2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

# EVALUATING THE NEUROIMAGING-GENETIC PREDICTION OF SYMPTOM CHANGES IN INDIVIDUALS WITH ADHD

PRANAV SURESH[1,2], BHASKAR RAY[1,2] KUAIKUAI DUAN[1,7], JIAYU CHEN[1], GIDO SCHOENMACKER[3], BARBARA FRANKE[3,4], JAN K. BUITELAAR[5], EMMA SPROOTEN[5], ALEJANDRO ARIAS-VASQUEZ[3,4], JESSICA A. TURNER[1,6], JINGYU LIU[1,2]

*Abstract*— **Attention-deficit/hyperactivity disorder (ADHD) is a neurodevelopmental disorder that could persist into adulthood with known abnormalities in brain structure. Genetics also play an important role in the etiology of the disorder and could affect the disorder trajectory. In this study, we investigated the prediction power of brain image and genomic features for symptom change in 77 individuals with ADHD as part of NeuroIMAGE cohort. Gray matter components and working memory assessments at baseline, as well as gene scores of interest, were used to predict the changes in the two symptom domains: inattentive and hyperactive/impulsive, an average of 4 years. A linear regression model coupled with various feature selection approaches, including leave-one-out-cross-validation (LOOCV), stability selection with resampling, and permutation tests, was implemented to mitigate the overtraining potential caused by small sample sizes. Results showed that traditional LOOCV overestimated the prediction power. We proposed a novel stability selection with the threshold set by permutation tests, which provided more objective assessment. Using our proposed procedure, we identified a statistical promising prediction model for inattention symptom change; the consistent correlation between predicted values and measured values during model training, validating and hold out testing (r=0.64, 0.53, 0.46, respectively), but the p value is not significant in the holdout test. The selected features include age, gray matter in the insula, genes OSBPL1A, CTNNB1, PRPSAP2, ACADM, and polygenic risk score of education attainment, which have been previously reported to be associated with ADHD. We speculate that significant associations may be observed with a large sample size.**

## I. INTRODUCTION

Attention-deficit/hyperactivity disorder (ADHD) is one of the most common neurodevelopmental disorders of childhood. It is usually first diagnosed in childhood and could persist into adulthood [1]. Children with ADHD may have trouble paying attention, controlling impulsive behaviors (may act without thinking about what the result will be), or be overly active [2], and may also have various types of cognitive impairments [2]. A meta-analysis of follow-up studies has shown that in about 15% children with ADHD the disorder persists into adulthood, and the persistence percentage increases to 65% if partially remitted patients are considered [2]. Although the classification of ADHD and the persistence of ADHD are binary, highly dependent on the threshold used by the clinicians following the Diagnostic and Statistical Manual of Mental Disorders (DSM) or International Classification of Diseases(ICD), the symptoms themselves are continuously distributed among individuals, as are the symptom changes. Predicting the trajectories of ADHD symptoms along the disorder progression can have huge impact in the development of effective prevention and treatment; it can classify individuals whose symptoms aggravate in the future, and thus early intervention can be provided in time.

Many longitudinal studies have been devoted to uncovering the the factors influencing the course of ADHD symptoms and to improve the prediction of symptom trajectories [4]. As reviewed by Caye, et al. a meta-analysis summarized the consistent predictors of symptom trajectory including characteristics of the clinical syndrome, ADHD symptom severity, treatment, comorbidities, and parental mental health problems, etc. [5]. However, little is known about the relation between brain structure and function factors and the symptom trajectory, in spite of abundant evidence support association of brain anomalies with ADHD and it's symptoms. Structural MRI studies have suggested that gray- and/or white-matter structural underdevelopment in frontal lobe, thalamus, and striatum significantly contribute to the emergence of ADHD during childhood [6-9]. Furthermore, persistence of ADHD symptoms is linked to reduced regional cortical gray matter thickness in frontal and parietal cortices [10,11]. Our team also found frontal and cerebellum gray matter variations consistently associated with working memory deficit and inattention symptoms in both adolescents and adults with ADHD[1-3]. Functionally, lower connection efficiency in right inferior frontal gyrus and left-side frontal-parietal functional interactions were observed in both adult remitters and persisters, and unique lower connection efficiency in right middle frontal gyrus and hyper-interactions

*This study was supported by National Institutes of Health, grant 1R01MH106655 to JL. We would like to thank the Dutch NeuroIMAGE team to share the data.

[1]Tri-Institutional Center for Translational Research in Neuro Imaging and Data Science (TreNDS). [2]Department of Computer Science, Georgia State University, Atlanta, USA. [3]Department of Human Genetics, Donders Institute for Brain, Cognition and Behavior, Radboud University Medical Center, Nijmegen, The Netherlands. [4]5Department of Psychiatry, Donders Institute for Brain, Cognition and Behavior, Radboud University Medical Center, Nijmegen, The Netherlands. [5]Department of Cognitive Neuroscience, Donders Institute for Brain, Cognition and Behavior, Radboud University Medical Center, Nijmegen. [6]Department of Psychology, Georgia State University, Atlanta, USA. [7] Department of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, USA.

between bilateral middle frontal gyrus in persisters [12,13]. How brain structural and functional alterations relate to symptom trajectory is yet to be studied[14].

The heritability of ADHD is estimated between 30%-80% in twin and family studies [15]. Strongly increased risks for ADHD (57%) among the offspring of adults with ADHD have been reported [16]. Longitudinal studies that investigated the genetic contributions to the long-term ADHD suggested while persistence of ADHD symptoms is predominantly due to the same genetic influences as its onset, changes of symptoms are to a large extent due to new genetic effects beginning in early adolescence, as well as environmental factors [17]. More recently, large sample genome-wide association studies (GWAS) have reported several genetic risk loci for ADHD [18]. Polygenic risk score (PRS) based on GWAS risk profile, estimating an individual's genetic liability for a particular disorder or trait, were able to explain significant variance (5.5%) in ADHD [18]. Genetic risk for ADHD is highly correlated to the risk to other disorders or traits, and one of the highest correlations are from genetics for college completion [18]. To what extent the genetics could influence the course of ADHD symptom is largely unstudied.

Although various genetic, cognitive, and neural factors have been associated with ADHD, very little is known about their combined ability to predict the symptom trajectory. In this study we leveraged longitudinal data collected from individuals with ADHD and investigated the prediction power of brain structure and genomic features, as well as cognition assessments, for future symptom changes. Based on the previous findings we hypothesize that the interplay of the examined factors might better explain the trajectory of symptoms in both domains (inattention and hyperactivity) than any individual feature set. Assessing the neural, genetic, and cognitive factors of subjects can enrich our understanding of the symptom trajectory and thereby aid in creating personalized prevention and treatment.

## II. Materials and Methods

### A. Participants

We employed a subset of data from NeuroIMAGE project [19]. The NeuroIMAGE is a multi-site prospective cohort study designed to investigate the course of ADHD, its genetic and environmental determinants, its cognitive and neurobiological underpinnings, and its presentations in adolescence and adulthood. The study was approved by regional ethics committee and the medical ethical committee of the VU University Medical Center. All participants provided a written consent form. From all participants, we selected 77 participants, including 43 male and 34 female, who 1) met the ADHD diagnostic criteria based on DSM-IV at one time point ( here named baseline), 2) provided good quality neuroimaging and genetic data at baseline, and 3) had cognitive and symptom assessments at both baseline and follow up time points. The average age of the participants was 16.30 and 19.97 years old for the baseline and follow up timepoint respectively.

Symptoms were measured in both domains: inattention and hyperactivity/impulsivity. Symptom change between the two time points reflects the progression of disorder and is the variable we want to predict. WAIS Digit Span test (maximum forward and maximum backward scores) was utilized to gauge working memory capacity, as it showed persistent impairment in adolescents and adults with ADHD [1,20]. Base-line working memory scores are the features tested for prediction.

### B. Neuroimaging data and features

T1-weighted MRI images after quality control were normalized, modulated, segmented, and smoothed with 6mm Gaussian kernel using SPM12, and the resultant gray matter maps were further regressed out age, sex, and site effects. Independent component analysis was then applied to the whole brain voxels with gray matter density $>0.2$, resulting in 24 components. The details of preprocess can be seen in [1].

Each component is a brain network and the relative gray matter density of this network is measured by the component loadings. Also, we added 5 more components we identified in our previous studies that are associated with adult ADHD symptom and cognitive impairment [1]. These 29 gray matter components' loadings were the input features for prediction model.

### C. Genomic data and features

From genomic SNP data after imputation, we computed two sets of genetic scores: Quantative Genetic Score (QGS) [20], and Polygenic Risk Score (PRS) [21]. QGS assigns a numeric value $0 \leq \text{QGS} \leq 1$ to any preselected genetic region, based on the average difference between an individual's genetic information (in the form of genotypes) and that of a reference population (i.e. the same reference used to impute genetic information to said individual). QGS can be interpreted as a measure of individual's genetic "distance" to the reference population: a lower score indicates higher similarity to the reference population, whereas a high score indicates a lower similarity [20]. We selected 29 genes whose QGS scores were stably associated with the five gray matter components underlying adult ADHD symptoms and working memory impairments [1]. See details of QGS method in [20]. PRS is the weighted summary score of individual SNPs' risk to a specific disease or trait based on genome-wide association study results. We computed PRS for education attainment [22], intelligence [23], ADHD [18], and major disorder [24] using PRsice2 [21].

### D. Data Analysis

In this study we focused on using baseline brain images, working memory tests, and genetic scores to predict symptom changes in both inattention and hyperactivity/impulsivity domains. Specifically, tested predictors include age, sex, baseline 29 gray matter components, QGS of 29 genes, four PRS, baseline working memory scores, and the interval between two timepoints. To build a reliable prediction model,

we implemented stepwise linear regression with forward feature selection, LOOCV, stability selection with resampling (two resampling strategies: subsamples and bootstrapping with replacement), permutation test. We combined stepwise linear regression feature selection with LOOCV (FS-LOOCV), which is useful when a small sample size cannot afford withholding data from the training set [25]. As in Figure 1, we performed FS-LOOCV in full training samples. To further reduce overfitting, we performed stability selection with resampling, and the permutation tests with different feature sets. Finally, we performed validation test on independent holdout samples. The details of each step are explained next.

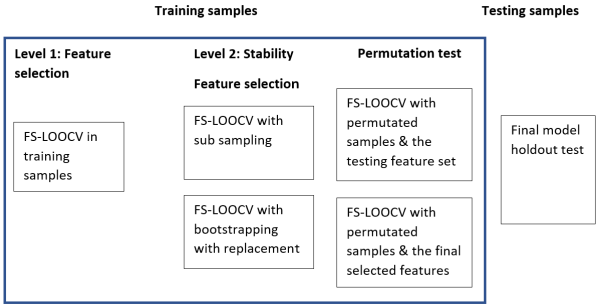



Fig. 1: Analyses performed on the data

FS-LOOCV: The performance of any machine learning model is sensitive to the set of features that are used in the training of the model. Determining the best set of features for the final model is called feature selection. Here we used forward stepwise regression, a procedure that selects the best set of features iteratively. Specially for a prediction model, starting with no predictors in the working set, at each iteration the algorithm tests model accuracies for individual predictors when added to the working set, and selects only one predictor with the best model fit to the working set. We used LOOCV $R^2$ to measure model fit. All training samples except one were used to build a linear regression model, which was then used to predict the value of the one sample not used. Repeat this procedure for N times (N = size of training samples) to generate predicted values for all training samples. LOOCV $R^2$ is then computed as

$$LOOCV R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$$

Where, $SS_{res} = \sum_i (y_i - f_i)^2$; $f_i$ = predicted value and $y_i$ is true value
$SS_{tot} = \sum_i (y_i - \bar{y})^2$; $y_i$ = true value and $\bar{y}$ = mean of true values

The pseudocode of the FS-LOOCV stepwise regression is as follows:

1) Starting with an empty working set, **S**, and all available predictors, **Predictors Set**
2) Iterate over available predictors in the **Predictors Set**
   a) Add each predictor to the working set **S**
   b) Test the model estimate using LOOCV when the predictor is added
   c) Remove the added predictor from **S**
3) Add the best predictor to the working set **S**
4) Remove the best predictor from **Predictors Set**
5) Repeat Step 2, 3 and 4.

To further reduce possible overfitting, we performed stability selection with resampling for the feature set after FS-LOOCV. Stability selection identifies the most stable predictors by assuming that the same algorithm should yield similar results on similar datasets if the results are "stable" [26]. To generate similar data, we implemented subsampling and bootstrapping strategies. For sub-sampling, sample sizes of 50, 55, 60, and 65 were selected. 64 sub-samples, including 16 random sub-samples of each sample size, are generated and FS-LOOCV is applied. The features selected from each sub-sample are aggregated. Frequency of each feature being selected indicates its stability. Whereas for bootstrapping, in each iteration, an instance is drawn from the same original dataset such that certain instance may appear more than once in a bootstrap sample [27]. We applied FS-LOOCV to each of the bootstrap samples, aggregate the features selected, and compute the feature frequency. For both resampling strategies, a preset threshold is used to select stable features; i.e., features with frequency higher than the threshold will be selected as stable features.

### *E. Permutation Tests*

To select the best threshold for stability frequency and empirical significance, we performed permutation test. We performed two permutation tests using, 1) all the 65 features in the dataset to set up the best threshold on stability selection and to test the power of samples to select the features, and 2) the selected features from the selected threshold to test the empirical significance of model prediction on symptom change.

For the permutation test using the full 65 features, we generated 100 datasets by randomly permuting the sample symptom scores of the full training data. For each dataset, the analysis methods used on the original data were applied, i.e. FS-LOOCV stepwise regression followed by stability selection. On the aggregated features derived from multiset resampling, different thresholds (50% to 100% in steps of 5) were applied to select stable features, and then using selected features the model was trained and tested using LOOCV. The model performance was measured by $R^2$. For each threshold, we calculated p-value as the probability of obtaining the $R^2$ on permuted samples equal to or greater than the observed $R^2$ in the original data at the same threshold. For example, a p-value of 0.23 at threshold 90% indicates a 23% chance of getting higher LOOCV explained

variance ($R^2$) in the randomly permuted samples than the observed explained variance in original data with the features selected at the same threshold of 90% ($R^2$ calculated using the same analysis method). The threshold that gave the best p-value (smallest p value) was used to set the threshold for selecting final features.

Another permutation test was performed in the same manner but only using the selected features of all training samples. The null hypothesis is the selected features cannot predict dependent variable. Analysis was performed on data with the selected features and permuting the symptom scores [28] to test the empirical significance.

### F. Holdout Test

Finally, we used the 6 hold out samples to verify the final prediction model, which is estimated using the linear regression on all 71 samples with the final selected features.

We calculated the model estimate on the holdout dataset using the linear model trained on all 71 samples with the final selected features. We used 6 datapoints that were not included in the original dataset.

## III. Results

First, using feature selection we reduced the feature space from 65 to 25 features in inattention domain (14 in the hyperactivity domain). Further, using stability selection with subsampling, to counter overfitting, we further reduced the feature space to 7 features. In the bootstrap method of stability selection, resulting frequencies of each aggregated features is very low. Standard data analysis was performed (training, testing, and holdout testing) and selected features explained significant variance in training and testing in both domains. However, the holdout testing in both the domains was not significant.

### A. Inattention Domain

Using the FS-LOOCV stepwise regression we selected 25 features that gave maximum LOOCV $R^2$ . Figure 2(a) plots the $R^2$ of the linear model trained with all the samples and LOOCV $R^2$ at each iteration of stepwise regression. After the $25^{th}$ iteration with 25 features the LOOCV R-square reaches the highest values and with each additional iteration $R^2$ value starts diminishing.

Using stability selection with 64 sub-samples, we computed the frequency of the 25 features from the stepwise regression. Figure 2(b) shows the frequency distribution of the features. There are 5 features which were included by all the stability models i.e. these features are strongly associated with symptom change. 19 features were selected at least by 50% of the subsample models. In contrast, Figure 2(c) shows the frequency distribution of features computed using bootstrapping (subsampling with replacement). The maximum frequency attained by the bootstrap method is 33%. Most of the features are selected only by 20% of the bootstrap models.

Figure 2(d) shows the frequency thresholds and their respective p-values calculated using permutation tests (tests performed using the full feature set). The p-value is highest at 50% threshold and starts decreasing while the threshold increases. The most significant result among all feature sets is p-value of 0.23 achieved by the features thresholding at frequency of 90%. Of the 25 features selected in stepwise regression with forward selection, 7 features had frequency greater than 90%.

The 7 features include Age, gene OSBPL1A, CTNNB1, GM in Insula region (Fig 3) which are negatively correlated to the symptom change, while genes PRPSAP2, ACADM, and PRS of education attainment were positively correlated to the symptom change. The permutation test performed using these 7 selected features has a p-value $< 0.05$. Table 1 summarizes the model training, testing, and holdout results in the inattention domain. The training model fit using all points with these 7 selected features has $R^2$ = 0.418 whereas the LOOCV testing $R^2$ = 0.26. Also, the correlation between the predicted values and true value in both the phases, training and testing, is 0.64 and 0.53 respectively. On the other hand, in the holdout test, the r-square is negative (-0.018) while the correlation is positive (0.46). The empirical p-value obtained by the permutation tests with these selected 7 features is less than 0.05.

TABLE I: Training, LOOCV testing, and holdout testing $R^2$, correlation, and MSE between true and predicted value

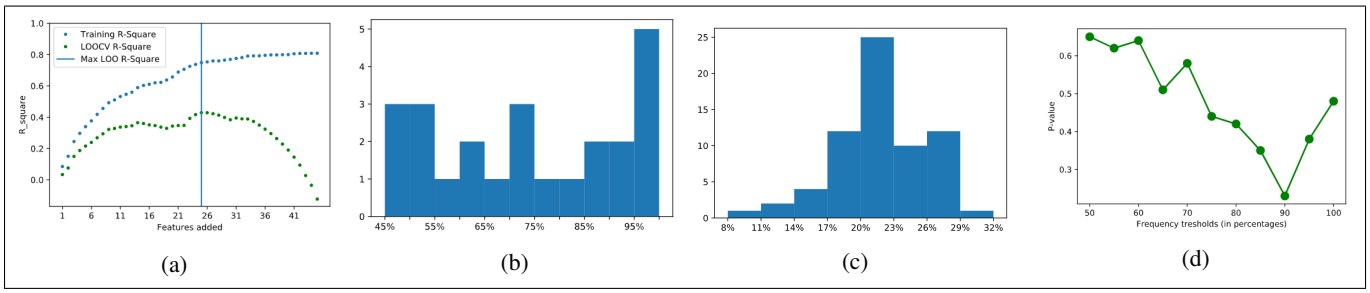
| Phase | $R^2$ | Correlation | MSE |
|---|---|---|---|
| Training | 0.41 | 0.64 | 3.31 |
| LOOCV Testing | 0.26 | 0.53 | 4.20 |
| Holdout Testing | -0.018 | 0.46 | 10.04 |

### B. Hyperactivity Domain

Similar to inattention domain, in the hyperactivity domain, using FS-LOOCV stepwise regression we selected 14 features that gave LOO $R^2$. The plot shown in Figure 4(a) compares the training $R^2$ vs the LOOCV testing $R^2$. The LOOCV $R^2$ is maximum after 15 iterations and starts decreasing when further features are added. Moreover, the $R^2$ becomes negative on further iterations.

Using stability selection with 64 sub-samples, we computed the frequency of the 14 features from the stepwise regression. Figure 4(b) shows the frequency distribution of the features. There are 5 features which were included by all the stability models i.e. their frequency is greater than 95%. 11 features were selected at least by 50% of the subsample models. On the other hand, Figure 4(c) shows the frequency distribution of features computed using bootstrapping (subsampling with replacement). The maximum frequency attained by the bootstrap method is 32%. Most of the features are selected with equal frequency.

Figure 4(d) shows the frequency thresholds and their respective p-values calculated using permutation tests (tests performed using the full feature set). The p-value is highest at 50% threshold and starts decreasing while the threshold




(a) (b) (c) (d)

Fig. 2: Results from prediction of inattention symptom change. a) comparison of LOCCV-$R^2$ and training $R^2$. b) Frequency distribution of features selected using subsampling method of stability selection. C) Frequency distribution of features selected using the bootstrapping method of stability selection. d) The thresholds for stable feature selection and their respective permutation test p-value

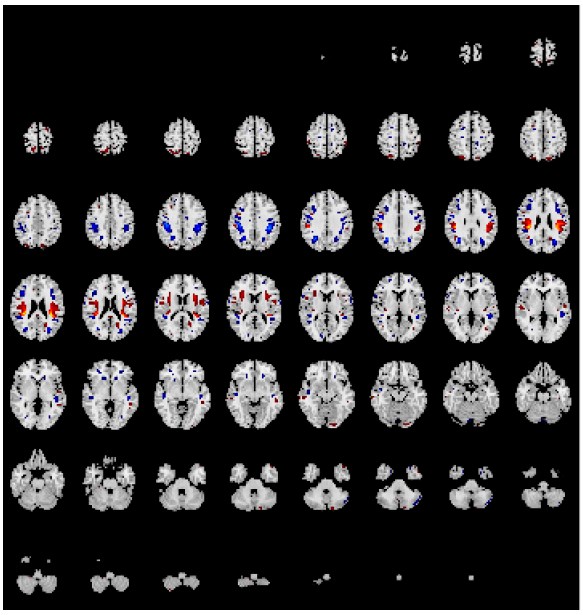


Fig. 3: Grey Matter(GM) component from insula region

increases. The most significant result among all feature sets is p-value of 0.33 achieved by the features that are threshold at frequency of 80%. Of the 14 features selected in stepwise regression with forward selection, 6 features had frequency greater than 80%.

At threshold of 80%, we selected 6 features that include genes CTNNB1, DYNC1I1, GM component (negatively correlated), genes PRPSAP2, LING02, and AC104662_2 (positively corelated) were selected. The permutation test performed using these 6 features has a p-value $< 0.05$. Model fit with all data points with these features has an $R^2$ of 0.31 and in the LOOCV testing the $R^2$ is 0.17. But the permutation p-value is 0.38. The holdout test result showed correlation r= 0.017 whereas $R^2$ = -2.82.

## IV. Discussion

In this study, we investigated the ability to predict the trajectory of the symptoms in both domains——inattention, and hyperactivity——using features from sMRI images and genomics of ADHD cohort of 77 subjects. Using stepwise regression with forward selection, we selected features that explain maximum variance in the symptom change. But testing results indicated the model to be overfitting. Using stability selection coupled with permutation tests, we further reduced the feature space. Permutation tests suggested we do not have the power for selecting features. But given the final selected features, for inattention, the prediction performance is very promising within our samples. In inattention domain, we identified age, genes OSBPL1A, CTNNB1, PRPSAP2, ACADM, and one GM component in the insula region associated with symptom change.

TABLE II: Training, LOOCV testing, and holdout testing $R^2$, correlation, and mean squared error between true and predicted value in hyperactivity domain

| Phase | $R^2$ | Correlation | MSE |
|---|---|---|---|
| Training | 0.31 | 0.56 | 4.48 |
| LOOCV Testing | 0.17 | 0.43 | 5.43 |
| Holdout Testing | -2.82 | 0.025 | 3.40 |

In both symptom domains, using the stepwise regression with forward selection we selected the set of features that has maximum testing LOOCV $R^2$. Even though this is a standard approach for feature selection and model training, however, as shown in the Figures 2(a) and 3(a), the improvement of testing LOOCV $R^2$ from a lower feature set (16 for inattention and 10 for hyperactivity) to the top feature set is small compared to the gap between training $R^2$ and testing LOOCV $R^2$. These results indicate that 1) set of features that gave maximum LOOCV $R^2$ are overfitting, and 2) a smaller set of features can explain similar variance.

Thus, we applied stability selection in combination with permutation tests to further reduce feature set to alleviate overfitting. Two sampling strategies were used for stability selection. In most studies the threshold value for stable feature selection has been a tuning parameter i.e. no objective way to determine the value. In this study, we used permutation tests to determine the threshold such that the variance explained by features selected at the threshold is significant

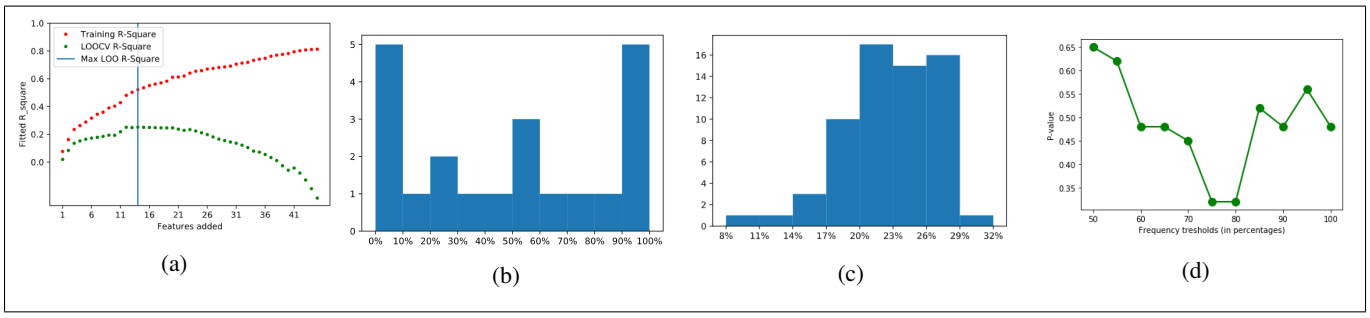



Fig. 4: Results from prediction of hyperactivity symptom change. a) comparison of LOCCV-$R^2$ and training $R^2$. b) Frequency distribution of features selected using subsampling method of stability selection. C) Frequency distribution of features selected using the bootstrapping method of stability selection. d) The thresholds for stable feature selection and their respective permutation test p-value

(or the most significant) compared to the null distribution simulated by permuted samples using features selected at same threshold. The difference between the training and testing $R^2$ results summarized in the Table 1 and Table 2 show that the smaller set of features has reduced overfitting.

However, the best p-value obtained using the permutation tests to identify the threshold of frequency, are 0.23 and 0.36 in the inattention and hyperactivity domains respectively. These results indicate that the selection of features using the threshold are non-significant; we do not have the power to select the features. On the contrary, the permutation tests performed using the selected features has empirical p-value $< 0.05$ which indicate that these features are significantly associated with the symptom changes within our training samples, but we do not know whether this is true for other independent samples.

On the other hand, the bootstrap method for stability selection was unable to select stable features due to heterogeneity in the data sample. We observed each data sampling produces a model with similar performance ($R^2$) but different feature sets. We speculate that within our small samples there are large heterogenous properties, so that each bootstrapping sampling has different property distribution, leading to a different feature set and model. Compared to subsampling strategy, common samples between any two samplings in bootstrapping is less, leading to low frequencies as shown in Figure 2(c) and 4(c).

The LOOCV testing results in both the domains were promising. However, LOOCV is known to have inferior performance for model estimation, risk for overfitting [29]. The additional validation test results are not significant and suggest the features selected lack generalizable power to significantly explain the symptom changes. Though the validation $R^2$ is small in the inattention domain, the correlation between the true value and the predicted value is high (r=0.46) and in the same level as training and testing results. The consistent effect size provides very promising indication of strong association among the features selected and the symptom change. This non-significance might be the result of small sample size of validation sets. We speculate that with large sample size these features may be proven significantly associated with symptom changes. On the other hand, the results in the hyperactivity domain indicate no association of features with the symptom change.

The features identified in the inattention domain include genes OSBPL1A and PRPSAP2 that have been previously reported to be associated with the ADHD [30]. The gene CTNNB1 was discovered recently to be responsible for developmental delay/intellectual disability [31]. CTNNB1 is important in the development and maturation of the brain and loss of its function causes learning and memory problems [31]. Aging effects on ADHD has been researched extensively [32,33]. Furthermore, the GM components reported in our results —component in the insula region, are also previously reported to be associated with ADHD problem [34]. However, the effect of postcentral gyrus on ADHD is yet to be studied.

The small sample size of 77 hinders our statistical analyses be generalizable to other data. This is demonstrated by the inability of obtaining significant results using the validation testing. To sum up, in this study we aimed to predict the trajectory of the ADHD symptoms in both the domains – inattention, and hyperactivity using genetics and neuroimaging data. Using data of 77 subjects from two time points (6 subjects used for holdout testing) we performed variable selection using stepwise regression using forward selection, leave one out cross validation. The selected variables were still overfitting, to further reduce overfitting we performed stability selection in combination with permutation tests and selected top features. In both the domains, the features selected were unable to explain the symptom change in the test samples. However, in the inattention domain, the features selected do have strong association with the symptom change and can be studied further to predict the trajectory of ADHD symptoms.

## REFERENCES

[1] Duan K, Chen J, Calhoun VD, Lin D, Jiang W, Franke B, et al. Neural correlates of cognitive function and symptoms in attention-deficit/hyperactivity disorder in adults. Neuroimage Clin 2018; 19: 374-383.

[2] Jiang W, Duan K, Rootes-Murdy K, Hoekstra PJ, Hartman CA, Oosterlaan J, et al. Structural brain alterations and their association with cognitive function and symptoms in Attention-deficit/Hyperactivity Disorder families. Neuroimage Clin 2020; 27: 102273.

[3] Liu J, Duan K, Jiang W, Rootes-Murdy K, Schoenmacker G, Buitelaar JK, et al. Gray matter networks associated with cognitive deficit in ADHD across adolescence and adulthood. Transl Psychiatry 2020; in press.

[4] Sasser TR, Kalvin CB, Bierman KL. Developmental trajectories of clinically significant attention-deficit/hyperactivity disorder (ADHD) symptoms from grade 3 through 12 in a high-risk sample: Predictors and outcomes. J Abnorm Psychol. 2016;125(2):207-219. doi:10.1037/abn0000112

[5] Caye A, Swanson J, Thapar A, et al. Life Span Studies of ADHD-Conceptual Challenges and Predictors of Persistence and Outcome. Curr Psychiatry Rep. 2016;18(12):111. doi:10.1007/s11920-016-0750-x

[6] Ellison-Wright I., Ellison-Wright Z., Bullmore E. Structural brain .2change in attention deficit hyperactivity disorder identified by meta-analysis. BMC Psychiatry. 2008;8:51

[7] Xia S., Li X., Kimball A.E., Kelly M.S., Lesser I., Branch C. Thalamic shape and connectivity abnormalities in children with attention-deficit/hyperactivity disorder. Psychiatry Res. 2012;204:161–167.

[8] Cupertino, R. B., Soheili-Nezhad, S., Grevet, E. H., Bandeira, C. E., Picon, F. A., Tavares, M., Naaijen, J., van Rooij, D., Akkermans, S., Vitola, E. S., Zwiers, M. P., Rovaris, D. L., Hoekstra, P. J., Breda, V., Oosterlaan, J., Hartman, C. A., Beckmann, C. F., Buitelaar, J. K., Franke, B., Bau, C., … Sprooten, E. (2020). Reduced fronto-striatal volume in attention-deficit/hyperactivity disorder in two cohorts across the lifespan. NeuroImage. Clinical, 28, 102403. https://doi.org/10.1016/j.nicl.2020.102403

[9] Damatac, C. G., Chauvin, R., Zwiers, M. P., van Rooij, D., Akkermans, S., Naaijen, J., Hoekstra, P. J., Hartman, C. A., Oosterlaan, J., Franke, B., Buitelaar, J. K., Beckmann, C. F., Sprooten, E. (2020). White Matter Microstructure in Attention-Deficit/Hyperactivity Disorder: A Systematic Tractography Study in 654 Individuals. Biological psychiatry. Cognitive neuroscience and neuroimaging, S2451-9022(20)30205-6. Advance online publication. https://doi.org/10.1016/j.bpsc.2020.07.015

[10] L.G. Almeida Montes, H. Prado Alcantara, R.B. Martinez Garcia, L.B. De La Torre, D. Avila Acosta, M.G. Duarte. Brain cortical thickness in ADHD: age, sex, and clinical correlations. J.Atten. Disord., 17 (2013), pp. 641-654

[11] Batty, M. J., Liddle, E. B., Pitiot, A., Toro, R., Groom, M. J., Scerif, G., Liotti, M., Liddle, P. F., Paus, T., Hollis, C. (2010). Cortical gray matter in attention-deficit/hyperactivity disorder: a structural magnetic resonance imaging study. Journal of the American Academy of Child and Adolescent Psychiatry, 49(3), 229–238.

[12] Francx W, Oldehinkel M, Oosterlaan J, et al. The executive control network and symptomatic improvement in attention-deficit/hyperactivity disorder. Cortex. 2015;73:62-72. doi:10.1016/j.cortex.2015.08.012

[13] Y. Luo, K.P. Schulz, T.L. Alvarez, J.M. Halperin, X. Li. Distinct topological properties of cue-evoked attention processing network in persisters and remitters of childhood ADHD. Cortex, 109 (2018), pp. 234-244

[14] Sudre, G., Sharp, W., Kundzicz, P., Bouyssi-Kobar, M., Norman, L., Choudhury, S., Shaw, P. (2020). Predicting the course of ADHD symptoms through the integration of childhood genomic, neural, and cognitive features. Molecular psychiatry, 10.1038/s41380-020-00941-x. Advance online publication. https://doi.org/10.1038/s41380-020-00941-x

[15] Pettersson E, Lichtenstein P, Larsson H, et al. Genetic influences on eight psychiatric disorders based on family data of 4 408 646 full and half-siblings, and genetic data of 333 748 cases and controls [published correction appears in Psychol Med. 2019 Jan;49(2):351]. Psychol Med. 2019;49(7):1166-1173. doi:10.1017/S0033291718002039

[16] Biederman J, Faraone SV, Mick E, Spencer T, Wilens T, Kiely K, Guite J, Ablon JS, Reed E, Warburton R. High risk for attention deficit hyperactivity disorder among children of parents with childhood onset of the disorder: a pilot study. Am J Psychiatry. 1995 Mar;152(3):431-5. doi: 10.1176/ajp.152.3.431. PMID: 7864271.

[17] Larsson JO, Larsson H, Lichtenstein P. Genetic and environmental contributions to stability and change of ADHD symptoms between 8 and 13 years of age: a longitudinal twin study. J Am Acad Child Adolesc Psychiatry. 2004;43(10):1267-1275. doi:10.1097/01.chi.0000135622.05219.bf

[18] Demontis, D., Walters, R.K., Martin, J. et al. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. Nat Genet 51, 63–75 (2019). https://doi.org/10.1038/s41588-018-0269-7

[19] Von Rhein, D., Mennes, M., van Ewijk, H. et al. The NeuroIMAGE study: a prospective phenotypic, cognitive, genetic and MRI study in children with attention-deficit/hyperactivity disorder. Design and descriptives. Eur Child Adolesc Psychiatry 24, 265–281 (2015). https://doi.org/10.1007/s00787-014-0573-4

[20] Gido Schoenmacker, Tom Claassen, Tom Heskes, Barbara Franke, Jan Buitelaar, Alejandro Arias-Vásquez, M17 - novel quantative method for genetic association testing, European Neuropsychopharmacology, Volume 29, Supplement 3, 2019, Pages S963-S964, ISSN 0924-977X

[21] Choi, S.W., Mak, T.S. O'Reilly, P.F. Tutorial: a guide to performing polygenic risk score analyses. Nat Protoc (2020). https://doi.org/10.1038/s41596-020-0353-1

[22] Lee, J.J., Wedow, R., Okbay, A. et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. Nat Genet 50, 1112–1121 (2018). https://doi.org/10.1038/s41588-018-0147-3

[23] Jansen, P.R., Nagel, M., Watanabe, K. et al. Genome-wide meta-analysis of brain volume identifies genomic loci and genes shared with intelligence. Nat Commun 11, 5606 (2020). https://doi.org/10.1038/s41467-020-19378-5

[24] Wray, N.R., Ripke, S., Mattheisen, M. et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. Nat Genet 50, 668–681 (2018). https://doi.org/10.1038/s41588-018-0090-3

[25] Hawkins, D. M., Basak, S. C., and Mills, D. (2003). Assessing model fit by cross-validation. Journal of Chemical Information and Computer Sciences, 43(2):579–586.

[26] Meinshausen, N. and Bühlmann, P. (2010), Stability selection. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 72: 417-473. https://doi.org/10.1111/j.1467-9868.2010.00740.x

[27] Efron, B. Bootstrap Methods: Another Look at the Jackknife. Ann. Statist. 7 (1979), no. 1, 1–26. doi:10.1214/aos/1176344552. https://projecteuclid.org/euclid.aos/1176344552

[28] Polina Golland, Feng Liang, Sayan Mukherjee, and Dmitry Panchenko. Permutation tests for classification. In Annual Conference on Learning Theory, pages 501–515, 2005.

[29] Elisseeff, A., Pontil, M. (2003). Leave-one-out error and stability of learning algorithms with applications. NATO science series sub series iii computer and systems sciences, 190, 111-130.

[30] Xu Y, Liu Y, Ridgway ND, McMaster CR. Novel members of the human oxysterol-binding protein family bind phospholipids and regulate vesicle transport. J Biol Chem. 2001 May 25;276(21):18407-14. doi: 10.1074/jbc.M101204200. Epub 2001 Feb 26. PMID: 11279184.

[31] Kharbanda M, Pilz DT, Tomkins S, et al. Clinical features associated with CTNNB1 de novo loss of function mutations in ten individuals. Eur J Med Genet. 2017;60(2):130-135. doi:10.1016/j.ejmg.2016.11.008

[32] Holland J, Sayal K. Relative age and ADHD symptoms, diagnosis and medication: a systematic review. Eur Child Adolesc Psychiatry. 2019;28(11):1417-1429. doi:10.1007/s00787-018-1229-6

[33] Biederman J, Faraone SV. Attention-deficit hyperactivity disorder. Lancet. 2005 Jul 16-22;366(9481):237-48. doi: 10.1016/S0140-6736(05)66915-2. Erratum in: Lancet. 2006 Jan 21;367(9506):210. PMID: 16023516.

[34] Lopez-Larson MP, King JB, Terry J, McGlade EC, Yurgelun-Todd D. Reduced insular volume in attention deficit hyperactivity disorder. Psychiatry Res. 2012;204(1):32-39. doi:10.1016/j.pscychresns.2012.09.009