Background
ADHD is a time-intensive diagnostic process due to its complex symptomatology, and patients have to wait up to one year for an ADHD evaluation in the US. This delay in diagnosis and treatment time leads to poor cognitive, behavioral, and social outcomes in children. Because early intervention improves clinical outcomes, there is a critical need to diagnose patients sooner by accelerating the diagnostic process.

The ADHD diagnostic questionnaire contains family history, demographic, and developmental data, as well as five standardized tests. Of these tests, the DuPaul ADHD Rating Scale is the primary metric used in the pre-screening process. How can the efficiency of ADHD diagnostic testing be increased using the diagnostic questionnaire?

Methods
1. 1464 ADHD Patients
   Kennedy Krieger Institute, Baltimore MD

   Questionnaire Data Collection and Preprocessing
   Data Balancing
   Feature Selection
   Correlation Analysis and Principal Component Analysis
   Random Forest

   
   $P_{\text{AUC}} = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k}}$
   Generalized Linear Model Generation

Results

Figure 1—GLM model p-values and coefficients. Features with the greatest significance are highlighted with red.


Figure 3—ROC: 5-fold cross-validation on the final model

Conclusion and Future Direction
We found that the DuPaul scores of inattention and hyperactivity comprised an adequate baseline model. However, upon adding different categories of features to the baseline model, we observed that the overall performance of the machine learning model to classify ADHD improved significantly.

We plan to deploy this risk assessment tool at the Kennedy Krieger Institute and analyze its efficacy by performing a sensitivity analysis over time, and potentially prioritize the patient waitlist based on risk.

References