Blavatnik School of Computer Science
Tel Aviv University

Sagol School of Neuroscience
Tel Aviv University

Predict Cognitive Abilities Based on fMRI Scans During a Cognitive Task

Ido Tavor's Lab

Moderator: Maya Kadushin

Submitted by – Koren Sarig, Gili Levy, Tal Yasdi

Submitted To – Prof. Amir Globerzon, Mr. Yoav Ger

Date of Submission – 04/09/23

Introduction

In the realm of cognitive neuroscience, where individuals vary widely in their cognitive abilities affecting learning, our study aims to predict intelligence abilities, as part of a study that aims to predict academic achievements. Our study aims to predict cognitive skills through task-related brain activations using machine learning. Using fMRI data from 47 participants who completed an N-back task, we explore how brain activity correlates with cognitive abilities, estimated by Raven's Advanced Progressive Matrices Test. This research contributes to understanding brain-cognition relationships, building on previous findings in brain connectivity and cognitive prediction.

Investigating the neural foundations of interindividual variability constitutes a fundamental pursuit within the field of cognitive neuroscience. Previous studies have shown that brain connectivity measurements can predict individual cognitive and intelligence scores. It was also found that task-induced brain activation maps outperform resting-state connectivity patterns in predicting individual intelligence, suggesting that a cognitively demanding environment improves prediction of cognitive abilities. Raven's cognitive test is found in correlation with cognitive abilities and therefore is used as the predicted value, The emotional N-back test is known for engaging emotion regulation processes and memory in different low conditions, and accordingly used as the study's task.

In our study, we want to examine whether functional connectivity patterns can predict individual intelligence abilities. Personality traits are known as can be predicted from brain connectivity measurements and cognitive abilities are in correlation with intelligence, therefore our measured and predicted score is Raven's test. The brain connectivity measurements are collected using functional magnetic resonance imaging (fMRI), while performing a cognitive task. The study is carried out as a within-subject, longitudinal experiment.

At first and as part of the process, the fMRI data was pre-processed and a data contrast was created based on the N-back tests brain measurements. Then, the fMRI and Raven scores data were extracted using python libraries. Our project included building a program that trains four different regression models (Linear Regression, Random Forest, Ridge and Elastic Net) based on three different data reduction methods (PCA, supervised feature selection and unsupervised feature selection) using hyper parameterization methods. Due to the small number of subjects, the models were trained using a leave-one-out cross validation.

For each combination of data reduction method and regression model, the Pearson correlation coefficient was measured. When the PCA reducing data method was used, the features that explained most of the data variance were extracted. The features were extracted from the PCA components that explained most of the variance, along with their normalized effect measurement.

Out of the models and reduced data methods that were tested, the best predictive model was found to be the PCA based Random Forest, with r = 0.364. A post analysis on the PCA components was made to retrieve which of the data voxels has affected the data variance in the most significant way, and it was found that the frontoparietal activation and default mode network deactivation - a brain activation pattern associated with executive processing and higher cognitive demand - had such effect.

Related Prior Work

Many studies have examined the relationship between cognitive intelligence and academic achievement. A meta-analysis by Ritchie and Bates (2013) found that there is a significant positive correlation between cognitive intelligence and academic achievement, as measured by standardized test scores and grade point averages (GPAs). This suggests that people with higher cognitive intelligence tend to have higher academic achievement.

The relationship between cognitive intelligence and other cognitive abilities has been a major focus of research. A study by Deary et al. (2010) found that there is a moderate to strong positive correlation between cognitive intelligence and a variety of cognitive functions, such as memory retention, logical reasoning, and the speed of information processing. This suggests that cognitive intelligence is closely linked to other cognitive abilities.

Raven's Advanced Progressive Matrices Test is a non-verbal cognitive test. The test consists of 36 multiple choice questions containing 8 visual geometric designs (Figure 1), that according to them the subject should conclude which of the 6 to 8 possible choices is the missing piece. The Raven test score is the amount of questions answered correctly (0 to 36). The Raven Progressive Matrices test is a widely used cognitive assessment tool that measures abstract reasoning, problem-solving, and general cognitive ability (John & Raven J., 2003). A study by Raven et al. (1998)

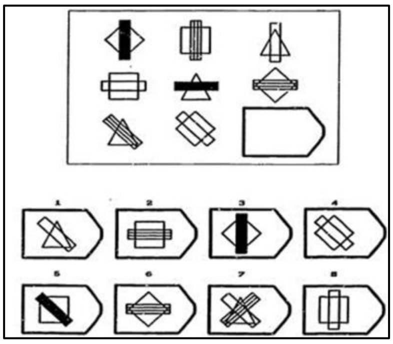


Figure 1: Raven test question example

found that the Raven test has strong predictive validity for these cognitive abilities. This means that performance on the Raven test is a good predictor of how well people will perform on other tasks that require abstract reasoning, problem-solving, and general cognitive ability.

Working memory is a crucial part of cognitive architecture. The working memory model, developed by Baddeley (1992), explains how working memory functions in a variety of cognitive tasks. This model emphasizes the importance of working memory in logical reasoning, comprehension, and problem-solving. Working memory is a fundamental cognitive ability that allows us to temporarily store and manipulate information.

The N-back task engages memory and emotion regulation processes (Owen A. M., et al., 2005). The task contains both high and low memory load conditions (the 2-back and 0-back tasks, accordingly). The stimuli includes a set of happy, fearful, neutral facial expressions and neutral places. The N-back session includes trials from each task. On each trial, participants are asked to respond as to whether the picture is a "Match" or "No Match". In the 2-back condition trials participants are instructed to respond "match" when the current stimulus is the same as the one shown two trials back. In the 0-back condition, participants are instructed to respond "match" when the current stimulus is the same as the target presented at the beginning of the block (Figure 2). Research on working memory has investigated the relationship between working memory training and performance on the N-back task. A study by Jaeggi et al. (2008) found that intensive N-back training can improve working memory capacity. This suggests that the N-back task is a valid measure of working memory capacity and that training on this task can improve working memory skills. In relation to our study, the brain activity during the task was studied in previous studies, however no prediction was yet found regarding the ability to predict intelligence based on it.

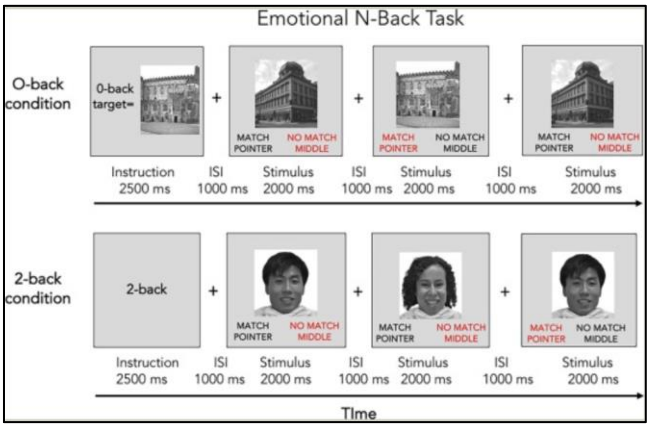


Figure 2: N-back task example

The advent of functional magnetic resonance imaging (fMRI) has made it possible to study the neural basis of cognitive functions, such as working memory and general cognitive abilities (GCA). A study by Sripada et al. found that tasks with a highly effective basis for prediction of GCA are the 2-back versus 0-back contrast. This contrast achieved high correlation with GCA scores, and 13 out of 15 task contrasts in the study afforded statistically significant prediction of GCA. A study by Owen et al. (2005) used fMRI to examine neural activity during the N-back task, a cognitive task that measures working memory capacity. The study found that the 2-back condition, which is more demanding than the 0-back condition, activates a network of brain regions that are involved in working memory, including the dorsolateral prefrontal cortex. This suggests that the contrast between 2-back task and 0-back task is a valid measure of working memory capacity and that it can be used to study the neural basis of working memory.

Previous studies have also shown that connectivity measurements derived from fMRI are predictive of cognitive scores as well as personality traits (Beaty et al., 2018; Cai et al., 2020; Dubois et al., 2018). When it comes to studying cognitive activity based on fMRI images while performing a task, and specifically when it comes to a contrast between a 2-back task and a 0-back task, Sripada et al. also found in their research that task contrasts that produce greater frontoparietal activation and default mode network deactivation—a brain activation pattern associated with executive processing and higher cognitive demand—are more effective in the prediction of GCA. In their research, Sirpada et al. extracted top three components based on statistical significance displayed so that greater expression of these components predicts higher GCA (Figure 3). These components include large activations in the supplementary motor area (SMA), precuneus, and dlPFC, as well as deactivations in anterior DMN.

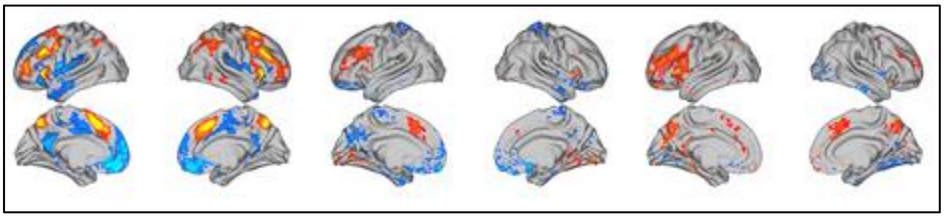


Figure 3: Visualization of the three components from the 2-Back versus 0-Back task contrast most predictive of general cognitive ability (GCA), as shown in Sirpada et al. article.

Our Research Problem and Data

In our research, we have studied the predictive relation between brain connectivity measurements and individual academic performance. Studying the predictive relationship between brain connectivity measurements and individual academic performance is important for several reasons. This knowledge could lead to personalized educational strategies and interventions that cater to the specific cognitive strengths and weaknesses of each student. The study of the relationship between brain connectivity metrics and individual academic performance is important for personalized education, early identification of learning disabilities, and advancing our understanding of the neurobiology of learning. The insights gained from this research can be used to develop more effective teaching methods and inform education policy formulation.

In order to examine the relation, prior to our study, imaging and behavioral data was collected from a cohort of undergraduate students at Tel Aviv University, who major in both Social Sciences and Life or Exact Sciences. The data was collected from 47 students at the beginning of the subjects' first academic year and will be tested against multiple timepoints throughout their undergraduate studies, as a longitudinal experiment. One participant was measured and excluded from the analysis, due to an identified pathology (subject 31).

As part of the collected measures, the participants underwent an MRI session at Tel-Aviv University's Strauss Center for Computational Neuroimaging, where brain derived measurements were collected. The measures included functional connectivity derived from resting-state fMRI, functional connectivity derived from naturalistic movie watching fMRI and task evoked brain activity, acquired while performing emotional N-back task (Casey et al., 2018). Each of the subjects' fMRI data included a 3-dimensional image that is represented by 91,282 voxels, each holding the captured mental activity. In our study, the brain based data used to predict the cognitive scores is a data contrast that was created based on the N-back tests brain measurements. The comparison was calculated based on the fMRI data recording while performing a 2-back task, subtracting the recorded brain activity from the fMRI data recording while performing a 0-back task.

Other than the brain measures, behavioral measurements were also collected. The behavioral data included each subject's Grade Point Average (GPA) and averaged scores in specific disciplines (Social Science/humanities vs. Exact or Life Sciences). The subjects have performed a Raven's Advanced Progressive Matrices Test (Raven, 1989) and their scores were collected. Participants' scores in psychological questionnaires were also recorded, including State-Trait Anxiety Inventory (STAI) (Spielberger, 1983), Patient Health Questionnaire-9 (PHQ-9) (Kroenke et al., 2001) and The Satisfaction with Life Scale (SWLS) (Diener et al., 1985). As part of the behavioral data, participants have reported social connections within their academic community.

Technical Aapproach

The code implements a data analysis pipeline (Figure 4) designed to predict cognitive test scores by fMRI data (Appendix 1). The primary goal is to illuminate the intricate relationship that exists between brain activity patterns captured by fMRI scans and cognitive performance scores. To achieve this, the code gathers machine learning techniques including dimensionality reduction, feature selection, and regression models.

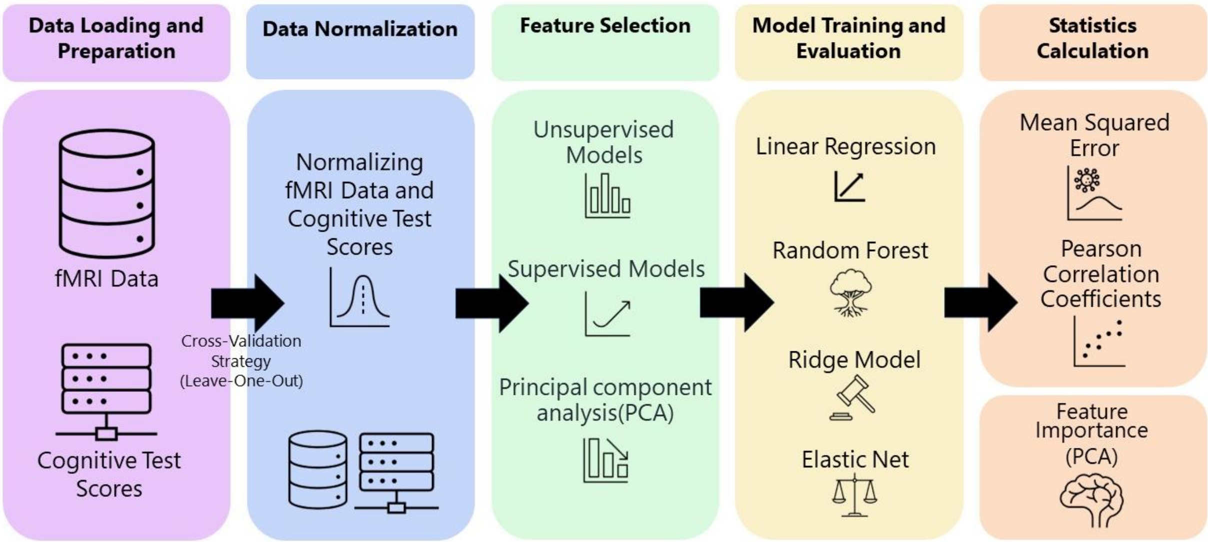


Figure 4: The pipeline design of the data analysis

At its technical core, the code starts by importing indispensable libraries, including nibabel for efficient fMRI data manipulation, numpy for essential numerical operations, and an assortment of machine learning modules from scikit-learn, including regression, feature selection, and cross-validation.

Data Loading and Preparation

The code starts by loading and preprocessing data. This involves extracting fMRI data from NifTi files, the file format used for neuroimaging. With an emphasis on accuracy, the data preprocessing tasks include loading the subjects' fMRI data. This step prepares the data for subsequent analyses.

In this section of the code, a cross-validation strategy known as Leave-One-Out (LOO) is employed to rigorously evaluate the performance of various machine learning models. The LOO approach systematically withholds one subject's data from the training set and employs the remaining subjects' data to train predictive models. The withheld subject's data are then used for testing, and this process is iterated over the entire dataset, thus facilitating an exhaustive evaluation process.

Data Normalization

To ensure uniformity in data representation, the fMRI data within the training and testing subsets undergo normalization. Normalization brings all features to a common scale, ensuring that each feature contributes proportionally to the selection process. This balanced treatment of features enhances the accuracy and fairness of feature selection by preventing one feature from exerting undue influence due to its larger magnitude. The normalization of the fMRI data is applied only when using either supervised or unsupervised feature selection methods and is not applied when using PCA, because the PCA involves computing the covariance matrix of the input data. The covariance matrix describes how the different features in the dataset vary in relation to one another. Variance-covariance matrices are scale-invariant, meaning they are not affected by changes in the scale of the data. The training scores are normalized through z-score standardization, where the scores are adjusted to have a mean of zero and a standard deviation of one.

Feature Selection, Model Training and Evaluation

In the implementation, a comprehensive exploration is conducted to uncover the most effective combination of feature selection method and machine learning model for predicting the cognitive test scores. The research approach involves systematically testing different combinations of feature selection techniques, including supervised and unsupervised methods, alongside Principal Component Analysis (PCA), a dimensionality reduction technique that captures the most relevant information from the fMRI data. These techniques are paired with a range of machine learning models, such as Linear Regression, Random Forest Regression, Ridge Regression, and ElasticNet Regression. By applying various feature selection methods and dimensionality reduction techniques to different machine learning models, the study aims to identify the optimal configuration that yields the highest predictive accuracy.

If PCA reduction is employed, the data undergo dimensionality reduction. This technique is applied to fMRI data, reducing its complexity while retaining the most salient brain activity patterns. PCA aims to find orthogonal axes that best represent the variance in the data. With the selection of 20 principal components, PCA endeavors to capture the variance within the data, thus rendering a compact representation of the neural dynamics. Alternatively, if either supervised or unsupervised feature selection methods are opted for, relevant features are retained. Supervised feature selection utilizes a RandomForestRegressor, while unsupervised selection employs SelectKBest.

The data then serves as input for subsequent regression models encompassing Linear Regression, Random Forest Regression, Ridge Regression, and ElasticNet Regression. Linear Regression is a fundamental model that establishes a linear relationship between the input features and the target variable. Random Forest Regression, on the other hand, is an ensemble technique that constructs multiple decision trees to aggregate predictions and capture intricate nonlinear relationships within the data. Ridge Regression introduces regularization to mitigate potential overfitting by adding a penalty term to the loss function, thereby achieving a balance between feature selection and preventing multicollinearity. The ElasticNet Regression extends Ridge Regression by combining L1 and L2 regularization, offering a robust solution for datasets with numerous features and collinearity.

In pursuit of optimizing model performance, a grid search strategy, known as hyperparameter tuning, is used to refine the Random Forest Regression model. Hyperparameters are the parameters that determine the behavior of a machine learning algorithm, and their values significantly impact a model's predictive capabilities. The grid search systematically iterates through a predefined set of hyperparameter combinations, evaluating each configuration's impact on the model's predictive accuracy. By considering a range of values for key parameters such as the number of trees, maximum depth, and maximum features, the grid search facilitates the identification of the most favorable combination that maximizes predictive efficacy. Hyperparameter tuning fine-tunes the model to its optimal configuration, enhancing its ability to generalize patterns from the training data to new, unseen data.

Statistics Calculation

The code employs an evaluation process to estimate the performance of the regression models and feature selection techniques. Mean squared error (MSE) serves as a pivotal metric to quantify the accuracy of predictions, representing the average squared difference between the predicted and actual cognitive test scores. A lower MSE indicates a better fit between the model's predictions and the true scores, signifying higher predictive precision. Additionally, Pearson correlation coefficients are computed to assess the linear relationship between the predicted and actual scores. A positive correlation coefficient denotes a direct linear association, with values closer to 1 indicating stronger positive relationships. These statistical assessments provide insights into the models' capabilities in capturing the complex interplay between fMRI data patterns and cognitive performance outcomes. The code produces a matrix of the correlation coefficients for each combination of regression models and feature selection methods. This matrix encapsulates the predictive outcomes of Linear Regression, Random Forest Regression, Ridge Regression, and ElasticNet Regression, coupled with Principal Component Analysis (PCA) and supervised/unsupervised feature selection. The resulting scores matrix is a reflection of the model's ability to infer cognitive test scores from the fMRI data. These predictions are essential in offering an overview of how each model configuration performs, providing valuable insights into which techniques yield the most accurate results.

Additional statistical calculation the code performs is measuring the importance of the features. It allows us to identify and rank the input variables based on their influence on model outcomes or dataset variance. In the code, the "BestFeatures" class implements this concept by leveraging principal component analysis (PCA). The principal components are ordered by their explained variance, which is the amount of variance in the original dataset that they explain. The "BestFeatures" class calculates the feature importance scores by accumulating the contributions of each feature to the explained variance ratio. This means that the more a feature contributes to the explained variance, the more important it is considered to be. The class also allows the user to specify whether to consider only positive or negative feature weights. This is useful because some features may have negative weights, which means that they are inversely correlated with the target variable. Overall, the class provides a way to extract feature importance insights from PCA results. This can help us to better understand the underlying dynamics of a dataset and make more informed decisions about data-driven projects.

Results

The study model consists of different implementation variations, including data reducing methods and regression models. As mentioned above, the data reducing methods include supervised feature selection, unsupervised feature selection and PCA. The implemented regression models are Linear Regression, Random Forest, Ridge and Elastic Net.

Each combination was trained and tested on the leave-one-out cross validation splitted data, and statistical measurements were extracted. The statistical values of each variation were calculated based on the model predicted normalized scores against the original normalized scores, and for each model the Pearson correlation coefficient value (r) was extracted (Figure 5).

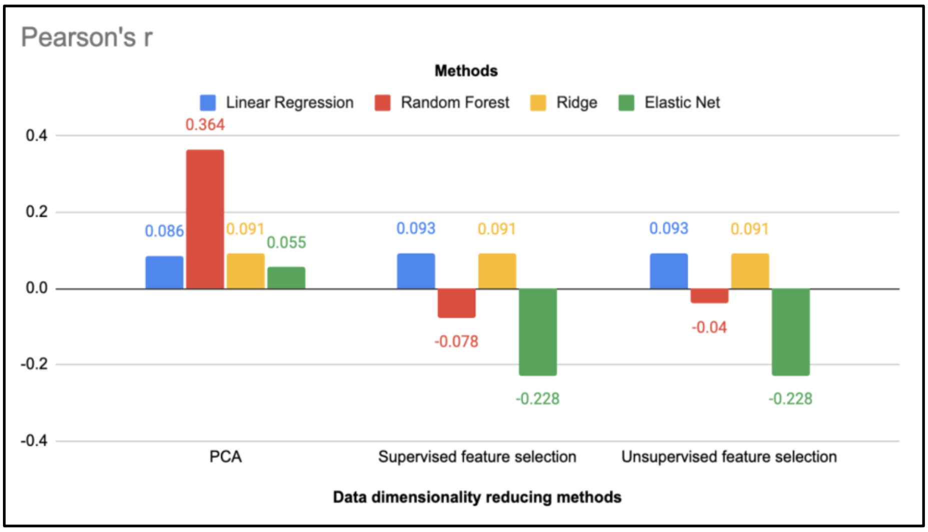


Figure 5: Pearson correlation coefficient values

Along with the Pearson correlation coefficient value, the associated p-value and the Mean squared error (MSE) were conducted. The model that was found to predict the test scores in the most accurate way according to the statistical values, was a model that includes PCA as the data reducing method and Random Forest as the regression model ($r=.364$, $p=.011$, $MSE=.975$).

In order to validate the association found according to the Pearson correlation coefficient value, a permutation test was conducted. As part of the procedure, the participants' scores data was randomly shuffled, which created different combinations of the original Raven test scores paired with random original brain activity data. The model was trained and tested on the shuffled data multiple times, which allowed creating a range of possible Pearson correlation coefficient values and examining where our initial statistic falls within the distribution (Figure 6). Based on the permutation test data and out of 53 total model executions, the original model reached the highest Pearson correlation coefficient value, and therefore the p-value stands at 1.88%.

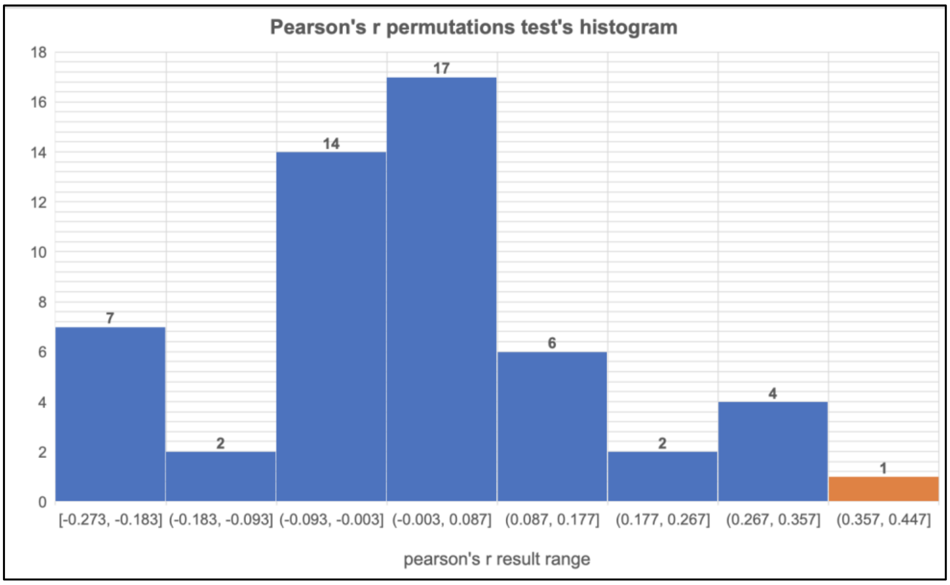


Figure 6: Permutation test distribution

After the PCA was found to be part of the most predictive model and as mentioned above, the variance explained by the PCA components was analyzed in order to understand which are the brain areas that most affect the model learned data. The 20 principal components that were created by the PCA were found to hold 69% of the data variance. Based on these components, the voxels' effect on the variance was extracted in the method described above, in order to identify which are the brain areas that hold an impact. In the analysis, the data was splitted into its positive and negative values, and was presented on a mid-thickness surface visualization (Figure 7).

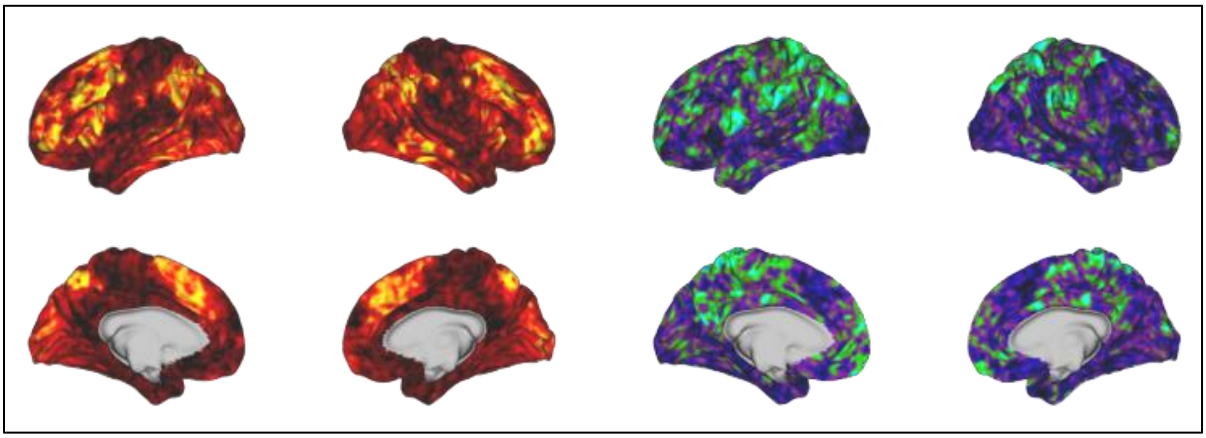


Figure 7: The variance explained by the PCA components as shown in voxels imaging. Right figure – positive values. Left figure – negative values.

According to the visualization (Appendix 2), it appears that the areas that most explain the data variance according to the PCA components calculation correspond to the previous knowledge found in science and indeed contain frontoparietal activation and default mode network deactivation (DMN). Furthermore, the motoric part, for example, is part of the areas that are inversely correlated with the target variable.

Conclusion & Discussion

In our study, a machine learning model was built in order to predict academic performance based on fMRI brain measures. The models' predicted value is the subjects' Raven test score as a cognitive test that is found in correlation with academic success. The models' data is built upon brain measurements data that were collected while performing the N-back task. The implementation includes the use of three different data reducing methods and four regression models. The model that uses PCA as the data reducing method and Random Forest as the regression model was found to have the best Pearson correlation coefficient value, with r = 0.364. The voxels that were found as expressing most of the variance in the data are frontoparietal and related to default mode network deactivation.

During the study construction, some assumptions were taken in conclusion and a few relevant further steps were not implemented due to the study limits. In this chapter, we will review some of the potential areas of development we recommend implementing in further research.

In our analysis, a contrast was created based on the recorded brain measurements while performing the N-back tests, and the contrast was used as the model data. We suggest exploring the use of other possible contrasts as the model data, such as creating a contrast between different facial expressions, as the N-back task can also be an emotional task. In that case, we suggest ascertaining whether it aids in prediction. Another possible way of action is using more of the collected measurements, other than brain activity, as the model input. For example, combining other cognitive information such as psychometric results, STAI, SWLS or PHQ-9 questionnaires might result in better predictive outcomes.

In our research, the Raven scores data was normalized based on the real scores that were defined as the test set in each Leave-one-out iteration. We believe that normalizing based on the log value of the scores might result in predicting the log train test in a more accurate way (Yang Y. & Eisenstein J., 2013).

As explained above, our model uses hyperparameter tuning in order to choose the appropriate hyperparameter values for optimal model performance results. In our model, we have used a pre-defined PCA components amount. In further research, we suggest using hyperparameter tuning to define the amount of components when choosing to use PCA as a data reducing method.

Regarding the used models, we recommend applying the hyperparameter tuning to all of the implemented models, while currently only the Random Forest model uses this approach. We also believe that other models should be considered to be implemented,; as they might improve the scores' prediction.

When using PCA as the data reducing method, the features that best explain the variance are extracted. Another possible way of understanding which brain areas are related to the ability to predict academic success, is by looking at the features that the model marks as the most relevant for prediction. In the current implementation, the PCA reduces the dimensionality of the data by finding the components that best represent the data variance. When creating the components, each feature holds its contribution to the component. While this method helps us understand the general effect of the voxels on the data, we recommend using the model based features importance for better understanding of the brain areas relevance to the prediction process.

In order to make sure the model is not overfitted due to the small number of subjects, we recommend on predicting other tests' scores instead of the Raven test. Given a different test that is found in correlation with cognitive abilities, we assume the trained model is supposed to predict the subjects' score in a way that is similar to the current model. Another possible variation for this study is exploring the option to replace the tasks the subjects perform during the fMRI scan.

As part of the statistical analysis made in our project, the p-value was verified by performing a permutation test that included randomly shuffling the train data, and training the model according to it. Our permutation test included 53 model executions, while normally it is expected to perform the test with a larger number of runs, for example, running it 1,000 times. Due to time limitations caused by our limited resources and the model execution time, we did not make it to a high number of runs, but recommend further studies to perform a wider test.

As for all machine learning models, the more data the merrier. For example, expanding the amount of the subjects and by that increasing collected data will allow building the model with different cross validation methods, such as using k-fold to split the data to the train and test sets. Using other resampling methods will reduce the possible model over-fitting.

Furthermore, as part of our current longitudinal research, the subjects' academic success should be collected, in order to validate their Raven scores correlation to their academic abilities and measure the model success.

Sources

Baddeley, A. (1992). Working memory: The interface between memory and cognition. *Journal of cognitive neuroscience*, *4*(3), 281-288.

Beaty, R. E., Kenett, Y. N., Christensen, A. P., Rosenberg, M. D., Benedek, M., Chen, Q., ... & Silvia, P. J. (2018). Robust prediction of individual creative ability from brain functional connectivity. Proceedings of the National Academy of Sciences, 115(5), 1087-1092.

Cai, H., Chen, J., Liu, S., Zhu, J., & Yu, Y. (2020). Brain functional connectome-based prediction of individual decision impulsivity. Cortex, 125, 288-298.

Dubois, J., Galdi, P., Paul, L. K., & Adolphs, R. (2018). A distributed brain network predicts general intelligence from resting-state human neuroimaging data. Philosophical Transactions of the Royal Society B: Biological Sciences, 373(1756), 20170284.

Duncan, J., Seitz, R. J., Kolodny, J., Bor, D., Herzog, H., Ahmed, A., ... & Emslie, H. (2000). A neural basis for general intelligence. *Science*, *289*(5478), 457-460.

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences*, *105*(19), 6829-6833.

Owen, A. M., McMillan, K. M., Laird, A. R., & Bullmore, E. (2005). N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. Human brain mapping, 25(1), 46-59.

Raven, J. (2003). Raven Progressive Matrices. In: McCallum, R.S. (eds) Handbook of Nonverbal Assessment. Springer, Boston, MA.

Raven, J. C., & John Hugh Court. (1998). *Raven's progressive matrices and vocabulary scales* (pp. 223-237). Oxford: Oxford Psychologists Press.

Ritchie, S. J., & Bates, T. C. (2013). Enduring links from childhood mathematics and reading achievement to adult socioeconomic status. *Psychological science*, *24*(7), 1301-1308.

Sripada, C., Angstadt, M., Rutherford, S., Taxali, A., & Shedden, K. (2020). Toward a "treadmill test" for cognition: Improved prediction of general cognitive ability from the task activated brain. Human Brain Mapping, 41(12), 3186-3197.

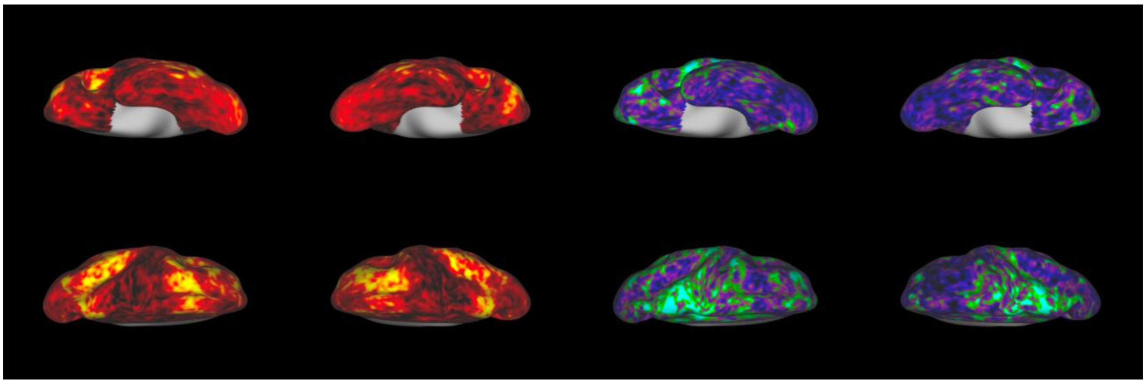
Yang, Y., & Eisenstein, J. (2013, October). A log-linear model for unsupervised text normalization. In Proceedings of the 2013 conference on empirical methods in natural language processing (pp. 61-72).

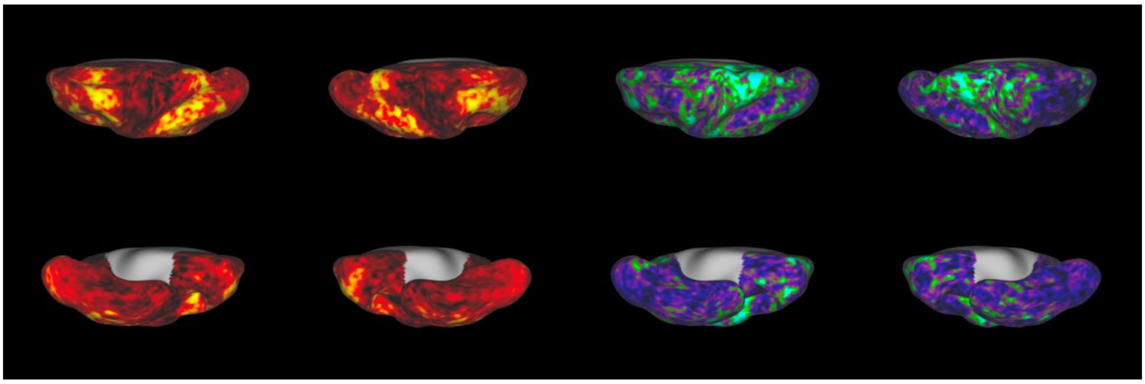
Appendices

Appendix 1 – The Code of The Project

https://github.com/Korensarig/final_project/

Appendix 2 - Voxels Imaging





In The Figures: The variance explained by the PCA components as shown in voxels imaging. Right figure – positive values. Left figure – negative values.