



Automatika

Journal for Control, Measurement, Electronics, Computing and Communications

ISSN: (Print) (Online) Journal homepage: <https://www.tandfonline.com/loi/taut20>

Classification and feature analysis of the Human Connectome Project dataset for differentiating between males and females

Jelena Bozek, Ivan Kesedzic, Leonard Novosel & Tomislav Bozek

To cite this article: Jelena Bozek, Ivan Kesedzic, Leonard Novosel & Tomislav Bozek (2021) Classification and feature analysis of the Human Connectome Project dataset for differentiating between males and females, *Automatika*, 62:1, 109-117, DOI: [10.1080/00051144.2021.1885890](https://doi.org/10.1080/00051144.2021.1885890)

To link to this article: <https://doi.org/10.1080/00051144.2021.1885890>



© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 09 Feb 2021.



Submit your article to this journal [↗](#)



Article views: 254



View related articles [↗](#)



View Crossmark data [↗](#)



Classification and feature analysis of the Human Connectome Project dataset for differentiating between males and females

Jelena Bozek ^a, Ivan Kesedzic^a, Leonard Novosel^a and Tomislav Bozek^{b,c}

^aFaculty of Electrical Engineering and Computing, University of Zagreb, Zagreb, Croatia; ^bSchool of Medicine, University of Zagreb, Zagreb, Croatia; ^cClinical Hospital Merkur, University Clinic Vuk Vrhovac, Zagreb, Croatia

ABSTRACT

We analysed features relevant for differentiation between males and females based on the data available from the Human Connectome Project (HCP) S1200 dataset. We used 354 features containing cognitive and emotional measures as well as measures derived from task functional magnetic resonance imaging (MRI) and structural brain MRI. The paper presents a thorough analysis of this extensive set of features using a machine learning approach with a goal of identifying features that have the ability to differentiate between males and females. We used two state of the art classification algorithms with different properties: support vector machine (SVM) and random forest classifier (RFC). For each classifier the hyperparameters were obtained and classifiers were optimized using nested cross validation and grid search. This resulted in the classification performance of 91% and 89% accuracy using SVM and RFC, respectively. Using SHAP (SHapley Additive exPlanations) method we obtained relevance of features as indicators of sex differences and identified features with high discriminative power for sex classification. The majority of top features were brain morphological measures, and only a small proportion were features related to cognitive performance. Our results demonstrate the importance and advantages of using a machine learning approach when analysing sex differences.

ARTICLE HISTORY

Received 4 February 2020
Accepted 13 June 2020

KEYWORDS

sex differences; features; classification; HCP; brain MRI; machine learning

1. Introduction

Differences between male and female brains have been the research focus of many neurological studies. Sex difference analyses include various combinations of different data types such as different magnetic resonance imaging (MRI) modalities (structural, functional, diffusion), as well as different behavioural, cognitive, psychological and other various non-imaging sex-related measures. Some studies analysed psychological, anatomical and structural differences and correlations [1], while others analysed structural and functional differences [2–4]. A common way of performing such analyses is multivariate and statistical analysis, and assessment of correlations between measures of variables. Recently, other methods have been introduced, such as machine learning approaches.

Tunc et al. [5] used multivariate analysis of brain networks and their connection with behavioural sex differences and showed an increasing separation between males and females in behavioural patterns and in brain structure. Szalkai et al. [1] analysed Human Connectome Project's 500-subject dataset using a maximum spanning tree method. They created braingraphs and analysed the pairwise correlations between 717 psychological, anatomical and structural connectome properties. They detected numerous natural correlations,

which describe parameters computable or approximable from one another, but have also found several significant, novel correlations in the dataset. They found that men, on average, have greater grip strength, brain volume and are physically more aggressive. Ritchie et al. [2] analysed structural and functional sex differences from a large sample comprising 5216 older adult participants (age range 44–77 years) from the UK Biobank project. They mapped sex differences in brain volume, surface area, cortical thickness, diffusion parameters and functional connectivity. Their results showed that males had higher raw volumes, raw surface areas and white matter fractional anisotropy, while females had higher raw cortical thickness and higher white matter tract complexity. Lotze et al. [4] performed statistical analysis and examined grey matter (GM) volume of 2838 adult brains (age range 21–90 years) to assess sex differences. Their findings revealed that females had more GM volume in medial and lateral prefrontal areas, the superior temporal sulcus, the posterior insula and orbitofrontal cortex, and that male brains had more GM volume in subcortical temporal structures (such as the amygdala and hippocampus), temporal pole, fusiform gyrus, visual primary cortex and motor areas (premotor cortex, putamen, anterior cerebellum).

CONTACT Jelena Bozek jelena.bozek@fer.hr Faculty of Electrical Engineering and Computing, University of Zagreb, Unska 3, 10000 Zagreb, Croatia

The usage of machine learning approaches and classification algorithms has an expanding role in sex differentiation or predicting cognitive outcomes. Joel et al. [3] used an anomaly detection algorithm to test whether the brain types typical of one sex category are also typical of the other sex category. This was complemented by unsupervised clustering to find clusters that best describe variability in a population of human brains regardless of sex category. Their study showed that it is possible to use one's brain architecture to predict whether this person is female or male with an accuracy of 80%, although one's sex category provides very little information on the likelihood that one's brain architecture is similar to or different from someone else's brain architecture. Luo et al. [6] used multivariate pattern analysis of cortical three-dimensional (3D) morphology and derived discriminative morphological features to identify gender effectively. Xin et al. [7] applied deep learning on the diffusion MRI data to analyse morphological differences between men and women. They used 3D convolutional neural network (3D CNN) on the maps of fractional anisotropy (FA) and compared it with support vector machine (SVM) and tract-based spatial statistics methods. Their proposed 3D CNN yielded a better classification result (93.3%) than the SVM (78.2%) on the whole-brain FA images. Their results indicate gender-related differences in the whole brain as well as in several specific brain regions. Van Putten et al. [8] used a deep convolutional neural network to predict sex from brain rhythms obtained in EEG (electroencephalogram) data and achieved an accuracy of 81%. Azevedo et al. [9] applied machine learning approach for predicting individual differences in cognitive functioning by using features derived from brain surface-based morphometry and cortical myelin estimates. They reduced the number of features into 23 sets using Factor Analysis, out of which they extracted nine factors that represented 70% of cumulative variance using Principal Component Analysis. They used nested cross validation and XGBoost (Extreme Gradient Boosting) method to predict the factors and applied SHAP (SHapley Additive exPlanations) analysis for interpreting predictions. Their results revealed that the prediction of the sex-related factor yielded the best Pearson-r correlation values compared to the other eight factors that comprised different cognitive measures.

The aim of our study was to analyse anatomical as well as behavioural factors as indicators of sex differences and the ability of the machine learning approaches to separate between male and female participants based on the used features. With that goal in mind we have applied two state of the art classifiers, support vector machine and random forest classifier, to detect underlying patterns and subtle information. The purpose of the applied classification algorithms and subsequent analysis of features was to identify features

relevant for discriminating between males and females. We performed the analysis of feature relevance using SHAP analysis. Results revealed features that had strong distinguishing power between males and females.

The significance of this work is that it uses a big dataset of young adults and a broad set of different features, from neuroimaging measures, to psychological and cognitive measures. Further, it uses state of the art machine learning approaches to analyse feature relevance for sex differentiation. Our results, that agree with statistical findings of sex differences, show the utility and power of the applied machine learning approach.

2. Materials and methods

2.1. Materials

Data used in this study were obtained from the S1200 public dataset released as part of the Human Connectome Project (HCP)¹ [10]. After accounting for missing information, the total number of subjects included in our analysis was 863, where 410 were males and 453 were females. Age distribution of male and female participants is shown in Table 1.

A total number of the analysed features was 354, as summarized below, and additional details can be found in the HCP Data Dictionary² and in Barch et al. [11].

Features can be separated into non-imaging and imaging features. Non-imaging features include the following:

- Health factor included sleep quality measured using the Pittsburgh Sleep Quality Index (PSQI) [12].
- Cognitive measures included NIH Toolbox tests [13] as well as non-toolbox tests [14,15] that sampled a broad spectrum of domains, and final composite scores.
- Motor skills were tested using NIH Toolbox endurance, locomotion, dexterity and strength tests.
- Sensory abilities were tested using NIH Toolbox (audition, olfaction, pain, taste) and Electronic Visual Acuity and Mars contrast tests (vision).
- Emotional processing measures included the Penn Emotion Recognition test and several NIH Toolbox surveys.
- Tasks during the fMRI experiment included emotion, gambling, language, relational, social and working memory tasks. Variables that capture

Table 1. Age distribution of male and female participants.

Age range	Male	Female
22–25	124	67
26–30	173	197
31–35	110	184
36+	3	5
Total	410	453

reaction times, difficulty level of stimuli and various accuracy measures were derived for each task in the scanner.

- Personality trait scores (agreeableness, openness, conscientiousness, neuroticism, extraversion) were assessed using the Five Factor Inventory (NEO-FFI) [16].

Listed NIH Toolbox measures were reported both age-adjusted (participant score was normed using the age appropriate band of Toolbox Norming Sample) and age-unadjusted (participant score was normed to those in the entire NIH Toolbox Normative Sample).

Imaging features included FreeSurfer (FS) derived structural measures of the brain comprising cortical and subcortical volumes, cortical surface thickness and area across different brain regions. These were obtained from the FreeSurfer part [17] of the minimal processing pipelines developed for the analysis of the HCP data [18].

2.2. Methods

The goal of the analysis was to identify features that have discriminative power for classifying male and female participants. We applied two state of the art classification algorithms with different properties from the scikit-learn package [19]:

- (1) Support Vector Machine (SVM) [20] with radial basis function (RBF) kernel finds a hyperplane that helps to classify the data points. The dimension of the hyperplane depends on the number of features N and results in a $N-1$ dimensional space. The position and the orientation of the hyperplane are defined by support vectors which are the data points closer to the hyperplane.
- (2) Random Forest Classifier (RFC) creates a set of decision tree classifiers from randomly selected subsets [21]. Each tree in the ensemble is built from a sample drawn with replacement (i.e. a bootstrap sample) from the training set. In contrast with the original method where each classifier votes for a single class, the used scikit-learn implementation combines classifiers by averaging their probabilistic prediction to decide the final class.

Before applying classification, the variables were first standardized by removing the mean and by scaling to unit variance. This reduces the chances for classifiers to be overly biased by variables that show the most variance in the original data.

We have split the dataset into two equal parts: (i) development set, which was further split into training and testing sets, that were used for optimizing the classifier models using grid search, and (ii) an evaluation set (comprising held-out samples that were not seen

during the optimization) for reporting final classifier performance and obtaining the feature rankings.

Each classifier was optimized using grid search for finding the best set of hyperparameters relevant for the classifier with a goal of achieving the best cross validation (CV) score. GridSearchCV in scikit-learn package exhaustively considers all parameter combinations. Hyperparameters that were optimized for SVM classification were regularizer C and a kernel coefficient γ . For RFC we tuned number of trees in random forest, number of features to consider at every split, maximum number of levels in tree and criterion that measures the quality of each split.

In order to avoid over-fitting of the model and to reduce the subsequent selection bias [22], we used a nested CV with 4 folds to do hyperparameter tuning and to assess the performance of the best model. The bias is reduced by passing only training set from the outer loop to the inner loop, while the testing set in the outer loop is held back. The performance of the model is assessed in the outer loop, and the selection of the best model is done in the inner loop. The model is selected on each outer-training set (using the inner CV loop) and its performance was measured on the corresponding outer-testing set. In the inner loop (executed by GridSearchCV), the score was approximately maximized by fitting a model to each training set, and then directly maximized in selecting hyperparameters over the testing set. In the outer loop (executed using `cross_val_score`), generalization error was estimated by averaging test set scores over dataset splits.

Classification results and the performance of each classifier were analysed on the evaluation set in terms of accuracy:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \quad (1)$$

where TP (true positive) is the number of positive classes correctly predicted (or classified) as positive, TN (true negative) is the number of negative classes correctly classified as negative, FP (false positive) is the number of negative classes misclassified as positive, and FN (false negative) is the number of positive classes misclassified as negative. Positive and negative are generic names for the predicted classes. Accuracy is often expressed as a percentage value obtained by multiplying the above expression with 100.

Additional measures that were used for assessing the performance include precision, recall and F1-score. These are defined as follows:

$$\text{Precision} = \frac{TP}{TP + FP}, \quad (2)$$

$$\text{Recall} = \frac{TP}{TP + FN}, \quad (3)$$

$$\text{F1 - score} = 2 \frac{\text{Recall Precision}}{\text{Recall} + \text{Precision}}. \quad (4)$$

In other words, precision gives a rate of positive predictions, and recall gives a fraction of correctly classified positives. F1-score combines the two measures and is their weighted average, with 1 being the best score, and 0 being the worst.

For determining the impact of a feature on the output of the model, i.e. the feature importance for the data separation task, we used SHAP method [23]. It is based on the game theory and the estimation of Shapley values [24]. SHAP was used for explaining and analysing individual predictions by computing the contribution of each feature to the prediction. KernelSHAP [23], which is a kernel-based estimation approach for Shapley values, was used for analysing SVM outputs. TreeSHAP [25], which is an efficient implementation for tree-based models, was used for analysing RFC outputs. TreeSHAP is a fast method and computes exact Shapley values, while KernelSHAP is computationally more expensive and only approximates the actual Shapley values.

3. Results

Classification accuracies of SVM and RFC are 91.44% and 89.35%, respectively. These results show that the features used for the classification have good discriminative power for classification between males and females. Precision, recall and F1-score are listed in the classification report shown in Table 2, separately for the two classes (males and females) and as the final overall score expressed through weighted average. Comparing precision and recall over two classes may be obsolete since it only shows that a classifier has better ability to find one class compared to the other. F1-score, which combines precision and recall, shows that both classifiers are slightly better in predicting females than males, 0.5% better in case of SVM and 1% better in case of RFC. However, in the final result that small difference has disappeared. Overall, classifier performance expressed in terms of weighted averages of precision, recall and F1-score are all approximately 0.91 for SVM and 0.89 for RFC. This indicates that each classification model and our dataset are balanced, i.e. that classifier's ability to correctly classify males is equivalent to its ability to correctly classify females.

Summary plot [23] of SHAP values of every feature for every sample gives an overview of features that are most important for the classifier. It sorts features

by the sum of SHAP value magnitudes over all samples, and uses SHAP values to show the distribution of the impacts each feature has on the model output. The colour represents the feature value (red is high, blue is low). Due to space limitations we plotted only top 20 features. Full feature names and explanations are available in the HCP Data Dictionary. In summary, feature names starting with FS are FreeSurfer measures of volume (Vol), cortical thickness (Thck) and surface area (Area) for different brain regions for left (L) and right (R) hemisphere. Top 20 features for classifying males and females obtained with SVM and RFC are presented in separate summary plots for each sex in Figures 1 and 2, respectively. In the summary plots, it can be noticed that the results for males and females are mirrored for a particular feature. High values (red) of a certain feature cause higher predictions for males, and low values (blue) cause low predictions for males, while at the same time for females low values of the same feature cause high predictions and high values cause low predictions. For example, higher values of grip strength show higher prediction for a male, and lower prediction for a female.

Mean absolute SHAP values for top 20 features for SVM and RFC classifiers are presented in summary plots with stacked bars for both sexes in Figure 3. The overlap between those top 20 features for SVM and RFC, based on SHAP values, is 6 features, or 30%, while for top 177 features (50% of all features) the overlap is 52%, or 92 features. However, a closer look at plots in Figure 1, with plots for SVM output for male and female SHAP values, shows that the relevance is not identical for male and female SHAP values either, and that the overlap between the two is different. This discrepancy may be in the way KernelSHAP (used for analysing SVM outputs) computes the resulting SHAP values, which are only an approximation of Shapley values, while TreeSHAP (used for RFC outputs) computes the exact Shapley values. Thus, for the final feature ranking we computed arithmetic mean of the feature importance rank across the two classifiers and presented them, along with the standard deviation (SD), in Table 3. Positions are rounded to the closest integer value. Due to space limitations, we included only the top 50 features.

The top three features, which achieved the same ranking using both SVM and RFC, are the grip strength (age-adjusted and unadjusted) and FreeSurfer total brain mask volume. Among the other top 50 relevant features are mostly imaging derived features from FreeSurfer (37 out of 50), with the majority being volumetric measures: total GM volume, cortical volume of cerebellum, volume of the third ventricle, total cortical GM volume, left and right cortical GM volume, etc. It is interesting to note that volumes of subcortical structures are not bilaterally included in the top 50 features. An example is the amygdala, where only the volume of

Table 2. Classification report for SVM and RFC.

	SVM			RFC		
	Precision	Recall	F1-score	Precision	Recall	F1-score
Males	0.923	0.901	0.911	0.915	0.864	0.889
Females	0.906	0.927	0.916	0.874	0.922	0.898
Weighted average	0.915	0.914	0.914	0.895	0.894	0.893

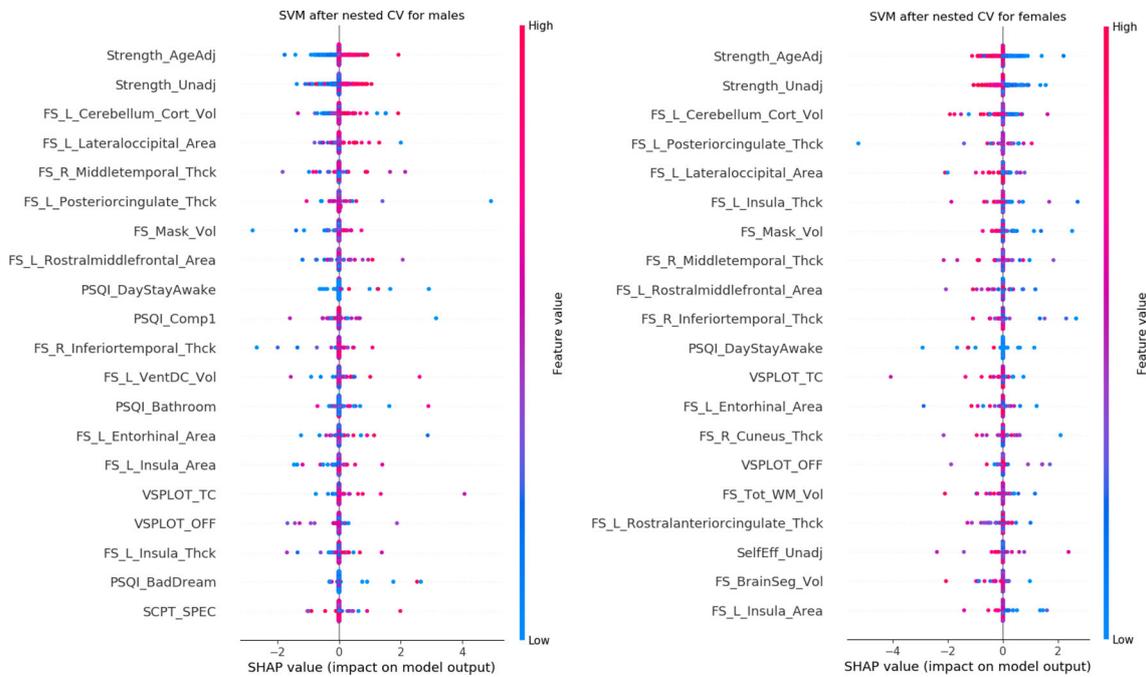


Figure 1. Summary plot of SHAP values for classifying males and females using SVM for the top 20 features.

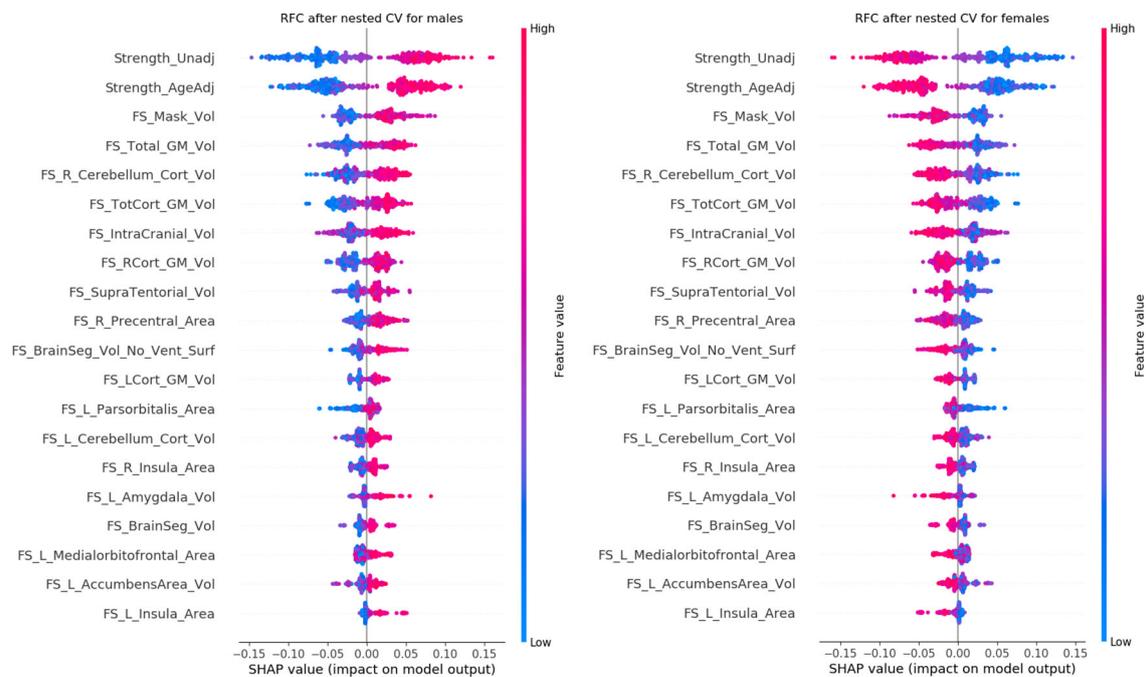


Figure 2. Summary plot of SHAP values for classifying males and females using RFC for the top 20 features.

the left one is among top 50 with its mean position of 23 (± 10.58 SD), while the volume of right amygdala has the mean position of 163 (± 57.27 SD). Also, the volume of left hippocampus is among the top 50 features, with its mean position being 25 (± 8.49 SD), while the volume of right hippocampus has the mean position of 153 (± 106.77 SD).

Further FreeSurfer features among the top 50 are surface areas of various regions, including right lateral occipital region, left posterior cingulate, right and left precentral and left postcentral, left entorhinal, right

inferiorparietal and left fusiform region. Surface area of both left and right insula is also among the top 50 features.

The following FreeSurfer derived measures of cortical thickness are in the top 50 relevant features: thickness of right precuneus region, left medial orbitofrontal, right superior temporal, right temporal pole, left cuneus and right parstriangularis.

Aside already mentioned grip strength, which is on the top position, other non-imaging features among the top 50 relevant features are: in-scanner task

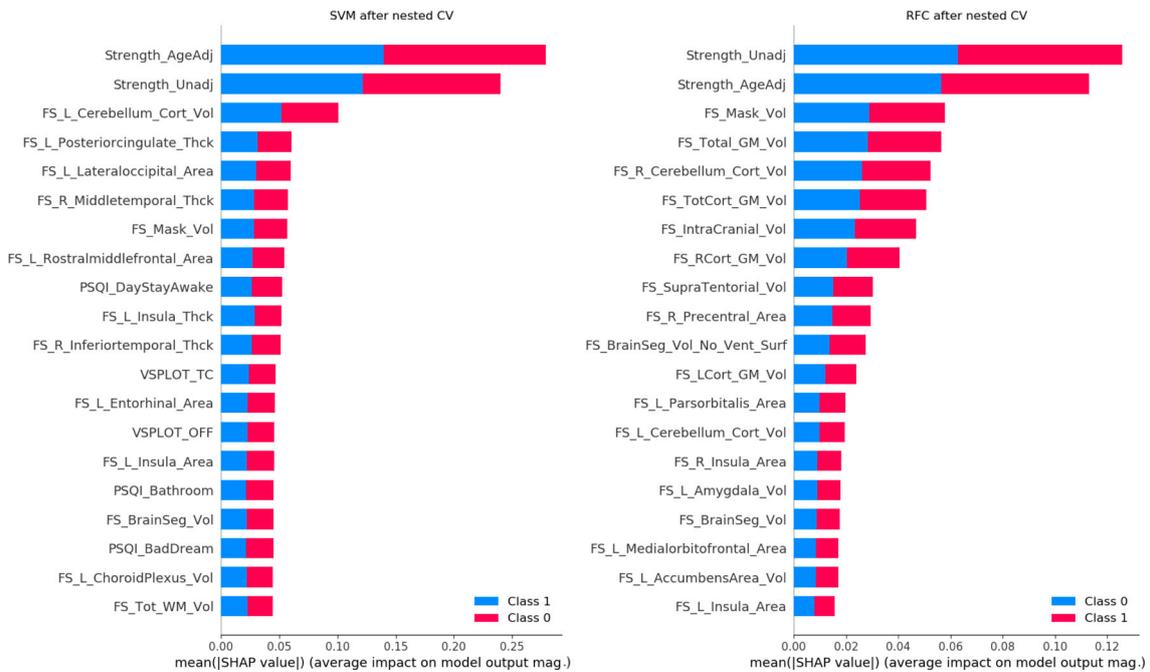


Figure 3. Mean absolute value of the SHAP values stacked for both males (Class 0) and females (Class 1) for SVM and RFC for the top 20 features. (mag. = magnitude).

Table 3. Mean positions and standard deviations (SD) across SVM and RFC of the top 50 features.

Feature	Mean position	SD	Feature	Mean position	SD
Strength_Unadj	1	0	FS_R_Precentral_Area	57	67
Strength_AgeAdj	2	0	Flanker_Unadj	57	61
FS_Mask_Vol	3	0	SCPT_TN	57	3
FS_Total_GM_Vol	10	8	FS_BrainSeg_Vol	58	59
FS_L_Cerebellum_Cort_Vol	11	4	FS_L_Entorhinal_Area	59	53
FS_R_Cerebellum_Cort_Vol	12	10	FS_R_Cerebellum_WM_Vol	59	13
FS_3rdVent_Vol	22	6	FS_BrainStem_Vol	60	52
FS_L_Amygdala_Vol	23	11	FS_L_Postcentral_Area	62	80
FS_LCort_GM_Vol	25	18	Language_Task_Story_Median_RT	64	53
FS_R_Lateraloccipital_Area	25	30	FS_L_Precentral_Area	65	30
FS_L_Hippo_Vol	27	8	FS_R_Superiortemporal_Thck	67	18
FS_L_Medialorbitofrontal_Area	31	18	FS_OpticChiasm_Vol	69	62
FS_BrainSeg_Vol_No_Vent_Surf	32	30	FS_R_Temporalpole_Thck	69	1
Relational_Task_Rel_Median_RT	34	38	FS_L_Cuneus_Thck	70	81
SCPT_SPEC	35	13	FS_R_Inferioparietal_Area	72	5
FS_L_Lateraloccipital_Area	39	9	NEOFAC_A	73	74
FS_TotCort_GM_Vol	44	54	Emotion_Task_Median_RT	74	18
FS_SupraTentorial_Vol	47	54	VSPLOT_OFF	74	50
FS_R_Precuneus_Thck	51	28	FS_R_Parstriangularis_Thck	76	35
FS_RCort_GM_Vol	52	62	FS_CSF_Vol	77	66
FS_IntraCranial_Vol	55	68	FS_R_Insula_Area	77	88
FS_L_Posteriorcingulate_Area	55	39	GaitSpeed_Comp	77	15
FS_L_Cerebellum_WM_Vol	56	20	FS_L_Fusiform_Area	79	71
FS_L_Insula_Area	56	52	NEOFAC_N	82	23

performance including median reaction time for correct trials during relational blocks in relational task (Relational Task Rel Median RT), and during story condition in language task (Language Task Story Median RT), and average of median reaction times from each condition in emotion task (Emotion Task Median RT); cognitive measures including specificity (SCPT SPEC) and true negatives (SCPT TN) of the sustained attention from the Short Penn Continuous Performance Test, Flanker measure of executive function and spatial orientation (VSPLOT OFF); personality traits agreeableness (NEOFAC A) and neuroticism (NEOFAC N); motor skills that included locomotion (GaitSpeed Comp); and alertness measure regarding sleep

habits of having a bed partner or roommate (PSQI BedPtnrRmate).

4. Discussion

This paper presents a machine learning approach to identify psychological and brain anatomical features and their relevance for distinguishing between males and females. We used two classifiers with different properties (SVM and RFC) to obtain relevance of 354 features and identify features with high discriminative power for determining sex on the set of 863 subjects from the publicly available HCP dataset. Hyperparameters for each classifier were optimized through a nested

cross validation and grid search with a goal of achieving optimal performance and reducing classification bias. The resulting models were used to identify features that had the highest contribution in the classification task.

High accuracy of the used classifiers (91% for SVM and 89% for RFC) shows that the features used for the classification have good discriminative power for classification between males and females. It is important to note that the performance was high due to the combination of features, and single features would likely perform worse. However, the purpose of this study has not been the selection of the best individual feature or subsets of features, but to use classifiers to identify features relevant for discriminating between males and females and obtain their ranking. Applying standard procedures for feature selection prior to the classification, such as e.g. recursive feature elimination, would reduce the number of features and retain only the best set of features for a certain classifier. Feature selection would identify which features (and their combinations) contribute the most in class prediction. However, here we wanted to assess the individual top features and their contribution to the classification, while retaining all features during the process.

Another point worth noting is that SVM outperformed RFC. This may be due to the demands of the given task being a two class problem. Namely, RFC is suited for multiclass problems, while SVM is intrinsically a two-class problem classifier.

Our classification analysis revealed that non-imaging features with strong discriminative power among the top 50 features are grip strength, reaction times during tasks in the scanner, cognitive measures of sustained attention, executive function and spatial orientation, personality traits agreeableness and neuroticism, locomotion and alertness measure. Many FreeSurfer measures derived from structural MR images were revealed as relevant features (37 among the top 50) for distinguishing between males and females, including volumetric, thickness and surface area measures.

Our results of top relevant features agree with findings from various statistical analysis that show sex differences in the regions that our classifiers and SHAP analysis revealed as top ranking features. Analysis by Szalkai and Grolmusz [26] showed that females have significantly larger numerous subcortical areas and most cortical areas. Further, analysis by Ritchie et al. [2] revealed that males generally had larger volumes and surface areas of the brain, whereas females had thicker cortices. They also found that volume and surface area mediated nearly all of the small sex difference in reasoning ability, but far less of the difference in reaction time. Results by Szalkai et al. [1] showed that men, on average, have greater grip strength and brain volume.

Beside a benefit of classification algorithms in helping identifying features relevant for discriminating sex, they also reduce the possible drawbacks of statistical

analysis. Namely, an issue of statistical analysis of bigger datasets has been brought out by Smith and Nichols [27], where caution is needed since a bigger sample size renders significant individual associations meaningless. However, where many variables are considered simultaneously, a reasonable total percentage of variance prediction may be found. Furthermore, a big dataset may show high sensitivity to artifactual associations due to confounding effects, and even when a real association exists, confounds can bias the estimate of the correlation. To tackle such issues of big data in neuroimaging, Smith and Nichols [27] propose multivariate analysis as well as rigorous cross validation based on held-out data. Thus, our classification approach is in line with the stated demands and issues, where the proportion of data was held-out for the final evaluation and feature ranking, and models were trained using a nested CV.

Lotze et al. [4] pointed out that assessing only volumes of brain structures, i.e. voxel-based morphometry may not include global or local changes of vertex-based measures (cortical thickness, surface area) in different directions (e.g. increase of cortical thickness, decrease of surface area). Thus, a study, such as ours, that includes and combines vertex-based measures such as cortical thickness and surface area along with the voxel-based measurements of volumes of brain structures has an increased value.

To our knowledge, our study is the first one that included cognitive, emotional as well as structural brain measures for the analysis of important features for distinguishing between males and females and applied machine learning approach for that purpose. Our results show the importance of using a machine learning approach when analysing sex differences, and the approach may well be expanded to other hypotheses designs in future studies. Future work should extend the analysis of relevant features by training separate classifiers for the cognitive and imaging measures, to determine the importance of each in differentiating sex and assessing the relevance of each, both separately and jointly. Finally, our results verify the findings from various statistical analyses of sex differences and thereby demonstrate the power and advantages of the applied machine learning approach.

Notes

1. <https://www.humanconnectome.org/>
2. <https://wiki.humanconnectome.org/display/PublicData/HCP+Data+Dictionary+Public+Updated+for+the+1200+Subject+Release>

Acknowledgements

JB has been supported through the funding from the European Regional Development Fund under the grant KK.01.1.1.01.0009 (DATACROSS). IK has been supported through the funding from the Croatian Science Foundation. Data were provided in part 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.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

JB has been supported through the funding from the European Regional Development Fund under the grant [KK.01.1.1.01.0009] (DATACROSS). IK has been supported through the funding from the Croatian Science Foundation.

Data availability

Publicly available datasets were analysed in this study. This data can be accessed at <https://www.humanconnectome.org/>. The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Compliance with ethical standards

This article does not contain any studies with human participants or animals performed by any of the authors.

ORCID

Jelena Bozek  <http://orcid.org/0000-0002-5986-5442>

References

- [1] Szalkai B, Varga B, Grolmusz V. Mapping correlations of psychological and structural connectome properties of the dataset of the human connectome project with the maximum spanning tree method. *Brain Imaging Behav.* August 2018; 1–8. ISSN 1931-7565. doi:10.1007/s11682-018-9937-6.
- [2] Ritchie SJ, Cox SR, Shen X, et al. Sex differences in the adult human brain: evidence from 5216 UK Biobank participants. *Cereb Cortex.* August 2018;28(8): 2959–2975.
- [3] Joel D, Persico A, Salhov M, et al. Analysis of human brain structure reveals that the brain types typical of males are also typical of females, and vice versa. *Front Hum Neurosci.* 2018;12:399, ISSN 1662-5161. doi:10.3389/fnhum.2018.00399.
- [4] Lotze M, Domin M, Gerlach FH, et al. Novel findings from 2,838 adult brains on sex differences in gray matter brain volume. *Sci Rep.* February 2019;9:1–7.
- [5] Tunc B, Solmaz B, Parker D, et al. Establishing a link between sex-related differences in the structural connectome and behaviour. *Philos Trans R Soc Lond, B, Biol Sci.* 2016;371(1688):20150111.
- [6] Luo Z, Hou C, Wang L, et al. Gender identification of human cortical 3-D morphology using hierarchical sparsity. *Front Hum Neurosci.* 2019;13:29.
- [7] Xin J, Zhang Y, Tang Y, et al. Brain differences between men and women: evidence from deep learning. *Front Neurosci.* 2019;13:185.
- [8] Van Putten MJAM, Olbrich S, Arns M. Predicting sex from brain rhythms with deep learning. *Scientific Reports.* February 2018;8(3069):1–7.
- [9] Azevedo T, Passamonti L, Lio P, et al. A machine learning tool for interpreting differences in cognition using brain features. *bioRxiv.* 2019; doi:10.1101/558403.
- [10] Van Essen DC, Smith SM, Barch DM, et al. The WU-Minn Human Connectome Project: an overview. *NeuroImage.* 2013;80:62–79. ISSN 1053-8119. Mapping the Connectome.
- [11] Barch DM, Burgess GC, Harms MP, et al. Function in the human connectome: task-fMRI and individual differences in behavior. *NeuroImage.* 2013;80:169–189. Mapping the Connectome.
- [12] Buysse DJ, Reynolds CF, Monk TH, et al. The Pittsburgh sleep quality index: a new instrument for psychiatric practice and research. *Psychiatry Res.* 1989;28(2):193–213. ISSN 0165-1781.
- [13] Gershon RC, Wagster MV, Hendrie HC, et al. NIH toolbox for assessment of neurological and behavioral function. *Neurology.* 2013;80(11 Supplement 3):S2–S6. doi:10.1212/WNL.0b013e3182872e5f.
- [14] Gur R, Ragland J, Moberg P, et al. Computerized neurocognitive scanning: I. Methodology and validation in healthy people. *Neuropsychopharmacology.* 2001;25(5):766–776.
- [15] Gur R, Richard J, Hughett P, et al. A cognitive neuroscience-based computerized battery for efficient measurement of individual differences: standardization and initial construct validation. *J Neurosci Methods.* 2010;187(2):254–262.
- [16] McCrae RR, Costa PT. A contemplated revision of the neo five-factor inventory. *Pers Individ Differ.* 2004;36(3):587–596. ISSN 0191-8869. doi:10.1016/S0191-8869(03)00118-1.
- [17] Fischl B. Freesurfer. *Neuroimage.* August 2012;62: 774–781.
- [18] Glasser MF, Sotiropoulos SN, Wilson JA, et al. The minimal preprocessing pipelines for the Human Connectome Project. *NeuroImage.* 2013;80:105124.
- [19] Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res.* 2011;12:2825–2830.
- [20] Chang C-C, Lin C-J. LIBSVM: a library for support vector machines. *ACM Trans Intell Syst Technol.* April 2011;3.
- [21] Breiman L. Random forests. *Mach Learn.* 2001;45: 5–32.
- [22] Cawley GC, Talbot NLC. On over-fitting in model selection and subsequent selection bias in performance evaluation. *J Mach Learn Res.* 2010;11:2079–2107.
- [23] Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. In: I Guyon, UV Luxburg, S Bengio, H Wallach, R Fergus, S Vishwanathan, R Garnett, editors. *Advances in neural information processing systems 30.* Curran Associates, Inc.; 2017. p. 4765–4774. URL <http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf>.
- [24] Shapley LS. A value for n-person games. *Contrib Theory Games.* 1953;2(28):307–317.

- [25] Lundberg SM, Erion G, Chen H, et al. From local explanations to global understanding with explainable ai for trees. *Nat Mach Intell.* 2020;2(1):2522–5839.
- [26] Szalkai B, Grolmusz V. Human sexual dimorphism of the relative cerebral area volumes in the data of the human connectome project. *Eur J Anat.* 2018;22(3): 221–225. ISSN 2340-311X.
- [27] Smith SM, Nichols TE. Statistical challenges in big data human neuroimaging. *Neuron.* 2018;97(2):263–268. ISSN 0896-6273.