Predicting ADHD questionnaire scores from motor behavior using machine learning in Python

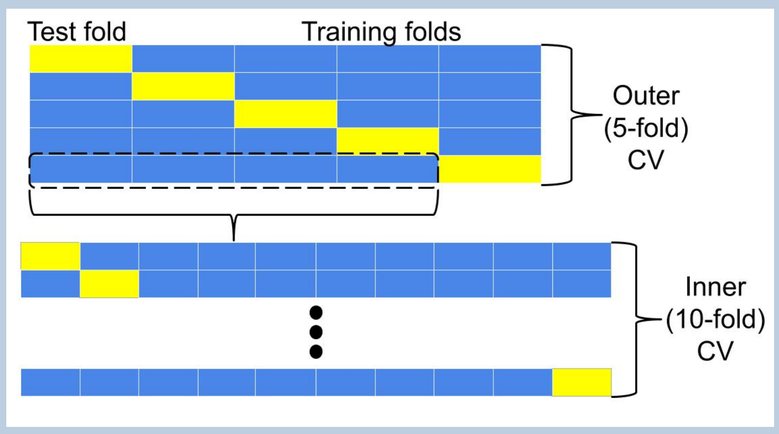
Anton Leontyev, Takashi Yamauchi, Moein Razavi
a.g.leontiev@tamu.edu, takashi-yamauchi@tamu.edu, moeinrazavi@tamu.edu

Introduction

Attention-deficit/Hyperactivity disorder (ADHD) affects the quality of life worldwide. It is commonly diagnosed and studied with specialized questionnaires and behavioral tests. However, behavioral measures (such as the standard stop-signal task, s-SST) often fail to gauge the deficiencies well-highlighted by questionnaires. This lack of sensitivity in behavioral tests is problematic because relying on questionnaires alone can introduce bias (e.g., cultural). Mouse movement measures, being continuous and dynamic, can capture the individual differences in ADHD better than traditional measures. Furthermore, common methods of establishing relationship (e.g., simple correlation) are often insufficient to uncover a complex relationship between behavior in a cognitive task and questionnaire scores.

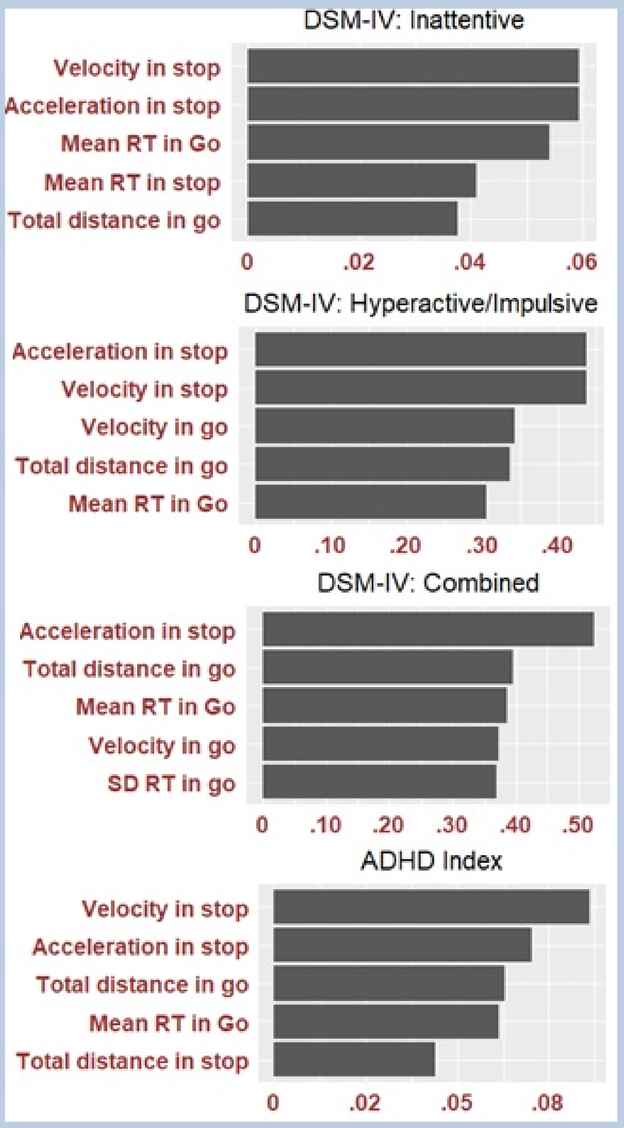
Analysis techniques

We used Python package *scikit-learn* to train **Ridge regression**, **Random Forest regression** and **Support vector Machine** models and predict questionnaires scores. Best hyperparameters were selected using nested k-fold cross-validation approach:



Feature importance

We have also investigated which behavioral and motor measures are most important for predicting questionnaire scores:

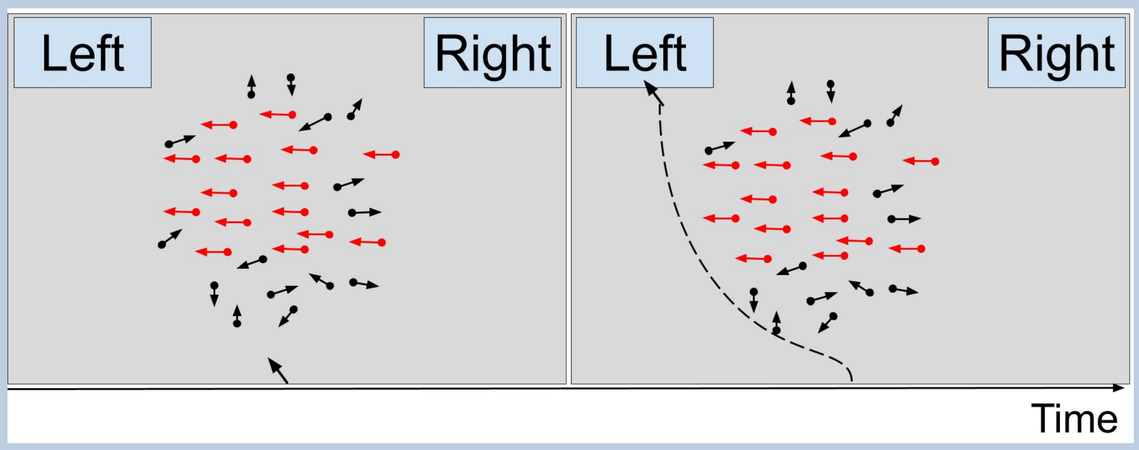


Participants & Design

Total: 100 participants

1. **s-SST**: 50 (27 males and 23 females)
2. **Mouse-SST**: 50 (15 males and 35 females)

To study behavior, we used the Stop-signal task. In this task, participants are presented with two types of trials: "**go**" (requiring response) and "**stop**" (requiring withholding a response). Responses in s-SST are made using arrow keys on the keyboard; responses in Mouse-SST are made by clicking a button on screen. In "go" trials, participants are required to identify the direction of coherently moving dots shown on screen (depicted as red dots on the figure). In "stop" trials, stop-signal (audio tone) was delivered after a variable delay.



Correlations between predicted & observed/permuted data

The table below summarizes the median values of correlation coefficients between questionnaire scores predicted using each of three techniques and actual questionnaire scores. The median values were selected from 20 correlations performed on real and predicted scores (i.e., on data obtained from each of 20 runs):

	DSM-IV: Inattentive	**DSM-IV: Hyperactive/ Impulsive**	**DSM-IV: Combined**	**ADHD Index**
		s-SST		
Ridge	-0.27 (0.07)	-0.11 (0.12)	-0.24 (0.07)	-0.22 (0.06)
Random forest	0.02 (0.06)	0.25 (0.05)	0.19 (0.06)	-0.06 (0.1)
SVM	-0.28(0.11)	0.09 (0.13)	-0.12 (0.11)	-0.21 (0.09)
		Mouse-SST		
Ridge	0.25 (0.06)*	0.20 (0.07)	0.30 (0.04)*	0.34 (0.05)**
Random forest	0.29 (0.07)	0.13 (0.1)	0.22 (0.08)	0.16 (0.06)
SVM	-0.09 (0.13)	-0.25 (0.13)	0.16 (0.1)	-0.1 (0.1)

*p <.05, ** p<.01.

To make sure that our models did not overfit, we predicted ADHD scores using randomly permuted dependent variables using nested CV 20 times. We then examined Spearman's correlation between predicted and permuted data. The table contains median Spearman's correlation values:

	DSM-IV: Inattentive	**DSM-IV: Hyperactive/ Impulsive**	**DSM-IV: Combined**	**ADHD Index**
		s-SST		
Ridge	-0.16 (0.12)	-0.23 (0.16)	-0.16 (0.14)	-0.23 (0.13)
Random forest	-0.01 (0.19)	-0.14 (0.12)	-0.009 (0.19)	0.04 (0.13)
SVM	-0.16 (0.11)	-0.18 (0.18)	-0.19 (0.11)	-0.17(0.1)
		Mouse-SST		
Ridge	-0.02 (0.27)	-0.11 (0.18)	-0.09(0.17)	-0.08(0.23)
Random forest	-0.09 (0.19)	-0.12 (0.15)	-0.08 (0.15)	-0.04 (0.17)
SVM	-0.14(0.13)	-0.1 (0.17)	-0.21 (0.17)	-0.29 (0.14)

Conclusions

1. Ridge regression produces the strongest and most stable associations between behavioral measures and questionnaire scores.
2. Introducing mouse movement measures improves predictive accuracy.