Received: 3 April 2020 | Accepted: 27 April 2020

DOI: 10.1002/hbm.25030

RESEARCH ARTICLE

WILEY

Sex classification using long-range temporal dependence of resting-state functional MRI time series

Elvisha Dhamala[1,2] | Keith W. Jamison[1] | Mert R. Sabuncu[3,4] | Amy Kuceyeski[1,2]

[1]Department of Radiology, Weill Cornell Medicine, New York, New York

[2]Brain and Mind Research Institute, Weill Cornell Medicine, New York, New York

[3]School of Electrical and Computer Engineering, Cornell University, Ithaca, New York

[4]Meinig School of Biomedical Engineering, Cornell University, Ithaca, New York

Correspondence
Elvisha Dhamala, 526 Campus Road, Biotechnology Building, Room 101D, Cornell University, Ithaca, NY, 14850.
Email: eld2024@med.cornell.edu

Funding information
NIH, Grant/Award Number: R01 NS102646-01A1 (AK), R21 NS104634-01 (AK), 1R21AG050122 (MRS), R01LM012719 (MRS), R01AG053949 (MRS); NSF CAREER, Grant/Award Number: 1748377 (MRS); NSF NeuroNex, Grant/Award Number: 1707312 (MRS)

Abstract

A thorough understanding of sex differences that exist in the brains of healthy individuals is crucial for the study of neurological illnesses that exhibit phenotypic differences between males and females. Here we evaluate sex differences in regional temporal dependence of resting-state brain activity in 195 adult male–female pairs strictly matched for total grey matter volume from the Human Connectome Project. We find that males have more persistent temporal dependence in regions within temporal, parietal, and occipital cortices. Machine learning algorithms trained on regional temporal dependence measures achieve sex classification accuracies up to 81%. Regions with the strongest feature importance in the sex classification task included cerebellum, amygdala, and frontal and occipital cortices. Secondarily, we show that even after strict matching of total gray matter volume, significant volumetric sex differences persist; males have larger absolute cerebella, hippocampi, parahippocampi, thalami, caudates, and amygdalae while females have larger absolute cingulates, precunei, and frontal and parietal cortices. Sex classification based on regional volume achieves accuracies up to 85%, highlighting the importance of strict volume-matching when studying brain-based sex differences. Differential patterns in regional temporal dependence between the sexes identifies a potential neurobiological substrate or environmental effect underlying sex differences in functional brain activation patterns.

KEYWORDS
classification, functional MRI, machine learning, neuroimaging, sex differences, temporal dependence

1 | INTRODUCTION

The study of sex differences in the brain is one of the most long-standing and debated themes in neuroscience. A compelling reason to investigate sex differences in the brain is that for many neurodevelopmental, neuropsychiatric, and neurodegenerative illnesses, the age of onset, prevalence, and symptomatology varies between the sexes. Furthermore, insight into the etiology of sex differences in the healthy brain provides an important foundation with which to delineate sex-specific pathophysiological mechanisms in different disorders and to guide the development of sex-specific treatment.

Even at rest, the brain generates an ever-changing pattern of activity that can be measured using fMRI. This activity is characterized by long-range temporal dependence such that signal fluctuations at the present time influence signal dynamics up to several minutes in the future. The Hurst exponent (HE) is a scalar measure of long-range

temporal dependence of time series that quantifies the tendency of a time series to either regress to the mean or persistently cluster in a direction (Tagliazucchi et al., 2016). The value of HE can be used to characterize the fMRI signal into three different scenarios: (a) when 0.5 < HE <1 there exist long-range temporal correlations such that a high value in the time series is likely to be followed by a second-high value and this tendency to cluster in one direction relative to the mean is likely to persist in the future; (b) when H = 0.5 there exists uncorrelated temporal activity such that the time series is similar to random noise; and, (c) when 0 < HE <0.5 there exist long-range anticorrelations such that a high value in the time series is likely to be followed by a low value and then another high value, and this tendency to fluctuate between high and low values is likely to persist into the future (Tagliazucchi et al., 2016). Based on these characteristics of the HE, it has been suggested that HE can represent signal complexity of brain activity, with a higher HE corresponding to lower signal complexity (A. M. Wink, Bernard, Salvador, Bullmore, & Suckling, 2006; Alle Meije Wink, Bullmore, Barnes, Bernard, & Suckling, 2008).

The use of HE to study signal fluctuations in functional MRI (fMRI) in both healthy and clinical populations has recently emerged (Dong et al., 2018; Lai et al., 2010; Tagliazucchi et al., 2016; Wei et al., 2013; A. M. Wink et al., 2006). In recent years, it has been observed that a baseline state of wakefulness in healthy subjects is associated with critical dynamics in the resting state time series and unconscious brain states (e.g., asleep or sedated) exhibit a departure from critical dynamics (Tagliazucchi et al., 2016). This departure corresponds to a reduction in long-range temporal dependence in the frontal lobe, salience network, and thalamus during unconsciousness (Tagliazucchi et al., 2016), suggesting that there is a relationship between consciousness and regional temporal dependence in the brain. Another recent study looked at the relationship between long-range temporal dependence and sex in healthy subjects and observed sex differences in frontal, parietal, occipital, and limbic lobes (Dong et al., 2018). In order to understand clinical or behavioral implications of HE, we must first fully quantify any baseline sex differences in the temporal dependence of resting-state time series in healthy adults.

In recent years, machine-learning techniques have increasingly been used in the analysis of resting-state fMRI data (Khosla, Jamison, Kuceyeski, & Sabuncu, 2019). Supervised methods have been successfully applied to make subject-level predictions in both healthy and clinical populations (Khosla, Jamison, Ngo, Kuceyeski, & Sabuncu, 2019). Despite the widespread interest on sex differences in the brain, very few studies have focused on sex classification using structural and/or functional information from the healthy brain. Using whole-brain functional connectomes and a partial least squares approach applied to data from the Human Connectome Project (HCP), sex classification has been performed with a maximum accuracy of 79–86% (Zhang, Dougherty, Baum, White, & Michael, 2018). Another study using a subset of the HCP data compared sex classification accuracies of 436 different models, each one based on a region's connectivity to the rest of the brain, in an effort to reduce the overall dimensionality. They reported mean accuracies of 60.0 to 68.7%, and a maximum accuracy of 75.1% (Weis et al., 2020). While sex classification using functional connectivity profiles has achieved varying levels of accuracy (Weis, Patil, et al., 2020; Zhang et al., 2018), it has not yet been attempted, to the best of our knowledge, using temporal dependence measures of brain activation like HE.

The present study evaluated sex differences in regional long-term temporal dependence of the resting-state fMRI time series in a GM-volume-matched subset of 390 subjects from the HCP (Van Essen et al., 2013) data set. The project's primary goals were: (a) to quantify regional sex differences in temporal dependence of cortical and subcortical grey matter (GM) regions and (b) to test if regional temporal dependence information is sufficient to successfully classify subjects based on sex. As a secondary goal, this study also evaluated whether regional volumetric sex differences persisted even after matching subjects on total GM volume.

2 | METHODS

An outline of our overall workflow can be found in Figure 1. All codes used for data analysis are available on GitHub (https://github.com/elvisha/HurstSexDifferences).

2.1 | Data set

Publicly available high-resolution, preprocessed magnetic resonance imaging data from the Human Connectome Project—Young Adult S1200 release (Van Essen et al., 2013) were used in this study. We examined time series from 1,003 volume-matched healthy adults having four complete resting-state fMRI runs (1,200 volumes each). The subjects had a mean age of 28.71 years (range: 22–37 years, *SD* = 3.71, median = 29) and included 469 males (46.8%). In order to avoid biases introduced by volumetric differences between males and females, a subset of 390 volume-matched subjects (190 non-overlapping male–female pairs) were identified and all analyses for this study were performed solely on that subset unless otherwise specified. The mean age of the 390 subjects was 28.6 years (range 22–37 years, *SD* = 3.7, median = 29).

Although the term “gender” is used in the HCP data dictionary, we use the term “sex” in this article because the database collected subject self-reported biological sex as opposed to gender identification. Genetic information was not used to verify the self-reported biological sex.

2.2 | Preprocessing of fMRI data

HCP MR imaging data were acquired on a Siemens Skyra 3 T scanner at Washington University in St. Louis. Each subject underwent four gradient-echo EPI rfMRI runs (TR = 720 ms, TE = 33.1 ms, 2.0 mm isotropic voxels, FoV = $208 \times 180\ mm^2$, flip angle = 52°, matrix = 104 × 90, 72 slices) of about 15 min each over two sessions: two runs in the first sessions (REST1_LR and REST1_RL) and two runs

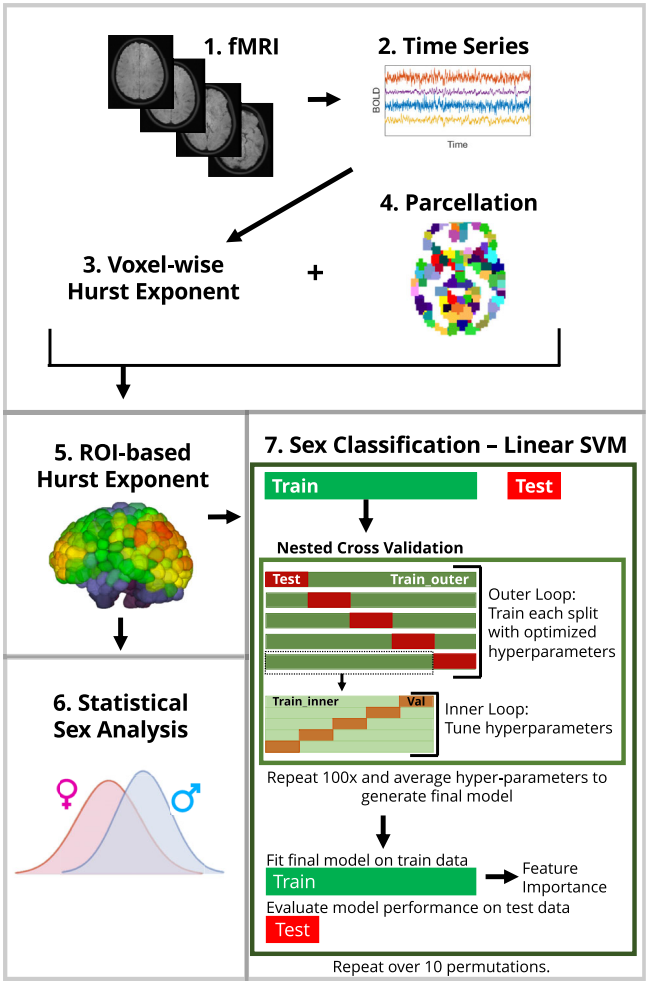

FIGURE 1 Overall workflow of analysis. (1) Functional MRI images from the Human Connectome Project were used to extract (2) voxel-wise time series. (3) Voxel-wise Hurst exponents were computed. (4) Parcellations of the voxels were generated for seven different atlases and (5) Hurst exponents were averaged for all voxels within a given ROI to generate ROI-based Hurst exponents. (6) Sex differences in regional Hurst exponent were analyzed using a Student's *t* test. (7) Prediction of sex was performed using a linear SVM classifier. Nested cross validation was used to optimize hyperparameters and a final model was fitted to the train data and evaluated on the test data

in the second session (REST2_LR and REST2_RL). The data consisted of 1,200 volumes for each run for a total of 4,800 volumes for each subject over the four runs. Each run of each subject's rfMRI was preprocessed by the HCP consortium (Smith et al., 2013). The data were minimally preprocessed (Glasser et al., 2013) and had artefacts removed using ICA + FIX (Griffanti et al., 2014; Salimi-Khorshidi et al., 2014).

2.3 | Motion

Framewise displacement (Power et al., 2014) (FD) was computed for each subject. Sex differences in framewise displacement for the specific windows used to calculate HE were evaluated using a two-sample *t* test. There were no sex differences identified in motion.

2.4 | Parcellations

As part of the HCP preprocessing pipeline (Glasser et al., 2013), FreeSurfer's recon-all pipeline (Dale, Fischl, & Sereno, 1999; Fischl et al., 2002; Fischl et al., 2008; Fischl, Liu, & Dale, 2001; Fischl, Sereno, & Dale, 1999; Fischl, Sereno, Tootell, & Dale, 1999; Segonne, Grimson, & Fischl, 2005) was optimized for the high-resolution HCP data. Structural T1-weighted and T2-weighted images (0.7 mm isotropic) were used with the optimized HCP FreeSurfer pipeline to automatically segment subcortical structures (Fischl et al., 2002) and cortical sulci and gyri (Desikan et al., 2006). These segmentations were used to reduce the fMRI data of the whole brain into cortical and subcortical GM regions of interest (ROI). All analyses were repeated seven times using the following seven different parcellations, where R denotes the number of ROIs: Automated Anatomical Labeling (Tzourio-Mazoyer et al., 2002) (AAL, R = 116), Craddock 200 (Craddock, James, Holtzheimer 3rd, Hu, & Mayberg, 2012) (CC200, R = 200), Craddock 400 (Craddock et al., 2012) (CC400, R = 392), Eickhoff-Zilles (Eickhoff et al., 2005) (EZ, R = 116), Harvard-Oxford (Goldstein et al., 2007; Makris et al., 2006) (HO, R = 110), Talairach and Tournoux (Lancaster et al., 2000) (TT, R = 94), as well as a FreeSurfer aparc+aseg parcellation (FS86, R = 86) in which cortical GM was divided into 34 regions per hemisphere as per the Desikan-Killiany atlas (Desikan et al., 2006), and subcortical GM was divided into nine regions per hemisphere. Results from the CC400 atlas are presented in the main article, while results from all other atlases are shown in the supplemental materials.

2.5 | Hurst exponent

Detrended fluctuation analysis (DFA) was used to analyze the temporal correlations of BOLD fluctuations, and estimate long-range temporal dependence in time series while accounting for nonstationarities. The obtained exponent from DFA is a generalization of the HE, with the exception that DFA may also be applied to time series whose underlying statistics or dynamics are nonstationary. In order to estimate the HE, DFA was used as previously described (Tagliazucchi et al., 2016) and summarised below.

1 Given x_t, series of T measurements, $1 \le t \le T$, subtract mean of signal, $<x>$, and compute cumulative sum, $X_t = \sum_{i=1}^{t}(x_i - < x >)$.
2 Divide signal into nonoverlapping windows of length L, with each segment labeled as Y_j^k, with j indexing time within segment ($1 \le j \le L$) and k indexing segment number ($1 \le k \le T/L$).
3 For each segment, fit a linear function using the least squares to determine slope, a^k and intercept, b^k.Subtract best fitting linear trend and compute fluctuation from mean of resulting signal,

$$F^k = \sqrt{\frac{1}{L}\sum_{j=1}^{L}\left(Y_j^k - a^k j - b^k\right)^2}.$$

4 Average F^k for all segments at temporal scale L to yield fluctuation function, $F(L)$.
5 Repeat 2–4 using different window lengths, L = [10, 15, 25, 30].
6 The slope of $F(L)$ in logarithmic scale versus L is the HE.

For every subject, DFA was applied to 32 segments (8 segments per scan) of 150 volumes each of the BOLD time series of every voxel. The HE computed over the 32 segments were averaged together to generate a single mean HE for each voxel. The HE was averaged over voxels within each ROI to generate a single HE value per ROI.

2.6 | Volume

To compare ROI size differences, subject-specific ROI volumes were computed by counting voxels in each subject's native 0.7 mm anatomical space. The FreeSurfer-derived FS86 atlas was already defined in this native space. The remaining atlases were defined in 2 mm MNI152 space, and were first transformed from MNI152 to each subject's native space using the standard2acpc_dc nonlinear warp provided by HCP and nearest-neighbor interpolation.

To avoid biases introduced by volumetric sex differences, 390 subjects (190 nonoverlapping male–female pairs) were selected such that each pair had a matched total GM volume (cortical and subcortical) with a percent difference less than or equal to 1%. The final volume-matched sample did not differ in total GM volume ($p > .05$) between males and females.

2.7 | Sex differences in HE and volume

For each atlas, sex differences in HE across the ROIs were evaluated. For each ROI, the difference in means between males and females was analyzed using a two-tailed two-sample t test. This was repeated with all atlases. p-values were corrected for multiple comparisons across all regions in all atlases (i.e., corrections accounted for 1,114 regions which is the total number of regions across all atlases) using the Benjamini-Hochberg procedure (Benjamini & Hochberg, 1995) to decrease the false discovery rate. This same process was repeated for volumetric sex differences.

2.8 | Sex classification

In order to classify subjects in the GM-volume-matched subset into males and females using ROI-based HE or regional volume from each atlas, a linear support vector machines (SVM) approach was implemented (Pedregosa et al., 2011). SVMs (Cortes & Vapnik, 1995) are supervised learning models which can be used for classification tasks. Given labeled training data with n features, the algorithm outputs an optimal hyperplane which can be used to separate the classes with a maximal margin. If the data are linearly separable, linear SVMs can be used to find this optimal hyperplane. Linear ridge, SVM with radial basis function kernel, random forest, and neural network classifiers were also evaluated in this work. However, the performance metrics obtained from those classifiers were worse than those obtained for linear SVM so the results for this article focus on linear SVM.

GM-volume-matched subjects (n = 390) were separated into stratified train (80%) and test (20%) subsets and the same train and test subsets were used across all atlases so that results could be directly compared with one another. For each atlas, hyperparameter tuning of penalty parameter C was conducted using nested cross validation in which both the inner and outer folds were randomly split into five stratified groups. A single best model and the hyperparameter C corresponding to it were identified. A coarse grid search followed by a finer grid search across the same hyperparameter space was conducted for all atlases. This nested grid search cross validation was repeated 100 times to generate 100 separate values for C. The final model was created by averaging the hyperparameter C obtained across 100 iterations of the finer grid search. This final model was evaluated on the test subset and the corresponding accuracy, area under the receiver operating characteristics curve (AUC), and feature importance are reported. This was repeated using the same 10 permutations of randomized training and testing splits for all atlases to get a distribution of overall performance metrics for both HE- and volume-based classification. The feature weights vector obtained from the linear SVM classifier for each atlas was averaged across the permutations and the absolute value of the average scaled such that feature importance for each ROI falls between zero and one, and it is easier to interpret across atlases. An ensemble model that combined prediction probability for each test subject's sex across all atlases (Khosla, Jamison, Kuceyeski, & Sabuncu, 2019) was also implemented for both HE- and volume- based sex classification separately. Atlas-specific ensemble models that combined prediction probability from the HE- and volume- based classification for each test subject's sex was also implemented. Lastly, a final ensemble model combined prediction probabilities from both the HE- and volume- based classification for each test subject's sex across all atlases. Differences in the classification AUC and accuracy between the HE-based classification, volume-based classification, and ensemble model classification based on both HE and volume were evaluated using a one-way analysis of variance (ANOVA).

Sex classification based on HE was also performed for the entire data set and randomly sample-size-matched subsets using the CC400 atlas. Ten permutations of sex classification, as described above, were completed using stratified splits of the entire data set (n = 1,003, 46.8% male) to determine the overall effect of GM volume matching on the performance metrics. Additional permutations of sex classification using randomly generated subsets were also performed to determine how an overall decrease in sample size (unrelated to GM volume matching) may affect the results. Ten distinct subsets (n = 390) were

randomly generated by selecting 195 males and 195 females from the entire data set. For each of the 10 distinct subsets, 10 permutations of sex classification, as described above, were completed. Balanced accuracy and AUC distributions were generated by averaging the results across the subsets. Finally, differences in the classification AUC and balanced accuracy between the GM-volume-matched subset, randomly sample-size-matched subset, and entire data set were evaluated using a one-way analysis of variance (ANOVA).

3 | RESULTS

We present results from a cohort of 390 GM-volume-matched healthy young adults (ages 22–37; 190 males) from the HCP—Young Adult S1200 data set (Van Essen et al., 2013). Each subject had four complete resting state fMRI scans with 1,200 volumes each that were acquired over two sessions and preprocessed (Smith et al., 2013). Each of the four fMRI scans was divided into eight segments of 150 volumes each, and HE was estimated using detrended fluctuation analysis (Tagliazucchi et al., 2016) on each segment. Voxel-wise HE maps were obtained by averaging HE over all segments from each of the four scans ($8 \times 4 = 32$ segments total) to improve signal-to-noise ratio. Regional HE was obtained by averaging the HE values of each voxel within a given region. Two sample *t* tests were used to evaluate regional sex differences in HE and all *p*-values were corrected for multiple comparisons (Benjamini & Hochberg, 1995). A linear SVM was optimized using 100 iterations of nested grid search with fivefold inner and outer cross validation and fit on a training subset of the total data (80%). Classification performance is calculated based on the final model's performance on the remaining 20% hold-out test data. This procedure was repeated over 10 permutations to assess robustness to training/test set assignments. To ensure that results were not subjected to biases introduced by a single GM parcellation scheme, all analyses were replicated over seven different atlases. An ensemble model that took the average of each of the individual atlas model predictions was also created. Finally, to quantify the effect of sample size and volume-matching on results, HE-based sex classification using the CC400 atlas was also performed: (a) for the entire data set (n = 1,003) and (b) for 10 randomly selected (not volume-matched) subsets of equal size to the volume-matched sample (n = 390). Additionally, as a secondary analysis, regional volume was used to perform sex classification to determine whether volume alone, even after strict GM volume matching, can still identify sex. Details are given in Methods section and an outline of the workflow can be seen in Figure 1. For clarity, we primarily present the results for the CC400 atlas in this article and results for all other atlases are shown in the Supplementary Materials.

3.1 | Sex differences in HE

Significant sex differences (p-corrected<.05) in HE was observed in 28 cortical and subcortical GM regions of the CC400 atlas, including the left (L) and right (R) superior and middle temporal gyri, L and R precentral and postcentral gyri, L and R paracentral lobules, L middle occipital gyrus, R parahippocampal gyrus, L cingulate gyrus, and R cuneus. In all regions found to be significantly different between the sexes, males exhibited a higher HE than females. Group average HE values for males and females are shown in Figure S1. The t-statistics for regions exhibiting significant sex differences in HE are shown in Figure 2a for the CC400 atlas. Results from other atlases demonstrate similar regional patterns of significant sex differences and are shown in the Supplementary Materials.

3.2 | Sex classification using HE

ROC curves for HE-based sex classification for all atlases can be seen in Figure 2b, and accuracy and AUC can be found in Table 1. For HE-based classification, a mean accuracy of 81.2% (SD = 3.5%) and a mean AUC of 0.87 (SD = 0.033) was achieved using the CC400 atlas. Regions exhibiting the strongest feature importance were the cerebellum, amygdala, frontal cortex, and occipital cortex, and a feature importance map is visualized in Figure 2c. Results from other atlases identified similar regions as important features for sex classification and are shown in the Supplementary Materials.

3.3 | Sex differences and classification using volume

As a secondary set of analyses, sex differences in regional volume were analyzed, and sex classification was repeated using regional volume information. Significant sex differences (p-corrected<.05) in regional volume was observed in 149 cortical and subcortical GM regions of the CC400 atlas. Females exhibited larger absolute volumes in the bilateral cingulate cortices, pre- and postcentral gyri, precunei, and frontal gyri, while males exhibited larger absolute volumes in bilateral cerebella, hippocampi and parahippocampi, thalami, caudates, and amygdalae. The t-statistics for regions exhibiting significant sex differences in volume are shown in Figure 3a for the CC400 atlas. Again, the other atlases demonstrate similar patterns of volumetric sex differences and these results are shown in the Supplementary Materials.

ROC curves for volume-based sex classification for all atlases can be seen in Figure 3b, and accuracy and AUC can be found in Table 1. For volume-based sex classification, a mean accuracy of 85.4% (SD = 2.5%) and a mean AUC of 0.92 (SD = 0.023) was achieved using the CC400 atlas. Regions exhibiting the strongest feature importance were the cerebellum, cingulate cortex, and temporal cortex and a feature importance map is shown in Figure 3c. Results from other atlases identified similar regions as important features for sex classification and are shown in the Supplementary Materials. Feature importance for HE-based models and volume-based models were not correlated for any of the atlases (p-corrected >.05).

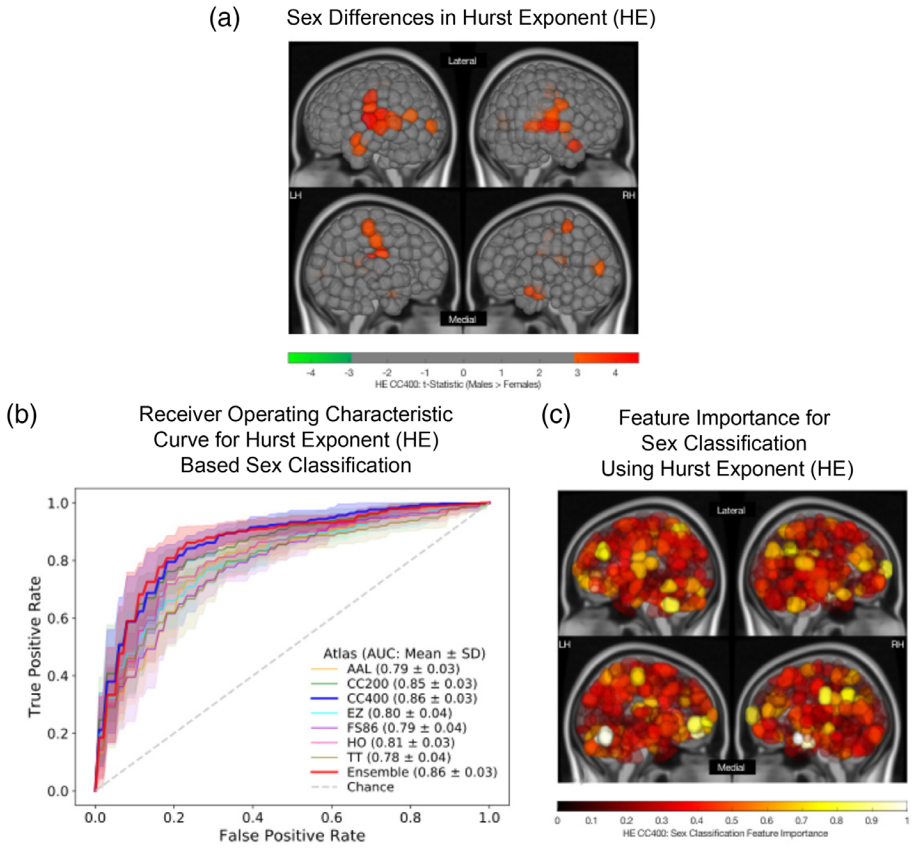

FIGURE 2 Sex differences and classification using regional Hurst exponent (HE). (a) Region-wise sex differences in HE for the CC400 atlas. Lateral (top) and medial (bottom) sides of the left (LH) and right (RH) hemispheres are shown. Regional t-statistics are shown as per the color scale for all significantly different (*p*-corrected <.05) areas. Nonsignificant areas are shown in grey. A positive t-statistic indicates that males have a higher mean value in that region than females. (b) Receiver operating characteristic curves for linear support vector machine (SVM) sex classification for all atlases based on HE. Mean and *SD* of the area under the curve (AUC) values for each atlas are indicated. (c) Feature importance map for a linear SVM classifier used to predict sex using HE computed on the CC400 atlas. Lateral (top) and medial (bottom) sides of the left (LH) and right (RH) hemispheres are shown. The absolute value of feature weights obtained from the linear SVM were scaled to generate normalized feature importance values as plotted per the color scale. Values closer to one indicate greater importance in the overall classification

TABLE 1 Sex classification results from grey-matter-volume-matched data set (n = 390)

	Hurst exponent		GM region volume		Ensemble predictions (Hurst exponent + GM region volume)	
Atlas (# of regions)	**Accuracy (%)**	**AUC**	**Accuracy (%)**	**AUC**	**Accuracy (%)**	**AUC**
FS86 (86)	72.95 ± 3.33	0.79 ± 0.04	72.18 ± 6.25	0.79 ± 0.05	75.77 ± 6.34	0.86 ± 0.04
TT (94)	72.95 ± 4.89	0.78 ± 0.04	73.08 ± 3.47	0.82 ± 0.03	78.59 ± 2.57	0.87 ± 0.03
HO (110)	74.23 ± 5.25	0.81 ± 0.03	73.08 ± 6.16	0.78 ± 0.05	77.69 ± 4.70	0.86 ± 0.04
AAL (116)	74.10 ± 2.08	0.79 ± 0.03	77.82 ± 3.31	0.86 ± 0.04	82.31 ± 4.28	0.89 ± 0.03
EZ (116)	74.49 ± 4.81	0.80 ± 0.04	78.08 ± 3.94	0.86 ± 0.04	82.56 ± 4.34	0.90 ± 0.03
CC200 (200)	77.56 ± 4.69	0.85 ± 0.04	81.92 ± 4.16	0.90 ± 0.03	86.15 ± 4.24	0.93 ± 0.03
CC400 (392)	**81.15 ± 3.53**	**0.87 ± 0.03**	**85.38 ± 2.51**	**0.92 ± 0.02**	**87.31 ± 3.22**	**0.94 ± 0.02**
Ensemble model (all atlases)	**81.15 ± 3.39**	0.86 ± 0.03	81.41 ± 4.84	0.91 ± 0.03	86.67 ± 4.22	0.93 ± 0.03

Note: Mean balanced accuracy ± *SD* and mean area under the curve (AUC) ± *SD* over 10 outer-loop permutations for each model are shown. Columns 2–3 show results for sex classification based on the Hurst exponent. Columns 4–5 show results for sex classification based on GM region volume. Columns 6–7 show results for sex classification based on ensemble models that combine prediction probabilities from Hurst exponent—and volume—based classification models across the same single atlas. The last row shows the results for sex classification based on an ensemble model that combines predictions from all seven atlases. Bold text indicates the model with the highest accuracy and AUC in each column

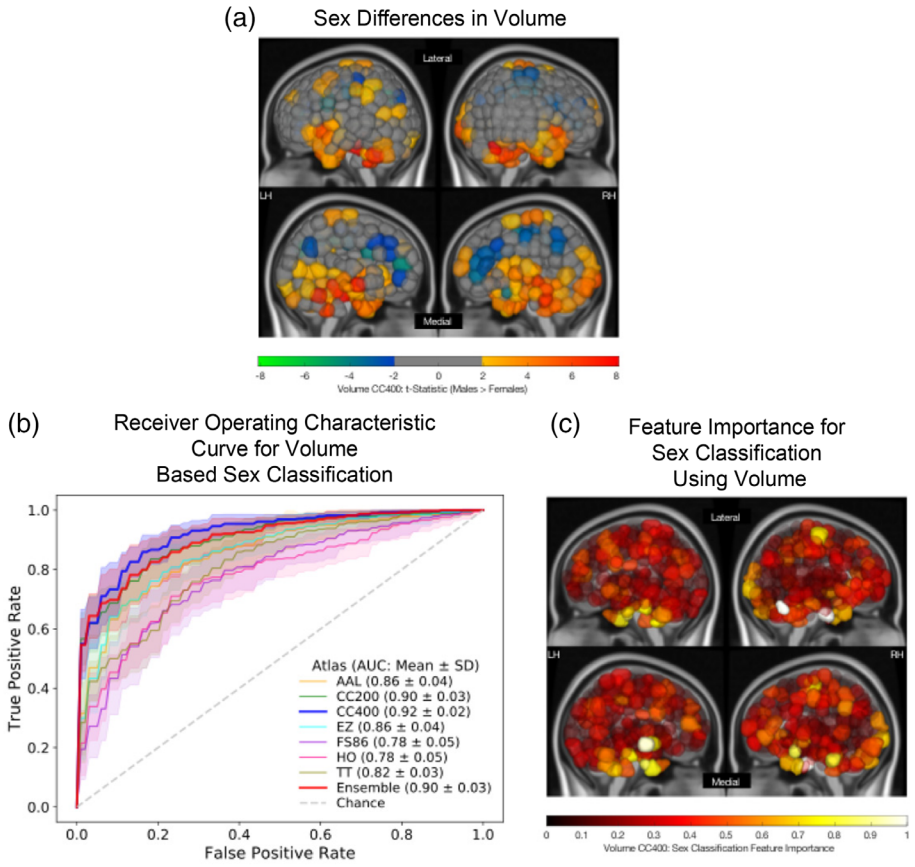

FIGURE 3 Sex differences and classification results using regional volume. (a) Region-wise sex differences in volume for the CC400 atlas. Lateral (top) and medial (bottom) sides of the left (LH) and right (RH) hemispheres are shown. Regional t-statistics are shown as per the color scale for all significantly different (*p*-corrected <.05) areas. Nonsignificant areas are shown in grey. A positive t-statistic indicates that males have a higher mean value in that region than females. (b) Receiver operating characteristic curves for linear support vector machine (SVM) sex classification for all atlases based on volume. Mean and *SD* of the area under the curve (AUC) values for each atlas are indicated. (c) Feature importance map for a linear support vector machine (SVM) classifier used to predict sex using volume of each region in the CC400 atlas. Lateral (top) and medial (bottom) sides of the left (LH) and right (RH) hemispheres are shown. The absolute value of feature weights obtained from the linear SVM were scaled to generate normalized feature importance values. Values closer to 1 indicate greater importance in the overall classification

3.3.1 | Ensemble models for sex classification

Atlas-specific ensemble models for sex classification were generated by combining the prediction probabilities from the atlas-specific HE- and volume- based classification models. Results from the ensemble models are shown in Table 1. The CC400 atlas' ensemble model performed with a mean accuracy of 87.3% (*SD* = 3.2%) and a mean AUC of 0.94 (*SD* = 0.015). The atlas-specific ensemble models significantly outperformed the HE-based models in terms of accuracy ($p < .05$) and AUC ($p < .01$). While these atlas-specific ensemble models also had a higher mean accuracy and AUC than the volume-based models, this difference was not significant ($p > .05$).

3.3.2 | Effect of volume-matching and sample size reduction on classification performance

Classification based on HE for entire data set (n = 1,003) performed with a mean balanced accuracy of 85.4% (*SD* = 2.1%) and mean AUC of 0.94 (*SD* = 0.012) for the CC400 atlas. Models trained on HE from the CC400 atlas for 10 randomly selected (not volume-matched) sample-size-matched subsets (n = 390) performed with a mean balanced accuracy of 84.0% (*SD* = 1.7%) and a mean AUC of 0.92 (*SD* = 0.012). A comparison of the balanced accuracy and AUC distributions obtained from sex classification models based on HE on the GM-volume-matched subset, random sample-size-matched subset, and the entire data set are shown in Figure 4. Mean balanced accuracies achieved by atlas-specific models trained with the GM-volume-matched subset were significantly worse than atlas-specific models trained with the entire data set ($p < .01$). Mean balanced accuracies of atlas-specific models trained with sample-size-matched subsets were not significantly different from atlas-specific models trained with the GM-volume-matched subset ($p > .05$) or atlas-specific models trained with the entire data set ($p > .05$). Mean AUCs achieved by atlas-specific models trained with the GM-volume-matched subset were significantly worse than atlas-specific models trained with the sample-size-matched subsets ($p < .001$) and atlas-specific models trained with the entire data set ($p < .001$). Mean AUCs of models trained on the

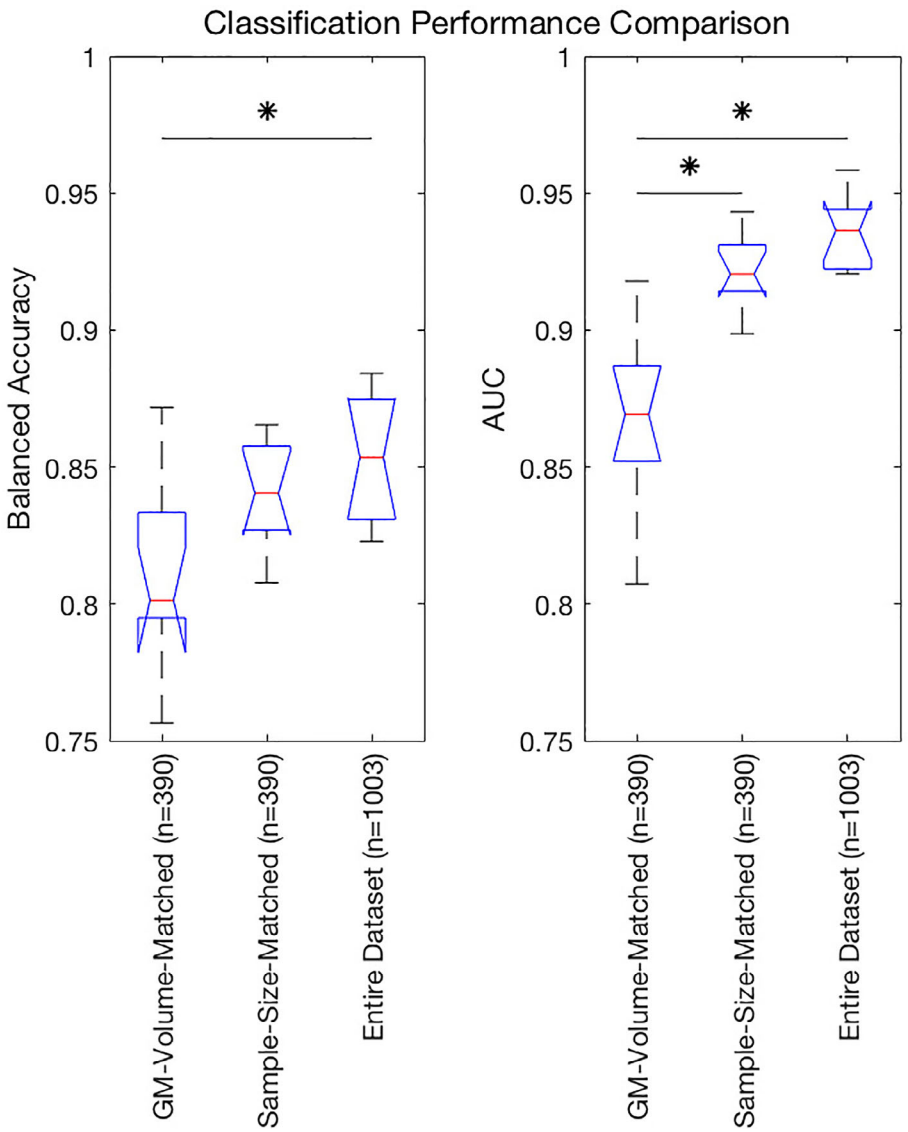

FIGURE 4 Balanced accuracy (left) and area under the curve (AUC) (right) distributions obtained from sex classification models based on HE on the grey-matter-volume-matched subset (n = 390), randomly selected sample-size-matched-subset (n = 390), and the entire data set (n = 1,003). Significant differences in means (p < .05) are denoted by *

sample-size-matched subsets and the entire data set were not significantly different (p > .05).

4 | DISCUSSION

Our findings reveal sex differences in long-range temporal dependence in healthy young adults, with males exhibiting more persistent long-range temporal dependence (higher HE) than females in the temporal, parietal, and occipital cortices. We performed sex classification based on regional temporal dependence using a linear SVM and achieved mean accuracy as high as 81.1% and mean AUC as high as 0.86. In our secondary analyses, we found that even when males and females are strictly matched on total GM volume, significant sex differences persist in regional volume across many cortical and subcortical regions. In fact, sex classification based on regional volume using a linear SVM and achieved mean accuracy of 85.3% and mean AUC of 0.92.

Male and female brains are similar in many respects but may be different in others (Cosgrove, Mazure, & Staley, 2007). Understanding the biology of male and female brain functionality in healthy individuals is crucial to elucidating mechanisms and determining more effective interventions for many neurodevelopmental, neurological, and psychiatric illnesses, which exhibit sex differences in prevalence, age of onset, and symptomatology (Giedd, Raznahan, Mills, & Lenroot, 2012). Both genetic and hormonal influences play a role in brain development, but any sex difference in neural structure may also be shaped through experience, practice, and neural plasticity (Cosgrove et al., 2007). In this study, sex differences were observed in the long-range temporal correlation of resting-state time series. However, the relationship between genetic information, hormonal measures, and HE was not evaluated in this study. In the absence of proof of genetic or hormonal influence underlying these observed sex differences in HE, it is important to note that any existing sex difference may have been shaped through experience, practice, or neural plasticity (Eliot, 2011).

Biological substrates driving HE are not well understood. Although sex differences at cellular and molecular levels of the nervous system have been observed (Cosgrove et al., 2007), we do not know whether it is gene expression, neuronal signaling, glial activity, anatomical structure, regional blood flow, hormone effects or other factors that are driving them. It is even less clear how these factors may contribute to sex differences in temporal dependence, as measured by HE, within the resting-state time series. As a result, it is difficult to interpret what these results may represent in a biological or behavioral domain. Previous work conjectured that sex differences in HE may be related to differences in behavioral and cognitive domains between males and females (Dong et al., 2018), however, no work yet has analyzed the relationship between behavior and/or cognitive domains and HE. It is important, however, to acknowledge that sex differences in cerebral blood flow have been observed at both a global (R. C. Gur & Gur, 2017; Song et al., 2019) and a regional (Rodriguez, Warkentin, Risberg, & Rosadini, 1988) level. Since the fMRI signal is a measure of the BOLD signal, it is possible that sex differences in HE are related to the sex differences in regional cerebral blood flow that would then influence the BOLD signal. Recent work has also investigated sex differences in low-frequency fluctuations (Yu et al., 2019), and eigenvector centrality dynamics (A. M. Wink, 2019) of resting-state fMRI. While these works do not attempt sex classification, they do report sex differences in these measures to varying degrees. Hence, additional properties of the temporal signal may also be contributing to the sex differences observed here.

While no previous studies, to our knowledge, have attempted sex classification using HE, one study did perform statistical comparisons of HE between males and females across a wide age range (Dong et al., 2018). They found HE was higher in females than males in the parietal lobe (Dong et al., 2018). Here, we report the opposite finding, that is, that HE is higher in males than females in many regions of the brain, including parietal lobes. The disagreement in our results and those of Dong et al. (2018) may be attributable to the demographic differences in the data sets used; their study had a much wider age range than the HCP data set and had a smaller sample size (N = 116). Other studies using the HCP data set in particular have performed sex classification using ROI-based resting-state functional connectivity measures (Weis, Patil, et al., 2020; Zhang et al., 2018). Those studies reported maximum test accuracies of 86.6% (AUC of 0.93) using whole-brain functional connectivity (Zhang et al., 2018) and 72.6% using region-specific functional connectivity (Weis, Patil, et al., 2020), which are comparable to the results reported here. This shows that while HE and functional connectivity represent different information about activation patterns in the brain, they share the ability to make accurate predictions about sex. HE is inherently lower in dimensionality, thus avoiding the curse of dimensionality, and can be mapped to a single specific region in the brain, making it potentially easier to interpret than functional connectivity that represents pairs of regions.

One key difference between our study and previous ones is that we strictly control for volume differences between the sexes. This is a major factor to consider when studying sex differences as there is inherent bias introduced by volumetric differences between males and females (Cosgrove et al., 2007; Ruben C Gur et al., 1999). In an initial analysis performed on the entire HCP data set (n = 1,003), we found that 725 regions (out of 1,114 across seven atlases) exhibited significant sex differences in HE. However, upon repeating the analysis on the grey-matter volume matched subset (n = 390) to eliminate the effects of volumetric differences, we found that only a subset of the originally identified regions were significantly different between males and females, suggesting that these differences must be accounted for when analyzing sex differences.

In terms of the influence of volumetric differences on sex classification, a previous study using functional connectivity to classify sex differences reported a decrease in performance in a GM-volume-matched subset compared with their entire data set (Weis, Patil, et al., 2020). However, they attributed the decrease in performance to the overall decrease in sample size. Our results show that the mean AUC for the sex classification task was significantly lower for models trained with the GM-volume-matched subset compared to both the entire data set and the randomly selected (not volume-matched) subsets of the same size. This suggests that, in our case, the decrease in AUC cannot be attributed to the decrease in sample size; rather, volume-matching reduced the classification performance of our fMRI-only based biomarker. Future studies that investigate sex differences in functional activation patterns must be aware that results may be inflated if they are not strictly volume-matching their female and male populations.

Studies often tend to use global volume matching strategies to reduce sex effects on volume. We show here that even after matching subjects on total GM volume, significant volumetric sex differences persist in many cortical and subcortical regions. Males had significantly larger volumes in the cerebellum, hippocampus, parahippocampus, thalamus, caudate, and amygdala while females had significantly larger volumes in the cingulate, precuneus, frontal cortex, and parietal cortex. This demonstrates that sex differences in regional brain volume exist even after matching subjects for GM volume. Finally, atlas-specific ensemble models that combined the prediction probabilities from the HE- and volume-based models had significantly better performance than HE based models and trends for better performance than volume-based models. This suggests that HE and volume measures are capturing distinct information that distinguishes males and females.

In this work, we repeated all of our analyses on sex differences and sex classification using seven different parcellation schemes. This was to ensure that sex differences observed in HE could be replicated with different atlases and were not subjected to biases introduced by one particular parcellation. The parcellations were used to segment the brain's GM into varying number of regions ranging from 86 to 392. For sex classification based on HE, we find that the classification performance, as measured by balanced accuracy, is significantly correlated with the number of regions (r = .98, p = .0001) in the atlas on which each model is based. In other words, as the number of regions increases, the classification performance also improves, suggesting that the algorithm may benefit from the increased precision of the parcellation, which also increases data dimensionality. While not

explicitly analyzed, a similar trend has been observed in previous work (Khosla, Jamison, Kuceyeski, et al., 2019). This observed correlation between number of regions and classification accuracy also holds for sex classification performed in our secondary analyses using regional volume (r = .88, p = .0094) and ensemble approaches (r = .79, p = .0331). These results suggest that the classification models benefit from the increased dimensionality of data. This effect of parcellation scheme might be important to consider in future work attempting to classify subjects based on neuroimaging data.

The choice of how to compute HE may also influence overall results. Here, we estimated HE using DFA applied to voxel-wise time series and then averaged HE across voxels to get each region's HE. An alternative approach is to calculate a single HE for each region using the average time series over each voxel in that region. All analyses presented in this article using HE computed on voxel-wise time series were repeated using HE computed on ROI time series and similar results were obtained for sex differences and sex classification with the CC400 atlas. These results are presented in detail in the Supplemental Materials, which largely showed similar results to what is presented in the main text.

4.1 | Limitations

Machine learning problems based on neuroimaging data are prone to the curse of dimensionality. Voxel-wise data are on the order of hundreds of thousands of features, and even when examining ROI-based data, there can be several hundred features. To avoid the curse of dimensionality and remove noise from the data, we used parcellations of the brain to generate a single measure per ROI for each subject. This resulted in the dimensionality being reduced from hundreds of thousands to between 86 and 392, depending on the atlas. However, by drastically reducing the overall dimensionality, information may have been lost and biases may have been introduced that limit the overall classification performance. In an attempt to mitigate possible biases introduced by atlas selection, the findings described in this article were evaluated over seven different parcellation schemes. However, to further reduce possible bias introduced by the use of atlases, a voxel-wise analysis of sex differences could be examined in future work. Finally, a single scaling parameter such as the HE is limited in its ability to characterize temporal features due to the complex nature of fMRI time series (Shimizu, Barth, Windischberger, Moser, & Thurner, 2004). An alternative to the HE is to use a whole set of fractal Hölder exponents instead which can also account for local intensity fluctuations (Shimizu et al., 2004).

Sex differences in functional brain activity may be due to an underlying hormonal effect (Jacobs & Goldstein, 2018; Pritschet et al., 2019). While various studies have shown that resting-state functional connectivity fluctuates across the menstrual cycle in women (De Bondt et al., 2015; Hjelmervik, Hausmann, Osnes, Westerhausen, & Specht, 2014; Weis, Hodgetts, & Hausmann, 2019), the effect of hormones on the HE has not yet been studied. Hormonal measures which would allow the study of the relationship between females' menstrual cycles and regional HE were not available in the data set used for this work. Future data sets should aim to collect hormonal levels such that a thorough investigation on the effect of menstrual cycle on HE can be examined. It is also important to note that hormonal fluctuations in regional HE may result in females exhibiting increased variability in those regions, as conjectured in (Weis, Patil, et al., 2020). Consequently, this could influence classification performance such that classification is more accurate in women during certain points of their cycles than others.

Studies examining sex differences often do not consider an individual's gender identity and fluidity. This study only used information about each subject's self-reported sex in the absence of gender identification. Males and females are exposed to different expected gender roles and a lifetime of gender-differentiated experience could be the underlying cause to sex differences in neuroimaging biomarkers (Eliot, 2011). These social factors may be partially responsible for the regional group differences in temporal dependence and overall classification performance reported in this study. Future data sets should aim to collect data about gender identity and fluidity for subjects and studies should incorporate this information into their work.

This study only used data from the HCP data set. Time series obtained from fMRI can be sensitive to scan parameters and preprocessing pipelines. Although test hold-out sets were used exclusively for evaluating the models generated for sex classification, the overall results may not be entirely generalizable to other data sets. A recent study used resting state functional connectivity data from HCP data set to make predictions about behavioral function (Li et al., 2019). The study identified that using functional connectivity with global signal regression improves prediction of a wide range of behavioral phenotypes compared to functional connectivity without global signal regression. It is therefore important to acknowledge that preprocessing steps used in the HCP data as well as in computation of HE for this study may have an influence on the overall results obtained.

5 | CONCLUSION

Understanding sex-specific brain differences in healthy individuals is a critical first step towards understanding sex-dependent variation in neurological, developmental, and psychiatric disorders and, possibly, the use of this information to develop personalized interventions. In this study, we observe that males exhibit higher temporal dependence of resting-state time series than females in the temporal, parietal, and occipital lobes. Furthermore, using ROI-based information about temporal dependence, we were successfully able to classify males and females using a linear SVM algorithm with a maximum mean accuracy of 81.1% and mean AUC of 0.86. We also identify that regional volume differences between males and females persist even after matching for cortical and subcortical grey-matter volume. Regional volume can also be used to successfully classify males and females with a maximum mean accuracy of 85.3% and mean AUC of 0.92. Finally, we demonstrate that

matching male and female groups for GM volume can decrease sex-classification accuracy of functionally-derived biomarkers, an issue that must be a carefully considered in future studies. Additional research is needed to understand the biological substrates underlying the observed sex differences in temporal dependence of resting-state fMRI.

6 | CITATION GENDER DIVERSITY STATEMENT

Recent work in neuroscience and other fields has identified a bias in citation practices such that papers from women and other minorities are under-cited relative to the number of such papers in the field (Caplar, Tacchella, & Birrer, 2017; Chakravartty, Kuo, Grubbs, & McIlwain, 2018; Dion, Sumner, & Mitchell, 2018; Dworkin et al., 2020; Maliniak, Powers, & Walter, 2013; Thiem, Sealey, Ferrer, Trott, & Kennison, 2018). Here we sought to proactively consider choosing references that reflect the diversity of the field in thought, form of contribution, gender, and other factors. We used classification of gender based on the first names of the first and last authors (Dworkin et al., 2020), with possible combinations including male/male, male/female, female/male, and female/female. Excluding self-citations to the first and last authors of our current article, the references contain 59.6% male/male, 12.8% male/female, 19.1% female/male, and 8.5% female/female. We look forward to future work that could help us to better understand how to support equitable practices in science.

ACKNOWLEDGMENTS

Data were provided and made available by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University. Data can be accessed at https://www.humanconnectome.org/study/hcp-young-adult/document/1200-subjects-data-release. This research was funded by NIH grants R01 NS102646-01A1 (AK), R21 NS104634-01 (AK), 1R21AG050122 (MRS), R01LM012719 (MRS), R01AG053949 (MRS); and, NSF CAREER 1748377 (MRS) and NSF NeuroNex Grant 1707312 (MRS). The authors would also like to acknowledge Clive Aaron D'Souza and Shashank Pathak from Cornell University for their contributions to an earlier version of this work.

CONFLICT OF INTEREST

All authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in Human Connectome Project at https://www.humanconnectome.org/study/hcp-young-adult/document/extensively-processed-fmri-data-documentation (Van Essen et al., 2013).

ORCID

Elvisha Dhamala https://orcid.org/0000-0002-8253-6962
Amy Kuceyeski https://orcid.org/0000-0002-5050-8342

REFERENCES

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate - A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, *57*(1), 289–300.

Caplar, N., Tacchella, S., & Birrer, S. (2017). Quantitative evaluation of gender bias in astronomical publications from citation counts. *Nature Astronomy*, *1*(6), 1–5.

Chakravartty, P., Kuo, R., Grubbs, V., & McIlwain, C. (2018). # CommunicationSoWhite. *Journal of Communication*, *68*(2), 254–266.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273–297. https://doi.org/10.1023/A:1022627411411

Cosgrove, K. P., Mazure, C. M., & Staley, J. K. (2007). Evolving knowledge of sex differences in brain structure, function, and chemistry. *Biological Psychiatry*, *62*(8), 847–855. https://doi.org/10.1016/j.biopsych.2007.03.001

Craddock, C. R., James, G. A., Holtzheimer, P. E., 3rd, Hu, X. P., & Mayberg, H. S. (2012). A whole brain fMRI atlas generated via spatially constrained spectral clustering. *Human Brain Mapping*, *33*(8), 1914–1928. https://doi.org/10.1002/hbm.21333

Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical surface-based analysis. I. segmentation and surface reconstruction. *NeuroImage*, *9*(2), 179–194. https://doi.org/10.1006/nimg.1998.0395

De Bondt, T., Smeets, D., Pullens, P., Van Hecke, W., Jacquemyn, Y., & Parizel, P. M. (2015). Stability of resting state networks in the female brain during hormonal changes and their relation to premenstrual symptoms. *Brain Research*, *1624*, 275–285. https://doi.org/10.1016/j.brainres.2015.07.045

Desikan, R. S., Segonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., … Killiany, R. J. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*, *31*(3), 968–980. https://doi.org/10.1016/j.neuroimage.2006.01.021

Dion, M. L., Sumner, J. L., & Mitchell, S. M. (2018). Gendered citation patterns across political science and social science methodology fields. *Political Analysis*, *26*(3), 312–327.

Dong, J., Jing, B., Ma, X., Liu, H., Mo, X., & Li, H. (2018). Hurst exponent analysis of resting-state fMRI signal complexity across the adult lifespan. *Frontiers in Neuroscience*, *12*, 34. https://doi.org/10.3389/fnins.2018.00034

Dworkin, J. D., Linn, K. A., Teich, E. G., Zurn, P., Shinohara, R. T., & Bassett, D. S. (2020). The extent and drivers of gender imbalance in neuroscience reference lists. *arXiv* preprint *arXiv:2001.01002*.

Eickhoff, S. B., Stephan, K. E., Mohlberg, H., Grefkes, C., Fink, G. R., Amunts, K., & Zilles, K. (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *NeuroImage*, *25*(4), 1325–1335. https://doi.org/10.1016/j.neuroimage.2004.12.034

Eliot, L. (2011). The trouble with sex differences. *Neuron*, *72*(6), 895–898. https://doi.org/10.1016/j.neuron.2011.12.001

Fischl, B., Liu, A., & Dale, A. M. (2001). Automated manifold surgery: Constructing geometrically accurate and topologically correct models of the human cerebral cortex. *IEEE Transactions on Medical Imaging*, *20*(1), 70–80. https://doi.org/10.1109/42.906426

Fischl, B., Rajendran, N., Busa, E., Augustinack, J., Hinds, O., Yeo, B. T., … Zilles, K. (2008). Cortical folding patterns and predicting cytoarchitecture. *Cerebral Cortex*, *18*(8), 1973–1980. https://doi.org/10.1093/cercor/bhm225

Fischl, B., Salat, D. H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., … Dale, A. M. (2002). Whole brain segmentation: Automated labeling of

neuroanatomical structures in the human brain. *Neuron*, *33*(3), 341–355.

Fischl, B., Sereno, M. I., & Dale, A. M. (1999). Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. *NeuroImage*, *9*(2), 195–207. https://doi.org/10.1006/nimg.1998.0396

Fischl, B., Sereno, M. I., Tootell, R. B., & Dale, A. M. (1999). High-resolution intersubject averaging and a coordinate system for the cortical surface. *Human Brain Mapping*, *8*(4), 272–284.

Giedd, J. N., Raznahan, A., Mills, K. L., & Lenroot, R. K. (2012). Review: Magnetic resonance imaging of male/female differences in human adolescent brain anatomy. *Biology of Sex Differences*, *3*(1), 19. https://doi.org/10.1186/2042-6410-3-19

Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., … WU-Minn HCP Consortium. (2013). The minimal preprocessing pipelines for the human connectome project. *NeuroImage*, *80*, 105–124. https://doi.org/10.1016/j.neuroimage.2013.04.127

Goldstein, J. M., Seidman, L. J., Makris, N., Ahern, T., O'Brien, L. M., Caviness, V. S., Jr., … Tsuang, M. T. (2007). Hypothalamic abnormalities in schizophrenia: Sex effects and genetic vulnerability. *Biological Psychiatry*, *61*(8), 935–945. https://doi.org/10.1016/j.biopsych.2006.06.027

Griffanti, L., Salimi-Khorshidi, G., Beckmann, C. F., Auerbach, E. J., Douaud, G., Sexton, C. E., … Smith, S. M. (2014). ICA-based artefact removal and accelerated fMRI acquisition for improved resting state network imaging. *NeuroImage*, *95*, 232–247. https://doi.org/10.1016/j.neuroimage.2014.03.034

Gur, R. C., & Gur, R. E. (2017). Complementarity of sex differences in brain and behavior: From laterality to multimodal neuroimaging. *Journal of Neuroscience Research*, *95*(1–2), 189–199. https://doi.org/10.1002/jnr.23830

Gur, R. C., Turetsky, B. I., Matsui, M., Yan, M., Bilker, W., Hughett, P., & Gur, R. E. (1999). Sex differences in brain gray and white matter in healthy young adults: Correlations with cognitive performance. *Journal of Neuroscience*, *19*(10), 4065–4072.

Hjelmervik, H., Hausmann, M., Osnes, B., Westerhausen, R., & Specht, K. (2014). Resting states are resting traits - An fMRI study of sex differences and menstrual cycle effects in resting state cognitive control networks. *PLoS ONE*, *9*(7), e103492. https://doi.org/10.1371/journal.pone.0103492

Jacobs, E. G., & Goldstein, J. M. (2018). The middle-aged brain: Biological sex and sex hormones shape memory circuitry. *Current Opinion in Behavioral Sciences*, *23*, 84–91. https://doi.org/10.1016/j.cobeha.2018.03.009

Khosla, M., Jamison, K., Kuceyeski, A., & Sabuncu, M. R. (2019). Ensemble learning with 3D convolutional neural networks for functional connectome-based prediction. *NeuroImage*, *199*, 651–662. https://doi.org/10.1016/j.neuroimage.2019.06.012

Khosla, M., Jamison, K., Ngo, G. H., Kuceyeski, A., & Sabuncu, M. R. (2019). Machine learning in resting-state fMRI analysis. *Magnetic Resonance Imaging*, *64*, 101–121. https://doi.org/10.1016/j.mri.2019.05.031

Lai, M. C., Lombardo, M. V., Chakrabarti, B., Sadek, S. A., Pasco, G., Wheelwright, S. J., … Consortium, M. A. (2010). A shift to randomness of brain oscillations in people with autism. *Biological Psychiatry*, *68*(12), 1092–1099. https://doi.org/10.1016/j.biopsych.2010.06.027

Lancaster, J. L., Woldorff, M. G., Parsons, L. M., Liotti, M., Freitas, C. S., Rainey, L., … Fox, P. T. (2000). Automated Talairach atlas labels for functional brain mapping. *Human Brain Mapping*, *10*(3), 120–131.

Li, J., Kong, R., Liegeois, R., Orban, C., Tan, Y., Sun, N., … Yeo, T. (2019). Global signal regression strengthens association between resting-state functional connectivity and behavior. *NeuroImage*, *196*, 126–141. https://doi.org/10.1016/j.neuroimage.2019.04.016

Makris, N., Goldstein, J. M., Kennedy, D., Hodge, S. M., Caviness, V. S., Faraone, S. V., … Seidman, L. J. (2006). Decreased volume of left and total anterior insular lobule in schizophrenia. *Schizophrenia Research*, *83*(2–3), 155–171. https://doi.org/10.1016/j.schres.2005.11.020

Maliniak, D., Powers, R., & Walter, B. F. (2013). The gender citation gap in international relations. *International Organization*, *67*(4), 889–922.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., … Duchesnay, E. (2011). Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Power, J. D., Mitra, A., Laumann, T. O., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2014). Methods to detect, characterize, and remove motion artifact in resting state fMRI. *NeuroImage*, *84*, 320–341. https://doi.org/10.1016/j.neuroimage.2013.08.048

Pritschet, L., Santander, T., Layher, E., Taylor, C. M., Yu, S., Miller, M. B., … Jacobs, E. G. (2019). Functional reorganization of brain networks across the human menstrual cycle. *bioRxiv*, 866913.

Rodriguez, G., Warkentin, S., Risberg, J., & Rosadini, G. (1988). Sex differences in regional cerebral blood flow. *Journal of Cerebral Blood Flow and Metabolism*, *8*(6), 783–789. https://doi.org/10.1038/jcbfm.1988.133

Salimi-Khorshidi, G., Douaud, G., Beckmann, C. F., Glasser, M. F., Griffanti, L., & Smith, S. M. (2014). Automatic denoising of functional MRI data: Combining independent component analysis and hierarchical fusion of classifiers. *NeuroImage*, *90*, 449–468. https://doi.org/10.1016/j.neuroimage.2013.11.046

Segonne, F., Grimson, E., & Fischl, B. (2005). A genetic algorithm for the topology correction of cortical surfaces. *Information processing in medical imaging*, *19*, 393–405.

Shimizu, Y., Barth, M., Windischberger, C., Moser, E., & Thurner, S. (2004). Wavelet-based multifractal analysis of fMRI time series. *NeuroImage*, *22*(3), 1195–1202. https://doi.org/10.1016/j.neuroimage.2004.03.007

Smith, S. M., Beckmann, C. F., Andersson, J., Auerbach, E. J., Bijsterbosch, J., Douaud, G., … Consortium, W. U.-M. H. (2013). Resting-state fMRI in the human connectome project. *NeuroImage*, *80*, 144–168. https://doi.org/10.1016/j.neuroimage.2013.05.039

Song, D., Chang, D., Zhang, J., Ge, Q., Zang, Y.-F., & Wang, Z. (2019). Associations of brain entropy (BEN) to cerebral blood flow and fractional amplitude of low-frequency fluctuations in the resting brain. *Brain Imaging and Behavior*, *13*(5), 1486–1495.

Tagliazucchi, E., Chialvo, D. R., Siniatchkin, M., Amico, E., Brichant, J. F., Bonhomme, V., … Laureys, S. (2016). Large-scale signatures of unconsciousness are consistent with a departure from critical dynamics. *Journal of the Royal Society Interface*, *13*(114), 20151027. https://doi.org/10.1098/rsif.2015.1027

Thiem, Y., Sealey, K. F., Ferrer, A. E., Trott, A. M., & Kennison, R. (2018). Just Ideas? The Status and Future of Publication Ethics in Philosophy: A White Paper

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., … Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*, *15*(1), 273–289. https://doi.org/10.1006/nimg.2001.0978

Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E., Yacoub, E., Ugurbil, K., & Consortium, W. U.-M. H. (2013). The WU-Minn human connectome project: An overview. *NeuroImage*, *80*, 62–79. https://doi.org/10.1016/j.neuroimage.2013.05.041

Wei, M. B., Qin, J. L., Yan, R., Li, H. R., Yao, Z. J., & Lu, Q. (2013). Identifying major depressive disorder using Hurst exponent of resting-state brain networks. *Psychiatry Research: Neuroimaging*, *214*(3), 306–312. https://doi.org/10.1016/j.pscychresns.2013.09.008

Weis, S., Hodgetts, S., & Hausmann, M. (2019). Sex differences and menstrual cycle effects in cognitive and sensory resting state networks. *Brain and Cognition*, *131*, 66–73. https://doi.org/10.1016/j.bandc.2017.09.003

Weis, S., Patil, K. R., Hoffstaedter, F., Nostro, A., Yeo, B. T. T., & Eickhoff, S. B. (2020). Sex classification by resting state brain

connectivity. *Cerebral Cortex*, *30*(2), 824–835. https://doi.org/10.1093/cercor/bhz129

Wink, A. M. (2019). Eigenvector centrality dynamics from resting-state fMRI: Gender and age differences in healthy subjects. *Frontiers in Neuroscience*, *13*(648). https://doi.org/10.3389/fnins.2019.00648

Wink, A. M., Bernard, F., Salvador, R., Bullmore, E., & Suckling, J. (2006). Age and cholinergic effects on hemodynamics and functional coherence of human hippocampus. *Neurobiology of Aging*, *27*(10), 1395–1404. https://doi.org/10.1016/j.neurobiolaging.2005.08.011

Wink, A. M., Bullmore, E., Barnes, A., Bernard, F., & Suckling, J. (2008). Monofractal and multifractal dynamics of low frequency endogenous brain oscillations in functional MRI. *Human Brain Mapping*, *29*(7), 791–801.

Yu, Y., Sun, Y., Qian, Z., Tao, L., Wei, G., Shi, Q., … Yang, Y. (2019). Amplitude of low-frequency fluctuation differences between the brains of young men and young women: A resting-state functional magnetic resonance imaging study. *Journal of Medical Imaging and Health Informatics*, *9*(4), 807–812.

Zhang, C., Dougherty, C. C., Baum, S. A., White, T., & Michael, A. M. (2018). Functional connectivity predicts gender: Evidence for gender differences in resting brain connectivity. *Human Brain Mapping*, *39*(4), 1765–1776. https://doi.org/10.1002/hbm.23950

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Dhamala E, Jamison KW, Sabuncu MR, Kuceyeski A. Sex classification using long-range temporal dependence of resting-state functional MRI time series. *Hum Brain Mapp*. 2020;41:3567–3579. https://doi.org/10.1002/hbm.25030