Autism classified by magnetic resonance imaging: A pilot study of a potential diagnostic tool

Darko Sarovic1,2 | Nouchine Hadjikhani1,3 | Justin Schneiderman2,4 | Sebastian Lundström1 | Christopher Gillberg1,5

1Gillberg Neuropsychiatry Centre, Department of Psychiatry and Neurochemistry, Institute of Neuroscience and Physiology, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden
2MedTech West, Gothenburg, Sweden
3Department of Radiology, Athinoula A. Martinos Center for Biomedical Imaging, Harvard University, Charlestown, Massachusetts, USA
4Department of Clinical Neurophysiology, Institute of Neuroscience and Physiology, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden
5Institute of Health & Wellbeing, University of Glasgow, Glasgow, Scotland, UK

Correspondence
Darko Sarovic, Gillberg Neuropsychiatry Centre, Institute of Neuroscience and Physiology, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden.
Email: darko.sarovic@gu.se

Funding information
Vetenskapsrådet, Grant/Award Number: 2017-03359, 621-2012-3673; Avtal om Läkarutbildning och Forskning, Västra Götalandsregionen, Grant/Award Number: ALF-GBG716351; Barnancerfonden, Grant/Award Number: MT2014-007; Fredrik O Ingrid Thurings Stiftelse, Grant/Award Number: 2018-00419; Torsten Söderbergs Stiftelse, Grant/Award Number: M141/14

Abstract

Objectives: Individual anatomical biomarkers have limited power for the classification of autism. The present study introduces a multivariate classification approach using structural magnetic resonance imaging data from individuals with and without autism.

Methods: The classifier utilizes z-normalization, parameter weighting, and interindividual comparison on brain segmentation data, for estimation of an individual summed total index (TI). The TI indicates whether the gross morphological pattern of each individual's brain is in the direction of cases or controls.

Results: Morphometric analysis found significant differences within subcortical gray matter structures and limbic areas. There was no significant difference in total brain volume. A case-control pilot-study of TIs in normally intelligent individuals with autism (24) and without (21) yielded a maximal accuracy of 78.9% following cross-validation. It showed a high accuracy compared with machine learning methods when tested on the same dataset. The TI correlated well with the autism quotient (R = 0.51) across groups.

Conclusion: These results are on par with studies on autism using machine learning. The main contributions are its transparency and simplicity. The possibility of including additional neuroimaging data further increases the potential of the classifier as a diagnostic aid for neuropsychiatric disorders, as well as a research tool for neuroscientific investigations.

Keywords
autism spectrum disorder, multivariate classification, structural magnetic resonance imaging, voxel-based morphometry
1 | INTRODUCTION

Autism spectrum disorder (ASD) is characterized by difficulties in social communication, as well as repetitive and restricted behaviors and interests (American Psychiatric Association, 2013). Recent research on the neurobiology of ASD has targeted biomarker discovery that can aid in diagnostics. Having a biomarker as a compliment to the clinical interview could lower diagnostic costs, decrease the time until the individual receives a diagnosis, and potentially increase diagnostic accuracy. It could also help to define endophenotypes that can guide genetic research and evaluation of pharmacotherapy. The search for biomarkers in ASD casts a wide net. Proposed neuroanatomical biomarkers (Donovan & Basson, 2017; Ecker, 2017) include enlarged amygdalae (Sparks et al., 2002), increased cerebellar with decreased vermal size (Hardan, Minshew, Harenksi, & Keshavan, 2001; Kaufmann et al., 2003), larger caudate nuclei (Holland et al., 2005), atypical gyration (Levitt et al., 2003; Piven et al., 1990), as well as changes in hippocampal volume and shape (Nicolson et al., 2006; Schumann et al., 2004). Functional neurophysiological biomarkers (Luckhardt, Jarczek, & Bender, 2014) include an increased excitation/inhibition ratio (Rubenstein & Merzenich, 2003), impairments in the mirror neuron system (Williams, White, Suddendorf, & Perrett, 2001), local hyperconnectivity and global hypoconnectivity in coherence analyses (Belmonte et al., 2004; Catarino et al., 2013), and various changes in amplitude and timing of event-related potentials (Dawson, Webb, Carver, Panagiotides, & McPartland, 2004).

One reason for the many conflicting results in the neuroimaging of ASD is the heterogeneity of underlying causes and thus the phenotypic expression of the disorder, which necessitates identification of biologically relevant endophenotypes for deconstructing ASD (Bernhardt, Di Martino, Valk, & Wallace, 2017). Furthermore, neuroradiological studies have traditionally been hampered by manual segmentation and volumetry, which are not only time consuming (Collier et al., 2003) but also have high intra- and inter-operator variability (Despotovic, Goossens, & Philips, 2015). The advent of automated segmentation of the brain, using programs such as FreeSurfer (Fischl, 2012), has allowed researchers to segment and compare brains on a larger scale than previously possible with manual volumetry. This has allowed for the recent transition from single area volumetry to whole brain segmentation.

Attempts to reconcile the general principle of combining data from multiple biomarkers with the development of automated segmentation via factor analysis and machine learning have yielded varying amounts of success. Within the field of psychiatric neuroradiology, several machine learning paradigms have been utilized on combinations of data from magnetic resonance imaging, diffusion tensor imaging, and functional magnetic resonance imaging for classification (Orru, Pettersson-Yeo, Marquand, Sartori, & Mechelli, 2012). These include support vector machines (Ecker et al., 2010; Ecker et al., 2010; Ingahalikkar, Parker, Bloy, Roberts, & Verma, 2011; Libero, DeRamus, Lahti, Deshpande, & Kana, 2015; Uddin et al., 2011), generalized linear classifiers (Nielsen et al., 2013), logistic model trees (Jiao et al., 2010), and random forests (Zhou, Yu, & Duong, 2014).

The major strength of employing machine learning for classification is that one can analyze vast amounts of data both within and between modalities, thereby increasing the diagnostic yield of gathered data from participants. Despite this, many studies have utilized only one modality. It is likely, however, that not only the macroscopic gray matter structure, but also white matter tract thickness and connectivity pattern, as well as microscopic cytoarchitecture, functional neurophysiological aspects, and so on affect the expressed behavioral phenotype. Any complete functional model of the brain must therefore incorporate all such parameters. Recent studies utilize several modalities at once; one can expect that a greater proportion of future studies will follow this trend. For example, Libero et al. (2015) reported on a support vector machine implemented on combined magnetic resonance imaging, diffusion tensor imaging, and neurochemistry data that resulted in an accuracy of 91.9%, compared to Uddin et al. (2011)’s 88% and Ingahalikkar et al. (2011)’s 79% that were limited to one modality (magnetic resonance imaging and diffusion tensor imaging, respectively). Similarly, Zhou et al. (2014) implemented a random forest on structural and functional magnetic resonance imaging data with an accuracy of 70%, compared to Sabuncu et al.’s (2015) 59% using only structural magnetic resonance imaging.

Despite these advances, limitations in the existing literature include small sample sizes and lack of replication. Furthermore, several pitfalls specifically regarding the practical application of machine learning have also been noted (Bone et al., 2016; Kassraian-Fard, Mathis, Balsters, Maathuis, & Wenderoth, 2016). One of the most significant limitations to more widespread clinical use of machine learning is the need for a deep understanding of the theoretical underpinnings of machine learning, familiarity with the software, and rigorous testing of results in order to identify and correct errors such as overfitting. Such expertise is rarely available outside the computer science community. A transparent and easy-to-use statistical method for multivariate classification would therefore be more accessible to researchers outside the field of machine learning and, perhaps more importantly, to clinicians.

The aims of the present study were to develop a user-friendly multivariate statistical method for classification that does not rely on machine learning and apply it to brain magnetic resonance imaging segmentations in an attempt to classify a clinical case-control cohort. Herein, we present such a method and its implementation for classification of autism.

2 | METHOD

2.1 | Population

The sample consisted of 45 adult males (see Table 1): 24 autistic and 21 typically developed (TD) subjects, group matched for age and IQ. Exclusion criteria were IQ < 80 for both groups, the presence of any registered psychiatric diagnosis for the TD group, and comorbid Attention Deficit Hyperactivity Disorder (ADHD) for the ASD group. Informed consent forms were provided by all participants. The study
was ethically approved by the regional ethical board in Gothenburg (DNR: 552-14).

TD cases were recruited from the website at the Gillberg Neuropsychiatry Centre (GNC; www.gnc.gu.se) and through flyers. All but one of the 24 ASD cases were recruited from two ongoing longitudinal studies at the GNC that have been described elsewhere (Davidsson et al., 2017; Helles, Gillberg, Gillberg, & Billstedt, 2015). Briefly, this is a well-characterized sample that has been longitudinally assessed, having been assigned a diagnosis of autism on at least two occasions, separated by at least 5 years, using the Diagnostic and Statistical Manual of Mental Disorder-4 (DSM; American Psychiatric Association, 1994) and the International Classification of Diseases-10 (World Health Organization, 1992). One patient was recruited via advertisement, and his medical records were obtained and scrutinized by an experienced senior child and adolescent psychiatrist to verify that diagnosis and exclusion criteria were met. All individuals in both groups were seen by a medical doctor and a psychologist. The TD subjects were screened for neurological and psychiatric disorders using a brief neurological examination and a medical/psychiatric checklist.

The participants also underwent intelligence testing (Wechsler Abbreviated Scale of Intelligence [Wechsler, 1999] or the Wechsler Adult Intelligence Scale-IV [Wechsler, 2008]) and completed the autism quotient (AQ; Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001). IQ-test results were missing for two ASD cases and one TD case.

### 2.2 | Data acquisition

The participants’ heads were scanned using 3-Tesla magnetic resonance imaging systems at three institutions using the recommended sequences for Freesurfer segmentation MPRAGE for Siemens, the MPRAGE equivalents FSPGR-BRAVO for GE, and TFE-SENSE for Philips. T1-weighted 3D-encoded images, consisting of 176 sagittal 1 mm slices, were used for the subsequent segmentation. The specific acquisition parameters are listed in Table 2.

All scans were reviewed locally by a neuroradiologist for image quality and possible pathologic findings that, if present, were communicated to the participant and referred to the corresponding hospital for follow-up.

### 2.3 | Data processing

This study utilizes data generated by FreeSurfer (http://surfer.nmr.mgh.harvard.edu/) segmentations from the T1-weighted structural brain magnetic resonance imaging scans. FreeSurfer segmentation provides data about the volumes, areas (Ar), and thicknesses (Th) of cortical regions of interest (ROIs), as well as the volumes of subcortical gray matter (SCV) and white matter (WM) structures. The FreeSurfer recon-all pipeline with default settings was used to segment the brains. All FreeSurfer segmentations were visually inspected to ensure accurate cortical parcellations and subcortical segmentations. See Figures 1–3 for the parcellation (cortical ROIs) and segmentation results.
(subcortical ROIs) of a representative individual. The same FreeSurfer version (v5.3.0) and computer (MacOS 10.11.6) were used for all segmentations. All cortical parcellations were used. Specific segmentations were excluded as they were deemed irrelevant: WM and non-WM hypointensities, left and right vessel, optic chiasm, fifth ventricle, and left and right choroid plexus (see appendices for complete list of included data). The technical details have been explained in previous publications (Dale, Fischl, & Sereno, 1999; Fischl, Sereno, & Dale, 1999). The accuracy of FreeSurfer segmentations has been validated by both manual segmentation (Kuperberg et al., 2003) and histopathological specimens (Cardinale et al., 2014; Rosas et al., 2002). FreeSurfer is relatively insensitive to acquisition across platforms and magnetic resonance imaging manufacturers. Results are similar to within-scanner results (Han et al., 2006; Jovicich et al., 2009), with some brain areas showing lower variability (hippocampi and thalami) than others (amygdalae and accumbens areas; Morey et al., 2010).

2.4 Statistical method used for analyzing segmented magnetic resonance imaging data

We employ the sign function (defined by Equation (1)) to determine which group average is larger for each given parameter (segmented ROI). We empirically use it to return +1 for those parameters for which the ASD group has a larger average size than the TD group and −1 for those for which it is smaller.

$$\text{sgn}(\Delta g^p) = \frac{\mu^p_{\text{ASD}} - \mu^p_{\text{C}}}{|\mu^p_{\text{ASD}} - \mu^p_{\text{C}}|}$$  \hspace{1cm} (1)

where $\Delta g^p$ is the group difference for parameter $p$, and $\mu^p_{\text{ASD}}$ and $\mu^p_{\text{C}}$ are the averages for the ASD and TD groups respectively.

We define a weight for each parameter according to the square of its effect size ($d^2$) so that those parameters that have the greatest predictive value contribute the most to the total index (TI). The sign function is employed in order to maintain group directionality (i.e., positive for ASD and negative for TD) following the squaring of the effect size, which otherwise yields an absolute value. For a dataset with many parameters, one can manipulate the exponent of the effect size to increase diagnostic accuracy. By increasing the exponent, from the square to the cube for example, the weighting on the parameters that are more informative for the grouping increases. In other words, by increasing the exponent, the number of parameters that affect the TI decreases as it diminishes the importance of those parameters that are not significant in separating the groups relative to those that do. Not only integers, but also functions, such as the logarithm, can be used for effect size weighting.

$$d^2 = \text{sgn}(\Delta g^p) \times \frac{|\mu^p_{\text{ASD}} - \mu^p_{\text{C}}|^2}{SD^p_{\text{ASD}}}$$  \hspace{1cm} (2)

The value of each parameter and participant is subtracted from the mean value of the groups for that parameter and Z-normalized.
entails calculation of parameter averages, standard deviations, and used for the determination of diagnostic accuracy of the method. It

The exhaustive form of leave

2.5

case of all structural data, weight the parameters with the square of their respective effect size (Equation (2), last right-hand side term in Equation (3)) while maintaining their sign (Equation (1), middle right-hand side term in Equation (3)). As such, the sign determines the direction of each parameter (positive for ASD and negative for TD) and the magnitude is a measure of how group specific that parameter is for the participant; a large positive value for a parameter indicates that the size of the participant's ROI is in the direction of ASD and that the ROI has a distinct size difference between the groups. We calculate this for each parameter and subject:

where $X^p_n$ is the resulting normalized and weighted index value for parameter $p$ and subject $n$ compared to the mean of both groups. $X^p_n$ is the weighted value of parameter $p$ (there are $q$ of these; for the case of all structural data, $q = 230$ for participant $n$ (there are $m$ of these; for the whole group analysis in this study, $m = 45$); when we run leave-one-out cross-validation [LOOCV], $m = 44$). $\mu^p_n$ is the average value between the group averages for the parameter in question ($\mu^p$, or the average of parameter $p$ across all subjects, is used if group sizes are equal). $\mu^p_{ASD}$ and $\mu^p_{TD}$ are the group averages for ASD and TD groups for each parameter $p$.

The sum of the index values for each parameter is calculated by

$$T_{ln} = \frac{1}{q} \sum_{p=1}^{q} (X^p_n)$$

(3)

to generate the total index, $T_{ln}$, for subject $n$.

In general, TI is a singular measure of an individual's position in a spectrum of multidimensional parameter spaces collapsed onto one axis. Those parameters that best separate the groups add the most to the TI while those that do not separate the groups have little, if any, contribution to it. The TIs for our dataset summarize the pattern of brain volumes and thicknesses for each participant in relation to the sample as a whole. The midpoint between the group averages, $T_{I} = 0$, represents a standard cutoff value for classification. The TIs for this dataset—based only on magnetic resonance imaging segmentation measures—are a quantification of how "autism-like" each brain is, with positive values representing an autistic-like pattern. Thus, employing a diagnostic cutoff of $T_{I} > 0$ increases specificity and decreases sensitivity. An example spreadsheet with the statistical method and information regarding it are found in the Supplementary materials.

2.4, with one participant excluded. The method is then applied to that participant to yield the TI of the "unknown" participant. In other words, the testing set is separated from the training set. This process was iterated for each of the 45 participants in the study, and for each of the datasets and all datasets together.

We use the unweighted average recall (UAR), which is the arithmetic mean of the values for specificity and sensitivity, as a performance metric. For unequally sized groups, this metric is preferable over accuracy, since it places equal weight on both specificity and sensitivity.

Linear regression was performed on the TI and AQ data and the Pearson’s $R$, coefficient of determination ($R^2$), and statistical significance are presented.

2.6 | Comparison with machine learning methods

Four different machine learning algorithms were applied on the segmentation data (all datasets: parcellations and segmentations using FreeSurfer [see appendices for complete list]) to compare classification accuracies with the presented method: decision tree classifier, support vector machine, logistic regression, and neural network. They were implemented using the Python-based Scikit-learn module (Pedregosa et al., 2011). Although a neural network is not expected to work well for segmented, preprocessed data and small sample sizes, it was included for completion. LOOCV was used to predict the diagnostic status of the left-out individual, and the percentage of correct classifications for each method was calculated following 5000 iterations of training and testing (a 44/1 split for training/testing with split randomization and no random seed for training).

3 | RESULTS

3.1 | Behavioral phenotype

The average AQs for the ASDs and TDs were similar to previous reports in the literature (Lugnegard, Hallerback, & Gillberg, 2015; Ruzich et al., 2015) with $23.0 \pm 9.3$ and $12.2 \pm 7.0$, respectively ($p = 0.00009, d = 1.31$). Table 3 shows the linear regression results between AQ and cross-validated TIs for each of the data sets. Figure 4 illustrates the correlation using SCV data, including the regression line.

3.2 | Brain segmentation results

Following Bonferroni correction within the family of macroscopic segmentations (total intracranial volume, brain volume, cortical gray matter volume, WM, and SCV; $p < 0.01$), only SCV achieved significance ($p = 0.0009$). The results of the macroscopic segmentations are listed in Table 1. Segmentation results with group averages
Individuals in our study were classifiable as having autism or not with high accuracy using only the pattern of gray matter sizes from structural brain magnetic resonance imaging. It is our belief that the potential of this method is considerable, and that this study acts as a proof-of-concept that one could classify psychiatric disorders neuroradiologically, even when dealing with such a heterogeneous disorder as ASD. ASD is a neurodevelopmental disorder with subtle diffuse neuroanatomical differences, and for such disorders, one cannot expect individual ROIs or neurophysiological biomarkers to have particularly good predictive values at the individual level, which is why multivariate models outperform univariate models in terms of prediction (Sabuncu et al., 2015).

The performance of the current method could be further improved by incorporating data about white matter and functional neurophysiology, somatic biomarkers, and behavioral questionnaires; any quantitative measure that reliably differentiates between ASDs and TDs can be implemented to increase its diagnostic accuracy.

Our present morphometric results are mostly in line with previous research. A review by Amaral, Schumann, and Nordahl (2008) showed that although children tend to have larger total brain, gray and white matter volumes, these tend to normalize toward adulthood. Our results for these measures were insignificant following Bonferroni correction. Also, we did not find the corpus callosum to be smaller in ASDs, as in previous studies (Bellani, Calderoni, Muratori, & Brambilla, 2013). Moreover, specific cortical ROIs that we found to have altered thicknesses and areas (see Appendices A–C, respectively) correspond well both to those in other neuroanatomical studies, as well as to the ROIs that have been implicated in neurophysiological and histopathological studies. The relevant differences identified in this study clustered around the cingulate gyrus, temporal cortex (including the fusiform and entorhinal gyri), and the parahippocampal gyrus, all of which are parts of the limbic cortex. The limbic system and temporal cortex have consistently been shown to be impacted in ASD.

A comment regarding the discrepancy in the classification results using all or individual datasets is warranted. Using SCV, Ar, and WM data alone in the classification yielded higher diagnostic accuracy compared to when used in conjunction with all datasets, perhaps due to the particularly poor discrimination by Th data where the group TIs overlap (see Figure 5). This could indicate the presence of different anatomical endophenotypes in our sample; we did not, however, have enough power to perform subgroup analyses. It could also represent an idiosyncrasy due to the small sample size, which is another reason to include a larger sample in a replication study. Given that the SCV dataset had the highest correlation to the AQ, it is possible that this represents a behavioral endophenotype.

### 3.3 Overall classifier performance

The method was applied both on individual FreeSurfer segmentations (SCV, Th, Ar, and WM) and all the datasets together. The resulting TI values were statistically compared and the results are presented in Table 4. Figure 5 shows the receiver operator characteristic (ROC) curve, including area under the ROC (AUC), for each dataset. Figure 6 shows the TIs of the individuals using all structural data, together with group averages (ASD: 30.5 \(\pm\) 32.7, C: \(-30.5 \pm 41.4\)) and 95% confidence intervals (ASD: 15.0, C: 21.0).

The greatest diagnostic accuracies were obtained using individual structural datasets (SCV, Ar, and WM), while the greatest effect sizes and significance levels were obtained from SCV and all datasets combined. The AUC was highest for SCV and all datasets showing that they confer the greatest diagnostic accuracy across the spectrum of TIs.

Compared with the machine learning methods (see Table 5), the present method was minimally outperformed only by the decision tree classifier (accuracy 67.5% compared with 66.1%). However, when optimizing the cutoff value, it outperformed all the algorithms (73.2%). The optimization of the cutoff value (selection of a threshold for the TI that produces the highest accuracy) was performed with one individual excluded (using the TIs for \(n-1\) subjects), so one cannot expect a “learning” effect.

### 4 DISCUSSION

### TABLE 3 Linear regression results for total index using different data sets and the autism quotient

| Data set | \(R\) | \(R^2\) | \(p\)   | \(F\text{-test}\) |
|----------|-------|--------|--------|-----------------|
| SCV      | 0.51  | 0.26   | <0.0005| \(F(1,43) = 15.140\) |
| Th       | 0.16  | 0.03   | 0.28   | \(F(1,43) = 1.192\) |
| Ar       | 0.37  | 0.14   | 0.01   | \(F(1,43) = 6.869\) |
| WM       | 0.35  | 0.12   | 0.02   | \(F(1,43) = 6.096\) |
| All data | 0.44  | 0.19   | 0.003  | \(F(1,43) = 10.089\) |

Note: Pearson’s \(R\), \(R^2\), and significance level are presented for the linear correlation between the total index for different data sets and the autism quotient.

Abbreviations: Ar, cortical area; SCV, subcortical volume; Th, cortical thickness; WM, white matter volume.

and significance values can be found for SCV, Th, Ar, and WM in Appendices A–D, respectively. For these segmentations, group differences were apparent after Bonferroni correction for several ROIs.

Within the family of SCV, the ASD group had enlarged hippocampi bilaterally, right thalamus, and left nucleus pallidus. From the cortical ROIs, significant differences in thickness were found for the left caudal anterior, and rostral anterior cingulate gyrus, right fusiform gyrus, bilateral entorhinal gyri, right parahippocampal gyrus, right inferior and superior temporal gyri, as well as the pericalcarine gyri bilaterally. For Ar, only the inferior temporal gyri bilaterally and the left banks of the superior temporal sulcus achieved significance. No WM segmentations were significantly different following Bonferroni correction.

The greatest diagnostic accuracies were obtained using individual FreeSurfer segmentations (SCV, Th, Ar, and WM) and all the datasets together. The resulting TI values were statistically compared and the results are presented in Table 4. Figure 5 shows the receiver operator characteristic (ROC) curve, including area under the ROC (AUC), for each dataset. Figure 6 shows the TIs of the individuals using all structural data, together with group averages (ASD: 30.5 \(\pm\) 32.7, C: \(-30.5 \pm 41.4\)) and 95% confidence intervals (ASD: 15.0, C: 21.0).

The greatest diagnostic accuracies were obtained using individual structural datasets (SCV, Ar, and WM), while the greatest effect sizes and significance levels were obtained from SCV and all datasets combined. The AUC was highest for SCV and all datasets showing that they confer the greatest diagnostic accuracy across the spectrum of TIs.

Compared with the machine learning methods (see Table 5), the present method was minimally outperformed only by the decision tree classifier (accuracy 67.5% compared with 66.1%). However, when optimizing the cutoff value, it outperformed all the algorithms (73.2%). The optimization of the cutoff value (selection of a threshold for the TI that produces the highest accuracy) was performed with one individual excluded (using the TIs for \(n-1\) subjects), so one cannot expect a "learning" effect.

### 4 DISCUSSION

Individuals in our study were classifiable as having autism or not with high accuracy using only the pattern of gray matter sizes from structural brain magnetic resonance imaging. It is our belief that the potential of this method is considerable, and that this study acts as a proof-of-concept that one could classify psychiatric disorders neuroradiologically, even when dealing with such a heterogeneous disorder as ASD. ASD is a neurodevelopmental disorder with subtle diffuse neuroanatomical differences, and for such disorders, one cannot expect individual ROIs or neurophysiological biomarkers to have particularly good predictive values at the individual level, which is why multivariate models outperform univariate models in terms of prediction (Sabuncu et al., 2015).

The performance of the current method could be further improved by incorporating data about white matter and functional neurophysiology, somatic biomarkers, and behavioral questionnaires; any quantitative measure that reliably differentiates between ASDs and TDs can be implemented to increase its diagnostic accuracy.

Our present morphometric results are mostly in line with previous research. A review by Amaral, Schumann, and Nordahl (2008) showed that although children tend to have larger total brain, gray and white matter volumes, these tend to normalize toward adulthood. Our results for these measures were insignificant following Bonferroni correction. Also, we did not find the corpus callosum to be smaller in ASDs, as in previous studies (Bellani, Calderoni, Muratori, & Brambilla, 2013). Moreover, specific cortical ROIs that we found to have altered thicknesses and areas (see Appendices A–C, respectively) correspond well both to those in other neuroanatomical studies, as well as to the ROIs that have been implicated in neurophysiological and histopathological studies. The relevant differences identified in this study clustered around the cingulate gyrus, temporal cortex (including the fusiform and entorhinal gyri), and the parahippocampal gyrus, all of which are parts of the limbic cortex. The limbic system and temporal cortex have consistently been shown to be impacted in ASD.

A comment regarding the discrepancy in the classification results using all or individual datasets is warranted. Using SCV, Ar, and WM data alone in the classification yielded higher diagnostic accuracy compared to when used in conjunction with all datasets, perhaps due to the particularly poor discrimination by Th data where the group TIs overlap (see Figure 5). This could indicate the presence of different anatomical endophenotypes in our sample; we did not, however, have enough power to perform subgroup analyses. It could also represent an idiosyncrasy due to the small sample size, which is another reason to include a larger sample in a replication study. Given that the SCV dataset had the highest correlation to the AQ, it is possible that this represents a behavioral endophenotype.
The total index (TI), based on subcortical volume data (SCV), and the autism quotient (AQ) showed a moderate statistically significant correlation ($F(1,42) = 15.140, p < 0.0005$), with an $R^2$ of 0.26. Removal of the typically developing (TD) outlier (a residual of 3.11 standard deviations from the regression line) slightly improved the regression results ($F(1,43) = 19.687, R^2 = 0.32, p < 0.0005$). ASD, autism spectrum disorder.

**TABLE 4** Statistical results for TI values obtained following LOOCV

| Dataset | Student's t-test (p) | Cohen's $d$ | AUC | UAR (%) | Maximal UAR (cutoff) |
|---------|---------------------|-------------|-----|---------|---------------------|
| SCV     | 0.0013              | 0.98        | 0.792 | 72.9    | 78.9% (−6)         |
| Th      | 0.0026              | 0.88        | 0.732 | 66.4    | 72.6% (−5)         |
| Ar      | 0.026               | 0.60        | 0.742 | 75.6    | 77.7% (−1)         |
| WM      | 0.050               | 0.50        | 0.714 | 75.9    | 75.9% (0)          |
| All data| 0.0013              | 0.95        | 0.789 | 66.1    | 73.2% (12)         |

Note: Unpaired, one-tailed Student's t-test.
Abbreviations: Ar, cortical area; AUR, area under curve; LOOCV, leave-one-out cross-validation; SCV, subcortical volume; Th, cortical thickness; TI, total index; UAR, unweighted average recall; WM, white matter volume.

$^a$UAR presented using a standard cutoff value of TI = 0.

$^b$Maximal UAR obtained when optimizing the cutoff value for TI to yield the highest UAR.

### 4.1 Limitations of the study

Our sample was not population based, which could inflate diagnostic accuracy due to lower than normal variation. The TD group had similar AQs as other TD groups in previous research. However, the AQs for our ASD cases was similar to, or lower than those of previous studies, which might reflect a milder phenotype and thus underestimate the results. As with previous studies, ours has a rather small sample, and replications of the present method should aim to include larger samples for more robust baselines with which to compare potential patients; this would also improve sampling issues. Furthermore, the classification was only used on a sample with autism. As such, conclusions about its specificity for autism cannot be drawn definitively. Further work should include a sample with other neurodevelopmental disorders on which our method can be applied in order to ensure that it does not classify neurodevelopmental disorders in general, rather than ASD specifically. Since we did not include individuals with IQ < 80, generalization of these results to that group is precluded. Due to differences in neuroanatomy between the sexes (Lenroot & Giedd, 2010), females were not included. Moreover, the sample only included adult participants. The continuous brain development throughout childhood leads to higher interindividual variation. Given that autism is often diagnosed at an early age, it would be interesting to investigate how the method would perform on a very young sample. However, less stable diagnoses and differing neurodevelopmental trajectories require larger sample sizes than doing the same for adults. Finally, despite studies showing the reliability of FreeSurfer segmentation results across pipelines, a potential issue is that several imaging centers were used in this study. It would have been preferable to use one imaging center.
or to regress out site effects. Unfortunately, since these data were recorded before the development of the method, the groups were not site matched, which precludes the use of regression analysis; lest one also regress out the effect of the presence of diagnosis. We recommend that a replication study should aim to employ a homogenous data acquisition and analysis pipeline to reduce the possible influence of random error. On the other hand, positive findings across centers points to the robustness of the neuroanatomical differences between the groups.

4.2 | Use in the study of ASD

A currently unsolved issue complicating research surrounding ASD in general, and its neuroimaging in particular, is the fact that it lacks both biological and construct validity (Waterhouse, London, & Gillberg, 2017); there is no agreed upon brain-based model of ASD and the behavioral symptoms have been found to be highly heterogeneous across patients. Until subgroups of ASD have been identified and defined, this will remain a limiting factor for the field. A further complicating issue is that inter-rater reliability of the diagnosis is less than 100%. Any classification system is intricately dependent on the validity of the clinical diagnosis, and it would be a logical fallacy to assume that a classification system can outperform the clinical interview, given that the classification grouping is based on the clinical diagnosis.

The heterogeneity in ASD is a complication for multivariate analyses, such as this one, especially with small sample sizes. Defining subgroups of autism is central to ASD diagnostics in general and also of importance if brain-based classification systems are to be accurate enough at the individual level. Given that subgroups of ASD have different neuroanatomical patterns, actual case-control differences from multivariate analyses will be underestimated. One way to minimize this effect would be to standardize the clinical diagnostic procedure; if possible, the same psychiatrist should diagnose all included cases, as in this sample. Furthermore, a majority of patients with ASD have neuropsychiatric comorbidities, such as ADHD. Such patients were excluded in the present study. However, given the clinical overlap of neuropsychiatric disorders, a replication study with pairwise comparison of cases with autism, ADHD, and both is necessary to determine specificity of the method, and hence its clinical relevance.

The TI from this method is a continuous variable that resembles the nondichotomous clinical expression of ASD and allows for correlational studies with regard to clinical data. Ecker et al. (2010) were the first to show a correlation between the results of a
the comparison between multivariate analyses and questionnaires one can make them more objective and show their biological validity.

Macrocephaly in ASD has been associated with regressive symptoms and low IQ (Amaral et al., 2017). Since this sample had normal IQ and no significant brain volume difference, normalization for brain volume was not performed. Even though the macroscopic segmentations in this study (except SCV) did not reach significance, a frequent neuroanatomical finding in ASD is an increased intracranial volume. A replication study using a population-based sample could shed light on whether these findings are due to sampling differences. The question of regressing out intracranial or brain volume for ASD remains open, since in doing so, one could remove the effect of a rather stable neuroanatomical biomarker. The loss of information must be weighed against what is gained, and the decision will ultimately depend on the specific classification method employed and whether or not it improves its accuracy.

### 4.3 Applicability of the method on other psychiatric conditions

A possible future direction is to use the method as a general diagnostic aid. By compiling data from several disorders, one can apply the method on an individual for each disorder to get a neurological profile based on disorder-specific TIs and develop best fits for different disorders (like a risk profile), as well as for behavioral measures in the healthy population, such as personality traits. This would be similar to how the Mini-International Neuropsychiatric Interview (Sheehan et al., 1998) and the Structured Clinical Interview for DSM (First, Williams, Karg, & Spitzer, 2015) are being used currently during initial psychiatric assessments to guide further evaluations, but from an objective neurobiological viewpoint rather than from self-report questionnaires.

### 4.4 Methodological discussion

Rather than just being a binary classifier, the outcome measure (TI) is a continuous variable. This is beneficial since it gives a probabilistic estimate of the association with the underlying disorder and allows for correlational studies to be performed.

Utilization of machine learning methods requires specific expertize, limiting its practical usability, and rendering it unavailable to most scientists and clinical practitioners. The presented method requires only elementary mathematical knowledge and a spreadsheet, making it widely available. Despite this, the diagnostic accuracy of the present method is comparable to studies using machine learning on neuroimaging data: a maximal cross-validated classification accuracy of 78.9% compared to 59%-88% for studies using magnetic resonance imaging.

Careful consideration of the underlying neurobiology should be made regarding the choice of parameter selection. For example, since

---

**TABLE 5** Classification results, following LOOCV, for machine learning algorithms when applied on all structural datasets (parcellations and segmentations from FreeSurfer)

| Method                          | Classification accuracy (%) |
|--------------------------------|----------------------------|
| Presented classification method| 66.1 (73.2% maximal UAR)   |
| Decision tree classifier        | 67.5                       |
| Support vector machine          | 56.8                       |
| Logistic regression             | 58.1                       |
| Neural network (multilayer perceptron) | 53.6             |

Abbreviations: LOOCV, leave-one-out cross-validation; UAR, unweighted average recall.
the function of white matter tracts more closely relates to tract thickness and connectivity than to the volume of a particular white matter segment, utilization of diffusion tensor imaging data may yield superior results compared with segmentation data. Similarly, some cortical parcellations (e.g., V1, primary motor cortex, and Broca’s area) rest on logical assumptions about underlying function, and as such parcellated data may yield higher accuracies than pure voxel-based analyses.

Previous studies have shown improvements in classification accuracy for ASD both when using functional data, as well as when using several modalities at once, lending credence to the potential of further improvement of this method with the inclusion of functional neurophysiological data.

While FreeSurfer segmentation data have been shown to be reliable, any analysis pipeline includes several steps (acquisition, preprocessing, and segmentation), each of which can induce small errors. As such, it is advisable that the pipeline within a study be standardized. In line with this, the previous studies with the lowest UAR (Nielsen et al., 2013; Sabuncu et al., 2015) were the ones using data from open databases containing data with different acquisition dates regarding hardware and software versions.

4.5 | Recommendations for future multivariate analyses

Based on the reasoning above, we recommend that future studies attempting multivariate analysis (1) include larger sample sizes, (2) focus on well-defined patient cohorts, preferably diagnosed by the same physician (at least for heterogeneous disorders such as ASD), (3) test the method on another clinical sample (e.g., another neuropsychiatric disorder if investigating ASD) to assess its specificity, (4) include clinically relevant behavioral data, with which classification results can be compared, and that can potentially also be used in the analysis, (5) employ a standardized data acquisition pipeline that is up-to-date regarding hardware and software versions, (6) utilize several imaging modalities at once, and (7) base the selection of parameters on careful consideration of the underlying neurobiological processes for each disorder.

5 | CONCLUSION

Our method was utilized on magnetic resonance imaging data and yielded a maximal diagnostic accuracy of 78.9% for ASD when compared to the clinical interview, which is the gold standard of clinical diagnosis. The main contribution of this study is the development of a novel and simple multivariate classification method that requires limited specific expertise without sacrificing diagnostic accuracy in comparison with machine learning methods. This study adds to previous studies indicating it might be an achievable goal to classify psychiatric disorders—even such a heterogeneous disorder as ASD—using neuroimaging methods.

ACKNOWLEDGEMENTS

The authors would like to thank Noreen Ward for her help regarding FreeSurfer. This work was supported by Vetenskapsrådet (grant number 2017-01359; 621-2012-3673), ALF (grant number ALFGBG716351), Barn cancerfonden (MT2014-007), Fredrik O Ingrid Thurings Stiftelse (2018-00419), and Torsten Söderbergs stiftelse (grant number M141/14). Other than financial support, there was no involvement in the study.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

ORCID

Darko Sarovic https://orcid.org/0000-0002-9302-4890
Nouchine Hadjikhani https://orcid.org/0000-0003-4075-3106
Justin Schneiderman https://orcid.org/0000-0002-4441-2360
Sebastian Lundström https://orcid.org/0000-0001-7235-8499
Christopher Gillberg https://orcid.org/0000-0001-8848-1934

REFERENCES

Amaral, D. G., Li, D., Libero, L., Solomon, M., Van de Water, J., Mastergeorge, A., ... Wu Nordahl, C. (2017). In pursuit of neurophenotypes: The consequences of having autism and a big brain. *Autism Research*, 10(5), 711–722. https://doi.org/10.1002/aur.1755
Amaral, D. G., Schumann, C. M., & Nordahl, C. W. (2008). Neuroanatomy of autism. *Trends in Neurosciences*, 31(3), 137–145. https://doi.org/10.1016/j.tins.2007.12.005
American Psychiatric Association. (1994). Diagnostic and statistical manual of mental disorders (4th ed.). Arlington, VA: American Psychiatric Publishing.
American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (5th ed.). Arlington, VA: American Psychiatric Publishing.
Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (AQ): Evidence from asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders*, 31(1), 5–17.
Bellani, M., Calderoni, S., Muratori, F., & Brambilla, P. (2013). Brain anatomy of autism spectrum disorders I. Focus on corpus callosum. *Epidemiology and Psychiatric Sciences*, 22(3), 217–221. https://doi.org/10.1017/S2045796013000139
Belmonte, M. K., Allen, G., Beckel-Mitchener, A., Boulanger, L. M., Carper, R. A., & Webb, S. J. (2004). Autism and abnormal development of brain connectivity. *Journal of Neuroscience*, 24(42), 9228–9231. https://doi.org/10.1523/JNEUROSCI.3340-04.2004
Bernhardt, B. C., Di Martino, A., Valk, S. L., & Wallace, G. L. (2017). Neuroimaging-based phenotyping of the autism spectrum. *Current Topics in Behavioral Neurosciences*, 30, 341–355. https://doi.org/10.1007/7854_2016_438
Bone, D., Bishop, S. L., Black, M. P., Goodwin, M. S., Lord, C., & Narayanan, S. S. (2016). Use of machine learning to improve autism screening and diagnostic instruments: Effectiveness, efficiency, and multi-instrument fusion. *Journal of Child Psychology and Psychiatry*, 57(8), 927–937. https://doi.org/10.1111/jcpp.12559
Cardinale, F., Chinnici, G., Bramerio, M., Mai, R., Sartori, I., Cossu, M., ... Ferrigno, G. (2014). Validation of FreeSurfer-estimated brain cortical thickness: Comparison with histologic measurements. *Neuroinformatics*, 12(4), 535–542. https://doi.org/10.1007/s12021-014-9229-2
Appendix A: Subcortical volume (SCV) segmentation results

| Region of interest | Laterality/part | ASD, mean ± SD (cm³) | TD, mean ± SD (cm³) | p-value |
|--------------------|----------------|----------------------|---------------------|---------|
|脑干 | - | 22.70 ± 2.4 | 21.28 ± 2.5 | 0.06 |
|小脑白质 | Left | 55.75 ± 6.0 | 55.49 ± 5.8 | 0.88 |
| | Right | 57.71 ± 6.3 | 55.23 ± 5.7 | 0.18 |
|小脑 | Left | 15.75 ± 2.3 | 16.32 ± 2.5 | 0.43 |
| | Right | 15.51 ± 2.6 | 16.11 ± 2.4 | 0.43 |
|脑干核 | Anterior | 0.92 ± 0.2 | 0.87 ± 0.2 | 0.33 |
| | Mid-anterior* | 0.51 ± 0.1 | 0.44 ± 0.1 | 0.04 |
| | Central* | 0.47 ± 0.1 | 0.40 ± 0.1 | 0.01 |
| | Mid-posterior | 0.46 ± 0.1 | 0.41 ± 0.1 | 0.05 |
| | Posterior | 1.00 ± 0.2 | 0.89 ± 0.2 | 0.11 |
|丘脑 | Left* | 8.43 ± 0.8 | 7.54 ± 1.2 | 0.007 |
| | Right** | 7.62 ± 0.7 | 6.85 ± 0.8 | 0.002 |
|尾状核 | Left* | 4.12 ± 0.6 | 3.69 ± 0.6 | 0.03 |
| | Right* | 4.25 ± 0.7 | 3.83 ± 0.6 | 0.03 |

How to cite this article: Sarovic D, Hadjikhani N, Schneiderman J, Lundström S, Gillberg C. Autism classified by magnetic resonance imaging: A pilot study of a potential diagnostic tool. Int J Methods Psychiatr Res. 2020;29:e1846. https://doi.org/10.1002/mpr.1846
**APPENDIX A** (Continued)

| Region of interest       | Laterality/part | ASD, mean ± SD (cm<sup>3</sup>) | TD, mean ± SD (cm<sup>3</sup>) | p-value  |
|--------------------------|----------------|----------------------------------|--------------------------------|----------|
| Nucleus pallidus         | Left**         | 1.78 ± 0.2                       | 1.44 ± 0.3                     | 0.0001   |
|                          | Right*         | 1.75 ± 0.2                       | 1.60 ± 0.2                     | 0.04     |
| Amygdala                 | Left*          | 1.65 ± 0.3                       | 1.41 ± 0.4                     | 0.02     |
|                          | Right          | 1.79 ± 0.2                       | 1.67 ± 0.4                     | 0.22     |
| Hippocampus              | Left**         | 4.28 ± 0.5                       | 3.69 ± 0.6                     | 0.0007   |
|                          | Right**        | 4.46 ± 0.5                       | 3.66 ± 0.6                     | 0.00002  |
| Putamen                  | Left           | 6.30 ± 0.9                       | 6.38 ± 0.6                     | 0.74     |
|                          | Right          | 5.97 ± 0.8                       | 5.97 ± 0.8                     | 0.98     |
| Accumbens area           | Left           | 0.67 ± 0.1                       | 0.68 ± 0.1                     | 0.84     |
|                          | Right          | 0.66 ± 0.1                       | 0.63 ± 0.1                     | 0.47     |
| Ventral diencephalon     | Left           | 4.23 ± 0.4                       | 4.00 ± 0.4                     | 0.07     |
|                          | Right*         | 4.20 ± 0.4                       | 3.75 ± 0.6                     | 0.004    |

Abbreviations: ASD, autism spectrum disorder; TD, typically developed.
*p < 0.05, **p < 0.002, **Significance level after Bonferroni correction within the family of SCV = 0.05/26 = 0.0019.

**APPENDIX B** Cortical thickness (Th) segmentation results

| Region of interest       | Laterality | ASD, mean ± SD (mm) | TD, mean ± SD (mm) | p-value  |
|--------------------------|------------|---------------------|--------------------|----------|
| Caudal anterior cingulate| Left**     | 2.69 ± 0.3          | 2.25 ± 0.3         | 0.00002  |
|                          | Right      | 2.40 ± 0.3          | 2.25 ± 0.3         | 0.07     |
| Rostral anterior cingulate| Left**    | 2.91 ± 0.3          | 2.51 ± 0.3         | 0.0001   |
|                          | Right*     | 2.71 ± 0.2          | 2.44 ± 0.3         | 0.002    |
| Isthmus cingulate        | Left       | 2.54 ± 0.3          | 2.39 ± 0.3         | 0.11     |
|                          | Right      | 2.49 ± 0.2          | 2.38 ± 0.2         | 0.15     |
| Posterior cingulate      | Left       | 2.53 ± 0.2          | 2.41 ± 0.3         | 0.07     |
|                          | Right*     | 2.51 ± 0.2          | 2.30 ± 0.3         | 0.004    |
| Insula                   | Left       | 3.08 ± 0.1          | 2.98 ± 0.2         | 0.08     |
|                          | Right      | 3.06 ± 0.2          | 2.98 ± 0.3         | 0.20     |
| Lingual                  | Left*      | 2.10 ± 0.1          | 2.23 ± 0.2         | 0.05     |
|                          | Right*     | 2.17 ± 0.2          | 2.24 ± 0.1         | 0.04     |
| Fusiform                 | Left*      | 2.69 ± 0.2          | 2.56 ± 0.2         | 0.03     |
|                          | Right**    | 2.79 ± 0.2          | 2.46 ± 0.2         | 0.000004 |
| Entorhinal               | Left**     | 3.29 ± 0.4          | 2.81 ± 0.4         | 0.0005   |
|                          | Right**    | 3.28 ± 0.4          | 2.70 ± 0.4         | 0.0005   |
| Parahippocampal          | Left*      | 2.75 ± 0.3          | 2.43 ± 0.4         | 0.008    |
|                          | Right**    | 2.75 ± 0.3          | 2.31 ± 0.4         | 0.0001   |
| Precuneus                | Left       | 2.40 ± 0.1          | 2.43 ± 0.2         | 0.58     |
|                          | Right      | 2.45 ± 0.1          | 2.43 ± 0.1         | 0.75     |
| Cuneus                   | Left*      | 1.96 ± 0.1          | 2.11 ± 0.2         | 0.004    |
|                          | Right*     | 1.98 ± 0.2          | 2.15 ± 0.2         | 0.002    |

(Continues)
### APPENDIX B (Continued)

| Region of interest                  | Laterality | ASD, mean ± SD (mm) | TD, mean ± SD (mm) | p-value |
|-------------------------------------|------------|---------------------|--------------------|---------|
| Superior frontal                    | Left       | 2.68 ± 0.4          | 2.70 ± 0.1         | 0.78    |
|                                     | Right      | 2.65 ± 0.1          | 2.66 ± 0.2         | 0.79    |
| Rostral middle frontal              | Left       | 2.40 ± 0.2          | 2.38 ± 0.2         | 0.59    |
|                                     | Right*     | 2.25 ± 0.1          | 2.36 ± 0.2         | 0.05    |
| Caudal middle frontal               | Left       | 2.58 ± 0.1          | 2.56 ± 0.1         | 0.49    |
|                                     | Right      | 2.49 ± 0.2          | 2.52 ± 0.1         | 0.49    |
| Lateral orbitofrontal               | Left       | 2.65 ± 0.2          | 2.59 ± 0.2         | 0.29    |
|                                     | Right      | 2.54 ± 0.2          | 2.51 ± 0.2         | 0.61    |
| Medial orbitofrontal                | Left       | 2.41 ± 0.2          | 2.32 ± 0.1         | 0.08    |
|                                     | Right      | 2.28 ± 0.2          | 2.25 ± 0.2         | 0.43    |
| Pars triangularis                   | Left       | 2.48 ± 0.1          | 2.46 ± 0.1         | 0.63    |
|                                     | Right      | 2.38 ± 0.1          | 2.50 ± 0.1         | 0.08    |
| Pars opercularis                    | Left       | 2.69 ± 0.2          | 2.56 ± 0.1         | 0.31    |
|                                     | Right      | 2.52 ± 0.2          | 2.54 ± 0.2         | 0.86    |
| Pars orbitalis                      | Left       | 2.78 ± 0.2          | 2.71 ± 0.2         | 0.34    |
|                                     | Right      | 2.58 ± 0.3          | 2.61 ± 0.3         | 0.70    |
| Frontal pole                        | Left       | 2.60 ± 0.3          | 2.65 ± 0.2         | 0.47    |
|                                     | Right      | 2.64 ± 0.3          | 2.69 ± 0.3         | 0.87    |
| Precentral                           | Left       | 2.59 ± 0.1          | 2.64 ± 0.1         | 0.54    |
|                                     | Right*     | 2.50 ± 0.2          | 2.60 ± 0.1         | 0.04    |
| Paracentral                         | Left       | 2.43 ± 0.1          | 2.43 ± 0.2         | 0.97    |
|                                     | Right      | 2.43 ± 0.1          | 2.40 ± 0.1         | 0.43    |
| Postcentral                         | Left       | 2.14 ± 0.1          | 2.19 ± 0.1         | 0.14    |
|                                     | Right      | 2.12 ± 0.1          | 2.17 ± 0.1         | 0.19    |
| Supramarginal                       | Left       | 2.51 ± 0.2          | 2.55 ± 0.1         | 0.35    |
|                                     | Right      | 2.54 ± 0.4          | 2.55 ± 0.2         | 0.87    |
| Banks of superior temporal sulcus   | Left       | 2.43 ± 0.1          | 2.39 ± 0.2         | 0.44    |
|                                     | Right*     | 2.64 ± 0.2          | 2.50 ± 0.2         | 0.02    |
| Inferior parietal                   | Left*      | 2.38 ± 0.1          | 2.47 ± 0.1         | 0.02    |
|                                     | Right      | 2.46 ± 0.1          | 2.50 ± 0.2         | 0.34    |
| Superior parietal                   | Left*      | 2.20 ± 0.1          | 2.26 ± 0.1         | 0.05    |
|                                     | Right      | 2.21 ± 0.1          | 2.28 ± 0.1         | 0.05    |
| Inferior temporal                   | Left*      | 2.65 ± 0.2          | 2.53 ± 0.2         | 0.04    |
|                                     | Right**    | 2.76 ± 0.2          | 2.43 ± 0.3         | 0.000001 |
| Middle temporal                     | Left*      | 2.80 ± 0.2          | 2.61 ± 0.2         | 0.003   |
|                                     | Right*     | 2.88 ± 0.2          | 2.65 ± 0.2         | 0.0008  |
| Superior temporal                   | Left*      | 2.79 ± 0.2          | 2.64 ± 0.2         | 0.009   |
|                                     | Right**    | 2.85 ± 0.2          | 2.65 ± 0.2         | 0.0005  |
| Transverse temporal                 | Left       | 2.45 ± 0.3          | 2.53 ± 0.3         | 0.36    |
|                                     | Right      | 2.52 ± 0.3          | 2.58 ± 0.2         | 0.40    |
### APPENDIX B
(Continued)

| Region of interest   | Laterality | ASD, mean ± SD (mm) | TD, mean ± SD (mm) | p-value |
|----------------------|------------|---------------------|--------------------|---------|
| Temporal pole        | Left       | 3.50 ± 0.6          | 3.21 ± 0.4         | 0.06    |
|                      | Right*     | 3.68 ± 0.5          | 3.16 ± 0.5         | 0.002   |
| Pericalcarine        | Left**     | 1.76 ± 0.2          | 1.97 ± 0.1         | 0.0002  |
|                      | Right**    | 1.74 ± 0.2          | 2.02 ± 0.2         | 0.000005|
| Lateral occipital    | Left*      | 2.11 ± 0.2          | 2.29 ± 0.2         | 0.0009  |
|                      | Right      | 2.25 ± 0.2          | 2.35 ± 0.1         | 0.06    |

Abbreviations: ASD, autism spectrum disorder; TD, typically developed.

*p < 0.05, **p < 0.0007, ***Significance level after Bonferroni correction within the family of Th = 0.05/68 = 0.00074.

### APPENDIX C
Cortical area (Ar) segmentation results

| Region of interest       | Laterality | ASD, mean ± SD (cm²) | TD, mean ± SD (cm²) | p-value |
|--------------------------|------------|----------------------|---------------------|---------|
| Caudal anterior cingulate| Left       | 7.11 ± 1.4           | 6.44 ± 1.0          | 0.07    |
|                          | Right      | 8.85 ± 2.4           | 8.10 ± 1.4          | 0.21    |
| Rostral anterior cingulate| Left*     | 8.99 ± 1.6           | 8.03 ± 1.3          | 0.03    |
|                           | Right*     | 7.65 ± 1.4           | 6.65 ± 1.7          | 0.03    |
| Isthmus cingulate        | Left       | 11.44 ± 2.3          | 11.67 ± 2.1         | 0.73    |
|                          | Right      | 10.20 ± 2.0          | 10.54 ± 1.9         | 0.56    |
| Posterior cingulate      | Left       | 12.83 ± 2.5          | 12.28 ± 1.3         | 0.12    |
|                          | Right      | 13.15 ± 2.4          | 12.28 ± 1.3         | 0.15    |
| Insula                   | Left       | 23.21 ± 4.1          | 21.77 ± 2.1         | 0.15    |
|                          | Right*     | 23.65 ± 3.8          | 21.18 ± 3.0         | 0.02    |
| Lingual                 | Left*      | 32.27 ± 4.7          | 29.38 ± 3.8         | 0.03    |
|                          | Right*     | 32.23 ± 4.1          | 29.72 ± 3.8         | 0.04    |
| Fusiform                 | Left       | 35.75 ± 3.3          | 33.29 ± 4.9         | 0.05    |
|                          | Right      | 35.43 ± 4.8          | 32.00 ± 4.8         | 0.10    |
| Entorhinal              | Left       | 4.56 ± 1.2           | 4.16 ± 0.8          | 0.20    |
|                           | Right*     | 3.85 ± 1.0           | 3.31 ± 0.8          | 0.05    |
| Parahippocampal          | Left       | 7.59 ± 1.3           | 7.11 ± 1.1          | 0.17    |
|                          | Right      | 7.61 ± 1.2           | 7.01 ± 0.9          | 0.08    |
| Precuneus                | Left       | 41.52 ± 4.7          | 39.89 ± 4.7         | 0.25    |
|                          | Right      | 43.42 ± 6.9          | 42.14 ± 5.6         | 0.50    |
| Cuneus                  | Left       | 15.48 ± 1.7          | 14.41 ± 2.1         | 0.07    |
|                          | Right      | 15.68 ± 2.1          | 15.48 ± 1.9         | 0.75    |
| Superior frontal         | Left       | 77.92 ± 9.1          | 76.51 ± 8.6         | 0.60    |
|                          | Right      | 76.08 ± 9.2          | 74.05 ± 9.5         | 0.47    |
| Rostral middle frontal   | Left       | 62.72 ± 6.5          | 58.70 ± 8.0         | 0.07    |
|                          | Right*     | 66.47 ± 8.0          | 59.40 ± 9.6         | 0.01    |
| Caudal middle frontal    | Left*      | 26.34 ± 4.8          | 23.50 ± 3.5         | 0.03    |
|                          | Right      | 22.71 ± 4.6          | 22.17 ± 3.8         | 0.67    |
| Lateral orbitofrontal    | Left*      | 28.05 ± 4.0          | 24.36 ± 3.0         | 0.001   |
|                          | Right*     | 27.99 ± 3.3          | 24.59 ± 3.8         | 0.003   |

(Continues)
| Region of interest                  | Laterality | ASD, mean ± SD (cm²) | TD, mean ± SD (cm²) | p-value |
|----------------------------------|------------|----------------------|---------------------|---------|
| Medial orbitofrontal            | Left       | 19.39 ± 2.5          | 18.82 ± 2.2         | 0.42    |
|                                  | Right      | 19.69 ± 2.8          | 18.72 ± 2.3         | 0.21    |
| Pars triangularis               | Left       | 13.56 ± 2.0          | 13.46 ± 2.1         | 0.87    |
|                                  | Right      | 15.87 ± 2.5          | 16.17 ± 2.6         | 0.70    |
| Pars opercularis                | Left       | 18.94 ± 4.0          | 17.52 ± 3.1         | 0.20    |
|                                  | Right      | 14.85 ± 2.4          | 14.22 ± 1.8         | 0.33    |
| Pars orbitalis                  | Left       | 6.92 ± 0.7           | 6.56 ± 0.9          | 0.15    |
|                                  | Right*     | 8.68 ± 1.0           | 7.92 ± 1.4          | 0.04    |
| Frontal pole                    | Left       | 2.39 ± 0.3           | 2.43 ± 0.3          | 0.65    |
|                                  | Right      | 3.15 ± 0.3           | 3.18 ± 0.5          | 0.84    |
| Precentral                      | Left       | 52.01 ± 6.3          | 50.12 ± 4.6         | 0.27    |
|                                  | Right      | 53.66 ± 5.1          | 51.14 ± 4.2         | 0.08    |
| Paracentral                     | Left       | 14.64 ± 2.4          | 14.44 ± 1.6         | 0.73    |
|                                  | Right      | 16.46 ± 2.8          | 16.31 ± 2.5         | 0.84    |
| Postcentral                     | Left       | 46.38 ± 5.2          | 43.56 ± 4.2         | 0.05    |
|                                  | Right      | 43.72 ± 5.7          | 41.64 ± 4.5         | 0.19    |
| Supramarginal                   | Left       | 43.17 ± 5.3          | 40.53 ± 4.5         | 0.08    |
|                                  | Right*     | 40.87 ± 6.4          | 37.11 ± 4.8         | 0.03    |
| Banks of superior temporal sulcus | Left**    | 11.69 ± 2.0          | 9.68 ± 1.6          | 0.0007  |
|                                  | Right*     | 10.66 ± 1.3          | 9.43 ± 1.6          | 0.008   |
| Inferior parietal               | Left       | 50.39 ± 5.4          | 47.03 ± 8.1         | 0.10    |
|                                  | Right      | 61.01 ± 8.5          | 57.99 ± 8.5         | 0.24    |
| Superior parietal               | Left       | 57.73 ± 5.8          | 57.02 ± 5.5         | 0.68    |
|                                  | Right      | 58.22 ± 5.5          | 57.23 ± 6.4         | 0.58    |
| Inferior temporal               | Left**     | 38.27 ± 4.3          | 31.84 ± 5.6         | 0.00009 |
|                                  | Right**    | 37.15 ± 4.6          | 31.01 ± 5.8         | 0.0003  |
| Middle temporal                 | Left*      | 34.74 ± 4.0          | 30.76 ± 4.3         | 0.002   |
|                                  | Right*     | 39.59 ± 4.7          | 33.70 ± 4.7         | 0.001   |
| Superior temporal               | Left       | 40.62 ± 4.9          | 38.31 ± 4.4         | 0.11    |
|                                  | Right      | 38.31 ± 4.7          | 36.06 ± 4.3         | 0.10    |
| Transverse temporal             | Left       | 4.77 ± 0.8           | 4.54 ± 0.7          | 0.32    |
|                                  | Right      | 3.68 ± 0.7           | 3.44 ± 0.5          | 0.20    |
| Temporal pole                   | Left       | 5.11 ± 0.8           | 5.49 ± 0.5          | 0.06    |
|                                  | Right      | 4.44 ± 0.8           | 4.35 ± 0.7          | 0.71    |
| Pericalcarine                   | Left       | 13.62 ± 2.4          | 12.67 ± 2.3         | 0.18    |
|                                  | Right      | 15.28 ± 2.5          | 13.98 ± 2.5         | 0.09    |
| Lateral occipital               | Left       | 52.15 ± 6.0          | 49.84 ± 5.1         | 0.17    |
|                                  | Right*     | 51.06 ± 6.1          | 46.79 ± 4.5         | 0.01    |

Abbreviations: ASD, autism spectrum disorder; TD, typically developed.
*p < 0.05, **p < 0.0007, **Significance level after Bonferroni correction within the family of $Ar = 0.05/68 = 0.00074$. 
## Appendix D

White matter (WM) segmentation results

| Region of interest               | Laterality | ASD, mean ± SD (cm$^3$) | TD, mean ± SD (cm$^3$) | p-value |
|----------------------------------|------------|--------------------------|-------------------------|---------|
| Caudal anterior cingulate        | Left       | 3.03 ± 0.5               | 2.81 ± 0.4              | 0.10    |
|                                  | Right*     | 3.46 ± 0.7               | 3.11 ± 0.4              | 0.05    |
| Rostral anterior cingulate       | Left       | 2.87 ± 0.5               | 2.73 ± 0.6              | 0.36    |
|                                  | Right      | 2.43 ± 0.4               | 2.18 ± 0.4              | 0.03    |
| Isthmus cingulate                | Left       | 4.22 ± 0.7               | 4.23 ± 0.8              | 0.96    |
|                                  | Right      | 3.79 ± 0.6               | 3.74 ± 0.7              | 0.80    |
| Posterior cingulate              | Left*      | 4.84 ± 0.6               | 4.45 ± 0.4              | 0.02    |
|                                  | Right      | 4.75 ± 0.8               | 4.39 ± 0.5              | 0.07    |
| Insula                           | Left       | 9.04 ± 1.6               | 8.14 ± 0.9              | 0.03    |
|                                  | Right*     | 9.33 ± 1.7               | 7.93 ± 1.1              | 0.002   |
| Lingual                          | Left*      | 5.95 ± 1.0               | 5.06 ± 0.7              | 0.002   |
|                                  | Right*     | 5.96 ± 1.0               | 5.13 ± 0.9              | 0.008   |
| Fusiform                         | Left       | 7.33 ± 0.7               | 6.80 ± 1.4              | 0.10    |
|                                  | Right      | 7.25 ± 0.9               | 6.77 ± 1.2              | 0.14    |
| Entorhinal                       | Left       | 0.95 ± 0.3               | 0.80 ± 0.2              | 0.07    |
|                                  | Right*     | 0.78 ± 0.2               | 0.62 ± 0.2              | 0.007   |
| Parahippocampal                  | Left*      | 1.78 ± 0.3               | 1.57 ± 0.3              | 0.03    |
|                                  | Right*     | 1.95 ± 0.4               | 1.59 ± 0.3              | 0.002   |
| Precuneus                        | Left       | 10.13 ± 1.4              | 9.59 ± 1.3              | 0.18    |
|                                  | Right      | 10.80 ± 1.8              | 10.19 ± 1.6             | 0.23    |
| Cuneus                           | Left*      | 2.57 ± 0.4               | 2.25 ± 0.5              | 0.02    |
|                                  | Right      | 2.52 ± 0.4               | 2.31 ± 0.4              | 0.09    |
| Superior frontal                 | Left       | 19.37 ± 2.3              | 18.82 ± 2.3             | 0.43    |
|                                  | Right      | 18.81 ± 2.0              | 18.07 ± 3.4             | 0.37    |
| Rostral middle frontal           | Left       | 13.61 ± 1.6              | 12.80 ± 1.9             | 0.14    |
|                                  | Right*     | 14.25 ± 2.2              | 12.68 ± 2.3             | 0.03    |
| Caudal middle frontal            | Left       | 7.39 ± 1.2               | 6.82 ± 1.0              | 0.10    |
|                                  | Right      | 6.17 ± 1.1               | 5.87 ± 1.1              | 0.36    |
| Lateral orbitofrontal            | Left*      | 7.13 ± 1.0               | 6.37 ± 0.9              | 0.009   |
|                                  | Right*     | 7.17 ± 0.9               | 6.57 ± 1.0              | 0.04    |
| Medial orbitofrontal             | Left       | 3.77 ± 0.7               | 3.78 ± 0.6              | 0.94    |
|                                  | Right      | 3.72 ± 0.5               | 3.74 ± 0.5              | 0.91    |
| Pars triangularis                | Left       | 3.08 ± 0.5               | 3.16 ± 0.6              | 0.59    |
|                                  | Right      | 3.46 ± 0.6               | 3.52 ± 0.5              | 0.73    |
| Pars opercularis                 | Left       | 4.09 ± 0.9               | 3.76 ± 0.8              | 0.21    |
|                                  | Right      | 3.55 ± 0.7               | 3.32 ± 0.4              | 0.16    |
| Pars orbitalis                   | Left       | 0.99 ± 0.1               | 0.91 ± 0.1              | 0.09    |
|                                  | Right*     | 1.31 ± 0.2               | 1.18 ± 0.3              | 0.11    |
| Frontal pole                     | Left       | 0.25 ± 0.1               | 0.25 ± 0.1              | 0.89    |
|                                  | Right      | 0.34 ± 0.1               | 0.34 ± 0.1              | 0.83    |

(Continues)
### APPENDIX D (Continued)

| Region of interest                        | Laterality | ASD, mean ± SD (cm³) | TD, mean ± SD (cm³) | p-value |
|-------------------------------------------|------------|-----------------------|---------------------|---------|
| Precentral                                | Left       | 13.53 ± 3.1           | 13.49 ± 1.7         | 0.96    |
|                                           | Right      | 13.88 ± 2.5           | 14.01 ± 1.6         | 0.84    |
| Paracentral                               | Left       | 3.97 ± 0.7            | 4.08 ± 0.6          | 0.61    |
|                                           | Right      | 4.81 ± 1.0            | 5.04 ± 0.8          | 0.39    |
| Postcentral                               | Left       | 8.36 ± 1.1            | 7.75 ± 0.9          | 0.05    |
|                                           | Right      | 8.05 ± 1.2            | 7.61 ± 1.1          | 0.21    |
| Supramarginal                             | Left       | 9.40 ± 1.4            | 8.93 ± 1.2          | 0.23    |
|                                           | Right      | 9.54 ± 1.6            | 8.83 ± 1.4          | 0.12    |
| Banks of superior temporal sulcus         | Left*      | 32.4 ± 0.8            | 25.4 ± 0.7          | 0.004   |
|                                           | Right*     | 3.13 ± 0.5            | 2.77 ± 0.7          | 0.05    |
| Inferior parietal                         | Left       | 10.92 ± 1.4           | 9.92 ± 1.7          | 0.04    |
|                                           | Right      | 13.03 ± 2.0           | 12.23 ± 1.7         | 0.16    |
| Superior parietal                         | Left       | 13.02 ± 1.6           | 12.17 ± 2.5         | 0.17    |
|                                           | Right      | 12.66 ± 1.6           | 12.13 ± 2.0         | 0.33    |
| Inferior temporal                         | Left*      | 7.06 ± 0.8            | 6.05 ± 1.3          | 0.002   |
|                                           | Right*     | 6.79 ± 0.8            | 58.1 ± 1.3          | 0.004   |
| Middle temporal                           | Left*      | 5.79 ± 0.8            | 5.25 ± 0.9          | 0.04    |
|                                           | Right*     | 6.73 ± 0.8            | 5.93 ± 1.2          | 0.01    |
| Superior temporal                         | Left       | 8.29 ± 1.3            | 7.89 ± 1.1          | 0.27    |
|                                           | Right      | 7.18 ± 1.2            | 6.75 ± 0.9          | 0.20    |
| Transverse temporal                       | Left       | 0.86 ± 0.2            | 0.82 ± 0.1          | 0.42    |
|                                           | Right      | 0.61 ± 0.1            | 0.60 ± 0.1          | 0.83    |
| Temporal pole                             | Left*      | 0.76 ± 0.1            | 0.85 ± 0.1          | 0.02    |
|                                           | Right      | 0.70 ± 0.2            | 0.70 ± 0.2          | 0.98    |
| Pericalcarine                             | Left       | 3.24 ± 0.8            | 2.92 ± 0.6          | 0.13    |
|                                           | Right*     | 3.39 ± 0.7            | 2.87 ± 0.6          | 0.01    |
| Lateral occipital                         | Left       | 9.85 ± 1.3            | 9.23 ± 1.3          | 0.12    |
|                                           | Right*     | 9.94 ± 1.5            | 8.73 ± 1.2          | 0.005   |

Abbreviations: ASD, autism spectrum disorder; TD, typically developed.
*p < 0.05.