rectangles were either of the same color or the opposite color, see Fig. 1A (Meyyappan et al., 2021). For this study, we mainly focused on the cue-related activity, reflecting preparatory or anticipatory attention, in the spatial attention conditions (Meyyappan et al., 2021); the feature attention conditions were not considered.

2.2. The emotion dataset

The experimental protocol was approved by the Institutional Review Board of the University of Florida. A total of 26 healthy volunteers with normal or corrected-to-normal vision gave written informed consent and participated in the study. Two participants withdrew from the experiment. Four additional participants were discarded due to excessive movements inside the scanner. The data from the remaining 20 participants were analyzed and reported here (10 men and 10 women; mean age: 20.4 ± 3.1 years) (Bo et al., 2021).

The participants viewed 60 gray-scaled pictures including 20 pleasant, 20 unpleasant, and 20 neutral pictures from the International Affective Picture System (IAPS) library while simultaneous EEG-fMRI was recorded (only the fMRI data were considered here (Lang et al., 1997)). There were 5 runs. The 60 pictures were presented in random order in each run, see Fig. 1B. The participants passively viewed the images in the scanner. No response was required from them. For this study, we mainly focused on the unpleasant pictures and neutral pictures.

2.3. fMRI acquisition and preprocessing

For both datasets, the functional MRI data were collected on a 3 T Philips Achieva scanner (Philips Medical Systems), with the following parameters: echo time, 30 ms; repetition time, 1.98 s; flip angle, 80° ; slice number, 36; field of view (FOV), 224 mm; voxel size, $3.5 \times 3.5 \times 3.5$ mm; matrix size, 64×64 . Preprocessing was carried out using the statistical parametric mapping toolbox (SPM) and custom scripts written in MATLAB, including the following steps: slice timing

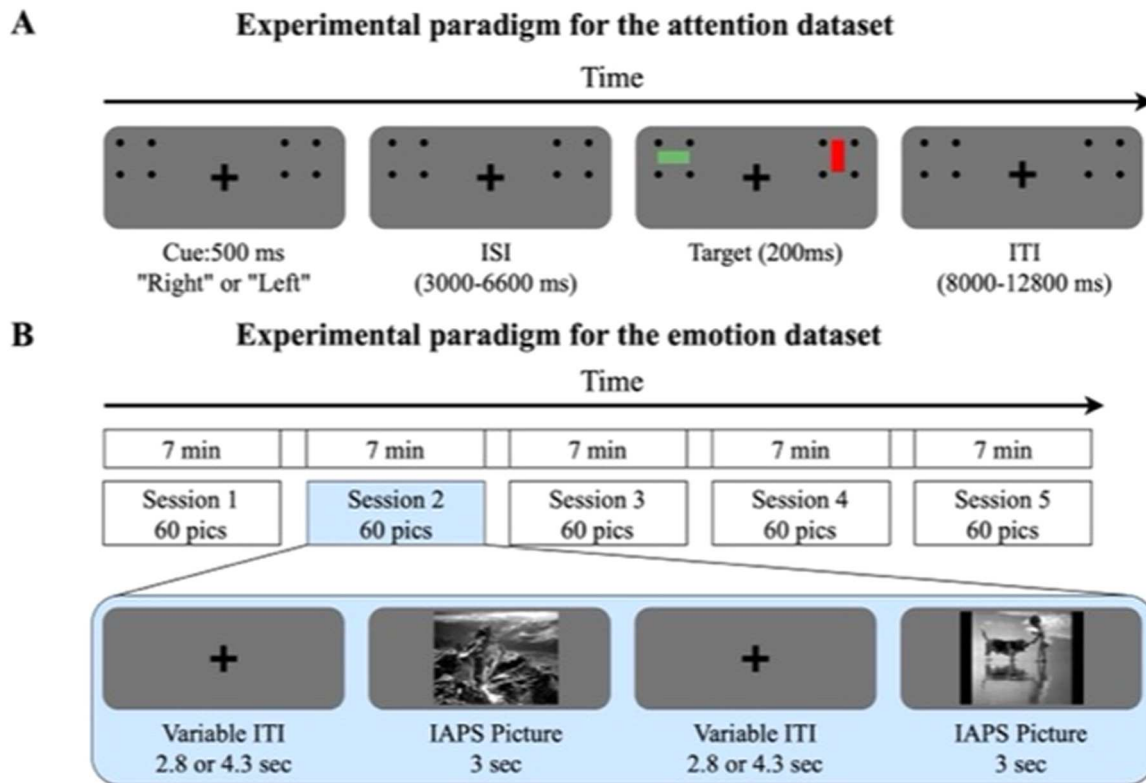


Fig. 1. Experimental paradigms. (A) Attention dataset. Each trial begins with an auditory cue lasting 500 ms, instructing the subject to covertly attend to a spatial location (left or right). After a variable cue-to-target delay ranging from 3000 to 6600 ms, two colored rectangles are displayed for 200 ms, one in each of the two peripheral locations. Participants are then asked to report the orientation of the rectangle (horizontal or vertical) appearing in the cued location. An intertrial interval (ITI) follows, randomly varied from 8000 to 12,800 ms after the target onset, before the start of the next trial. (B) Emotion dataset. The experiment consisted of five sessions, each lasting 7 min. In each session, a total of 60 International Affective Picture System (IAPS) pictures, including 20 pleasant, 20 unpleasant, and 20 neutral, were presented in random order. Each picture was displayed for 3 s, followed by a fixation period referred to as inter-trial interval (ITI) lasting 2.8 or 4.3 s. Participants were required to fixate on the cross in the center of the screen throughout the session.

correction, realignment, spatial normalization, and smoothing. Slice timing correction was conducted using sinc interpolation to correct for differences in slice acquisition time within an EPI volume. The images were then spatially realigned to the first image of each session by a 6-parameter rigid-body spatial transformation to account for head movement during data acquisition. Each participant's images were then normalized and registered to the MNI space. All images were further resampled to a voxel size of $3 \times 3 \times 3$ mm and spatially smoothed using a Gaussian kernel with 7 mm FWHM. The preprocessed signal was further passed through a high-pass filter with a cutoff frequency set at 1/128 Hz to attenuate lower frequency noise.

2.4. Regions of interest (ROIs)

For both datasets, the stimuli are visual, and as such the primary visual cortex (V1) plays an essential role in extracting and processing basic visual features of the stimuli. It is also known that V1 is subject to influence and modulation of higher-order visual structures depending on the task conditions. We first performed the comparison between SVM and CNN in primary visual cortex (V1). The goal was to examine whether SVM and CNN exhibit comparable classification performance and make classifications of cognitive states based on similar visual features. V1v (V1 ventral) and V1d (V1 dorsal) in a recently published probabilistic visual retinotopic atlas (Wang et al., 2015) were combined to form the V1 ROI (Fig. 2A). We next performed the comparison between SVM and CNN at the whole brain level because cognitive functions such as attention and emotion engage large-scale distributed brain networks. The goal was to examine whether SVM and CNN exhibit similar classification performance and make classifications of cognitive

states based on similar brain regions. The 129 brain regions in the Lausanne Brain Atlas are combined to form the whole brain ROI (Daducci et al., 2012); see Fig. 2B.

2.5. Univariate analysis with GLM

As a starting point, we first analyzed the data using the univariate approach, in which we adopted the standard GLM method, as implemented in the SPM toolbox. For the attention dataset, eight task-related events were included as regressors in the GLM model, including five regressors to model the cue-related BOLD activity, two regressors to account for BOLD responses evoked by validly and invalidly cued targets separately, and one regressor to model trials with incorrect responses. For the emotion dataset, four task-related events were included as regressors in the GLM analysis, including three stimuli-related regressors and one fixation regressor. For both datasets, the Hemodynamic Response Function (HRF) used in the GLM analysis was the default HRF provided by the SPM toolbox, with a delay of 6 s. At the group level, fMRI activation maps were obtained by applying a parametric one-sample t-test and thresholded at a significance level of $p < 0.05$ after correcting for multiple comparisons using the false discovery rate (FDR) method.

2.6. Single-trial estimation of fMRI-BOLD

Since MVPA is performed at the single-trial level, we applied a beta series regression method to estimate the BOLD response on each trial in every voxel for both attention and emotion datasets (Mumford et al., 2012). Specifically, for both datasets, the trial of interest was

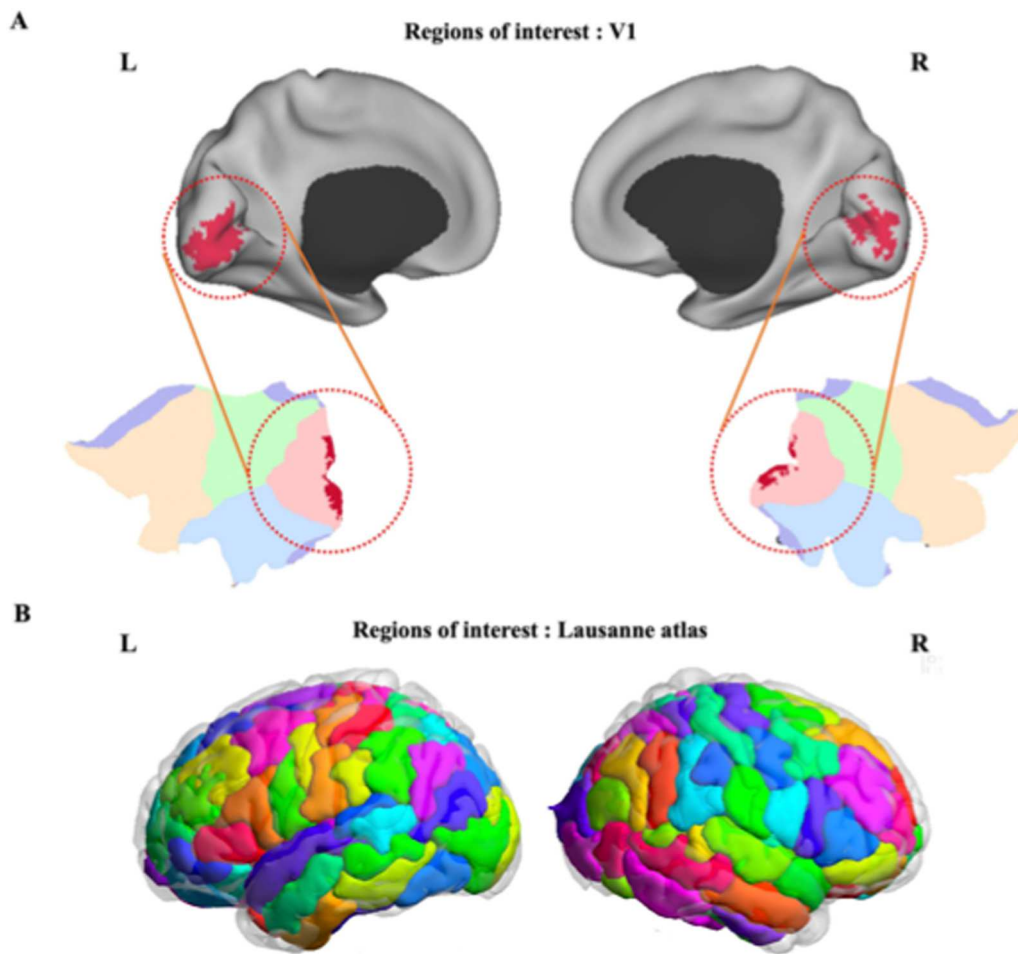


Fig. 2. Regions of interest. (A) V1 ROI. (B) Whole brain ROI based on Lausanne Atlas (129 regions).

represented by one regressor, and all the other trials were represented by another regressor. Six motion regressors were also included to account for any movement-related artifacts during the scan. The process was repeated for all trials. The single-trial beta responses were fed into SVM and CNN for decoding analysis and heatmap generation.

2.7. MVPA: SVM

The SVM algorithm as implemented in the LibSVM (Chang and Lin, 2011) was used to decode cue left vs. right (spatial attention control) for the attention dataset in both the V1 ROI and the whole brain. For the emotion dataset, we decoded the brain patterns evoked by unpleasant vs. neutral stimuli. A leave-one-participant-out across-subject approach was adopted. Specifically, for the 20 participants in each dataset, the data from 19 were used to train a SVM classifier, which was then tested on the remaining participant to obtain decoding accuracy. Here the decoding accuracy is defined by the below formula:

$$\text{Decoding Accuracy} = \frac{\text{Correctly predicted trials}}{\text{Total number of tested trials}}$$

This process was repeated 20 times and the 20 decoding accuracies were averaged to yield the group-level average decoding accuracy. The reason we took the across-subject decoding approach rather than the more conventional within-subject approach is because for training CNN models (see below), larger amount of data were required, which means that the within-subject analysis was not viable for CNN analysis and the across-subject approach needs to be used to increase the size of training data. Although the SVM approach has been mainly applied in a within-

subject fashion, for an equitable comparison, we applied the SVM decoding in this study using the same across-subject decoding approach as the CNN.

In addition to decoding accuracy, the heatmap is another critical aspect of the SVM technique, which can be used to attribute functional significance to voxels inside an ROI. To compute the SVM heatmap inside an ROI, we applied the transformation proposed by Haufe et al. to the weight vectors computed from the SVM (Haufe et al., 2014). The absolute value of the resulting weight in each voxel measures the contribution of the voxel to the decoding performance (Lee et al., 2010; Mourao-Miranda et al., 2005).

2.8. MVPA: CNN

Widely used CNNs such as ResNet (He et al., 2016) take two dimensional (2D) images, such as those from the ImageNet (Deng et al., 2009) or Cifar10 (Krizhevsky and Hinton, 2009), as input and produce recognition of the object in the image as output. But fMRI data from the brain are three dimensional (3D). We thus changed all the 2D convolutional filters in the ResNet18 structure to 3D convolutional filters; see Fig. 3A. The process of 3D convolutions entails the application of 3D convolutional filter to the 3D imaging dataset, with the filter traversing in three orthogonal directions (x, y, and z), to yield low-level feature representations used by subsequent layers for more in depth processing and classification. In addition, to make the comparison with the SVM approach equitable, we modified the kernel size of the first convolutional layer, changing it from the original kernel size of 7 to the kernel size of 1. The detailed filter and dimension size was listed in Table 1. The

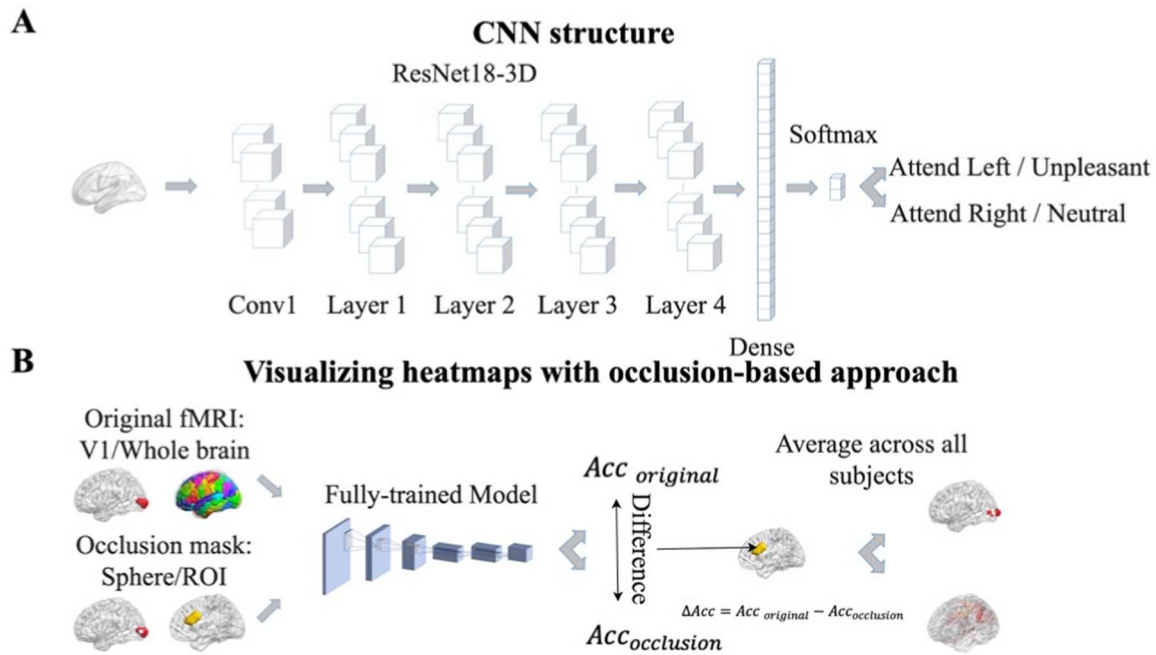


Fig. 3. Pipeline for Convolutional Neural Network (CNN) analysis. (A) CNN model was trained to predict attention control or emotional processing from fMRI data. (B) An occlusion approach was carried out to evaluate the contribution of different brain regions or voxels to CNN model classification performance. The detailed filter size and feature map dimension are listed in Table 1. Acc: accuracy.

Table 1

Detailed structure of the proposed ResNet-18. F is the number of feature channels, and N is the number of blocks in each layer.

Model	Block	Conv1	Layer1		Layer2		Layer3		Layer4		Average Pooling SoftMax
			F	N	F	N	F	N	F	N	
ResNet-18	ResNet	Conv, $1 \times 1 \times 1$, 64, stride 2, padding 3	64	2	128	2	256	2	512	2	

same leave-one-participant-out approach as in SVM analysis above was used here for both datasets. The decoding was also defined in the same way as the SVM decoding accuracy.

For heatmaps, we applied an occlusion approach. Specifically, for V1 decoding, after a CNN model was trained and the decoding accuracy obtained, we removed a sphere with a radius of 6 mm (2 voxels) around each voxel (i.e., change the activity level in the voxels inside the sphere to 0) and fed the fMRI BOLD into the trained CNN network to obtain decoding accuracy. The decrease in decoding accuracy, which is taken as the weight of the voxel in the heatmap, is considered a measure of the significance of the voxel. For whole-brain decoding, after a CNN model was trained and the decoding accuracy obtained, we removed a ROI by setting the activity level in all the voxels in the ROI to 0 and fed the fMRI BOLD into the trained CNN network to obtain decoding accuracy. The decrease in decoding accuracy is taken as a measure of the weight of the ROI in the heatmap, see Fig. 3B. The logic here is that the more the decoding accuracy decreased following occlusion, the more significant the voxel or ROI contributed to the decoding performance, and the more highly weighted the voxel or ROI is in the heatmap.

3. Results

3.1. Attention dataset: decoding attentional control in V1

According to the prevailing theory of attention control, following an attention directing cue, top-down signals from frontoparietal attention control areas propagate to visual cortex to bias the sensory neurons to enhance attended information and suppress distraction (Corbetta et al., 2008; Liu et al., 2003; Wang et al., 2016). To what extent such signals reach the level of the primary visual cortex (V1) is not well established.

Using V1 as the ROI, we decoded spatial attention control (cue left vs. cue right) using SVM and CNN, and compared the decoding accuracies. As shown in Fig. 4A, both SVM and CNN had decoding accuracy significantly above chance level of 50% ($p < 0.0001$), with the CNN decoding accuracy significantly higher than the SVM ($p < 0.002$), suggesting that attention control signals can reach the level of V1 and CNN has stronger abilities to detect these signals. Correlating the decoding accuracies between SVM and CNN, we found that the two decoding accuracies were not significantly correlated ($R=0.12$, $p = 0.62$), suggesting that these two methods may emphasize different data features to achieve their respective decoding performance, see Fig. 4B.

To further examine whether different features drove the decoding performance for the two different kinds of MVPA methods (SVM vs CNN), we compared the top 20% voxels (80 voxels) from the heatmap of SVM and that of CNN; see Fig. 4C. Intuitively the heatmaps look quite different. To quantify the extent of overlap/nonoverlap of two sets of voxels in the two heatmaps, the Jaccard index (JI) was computed, where a JI of 0 or 1 indicates no overlap or total overlap (Levandowsky and Winter, 1971). We found $JI=0.10$ to be the overlap between the top 80 voxels from the SVM heatmap and that from the CNN map. To better understand the meaning of this number, we randomly chose two sets of 80 voxels and computed the JI. Repeating this procedure 10000 times, the average random permutation JI was found to be 0.11 ± 0.025 . Comparing $JI=0.10$ from the actual data with the JI from random permutation yielded no significant difference ($p = 0.65$). We take this as evidence that SVM and CNN rely on different feature information for making classification of the two attentional states. Extending the percent of voxels included in the overlap analysis to [10%,90%], we found that the JI is an increasing function of the percent threshold, as shown in Fig. 4D, and at every threshold, the JI from the data is not significantly

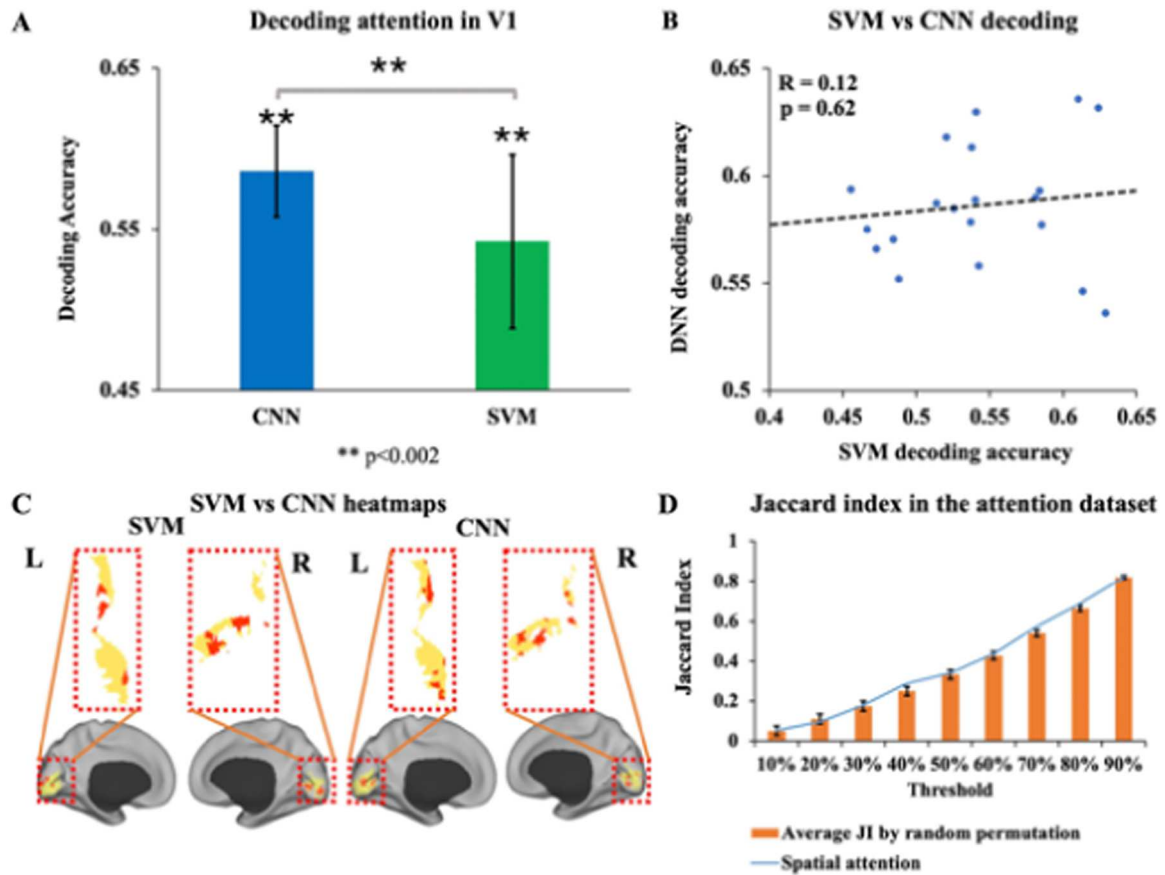


Fig. 4. Decoding attention control in V1. (A) SVM versus CNN decoding. Both decoding accuracies are significantly above chance level ($p < 0.0001$), with CNN decoding accuracy higher than SVM decoding accuracy ($p < 0.002$). (B) Correlation between SVM and CNN decoding accuracies. The decoding accuracies from the two methods are not significantly correlated ($R=0.12$, $p = 0.62$). (C) The heatmaps underlying spatial attention control by SVM and CNN. The two heatmaps have limited overlap for the top 20% of voxels ($JI=0.10$), suggesting that the two methods emphasize different features for making predictions. (D) The average random permutation JI and the JI from the data at different thresholds from 10% to 90%. None of the thresholds showed significant differences between the random permutation JI and the JI from the data. JI: Jaccard Index. * $p < 0.002$.

different from the JI of two sets of randomly chosen voxels ($p > 0.05$, FDR corrected). These additional experimental results confirm that SVM and CNN rely on different voxel features for representing different attentional control states (attend left vs attend right).

3.2. Attention dataset: decoding attentional control over the whole brain

Traditional whole-brain univariate analysis was done first to provide a basis for comparison with the subsequent multivariate analysis. As in previous work, the whole-brain responses evoked by spatial cues (left cue and right cue combined) were analyzed using the GLM method. As depicted in Fig. 5A, the spatial cues activated the precentral/postcentral cortex, which contains the frontal eye field (FEF), and superior parietal cortex, which contains the inferior parietal sulcus (IPS)/superior parietal lobule (SPL), which is consistent with previous findings (Giesbrecht et al., 2003; Rajan et al., 2021; Slagter et al., 2007); FEF and IPS/SPL comprise the dorsal attention network (DAN) which is known for its role in attention control. Additionally, other regions such as the putamen, superior frontal and inferior frontal were also activated, agreeing with our previous study (Meyyappan et al., 2021). A total of 9 regions are activated by the spatial cues (see Table 2). However, when the left cue is contrasted against the right cue or vice versa, no regions appeared in the activation map, suggesting that the regions that encode the specific attended information are not discovered by the univariate analysis; see Figs. 5B and 5C.

In contrast, both the MVPA methods SVM and CNN were able to decode the two attention control conditions (cue left vs. cue right) in the

whole-brain analysis ($p < 0.0001$), as shown in Fig. 6A, with the CNN decoding accuracy significantly higher than SVM decoding accuracy ($p < 0.05$). Unlike in V1, over the whole brain, the decoding accuracies between SVM and CNN were significantly correlated ($R = 0.77$, $p < 0.0001$), suggesting that at the whole brain level, the two decoding methods may rely more on common input features than totally different input features; see Fig. 6B.

To further examine the features that drove the classification performance, instead of focusing on voxels as in the case of V1, we focused on the regions of interest for ease of interpretation and comparison with univariate analysis. For SVM and CNN, we presented the top 9 regions in the heatmap (which was the same number of regions as those identified in the univariate analysis) from each method, as shown in Fig. 6C. The top 9 regions identified by the univariate and each of the two MVPA methods are listed in Table 2. There were very few overlapping ROIs among the methods. For example, only lateral occipital, fusiform and superior parietal regions were identified by both SVM and CNN. The lateral occipital, fusiform, and insula appeared in both univariate and SVM analyses, and the precentral, lateral occipital, fusiform and the rostral middle frontal were the areas appearing in both univariate and CNN analyses. In summary, each of the three methods provided insights that are largely not contained in the other methods, suggesting that combining these methods might be a way to obtain a more comprehensive understanding of spatial attention control.

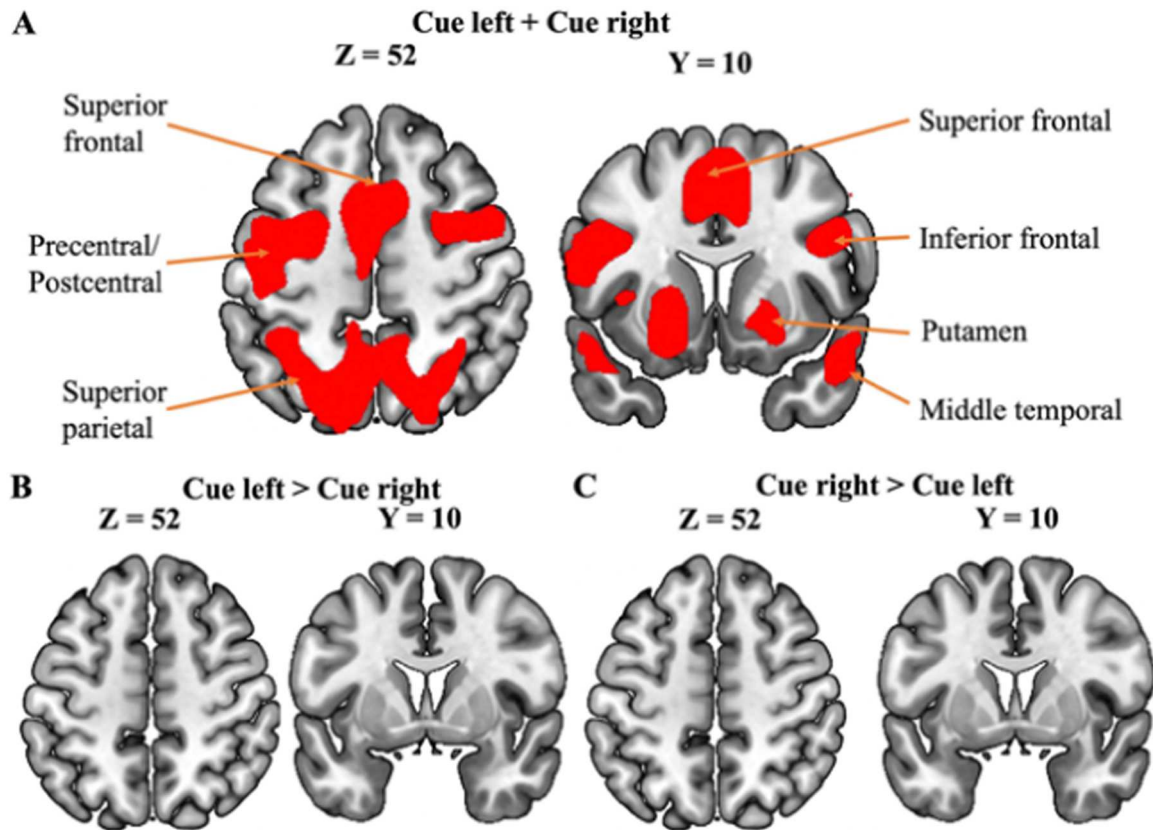


Fig. 5. Univariate analysis of attention control. (A) Cue-evoked activation map ($p < 0.05$, FDR) combining cue left and cue right. Several regions, including precentral/postcentral, superior parietal, inferior frontal, and putamen, were activated. (B) Cue left > cue right activation map ($p < 0.05$, FDR). No regions appeared in the map. (C) Cue right > cue left activation map ($p < 0.05$, FDR). No regions appeared in the map.

Table 2

Important regions underlying attention control identified by univariate and MVPA analysis.

	Cue left + Cue right	Cue left vs Cue right	SVM	CNN
Middle temporal	✓			
Bankssts	✓			
Lateral orbito frontal	✓			
Precentral	✓			✓
Lateral occipital	✓		✓	✓
Fusiform	✓		✓	✓
Insula	✓		✓	
Rostral middle frontal	✓			✓
Medial orbito frontal	✓			
Superior parietal			✓	✓
Parsorbitalis			✓	
Pericalcarine			✓	
Parstriangularis			✓	
Lingual			✓	
Cuneus			✓	
Inferior temporal				✓
Thalamus				✓
Inferior parietal				✓
Superior temporal				✓

3.3. Emotion dataset: decoding emotion processing in V1

Conventionally, the primary visual cortex (V1) is thought to mainly play the role of extracting basic visual features from sensory input and send them to higher order visual areas for further processing. Our recent work where decoding is applied in a within-subject manner found affect-specific neural representations in V1 when participants viewed natural

images containing varying degrees of affective content (Bo et al., 2021). Here, for the same dataset, as shown in Figs. 7A and 7B, both SVM and CNN, when applied in a across-subject fashion, exhibited decoding accuracies between unpleasant and neutral images that were significantly higher than chance level of 50% ($p < 0.0001$), with CNN decoding accuracy significantly higher than that of SVM ($p < 0.002$), further confirming that V1's role is not limited to extracting visual features and it forms an integral part of the neural network that represents the affective significance of natural images.

To test whether the same or different features drove the SVM vs CNN classification performance, we compared the heatmaps from the two methods. The top 20% of voxels (84 voxels) from the SVM heatmap and CNN heatmap were shown in Fig. 7C. Intuitively, there is little overlap between the two sets of voxels; quantitatively, the $JI = 0.09$. Choosing two random sets of 84 voxels and computing the JI 10000 times, the average JI was found to be 0.11 ± 0.024 . The JI of 0.09, which is not significantly different from the overlap between two sets of random voxels ($R = 0.28$, $p = 0.24$). Extending the voxel selection threshold from 10% to 90% and computing the JI index for each threshold, we found that the JI value increased as percentage threshold was increased, as would be expected, but for every threshold, the JI from the data is not significantly different from the overlap of two random sets of voxels ($p > 0.05$, FDR corrected); see Fig. 7D. These results suggest that the features driving the prediction of SVM and CNN differ.

3.4. Emotion dataset: decoding emotional processing over the whole brain

We first conducted a whole-brain univariate analysis by contrasting unpleasant images against neutral images. In comparison to neutral pictures, unpleasant pictures evoked activations in the precentral, insula, putamen, amygdala, and inferior temporal cortices, as shown in

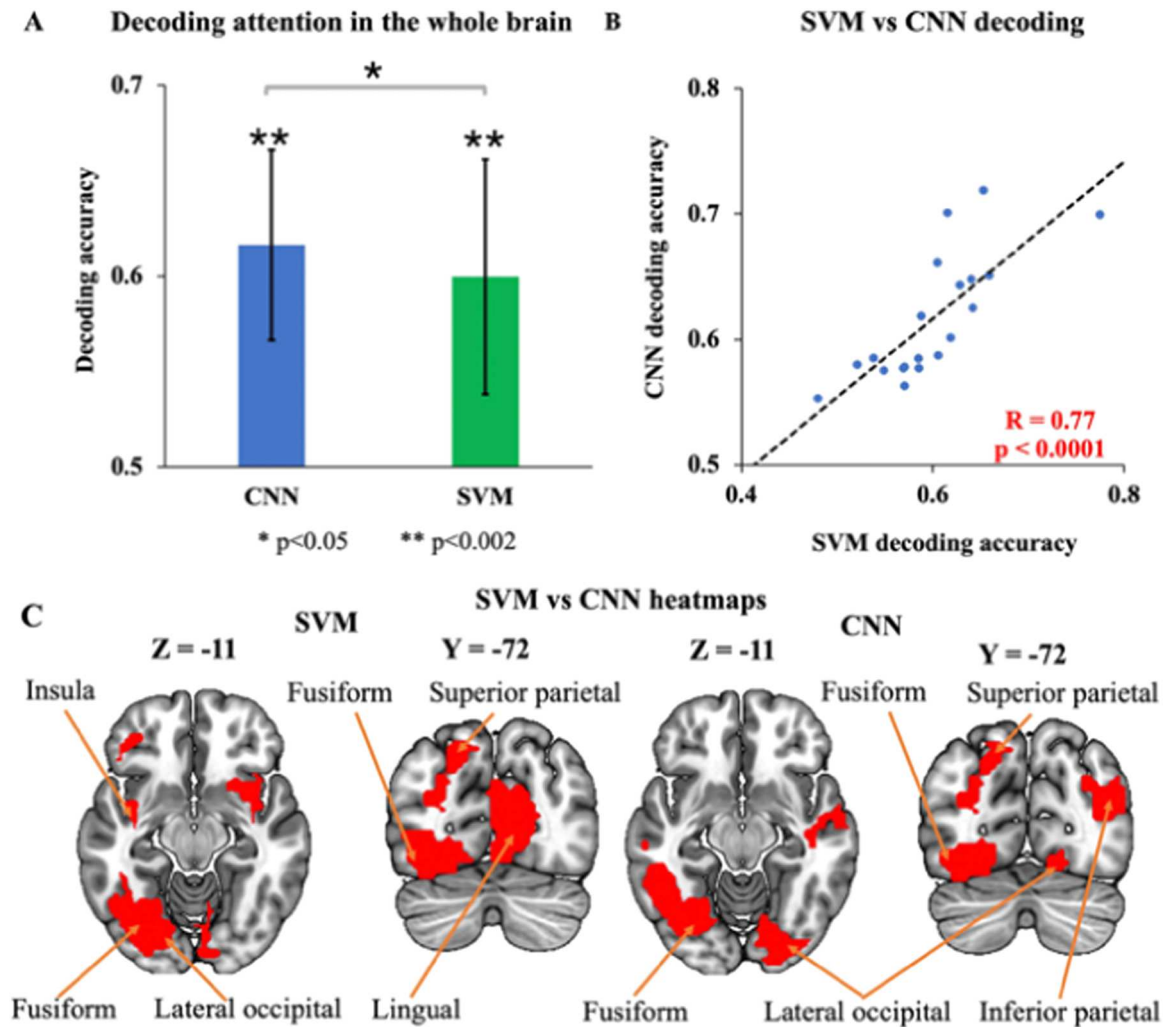


Fig. 6. Decoding attention control over the whole brain. (A) fMRI spatial attention decoding accuracy over the whole brain by SVM and CNN. Both decoding accuracies are significantly above chance level ($p < 0.0001$), with CNN decoding accuracy higher than SVM decoding accuracy ($p < 0.05$). (B) Correlation between SVM and CNN decoding accuracies for spatial attention over the whole brain. The decoding accuracies from the two methods are significantly correlated ($R=0.77$, $p < 0.0001$). (C) The heatmaps underlying spatial attention control. The two heatmaps have several overlapping ROIs, including the fusiform, lateral occipital, and superior parietal cortex. * $p < 0.002$ ** $p < 0.05$.

Fig. 8A. On the other hand, relative to unpleasant pictures, neutral pictures activated the lateral orbitofrontal, superior frontal, middle temporal and parahippocampal regions, as illustrated in **Fig. 8B**. The univariate activation map in **Fig. 8** was thresholded at $p < 0.05$ FDR. The regions in **Figs. 8A** and **8B** are given in **Table 3**. Prior empirical investigations have provided strong evidence for the pivotal roles of the amygdala, insula and parahippocampal in the processing of pain and emotion. Our univariate analysis results are thus consistent with the reports in the literature.

Both MVPA methods SVM and CNN achieved decoding accuracies that were significantly higher than chance level ($p < 0.0001$), with the CNN decoding accuracy significantly higher than that of SVM ($p < 0.05$); see **Fig. 9A**. No significant correlations were found between the decoding accuracies of SVM and CNN in **Fig. 9B** ($R=0.07$, $p = 0.78$). To what extent this finding implies that the two methods relied on different features to perform classifications was examined next using the heatmap.

To identify the essential ROIs driving the prediction performance in both SVM and CNN, we selected the top 11 ROIs (same as the number of ROI identified in the univariate analysis) based on the Lausanne atlas as critical ROIs for generating the heatmaps, as shown in **Fig. 9C**. Regions including fusiform, lateral occipital, inferior parietal and precuneus

were detected by both approaches. See **Table 3**.

4. Discussion

MVPA methods are now widely used in neuroimaging. SVM has been the mainstay in MVPA analysis of fMRI data since the early 2000 s. In more recent years DNNs especially CNNs have been added to the MVPA's toolchest and are gaining importance due to its ability to take into account of nonlinear relationships between neural patterns and cognitive functions. To date no systematic comparisons between these two classes of methods have been carried out. To fill this gap, using two well-characterized fMRI datasets, we compared the decoding performance and heatmaps derived from SVM and CNN, and reported the following results. First, both SVM and CNN are able to achieve above chance level decoding accuracies for attention control and emotion processing in both the primary visual cortex and the whole brain, and the CNN decoding accuracies were consistently higher than that of the SVM. Second, the SVM and CNN decoding accuracies are generally not correlated, and the heatmaps derived by SVM and CNN, which emphasize the importance of certain voxels or brain regions over others in decoding performance, were shown to be not significantly overlapping, suggesting that SVM and CNN may rely on different brain

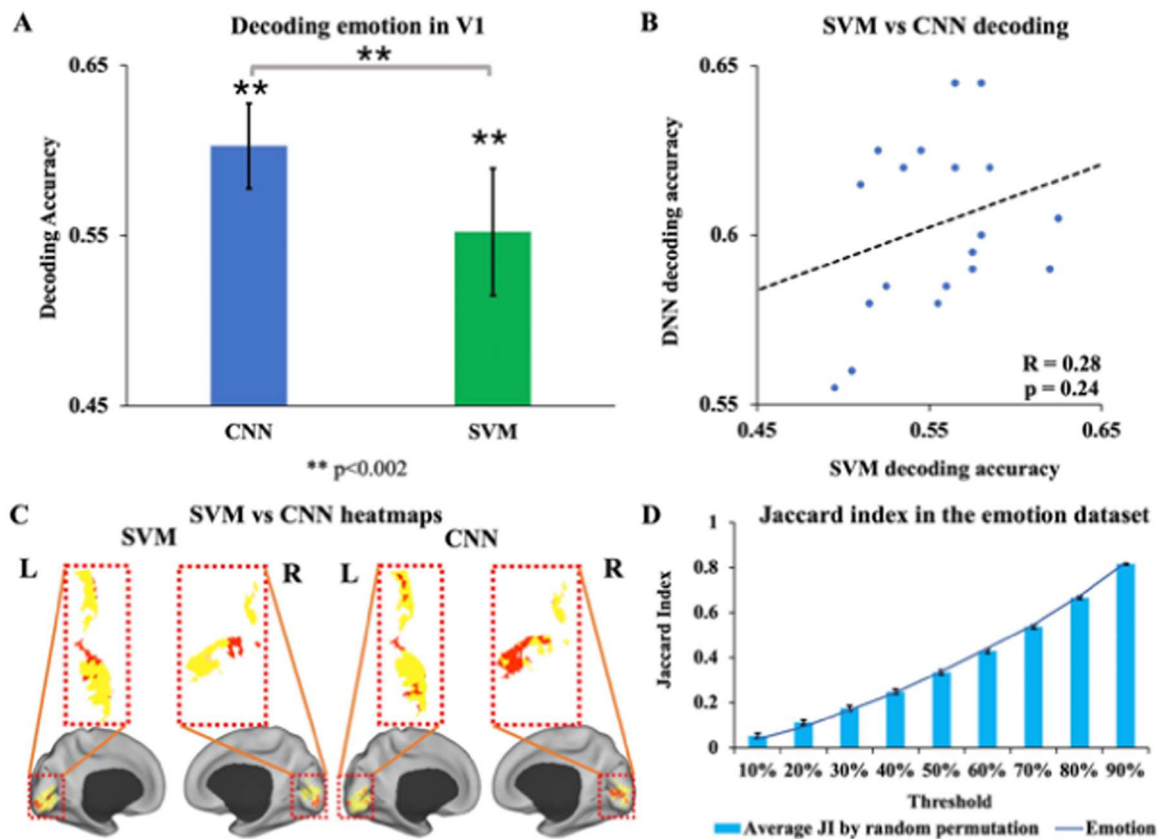


Fig. 7. Decoding emotion processing in V1. (A) Decoding unpleasant vs neutral images in V1 by SVM and CNN. Both decoding accuracies are significantly above chance level ($p < 0.0001$). CNN decoding accuracy is higher than SVM decoding accuracy ($p < 0.002$). (B) Correlation between SVM and CNN decoding accuracies. The decoding accuracies from the two methods are not significantly correlated ($R = 0.28$, $p = 0.24$). (C) The heatmaps underlying decoding by SVM and CNN. The two heatmaps have limited overlap for the top 20% of voxels ($JI = 0.09$), suggesting that the two methods may emphasize different features for making predictions. (D) The average random permutation JI and the JI from the data at different thresholds from 10% to 90%. None of the thresholds showed significant differences between the random permutation JI and the JI from the data. JI: Jaccard Index. * $p < 0.002$.

activity patterns to predict cognitive conditions.

4.1. Analysis in V1

The primary visual cortex (V1) is the first stage of the cortical visual processing system (Hubel and Wiesel, 1977). As such its role is often thought to be limited to extracting visual features from input and passing them up the visual hierarchy for further processing. For both datasets analyzed here, our results demonstrate that V1's role goes beyond simple feature extraction. In the case of visual spatial attention, while attentional modulation of neuronal responses has been reported in many extrastriate cortical areas such as V2, V4, temporal-occipital area (TEO), and middle temporal area (MT), evidence of attentional modulation in V1 has been relatively scarce (Pessoa et al., 2003), even though it is well known that V1 receives feedback connections from higher-level visual areas, allowing its activity to be modulated by top-down attentional control signals (Juan and Walsh, 2003). In the present study, both SVM and CNN achieved above-chance decoding accuracies between two preparatory attention conditions (attend left vs attend right) using only the voxels extracted from V1, demonstrating a role of V1 in attention control. Furthermore, owing to the across-subject decoding approach adopted here, the patterns of the attention control signals are found to be shared by the participants (Martinez et al., 1999). This finding extends a recent study by Meyyappan et al. where an within-subject decoding approach was employed to show that V1 is engaged in visual spatial attention control (Meyyappan et al., 2023). In the case of visual emotion processing, the role of V1 is not well understood, and findings from different studies have been inconsistent. Generally, univariate

BOLD analyses have not found differential activations by differently valenced emotional images (e.g., unpleasant vs neutral) in early visual areas such as V1 (Sabatinelli et al., 2009). Using a within-subject decoding approach, our previous analysis of the same emotion dataset shows that there are affect-specific neural representations in the retinotopic visual cortex, which includes V1 (Bo et al., 2021). In the present study, applying an across-subject approach using two different decoding methods, we again found evidence to support the notion that the primary visual cortex V1 plays a role in the representation of the emotional significance of natural images and these representations are shared by participants. It is worth noting that for both datasets, the heatmaps derived from SVM and CNN are largely nonoverlapping, suggesting that these two classes of methods may rely on different features for decoding.

4.2. Analysis over the whole brain

At the whole brain level, we first performed the traditional univariate analysis to provide a departure point for the comparison of the two MVPA methods. For the attention dataset, when attend left cue and attend right cue were combined, the frontal eye field and the IPS/SPL were found to be activated. These regions are part of the dorsal attention network (DAN) whose role in the control of spatial attention, feature attention, and object attention is well established (Corbetta et al., 2005; Giesbrecht et al., 2003; Morishima et al., 2009; Slagter et al., 2007). Additional areas activated by the combined cues include inferior frontal cortex, which are associated with cognitive control and response inhibition, both being integral parts of attention (Forstmann et al., 2008). When we contrasted attend left cue against attend right cue, however,

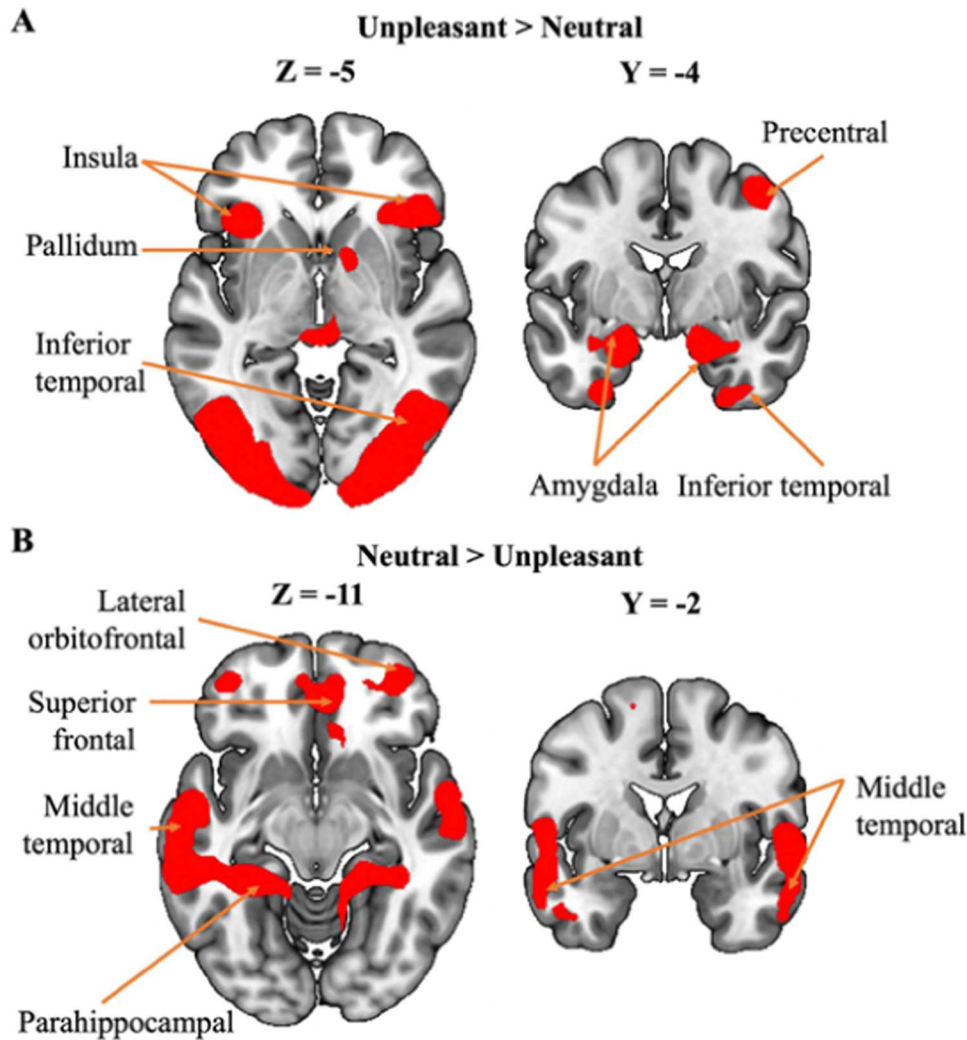


Fig. 8. Univariate analysis of emotion processing. Activation map ($p < 0.05$, FDR) contrasting (A) unpleasant > neutral (B) neutral > unpleasant pictures.

no brain activations were found, suggesting that the univariate analysis is not able to discover the signals that are specific to the attended information (e.g., attend left vs attend right). In contrast, both SVM and CNN were able to decode attend left vs attend right using the whole brain as the ROI. In particular, these MVPA analyses were able to unveil brain regions whose involvement in attention control was not uncovered by the conventional univariate analysis, including the inferior parietal cortex and inferior temporal cortex. Recent studies utilizing imaging techniques and lesion investigations have provided insights into the functional properties of inferior parietal regions and found that they contribute to the maintenance of attention, detection of salient events within a sequence, and exertion of control over attentional mechanisms (Husain and Nachev, 2007). Moreover, event-related fMRI investigations have identified a network of cortical areas, including the inferior parietal cortex, that play a role in top-down attentional control (Hopfinger et al., 2000) and attentional shifting (Corbetta et al., 1993). In terms of the univariate activation map and the heat maps derived from the two MVPA methods, similar to V1, there is limited overlap among the regions identified by each of the three methodologies (i.e., univariate, SVM, and CNN) as being important for spatial attention control, suggesting that combining these methods may afford us the ability to more comprehensively identify important neural substrate of cognitive functions. One observation worth noting is that the DAN did not appear in the heatmap of either of the two MVPA methods. This is in contrast to our previous decoding analysis in DAN, which, using a

within-subject approach (Rajan et al., 2021), demonstrated the existence of a microstructure for attention control in DAN, consistent with an extensive literature (Giesbrecht et al., 2003; Morishima et al., 2009). The findings reported here may be taken to imply that the microstructure of attention control exhibits a high degree of individual variability, which renders the across-subject decoding approach not effective in revealing its involvement in attentional control.

For the emotion dataset, the univariate analysis revealed activations in multiple brain areas when contrasting unpleasant images against neutral images, including the insula, precentral gyrus, amygdala, fusiform gyrus, superior frontal gyrus, and middle temporal gyrus, as reported in our previous study (Bo et al., 2021). The amygdala and insula are widely recognized as key brain regions involved in processing emotional information especially negative emotions (Kesler et al., 2001). The superior frontal region is implicated in the cognitive ability to infer the mental states of others and other social functions where emotion processing is critical (Mak et al., 2009; Völlm et al., 2006). Notably, the retinotopic visual cortex is not engaged in representing the emotional significance of visual processing according to the univariate analysis, contrary to the theoretical argument. Specifically, the signal reentry hypothesis suggested that subcortical structures, such as the amygdala, may send feedback signals into the visual cortex upon receiving initial sensory input to enhance the processing of emotionally salient visual stimuli (Keil et al., 2009; Lang and Bradley, 2010; Sabatinelli et al., 2009). In both SVM and CNN analysis, the visual cortex was

Table 3

Important regions underlying emotion processing identified by univariate and MVPA analysis.

	Unpleasant>Neutral	Neutral>Unpleasant	SVM	CNN
Inferior temporal	✓		✓	
Lateral orbito frontal	✓			
Amygdala	✓			
Superior frontal	✓			
Dorsal Acc	✓			
Brainstem	✓			✓
Pallidum	✓			
Cuneus		✓	✓	
Caudal middle frontal		✓		
Rostral middle frontal		✓		✓
Rostral Acc		✓		
Fusiform			✓	✓
Pericalcarine			✓	
Lingual			✓	
Lateral occipital			✓	✓
Supramarginal			✓	
Inferior parietal			✓	✓
Precunues			✓	✓
Posterior cingulate			✓	
Superior parietal			✓	
Middle temporal				✓
Pars opercularis				✓
Superior temporal				✓
Medial orbito frontal				✓
Parsorbitalis				✓

found to contribute to the decoding performance, in agreement with our previous study using a within-subject decoding approach (Bo et al., 2021). In light the across-subject employed here, this result further suggests that the neural representations of the emotional significance of natural images are shared across participants. Furthermore, the involvement of the precuneus in representing emotion was found by both SVM and CNN, but was not revealed by univariate analysis. The precuneus region, along with activity in the right anterior insular cortex and ventromedial prefrontal cortex (VMPFC), has been linked to the evaluation of emotional states (Terasawa et al., 2013). Critchley et al. (Critchley et al., 2001) provides evidence that the precuneus plays a critical role in transforming interoceptive information into subjective emotions.

4.3. Methodological considerations

Two methodological issues are worth considering. First, the findings of the current study suggest that CNN performed better than SVM in all decoding tasks. This advantage of CNN can be attributed to its superior learning abilities relative to SVM (Khodatars et al., 2021); SVM can be considered as having a single-layer architecture and is thus less computationally capable compared to a multi-layered system such as the CNN (Kim et al., 2019). Consistent with this, recent work has found that combining SVM with nonlinear kernels did not result in superior decoding performance (Misaki et al., 2010), suggesting that the superior learning ability made possible by the multiple-layered architecture of the CNN, along with CNN's ability to model nonlinear functions, underlie the CNN's superior performance observed in our study. However,

CNN, having multiple layers of computational units (neurons), has a large number of parameters; it is known that as the number of parameters is increased, overfitting is more likely to occur. We applied cross-validation, batch normalization, and dropout to minimize overfitting (Ioffe and Szegedy, 2015; Srivastava et al., 2014). Second, we utilized an across-subject decoding approach rather than the more traditional within-subject decoding approach. In the within-subject decoding approach, patterns of neural activity are collected for each individual during a specific experimental condition or task, and then subjected to decoding or classification analysis. The across-subject approach, which is adopted here to cope with the fact that training CNN models requires larger amount of data than can be expected from a single subject in typical neuroimaging studies, involves pooling data from multiple participants performing the same experiment, and then performing decoding or classification of the patterns; the decodable patterns in this case are patterns that are common across the group. The within-subject and across-subject decoding approaches each have their advantages and limitations. Within-subject decoding allows for investigating individual differences and can provide insights into the neural mechanisms that are specific to each individual. It is less affected by inter-individual variability and can potentially capture more fine-grained individual-specific information. The use of an across-subject approach in fMRI research also has advantages. For example, pooling data from multiple subjects can increase the statistical power of the analysis, making it more likely to detect subtle or complex effects that may not be apparent in individual subjects due to variability or noise in the data.

4.4. Summary and future directions

In this work we showed that: (1) both SVM and CNN are effective methods for analyzing neuroimaging data with CNN exhibiting superior performance by consistently outperforming SVM in terms of decoding accuracy, (2) SVM and CNN may rely on different neural features for making classifications of brain states, (3) Jaccard index can be combined with random permutation to assess the overlap between SVM and CNN heatmaps, and (4) SVM and CNN can yield novel neuroscience insights not possible with conventional fMRI analysis. It is worth noting that these results were obtained on datasets with modest sample size ($n = 20$). Therefore they should be viewed as being preliminary. Future work with substantially larger samples sizes is required to reach definitive conclusions. In addition, these results were achieved by applying SVM and CNN independently. A fruitful future direction may lie in combining the two methods in the same analysis pipeline: apply SVM first to model the linearly decodable components of the data and CNN next to model the residual nonlinear components of the data. Furthermore, regarding the application of CNNs to neuroimaging datasets, in addition to across-subject decoding and rigorous cross validation, methods such as L1 and L2 norm regularizations, as well as data augmentation techniques, may also be considered to address the issue of relatively small sample sizes. Finally, one of our main findings is that the features that drive SVM decoding and that drive CNN decoding appear to be not the same. The neuroscientific significance of these features remains not clear. Progress on this issue may lead to a better understanding of the neural substrate of cognitive phenomena and the development of novel cognitive neuroscience theories.

Funding

This study was supported by the NIH grants MH117991, MH125615, and MH112558 and NSF grant BCS2318984.

CRedit authorship contribution statement

Yun Liang: Writing – original draft, Writing – review & editing, Methodology, Investigation, Formal analysis. **Ke Bo:** Data curation.

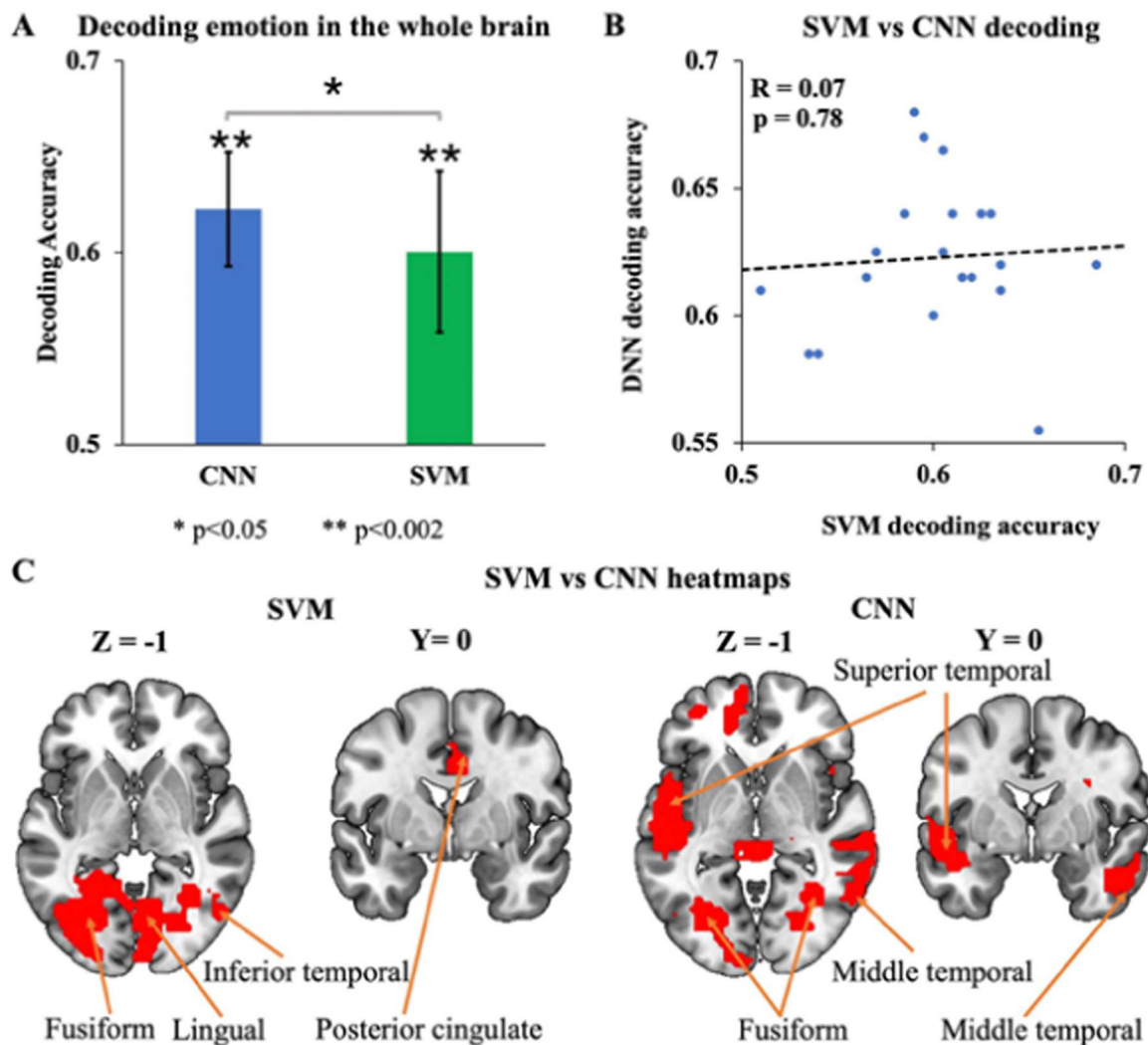


Fig. 9. Decoding unpleasant vs neutral IAPS images over the whole brain. (A) fMRI decoding accuracy in the whole brain by SVM and CNN. Both decoding accuracies are significantly above chance level ($p < 0.0001$). CNN achieved significantly higher decoding accuracy compared to SVM ($p < 0.05$). (B) Correlation between SVM and CNN decoding accuracies. The decoding accuracies from the two methods are not significantly correlated ($p = 0.78$). (C) The heatmaps underlying the decoding between unpleasant vs. neutral over the whole brain. Both heatmaps include the fusiform, lateral occipital, inferior parietal and precuneus as essential brain regions. * $p < 0.002$ * $p < 0.05$.

Sreenivasan Meyyappan: Data curation. **Mingzhou Ding:** Writing – review & editing, Writing – original draft, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation.

Declaration of Competing Interest

The authors declare no conflicts of interest.

Data availability

Data will be made available upon request.

References

- Abrol, A., Fu, Z., Salman, M., Silva, R., Du, Y., Plis, S., Calhoun, V., 2021. Deep learning encodes robust discriminative neuroimaging representations to outperform standard machine learning. *Nat. Commun.* 12, 1–17.
- Baucom, L.B., Wedell, D.H., Wang, J., Blitzer, D.N., Shinkareva, S.V., 2012. Decoding the neural representation of affective states. *Neuroimage* 59, 718–727.
- Birn, R.M., Saad, Z.S., Bandettini, P.A., 2001. Spatial heterogeneity of the nonlinear dynamics in the fMRI BOLD response. *Neuroimage* 14, 817–826.

- Bo, K., Yin, S., Liu, Y., Hu, Z., Meyyappan, S., Kim, S., Keil, A., Ding, M., 2021. Decoding neural representations of affective scenes in retinotopic visual cortex. *Cereb. Cortex* 31, 3047–3063.
- Bonnici, H.M., Sidhu, M., Chadwick, M.J., Duncan, J.S., Maguire, E.A., 2013. Assessing hippocampal functional reserve in temporal lobe epilepsy: a multi-voxel pattern analysis of fMRI data. *Epilepsy Res.* 105, 140–149.
- Chang, C.-C., Lin, C.-J., 2011. LIBSVM: a library for support vector machines. *ACM Trans. Intell. Syst. Technol. (TIST)* 2, 1–27.
- Corbetta, M., Miezin, F.M., Shulman, G.L., Petersen, S.E., 1993. A PET study of visuospatial attention. *J. Neurosci.* 13, 1202–1226.
- Corbetta, M., Tansy, A.P., Stanley, C.M., Astafiev, S.V., Snyder, A.Z., Shulman, G.L., 2005. A functional MRI study of preparatory signals for spatial location and objects. *Neuropsychologia* 43, 2041–2056.
- Corbetta, M., Patel, G., Shulman, G.L., 2008. The reorienting system of the human brain: from environment to theory of mind. *Neuron* 58, 306–324.
- Critchley, H.D., Mathias, C.J., Dolan, R.J., 2001. Neuroanatomical basis for first- and second-order representations of bodily states. *Nat. Neurosci.* 4, 207–212.
- Daducci, A., Gerhard, S., Griffa, A., Lemkaddem, A., Cammoun, L., Gigandet, X., Meuli, R., Hagmann, P., Thiran, J.-P., 2012. The connectome mapper: an open-source processing pipeline to map connectomes with MRI. *PLoS One* 7, e48121.
- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L., 2009. Imagenet: A large-scale hierarchical image database. 2009 IEEE conference on computer vision and pattern recognition. Ieee, 2009: 248–255.
- Desmond, J.E., Glover, G.H., 2002. Estimating sample size in functional MRI (fMRI) neuroimaging studies: statistical power analyses. *J. Neurosci. Methods* 118, 115–128.

- Farahani, F.V., Karwowski, W., Lighthall, N.R., 2019. Application of graph theory for identifying connectivity patterns in human brain networks: a systematic review. *Front. Neurosci.* 13, 585.
- Forstmann, B.U., Jahfari, S., Scholte, H.S., Wolfensteller, U., van den Wildenberg, W.P., Ridderinkhof, K.R., 2008. Function and structure of the right inferior frontal cortex predict individual differences in response inhibition: a model-based approach. *J. Neurosci.* 28, 9790–9796.
- Giesbrecht, B., Woldorff, M.G., Song, A.W., Mangun, G.R., 2003. Neural mechanisms of top-down control during spatial and feature attention. *Neuroimage* 19, 496–512.
- Haufe, S., Meinecke, F., Görgen, K., Dähne, S., Haynes, J.-D., Blankertz, B., Bießmann, F., 2014. On the interpretation of weight vectors of linear models in multivariate neuroimaging. *Neuroimage* 87, 96–110.
- Haxby, J.V., Guntupalli, J.S., Connolly, A.C., Halchenko, Y.O., Conroy, B.R., Gobbini, M. I., Hanke, M., Ramadge, P.J., 2011. A common, high-dimensional model of the representational space in human ventral temporal cortex. *Neuron* 72, 404–416.
- Haynes, J.-D., 2015. A primer on pattern-based approaches to fMRI: principles, pitfalls, and perspectives. *Neuron* 87, 257–270.
- He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* 770–778.
- Hopfinger, J.B., Buonocore, M.H., Mangun, G.R., 2000. The neural mechanisms of top-down attentional control. *Nat. Neurosci.* 3, 284–291.
- Hubel, D.H., Wiesel, T.N., 1977. Ferrier lecture-Functional architecture of macaque monkey visual cortex. *Proc. R. Soc. Lond. Ser. B. Biol. Sci.* 198, 1–59.
- Husain, M., Nachev, P., 2007. Space and the parietal cortex. *Trends Cogn. Sci.* 11, 30–36.
- Ioffe, S., Szegedy, C., 2015. Batch normalization: accelerating deep network training by reducing internal covariate shift. *Int. Conf. Mach. Learn. PMLR* 448–456.
- Juan, C.-H., Walsh, V., 2003. Feedback to V1: a reverse hierarchy in vision. *Exp. Brain Res.* 150, 259–263.
- Keil, A., Sabatinelli, D., Ding, M., Lang, P.J., Ihssen, N., Heim, S., 2009. Re-entrant projections modulate visual cortex in affective perception: evidence from Granger causality analysis. *Hum. Brain Mapp.* 30, 532–540.
- Kesler, M.L., Andersen, A.H., Smith, C.D., Avison, M.J., Davis, C.E., Kryscio, R.J., Blonder, L.X., 2001. Neural substrates of facial emotion processing using fMRI. *Cogn. Brain Res.* 11, 213–226.
- Khodatars, M., Shoeibi, A., Sadeghi, D., Ghaesemi, N., Jafari, M., Moridian, P., Khadem, A., Alizadehsani, R., Zare, A., Kong, Y., 2021. Deep learning for neuroimaging-based diagnosis and rehabilitation of autism spectrum disorder: a review. *Comput. Biol. Med.* 139, 104949.
- Kim, H.-C., Bandettini, P.A., Lee, J.-H., 2019. Deep neural network predicts emotional responses of the human brain from functional magnetic resonance imaging. *NeuroImage* 186, 607–627.
- Kotz, S.A., Kalberlah, C., Bahlmann, J., Friederici, A.D., Haynes, J.D., 2013. Predicting vocal emotion expressions from the human brain. *Hum. Brain Mapp.* 34, 1971–1981.
- Krizhevsky, A., Hinton, G., 2009. Learning multiple layers of features from tiny images. 2009.
- LaConte, S., Strother, S., Cherkassky, V., Anderson, J., Hu, X., 2005. Support vector machines for temporal classification of block design fMRI data. *NeuroImage* 26, 317–329.
- Lang, P.J., Bradley, M.M., 2010. Emotion and the motivational brain. *Biol. Psychol.* 84, 437–450.
- Lang, P.J., Bradley, M.M., Cuthbert, B.N., 1997. International affective picture system (IAPS): technical manual and affective ratings. NIMH Cent. Study Emot. Atten. 1, 3.
- Lee, S., Halder, S., Kübler, A., Birbaumer, N., Sitaram, R., 2010. Effective functional mapping of fMRI data with support-vector machines. *Wiley Online Libr.*
- Levandowsky, M., Winter, D., 1971. Distance between sets. *Nature* 234, 34–35.
- Lewis-Peacock, J.A., Norman, K.A., 2014. Multi-voxel pattern analysis of fMRI data. *Cogn. Neurosci.* 512, 911–920.
- Liang, Y., Zhao, Q., Hu, Z., Bo, K., Meyyappan, S., Neubert, J.K., Ding, M., 2023. Imaging the neural substrate of trigeminal neuralgia pain using deep learning. *Front. Hum. Neurosci.* 17, 1144159.
- Liu, T., Slotnick, S.D., Serences, J.T., Yantis, S., 2003. Cortical mechanisms of feature-based attentional control. *Cereb. Cortex* 13, 1334–1343.
- Mahmoudi, A., Takerkart, S., Regragui, F., Boussaoud, D., Brovelli, A., 2012. Multivoxel pattern analysis for fMRI data: a review. *Comput. Math. Methods Med.* 2012.
- Mak, A.K., Hu, Z.-g., Zhang, J.X., Xiao, Z.-w., Lee, T.M., 2009. Neural correlates of regulation of positive and negative emotions: an fMRI study. *Neurosci. Lett.* 457, 101–106.
- Martinez, A., Anllo-Vento, L., Sereno, M.I., Frank, L.R., Buxton, R.B., Dubowitz, D., Wong, E.C., Hinrichs, H., Heinze, H.J., Hillyard, S.A., 1999. Involvement of striate and extrastriate visual cortical areas in spatial attention. *Nat. Neurosci.* 2, 364–369.
- Meyyappan, S., Rajan, A., Mangun, G.R., Ding, M., 2021. Role of inferior frontal junction (IFJ) in the control of feature versus spatial attention. *J. Neurosci.* 41, 8065–8074.
- Meyyappan, S., Rajan, A., Yang, Q., Mangun, G.R., Ding, M., 2023. Top-down biasing of visual cortical activity encodes attended information and facilitates behavioral performance in visual spatial attention. *bioRxiv*, 2023.08. 05.552084.
- Misaki, M., Kim, Y., Bandettini, P.A., Kriegeskorte, N., 2010. Comparison of multivariate classifiers and response normalizations for pattern-information fMRI. *Neuroimage* 53, 103–118.
- Mitchell, T.M., Hutchinson, R., Just, M.A., Niculescu, R.S., Pereira, F., Wang, X., 2003. Classifying instantaneous cognitive states from fMRI data. *AMIA Annu. Symp. . Proc. Am. Med. Inform. Assoc.* 465.
- Morishima, Y., Akaishi, R., Yamada, Y., Okuda, J., Toma, K., Sakai, K., 2009. Task-specific signal transmission from prefrontal cortex in visual selective attention. *Nat. Neurosci.* 12, 85–91.
- Mourao-Miranda, J., Bokde, A.L., Born, C., Hampel, H., Stetter, M., 2005. Classifying brain states and determining the discriminating activation patterns: support vector machine on functional MRI data. *NeuroImage* 28, 980–995.
- Mumford, J.A., Turner, B.O., Ashby, F.G., Poldrack, R.A., 2012. Deconvolving BOLD activation in event-related designs for multivoxel pattern classification analyses. *Neuroimage* 59, 2636–2643.
- Norman, K.A., Polyn, S.M., Detre, G.J., Haxby, J.V., 2006. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends Cogn. Sci.* 10, 424–430.
- Pessoa, L., Kastner, S., Ungerleider, L.G., 2003. Neuroimaging studies of attention: from modulation of sensory processing to top-down control. *J. Neurosci.* 23, 3990–3998.
- Rajan, A., Meyyappan, S., Liu, Y., Samuel, I.B.H., Nandi, B., Mangun, G.R., Ding, M., 2021. The microstructure of attentional control in the dorsal attention network. *J. Cogn. Neurosci.* 33, 965–983.
- Sabatinelli, D., Lang, P.J., Bradley, M.M., Costa, V.D., Keil, A., 2009. The timing of emotional discrimination in human amygdala and ventral visual cortex. *J. Neurosci.* 29, 14864–14868.
- Said, C.P., Moore, C.D., Engell, A.D., Todorov, A., Haxby, J.V., 2010. Distributed representations of dynamic facial expressions in the superior temporal sulcus. *J. Vis.* 10, 11.
- Sain, S.R., 1996. *The Nature of Statistical Learning Theory*. Taylor & Francis.
- Saraf, S., DeSouza, D.D., Anderson, J., Tofghi, G., 2016. Initiative AsDN. DeepAD: Alzheimer's disease classification via deep convolutional neural networks using MRI and fMRI. *BioRxiv*, 070441.
- Sitaram, R., Lee, S., Ruiz, S., Rana, M., Veit, R., Birbaumer, N., 2011. Real-time support vector classification and feedback of multiple emotional brain states. *Neuroimage* 56, 753–765.
- Slagter, H.A., Giesbrecht, B., Kok, A., Weissman, D., Kenemans, J.L., Woldorff, M., Mangun, G., 2007. fMRI evidence for both generalized and specialized components of attentional control. *Brain Res.* 1177, 90–102.
- Song, S., Zhan, Z., Long, Z., Zhang, J., Yao, L., 2011. Comparative study of SVM methods combined with voxel selection for object category classification on fMRI data. *PLoS One* 6, e17191.
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* 15, 1929–1958.
- Tagliazucchi, E., Laufs, H., 2014. Decoding wakefulness levels from typical fMRI resting-state data reveals reliable drifts between wakefulness and sleep. *Neuron* 82, 695–708.
- Tagliazucchi, E., von Wegner, F., Morzelewski, A., Borisov, S., Jahnke, K., Laufs, H., 2012. Automatic sleep staging using fMRI functional connectivity data. *Neuroimage* 63, 63–72.
- Terasawa, Y., Fukushima, H., Umeda, S., 2013. How does interoceptive awareness interact with the subjective experience of emotion? An fMRI study. *Hum. Brain Mapp.* 34, 598–612.
- Vieira, S., Pinaya, W.H., Mechelli, A., 2017. Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: Methods and applications. *Neurosci. Biobehav. Rev.* 74, 58–75.
- Völlm, B.A., Taylor, A.N., Richardson, P., Corcoran, R., Stirling, J., McKie, S., Deakin, J. F., Elliott, R., 2006. Neuronal correlates of theory of mind and empathy: a functional magnetic resonance imaging study in a nonverbal task. *Neuroimage* 29, 90–98.
- Wang, C., Rajagovindan, R., Han, S.-M., Ding, M., 2016. Top-down control of visual alpha oscillations: sources of control signals and their mechanisms of action. *Front. Hum. Neurosci.* 10, 15.
- Wang, L., Mruczek, R.E., Arcaro, M.J., Kastner, S., 2015. Probabilistic maps of visual topography in human cortex. *Cereb. Cortex* 25, 3911–3931.