NeuroImage 172 (2018) 217–227

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Neuroanatomical morphometric characterization of sex differences in youth using statistical learning

Farshid Sepehrband [a,*], Kirsten M. Lynch [a,b], Ryan P. Cabeen [a], Clio Gonzalez-Zacarias [a,b], Lu Zhao [a], Mike D'Arcy [c], Carl Kesselman [c], Megan M. Herting [d,e], Ivo D. Dinov [a,f], Arthur W. Toga [a], Kristi A. Clark [a]

[a] Laboratory of Neuro Imaging, USC Mark and Mary Stevens Neuroimaging and Informatics Institute, Keck School of Medicine of USC, University of Southern California, Los Angeles, CA, USA
[b] Neuroscience Graduate Program, University of Southern California, Los Angeles, CA, USA
[c] USC Information Sciences Institute, University of Southern California, Los Angeles, CA, USA
[d] Department of Preventive Medicine, Keck School of Medicine of USC, University of Southern California, Los Angeles, CA, USA
[e] Department of Pediatrics, Keck School of Medicine of USC, University of Southern California, Los Angeles, CA, USA
[f] Statistics Online Computational Resource, Department of Health Behavior and Biological, University of Michigan, Ann Arbor, MI, USA

ARTICLE INFO

Keywords:
Sex difference
Cortical morphology
Multivariate modeling
SVM
GLM
FreeSurfer

ABSTRACT

Exploring neuroanatomical sex differences using a multivariate statistical learning approach can yield insights that cannot be derived with univariate analysis. While gross differences in total brain volume are well-established, uncovering the more subtle, regional sex-related differences in neuroanatomy requires a multivariate approach that can accurately model spatial complexity as well as the interactions between neuroanatomical features. Here, we developed a multivariate statistical learning model using a support vector machine (SVM) classifier to predict sex from MRI-derived regional neuroanatomical features from a single-site study of 967 healthy youth from the Philadelphia Neurodevelopmental Cohort (PNC). Then, we validated the multivariate model on an independent dataset of 682 healthy youth from the multi-site Pediatric Imaging, Neurocognition and Genetics (PING) cohort study. The trained model exhibited an 83% cross-validated prediction accuracy, and correctly predicted the sex of 77% of the subjects from the independent multi-site dataset. Results showed that cortical thickness of the middle occipital lobes and the angular gyri are major predictors of sex. Results also demonstrated the inferential benefits of going beyond classical regression approaches to capture the interactions among brain features in order to better characterize sex differences in male and female youths. We also identified specific cortical morphological measures and parcellation techniques, such as cortical thickness as derived from the Destrieux atlas, that are better able to discriminate between males and females in comparison to other brain atlases (Desikan-Killiany, Brodmann and subcortical atlases).

Introduction

The study of sex differences is of considerable scientific interest. Previous work has discovered links between sex differences and many phenotypic traits, such as behavior and susceptibility to disease (Gobinath et al., 2017; Rutter et al., 2003). In fact, several neuropsychiatric and developmental disorders manifest differently in males and females. For example, autism spectrum disorders (ASD), attention deficit and hyperactivity disorder (ADHD) and oppositional defiant disorder are more common in males (Baron-Cohen et al., 2011; Munkvold et al., 2011; Nøvik et al., 2006); while depression and anxiety are more prevalent in females (Schuch et al., 2014; Altemus et al., 2014). Moreover, because cognitive processes are rooted in neuronal architecture, the evaluation of sex differences in brain structure may provide a neuroanatomical basis for the sex differences in behavior and susceptibility to certain psychiatric disorders (Baron-Cohen et al., 2005; Gur et al., 1999; Gur and Gur, 2016). Specifically, identification of neurological structures underlying sexually dimorphic relationships may provide important insight into disease etiology and potential targets for treatment.

Previous studies of brain structure *in vivo* using magnetic resonance imaging (MRI) have revealed consistent differences in whole brain tissue volume between the sexes, with total brain volume significantly larger in

* Corresponding author.
E-mail address: farshid.sepehrband@loni.usc.edu (F. Sepehrband).

https://doi.org/10.1016/j.neuroimage.2018.01.065
Received 8 November 2017; Accepted 25 January 2018

1053-8119/© 2018 Elsevier Inc. All rights reserved.

males compared to females across all ages (Giedd et al., 1997; Goldstein, 2001; Gur and Gur, 2016; Ingalhalikar et al., 2014; Nopoulos et al., 2000; Ritchie et al., 2017). While these gross neuroanatomical differences are well-documented, more subtle regional differences in brain architecture are unclear. Previous studies using univariate parametric approaches, such as voxel-based morphometry (VBM), have yielded mixed results in cortical and subcortical structures, such as the amygdala (Andreano and Cahill, 2009; Hines, 2010; Marwha et al., 2017; Ruigrok et al., 2014), hippocampus (Cahill, 2006; Neufang et al., 2009; Ruigrok et al., 2014; Tan et al., 2016), and thalamus (Koolschijn and Crone, 2013; Ruigrok et al., 2014; Sowell et al., 2002). Discrepancies between studies could be due to differences in methodology, such as using different age ranges or different sample sizes. However, perhaps more fundamentally, the differences between studies could be due to the limitations inherent in using a univariate approach to studying sex differences. Because univariate methods neglect interactions between neuroanatomical features, they fail to account for differences with high spatially complexity (Davatzikos, 2004). This limitation could be overcome by employing a multivariate model; however, incorporating too many covariates into a generalized linear model (GLM) is not recommended, because high-dimensional modeling requires prohibitively large number of observations (Bellman, 1957; Hastie et al., 2009).

Compared to GLM approaches, multivariate statistical learning may have several advantages in establishing neuroanatomical differences between various groups, including the sexes. Specifically, multivariate statistical learning is theoretically a better approach since the problem of dimensionality can be overcome by considering the high-dimensional morphological profile as a single entity and optimizing parameters in order to reduce dimensionality (Davatzikos, 2004; Rosenblatt, 2016). Specifically, linear support vector machine (SVM) classifiers have been used to identify group differences in neuroimaging features for several neurological disorders (Bendfeldt et al., 2012; Ecker et al., 2010; Wendler, 2013). Also, studies using SVM classifiers have shown a correlation between age-related and sex-related differences in brain connectivity and cognition (Satterthwaite et al., 2015; Tunc et al., 2016). Thus, a multivariate approach using SVM may be especially useful for identifying neuroimaging features that reflect distinct neuroantomic differences between the sexes (Chekroud et al., 2016; Del et al., 2016; Rosenblatt, 2016), not previously detected using explanatory analysis (Joel et al., 2015).

To date, only a few studies have used multivariate classification approach to look at neuroanatomical differences between males and females. Wang et al. established discriminative neuroanatomical maps between sexes from anatomical and functional neuroimaging datasets of 140 healthy subjects (70 females, age range: 18–26) utilizing an SVM voxel-wise approach (i.e. each voxel was treated as a feature) (Wang et al., 2012). Similarly, Feis et al. created sex discriminative maps from anatomical and diffusion imaging datasets of 121 healthy subjects (67 females, age range: 20–30) using a voxel-based SVM approach (Feis et al., 2013). Both studies evaluated accuracy of the model in predicting an individual's sex using cross-validation (CV) on the same cohort (CV accuracies of 96% and 89%, respectively). However, limitations common to both of these studies were relatively small sample size and failure to test the model on an independent dataset. Furthermore, while both studies examined features on a voxel-wise level, it may also be valuable to identify regional neuroanatomical differences that can be used to discriminate between sexes.

The current study aims to expand upon previous findings in order to identify differences in regional neuroanatomical features between the sexes derived from structural MRI datasets of 967 youth (age range: 8–22) using a multivariate model tested on an independent multi-site cohort of 682 children and youth (age range: 3–21). Specifically, we built a linear SVM classifier comprised of cortical features, including curvature, thickness, volume and surface area, extracted from standard atlases. The SVM model for sex classification was first developed using the large single-site Philadelphia Neurodevelopmental Cohort (PNC) study (Satterthwaite et al., 2016, 2014), and then validated by applying it to the independent, multi-site Pediatric Imaging, Neurocognition and Genetics (PING) dataset (http://ping.chd.ucsd.edu/). The statistical parameters derived from applying our model to this dataset were compared against those derived from GLM. In summary, the methodology outlined in this study aims to do the following: quantify neuroanatomical differences between sexes using a multivariate SVM classifier model based on cortical morphology, determine to what extent these sex-related differences derived from this multivariate approach coincide and/or differ with those obtained from a GLM-based approach.

Materials and methods

We utilized the Big Data for Discovery Science (BDDS: http://bd2k.ini.usc.edu) (Toga et al., 2015) toolset to pre-process datasets from two independent cross-sectional youth cohorts (one single-site and one multi-site). Support vector machine (SVM) classification with a linear kernel was applied to the single-site cohort dataset to build a model for sex classification based solely on neuroimaging features. The generalizability of this model was tested on the independent, multi-site dataset ($n = 682$). We then compared the parameters derived from this SVM model to the statistical measures obtained from conventional generalized linear models (GLM). The source code for all the statistical analyses, including multivariate statistical learning and independent validation on the PING dataset, as well as GLM analysis is available on GitHub (https://github.com/sepehrband/Mining_NeuroAnat).

Datasets

Inferential and exploratory (i.e. training) analyses were performed on the Philadelphia Neurodevelopmental Cohort (PNC) dataset. The Pediatric Imaging, Neurocognition and Genetic (PING) dataset was used only as an independent dataset for testing the generalizability of the multivariate statistical learning model.

PNC dataset

Cross-sectional neuroimaging data from 997 healthy subjects from the PNC, ages 8–21 years (mean age $\pm$ SD $= 14.64 \pm 3.44$ y), including 512 females, were acquired through the database of Genotypes and Phenotypes (dbGaP) (Satterthwaite et al., 2016, 2014). Detailed acquisition parameters are described elsewhere (Satterthwaite et al., 2014). For this study, we used the three-dimensional (3D) T1-weighted structural MRI scans, acquired using a T1-weighted magnetization prepared, rapid-acquisition gradient-echo (MPRAGE) sequence with the following parameters: TR = 1810 ms, TE = 3.5 ms, FOV = 180×240 mm^2, matrix = 256×192, 160 slices, TI = 1100 ms, flip angle = 9°, effective voxel resolution = $0.9 \times 0.9 \times 1$ mm^3. For the PNC subjects, all data were collected using the same protocol on the same scanner (3T Siemens Tim Trio whole-body MRI, Erlangen, Germany; with 32-channel head coil). Of the 997 subjects, 30 subjects (14 females) were excluded because of missing demographic data, poor raw image quality, or failure in pre-processing, leaving 967 subjects for the present analysis. Demographics of the subjects included is presented in Table 1.

PING dataset

Cross-sectional structural T1-weighted MRI images were acquired

Table 1
Demographics of the subjects included from the Philadelphia Neurodevelopmental Cohort (PNC) and the Pediatric Imaging, Neurocognition and Genetic (PING) datasets.

Dataset	Number	Mean age (SD)	Age range
PNC	967	14.7 (3.4)	8.3–22.6
Female	498	15.0 (3.4)	8.6–22.6
Male	469	14.4 (3.5)	8.3–21.7
PING	682	12.0 (5.0)	3.2–21.0
Female	322	12.1 (5.1)	3.2–21.0
Male	360	12.0 (4.9)	3.2–21.0

Download English Version:

https://daneshyari.com/en/article/8687019

Download Persian Version:

https://daneshyari.com/article/8687019

Daneshyari.com