Predicting autism spectrum disorder based on regional measures of cortical thickness

Jennifer Fedor

August 2, 2019

INFSCI 0019 Final Project

Abstract

Autism spectrum disorder (ASD) is a prevalent neurodevelopmental disorder characterized by an early onset and life-long impairments in social interaction, communication, and behavior. Although differences in the thickness of the cerebral cortex have been reported in ASD relative to typically developing (TD) controls, both overall as well as in specific regions, these results have been inconsistent. Therefore, the goal of the present study was to determine if regional measures of cortical thickness could be used to reliably classify individuals with ASD from TD controls, and thus demonstrate potential utility as a biomarker of the disorder. The sample consisted of 1035 participants, 505 with ASD and 530 TD controls, in the Autism Brain Imaging Data Exchange dataset. Cortical thickness measures in 25 regions of interest (ROIs) were calculated from structural MRI data and adjusted for confounds for each participant. Four machine learning algorithms including naive Bayes, support vector machines (SVM), k-nearest neighbors, and random forest, were used to generate initial classification models. Among the four algorithms, SVM was the best-performing, although its performance was poor with accuracy = 58.45%. Dimension reduction techniques including recursive feature elimination determined that adjusted cortical thickness in a subset of 12 ROIs were the most important features; however, a classification model using SVM and these 12 features still performed poorly (ten-fold cross-validated accuracy = 53.91%, sensitivity = 0.5408, specificity = 0.6055). The results of this analysis suggest that regional cortical thickness is not capable of classifying a heterogenous group of individuals with ASD from TD individuals.

Autism spectrum disorder (ASD) is a neurodevelopmental disorder characterized by impairments in social interaction and communication, as well as restricted patterns of behavior and interests (American Psychiatric Association, 2000). ASD affects about one in every 59 children in the US, and is about four times more prevalent in males than in females (CDC, 2019). Although typically diagnosed early on in childhood, the symptoms of ASD are life-long and associated with significant impacts on social functioning, physical health, and overall quality of life (Buescher, Cidav, Knapp, & Mandell, 2014). Therefore, methods to facilitate the early detection and diagnosis of ASD, such as through non-invasive brain imaging, are crucial.

There is still relatively little consensus on the neurobiological bases of ASD, although differences in brain structure and function have been widely studied and reported in individuals with the disorder. One widely reported finding in studies of brain structure in ASD is a difference in the thickness of the cerebral cortex, the outermost and most highly developed layer of the brain, which typically varies in thickness between 1 and 4.5 mm (Fischl & Dale, 2000). Differences in cortical thickness in ASD have been reported both overall across the cerebrum as well as within specific regions (van Rooij et al., 2018), such as the cingulate gyrus, which is important for emotion regulation, reward processing, and motivation (Hayden & Platt, 2010; Heilbronner & Hayden, 2016), and the fusiform gyrus, which is crucial for visual object processing, particularly face recognition (Weiner & Zilles, 2016). Abnormalities in cortical grey matter thickness in these areas have been associated with deficits in cognitive functioning, such as impairments in social cognition, in ASD (Hadjikhani, Joseph, Snyder, & Tager-Flusberg, 2006).

However, results of these studies have been inconsistent, with some reporting increased cortical thickness (Ecker et al., 2013; Hardan, Muddasani, Vemulapalli, Keshavan, & Minshew, 2006), and others reporting decreased cortical thickness (Hadjikhani et al., 2006; Laidi et al., 2019), in ASD relative to typically developing (TD) controls. Additionally, studies of brain structure in ASD are often limited by small sample sizes typical of neuroimaging research, and low generalizability, as many focus exclusively on subpopulations of ASD (e.g., high-functioning individuals with average IQs). Thus it remains to be determined whether regional cortical thickness reliably differs in ASD, and if these measures could potentially represent a biomarker of ASD, which could aid in the diagnosis and treatment of the disorder.

Therefore, the goal of the present study was to determine if regional measures of cortical thickness could be used to classify individuals with ASD from TD controls using supervised machine learning with a large-scale, multi-site dataset. Because cortical thickness changes with age (Zielinski et al., 2014) and differs across gender (Luders et al., 2006), and because sequences and scanner parameters differed between study sites, cortical thickness measures were first adjusted for these potential confounds.

Methods

Sample

The initial sample consisted of 1112 participants, 539 with ASD and 573 TD controls, from the Autism Brain Imaging Data Exchange (ABIDE), a neuroimaging data-sharing collaboration between 16 international sites (Di Martino et al., 2014). The dataset consists of 3T structural MRI and resting state fMRI data, and phenotypic information such as age and IQ. Usable MRI data was not available for 77 participants (e.g., due to incomplete scans or excessive motion artifact), resulting in a final sample size of 1035, including 505 individuals with ASD and 530 TD controls. Although study protocols and inclusion/exclusion criteria varied across sites, TD controls were generally free of any psychiatric or neurological disorder, and all diagnoses of ASD were made according to DSM-IV and ICD-10 criteria by "gold standard" diagnostic instruments and/or expert clinical assessment. The final sample consisted mostly of males (84.83% , N = 878) due to the prevalence of ASD, and had an average age at scan of 16.95 years (SD = 8.00, min = 6.47, max = 64.0) and an average full-scale IQ of 108.31 (SD = 14.51, min = 41.0, max = 148.0).

Predictors

Structural MRI data from the ABIDE dataset was preprocessed by and available through the Preprocessed Connectomes Project (Craddock et al., 2013). Cortical parcellation and ROI labelling was performed according to the Desikan-Killiany-Tourville protocol (Klein & Tourville, 2012), and cortical thickness (in mm) was calculated using the Advanced Normalization Tools pipeline (Tustison et al., 2013), resulting in average cortical thickness values for a total of 70 (35 bilateral) regions of interest (ROIs) for each participant. Average cortical thickness values for 25 of these 70 ROIs were considered as candidate predictors in this analysis, including: L lateral occipital gyrus, L superior temporal gyrus, L and R entorhinal gyrus, L and R fusiform gyrus, L and R middle temporal gyrus, L and R pars triangularis , L and R rostral middle frontal gyrus, L and R superior frontal gyrus, R caudal middle frontal gyrus, R cuneus, R inferior temporal gyrus, R insula, R isthmus of the cingulate gyrus, R medial orbitofrontal gyrus, R parahippocampal gyrus, R pars orbitalis, R posterior cingulate gyrus, R rostral anterior cingulate gyrus, and R transverse temporal gyrus. These ROIs were chosen because average cortical thickness was found to be significantly different in ASD compared to TD controls in these regions in a recent, large-scale study of brain morphometry in ASD (van Rooij et al., 2018).

Cortical thickness data for these 25 ROIs was first adjusted to control for the potential confounds of age, gender, and study site as described in Rao, Monteiro, & Mourao-Miranda (2017). A separate

linear regression model was fit for each ROI, with average cortical thickness in that region as the dependent variable, and age at scan and dummy variables for gender and study site as independent variables. Pearson residuals from each linear model, standardized to have a mean of 0 and standard deviation of 1, were saved and are considered to represent cortical thickness in each region after adjusting for confounds. All subsequent analyses were performed using this adjusted data.

Analysis

All analyses were performed in Jupyter Notebooks using Python version 3.7.3 and the *scikit-learn* module. The dataset was first inspected for missing values. There were no missing values for diagnostic group, age, gender, study site, or any of the cortical thickness measures, although missing data was present for some phenotypic variables that were potentially of interest (i.e., handedness, full-scale IQ). Because handedness was missing for 31.5% of the observations in the dataset (N = 326), this feature was ultimately excluded, while missing full-scale IQ values were replaced with group-specific means. Initial graphical analyses were used to examine the data for outliers and skewed distributions. Outlying values were investigated but not dropped if there was no evidence that they were incorrect or physiologically impossible. Transformations were attempted for data that did not appear acceptably normal.

Following any necessary cleaning and transformation, the dataset was split into training and testing datasets using an 80/20 split, with diagnostic group as the outcome and adjusted cortical thickness in the 25 ROIs as features. Using this data, four separate classification models were generated, each using one of the following algorithms with its default parameters unless otherwise specified: naive Bayes, support vector machines (SVM), k-nearest neighbors (kNN; k = 5), and random forest. Accuracy score and area under the receiver operating characteristic curve (AUC) were considered in determining which algorithm performed best. Once the best-performing algorithm was selected, dimension reduction techniques were then performed in an effort to improve model performance by considering: (1) features with low variance; (2) high correlations ($r > 0.8$) between features which may indicate collinearity or redundancy; (3) features eliminated through recursive feature elimination, which were determined by fitting a logistic regression model to the entire dataset, with diagnostic group as the outcome and adjusted cortical thickness in the 25 ROIs as predictors; and (4) features that were not significantly associated with the outcome ($p > 0.05$), which were determined using univariate feature selection with an ANOVA score function. A series of classification models were subsequently generated using the chosen algorithm and each of the feature sets identified through the dimension reduction methods as well as the original set, and each model was validated using ten-fold cross-validation. Final model selection and evaluation was then performed by considering the cross-validated accuracy score and AUC.

Results

Initial algorithm selection

Results from the four initial models, each using naive Bayes, SVM, kNN, or random forest algorithms, with adjusted cortical thickness values from all 25 candidate ROIs as predictors, are reported in Table 1. All four models performed similarly poorly in classifying individuals with ASD from TD controls. In particular, the random forest and naive Bayes algorithms performed the worst, with 52.17% and 53.14% accuracy, respectively. The latter result was not entirely unexpected, as the independence assumption of naive Bayes was likely violated, given that brain regions are inherently related. SVM was the best performing model, with an AUC of 0.5828 and an accuracy score of 58.45%, although these scores were still very low and only marginally better than random chance.

Table 1. Performance of the four initial classification models

Algorithm	**AUC**	**Accuracy**
Naive Bayes	0.5216	53.14%
SVM	0.5828	58.45%
kNN	0.5517	55.56%
Random forest	0.5232	52.17%

Dimension reduction

Features with (1) low variance, (2) high correlation coefficients, (3) those eliminated through a recursive feature elimination procedure, and/or (4) those not significantly associated with the outcome were considered for removal in order to reduce dimensionality and potentially improve the performance of the SVM model. All 25 features had variances approximately equal to 1, because the data had been standardized such that adjusted cortical thickness in each of the 25 ROIs had a mean of 0 and standard deviation of 1. Therefore, low variance was not able to be used as a criterion for dimension reduction.

Adjusted cortical thickness was generally very weakly correlated among the 25 ROIs with few exceptions (Figure 1). In particular, there were high correlations between the L and R fusiform gyri ($r = 0.84$), and between the L and R superior frontal gyri ($r = 0.9$). This result was not surprising as these regions are paired across hemispheres of the brain. Although below the threshold of $r = 0.8$, some of the

frontal regions were also relatively strongly correlated with one another, which again was not surprising as these regions are located within the same lobe of the brain. Unfortunately, it was not particularly obvious which, if any, of these regions to exclude in order to mitigate potential problems with collinearity or redundancy. Therefore, none of the highly correlated features were excluded at this step.

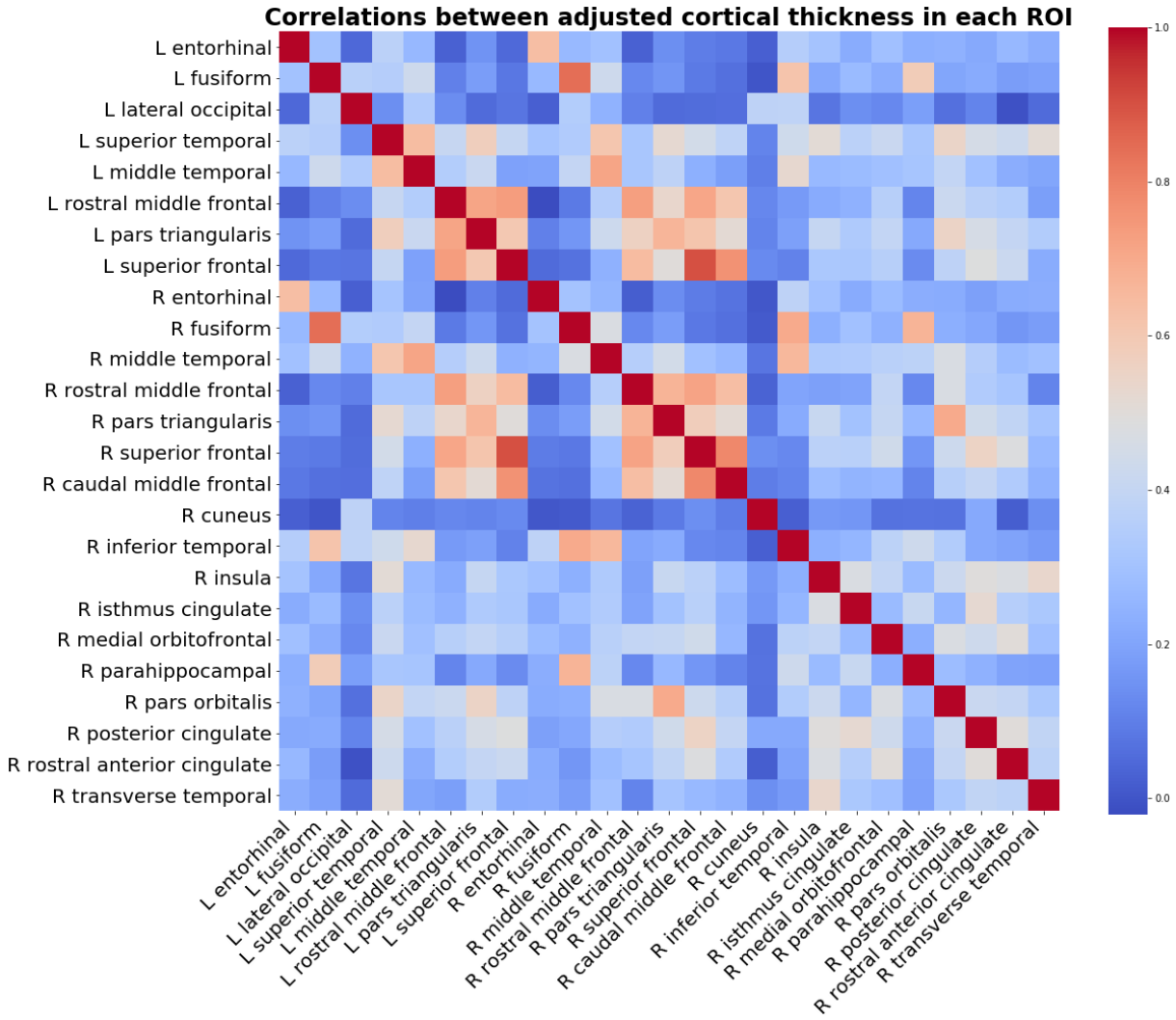

Figure 1. Correlation matrix showing correlation coefficient (r) for each pair of ROIs; darker blue values represent weaker positive correlations and darker red values represent stronger positive correlations. Adjusted cortical thickness was strongly correlated between the L and R fusiform gyri, between the L and R superior frontal gyri, and among some of the other frontal regions

After fitting a logistic regression model to the entire dataset, with diagnostic group as the outcome and adjusted cortical thickness in the 25 ROIs as predictors, recursive feature elimination (RFE) determined that 12 of the 25 predictors were the best-performing features (Figure 2). The selected

features included adjusted cortical thickness in the R rostral middle frontal gyrus, R caudal middle frontal gyrus, R fusiform gyrus, R middle temporal gyrus, R pars triangularis, R insula, R parahippocampal gyrus, R pars orbitalis, L entorhinal gyrus, R transverse temporal gyrus, L pars triangularis, and R entorhinal gyrus. Although adjusted cortical thickness in the R rostral and caudal middle frontal gyri appeared to have relatively greater feature importance compared to the other ROIs, all 12 of the selected features were included in a candidate feature set .

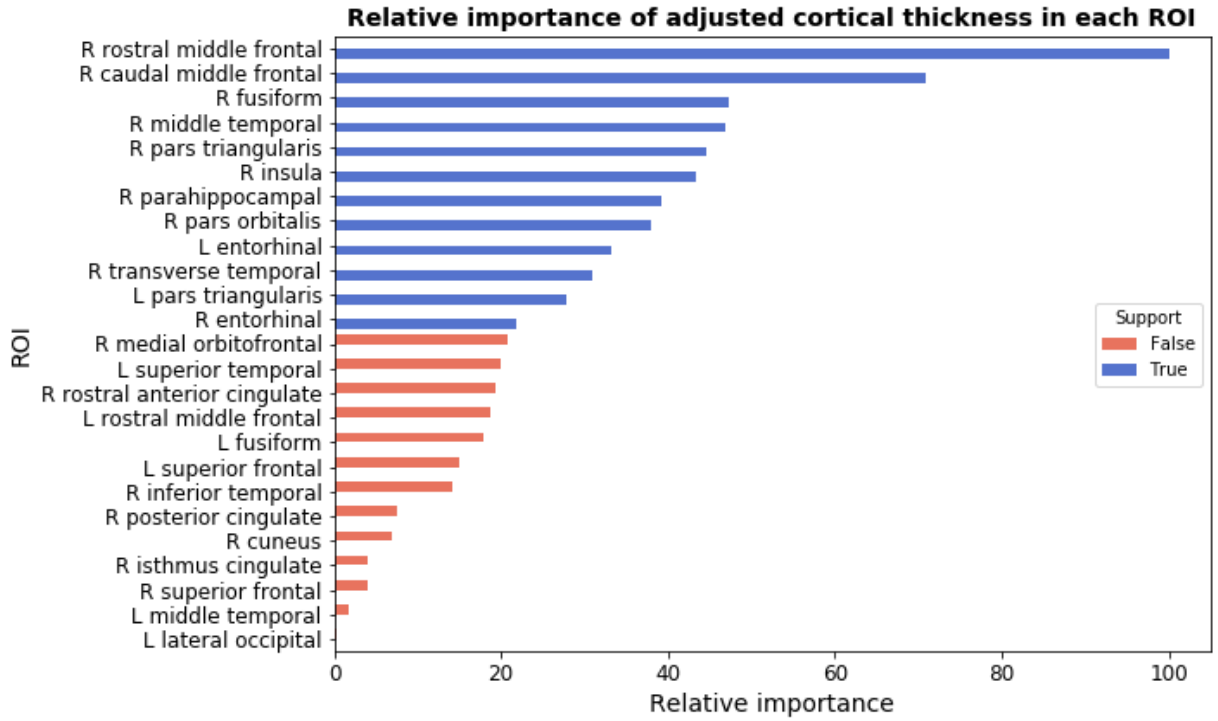

Figure 2. Relative feature importance and recursive feature elimination outcome for adjusted cortical thickness in each of the 25 ROIs; the 12 selected features are displayed in blue

Univariate feature selection using ANOVA revealed that adjusted cortical thickness was significantly different between the ASD group and the TD group in three ROIs, including the R insula ($p = 0.0076$), the R rostral middle frontal gyrus ($p = 0.0121$), and the R middle temporal gyrus ($p = 0.0200$). Notably, these three ROIs were also identified as important features through recursive feature elimination. There were no significant between-group differences in adjusted cortical thickness for any of the other ROIs (all p's > 0.05). The three ROIs in which adjusted cortical was significantly different between diagnostic groups were included in a separate candidate feature set.

Final model selection and evaluation

Results of three classification models, each using SVM and one of the candidate feature sets (the original set of 25 ROIs and the two dimension-reduced sets) are reported in Table 2. Neither of the two dimension-reduced models performed better than the initial SVM model that included adjusted cortical thickness in all 25 ROIs as features, although all three models performed relatively similarly in classifying individuals with ASD from TD controls. The model including only the features that were significantly different between the ASD group and the TD group performed the worst, correctly classifying only 55.07% of observations in the test dataset, although this model had the highest sensitivity (0.5816) compared to the other two models. The model with the original set of features appeared to perform the best overall, with the highest AUC (0.5829), specificity (0.6147), and classification accuracy for observations in the test dataset (58.45%). However, following ten-fold cross-validation, the accuracy of this model decreased to 53.71%, which suggests that the model may be overfit. Although the model containing only the features selected through recursive feature elimination had a slightly higher cross-validated accuracy score than did the model containing the original set of features (53.91% compared to 53.72%), this difference is quite small and thus likely insignificant.

Table 2. Performance of models using SVM and each of the three candidate feature sets

Feature set selection method	**N features**	**AUC**	**Sensitivity**	**Specificity**	**Accuracy**	**Cross-validated accuracy**
Original set	25	0.5829	0.5510	0.6147	58.45%	53.72%
RFE	12	0.5732	0.5408	0.6055	57.49%	53.91%
ANOVA	3	0.5523	0.5816	0.5229	55.07%	52.56%

Altogether, these results suggest that the model using an SVM algorithm and including adjusted cortical thickness values in all 25 ROIs as features is the best model for classifying individuals with ASD from TD controls. However, the model's performance was still very poor, with a cross-validated accuracy only marginally better than random chance, and a small AUC (Figure 3). Ultimately, the final model is not capable of classifying ASD from typical development.

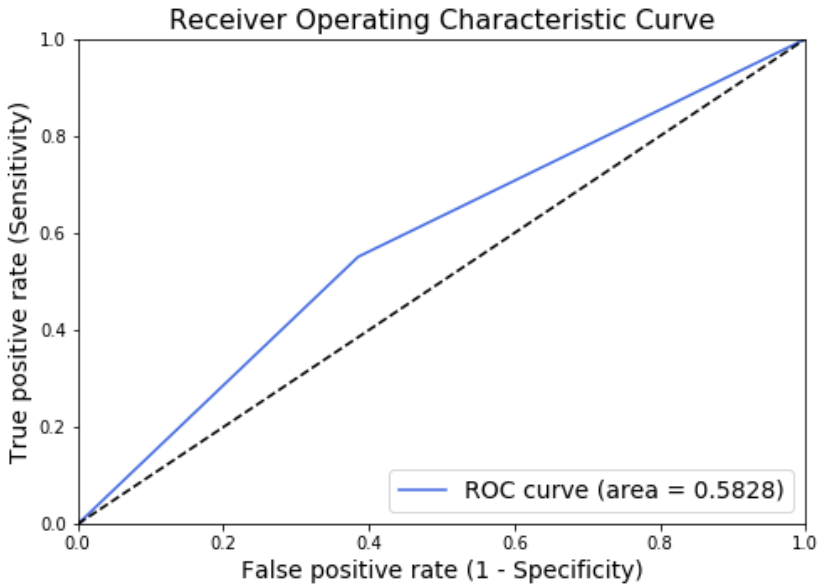

Figure 3. Receiver operating characteristic (ROC) curve for the final SVM model including adjusted cortical thickness in all 25 ROIs as features. The area under the curve was very low at 0.5829, indicating that the model was not capable of discriminating between ASD and TD

Discussion

The goal of the present study was to determine if regional measures of cortical thickness, adjusted for the potential confounds of age, gender, and study site, could classify individuals with ASD from TD controls using supervised machine learning with a large-scale dataset. Using features including adjusted cortical thickness in 25 regions of interest, none of the four machine learning algorithms were capable of accurately classifying individuals with ASD from TD controls. Among the four algorithms, SVM performed the best (albeit still poorly), correctly classifying diagnostic group for 58.45% of the observations in the testing dataset. Further efforts to improve the model's performance through dimension reduction were unsuccessful. Recursive feature elimination determined that adjusted cortical thickness in 12 regions were the best-performing features, while univariate feature selection revealed that adjusted cortical thickness was significantly different between diagnostic groups in three regions. Subsequent SVM models including these sets of 12 and three features also performed poorly in classifying ASD from TD, with average accuracy scores following ten-fold cross-validation of only 53.91% and 52.56%, respectively. Compared to these two dimension-reduced models, the original model including adjusted cortical thickness in all 25 regions appeared to perform better in classifying ASD from

TD, with a relatively higher area under the ROC curve, and similar but slightly lower cross-validated accuracy of 53.72%, although there was evidence that the model was potentially overfit. Ultimately, even this "best-performing" model was not capable of accurately classifying individuals with ASD from TD controls.

Although unsuccessful in improving model performance, the recursive feature elimination and univariate feature selection methods did reveal some interesting results. In particular, adjusted cortical thickness in the R rostral middle frontal gyrus was significantly different between the diagnostic groups, and, along with the R caudal middle frontal gyrus, had the greatest relative feature importance in classifying ASD from TD. These two regions are involved in the ventral attention network, which is important for attention and orientation to stimuli (Fox, Corbetta, Snyder, Vincent, & Raichle, 2006), and have more recently been implicated in "theory of mind" cognition, or reasoning about others' mental states (Corbetta, Patel, & Shulman, 2008). This is consistent with evidence that theory of mind is impaired in ASD and may be related to deficits in social functioning characteristic of the disorder (Mazza et al., 2017).

This study likely suffered from several limitations. In particular, aside from excluding 77 subjects whose data was unable to be preprocessed, no other quality assurance steps, such as controlling for motion, were performed in order to ensure that cortical thickness values calculated from the structural MRI data were accurate. This is potentially problematic, because in-scanner head motion is significantly greater in younger individuals and clinical populations, and has been found to bias estimates of cortical thickness and other measures of brain morphometry in these groups (Pardoe, Kucharsky Hiess, & Kuzniecky, 2016). Additionally, although this study benefited from a large sample size, substantial heterogeneity in terms of symptom severity, IQ, comorbidity, and other factors in the group with ASD may have also hindered analyses; it's possible that cortical thickness may be capable of classifying TD from specific subpopulations of ASD, such as lower-functioning individuals or those with more severe symptomatology, but not ASD more generally.

In sum, the findings of the present study suggest that regional cortical thickness is not a reliable measure for classifying ASD from typical development, and therefore likely has little utility as a potential biomarker for detection or diagnosis of the disorder. Future research could perhaps investigate if other measures of brain structure, such as grey matter volume or white matter connectivity, are better predictors of ASD than are measures of cortical thickness.

References

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders: DSM-IV*. Washington, DC.

Buescher, A. V. S., Cidav, Z., Knapp, M., & Mandell, D. S. (2014). Costs of Autism Spectrum Disorders in the United Kingdom and the United States. *JAMA Pediatrics*, *168*(8), 721–728. https://doi.org/10.1001/jamapediatrics.2014.210

CDC. (2019, April 5). Data and Statistics on Autism Spectrum Disorder | CDC. Retrieved July 28, 2019, from Centers for Disease Control and Prevention website: https://www.cdc.gov/ncbddd/autism/data.html

Corbetta, M., Patel, G., & Shulman, G. L. (2008). The Reorienting System of the Human Brain: From Environment to Theory of Mind. *Neuron*, *58*(3), 306–324. https://doi.org/10.1016/j.neuron.2008.04.017

Craddock, C., Benhajali, Y., Chu, C., Chouinard, F., Evans, A., Jakab, A., … Bellec, P. (2013). The Neuro Bureau Preprocessing Initiative: open sharing of preprocessed neuroimaging data and derivatives. *Frontiers in Neuroinformatics*.

Di Martino, A., Yan, C.-G., Li, Q., Denio, E., Castellanos, F. X., Alaerts, K., … Milham, M. P. (2014). The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. *Molecular Psychiatry*, *19*(6), 659–667. https://doi.org/10.1038/mp.2013.78

Ecker, C., Ginestet, C., Feng, Y., Johnston, P., Lombardo, M. V., Lai, M.-C., … MRC AIMS Consortium. (2013). Brain surface anatomy in adults with autism: the relationship between surface area, cortical thickness, and autistic symptoms. *JAMA Psychiatry*, *70*(1), 59–70. https://doi.org/10.1001/jamapsychiatry.2013.265

Fischl, B., & Dale, A. M. (2000). Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proceedings of the National Academy of Sciences*, *97*(20), 11050–11055. https://doi.org/10.1073/pnas.200033797

Fox, M. D., Corbetta, M., Snyder, A. Z., Vincent, J. L., & Raichle, M. E. (2006). Spontaneous neuronal activity distinguishes human dorsal and ventral attention systems. *Proceedings of the National Academy of Sciences*, *103*(26), 10046–10051. https://doi.org/10.1073/pnas.0604187103

Hadjikhani, N., Joseph, R. M., Snyder, J., & Tager-Flusberg, H. (2006). Anatomical differences in the mirror neuron system and social cognition network in autism. *Cerebral Cortex (New York, N.Y.: 1991)*, *16*(9), 1276–1282. https://doi.org/10.1093/cercor/bhj069

Hardan, A. Y., Muddasani, S., Vemulapalli, M., Keshavan, M. S., & Minshew, N. J. (2006). An MRI study of increased cortical thickness in autism. *The American Journal of Psychiatry*, *163*(7), 1290–1292. https://doi.org/10.1176/ajp.2006.163.7.1290

Hayden, B. Y., & Platt, M. L. (2010). Neurons in Anterior Cingulate Cortex Multiplex Information about Reward and Action. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *30*(9), 3339–3346. https://doi.org/10.1523/JNEUROSCI.4874-09.2010

Heilbronner, S. R., & Hayden, B. Y. (2016). Dorsal Anterior Cingulate Cortex: A Bottom-Up View. *Annual Review of Neuroscience*, *39*, 149–170. https://doi.org/10.1146/annurev-neuro-070815-013952

Klein, A., & Tourville, J. (2012). 101 Labeled Brain Images and a Consistent Human Cortical Labeling Protocol. *Frontiers in Neuroscience*, *6*. https://doi.org/10.3389/fnins.2012.00171

Laidi, C., Boisgontier, J., de Pierrefeu, A., Duchesnay, E., Hotier, S., d'Albis, M.-A., … Houenou, J. (2019). Decreased Cortical Thickness in the Anterior Cingulate Cortex in Adults with Autism. *Journal of Autism and Developmental Disorders*, *49*(4), 1402–1409. https://doi.org/10.1007/s10803-018-3807-3

Luders, E., Narr, K. L., Thompson, P. M., Rex, D. E., Woods, R. P., DeLuca, H., … Toga, A. W. (2006). Gender effects on cortical thickness and the influence of scaling. *Human Brain Mapping*, *27*(4), 314–324. https://doi.org/10.1002/hbm.20187

Mazza, M., Mariano, M., Peretti, S., Masedu, F., Pino, M. C., & Valenti, M. (2017). The Role of Theory of Mind on Social Information Processing in Children With Autism Spectrum Disorders: A Mediation Analysis. *Journal of Autism and Developmental Disorders*, *47*(5), 1369–1379. https://doi.org/10.1007/s10803-017-3069-5

Pardoe, H. R., Kucharsky Hiess, R., & Kuzniecky, R. (2016). Motion and morphometry in clinical and nonclinical populations. *NeuroImage*, *135*, 177–185. https://doi.org/10.1016/j.neuroimage.2016.05.005

Rao, A., Monteiro, J. M., & Mourao-Miranda, J. (2017). Predictive modelling using neuroimaging data in the presence of confounds. *NeuroImage*, *150*, 23–49. https://doi.org/10.1016/j.neuroimage.2017.01.066

Tustison, N. J., Avants, B. B., Cook, P. A., Song, G., Das, S., Strien, N. van, … Gee, J. C. (2013). The ANTs cortical thickness processing pipeline. *Medical Imaging 2013: Biomedical Applications in Molecular, Structural, and Functional Imaging*, *8672*, 86720K. https://doi.org/10.1117/12.2007128

van Rooij, D., Anagnostou, E., Arango, C., Auzias, G., Behrmann, M., Busatto, G. F., … Buitelaar, J. K. (2018). Cortical and Subcortical Brain Morphometry Differences Between Patients With Autism Spectrum Disorder and Healthy Individuals Across the Lifespan: Results From the ENIGMA ASD Working Group. *The American Journal of Psychiatry*, *175*(4), 359–369. https://doi.org/10.1176/appi.ajp.2017.17010100

Weiner, K. S., & Zilles, K. (2016). The anatomical and functional specialization of the fusiform gyrus. *Neuropsychologia*, *83*, 48–62. https://doi.org/10.1016/j.neuropsychologia.2015.06.033

Zielinski, B. A., Prigge, M. B. D., Nielsen, J. A., Froehlich, A. L., Abildskov, T. J., Anderson, J. S., … Lainhart, J. E. (2014). Longitudinal changes in cortical thickness in autism and typical development. *Brain: A Journal of Neurology*, *137*(Pt 6), 1799–1812. https://doi.org/10.1093/brain/awu083