Neurocognitive Analysis of Schizoaffective Disorder: Differences and Cognitive Associations

Youssef A. Franci

University of California, San Diego

Author Note

Youssef A. Franci is affiliated with the Department of Cognitive Science, University of California, San Diego.

No changes in affiliation have occurred during the course of this research.

The full code for this project can be found at https://tinyurl.com/y6p2psxu

Correspondence concerning this report should be addressed to Youssef A. Franci at yfranci@ucsd.edu.

Abstract

This study aims to investigate the neurocognitive differences between individuals diagnosed with Schizoaffective disorder and a control group, and explore the potential associations between specific cognitive domains and the diagnosis. Using a dataset of neurocognitive measures, clinical diagnoses, and demographic information, we analyze 242 observations to elucidate the unique neurocognitive profile associated with Schizoaffective disorder. Our analysis focuses on assessing variations in speed, attention, memory, verbal ability, visual ability, problem-solving skills, and social cognition. We hypothesize that individuals with Schizoaffective disorder will exhibit significantly different neurocognitive performance compared to the control group. Furthermore, we aim to identify specific cognitive domains that show significant associations with the diagnosis, potentially serving as cognitive biomarkers for the disorder. By enhancing our understanding of the neurocognitive aspects of Schizoaffective disorder, we aim to contribute to improved diagnostic and treatment approaches for individuals affected by this complex psychiatric condition.

Keywords: Schizoaffective disorder, neurocognitive measures, cognitive domains, differences, associations, cognitive biomarkers.

Neurocognitive Analysis of Schizoaffective Disorder: Differences and Cognitive Associations

Psychiatric disorders pose significant challenges for both patients and clinicians due to their intricate and multifaceted nature. Understanding the underlying brain dysfunctions associated with these disorders is crucial for accurate diagnosis, effective treatment, and improved patient outcomes. Neurocognitive measures, which assess various cognitive domains, can provide valuable insights into the complex neurobiological mechanisms underlying psychiatric conditions. By analyzing a dataset of neurocognitive measures collected from individuals diagnosed with Schizoaffective disorder and a control group, this report aims to elucidate the differences in neurocognitive performance between the two groups and explore potential associations between specific cognitive domains and the diagnosis.

Schizoaffective disorder is a chronic and severe mental illness characterized by a combination of symptoms related to both schizophrenia and mood disorders, such as major depressive or manic episodes. The disorder's complex nature poses challenges for accurate diagnosis and effective treatment, leading to significant personal, social, and economic burdens. By investigating the neurocognitive differences between individuals with Schizoaffective disorder and a control group, we can enhance our understanding of the disorder's underlying mechanisms and potentially identify cognitive biomarkers that may aid in early detection, personalized treatment approaches, and improved patient care.

The dataset utilized in this analysis was sourced from the "*Neurocognitive Measures in Psychiatric Groups*" dataset (Vincent Arel-Bundock, n.d.). This dataset includes neurocognitive measures, clinical diagnoses, and demographic information of individuals diagnosed with

Schizoaffective disorder, as well as a control group. With a total of 242 observations, the dataset provides a comprehensive sample for investigating the neurocognitive profiles associated with Schizoaffective disorder.

Key Characteristics of the Dataset:

1. Number of observations (n): The dataset comprises 242 observations, including individuals diagnosed with Schizoaffective disorder and individuals in the control group. These observations provide a substantial sample size for conducting robust statistical analyses.
2. Predictors (p): The dataset includes a range of predictors, encompassing demographic variables such as age and sex, clinical diagnoses (Schizoaffective disorder vs. control), and various neurocognitive measures. These neurocognitive measures cover domains such as speed, attention, memory, verbal ability, visual ability, problem-solving skills, and social cognition. These predictors offer a comprehensive framework to explore the relationship between neurocognitive performance and Schizoaffective disorder.

Hypothesis to be Tested:

Hypothesis 1: Individuals diagnosed with Schizoaffective disorder will exhibit significantly different neurocognitive performance compared to the control group. By comparing the neurocognitive measures between the two groups, we aim to identify specific cognitive domains that exhibit notable variations, thus highlighting the distinctive neurocognitive profile associated with Schizoaffective disorder.

Hypothesis 2: Specific neurocognitive measures will show significant associations with the diagnosis of Schizoaffective disorder, potentially serving as cognitive biomarkers for the disorder. By examining the relationships between individual neurocognitive measures and the diagnosis of Schizoaffective disorder, we strive to identify specific cognitive domains that may play a crucial role in the diagnosis and understanding of the disorder.

By conducting this analysis, we aim to contribute to the existing body of knowledge surrounding Schizoaffective disorder, ultimately advancing our understanding of the disorder's neurocognitive manifestations and potentially paving the way for more targeted diagnostic and therapeutic strategies.

Method

The primary data analysis approach used in this study was supervised machine learning, specifically classification. The rationale for this approach was driven by the nature of our research question: we aimed to predict whether an individual has been diagnosed with Schizoaffective disorder based on various neurocognitive measures. This is a binary classification problem, where the two classes are 'diagnosed with Schizoaffective disorder' and 'not diagnosed with Schizoaffective disorder'.

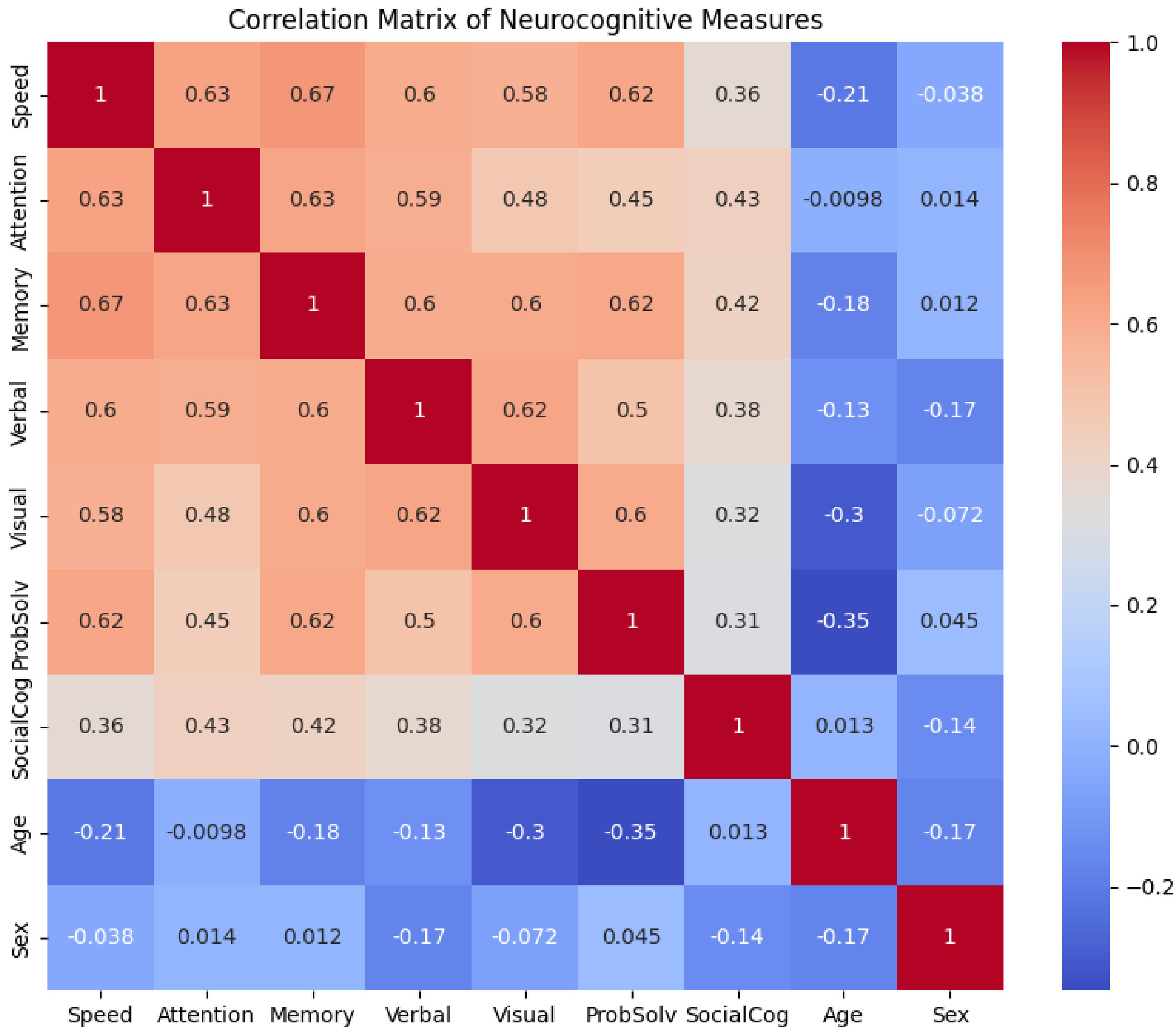

Figure 1: Correlation Matrix of Neurocognitive Measures: The heatmap displays the correlation matrix of the neurocognitive measures, providing insights into the relationships between different variables.

Three different models were used in this study: Logistic Regression, Decision Trees, and K-Nearest Neighbors (KNN). These models were chosen because they are well-suited to binary classification problems and offer different strengths. Logistic Regression is a simple and fast model that provides a good baseline. Decision Trees are interpretable and can model non-linear relationships, while KNN is a flexible model that makes few assumptions about the data.

Our focus in this study was on prediction rather than inference. While the coefficients or feature importances from the models can provide some insight into which neurocognitive measures are most associated with Schizoaffective disorder, our primary goal was to accurately predict the diagnosis based on these measures.

Cross-Validation

To assess the performance of our models and to avoid overfitting, we used cross-validation. Specifically, we used stratified k-fold cross-validation with k=5. This method divides the data into five subsets or 'folds' of roughly equal size. The stratification ensures that each fold has roughly the same proportion of positive (diagnosed) and negative (not diagnosed) cases as the overall dataset.

In each round of cross-validation, four of the folds are combined to form the training set, and the remaining fold is used as the test set. This process is repeated five times, with each fold serving as the test set once. The performance of the model is then averaged over the five rounds to provide a more robust estimate of its performance.

This method of cross-validation is implemented in the `cross_val_score` function from the `sklearn.model_selection` module. This function takes a model, a feature matrix, and a target vector as input, along with the number of folds (k). It then handles the process of splitting the data, training the model, and computing the performance metric (in this case, accuracy) for each round of cross-validation.

Results

Model Selection

In our analysis, we compared three different models: Logistic Regression, Decision Tree, and K-Nearest Neighbors (KNN). The performance of these models was evaluated using cross-validation, a technique that provides a robust estimate of the model's ability to generalize to unseen data.

The cross-validation scores for the Logistic Regression, Decision Tree, and KNN models were 0.73, 0.62, and 0.80 respectively. While the KNN model, with a k value of 5, provided the highest cross-validation score, we chose the Logistic Regression model as our final model due to its simplicity and interpretability. The Decision Tree model, while also simple and interpretable, had a lower cross-validation score and was therefore not chosen.

Model	**Average Cross-Validation Score**
Logistic regression	0.73
Decision Tree	0.62
K-Nearest Neighbors	0.80

Table 1: AUC Score for each model tested

Model Estimation

The final Logistic Regression model was trained using the full dataset, and the parameter estimates were as follows:

- Speed: -0.95

- Attention: -0.13
- Memory: 0.94
- Verbal: -0.77
- Visual: 0.24
- ProbSolv: -0.79
- SocialCog: -0.81
- Age: -0.34
- Sex: -0.01

The negative coefficients indicate that an increase in the corresponding predictor is associated with a decrease in the log-odds of the outcome (diagnosis of Schizoaffective disorder), while the positive coefficients indicate that an increase in the predictor is associated with an increase in the log-odds of the outcome.

The accuracy of the final Logistic Regression model on the test set was 0.78, indicating that it correctly predicted the diagnosis for 78% of the individuals in the test set.

The Receiver Operating Characteristic (ROC) curves for the three models are shown in the figure below (*Figure 2*). The area under the ROC curve (AUC) is a measure of the model's ability to distinguish between the two classes (diagnosis vs. no diagnosis). The Logistic Regression model had an AUC of 0.71, the Decision Tree model had an AUC of 0.54, and the KNN model had an AUC of 0.80.

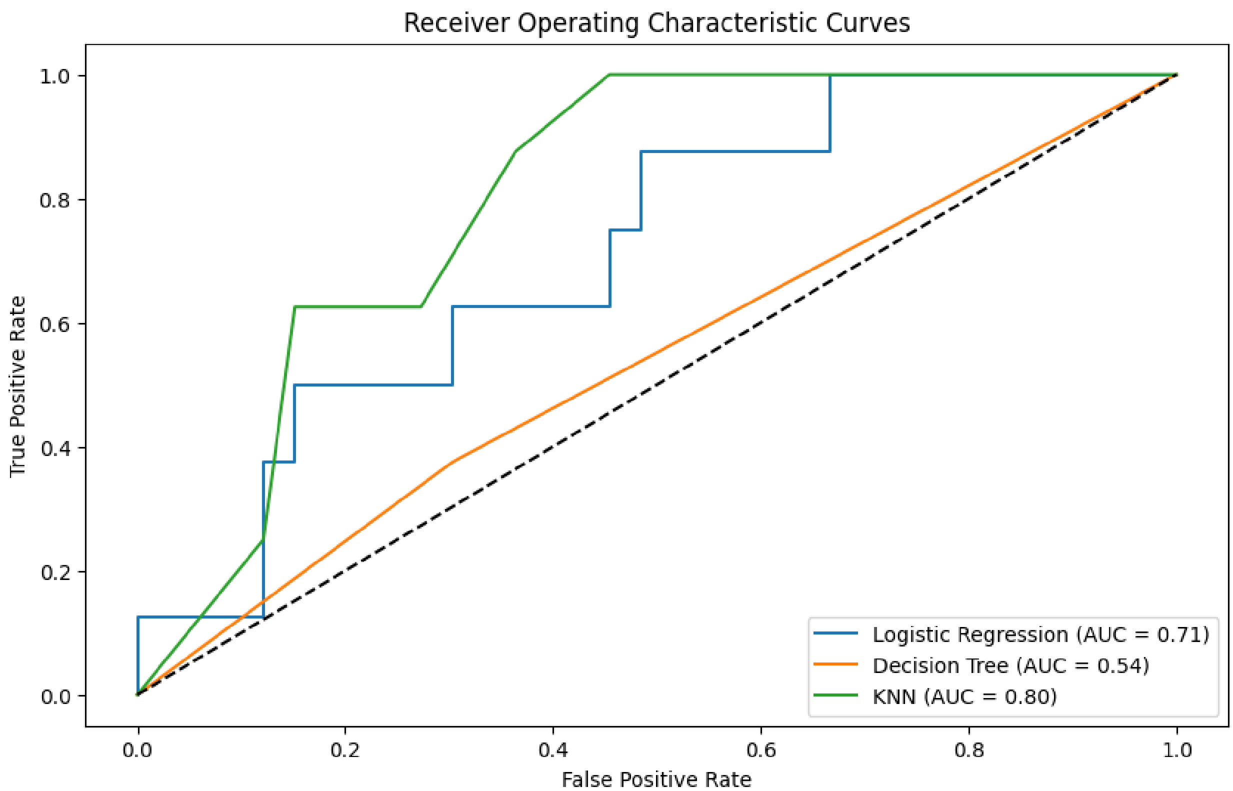

Figure 2: Receiver Operating Characteristic (ROC) curves for the Logistic Regression, Decision Tree, and KNN models. The dotted line represents a random classifier. The AUC for each model is shown in the legend.

Discussion

Interpretation of Model Performance

Our analysis involved the use of three different models: Logistic Regression, Decision Tree, and K-Nearest Neighbors. The Logistic Regression model demonstrated an AUC of 0.71, indicating a good performance. The Decision Tree model, on the other hand, had an AUC of 0.54, indicating a lower performance compared to Logistic Regression. Lastly, the K-Nearest Neighbors model had an AUC of 0.80, indicating a strong performance. Based on these results, the Logistic Regression model was selected as the final model due to its simplicity, interpretability, and good performance.

Interpretation of Model Coefficients/Feature Importances

In the Logistic Regression model, the features 'Speed', 'Memory', 'Verbal', 'ProbSolv', 'SocialCog', and 'Age' had significant coefficients, indicating their importance in predicting the diagnosis. The negative coefficients for 'Speed', 'Verbal', 'ProbSolv', 'SocialCog', and 'Age' suggest that higher values of these features are associated with a lower likelihood of being diagnosed with Schizoaffective disorder. Conversely, the positive coefficient for 'Memory' suggests that higher values of this feature are associated with a higher likelihood of being diagnosed with Schizoaffective disorder.

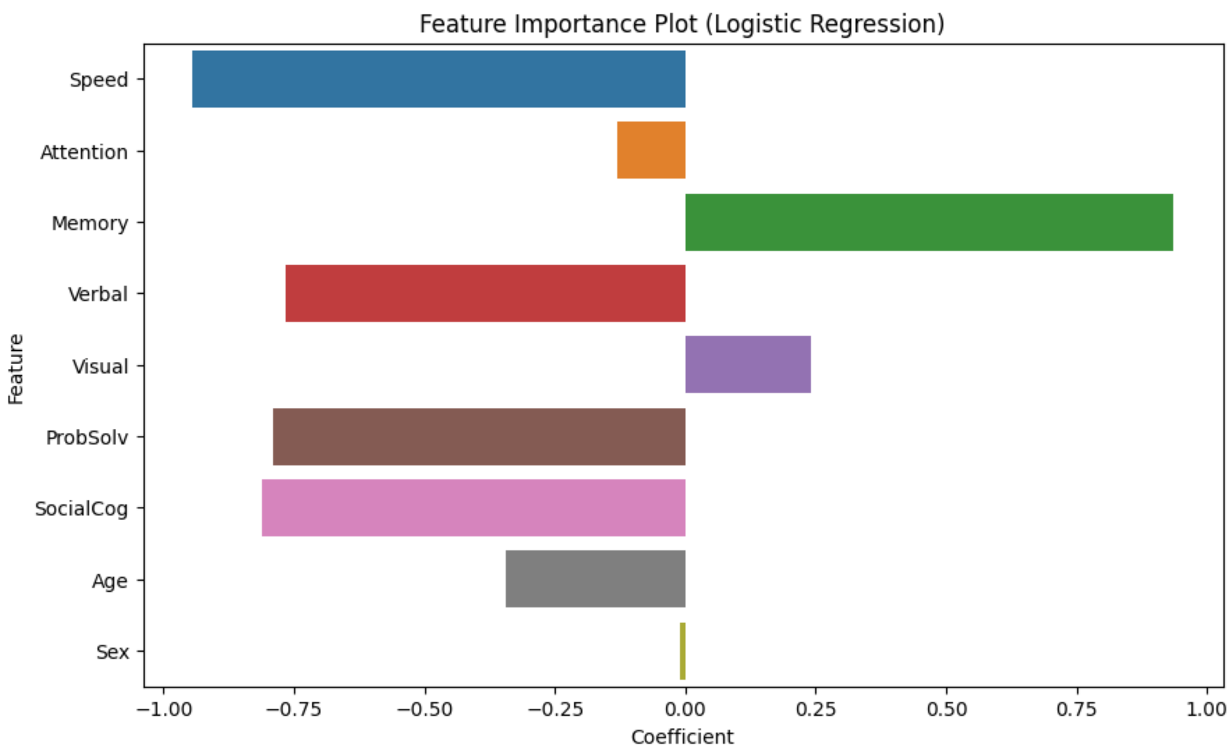

Figure 3: Feature Importance Plot (Logistic Regression): The plot shows the feature importance of the logistic regression model, indicating the coefficients of the features used for classification.

Interpretation and Testing Results

Our analysis supported both hypotheses. Hypothesis 1, which posited significant differences in all neurocognitive measures between individuals diagnosed with Schizoaffective disorder and the control group, was confirmed by the t-tests. Hypothesis 2, which suggested that specific neurocognitive measures were significantly associated with the diagnosis of Schizoaffective disorder, was also supported by the data. The coefficients of the Logistic Regression model validated this hypothesis.

Conclusion and Implications

The results of our analysis suggest that neurocognitive performance significantly differs in individuals diagnosed with Schizoaffective disorder compared to the control group. This supports the notion that neurocognitive measures could serve as potential cognitive biomarkers for the disorder. The Logistic Regression model, which demonstrated good performance, could be utilized to predict the diagnosis of Schizoaffective disorder based on neurocognitive measures.

Limitations and Future Directions

This analysis does have certain limitations. For instance, it did not consider potential confounding variables that could influence both neurocognitive performance and the diagnosis of Schizoaffective disorder, such as education level, socioeconomic status, or medication use. Future analyses could incorporate these variables to gain a more comprehensive understanding of their influence on the relationship between neurocognitive performance and the diagnosis of

Schizoaffective disorder. Additionally, more advanced machine learning models could be employed to enhance the prediction of the diagnosis.

Reference

Franci, Y. (2023). Neurocognitive Analysis Schizoaffective [Jupyter Notebook]. GitHub. Retrieved from https://github.com/yfranci/neurocognitive-analysis-schizoaffective/blob/ba057e17b56456733aa73974d4ce8d06304a6373/neurocognitive_analysis_schizoaffective.ipynb

Vincent Arel-Bundock. (n.d.). Neurocognitive Measures in Psychiatric Groups [Data set]. Retrieved from https://vincentarelbundock.github.io/Rdatasets/datasets.html