



Utilizing Machine Learning to Identify Autism Spectrum Disorder-Associated Actimetry Traits

J. Dylan Weissenkampen¹, Arpita Ghora¹, Jing Zhang¹, ASPE Recruitment Team², Daniel Rader¹, Ted Brodtkin², Laura Almasy³, Ronnie Sebro⁴, Maja Bucan¹

¹Department of Genetics, University of Pennsylvania; ²Department of Psychiatry, University of Pennsylvania; ³Children's Hospital of Philadelphia; ⁴Mayo Clinic, Jacksonville, FL.

Contact Email: James.VanKampen@penmedicine.upenn.edu

Introduction

Autism Spectrum Disorders (ASD) are highly heritable neurodevelopmental disorders characterized by impaired social interactions and stereotyped behavior. However, ASD typically presents with a large phenotypic heterogeneity. Even so, between 44-80% of ASD individuals report sleep disturbances. These sleep disturbances may share an underlying neurological basis with ASD, while providing an objective and measurable trait for analysis. Here we hypothesize that using machine learning variable selection techniques, we will identify sleep and circadian traits associated with ASD status.

Methods

- ASD adults and their families (N=360) were recruited and sent a wearable actimetry measure to wear for two weeks
- Actimetry data was analyzed using the GGIR package in R. Data was split into even and odd days for semi-independent analysis
- Medians and standard deviations (SD) were calculated for each variable
- Performed 5-fold cross-validation within the odd day sets, and validated in even day sets with least absolute shrinkage and selection operator (LASSO) and Random Forest using the R packages *cv.glmnet* and *randomForest*, respectively

Figure 1

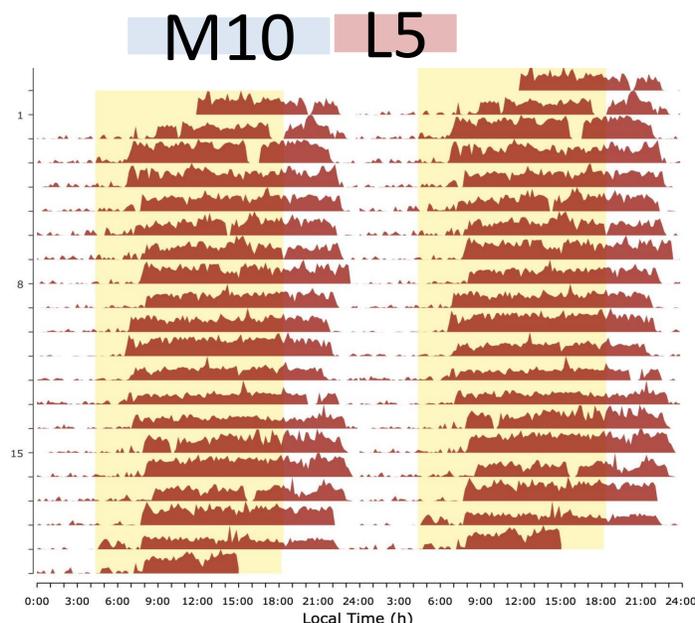


Figure 1. Actimetry measures in one participant for two weeks. Most active 10 hour period (M10) and least active five hour period (L5) are shown. Sunlight period is shown in yellow.

Results

- The LASSO and Random Forest models work well in the test and validation sets
- The minimum model for LASSO selected 13 variables with the more conservative model within 1 standard error of the minimum model containing 9 variables
- LASSO performed better than Random Forest in even and odd day sets (Figure 3B)
 - LASSO showed some loss in prediction on the semi-independent even day data, while Random Forest was consistent between sets

Figure 2

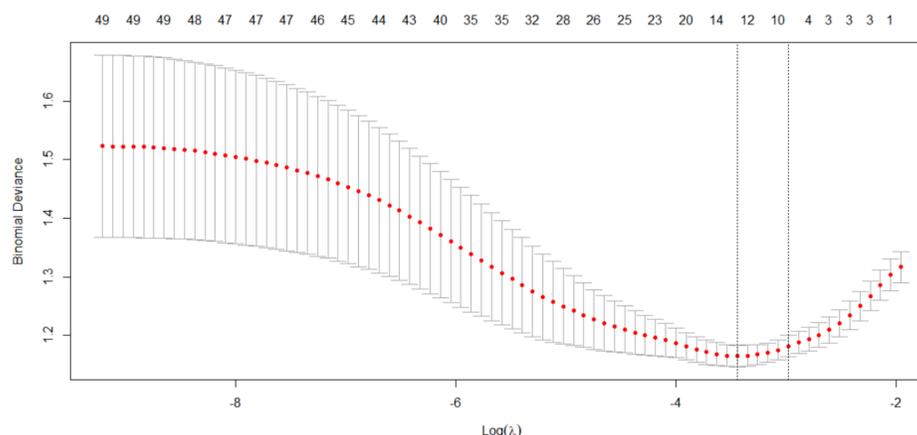


Figure 2. LASSO model selection. Binomial deviance values were used to judge the model performance for LASSO. The minimum model contained 13 variables, while the more conservative 1 standard error model contained 9.

Figure 3

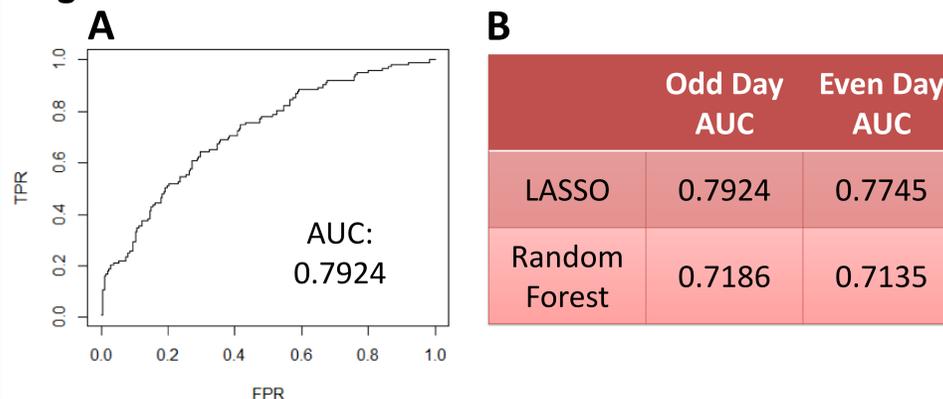


Figure 3. A) Receiver-Operator-Curve (ROC) for the LASSO minimum model in the test set for the odd day data. The true positive rate (y-axis) and the false positive rate (x-axis) are shown, with the area under the curve (AUC) of 0.7924. B) Table of AUC values in the odd and even day datasets using LASSO and Random Forest

Table 1

Direction	Variable
-	(Intercept)
+	Number of nights (SD)
-	Duration of light activity (Median)
+	Duration of waking hours (SD)
+	Sleep efficiency (SD)
+	Average activity level during light activity (Median)
+	Average activity level during moderate activity (SD)
+	Average activity level during vigorous activity (SD)
+	Level of activity for L5 (SD)
+	Level of activity for L10 (SD)
-	Level of activity for M10 (SD)
+	Relative amplitude (SD)
-	Age
+	Sex

Table 1. Selected variables associated with ASD status. Here the variables selected in the LASSO minimum model and their directions of effect are listed. Whether the variable selected corresponded to the median or the standard deviation of the variable is listed in () next to the variable. Age and sex were also predictors in our data.

Discussion & Future Directions

- Here we demonstrate that several sleep and activity measures associated with ASD status with decent predictive accuracy.
 - This predictive accuracy only slightly lessened when removing age and sex from the analysis.
- Variables' standard deviations suggest that there may be different variability within these traits for ASD individuals compared to their control family members.
- By investigating the neurological basis of the traits we identified, we may elucidate common pathways with ASD.

Funding & Acknowledgements

This work was supported by the Autism Spectrum Program of Excellence (Research Gift to the University of Pennsylvania; Dr. Rader, PI) and the NIH grant (NIMH R21MH093415 to MB). Work is also supported by the T32ES019851 grant.