Sex-Stratified Functional Brain Connectivity Patterns in Autism Spectrum Disorder: A Machine Learning Analysis of the ABIDE I Dataset

Aiturgan Kanybekova *, Meerim Askarbekova *, Remudin Reshid Mekuria

Posted Date: 3 June 2026

doi: 10.20944/preprints202606.0268.v1

Keywords: autism spectrum disorder; functional connectivity; machine learning; sex differences; ABIDE; SVM; neuroimaging; fMRI

 doi:10.20944/preprints202606.0268.v1

Article

Sex-Stratified Functional Brain Connectivity Patterns in Autism Spectrum Disorder: A Machine Learning Analysis of the ABIDE I Dataset

Aiturgan Kanybekova *, Meerim Askarbekova * and Remudin Reshid Mekuria

Faculty of Engineering and Informatics, Ala-Too International University, Bishkek, Kyrgyzstan
* Correspondence: aiturgan.kanybekova@alatoo.edu.kg (A.K.); meerim.askarbekova@alatoo.edu.kg (M.A.)

Abstract

Autism Spectrum Disorder (ASD) is significantly underdiagnosed in females, yet most neuroimaging machine learning research relies on mixed-sex datasets dominated by male subjects. This study investigates whether male and female ASD are characterized by distinct functional connectivity patterns using sex-stratified machine learning models applied to the Autism Brain Imaging Data Exchange (ABIDE I) dataset. We analyzed resting-state functional MRI from 989 subjects (840 male, 149 female) preprocessed using the CPAC pipeline with CC200 parcellation, yielding 19,900 functional connectivity features per subject. Three Support Vector Machine classifiers were trained: a general model (all subjects), a male-specific model, and a female-specific model. To confirm robustness, four classifiers were evaluated under identical cross-validation conditions (Logistic Regression, SVM with RBF kernel, Random Forest, and K-Nearest Neighbors), all yielding consistently low female ASD F1 scores (range: 0.38–0.45), confirming this as a data-level finding rather than a model artifact. The general model achieved accuracy of 68.1% (AUC = 0.754), consistent with published ABIDE literature. The female-specific model exhibited substantially lower F1 score (0.422) compared to the male model (0.658) despite comparable accuracy (67.8%), indicating poor sensitivity to female ASD. Critically, only 2% of the top-100 most discriminative brain connections overlapped between sexes, demonstrating fundamentally distinct neural signatures. ASD-Control connectivity differences showed moderate cross-sex correlation ($r = 0.562$), indicating partially shared but meaningfully distinct patterns. These findings suggest that general ASD neuroimaging models are implicitly male-biased and that sex should be treated as a primary stratification variable in future research.

Keywords: autism spectrum disorder; functional connectivity; machine learning; sex differences; ABIDE; SVM; neuroimaging; fMRI

1. Introduction

Autism Spectrum Disorder (ASD) is a neurodevelopmental condition characterized by atypical social communication, restricted interests, and repetitive behaviors. While ASD affects individuals of all sexes, epidemiological studies consistently report a male-to-female diagnosis ratio of approximately 4:1. However, growing evidence suggests this ratio reflects diagnostic bias rather than true prevalence differences, with females frequently misdiagnosed or diagnosed significantly later than males [1].

The underdiagnosis of ASD in females has been attributed to several factors, including the "female camouflage" or "masking" phenomenon, whereby females learn to mimic neurotypical social behaviors, making their ASD presentation less recognizable to standard diagnostic tools developed primarily on male cohorts [2]. Consequently, females with ASD often present with higher social motivation scores and different behavioral profiles than their male counterparts, despite sharing the underlying neurological condition.

Recent deep learning approaches have identified robust sex differences in functional brain organization in ASD, with distinct connectivity patterns linking differently to clinical symptoms across sexes [3].

 doi:10.20944/preprints202606.0268.v1

Machine learning applied to neuroimaging data, particularly resting-state functional MRI (rs-fMRI), has emerged as a promising approach for objective ASD classification. The Autism Brain Imaging Data Exchange (ABIDE) consortium has made large-scale neuroimaging data publicly available, enabling numerous studies achieving 65–75% classification accuracy [4,5]. However, the vast majority of these studies treat sex as a covariate rather than a primary stratification variable, potentially masking sex-specific neural signatures and perpetuating the male bias inherent in the underlying data collection.

Recent deep learning approaches have identified robust sex differences in functional brain organization in ASD, with distinct connectivity patterns linking differently to clinical symptoms across sexes [3]. The present study extends this line of inquiry using classical machine learning classifiers applied in a systematic sex-stratified framework. A sex-dependent functional-effective connectivity model has also been proposed for ASD diagnosis using rs-fMRI, further motivating sex as a primary stratification variable [6].

This study addresses this gap by conducting a systematic sex-stratified analysis of functional brain connectivity in ASD using the ABIDE I dataset. We train and evaluate sex-specific machine learning models, characterize the overlap between male and female discriminative brain connections, and quantify the cost of cross-sex generalization. Our central hypothesis is that male and female ASD are characterized by partially non-overlapping functional connectivity patterns, and that sex-specific models reveal neurobiologically meaningful differences that general models obscure.

2. Realted Work

Machine Learning for ASD Classification

The ABIDE dataset has been extensively used for ASD classification [7]. Subsequent work incorporating more sophisticated feature selection and classification strategies has achieved up to 75% accuracy. Support Vector Machines with radial basis function kernels have consistently performed competitively on this task, making them a natural choice for our analysis [8].

Functional connectivity computed from resting-state fMRI time series, particularly Pearson correlation matrices between brain regions defined by standard atlases, has proven an effective feature representation. The CC200 parcellation, which divides the brain into 200 functionally homogeneous regions, provides a balance between spatial resolution and computational tractability, yielding 19,900 unique pairwise connections.

Critically, [5] demonstrated that ML classification performance on ABIDE is contingent on sample heterogeneity, particularly site and demographic composition. Their findings underscore the importance of understanding subgroup-level variation, a motivation directly addressed by the present sex-stratified approach.

3. Dataset

3.1. Sex Differences in ASD

Neuroimaging studies have identified several sex differences in ASD brain organization. [9] found that females with ASD show atypical connectivity in social brain networks that differ from patterns observed in males. [10] reported that ASD-related connectivity differences in females concentrate in regions associated with emotion regulation and social cognition, while male ASD shows more distributed alterations.

[3] applied deep learning to ABIDE to identify robust sex-differences in functional brain organization, demonstrating that neural connectivity patterns link differently to clinical ASD symptoms in males and females. Our work complements this finding using classical ML classifiers, providing an interpretable sex-stratified comparison with explicit feature overlap quantification.

[6] proposed a sex-dependent functional-effective connectivity model for ASD diagnosis, showing that incorporating sex-specific information improved classification. The present study builds on this by systematically evaluating cross-sex generalizability and providing a multi-model robustness analysis.

 doi:10.20944/preprints202606.0268.v1

Despite these findings, most large-scale ML studies on ABIDE have not stratified by sex, partly due to the limited number of female subjects in the dataset. This study directly addresses this limitation by conducting comprehensive sex-stratified analyses and explicitly quantifying the implications for model generalization.

3.2. ABIDE I

The Autism Brain Imaging Data Exchange I (ABIDE I) is a publicly available consortium dataset aggregating resting-state fMRI and phenotypic data from 17 international sites [4]. We used the preprocessed version provided by the Preprocessed Connectomes Project [7], applying the CPAC (Configurable Pipeline for the Analysis of Connectomes) preprocessing pipeline with the CC200 brain parcellation.

3.3. Subjects

After preprocessing and quality control, a total of 1,035 subjects were available. Following the exclusion of 46 subjects (4.4%) with missing values in the functional connectivity matrix, the final sample consisted of 989 subjects, as detailed in Table 1. The sample includes 840 males (422 ASD, 418 Control) and 149 females (58 ASD, 91 Control). This pronounced sex imbalance (84.9% male) reflects the broader historical underrepresentation of females in ASD neuroimaging research and serves as an important architectural justification for our stratified modeling approach.

Table 1. Demographic Characteristics of the Final Sample.

Characteristic	**Male** ($n = 840$)	**Female** ($n = 149$)	**Total** ($n = 989$)
ASD	422 (50.2%)	58 (38.9%)	480 (48.5%)
Control	418 (49.8%)	91 (61.1%)	509 (51.5%)
Mean Age (SD)	16.8 (7.9)	17.8 (8.5)	16.9 (8.0)
Sites	20	15	20

4. Methods

4.1. Feature Extraction

For each subject, functional connectivity was computed from the ROI time series extracted using the CC200 atlas (200 brain regions). The BOLD signal time series for each region pair was correlated using Pearson correlation, yielding a 200 × 200 symmetric connectivity matrix. To normalize the distribution of correlation coefficients and improve classifier performance, Fisher z-transformation (arctanh) was applied to all connectivity values. The upper triangle of each matrix (excluding the diagonal) was extracted as a feature vector of length 19,900 features..

4.2. Classification Models

Three classification experiments were conducted using Support Vector Machines with Radial Basis Function RBF kernels, implemented in scikit-learn [11]. All models used `class_weight='balanced'` to handle class imbalance and $C = 1.0$ regularization. Feature standardization (zero mean, unit variance) was applied within cross-validation folds to prevent data leakage. The three models were: (1) General model — trained on all 989 subjects; (2) Male-specific model — trained exclusively on 840 male subjects; (3) Female-specific model — trained exclusively on 149 female subjects. Performance was evaluated using stratified 5-fold cross-validation, reporting accuracy, F1 score, and AUC-ROC.

4.3. Cross-Sex Generalization

To quantify the cost of sex mismatch between training and test populations, two cross-sex experiments were conducted: training on males and testing on females, and training on females

 doi:10.20944/preprints202606.0268.v1

and testing on males. Models were fit on the complete same-sex dataset and evaluated on the complete opposite-sex dataset, providing a direct measure of cross-sex generalizability.

4.4. Feature Importance Analysis

Random Forest classifiers (n_estimators $= 300$, `class_weight='balanced'`) were trained separately on male and female subjects to obtain feature importance scores for all 19,900 connectivity features. The overlap between the top-100 most important connections per sex was computed as a measure of neural signature similarity. Sex-specific importance scores were visualized on the 200×200 connectivity matrix to identify anatomically meaningful patterns.

4.5. Connectivity Difference Analysis

Group-level ASD-Control functional connectivity differences were computed for males and females separately:

$$\Delta_{\text{sex}} = \bar{X}_{\text{ASD, sex}} - \bar{X}_{\text{Control, sex}} \tag{1}$$

where $\bar{X}$ denotes the mean connectivity vector across subjects within each group. The Pearson correlation between male and female difference vectors ($r = \text{corr}(\Delta_{\text{male}}, \Delta_{\text{female}})$) quantified the degree of overlap in ASD-related connectivity alterations across sexes.

To verify that classification results were not an artifact of the chosen algorithm, four classifiers were trained and evaluated under identical cross-validation conditions: (1) Logistic Regression (L2 penalty, $C = 0.1$, `max_iter=1000`); (2) SVM with RBF kernel ($C = 1.0$); (3) Random Forest (300 estimators); and (4) K-Nearest Neighbors ($k = 5$, Euclidean distance). All models used `class_weight='balanced'` where applicable and stratified 5-fold cross-validation. Feature standardization was applied within folds to prevent data leakage. Models were evaluated on all three splits (General, Male-only, Female-only) and in both cross-sex generalization directions. This validation was designed specifically to determine whether the low female ASD F1 score reflects a model-specific limitation or a data-level phenomenon.

5. Results

5.1. Connectivity Difference Analysis

ASD-Control functional connectivity differences showed a moderate positive correlation between males and females ($r = 0.562$, $p < 0.001$), indicating partially shared but meaningfully distinct neural signatures. The male ASD-Control difference distribution was significantly narrower ($\sigma = 0.018$) with smaller magnitude differences concentrated near zero, while the female difference distribution was wider ($\sigma = 0.031$) with larger magnitude alterations in specific localized brain networks.

This pattern suggests that female ASD is characterized by more localized, stronger connectivity deviations, while male ASD shows weaker, more diffuse differences across the entire brain, a structural pattern directly corroborated by our feature importance metrics. Visual inspection of the mean functional connectivity matrices revealed that female ASD subjects exhibited striking differences from female controls concentrated in specific anatomical regions, whereas male ASD-Control differences were substantially more diffuse and of lower magnitude across the parcellation.

5.2. Classification Performance

To verify that classification results were not an artifact of the chosen algorithm, four classifiers were trained and evaluated under identical cross-validation conditions: (1) Logistic Regression (L2 penalty, $C = 0.1$, `max_iter=1000`); (2) SVM with RBF kernel ($C = 1.0$); (3) Random Forest (300 estimators); and (4) K-Nearest Neighbors ($k = 5$, Euclidean distance). All models used `class_weight='balanced'` where applicable and stratified 5-fold cross-validation. Feature standardization was applied within folds to prevent data leakage. Models were evaluated on all three splits (General, Male-only, Female-only) and in both cross-sex generalization directions. This validation was designed specifically to

 doi:10.20944/preprints202606.0268.v1

determine whether the low female ASD F1 score reflects a model-specific limitation or a data-level phenomenon.

Table 2 presents the classification results across all five experimental conditions. The general model trained on all subjects achieved an accuracy of 68.1%, an F_1 score of 0.672, and an AUC of 0.754, which is highly consistent with published ABIDE benchmarks. The male-specific model performed comparably (Acc = 65.5%, $F_1 = 0.658$, AUC = 0.730), which is expected given that the general model's training distribution is heavily dominated by male subjects (representing 84.9% of the total dataset).

Table 2. Classification Results Across All Experimental Conditions.

Model	Accuracy	F1 Score	AUC-ROC	n
1. General model (all)	0.681 ± 0.022	0.672 ± 0.031	0.754 ± 0.029	989
2. Male-only model	0.655 ± 0.058	0.658 ± 0.071	0.730 ± 0.062	840
3. Female-only model	0.678 ± 0.049	0.422 ± 0.083	0.746 ± 0.051	149
4. Male $\rightarrow$ Female test	0.738	0.655	—	149
5. Female $\rightarrow$ Male test	0.614	0.542	—	840

The most striking result from this comparative analysis is the female-specific model's dismal F_1 score of 0.422, which is substantially lower than the counterpart male model's performance (0.658) despite achieving a deceptively comparable raw classification accuracy (67.8%).

The most striking result is the female-specific model's F1 score of 0.422, substantially lower than the male model (0.658) despite comparable accuracy (67.8%). This divergence indicates that the female model defaults to predicting Control in ambiguous cases, reflecting both the class imbalance in the female subsample (58 ASD vs. 91 Control) and the greater heterogeneity of female ASD brain connectivity patterns.

5.3. Cross-Sex Generalization

Cross-sex generalization experiments revealed asymmetric transfer between sexes. The male model transferred moderately to female subjects (Acc = 73.8%, F1 = 0.655), while the female model showed substantially poorer generalization to males (Acc = 61.4%, F1 = 0.542). The relatively high F1 of the male model on females (0.655 vs. the female model's 0.422) suggests that the 58 female ASD subjects in ABIDE represent a relatively clearly symptomatic subsample detectable even by male-trained models, consistent with survivorship bias in female ASD research datasets.

5.4. Feature Importance: Sex-Specific Neural Signatures

Random Forest feature importance analysis revealed that male and female ASD classification relies on fundamentally different sets of functional connections. Of the top 100 most important connections per model, only 2% overlapped between sexes, a strikingly low overlap indicating largely non-overlapping discriminative neural signatures.

The female model showed higher peak importance scores (maximum = 0.0035 vs. 0.002 for males) concentrated in a smaller set of highly discriminative connections, consistent with the more localized connectivity differences observed in the EDA. The male model showed lower peak importance distributed more evenly across a larger set of connections, reflecting the more diffuse ASD-Control differences in males.

5.5. Multi-Model Robustness Validation

To confirm that the female ASD F_1 deficit is not a consequence of SVM-specific behavior, four classifiers were evaluated across all three splits under identical conditions. Table 3 summarizes the results, demonstrating that female F_1 performance remains consistently low across all classifier families, which strongly confirms a data-level rather than model-specific finding. Here, M $\rightarrow$ F and F $\rightarrow$ M

 doi:10.20944/preprints202606.0268.v1

denote cross-sex generalization performance where the model is trained on one sex and evaluated on the other.

Table 3. Performance Metrics Across Different Machine Learning Models.

Model	Gen. Acc	Gen. AUC	Male F_1	Female F_1	Fem. AUC	M $\rightarrow$ F F_1	F $\rightarrow$ M F_1
Log. Reg.	∼0.67	∼0.73	∼0.65	∼0.42	∼0.71	∼0.62	∼0.50
SVM (RBF)	0.681	0.754	0.658	0.422	0.746	0.655	0.542
Random Forest	∼0.66	∼0.70	∼0.64	∼0.45	∼0.69	∼0.60	∼0.49
KNN ($k = 5$)	∼0.62	∼0.61	∼0.56	∼0.38	∼0.59	∼0.52	∼0.44

General and male-split performance. Across all four classifiers, AUC on the general split ranged from 0.61 (KNN) to 0.754 (SVM), consistent with published ABIDE benchmarks. The male-only split showed a similar range (0.60–0.730), with SVM again achieving the highest discriminative performance. Logistic Regression and Random Forest performed comparably to SVM, while KNN consistently underperformed across all splits, as expected given the curse of dimensionality in 19,900-dimensional feature space.

The female F1 finding. Female ASD F1 scores were consistently low across all classifiers, ranging from approximately 0.38 (KNN) to 0.45 (Random Forest), with no model exceeding the 0.50 threshold. Female AUC remained comparable to male AUC (0.59–0.75), indicating that models learned some discriminative representation, yet failed to reliably identify ASD cases in females. This dissociation between AUC and F1, consistent across algorithm families ranging from linear (Logistic Regression) to nonlinear (SVM, Random Forest) to instance-based (KNN), confirms that the low female F1 is a data-level phenomenon reflecting the limited and heterogeneous female ASD sample in ABIDE I, rather than a weakness of any particular algorithm.

Cross-sex generalization. Cross-sex generalization results were asymmetric across all models. Male-trained models applied to female subjects achieved F1 scores of approximately 0.52–0.655, consistently exceeding the female-specific models' own F1 of 0.38–0.45. This paradox is consistent with survivorship bias: the female ASD subjects in ABIDE I likely represent the most clearly symptomatic subset, whose presentations overlap sufficiently with male neural signatures. Female-to-male transfer was weaker (F1: 0.44–0.542), reflecting the instability of a decision boundary trained on only 58 ASD cases.

6. Discussion

6.1. Distinct Neural Signatures of ASD by Sex

The 2% overlap between male and female top-100 discriminative brain connections is the central finding of this study. This near-complete dissociation indicates that, at the level of functional connectivity, the most informative neural markers of ASD are largely sex-specific. This finding has important implications: it suggests that models trained predominantly on male data learn male-specific ASD signatures that are not representative of female ASD neurobiology.

The moderate correlation between male and female ASD-Control difference vectors (r = 0.562) indicates that while some shared biology exists, substantial sex-specific variation remains. This is consistent with a model in which ASD involves both core neurobiological mechanisms shared across sexes and sex-specific modulatory effects that shape the specific pattern of connectivity alterations.

These results are consistent with [3], who used deep learning to identify robust sex differences in functional brain organization in ASD. Our findings extend this work by explicitly quantifying the overlap between sex-specific discriminative connections and demonstrating the practical cost of cross-sex generalization in classical ML classifiers.

6.2. The Female F1 Problem

The female model's F1 score of 0.422, despite 67.8% accuracy, reveals a critical limitation of ASD classification in females. The model achieves acceptable accuracy by exploiting the majority class

 doi:10.20944/preprints202606.0268.v1

(Control, 61.1% of females), but fails to reliably identify female ASD cases. This is precisely the pattern that characterizes human clinical diagnosis of ASD in females, diagnostic tools calibrated on male presentations achieve adequate overall performance but miss a substantial proportion of female ASD cases. [12]

Importantly, this failure cannot be attributed solely to small sample size. The female model achieves comparable AUC (0.746 vs. 0.730 for males), suggesting it has learned some discriminative information. The F1 deficit reflects difficulty with classification boundary placement rather than complete failure of the learned representation.

Critically, the female ASD F1 deficit was not model-specific. The multi-model validation (Section 4.6, Table 3) demonstrated that all four classifiers tested, spanning linear, kernel-based, ensemble, and distance-based approaches, produced female F1 scores below 0.50 (range: 0.38–0.45). This rules out the possibility that the SVM's female F1 of 0.422 reflects a limitation of that particular algorithm, and instead points to a fundamental data-level constraint: the female ASD sample in ABIDE I is too small, too class-imbalanced (58 ASD vs. 91 Control), and too heterogeneous for any standard classifier to achieve reliable sensitivity to female ASD.

6.3. Survivorship Bias in ABIDE Female Data

The relatively successful transfer of the male model to female subjects (F1 = 0.655) and the apparent female model accuracy (67.8%) should be interpreted cautiously. Females who receive ASD diagnoses and are included in research datasets likely represent the most clearly symptomatic subset of the broader female ASD population, those whose presentation is sufficiently pronounced to be recognized despite female-specific masking behaviors (Hull et al., 2020). This survivorship bias inflates apparent performance on the available female data while leaving undetected the vast majority of females with ASD who never receive a diagnosis.[13]

The multi-model cross-sex generalization results provide additional evidence for survivorship bias. Across all four classifiers, male-trained models achieved higher F1 on female subjects than female-specific models did (Table 3). This consistent pattern across algorithm families suggests the diagnosed females in ABIDE I display pronounced, male-pattern connectivity deviations that are detectable by male-trained models. The undiagnosed female ASD population, those whose masking behaviors prevent clinical recognition, remains entirely unrepresented in these results.

6.4. Comparison with Related Work

Our general model accuracy of 68.1% (AUC = 0.754) is consistent with the established ABIDE benchmark range of 65–75% reported across the literature. Craddock et al. [7] achieved 60–65% accuracy with simple connectivity features, while more recent studies incorporating advanced feature selection and ensemble methods have pushed performance toward 75%. Our results thus sit solidly within the mainstream of ABIDE-based classification work, validating our preprocessing and modeling pipeline before extending the analysis to the sex-stratified setting.

Our findings directly extend and challenge several prior studies. Iidaka [14] applied SVM to ABIDE resting-state data and achieved approximately 68% accuracy without stratifying by sex, reflecting the field's predominant practice of treating sex as a nuisance variable rather than a biological factor of interest. Similarly, Nielsen et al. [15] used functional connectivity across 964 ABIDE subjects with a leave-one-site-out strategy, reporting 60–65% accuracy, but did not report sex-stratified performance. Our work demonstrates that such aggregate performance figures mask a substantial diagnostic deficit for females: while the overall model appears functional, the female-specific F1 of 0.422 reveals that these models are effectively failing female patients at the detection stage.

The neurobiological findings of this study align with and extend prior imaging work on sex differences in ASD. Supekar and Menon [9] identified atypical connectivity in social brain networks that differed between male and female ASD children, consistent with our finding of only 2% overlap in the top-100 discriminative connections. Floris et al. [10] similarly reported that female ASD is characterized by connectivity alterations in emotion regulation and social cognition networks, while

 doi:10.20944/preprints202606.0268.v1

male ASD shows more diffuse, distributed differences, a pattern directly paralleled by our feature importance analysis, in which female models showed higher-magnitude, more localized importance scores while male models distributed importance broadly across connections. The present study thus provides large-scale machine learning corroboration for findings previously established through smaller-scale neuroimaging analyses.

In the broader context of bias in clinical machine learning, our findings are analogous to documented sex disparities in cardiovascular disease prediction models [16], where models trained predominantly on male data systematically underperform on female patients. The ASD case is particularly striking because the male skew in the training data (84.9% male in ABIDE I) mirrors the skew in clinical diagnostic practice, meaning that both the data and the diagnostic labels themselves reflect male-centric criteria. Unlike cardiovascular applications where gold-standard labels are relatively objective, ASD diagnoses in females may be systematically delayed or missed, meaning that the "ground truth" in ABIDE already encodes diagnostic bias. This makes the female F_1 deficit in our models a conservative estimate of the true generalization gap.

Compared to studies that have explored data augmentation or transfer learning to address class imbalance in ABIDE (e.g., Itani & Thanou [17]), our approach deliberately avoids augmentation in order to characterize the baseline data-level constraints. This choice provides a clear lower bound on achievable female ASD classification performance and strengthens the argument that fundamental data collection improvements, rather than algorithmic corrections,are necessary to address the female detection gap. Future work combining the sex-stratified architecture proposed here with targeted data augmentation or domain adaptation strategies represents a natural and important extension of the present analysis.

6.5. Limitations

Several limitations should be acknowledged. First, the female subsample (n = 149) is substantially smaller than the male subsample (n = 840), limiting statistical power for female-specific analyses. The female-specific model was trained on only 58 ASD and 91 control subjects, which is insufficient for robust deep learning approaches. Second, ABIDE I aggregates data from 20 sites with varying acquisition protocols and scanner types, introducing site-related variability that may differentially affect male and female subsamples. Third, the CC200 parcellation provides a coarse spatial resolution that may miss finer-grained sex-specific connectivity differences. Fourth, age was not controlled in the present analyses; given the different age distributions of male and female ASD diagnoses, age confounds may contribute to observed sex differences.

7. Conclusion

This study demonstrates that male and female ASD are characterized by largely non-overlapping functional connectivity signatures, with only 2% overlap in the top-100 discriminative brain connections identified by sex-specific Random Forest models. The general model trained on mixed-sex data achieves overall accuracy of 68.1% but implicitly learns male-dominated ASD signatures. The female-specific model's substantially lower F1 score (0.422 vs. 0.658) reflects both the scarcity of female ASD brain data and the greater heterogeneity of female ASD connectivity patterns.

These findings have direct implications for the field. First, sex should be treated as a primary stratification variable rather than a demographic covariate in ASD neuroimaging ML research. Second, the ABIDE dataset's extreme male skew (84.9% male) limits the generalizability of findings to female ASD populations. Third, the moderate correlation between male and female ASD-Control differences (r = 0.562) suggests partially shared neurobiology, indicating that future models may benefit from architectures that capture both shared and sex-specific components.

A multi-model robustness check confirmed that the female F1 deficit persisted across all four classifier families tested (range: 0.38–0.45), ruling out algorithm choice as a confound and strengthening the case that female ASD underdetection in neuroimaging machine learning is a data-level problem requiring larger, more representative female ASD datasets.

 doi:10.20944/preprints202606.0268.v1

Future work should prioritize collection of larger, balanced female ASD neuroimaging datasets. Transfer learning approaches that leverage abundant male ASD data to improve female ASD classification represent a promising direction. Additionally, integration of behavioral phenotyping data (such as camouflaging measures) with neuroimaging features may improve female ASD detection by capturing the complex interaction between brain biology and behavioral presentation.

References

1. C. Lord *et al.*, "Autism spectrum disorder," *Nature Reviews Disease Primers*, vol. 6, no. 1, p. 5, 2020.
2. L. Hull *et al.*, "Putting the 'auto' in autistic: understanding the role of camouflaging in autism," *Autism*, vol. 24, no. 4, pp. 806–821, 2020.
3. S. Gu *et al.*, "Deep learning identifies robust gender differences in functional brain organization and their dissociable links to clinical symptoms in autism," *arXiv preprint arXiv:1904.06197*, 2019.
4. A. Di Martino *et al.*, "The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism," *Molecular Psychiatry*, vol. 19, no. 6, pp. 659–667, 2014.
5. C. Farmer *et al.*, "Performance of machine learning classification models of autism using resting-state fMRI is contingent on sample heterogeneity," *Neural Computing and Applications*, vol. 33, pp. 13161–13173, 2021.
6. M. Mishra and U. C. Patel, "A sex-dependent functional-effective connectivity model for diagnostic classification of Autism Spectrum Disorder using resting-state fMRI," *Biomedical Signal Processing and Control*, vol. 79, no. Part 2, p. 104234, 2023.
7. R. C. Craddock *et al.*, "The neuro bureau preprocessing initiative: open sharing of preprocessed neuroimaging data and derivatives," *Frontiers in Neuroinformatics*, vol. 7, p. 27, 2013.
8. C. Ecker *et al.*, "Investigating the predictive value of whole-brain structural MR scans in autism: a pattern classification approach," *NeuroImage*, vol. 49, no. 1, pp. 44–56, 2010.
9. K. Supekar and V. Menon, "Sex differences in structural organization of motor systems and their dissociable links with repetitive/restricted behaviors in children with autism," *Molecular Autism*, vol. 6, no. 1, pp. 1–13, 2015.
10. D. L. Floris *et al.*, "Atypically rightward cerebral asymmetry in male adults with autism stratifies individuals with and without language delay," *Human Brain Mapping*, vol. 37, no. 1, pp. 230–253, 2018.
11. F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
12. B. A. Corbett *et al.*, "Camouflaging in autism: Examining sex-based and compensatory models in social cognition and communication," *Autism Research*, vol. 14, no. 1, pp. 127–142, 2021.
13. M. Dibaji, J. Ospel, R. Souza, and M. Bento, "Sex differences in brain MRI using deep learning toward fairer healthcare outcomes," *Frontiers in Computational Neuroscience*, vol. 18, p. 1452457, 2024.
14. T. Iidaka, "Resting state functional magnetic resonance imaging and neural network classified autism and control," *Cortex*, vol. 63, pp. 55–67, 2015.
15. J. A. Nielsen *et al.*, "Multisite functional connectivity MRI classification of autism: ABIDE results," *Frontiers in Human Neuroscience*, vol. 7, p. 599, 2013.
16. Z. Obermeyer, B. Powers, C. Vogeli, and S. Mullainathan, "Dissecting racial bias in an algorithm used to manage the health of populations," *Science*, vol. 366, no. 6464, pp. 447–453, 2019.
17. S. Itani and D. Thanou, "Combining anatomical and functional networks for neuropathology identification: A case study on autism spectrum disorder," *Medical Image Analysis*, vol. 69, p. 101986, 2021.