Exploring Working Memory Task Predictability using fMRI Data

Jennifer Hung
yhung@ucsd.edu

Sahana Narayanan
sanarayanan@ucsd.edu

Judel Ancayan
jancayan@ucsd.edu

Gabriel Riegner (Mentor)
gariegner@ucsd.edu

Armin Schwartzman (Mentor)
armins@ucsd.edu

Abstract

In our quarter-one project, our group performed a first-level analysis of functional magnetic resonance imaging (fMRI) data. This consisted of analyzing brain data in a 4-dimensional array and identifying areas of significant activation. In this project, we use voxel weights generated from individual first-level analysis to train various machine learning models to classify task types based on fMRI brain imaging data. With this model, we can address gaps in neuroscience literature on whether visual working memory tasks with varying task strategies use fundamentally different neural circuitry. Furthermore, it could suggest that models are able to "read minds" of individuals under different tasks. Multivoxel pattern analysis (MVPA) is an increasingly popular method of analyzing fMRI brain imaging by retaining the information of each voxel in the brain. However, recent work on MVPA was typically limited to Support Vector Machines (SVMs). With our project, we hope to explore other popular machine learning models to classify task type, look into the interpretability of models for task fMRI classification, and explore the use of confounds within brain imaging data to build more accurate models using the full fMRI data.

Code: https://github.com/yunchen-hung/fmriWorkingMemoryProject

1 Introduction

Multivoxel pattern analysis (MVPA) has been an increasingly popular method of functional magnetic resonance imaging (fMRI) brain imaging analysis in recent years. Instead of studying univariate analysis such as averaging activity across the entire brain, MVPA compares patterns of activation in clusters of voxels. It has also been considered as a method of "mind reading", as MVPA allows the decoding of representational states within the brain. A review study by (Mahmoudi et al. 2012) noted the effectiveness of Support Vector Machines (SVM) in MVPA of functional imaging data as a supervised classification problem, as it was equipped to deal with high dimensional data and flexibility in modeling various types of data. Recent literature by (Aglieri et al. 2021) performed MVPA on task fMRI data to decode which speaker participants were recalling in a speaker recognition task. They were able to identify temporal and extra-temporal regions associated with the task, displaying the capability of using SVM on task fMRI data in understanding the neural correlates involved in a task. Similarly, (Hof et al. 2021) trained SVMs on a task fMRI data to develop a sexual image classifier. They were able to distinguish between brain regions associated with response to sexual images and general response areas with high accuracy with MVPA. However, k-nearest neighbors (kNN), Gaussian Naive Bayes, and Logistic Regression are also potential models of interest for fMRI classification (Pereira, Mitchell and Botvinick 2009).

In our project, we specifically look into whether we can identify which visual working memory task subjects are performing based on their brain imaging data. The motivation for this lies in the fact that the two tasks within this dataset are highly similar to one another, with the only difference being the memory test section (For more information, refer to the dataset description section 2.1). Past neuroscience studies have yet to look into whether the neural pathways throughout the act of temporarily maintaining or storing perceptual information are different given different task strategies. We are interested in using modern machine learning methods to see if the neural correlates involved throughout the duration of the two memory tasks can be clearly separated and classified. We hypothesize that if the model accuracy is high, then it suggests that the neural circuitry behind the two tasks are different enough to be discriminated.

We aim to compare popular MVPA models such as kNN, Logistic Regression, and Naive Bayes to our baseline model of linear SVMs. In our quarter-one project, we took working memory task fMRI data and identified regions of significant activation in response to the various tasks. This was done by performing a voxel-wise linear regression on a dataset containing brain scans over time for each subject. Continuing forward, we will refer to this process as performing a first-level analysis of the data. Through these first-level analyses, we found that there are various differences in the activation depending on the corresponding task. This begs the question: are these differences in activation significant enough that we can distinguish between the tasks from the resulting fMRI data alone? Our supervised models are trained on the beta weights generated from first-level analysis for a full brain image in all runs (where a run is one trial of the task). This allows for easier interpretation as we can know which regions of the brain are used for each task. These models will

then be evaluated through various metrics, such as accuracy, precision, and recall. We also explore the effect of confounds, such as noise within the white matter, cerebral spinal fluid noise, and movements during fMRI scans, on model accuracy, as it may contain important information about the subjects ability to do each task.

Additionally, through training and comparing various machine learning models, we can further strengthen or improve the current literature on the optimal model for task fMRI classification with MVPA. Furthermore, we aim to focus more on increasing the interpretability of these models through analyzing brain regions relevant to the visual working memory tasks to not only aid other neuroscientists in understanding the neural correlates involved in such tasks, but also provide computer scientists with regions of interest to perform dimension reduction and to improve computation efficiency.

2 Methods

2.1 Dataset Description

The dataset in our project was sourced from the paper (Scimeca, Kiyonaga and D'Esposito 2018). Subjects were asked to perform two different visual working memory tasks, and they would alternate between each task for 4 runs each for a total of one hour in a functional magnetic resonance imaging (fMRI) machine. The two tasks were fairly similar but for the memory recall part. Subjects were asked to memorize three differently colored patches, then after a delay they were cued to recall the color of one of the squares by using a color wheel (task 'colorwheel') or performing a binary response (task 'samedifferent'). Due to the different levels of abstraction for each task, it is hypothesized that they may be using different areas of the brain to perform a memory task. However, whether the neural circuitry is distinguishable enough from each other is unknown.

In our model training and testing, we selected 34 subjects who had little to no problems during their brain scan. Each subject had 4 runs for both tasks, making it about 136 runs total for each task. In particular, one subject was missing a 4th run of their colorwheel task, making the final total number of runs 271 across both tasks. The raw BIDS formatted data was preprocessed with fMRIprep pipeline. The specific methods used can be found in Appendix A.2, which is sourced directly from the preprocessing boilerplate.

2.2 First-Level Analysis

To increase mechanistic interpretability, we chose to extract the significance of every weight within a subject's brain through the analysis of beta weights. By extracting the beta weights for a brain image, we can gain information about the amount of correlation between the task and the voxels within each brain over the task period. To achieve this, we performed a first-level analysis on individual participants by fitting a General Linear Model (GLM) for each run within two visual tasks. Our function took in information on the subject ID and

task type to calculate the design matrix with the subject's events file and brain imaging data. The events file stored information about the onset offset of cues and probes within the experiment. For the first level analysis model, we selected 'spm' and its time derivatives for the hemodynamic response function, with a repetition time of 2 seconds. Moreover, we smoothed the data with a Gaussian kernel width of 6 millimeters and used all but one of the available CPU cores to compute each first-level analysis.

We proceeded to do this process for each individual subject, for every run of their time-series brain scan. Notably, we chose to create two different datasets for separate model training: one with beta weights denoised and corrected for movement, and another with the following specific confounds included: white matter and cerebral spinal fluid noise, the rotation and translation across the 'xyz' axis. Next, we extracted the entire brain from 4-dimensional files containing beta weights with a MNI152 whole-brain mask. This ensured that our model was not training on empty white space outside the brain, and specifically selecting regions within the brain. Furthermore, using a mask allowed us to perform inverse transformation post-model training to obtain a 3-dimensional brain image from a 1-dimensional array of coefficients. We then trained our models either with or without the confounded beta weights.

2.3 Model Building

For each individual run in our dataset, we flattened the 3-dimensional beta weights into a 2-dimensional array, resulting in 235, 375 features for both the confounded and non confounded training data. Each feature in the model corresponds with a specific voxel found from the first level analysis. Our model should predict either the task 'color wheel' or the task 'same different'. To build the model, different approaches were considered for building the training and test sets. One approach involved splitting the dataset into 3 groups: a training set containing 60% of the sample size, a validation set containing 20%, and a test set containing 20%. To achieve this, scikit-learn's 'train test split' module was used twice. The first usage was to split the data into two groups consisting of 80% and 20% of the total data. Following this, the larger set was split once again into two groups containing 60% and 20% of the total data. The purpose of each set was as follows:

- Training set: used to train the separate models and hyperparameters
- Validation set: used to evaluate which set of hyperparameters produces the best model
- Testing set: used for the final evaluation of the best-performing model

Another approach was considered for splitting the data into train and test sets. This approach involved organizing each set by each subject's run number. As stated previously, each subject performed four runs under each task. The first three of these runs are used for the training set while the fourth would be used for the test set.

For our classification model training, we explored commonly used Multivoxel Pattern Analysis (MVPA) models. These models included SVM, Naive Bayes, logistic regression, and k-nearest neighbors. Each of these models are then evaluated and compared with a base-

line SVM model, which was generated using the 'LIBSVM' software (a library for support vector machines). The accuracies for each model were calculated in order to evaluate them. We determined that accuracy would be the most crucial metric for evaluating the performance of these models. This is due to the number of labels for each category being roughly equivalent to each other, therefore we would want our classifiers to predict the correct class and have the highest accuracy possible.

To further validate our models, we performed a simple permutation test by shuffling the category labels and checking for test accuracy. This allows us to see how the model would perform under a null distribution, where we would expect to see a model accuracy about 50%. This accuracy would be the approximate accuracy if there were no significant relationships between the model features and the predictions. Ensuring that our model matches this low accuracy with randomized labels makes sure that our models are correctly built, without any artificial increases in accuracy.

For the logistic regression model, performing the custom split of using 3 runs for the train set and 1 run for the test set per subject resulted in high accuracy and additional interventions were not necessary. However for the remaining models, we applied a few specific additional methods. These methods are as follows:

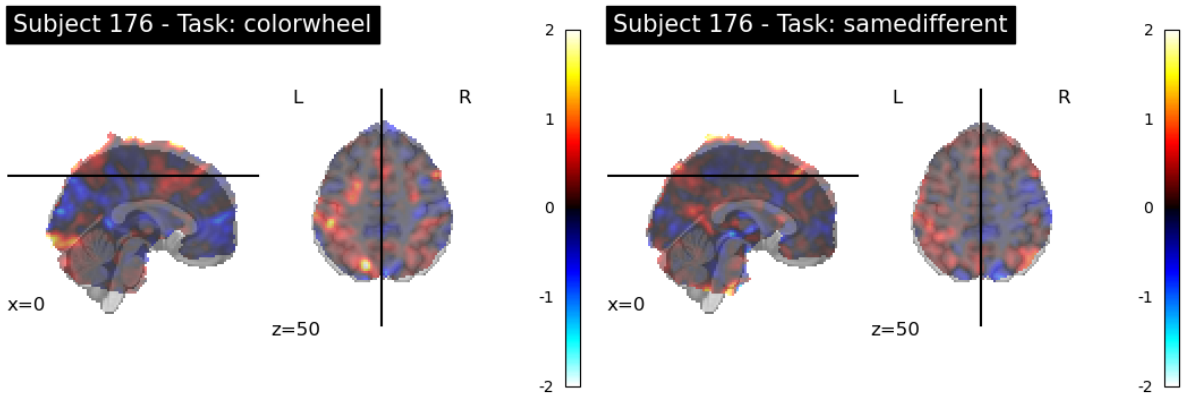
- Naive Bayes: To improve the accuracy of the naive bayes model, we used scikit-learn's gridsearchCV in order to find the optimal value for the variable smoothing hyperparameter. Following this, we applied a principal component analysis (PCA) on the features using 30 components, since Naive Bayes benefits a lot from having less features.
- kNN: The approach that produced the best accuracy for k-nearest neighbors was the custom split of using 3 runs for the train set and 1 run for the test set, per subject. Additionally, different values for the 'k' parameter were experimented with by running multiple iterations of the model while incrementing the value of 'k' by 2 each time. The optimal value of 'k' was found to be 19.

Finally, we chose to visualize the coefficient of the logistic regression model as it had the most interpretable weights amongst all the models we trained. Through this visualization by plotting the coefficients back to its corresponding brain regions, we can see which specific regions of the brain contribute to the model prediction. This was done by using scikit-learn's coefficient function to get the array of model coefficients, then inversely transforming with Nilearn's built in function to transform the 1D data back to an image in brain space. We then plotted the 3D image for visual analysis.

3 Results

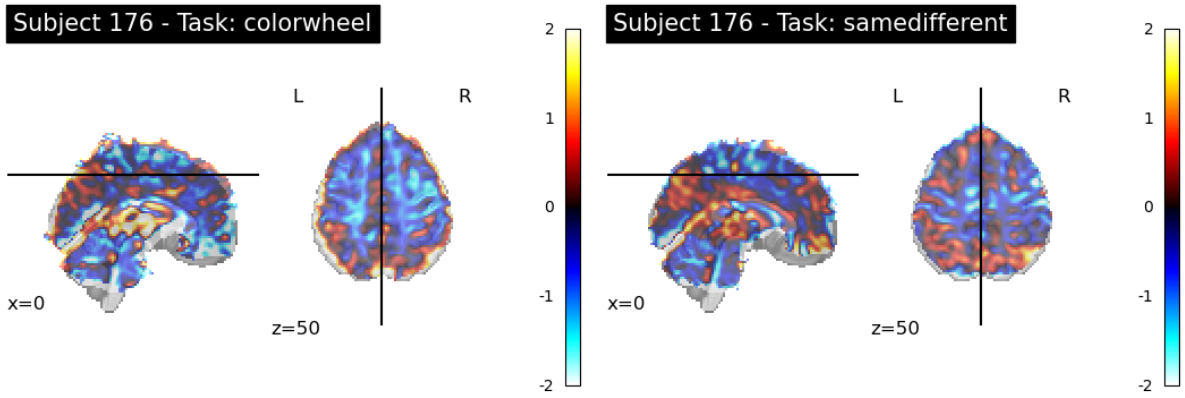
Figures 1a, 1b, 2a, 2b show the confounded and nonconfounded beta weights for a single subject in a random run of each task. We can observe that the confounded beta weights (Figures 1a, 1b) have more toned down weights due to extra noise information included in the calculation of the correlation. In the unconfounded beta weights, we can more clearly distinguish between blue and red regions, though the difference between the two tasks are

not quite separable to the human eye.



(a) Figure 1a (b) Figure 1b

Figure 1: The images of confounded beta weights for every voxel in a brain. This is the brain image for a single subject, for a single run.



(a) Figure 2a (b) Figure 2b

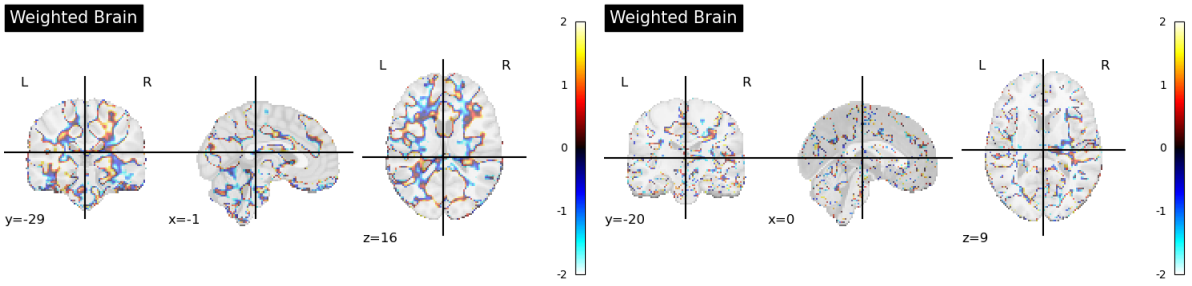
Figure 2: The images of unconfounded beta weights for every voxel in a brain. This is the brain image for a single subject, for a single run.

3.1 Specific Model Results

3.1.1 SVM

Our SVM classification model achieved an accuracy of 0.97 for unconfounded beta weights and 0.94 for confounded beta weights when training a random 3 runs from every subject, and tested on the remaining run. On the other hand, the model achieved a slightly lower accuracy at 0.84 for unconfounded and 0.92 for confounded when trained with the random 80% split of the training data. For the SVM models, we attempted to plot the coefficients by

extracting the dual coefficients and retrieving the support vectors to calculate the weights. The resulting graphs (Figures 3a, 3b) showed us that our trained SVM model weighs a lot more on the confounded regions (white matter) when trained with the extra confounds.

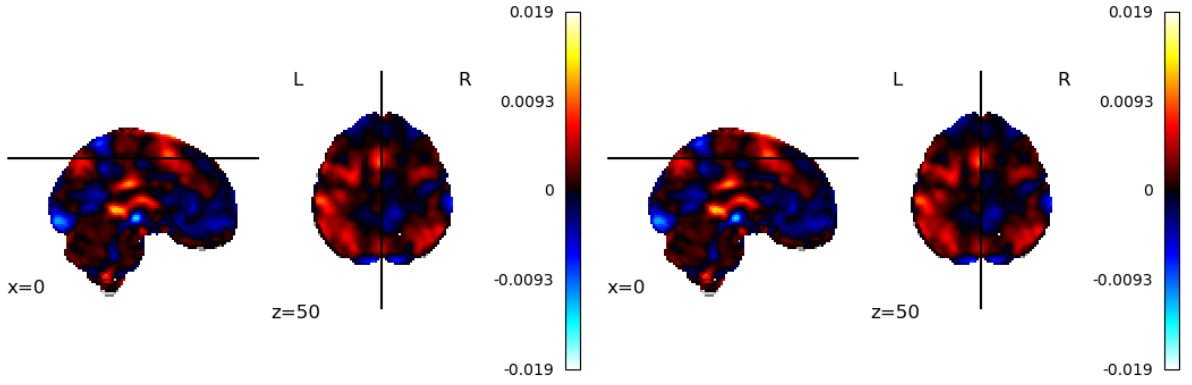


(a) Figure 3a (b) Figure 3b

Figure 3: Brain map of the SVM model coefficients. Data containing the confounded beta weights on the left and unconfounded on the right.

3.1.2 Logistic Regression

Our logistic regression model achieved an accuracy of 0.97 for unconfounded beta weights and 0.94 for confounded beta weights when using 3 runs for the training set and 1 run for the test set for each subject along with the masker. The model achieved a high accuracy of 0.96 for unconfounded beta weights even with a random 80% train, 20% test split, however the custom split produced the best results for the confounded and unconfounded data. We extracted the coefficients produced by the model and used the 'inverse transform' feature of the masker to plot 3D maps and examine the results.



(a) Figure 4a (b) Figure 4b

Figure 4: Brain map of the logistic regression model coefficients. Data containing the confounded beta weights on the left and unconfounded on the right.

3.1.3 Naive Bayes

The first iteration of naive bayes models trained was the default gaussian naive bayes model, with no changes in the variable smoothing hyperparameter. The default setting for this hyperparameter is 10^{-9}, almost 0. This results in an accuracy of 0.389 when training with confounds and 0.574 without confounds. Following this, we found the best value for variable smoothing through the use of a grid search with both confounds and nonconfounds. This results in a variable smoothing of 0.00231 for confounds and $6.58 * 10^{-5}$ for nonconfounds. This increases model accuracy, yielding a 0.759 accuracy for confounds and 0.630 for nonconfounds. Finally, we combined the results from the gridsearch with a PCA on the naive bayes features using 30 components. The final accuracies found using this method were 0.815 when using confounds and 0.556 when not using confounds.

3.1.4 kNN

The first iteration of kNN was trained using the default model and parameters, which uses a 'k' value of 5. Different values of 'k' were experimented with and 19 was ultimately decided on as it resulted in the highest accuracy. Once the custom split of using 3 runs for the training set and 1 run for the test set for each subject was applied along with the masker, the final results of the kNN model are as follows: 0.8 with confounds and 0.84 without confounds.

3.2 Final Model Results

The table below shows the final results for each model, with and without confounds. These results are specifically for the 3 runs train, 1 run test format.

Model	Accuracy	Accuracy (with confounds)
Baseline SVM	0.97	0.94
Logistic Regression	0.97	0.94
Naive Bayes	0.58	0.63
kNN	0.84	0.80

Figure 5: Comparison of task fMRI classification performances across 4 models

4 Discussion

Our project aimed to explore multiple different models commonly used in MVPA to explore whether two visual tasks that are fundamentally the same except for the final memory

retrieval stage can be distinctly classified. Furthermore, we explored the impact on model accuracy with the addition of confounds acquired through an fMRI scan in the training data.We also sought to increase mechanistic interpretability by plotting the model weights back into a brain image. Through these various experiments with data manipulation and model coefficients, we can understand more about which regions of the brain are relevant for specific tasks, rather than simply noting how accurately a model can predict some task.

4.1 Confounds Exploration

Our model performance only improved for Naive Bayes with the addition of noise to the denoise brain image, while decreasing in accuracy for all other models. We speculated that the confounds may be correlated to the two tasks, and our model may be highly overfitting to irrelevant noise. From the reconstructed coefficients retrieved from the SVM model (Figures 3a, 3b), we can see that the addition of confounds allows the model to overly rely on white matter regions rather than gray matter regions (which is where most neural activity lies). The addition of confounds decreases our ability to interpret the underlying factors in shaping prediction accuracy. Thus, we focused on the non-confounded dataset for the coefficient brain map.

4.2 Understanding Anatomical Relationships

The resulting coefficient brain image from logistic regression is aligned with the expected brain regions for the given two tasks. As the task 'samedifferent' requires subjects to respond in a binary 'yes or no' to whether a color was the same as they remembered, it is considered a more abstract, semantic task. On the other hand, the task 'colorwheel' requires subjects to use more visual detail strategies to recall a memorized color. (Hamidi, Tononi and Postle 2009) found that participants performing a change detection memory task had impaired behavior when their prefrontal cortex was inhibited but not when their intraparietal cortex was inhibited; however, (Mackey and Curtis 2017) found that inhibition to intraparietal cortex would impair continuous report memory task performance, but not inhibition to the prefrontal cortex. We expected to see parietal regions or early visual cortices be associated with continuous report tasks that demand more visual detail resources. Similarly, we expect to see that the prefrontal cortex that typically plays a role in later stages of visual processes to be highly correlated with more abstract, change detection visual memory tasks. We note that the task 'samedifferent' is used as 'Label 1' while the task 'colorwheel' is 'Label 0'. In Figures 4a we note that the more positive weights (red) should be associated with predicting Label 1, while more negative weights (blue) would indicate more weight on predicting Label 0. We can note that the red regions seem to align closely with the prefrontal cortex as seen in Figures 4b, while the blue regions are mostly clustered in the back of the brain, where the early visual cortex lies. From our model coefficients, we can interpret that predicting the 'samedifferent' task is highly correlated with the beta weights from the prefrontal cortex (red region), while the 'colorwheel' task is associated with the early visual cortex (blue region). These findings line up with previous literature on the neural correlates of such

visual working memory tasks.

4.3 Discussion of Specific Models

4.3.1 Logistic Regression Performance

The logistic regression model demonstrated the best performance, achieving an accuracy of 0.94 with confounds and 0.97 without confounds. Logistic regression is a linear model that estimates probabilities using a logistic function, therefore it is robust to the large number of features associated with our dataset. Despite having so many features, logistic regression can effectively weigh each feature's importance through its coefficients, helping to identify relevant patterns that distinguish between different cognitive tasks or conditions reflected in the brain images. Logistic regression is also considered to be less prone to overfitting compared to other models. Additionally, the coefficients that resulted from this model were interpretable and we were able to plot them in order to visualize which specific regions of the brain contributed to the model prediction. We can conclude that logistic regression is a good choice when analyzing data of this format.

4.3.2 Naive Bayes Performance

As stated in the results section, our Naive Bayes model was by far the worst performing model among the ones we tested. The reason for this poor performance is that the naive bayes models operate under the assumption that all features are independent of each other. However, this is not the case for the voxels within fMRI. In fMRI, the brain is mapped via hundreds of thousands of voxels, and these voxels are not necessarily independent of each other. Voxels within the same brain region (such as the interparietal cortex) are going to have highly correlated with each other. Within our model, this correlation is apparent for almost every feature. These correlations make the conditional independence assumption not true, worsening the accuracy of a naive bayes model.

The presence of highly correlated features also explains the fact that the Naive Bayes model performed slightly better after performing PCA on the features. PCA reduces the number of features by taking the top eigenvectors from the sample set. This reduces the amount of features in a given sample by reducing the amount of correlation within the features. This helps remedy the effect of the highly correlated features on the naive bayes model. However, PCA does not always maintain all of the relevant information while reducing the number of features. Therefore, while PCA does help solve the issue of conditional independence, it is not enough to make naive bayes a top model in MVPA.

4.3.3 kNN Performance

The performance of the kNN model on our data was average compared to the other models. kNN models generally do not perform too well when working with data with many features

or in high-dimensional spaces. As the number of features increase, the distance between points becomes less distinctive, making it challenging for kNN to effectively identify and classify neighboring points based on their similarity. The presence of confounds also made a clear difference for this model. Adding confounds added noise to the data and potentially hindered some of the underlying patterns that the model attempted to learn, therefore slightly reducing the accuracy when they are included. On the other hand, removing the confounds led to a clearer pattern and less noisy data, allowing the kNN model to perform slightly better and achieve an increase in accuracy of around 4%.

4.4 Future Work

Our project is only a preliminary work of exploration within MVPA for visual working memory tasks. Future work can be extended in both the domain of neuroscience and data science. Within neuroscience, we see that MVPA can be a way of locating brain regions for future analysis without making prior assumptions. Furthermore, it suggests that there may be fundamental differences in neural pathways between abstract versus visual detail working memory tasks. In terms of MVPA as a technique, we can use it to locate regions of interest and perform dimension reduction by reducing the amount of voxels in training data and improving computational efficiency. Another experiment can take the same data that we have and explore more into model performance without these highly weighted regions. A project dedicated to exploring the meaningfulness of fMRI confounds is also worth exploring, as our current analysis can only reduce it to a matter of model overfitting. It could be worth exploring further into whether these confounds actually contain important information about the brain under visual tasks. Another direction that we did not dive into was looking into the time series data involved in fMRI data. Future studies can explore how the classification of task-fMRI data is impacted with extracting the brain voxels during specific timestamps of stimulus or probe onset.

5 Conclusion

To conclude, our project demonstrates the capability of using different machine learning models to achieve high classification accuracy while distinguishing between specific brain regions associated with different visual memory tasks. We were able to successfully use voxel-based weights from first-level fMRI analyses for multivoxel pattern analysis (MVPA). The logistic regression model specifically provided a high level of interpretability through its coefficients, which allowed us to visually map specific brain regions with different cognitive tasks.

This result supports the utility of MVPA as a powerful tool in neuroscience for mapping relevant brain areas without prior assumptions about their functionality. Our findings suggest potential fundamental differences in neural circuitry between tasks that require abstract versus detail-oriented visual memory strategies, indicating distinct pathways for processing different types of visual information.

Additionally, our approach demonstrates how MVPA can be used to effectively identify and optimize the study of brain regions. By reducing the number of voxels included in the training data, we enhanced computational efficiency and were able to focus on analyzing the most crucial features.

These results not only confirm the utility of using such machine learning models in task fMRI classification, but also lead to possibilities for future research. We found that MVPA is able to set the stage for more detailed research into how the brain handles cognitive tasks and provide a guide for improving the accuracy and efficiency of analyzing fMRI data.

6 Contributions

Judel - started a paper draft with bullet points, wrote the Abstract and Methods section. Performed ⅓ of first-level analysis, and explored naive bayes models, worked with TA to help resolve DSMLP storage issues. Wrote about naive bayes results, methods and discussion. Worked on website background section, and general website formatting.

Jennifer - wrote the Introduction, the Dataset Description, the first-level analysis subsection and the coefficient brain map section in the Methods, and the Discussion section. Performed ⅓ of the first-level analysis, and explored SVM models, worked with TA Gabriel on data preprocessing and DSMLP storage issues. Worked on finalizing the poster and the writing for the website sections.

Sahana - Worked on the Methods, logistic regression and KNN results/discussion, and conclusion sections of the report. Performed ⅓ of first-level analysis, worked on feature vector generation, trained several different baseline models. Experimented with different methods and parameters for training logistic regression and KNN. Worked with the TA to resolve DSMLP storage issues. Converted the report into LaTex format and added the references and appendix.

7 Appendix

References

Abraham, Alexandre, Fabian Pedregosa, Michael Eickenberg, Philippe Gervais, Andreas Mueller, Jean Kossaifi, Alexandre Gramfort, Bertrand Thirion, and Gael Varoquaux. 2014. "Machine learning for neuroimaging with scikit-learn." *Frontiers in Neuroinformatics* 8. [Link]

Aglieri, Virginia, Bastien Cagna, Lionel Velly, Sylvain Takerkart, and Pascal Belin. 2021. "FMRI-based identity classification accuracy in left temporal and frontal regions predicts speaker recognition performance." *Scientific Reports* 11. [Link]

Avants, B.B., C.L. Epstein, M. Grossman, and J.C. Gee. 2008. "Symmetric diffeomorphic

image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain." *Medical Image Analysis* 12 (1): 26–41. [Link]

Behzadi, Yashar, Khaled Restom, Joy Liau, and Thomas T. Liu. 2007. "A component based noise correction method (CompCor) for BOLD and perfusion based fMRI." *NeuroImage* 37 (1): 90–101. [Link]

Ciric, R., William H. Thompson, R. Lorenz, M. Goncalves, E. MacNicol, C. J. Markiewicz, Y. O. Halchenko, S. S. Ghosh, K. J. Gorgolewski, R. A. Poldrack, and O. Esteban. 2022. "TemplateFlow: FAIR-sharing of multi-scale, multi-species brain models." *Nature Methods* 19: 1568–1571. [Link]

Esteban, Oscar, Ross Blair, Christopher J. Markiewicz, Shoshana L. Berleant, Craig Moodie, Feilong Ma, Ayse Ilkay Isik, Asier Erramuzpe, Mathias Kent, James D. and-Goncalves, Elizabeth DuPre, Kevin R. Sitek, Daniel E. P. Gomez, Daniel J. Lurie, Zhifang Ye, Russell A. Poldrack, and Krzysztof J. Gorgolewski. 2018. "fMRIPrep." *Software*. [Link]

Esteban, Oscar, Christopher Markiewicz, Ross W Blair, Craig Moodie, Ayse Ilkay Isik, Asier Erramuzpe Aliaga, James Kent, Mathias Goncalves, Elizabeth DuPre, Madeleine Snyder, Hiroyuki Oya, Satrajit Ghosh, Jessey Wright, Joke Durnez, Russell Poldrack, and Krzysztof Jacek Gorgolewski. 2019. "fMRIPrep: a robust preprocessing pipeline for functional MRI." *Nature Methods* 16: 111–116. [Link]

Evans, AC, AL Janke, DL Collins, and S Baillet. 2012. "Brain templates and atlases." *NeuroImage* 62 (2): 911–922. [Link]

Fonov, VS, AC Evans, RC McKinstry, CR Almli, and DL Collins. 2009. "Unbiased nonlinear average age-appropriate brain templates from birth to adulthood." *NeuroImage* 47, Supplement 1, p. S102. [Link]

Gorgolewski, K., C. D. Burns, C. Madison, D. Clark, Y. O. Halchenko, M. L. Waskom, and S. Ghosh. 2011. "Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in Python." *Frontiers in Neuroinformatics* 5, p. 13. [Link]

Gorgolewski, Krzysztof J., Oscar Esteban, Christopher J. Markiewicz, Erik Ziegler, David Gage Ellis, Michael Philipp Notter, Dorota Jarecka, Hans Johnson, Christopher Burns, Alexandre Manhães-Savio, Carlo Hamalainen, Benjamin Yvernault, Taylor Salo, Kesshi Jordan, Mathias Goncalves, Michael Waskom, Daniel Clark, Jason Wong, Fred Loney, Marc Modat, Blake E Dewey, Cindee Madison, Matteo Visconti di Oleggio Castello, Michael G. Clark, Michael Dayan, Dav Clark, Anisha Keshavan, Basile Pinsard, Alexandre Gramfort, Shoshana Berleant, Dylan M. Nielson, Salma Bougacha, Gael Varoquaux, Ben Cipollini, Ross Markello, Ariel Rokem, Brendan Moloney, Yaroslav O. Halchenko, Demian Wassermann, Michael Hanke, Christian Horea, Jakub Kaczmarzyk, Gilles de Hollander, Elizabeth DuPre, Ashley Gillman, David Mordom, Colin Buchanan, Rosalia Tungaraza, Wolfgang M. Pauli, Shariq Iqbal, Sharad Sikka, Matteo Mancini, Yannick Schwartz, Ian B. Malone, Mathieu Dubois, Caroline Frohlich, David Welch, Jessica Forbes, James Kent, Aimi Watanabe, Chad Cumba, Julia M. Huntenburg, Erik Kastman, B. Nolan Nichols, Arman

Eshaghi, Daniel Ginsburg, Alexander Schaefer, Benjamin Acland, Steven Giavasis, Jens Kleesiek, Drew Erickson, René Küttner, Christian Haselgrove, Carlos Correa, Ali Ghayoor, Franz Liem, Jarrod Millman, Daniel Haehn, Jeff Lai, Dale Zhou, Ross Blair, Tristan Glatard, Mandy Renfro, Siqi Liu, Ari E. Kahn, Fernando Pérez-García, William Triplett, Leonie Lampe, Jörg Stadler, Xiang-Zhen Kong, Michael Hallquist, Andrey Chetverikov, John Salvatore, Anne Park, Russell Poldrack, R. Cameron Craddock, Souheil Inati, Oliver Hinds, Gavin Cooper, L. Nathan Perkins, Ana Marina, Aaron Mattfeld, Maxime Noel, Lukas Snoek, K Matsubara, Brian Cheung, Simon Rothmei, Sebastian Urchs, Joke Durnez, Fred Mertz, Daniel Geisler, Andrew Floren, Stephan Gerhard, Paul Sharp, Miguel Molina-Romero, Alejandro Weinstein, William Broderick, Victor Saase, Sami Kristian Andberg, Robbert Harms, Kai Schlamp, Jaime Arias, Dimitri Papadopoulos Orfanos, Claire Tarbert, Arielle Tambini, Alejandro De La Vega, Thomas Nickson, Matthew Brett, Marcel Falkiewicz, Kornelius Podranski, Janosch Linkersdörfer, Guillaume Flandin, Eduard Ort, Dmitry Shachnev, Daniel McNamee, Andrew Davison, Jan Varada, Isaac Schwabacher, John Pellman, Martin Perez-Guevara, Ranjeet Khanuja, Nicolas Pannetier, Conor McDermottroe, and Satrajit Ghosh. 2018. "Nipype." *Software*. [Link]

Greve, Douglas N, and Bruce Fischl. 2009. "Accurate and robust brain image alignment using boundary-based registration." *NeuroImage* 48 (1): 63–72. [Link]

Hamidi, Massihullah, Giulio Tononi, and Bradley R. Postle. 2009. "Evaluating the role of prefrontal and parietal cortices in memory-guided response with repetitive transcranial magnetic stimulation." *Neuropsychologia* 47 (2): 295–302. [Link]

van 't Hof, Sophie R, Lukas Van Oudenhove, Erick Janssen, Sanja Klein, Marianne C Reddan, Philip A Kragel, Rudolf Stark, and Tor D Wager. 2021. "The brain activation-based sexual image classifier (BASIC): a sensitive and specific fMRI activity pattern for sexual image processing." *Cerebral Cortex* 32 (14): 3014–3030. [Link]

Jenkinson, Mark, Peter Bannister, Michael Brady, and Stephen Smith. 2002. "Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images." *NeuroImage* 17 (2): 825–841. [Link]

Jenkinson, Mark, and Stephen Smith. 2001. "A global optimisation method for robust affine registration of brain images." *Medical Image Analysis* 5 (2): 143–156. [Link]

Mackey, Wayne E., and Clayton E. Curtis. 2017. "Distinct contributions by frontal and parietal cortices support working memory." *Scientific Reports* 7 (1), p. 6188. [Link]

Mahmoudi, Abdelhak, Sylvain Takerkart, Fakhita Regragui, Driss Boussaoud, and Andrea Brovelli. 2012. "Multivoxel Pattern Analysis for fMRI Data: A Review." *Computational and Mathematical Methods in Medicine* 2012 (1), p. 961257. [Link]

Patriat, Rémi, Richard C. Reynolds, and Rasmus M. Birn. 2017. "An improved model of motion-related signal changes in fMRI." *NeuroImage* 144, Part A: 74–82. [Link]

Pereira, Francisco, Tom Mitchell, and Matthew Botvinick. 2009. "Machine learning classifiers and fMRI: A tutorial overview." *NeuroImage* 45 (1, Supplement 1): S199–S209. [Link]

Power, Jonathan D., Anish Mitra, Timothy O. Laumann, Abraham Z. Snyder, Bradley L. Schlaggar, and Steven E. Petersen. 2014. "Methods to detect, characterize, and remove motion artifact in resting state fMRI." *NeuroImage* 84 (Supplement C): 320–341. [Link]

Satterthwaite, Theodore D., Mark A. Elliott, Raphael T. Gerraty, Kosha Ruparel, James Loughead, Monica E. Calkins, Simon B. Eickhoff, Hakon Hakonarson, Ruben C. Gur, Raquel E. Gur, and Daniel H. Wolf. 2013. "An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data." *NeuroImage* 64 (1): 240–256. [Link]

Scimeca, Jason, Anastasia Kiyonaga, and Mark D'Esposito. 2018. "Dissociating the causal roles of frontal and parietal cortex in working memory capacity [Registered Report Stage 1 - Protocol]." *Springer Nature*. [Link]

Tustison, N. J., B. B. Avants, P. A. Cook, Y. Zheng, A. Egan, P. A. Yushkevich, and J. C. Gee. 2010. "N4ITK: Improved N3 Bias Correction." *IEEE Transactions on Medical Imaging* 29 (6): 1310–1320. [Link]

Zhang, Y., M. Brady, and S. Smith. 2001. "Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm." *IEEE Transactions on Medical Imaging* 20 (1): 45–57. [Link]

Appendices

A.1 Quarter 2 Project Proposal

Broad Problem Statement

Functional MRI (fMRI) allows researchers to map patterns of brain activity to specific thoughts or experiences, offering valuable insights into how the brain encodes and processes information. Working memory is the brain's ability to temporarily store and manipulate information while doing a task, and it is crucial for tasks such as reasoning, decision-making, and learning. One leading theory, the sensorimotor recruitment theory, proposes that working memory relies on the same brain regions responsible for perceiving the information. For example, visual details would be stored in the visual cortex, while abstract information would be maintained in high-level brain regions in the front.

To test whether this theory applies universally across task types, we propose developing a model to classify specific working memory tasks based on fMRI activity. By furthering our understanding of how and where the brain stores information, this work could provide groundbreaking insights into Working Memory and contribute to the growing potential of neuroscience to further understand the relationship between brain activity and function. For future work, the model can be used to further understand whether the sensorimotor recruitment theory is applicable to all working memory tasks that have brain imaging.

Narrow Problem Statement

Previous neuroimaging studies on visual Working Memory maintenance sensitivity to task type do not always target the same regions to analyze (for instance, looking into Early Visual Cortex versus intraparietal cortex for low-level, fine detail processing). Moreover, even when studies focus on specific regions (i.e. intraparietal for low-level and prefrontal for high-level), the exact areas are not clearly defined and can vary or overlap across studies. Thus, by delineating an exact voxel cluster of significance, we aim to deploy a computational model that can predict the visual Working Memory task type (high visual detail or abstract semantic) given a fMRI brain imaging.

In our Quarter 1 project, we mainly focused on locating those clusters for one given task (Change Detection, a task that uses more abstract, semantic demands). We want to expand on simply locating a certain brain region associated with some task for a single subject, and train a supervised neural network model to identify whether we can predict task type based

on BOLD signals in fMRI data. We plan to run the first level analysis on multiple subject data (as we did in Quarter 1) to get the weights for within-subject variability, then run our neural network on extracted features. We also plan on using the first three runs of each experiment as training data, and the last run for test data. As the two tasks in this experiment are fairly similar, we wonder if their neural circuitry can be discerned and classified. If the model performance is high, this may suggest that there are fundamentally different mechanisms and neural circuits. However, if the model performance is not better than chance, this suggests that the two tasks use similar circuitry and may lead to more discussion on whether fine-detail tasks versus abstract tasks inherently involve similar circuitry, but have different feedforward and feedback mechanisms.

Statement of primary output

For our quarter two project, we plan on creating a classification model that takes in brain imaging data (fMRI) and uses it to predict what task was done in order to produce that data. With this, we plan on being able to establish a strong relationship between the fMRI data and the task. To strongly show this, an ideal form of output would be one that easily supports visualization and interactivity. Therefore, we will create a website to represent our project. On the website, the user will be able to directly interact with our model, by controlling what data to use and classify. Being able to interact with a model, rather than simply reading about its results will strengthen the relation shown between fMRI data and its corresponding task classification.

A.2 fMRI Prep Boilerplate

Results included in this manuscript come from preprocessing performed using *fMRIPrep* 24.1.1 (Esteban et al. (2019); Esteban et al. (2018); RRID:SCR_016216), which is based on *Nipype* 1.8.6 (Gorgolewski et al. (2011); Gorgolewski et al. (2018); RRID:SCR_002502).

Anatomical data preprocessing A total of 1 T1-weighted (T1w) images were found within the input BIDS dataset. The T1w image was corrected for intensity non-uniformity (INU) with `N4BiasFieldCorrection` (Tustison et al. 2010), distributed with ANTs 2.5.3 (Avants et al. 2008, RRID:SCR_004757), and used as T1w-reference throughout the workflow. The T1w-reference was then skull-stripped with a *Nipype* implementation of the `antsBrainExtraction.sh` workflow (from ANTs), using OASIS30ANTs as target template. Brain tissue segmentation of cerebrospinal fluid (CSF), white-matter (WM) and gray-matter (GM) was performed on the brain-extracted T1w using `fast` (FSL (version unknown), RRID:SCR_002823, Zhang, Brady and Smith 2001). Volume-based spatial normalization to two standard spaces (MNI152NLin6Asym, MNI152NLin2009cAsym) was performed through nonlinear registration with `antsRegistration` (ANTs 2.5.3), using brain-extracted versions of both T1w reference and the T1w template. The following templates were were selected for spatial normalization and

accessed with *TemplateFlow* (24.2.0, Ciric et al. 2022): *FSL's MNI ICBM 152 nonlinear 6th Generation Asymmetric Average Brain Stereotaxic Registration Model* [Evans et al. (2012), RRID:SCR_002823; TemplateFlow ID: MNI152NLin6Asym], *ICBM 152 Nonlinear Asymmetrical template version 2009c* [Fonov et al. (2009), RRID:SCR_008796; TemplateFlow ID: MNI152NLin2009cAsym].

Functional data preprocessing For each of the 1 BOLD runs found per subject (across all tasks and sessions), the following preprocessing was performed. First, a reference volume was generated, using a custom methodology of *fMRIPrep*, for use in head motion correction. Head-motion parameters with respect to the BOLD reference (transformation matrices, and six corresponding rotation and translation parameters) are estimated before any spatiotemporal filtering using `mcflirt` (FSL , Jenkinson et al. 2002). The BOLD reference was then co-registered to the T1w reference using `mri_coreg` (FreeSurfer) followed by `flirt` (FSL , Jenkinson and Smith 2001) with the boundary-based registration (Greve and Fischl 2009) cost-function. Co-registration was configured with six degrees of freedom. Several confounding time-series were calculated based on the *preprocessed BOLD*: framewise displacement (FD), DVARS and three region-wise global signals. FD was computed using two formulations following Power (absolute sum of relative motions, Power et al. (2014)) and Jenkinson (relative root mean square displacement between affines, Jenkinson et al. (2002)). FD and DVARS are calculated for each functional run, both using their implementations in *Nipype* (following the definitions by Power et al. 2014). The three global signals are extracted within the CSF, the WM, and the whole-brain masks. Additionally, a set of physiological regressors were extracted to allow for component-based noise correction (*CompCor*, Behzadi et al. 2007). Principal components are estimated after high-pass filtering the *preprocessed BOLD* time-series (using a discrete cosine filter with 128s cut-off) for the two *CompCor* variants: temporal (tCompCor) and anatomical (aCompCor). tCompCor components are then calculated from the top 2% variable voxels within the brain mask. For aCompCor, three probabilistic masks (CSF, WM and combined CSF+WM) are generated in anatomical space. The implementation differs from that of Behzadi et al. in that instead of eroding the masks by 2 pixels on BOLD space, a mask of pixels that likely contain a volume fraction of GM is subtracted from the aCompCor masks. This mask is obtained by thresholding the corresponding partial volume map at 0.05, and it ensures components are not extracted from voxels containing a minimal fraction of GM. Finally, these masks are resampled into BOLD space and binarized by thresholding at 0.99 (as in the original implementation). Components are also calculated separately within the WM and CSF masks. For each CompCor decomposition, the k components with the largest singular values are retained, such that the retained components' time series are sufficient to explain 50 percent of variance across the nuisance mask (CSF, WM, combined, or temporal). The remaining components are dropped from consideration. The head-motion estimates calculated in the correction

step were also placed within the corresponding confounds file. The confound time series derived from head motion estimates and global signals were expanded with the inclusion of temporal derivatives and quadratic terms for each (Satterthwaite et al. 2013). Frames that exceeded a threshold of 0.5 mm FD or 1.5 standardized DVARS were annotated as motion outliers. Additional nuisance timeseries are calculated by means of principal components analysis of the signal found within a thin band (*crown*) of voxels around the edge of the brain, as proposed by (Patriat, Reynolds and Birn 2017). All resamplings can be performed with *a single interpolation step* by composing all the pertinent transformations (i.e. head-motion transform matrices, susceptibility distortion correction when available, and co-registrations to anatomical and output spaces). Gridded (volumetric) resamplings were performed using `nitransforms`, configured with cubic B-spline interpolation.

Many internal operations of *fMRIPrep* use *Nilearn* 0.10.4 (Abraham et al. 2014, RRID:SCR_001362), mostly within the functional processing workflow. For more details of the pipeline, see the section corresponding to workflows in *fMRIPrep*'s documentation.

A.2.1 Copyright Waiver

The above boilerplate text was automatically generated by fMRIPrep with the express intention that users should copy and paste this text into their manuscripts *unchanged*. It is released under the CC0 license.