

Representative Scale-invariant Characteristics of Male and Female Brains in Magnetic Resonance Images

Matthew Toews^{1*}, Talía Vázquez Romaguera¹,
William Wells III^{2,3} and Nikos Makris²

^{1*}Ecole de Technologie Supérieure, Montreal, Quebec, Canada.

²Harvard Medical School, Boston, MA, USA.

³Massachusetts Institute of Technology, Boston, MA, USA.

*Corresponding author(s). E-mail(s): matt.toews@gmail.com;
Contributing authors: talia.vazquez-romaguera.1@ens.etsmtl.ca;
sandywells@gmail.com; nikos@cma.mgh.harvard.edu;

Abstract

This paper investigates the link between sex and the human brain from anatomical MRI data, where a primary confound is the size difference between male and female groups. Anatomy is characterized by the 3D scale-invariant feature transform (SIFT), where features are salient image regions that are automatically identified and normalized according local size or scale. Experiments use T1-w MRI data of 422 healthy unrelated subjects from the Human Connectome Project (HCP) dataset (211 males, 211 females, 22-36 years of age). We found that brain sex may be predicted via image-to-image feature matching with 91.9% accuracy, that classification is driven largely by weakly-informative features distributed throughout the brain and shared by unique subsets of subjects, and that a pair of features identified in the right and left thalamic regions of 97% of subjects predicts sex with 74% accuracy. Misclassified subjects exhibit features typical of the opposite sex in one or both hemispheres.

Keywords: Human brain, anatomical MRI, sex differences

1 Introduction

Biological sex is a critical variable in neuroscience impacting both normal gender-related behaviour [1] and mental health disorders [2] including the variable prevalence of autism spectrum disorder[3] and schizophrenia[4]. Do male and female brains differ in terms of anatomy, and if so how? On the one hand, relatively subtle sex-related neuroanatomical differences are well-known, for example the interthalamic adhesion (ITA), aka massa intermedia, tends to be larger and more prevalent in women than men [5–7]. However, size is a major confound, as the male anatomy is on average larger than the female, and a recent survey suggested that few meaningful differences persist between male and female brains after correcting for size differences [8]. Deep network methods are able to discriminate between male and female brain MRIs with high accuracy even following normalization for size differences [9, 10], however these may be difficult to interpret due to the black box nature of deep network processing [11] and it is unclear which structures might be most representative of male and female brains [8–10]. While overall classification accuracy is of interest, our the primary goal of this study is to understand the neuroanatomical features driving sex classification independently from known sex-related size differences [8].

This paper reports on representative characteristics of male and female human brains identified using the 3D scale-invariant feature transform (SIFT) method [12, 13]. In contrast to typical analyses focusing on standard neuroanatomical structures of interest, the SIFT method is based on generic salient regions of interest that may not be shared by all subjects but that when present may be detected in a manner invariant to the image scale or resolution. The classic 2D SIFT method is widely used in the field of computer vision to identify point-to-point matches between different photographs of the same scene or object despite differences in size or image resolution [14]. Similarly, in our previous work we have found the 3D SIFT method to be effective in 3D brain MRI tasks, e.g. identifying pairs of family members [15] and repeated subject scans in large public datasets[16]. High sex classification accuracy (AUC=95) was reported using 3D SIFT [15] with data from Human Connectome Project (HCP) dataset [17]. These findings led us to hypothesize that sex classification was being driven by informative local features throughout the brain, and classification accuracy values were reproduced here to validate the current independent study into the nature of these features.

The remainder of this paper investigates sex classification of T1-w human brain MRIs using 3D SIFT features, where classification is driven by image-to-image feature matches identify the salient anatomical regions of interest. We note that in the context of this work, sex refers to self-reported binary categorization as either male or female, as provided by subjects in the HCP dataset [17]. Gender refers to socially constructed roles, behaviors and identities related to women, men and gender-diverse individuals, and is not studied here.

2 Methods

We investigate the sex classification of anatomical brain MRI data based on nearest-neighbour SIFT feature matching, as described in [15], including the overall classification accuracy and the interpretation of matches driving classification. Classification is based on a continuous sex prediction score estimated from nearest-neighbour feature matches between a subject and a set of labelled training subjects stored in memory [18]. Intuitively, feature matches indicate distinctive anatomical similarities between subjects, and the similarity between a subject and male or female groups is quantified by their proportions of shared feature matches using the Jaccard index.

Our analysis represents each image as a set of salient keypoints, which are identified via the 3D scale-invariant feature transform (SIFT) method [12, 13]. Each SIFT feature is an image region detected at a location (x, y, z) and scale σ that locally maximizes the spherically symmetric Laplacian-of-Gaussian operator:

$$\nabla^2 I(x, y, z, \sigma) = \nabla^2 * G(\sigma) * I(x, y, z), \quad (1)$$

where in Equation (1), ∇^2 is the Laplacian operator, $G(\sigma)$ is the Gaussian kernel of standard deviation σ , $I(x, y, z)$ is the image, and $*$ is the convolution operator. Once detected, the local image content surrounding (x, y, z) is normalized by scale σ to a standard resolution and encoded as a scale-invariant appearance descriptor from image gradient orientation information. The SIFT method was designed originally to identify corresponding keypoints in different photographs of the same scene, despite geometrical deformations due to changes in perspective and resolution [14], and was used to classify sex and interpret sex differences from human face photographs acquired from arbitrary viewpoints [19]. The 3D SIFT-Rank method extended scale invariance to volumetric image data [12] and enabled a number of applications, including our own work in chest CT classification [20], ultrasound registration for image-guided neurosurgery [21], and whole-body MRI segmentation [22], and in other work robust mid-sagittal plane detection [23], deformable registration evaluation [24] and ultrasound liver image stitching [25]. The strength of the SIFT method lies in the capacity to identify informative local image patterns in a manner invariant to translations, rotations and scalings of the image, with minimal image pre-processing. In studies of the human brain from anatomical MRI, it was used to identify discriminative features related to Alzheimer's disease [26], and proved highly accurate in identifying images of family members and same-subjects from large MRI datasets [15, 16] in comparison to neural networks [27].

With features extracted, a prediction score is derived from feature set overlap as follows. The similarity two feature sets $A = \{a_i\}$ and $B = \{a_i\}$ may be estimated via the Jaccard index or intersection-over-union $J(A, B)$:

$$J(A, B) = |A \cap B| / (|A| + |B| - |A \cap B|), \quad (2)$$

where in Equation (2), $|A \cap B|$ is the cardinality of the set intersection $A \cap B$ of features common to both A and B . The Jaccard index lies on the range $J(A, B) \in [0, 1]$, where $J(A, B) = 1$ indicates perfect overlap or similarity. Equivalence between a pair of features $a_i \in A$ and $b_i \in B$ is defined as according to the Euclidean distance $d(a_i, b_j) = \|a_i - b_j\|$ between their appearance descriptors. The exponential function $\exp^{-d^2(a_i, b_j)} \in [0, 1]$ provides a soft or non-binary measure of feature equivalence, accounting for noise and inter-subject variability, and is summed over feature matches to compute $|A \cap B|$. The larger the number of matches between sets (A, B) , the greater the Jaccard index. Our prediction score is based on the Jaccard indices $J(A, Female)$ and $J(A, Male)$ between a test subject feature set A and feature sets pooled from *Male* and *Female* subject MRIs. Noting that the logarithm transforms the Jaccard overlap to a distance measure $-\log J(A, B) \in [0, \infty]$, we define our prediction score as the difference of Jaccard distances $\log J(A, Female) - \log J(A, Male) \in \mathbb{R}^1$ from subject feature set A to either group,

$$Score(A) = \log J(A, Female) - \log J(A, Male), \quad (3)$$

where $Score(A)$ generally ranges from positive or negative depending on whether A is more representative of the *Male* or *Female* group, respectively.

Experiments analyzed 3D T1-weighted MRI brain scans from the human connectome project (HCP) dataset [17], aged 22-36 years (mean 29 years) from 434 families. A balanced set of 211 male and 211 female subjects were randomly selected from 422 different families, thereby removing potential family-related bias. Sex labels are determined based on self-reporting, HCP subjects self-reporting as female and providing complete menstrual history records (ten fields including menstrual irregularity, birth control usage, etc.) are considered female, others are considered males. Pre-processing steps such as registration or segmentation are generally not required for SIFT classification, however skull-stripped MRI data are used to restrict analysis to brain tissues.

Our analysis considers 3D SIFT features extracted from MRI volumes that are spatially registered to the MNI152 brain coordinate reference space [28] according to standard HCP processing pipelines [17], including rigid registration [29] which preserves brain size and shape, and non-linear registration which does not [30]. We compare classification from data following rigid and non-linear registration, in order to observe the effect of normalization on sex classification. We also consider classification with or without geometrical information, i.e. feature (x, y, z) centroid location and scale, where geometrically inconsistent feature matches are down-weighted in computing the Jaccard overlap and prediction score. We expect accuracy to be highest in the case of rigid registration, where size and shape information may play a role, and in the case where feature geometry information is included.

3 Results

Results here first establish baseline classification accuracy, then interpret classification in terms of prediction score and sex-informative 3D SIFT features. Table 1 summarizes the result of classification and feature extraction. As expected, accuracy is highest in the case of rigid registration and where feature geometry information is included (0.933) and lowest in the case of non-linear registration and no geometry information (0.832). In the case of non-linear registration, accuracy is approximately 8% lower than for rigid registration and feature counts are approximately equal. This is consistent with expectation, i.e. as non-linear registration removes size and shape differences between subjects, extracted features are more homogeneous and less informative. We note that in the case of rigid registration, male images contain 16% more features per image at a given resolution, these arise from fine-scaled image information only observable in images of larger brains which are typically male. They represent spurious noisy features that do not match systematically to either sex, and removing the excess of $292 - 252 = 40$ smallest-scaled features from male sets does not affect classification. We note that the SIFT-based classification results from rigidly registered data here are consistent with those previously reported [15], and that results from non-linearly registered data are new to this work.

Table 1 Sex classification of MRI brain features, given registration algorithm and inclusion of geometry information.

Registration	Accuracy		Features per Image	
	w/o Geometry	w Geometry	Female	Male
Rigid	91.9 ± 0.7	93.3 ± 0.0	252 ± 42	292 ± 42
Non-linear	83.2 ± 0.3	85.1 ± 0.5	315 ± 33	314 ± 33

Subjects may be interpreted lying along a prediction score continuum as in Figure 1, with subjects sorted from most female (left, orange) to most male (right, blue). Note the close linear fit with $R^2 = 0.98$, in comparison subjects distributed equally according to a binary-valued score would form a discontinuous step function with squared correlation of $R^2 = 0.75$. Visual examples of features (circles) associated with correct and incorrect classification are shown below. Circle colour indicates the affinity of each feature f in terms of the log likelihood ratio $\log \frac{p(f|Male)}{p(f|Female)}$, which ranges from female (orange) to neutral (white) to male (blue). In general each brain exhibits features associated with either male or female affinity, and it is their combination which determines classification [15]. Two larger features centred approximately in the right and left thalami are visually consistent with correct classification, as indicated by arrows. These are automatically identified in the right and left thalamic regions of 97% of subjects, and thus represent informative common structure shared by most of the population.

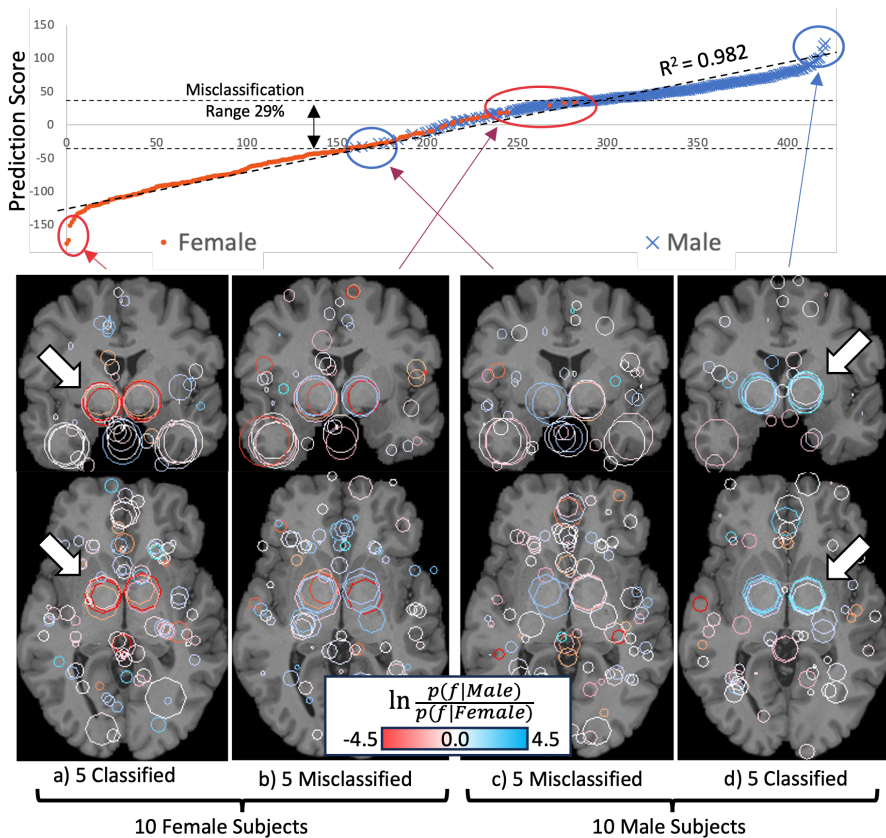


Fig. 1 The top graph plots 422 test subjects sorted according to female (left) vs male (right) prediction score, the value of $R^2 = 0.982$ indicates a near-linear continuum. The four lower figures display sex-informative features as circles in coronal (upper) and axial (lower) slices for the 5 strongest classifications (outer pair) and misclassifications (inner pair). Feature colour indicates the log likelihood of male (blue) vs female (orange) labels associated with individual features f , white indicates equal likelihood. Two features centred approximately upon the left and right thalami (arrows) are visibly consistent with correct classification.

Subjects may also be interpreted as clustering into prediction score distributions $p(\text{Score}|\text{Female})$ and $p(\text{Score}|\text{Male})$ conditioned on $\{\text{Male}, \text{Female}\}$ labels as in Figure 2, note that these exhibit a range of overlap associated with misclassification. Images below show examples of features associated with correctly classified subjects (a, d) and misclassified subjects (b, c). The upper and lower image rows show features that are either correctly classified or misclassified for the associated subject group. Note that misclassified subjects may exhibit thalamic features that are absent or consistent with the opposite sex, as indicated by arrows in Figure 2 lower row. Note that the distributions appear slightly shifted to the left, e.g. the optimal classification threshold at which distributions cross is slightly lower than 0. This is due to

the additional spurious male features, these increase the size of the male vs the female feature set $|Male| > |Female|$, which decreases the Jaccard index in Equation (2) for the male group.

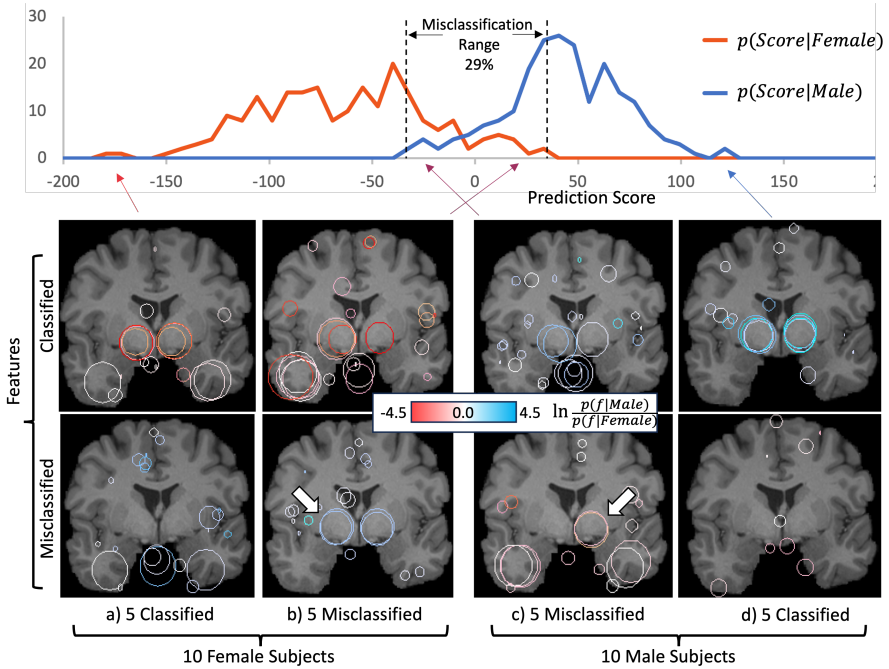


Fig. 2 The top graph shows distributions of prediction score conditioned on female and male labels, note that misclassifications occur within an overlap range containing 29% of subjects (dashed lines). The four lower figures display sex-informative features as circles in coronal slices for the 5 strongest classifications a) and d), and strongest misclassifications b) and c). Feature colour indicates the log likelihood of male (blue) vs female (orange) labels associated with individual features, white indicates equal likelihood. The upper and lower rows display features classified or misclassified, respectively. Note that some misclassified subjects (lower row) b) and c) exhibit thalamic features consistent with the opposite sex.

Finally, the geometry of the thalamic feature pair as identified in correctly classified subjects may be visualized together in Figure 3 for 5 males (blue) and 5 females (orange). Note that female features (orange) are located in closer proximity, consistent with an interthalamic adhesion more frequently observed in female brains [5–7]. These two features predict sex at an accuracy of 0.69 individually and 0.74 when combined as a pair.

4 Discussion

Size is a primary confound in identifying meaningful anatomical differences between male and female brains in MRI data, as the male anatomy is on

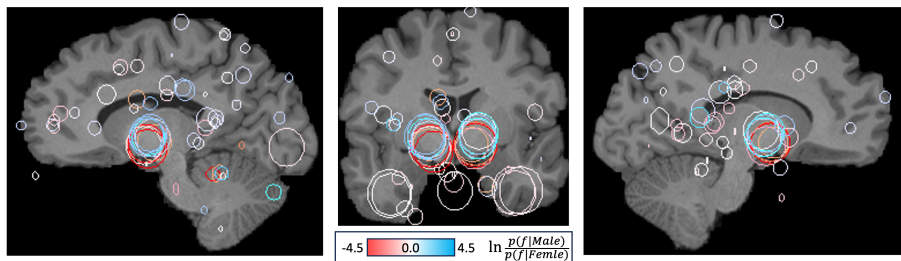


Fig. 3 Visualizing thalamic feature variation for the five most strongly classified females and males over a representative brain MRI. Circles indicate feature location and scale, and colour indicates the log likelihood of male (blue) vs female (orange) associated with individual features. The scale difference (11.6 ± 1.46 , 11.9 ± 1.34) between male and female features is statistically significant ($p=0.0402$), with a female/male size ratio of 0.968, similar to the ratio 0.970 estimated from image-to-image registration. A notable shift in z coordinate is observed due to rigid misregistration and size, this may be used to classify with 0.722 accuracy.

average larger than the female. Here we present the result of a unique feature-based analysis that recently reported high sex classification accuracy using 3D SIFT image feature matching [15]. In 3D MRI data, SIFT features represent maximally symmetric image patterns, e.g. blob-like regions of bulging image contrast, where size or scale is determined automatically by the Laplacian-of-Gaussian detector [12] and used to compute image descriptors that are invariant to the local scale of the anatomy. Furthermore, classification is achieved via feature matching, which requires no explicit training procedure and allows the result to be interpreted in terms of matches between specific anatomical regions of specific individual subjects.

Baseline classification accuracy is first established from spatially normalized brain volumes from the HCP dataset [17], with normalization achieved via rigid and non-linear image registration. Rigid registration preserves scale and shape differences, and accuracy from scale-invariant descriptors alone is 0.919, which represents an upper limit of accuracy given no scale or location information. Accuracy rises to 0.933 if geometry information i.e. feature location and scale are included, however this may be incorporating size-related information. Classifying non-linearly registered images is a more challenging context as scale and overall shape differences are removed. Accuracy following non-linear registration is 0.832 from invariant features alone, rising to 0.851 when geometry information is included. 0.851 thus represents a lower limit of accuracy given no scale or shape information, since assuming accurate registration, geometry information serves solely to disambiguate incorrect feature matches.

Local SIFT features driving classification may be interpreted as modes of anatomy that do not occur identically in all subjects, but with a probability across a population and within groups. For example in previous work, group-informative features were identified about the hippocampi and ventricles for Alzheimer’s disease patients [26], and distributed throughout the brain and

cortices for family members [16]. Here in the case of sex classification, the most notable group informative features are a pair located symmetrically in the left and right thalamic regions. Together they predict sex with accuracy 0.74, and as they are automatically detected in 97% of subjects, they represent informative common structure shared by most subjects. This finding is consistent with prior literature, where the thalamus has been linked to sex in the human brain [5–7] including neural networks and volume-normalized MRI data [9, 10], and also in brain MRI of other species such as mice [31]. However, overall 0.919 classification appears to be driven largely by features scattered throughout the brain, a result also found in logistic regression and neural network analysis methods [9, 10]. These are associated with feature matches shared by small and apparently random subsets of subjects, which are on average informative regarding sex, but which are less representative of the population as a whole.

In terms of classification accuracy, our result of 0.919 from scale-normalized feature matching appears consistent with recent literature on sex classification from volume-corrected MRIs. Logistic regression on a similar set of 399 HCP test subjects achieved 0.92 accuracy (with a model trained on an independent cohort of 3298 subjects) [9], and SVM classifiers applied volume-normalized grey matter maps of age-matched subjects including 396 HCP subjects (1614 in total) achieved 0.925 accuracy [32]. In terms of regions, the cerebellum and thalamus were found to be the most informative, however classification was found to be driven primarily by sex-informative features distributed throughout the brain rather than localized within individual neuroanatomical regions [9]. Our work also identified the thalamus as informative, we did not identify the cerebellum directly as it is a larger organ represented collections of smaller features, however these appear to impact overall classification similarly to other features distributed throughout the brain.

The consistency of our results with recent literature is particularly notable considering the very different prediction strategies, i.e. our method uses scale-normalized feature matching parameterized by a memory of approximately 114K features extracted from 422 subjects, vs neural network and logistic regression models with 1.3M parameters estimated 3299 volume-normalized subject images [9]. The features distributed throughout the brain and driving classification are of particular interest, observed here and in [9]. In our work, these appear to be similar features shared by small and seemingly random subsets of subjects, which when identified tend to be informative regarding sex. This seems consistent with the hypothesis of brain structure development affected by multiple independent mechanisms, leading to individual brains appearing as a unique mosaics of both female and male-informative features [33]. Feature matching offers a means of further investigating the variability of such sex-informative neuroanatomical structure that may only be observed in subsets of the population. For example, clusters of feature matches represent sparse correlations throughout the brain between subsets of subjects, and these could be characterized by alternative methods or imaging modalities to understand their role in strong overall sex prediction.

Various limitations are to be noted with regard to this work. Although features identified in the thalamic region appear informative regarding sex, we cannot further comment on their link to sex, and they may be simply the result of neurodevelopment in a larger or smaller cranial anatomy. Our investigation is limited to self-reported binary sex, and does not account for transgender or intersex individuals who do not necessarily fall into binary male or female categories. Our results are limited to the HCP cohort age range of 22-36 years, where the brain morphology is relatively stable, and does not include young or elderly subjects whose anatomy would be affected by growth and aging.

In conclusion, our findings are consistent with the interpretation that an individual level, each brain image may be viewed as mosaic of features that are generally associated with either sex[33]. No single brain region classifies sex definitively, and classification is driven primarily by features distributed throughout the brain that may be pooled to predict sex with an accuracy of approximately 0.92. Individual subjects may be assigned a continuous sex prediction score, in our work based on the Jaccard index of feature set overlap, in other work based on regression [9] or likelihood ratios associated with features [34]. Specific features such as the thalamic features observed here are identified in 97% of subjects and predict sex with an accuracy 0.74. Experiments with rigidly registered subjects reveal that these occur despite size differences between males and females, which are normalized out by the SIFT method here. The misclassification rate for whole-brain sex classification from feature matching lies conservatively between 9%-15%, for rigidly and non-linearly registered images, respectively. Misclassified brains are observed to exhibit features typically associated with the opposite sex. ¹

We note that this research was approved by the Research Ethics Committee at ETS, that all procedures conformed to the guidelines laid out in the most recent revision of the Declaration of Helsinki, and that all participants provided written informed consent prior to the acquisition and use of their data.

Acknowledgements. This work was supported by the Canadian National Sciences and Engineering Research Council (NSERC) Discovery Grant and NIH grant P41EB015902 (NAC). Data were provided 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.

References

- [1] Hines, M.: Human gender development. *Neuroscience & Biobehavioral Reviews* **118**, 89–96 (2020)

¹Code required to reproduce our results may be obtained at <https://github.com/3dsift-rank>.

- [2] McCarthy, M.M.: Multifaceted origins of sex differences in the brain. *Philosophical Transactions of the Royal Society B: Biological Sciences* **371**(1688), 20150106 (2016)
- [3] Ferri, S.L., Abel, T., Brodtkin, E.S.: Sex differences in autism spectrum disorder: a review. *Current psychiatry reports* **20**, 1–17 (2018)
- [4] Abel, K.M., Drake, R., Goldstein, J.M.: Sex differences in schizophrenia. *International review of psychiatry* **22**(5), 417–428 (2010)
- [5] de Macedo, F.F.: *De L'encéphale Humain Avec et Sans Commissure Grise: Essai Synthétique D'observations Anatomo-psychiques Post Mortem, et Leurs Relations Avec la Criminalité*, (1889)
- [6] Allen, L.S., Gorski, R.A.: Sexual dimorphism of the anterior commissure and massa intermedia of the human brain. *Journal of Comparative Neurology* **312**(1), 97–104 (1991)
- [7] Wong, A.K., Wolfson, D.I., Borghei, A., Sani, S.: Prevalence of the interthalamic adhesion in the human brain: a review of literature. *Brain Structure and Function* **226**, 2481–2487 (2021)
- [8] Eliot, L., Ahmed, A., Khan, H., Patel, J.: Dump the “dimorphism”: Comprehensive synthesis of human brain studies reveals few male-female differences beyond size. *Neuroscience & Biobehavioral Reviews* **125**, 667–697 (2021)
- [9] Ebel, M., Domin, M., Neumann, N., Schmidt, C.O., Lotze, M., Stanke, M.: Classifying sex with volume-matched brain mri. *Neuroimage: Reports* **3**(3), 100181 (2023)
- [10] Gao, K., Shen, H., Liu, Y., Zeng, L., Hu, D.: Dense-cam: Visualize the gender of brains with mri images. In: 2019 International Joint Conference on Neural Networks (IJCNN), pp. 1–7 (2019). IEEE
- [11] Rudin, C.: Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature machine intelligence* **1**(5), 206–215 (2019)
- [12] Toews, M., Wells III, W.M.: Efficient and robust model-to-image alignment using 3d scale-invariant features. *Medical image analysis* **17**(3), 271–282 (2013)
- [13] Rister, B., Horowitz, M.A., Rubin, D.L.: Volumetric image registration from invariant keypoints. *IEEE Transactions on Image Processing* **26**(10), 4900–4910 (2017)

- [14] Lowe, D.G.: Distinctive image features from scale-invariant keypoints. *International journal of computer vision* **60**, 91–110 (2004)
- [15] Chauvin, L., Kumar, K., Desrosiers, C., Wells, W., Toews, M.: Efficient pairwise neuroimage analysis using the soft jaccard index and 3d keypoint sets. *IEEE Transactions on Medical Imaging* **41**(4), 836–845 (2021)
- [16] Chauvin, L., Kumar, K., Wachinger, C., Vangel, M., de Guise, J., Desrosiers, C., Wells, W., Toews, M., ADNI: Neuroimage signature from salient keypoints is highly specific to individuals and shared by close relatives. *NeuroImage* **204**, 116208 (2020)
- [17] Van Essen, D.C., Smith, S.M., Barch, D.M., Behrens, T.E., Yacoub, E., Ugurbil, K., Consortium, W.-M.H., *et al.*: The wu-minn human connectome project: an overview. *Neuroimage* **80**, 62–79 (2013)
- [18] Cover, T., Hart, P.: Nearest neighbor pattern classification. *IEEE transactions on information theory* **13**(1), 21–27 (1967)
- [19] Toews, M., Arbel, T.: Detection, localization, and sex classification of faces from arbitrary viewpoints and under occlusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **31**(9), 1567–1581 (2008)
- [20] Toews, M., Wachinger, C., Estepar, R.S.J., Wells III, W.M.: A feature-based approach to big data analysis of medical images. In: *Information Processing in Medical Imaging: 24th International Conference, IPMI 2015, Sabhal Mor Ostaig, Isle of Skye, UK, June 28–July 3, 2015, Proceedings*, pp. 339–350 (2015). Springer
- [21] Machado, I., Toews, M., Luo, J., Unadkat, P., Essayed, W., George, E., Teodoro, P., Carvalho, H., Martins, J., Golland, P., *et al.*: Non-rigid registration of 3d ultrasound for neurosurgery using automatic feature detection and matching. *International journal of computer assisted radiology and surgery* **13**, 1525–1538 (2018)
- [22] Wachinger, C., Toews, M., Langs, G., Wells, W., Golland, P.: Keypoint transfer for fast whole-body segmentation. *IEEE transactions on medical imaging* **39**(2), 273–282 (2018)
- [23] Wu, H., Wang, D., Shi, L., Wen, Z., Ming, Z.: Midsagittal plane extraction from brain images based on 3d sift. *Physics in Medicine & Biology* **59**(6), 1367 (2014)
- [24] Paganelli, C., Peroni, M., Riboldi, M., Sharp, G.C., Ciardo, D., Alterio, D., Orecchia, R., Baroni, G.: Scale invariant feature transform in adaptive radiation therapy: a tool for deformable image registration assessment and re-planning indication. *Physics in Medicine & Biology* **58**(2), 287 (2012)

- [25] Ni, D., Qu, Y., Yang, X., Chui, Y.P., Wong, T.-T., Ho, S.S., Heng, P.A.: Volumetric ultrasound panorama based on 3d sift. In: International Conference on Medical Image Computing and Computer-assisted Intervention, pp. 52–60 (2008). Springer
- [26] Toews, M., Wells III, W., Collins, D.L., Arbel, T.: Feature-based morphometry: Discovering group-related anatomical patterns. *NeuroImage* **49**(3), 2318–2327 (2010)
- [27] Puglisi, L., Barkhof, F., Alexander, D.C., Parker, G.J., Eshaghi, A., Ravi, D.: Deepbrainprint: A novel contrastive framework for brain mri re-identification. arXiv preprint arXiv:2302.13057 (2023)
- [28] Mazziotta, J.C., Toga, A.W., Evans, A., Fox, P., Lancaster, J., *et al.*: A probabilistic atlas of the human brain: theory and rationale for its development. *Neuroimage* **2**(2), 89–101 (1995)
- [29] Jenkinson, M., Bannister, P., Brady, M., Smith, S.: Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage* **17**(2), 825–841 (2002)
- [30] Andersson, J.L., Jenkinson, M., Smith, S., *et al.*: Non-linear registration, aka spatial normalisation fmrib technical report tr07ja2. FMRIB Analysis Group of the University of Oxford **2**(1) (2007)
- [31] Spring, S., Lerch, J.P., Henkelman, R.M.: Sexual dimorphism revealed in the structure of the mouse brain using three-dimensional magnetic resonance imaging. *Neuroimage* **35**(4), 1424–1433 (2007)
- [32] Wiersch, L., Hamdan, S., Hoffstaedter, F., Votinov, M., Habel, U., Clemens, B., Derntl, B., Eickhoff, S.B., Patil, K.R., Weis, S.: Accurate sex prediction of cisgender and transgender individuals without brain size bias. *Scientific reports* **13**(1), 13868 (2023)
- [33] Joel, D.: Beyond the binary: Rethinking sex and the brain. *Neuroscience & Biobehavioral Reviews* **122**, 165–175 (2021)
- [34] Phillips, O.R., Onopa, A.K., Hsu, V., Ollila, H.M., Hillary, R.P., Hallmayer, J., Gotlib, I.H., Taylor, J., Mackey, L., Singh, M.K.: Beyond a binary classification of sex: an examination of brain sex differentiation, psychopathology, and genotype. *Journal of the American Academy of Child & Adolescent Psychiatry* **58**(8), 787–798 (2019)