National College of Ireland

Classifying Autism Spectrum Disorder via Brain Connectivity and Graph Neural Networks

Research Practicum
Msc Data Analytics

Valesca Frey Pereira
Student ID: 24190284

School of Computing
National College of Ireland

Supervisor: Dr. Anderson Simiscuka

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Valesca Frey Pereira
Student ID:	24190284
Programme:	Msc Data Analytics
Year:	2025
Module:	Research Practicum
Supervisor:	Dr. Anderson Simiscuka
Submission Due Date:	12/12/2025
Project Title:	Classifying Autism Spectrum Disorder via Brain Connectivity and Graph Neural Networks
Word Count:	XXX
Page Count:	20

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	
Date:	28th January 2026

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Dedicated to my big sister, who was late-diagnosed with ASD at age 32,
and who has always been a steady beacon of quiet kindness in my life.

Classifying Autism Spectrum Disorder via Brain Connectivity and Graph Neural Networks

Valesca Frey Pereira
24190284

Abstract

Autism Spectrum Disorder (ASD) is a neurodevelopmental condition where the one has challenges in social interactions, communication, and behaviours. Current researches show that male are more frequently diagnosed than female, with ratios up to 4:1. This imbalance raises an important concern: are the models adjusted in generalised data effectively classifying ASD in women? While the field has gained advances with the use of Graph Neural Network (GNN) models, the question of weather these models perform equally well considering sexes is largely unexplored. This study investigates whether autistic women exhibit different neural connectivity patterns that require sex-specific diagnostic modelling. To explore this, fMRI data from the ABIDE I dataset were converted into brain graphs using the AAL atlas, combining functional correlations with structural features. The study began by training a Graph Convolutional Networks (GCN) and a Graph Attention Networks (GAT) as baseline for the future experiments. A series of training scenarios were applied to verify the sex differences systematically. Models were trained on men only, on all subjects combined, on women only, and on a balanced subset. Their performance was evaluated through stratified cross-validation and held-out testing. Across all scenarios, accuracy was consistently lower for women. When trained on men subjects only, the precision showed a gap of 20% men compared to women. In the balanced subset, women's performance improved, likely because they were more equally represented, although the small dataset and a 10-fold split also introduced variability. In the women only trained model, the accuracy improved to 70%. Overall, the results indicate that autistic women might have distinct connectivity shapes and might therefore require diagnostic models that are explicitly sex-aware. This work highlights the need for better female representation in neuroimaging research and suggests that future ASD classification studies should consider sex-specific modelling approaches.

1 Introduction

Autism Spectrum Disorder (ASD) is a neurodevelopmental condition where the one has challenges in social interactions, communication, and behaviours. It could be expressed in different levels of severity. According to (Loomes et al., 2017), males are more frequently diagnosed than females, with a ratio of 3:1. While other studies suggest an even higher ratio of 4:1. Although some of this discrepancy may reflect genuine differences in prevalence, some lines of discussion infer the discrepancy to biased diagnosis, which often leads to underdiagnosis women, whose behavioural and neural profiles may differ from

males. Underdiagnosis often means consequence in one personal life, misunderstood feelings, increased social pressure, limited access to mental health service, delayed support within educational or workplace settings, and in many cases secondary mental-health issues such as anxiety and depression.

In addition to these clinical concerns, studies have increasingly explored brain connectivity differences in ASD. For instance, (Li et al., 2023) found reduced connectivity in the prefrontal cortex and bilateral precuneus in females compared to males with ASD. Additionally, (Supekar et al., 2022) used spatiotemporal deep neural network (stDNN) to demonstrated that males and females with ASD have distinct patterns of functional brain organisation, which correspond to their difference in the clinical symptoms. Another promising direction is the use of fMRI to model the brain as a graph, where regions are represented as nodes and functional correlations form weighted edges. Graph Neural Networks (GNNs) have shown promising results in capturing patterns within these connectivity networks, and several studies have used them to attempt automated ASD classification. However, while the literature have demonstrated growing success applying GNNs for ASD classification, the models have not been applied to explore sex-specific features of ASD, raising an important methodological concern: if the models are mainly learning male brain patterns, can they reliably identify ASD in women?

This project aims to explore the gap by assessing whether graph-based machine learning models show performance differences between male and female participants. The study begins with a general baseline comparison between two commonly used GNN architectures: Graph Convolutional Networks (GCN) and Graph Attention Networks (GAT). Once GAT is selected as the prefered model, several experimental scenarios were implemented to test how sex representation affects model performance. These include training on men only, training on both sexes together, training on women only, and training on a deliberately balanced subset containing equal numbers of men and women.

The overarching motivation of this study is to determine whether women may require sex-specific modelling due to differences in functional connectivity patterns. Understanding this issue is important not only for machine-learning applications but also for the broader clinical challenge of ensuring that ASD diagnostic tools reflect the diversity of those affected.

Although the methodological approach is aligned with current practice in neuroimaging and machine learning, this research was conducted independently with limited computational resources. This imposed practical constraints on model complexity, training duration, and hyper-parameter optimisation. As well as the reduced number of female subjects on the public dataset must be noted. Nonetheless, the experiments were sufficient to reveal consistent patterns across sexes and provide initial evidence to address the research question.

Finally, the research question guiding this study is: **Do Graph Neural Network models classify ASD differently in males and females, and do women require sex-specific modelling?** To answer the question, a few objects were made: compare baseline performance of GCN and GAT models; analyse how model performance changes when trained on men only, women only, or both combined; investigate whether a balanced male–female dataset alters classification behaviour; and evaluate whether observed differences support the need for sex-specific ASD diagnostic modelling.

The remainder of this report is structured as follows: Section 2 is related work. Section 3 describes the methodology, including dataset preparation, graph construction, model design, and evaluation strategy. Section 4 design specification. Section 5 Implementation.

Section 6 Evaluation and discussion. Section 7 describes the conclusions of the study, and suggests future directions.

2 Related Work

Autism research has recently benefited from advances in deep learning, particularly through the use of neuroimaging models for classification and biomarker discovery. In this section, the literature is organised into three main strands: graph neural networks applied to brain connectivity, sex differences in ASD, and graph attention networks (GAT) as a specialised class of GNN models. The section concludes with a summary of the remaining research gap and how the present study addresses it.

Traditional clinical diagnostic for Autism Spectrum Disorder (ASD) were based on behaviour observations and cognitive tests, what made the process a long journey and, many times, not accurate. Although recent advances in machine learning have introduced data-driven methodologies to the field. The exploration of models, such Graph Neural Networks (GNNs), to Magnetic Resonance Imaging (MRI) datasets emerged as a promising computational approach for analysing brain connectivity and facilitate diagnosis. (Yang et al., 2021) have presented optimistic accuracies to detect ASD, achieving 72.4% accuracy through GAN model, followed by (Pan et al., 2025), who used a multi-scale GNN model and achieved 74.8% accuracy. The results support the GNN effectiveness in ASD classification.

However, most of the studies focus on children, refusing the fact that there is a growing population of underdiagnosed ASD into adulthood, particularly among adult women. (McCrossin, 2022) have found that approximately 80% of women remain underdiagnosed up to the age 18. This disparity might be influenced by the tendency to women being taught to best behaviour according to society rules, and end up masking the symptoms since childhood ((Simply Psychology, 2025)). Furthermore the patterns of behaviour in autistic women differ compared to autistic men ((Harmens et al., 2022)), which may have influenced in the development of biased diagnostic frameworks (((Trafton, 2022)). This section aims to explore significant studies and key collaborations in the application of GNN models to ASD detection, as well as it reviews the literature related to sex differences in ASD.

2.1 Graph Neural Networks in Brain Connectivity

Graph Neural Networks (GNNs) are artificial neural networks designed to analyse graphs. They are structured in such a way that the nodes represent the brain regions and the edges represent the brain connections ((Pan et al., 2025)), which perfectly corresponded to the ASD analysis requirements. (Wang et al., 2023) built a model that achieved 79.78% accuracy. The superior performance is due to the models ability to incorporate individual brain region and inter-regional connectivity simultaneously. The temporal dimension of brain connectivity has also emerged as relevant for GNN applications. (Tang et al., 2024) developed a GNN-LSTM hybrid model that captures both spatial features through graph convolution and temporal dynamics through long short-term memory networks, achieving accuracies of 80.4% on ABIDE I and 79.63% on ABIDE II datasets. This approach addresses a fundamental static connectivity analysis by modelling the dynamic fluctuations in brain network organization. However, while these temporal GNN approaches show

promise, they primarily focus on functional connectivity, leaving structural connectivity patterns unexplored.

When analysing brain connectivity for ASD, it is valuable to consider the distinctions between structural and functional connections (Walker et al., 2012). This suggests that effective GNN models should incorporate both connectivity modalities in the classification, structural and functional connections. In recent study, (Cardenas-Hernandez et al., 2022) suggests that structural connectivity changes in ASD may be more evident across individuals than functional connectivity changes. (Jiang et al., 2023) using diffusion tensor imaging (DTI) and machine learning achieved 96.77% classification accuracy in a discovery dataset, with independent validation accuracies of 91.67% and 88.89%. These results, while not using GNNs specifically, suggest that structural connectivity patterns may provide more stable and reliable biomarkers for ASD diagnosis than functional measures alone. The superior performance of structural connectivity classifiers raises important questions about the relative underutilization of DTI data in current GNN approaches.

On the other hand, current GNN models to ASD neuroimaging have faced some limitations that might difficult the clinical application. Many studies are based on the same public datasets (ABIDE). While these datasets enable large-scale studies and cross-validation, they might have biases related to data collection, demographic characteristics, and diagnostic criteria. In addition, the predominance of male records on these datasets might represent a disadvantage for female specific studies ((Di Martino et al., 2014)).

Recent deep-learning approaches, such as multi-site convolutional or autoencoder-based models, also report high accuracies (often above 80%) when aggregating data across sites and diverse atlas (Subah et al., 2021). However, these methods typically optimise global performance without examining whether model accuracy is uniform across sexes, which is the central focus of the present work.

2.2 Sex Differences in Autism Spectrum Disorder

Traditional studies agree that the symptoms of ASD differ in severity between male and female individuals ((Lai et al., 2017)), but suggesting that these differences would be the unique fact related to sex. In contrast, contemporary studies have found that, behind the symptoms, variations in brain patterns associated with ASD are also linked to sex differences. (Supekar et al., 2022) applied deep neural networks (DNN) to demonstrate that males and females with ASD exhibit distinct patterns of functional brain organisation. In addition, their model was capable to distinguish images between male and female with ASD with high accuracy, but could not classify it for neurotypical individuals, reaffirming that autistic brain differ from normative, when related to sex differences.

In one large study of functional connectivity in pediatric females with ASD, (Lawrence et al., 2020)) releved that females have greater functional brain connectivity than males. This find might represent the compensatory mechanisms that females are taught to help mask social communication difficulties, and might lead inaccurate diagnosis. Those novel finds based on machine learning correlated with an older hypothesis. In past years (Robinson et al., 2013) have introduced the "*female protective effect*" hypothesis, which suggests females may develop compensatory neural strategies and, in consequence, cognitive behaviours, that hidden traditional diagnostic markers.

Deep learning applications are a promising approach for sex-specific ASD diagnosis. (Heo et al., 2025) have used deep learning to reveal sex-specific brain patterns in ASD

neuroimaging, surpassing traditional studies that several times ignore sex differences. (Supekar et al., 2022) proved that DNN can reach high level of accuracy when separating male and female individuals analysis. Sex specific GNN models for ASD diagnosis are a critical and unaddressed challenge in current studies.

Overall, this literature shows that sex differences are both behavioural and neurobiological, but most current diagnostic models implicitly assume no sex difference in performance, leaving space for testing GNN and GAT models separately in men and women.

2.3 Graph Convolutional Networks (GCN)

Graph Convolutional Networks (GCNs) are one of the most widely GNN architectures used. GCNs work by aggregating information of node feature by their neighbour, allowing each brain region to update the representation based on its connected areas. This makes GCN a common baseline for connectivity analysis. They model the overall structure of the brain network without make assumptions about which connections are more relevant. However, because GCNs treat all neighbour nodes with equal importance, they may overlook subtle or sex specific patterns in ASD functional connectivity. For this reason, many recent studies prefer attention-based models, such as GAT, that can assign different weights to different connections. In the present work, GCN is primarily used as an initial basis to assess whether a more expressive model such as GAT is necessary for capturing sex differences in ASD.

2.4 Graph Attention Networks (GAT)

Graph Attention Networks (GAT) can learn to weight the importance of different brain connections, rather than treating all neighbours equally. In this way, GAT have been particularly promising for ASD classification. (Yang et al., 2021) proposed a sparse functional brain network combined with a GAT classifier on the ABIDE I dataset and reported an accuracy of around 72.4%, outperforming baseline graph models. More recent work such as mGNN-bw (Pan et al., 2025) used multi-scale graph representations, achieving 74.8% accuracy with the CC200 atlas and 73.2% with the AAL atlas on ABIDE I.

(Hu et al., 2021) introduced GAT-LI, a two-stage framework where a GAT is added with GNNExplainer to interpret the most influential connections and regions for ASD classification. Their best configuration achieves accuracies in middle 60% . Other studies combine attention mechanisms with multi-atlas and report higher overall accuracies ($>$ 80%), but the focus is typically on maximising global precision rather than understanding failure modes in specific demographic groups such as adult women.

In addition, GAT architectures remain relatively lightweight compared to some 3D CNN or hybrid models, which is compatible with the computational constraints of running experiments for this study.

To provide an overview of how the present work aligns with existing ASD classification research, Table 1 summarises neuroimaging studies using ABIDE. The table highlights the diversity of modelling approaches and it shows that most reported accuracies are between 65% and 80%, depending on the atlas, model complexity, and preprocessing choices. While these studies demonstrate the potential of graph and deep learning methods, none evaluates performance differences between men and women, which reinforces the need for the sex-specific investigation carried out in this project.

Author	Year	Model	Layers	Accuracy	Location
Subah et al.	2021	DNN	2	88%	Bangladesh
Bhavna et al.	2024	Ex-NEGAT	2	81%	India
Yousefian et al.	2023	GNN	multiple	80%	Iran
Wenqiu Pan et al.	2025	mGNN-bw	3	75%	China
Yang et al.	2021	GAT	2	72%	China
Hu et al.	2021	GAT-LI	2	68%	China
Jiang, Hao et al.	2020	Hi-GCN	multiple	67%	China
Okeno-Storms et al.	2025	GAT	3	65%	USA

Table 1: Summary of related works and model performance

2.5 Research gap

Findings across the literature show that an important gap remains in ASD diagnostic research. Although works such as (Pan et al., 2025) present models with strong performance, most studies still overlook the role of sex differences. Age is another dimension that receives limited attention. A large portion of ASD research concentrates on children, even though many individuals reach adulthood without ever receiving a diagnosis ((Organization for Autism Research, 2018)). As a result, combining GNN-based methods with an adult female sample represents a largely unexplored research space. Despite clear theoretical motivations, no previous study has applied this approach.

The study proposed aims to offer contributions both to ASD research and to clinical practice. The study aim to support the development of diagnostic tools suited for female subjects, which are often overlooked.

Methodologically, the work also extends beyond ASD. Designing GNN modelsa dapted to neuroimaging data can inform broader research in computational neuroscience, potentially influencing future studies on neurological and psychiatric disorders.

Overall, existing literature shows that GNN and GAT architectures achieve competitive performance on ABIDE; however, they usually optimise results for mixed-sex datasets and rarely investigate whether the models perform equally well across genders. This study aims to fill that gap by applying GAT classifiers to explicitly compare performance between males and females under different training settings, treating women as a central focus rather than a secondary subgroup.

3 Methodology

This section shows the research methodology, including the dataset, graph construction, model architectures selected for analysis, and the the experimental setups. The methodology follows the technical steps required to investigate whether Graph Neural Network models classify Autism Spectrum Disorder (ASD) differently in male and female participants. Figure 1 represents the research methodology adopted.

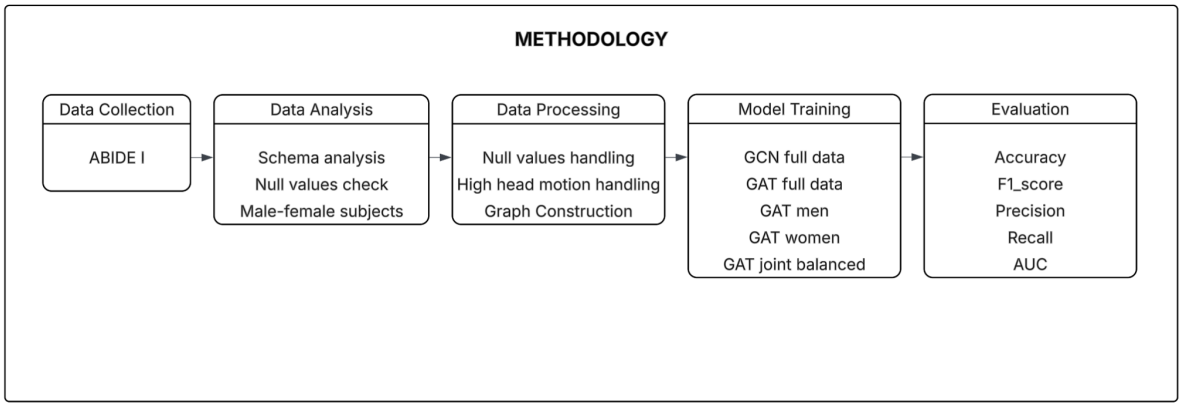

Figure 1: Research Methodology Schema

3.1 Dataset Description And Preprocessing

This study is based on the public Autism Brain Imaging Data Exchange (ABIDE I) dataset. The dataset includes Functional Magnetic Resonance Imaging (fMRI) from 17 international sites, as well as their phenotypic data.

The full ABIDE I dataset (Craddock et al. (2013)) collected was used for exploratory visualization, which initially contained approximately 1000 subjects. However, after quality control, we retained 610 subjects (511 males and 99 females) . We excluded participants with missing phenotypic data and/or with excessive head motion during fMRI (MeanFD>0.2). The action was taken to reduce noise and invalid correlations. This threshold is commonly used in fMRI studies and ASD research, as Mean Framewise Displacement values above 0.2 mm are most associated with motion related distortions which can bias the connectivity estimates (Power et al. (2012);Di Martino et al. (2014)).

The study made use of the preprocessed time series data provided by the Preprocessed Connectomes Project (PCP) ((Craddock et al., 2013) using the Configurable Pipeline for the Analysis of Connectomes (C-PAC) with the Automated Anatomical Labeling (AAL) atlas parcellation. The AAL divided the brain into 116 regions of interest (ROIs), providing consistent node definitions for graph construction. For each subject the preprocessed time series with 116 ALL ROIs was loaded, and the graphs constructed. Metadata used included sex, diagnostic labels (autistic or neurotypical), site, and IQ. After quality control, the final dataset comprised 610 participants. The sample was strongly male-dominant, with females representing only 16.2% of the cohort (Figure 2). Data were acquired from 14 imaging sites, reflecting the multi-site nature of the dataset (Figure 3). Diagnostic labels included 282 (46.2%) autistic and 328 (53.8%) neurotypical participants. IQ values were available vastly available for the subjects, with valid scores ranging from 98 to 148 and a median of approximately 109, indicating a relatively broad range of cognitive profiles in the sampled population. However, the dataset primarily represents cognitively able individuals and is unlikely to capture more severe ASD presentations or cases involving combined intellectual disability.

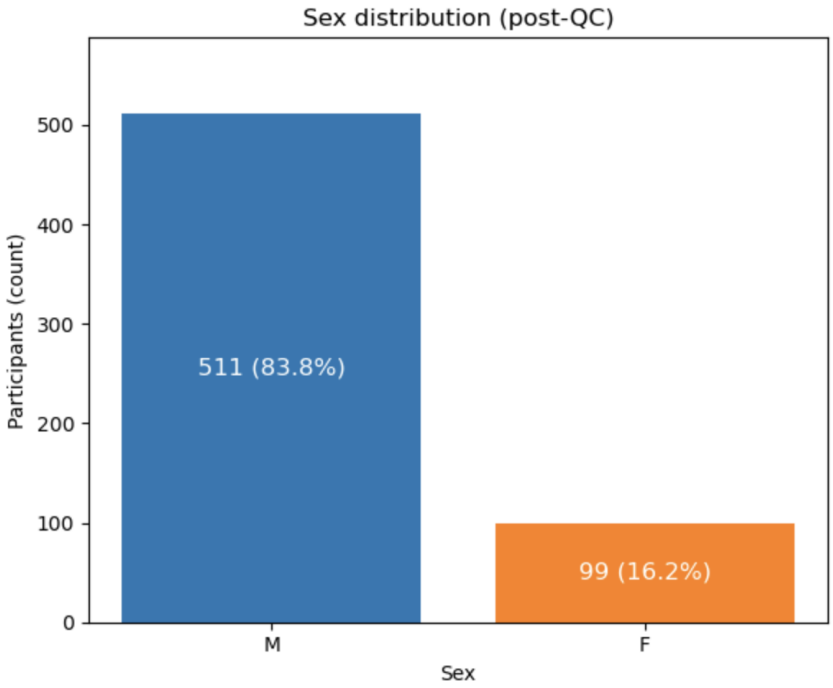

Figure 2: Sex distribution after quality control

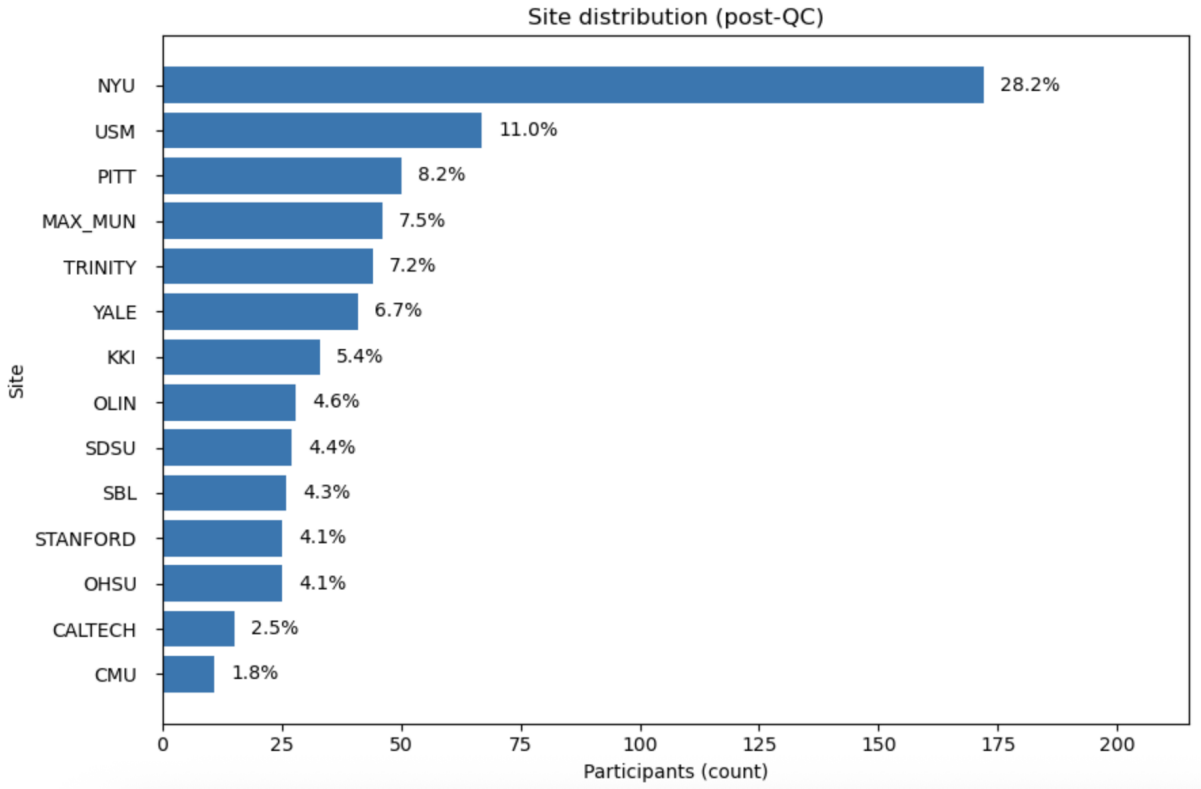

Figure 3: Site distribution after quality control

3.2 Model Architecture and training configuration

This study used Graph Convolutional Network (GCN) and Graph Attention Network (GAT) architectures.

3.2.1 Graph Convolutional Network (GCN)

GCN model used two layers. Each layer aggregates information from a ROI neighbours and then applies a non-linear activation function. And later performed binary classification.

3.2.2 Graph Attention Network (GAT)

A multi-head GAT model was built to learn which brain regions (ROI) are most important by assigning attention weights between them. Each layer used several attention heads, followed by non-linear activations, and dropout for regularisation, to combine all node features into a single representation.

3.3 Experimental Scenarios

The study baseline compared GCN and GAT architectures for choosing the most suitable for future experiments. As result, several experimental scenarios were designed using GAT to analyse the sex differences in ASD. All experiments were designed based on hardware limitations. Extensive hyperparameters or deeper models demonstrated not feasible for the study.

3.4 Evaluation Metrics

The models performance was evaluated through Accuracy, Precision, Recall, F1-score, and AUC. The metrics were computed separately for different groups as all test subjects; male test subjects; female test subjects. The distinguishing aims to address the study research question regarding sex awareness.

3.5 Method Limitations

The study lean on limitations. It is noted that the public dataset has multisite variability and less female than males data. Additionally, all experiments were conducted locally on a Macbook 8GB Ram and M3 chip, restricting model complexity, and training time.

4 Design Specification

This section describes the pipeline designed to classify Autism within the use on GNN models. The pipeline includes converting the fMRI time series loaded data into the graphs. Apply the GNN and GAT models, and aloow specific experiments of training subsets (men only, women only, balanced groups) to investigate sex aware performance of the models. Figure 4 represents the six main modules designed: data loading, graph construction, model, training, evaluation, and analysis.

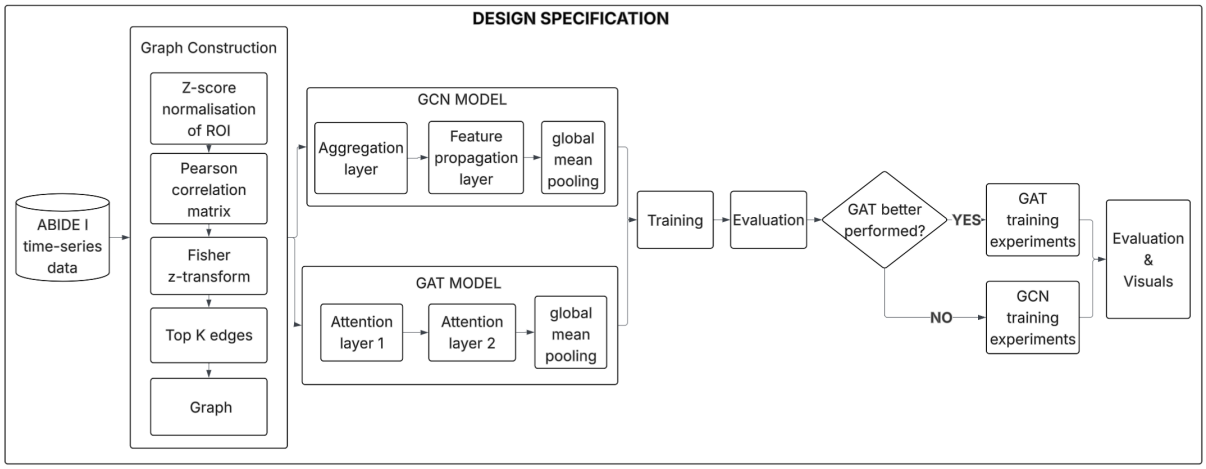

Figure 4: Design Architecture

4.1 Design requirements

The design considers the hardware limitations, sticking with modest graph construction (116 ROIs) and model depth. It also includes fixed random seeds and modular files structure for future reproducibility. Additionally, for each module the pipeline line should consider the following requirements:

- Data loading: load preprocessed data within C-PAC pipeline, and saved the phenotypic data.
- Graph construction: use Pytorch to convert the AAL matrices in graphs, store node attributes and edges weights.
- Model: allow configuration of hidden size,number of layers, dropout, learning rate, and epoch count.
- Training: support stratified 5 or 10 k-fold cross validation, and 80-20 train test split.
- Evaluation: sex aware evaluations through accuracy, precision, recall, F1, ROC-AUC.
- Analysis: build visual outputs.

5 Implementation

This section describes the implementation of the designed pipeline and all the tools used. The experiments were run locally on a standard 8GB Macbook.

5.1 Environment

The implementation used the tools and libraries listed bellow:

- Python 3.2;

- Jupyter Notebook;
- Pytorch Geometric; NetworkX;
- Numpy, SciPy, Pandas;
- scikit-learn;
- Matplotlib;

5.2 Implementation of Graph Construction

For each subject the preprocessed time series with 116 ALL ROIs was loaded, and then z-normalized aiming to reduce the differences among all the sites. For every participant, we extracted the average BOLD (Blood-Oxygen-Level-Dependent) signal over time for each of these regions. Pearson correlations were calculated between every pair of brain regions, using numPy. And then correlations were Fisher z-transformed. After all, we only retained the strongest edges (k = 15). This procedure resulted in a 116 x 116 weighted connectivity matrix for each subject. Each previous matrix was converted into a graph using PyTorch, representating the subject brains. Figure 5 shows an example of a graph built.

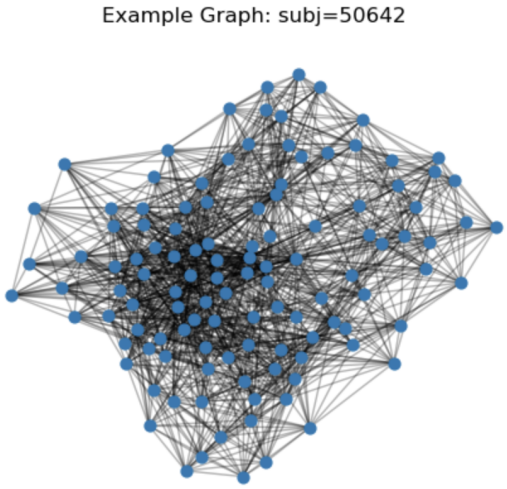

Figure 5: Example Brain Graph

5.3 Model Implementation

Graph Convolutional Network(GCN) and Graph Attention Network (GAT) were firstly used as baseline architecture.

5.3.1 GCN Implementation

GCN was implemented using Pytorch Geometric. It used 2 layers, 32 hidden dimensions, dropout at 0.25 and performed graph convolutions, reLU activation, and global mean pooling over nodes. Figure 6 shows the GCN architecture.

```python
class SimpleGCN(nn.Module):
    """
    Simple 2-layer GCN:
    - conv1: in_dim -> hidden
    - conv2: hidden -> hidden
    - global mean pooling
    - linear layer to 2 classes (ASD vs Control)
    """
    def __init__(self, in_dim, hidden, dropout):
        super().__init__()
        self.conv1 = GCNConv(in_dim, hidden)
        self.conv2 = GCNConv(hidden, hidden)
        self.dropout = nn.Dropout(dropout)
        self.fc = nn.Linear(hidden, 2)

    def forward(self, data):
        x, edge_index = data.x, data.edge_index
        batch = getattr(
            data, "batch",
            torch.zeros(x.size(0), dtype=torch.long, device=x.device)
        )

        # 1st GCN layer
        x = self.conv1(x, edge_index)
        x = F.relu(x)
        x = self.dropout(x)

        # 2nd GCN layer
        x = self.conv2(x, edge_index)
        x = F.relu(x)
        x = self.dropout(x)

        # Global pooling over nodes → one vector per subject
        x = global_mean_pool(x, batch)

        # Final classification
        out = self.fc(x)
        return out
```

Figure 6: GCN Architecture

5.3.2 GAT Implementation

GAT was implemented using Pytorch Geometric. It used multi-head attention layers to improve results over noisy connectivity graphs. The same GCN configuration was kept for comparison: 32 hidden dimensions, dropout at 0.25 Figure 7 shows the GAT architecture.

```python
class SmallGAT(nn.Module):
    def __init__(self, in_dim, hidden, heads, dropout):
        super().__init__()
        self.gat1 = GATConv(in_dim, hidden, heads=heads, concat=True, dropout=dropout)
        self.gat2 = GATConv(hidden * heads, hidden, heads=heads, concat=True, dropout=dropout)
        self.gat3 = GATConv(hidden * heads, hidden, heads=1, concat=False, dropout=dropout)
        self.dropout = nn.Dropout(dropout)
        self.fc = nn.Linear(hidden, 2)

    def forward(self, data):
        x, edge_index = data.x, data.edge_index
        batch = getattr(
            data, "batch",
            torch.zeros(x.size(0), dtype=torch.long, device=x.device)
        )

        x = F.elu(self.gat1(x, edge_index))
        x = self.dropout(x)
        x = F.elu(self.gat2(x, edge_index))
        x = self.dropout(x)
        x = F.elu(self.gat3(x, edge_index))
        x = self.dropout(x)

        x = global_mean_pool(x, batch)
        out = self.fc(x)
        return out
```

Figure 7: GAT Architecture

5.3.3 Training and Validation

A training function was implemented to iterate the batches, compute predictions, apply the cross-validation, and predict probabilities. The cross-validation loop used StratifiedKFold and AdamW optimiser. The settings include a learning rate of 5e-4, and a Weight decay of 1e-4. Figure 8 represents training the architecture.

```python
for epoch in range(1, cfg.epochs + 1):
    tr_loss, _, _ = epoch_run(model, train_loader, optim=optim,
                              device=device, class_weights=cls_w)
    va_loss, y_true, y_prob = epoch_run(model, val_loader, optim=None,
                                        device=device, class_weights=cls_w)
    m = metrics_from_probs(y_true, y_prob)

    if m["auc"] > best_auc + 1e-4:
        best_auc = m["auc"]
        best_state = {k: v.cpu().clone() for k, v in model.state_dict().items()}

    if epoch % 20 == 0 or epoch == 1:
        print(f"[ALL 80% Fold {fold_id}][Epoch {epoch}] "
              f"tr={tr_loss:.4f} va={va_loss:.4f} AUC={m['auc']:.3f}")

if best_state is not None:
    model.load_state_dict(best_state)
_, y_true, y_prob = epoch_run(model, val_loader, optim=None,
                              device=device, class_weights=cls_w)
m = metrics_from_probs(y_true, y_prob)
cv_rows.append({"fold": fold_id, **m})
```

Figure 8: Training architecture

6 Evaluation

This section presents the evaluation of the GNNs in this project. It reports the results of the experiments, compares performance across model architectures, and examines how sex representation within the training data influences ASD classification outcomes.

6.1 Experiment 1: GCN full data

The first case evaluated the two layers GCN trained on the full data set. The training used 80% of the dataset and left 20% for testing. The performance metrics were generated separately for men, women, and all combined. Figure 9 shows the bar chart correspondent to the results. Although the accuracy was modest for all tests, when analysed only for women, the accuracy got even worse. The model achieved 48% overall accuracy, while women performed was 36%, substantially lower than for men 50%. This first result already suggested possible sex-specific differences.

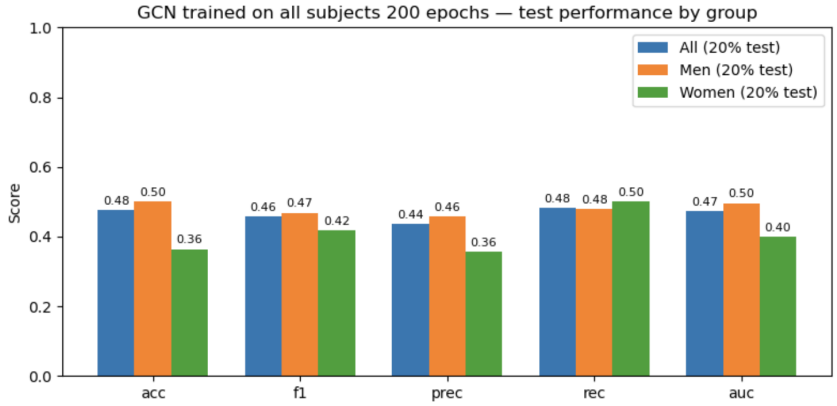

Figure 9: GCN full data

6.2 Experiment 2: GAT full data

The same 80-20 experimental setup from experiment 1 was used for GAT. Figure 10 shows the GAT model performance. It is observed that the accuracy improved related to previous GCN model, however the model still show a gap of 9% in performance when related to men. The AUC 57% indicate that when used attention, the model learns some signal.

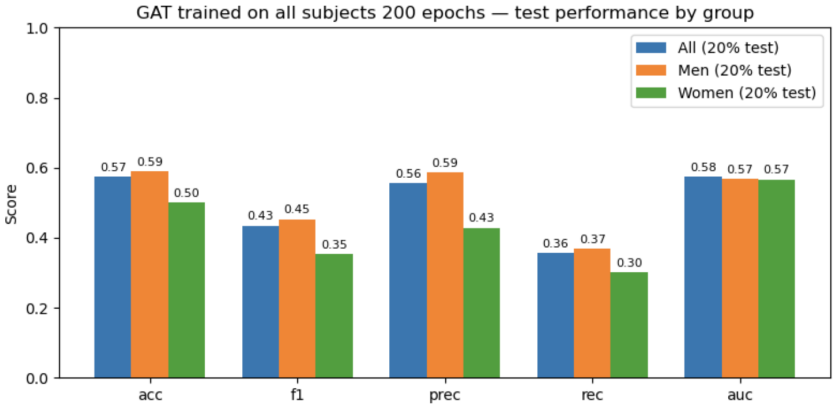

Figure 10: GAT full data

6.3 Comparison GCN and GAT

A comparison between GCN and GAT was made in order to decide which model would be applied for next experiments. Table 2 shows all the results obtained. Both models showed disparity between subjects, with the models performing better for men. However, GAT outperformed GCN in the overall results for this configuration. While GCN had an accuracy of 48%, the GAT had 57%. In this way, GAT was chosen for the following study.

Model	Group	ACC	F1	Precision	Recall	AUC
GCN	Women	0.36	0.42	0.36	**0.50**	0.40
	Men	0.50	**0.47**	0.46	0.48	0.50
	All	0.48	0.46	0.44	0.48	0.47
GAT	Women	0.50	0.35	0.43	0.30	0.57
	Men	**0.59**	0.45	**0.59**	0.37	0.57
	All	0.57	0.43	0.56	0.36	**0.58**

Table 2: Comparison between GCN and GAT

6.4 Experiment 3: GAT trained on male subjects

For this experiment, the model was trained only on men subjects using 5k fold, and women subjects were hold for external test. The overall accuracy dropped 11% when tested for women. The model can find 67% of autistic men cases, but for only 57% of women cases. This finds might indicate that the ASD male brain patters not transfer well for for female. And that female subjects might be underdiagnosed.

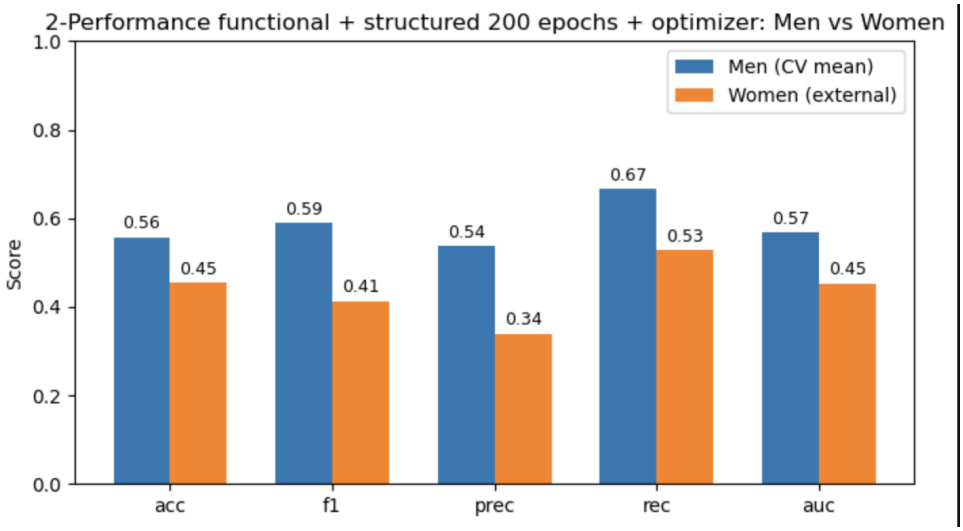

Figure 11: Men trained GAT

6.5 Experiment 4: GAT trained on female subjects

For this experiment, the model was trained only on 80% of the women subjects available using 5K fold, then tested on 20% and on men. Figure 12 shows the model performance.

The model had a strong performance for female subjects, within the overall accuracy of 70% and the recall of 71%. On the other hand, the recall tested on men subjects dropped to 34%. Those finds might indicate that female autistic brain patterns are not likely a subset of male patterns.

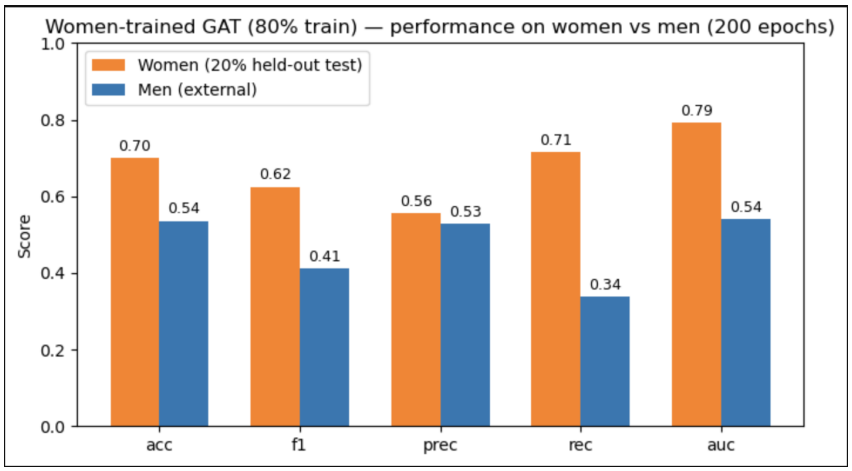

Figure 12: Women trained GAT

6.6 Experiment 5: GAT trained on balanced dataset 5k-fold (100 male + 99 female)

For this experiment, we selected 100 men subjects and joint it with the 99 women subjects, aiming to investigate the model with reduced sex imbalance. This experiment allow to observe the performance when both sexes contribute equally on training. The accuracy remained randomly comparable for both sexes. Women had a lower precision (20%) than men (45%). Unlike earlier experiments where men consistently outperformed women, the balanced scenario shows reduced male performance. This finds might suggest that in balanced datasets the learned feature space shifts, making male classification weaker but allowing women's classification to stabilise. This supports the hypothesis that male dominant models implicitly in favour of male autistic brain patterns.

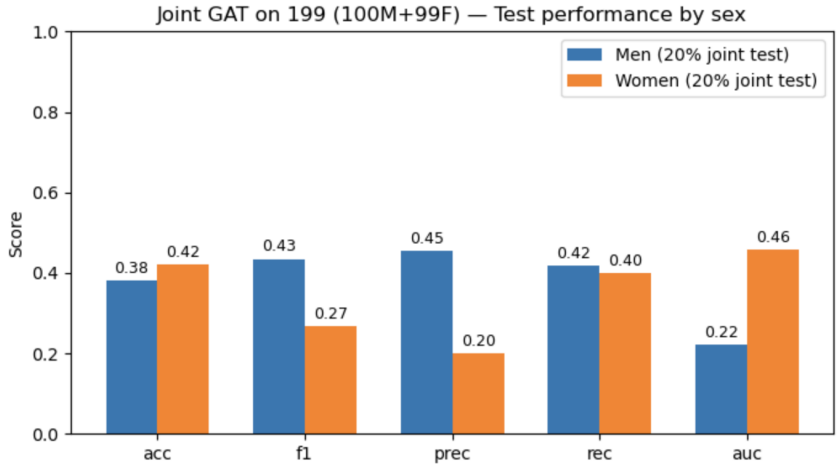

Figure 13: Balanced dataset 5k fold

6.7 Experiment 6: Exploratory GAT trained on balanced dataset 10k-fold (100 male + 99 female)

For this experiment, instead of 5k fold, we used 10k fold on the set of 100 men subjects and 99 subjects. Figure 14 shows the model performance. The model had a significant increase in the overall performance for women, achieving 80% overall accuracy, which most likely is due to the tiny fold sets. The men accuracy also drops, confirming the instability. These results were treated as exploratory only due to instability of the number of subjects per fold. Although the results still might suggest that women specific features become more learnable when women are equally represented, they are not reliable for conclusions.

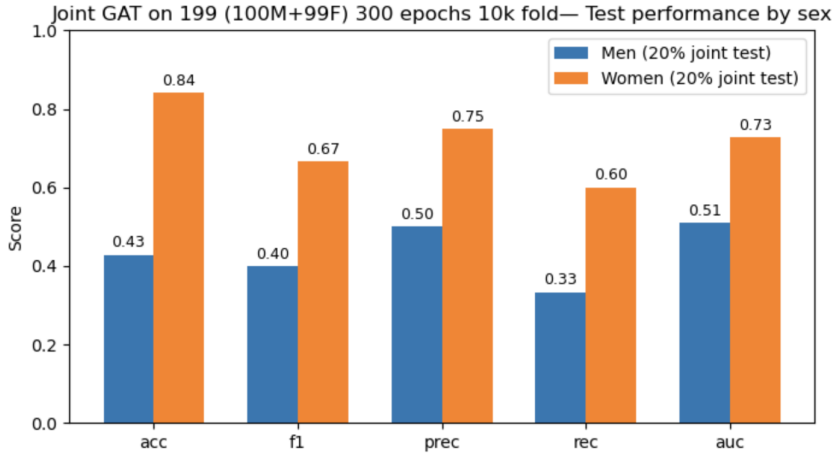

Figure 14: Balanced dataset 10k fold

6.8 Discussion

The table 3 summarize all the results of GAT experiments. It is observed that across all tests, models classify ASD more accurately in men than in women.

Experiment	Group	ACC	F1	Precision	Recall	AUC
General 80/20	Women	0.50	0.35	0.43	0.30	0.57
	Men	0.59	0.45	0.59	0.37	0.57
Men-only	Women	0.45	0.41	0.34	0.53	0.45
	Men	0.56	0.59	0.54	0.67	0.57
Women-only	Women	0.70	0.62	0.56	0.71	0.79
	Men	0.54	0.41	0.53	0.34	0.54
Balanced 5F	Women	0.42	0.27	0.20	0.40	0.46
	Men	0.38	0.43	0.45	0.42	0.22
Balanced 10F (Exploratory)	Women	0.84	0.67	0.75	0.60	0.73
	Men	0.43	0.40	0.50	0.33	0.51

Table 3: Summary of all GAT experiments

When analysed general models without sex awareness, as in mostly of current studies, models tends in favour of male subjects. The accuracy of 59% for men against 50% for women suggest that the model is implicitly learning a male-biased representation of ASD.

When analysed men-only trained models, the model underperform on women, dropping the accuracy to 45% and the precission to 34%. This find strongly support the hypothesis that ASD manifests differently in female brain connectivity.

When analysed women-only trained models, the model performs well on women but poorly on men. This indicates that female ASD patterns are distinct and not equivalent to male patterns.

When analysed subset balanced experiments, they confirm that representation matters in the ASD classification. In 5K-fold, men and women achieve similar accuracy. In 10K-fold, women outperform men (84% accuracy), although those are exploratory finds due to small folds.

Finally we reinforce that overall performance of the models are modest and previous expected. This happened mainly due to limited number of female subjects, multisite variability from dataset, and hardware constraints. Nevertheless, the sex pattern remains extremely consistent, strengthening the study's conclusions.

7 Conclusion and Future Work

This study aimed to investigate whether Graph Neural Network (GNN) models classify Autism Spectrum Disorder (ASD) differently in males and females, and whether women may require sex-specific modelling approaches. Using the ABIDE I dataset, connectivity graphs were constructed from fMRI and analysed through GCN and GAT models in diverse experimental scenarios. The results consistently showed that classification models behave differently depending on the sex composition of the training data, and that women are generally harder to classify under standard modelling pipelines.

The comparison between GCN and GAT demonstrated that GAT performs better for this study configuration, although both models achieved modest accuracies. These performances, however, are aligned with several recent studies using ABIDE data, where multisite variability and heterogeneous ASD presentations make the classification problem challenging. Therefore, model performance for women was consistently lower than for men in the general experiments. This indicates that mixed-sex models may implicitly learn representations that in favour of male ASD patterns.

The men-only trained model reinforces the bias towards males. Although it performed reasonably on male subjects, its performance dropped significantly when tested on women. This demonstrates that features learned from male data do not generalise well to female data, supporting the idea that female ASD may present differently at the connectivity level. The women-only model, despite the extremely limited sample size, performed relatively well on women but poorly on men. This mirrored the previous finding that training exclusively on one sex produces a model that is specialised to that sex and cannot generalise across groups.

Overall, the results across all scenarios consistently reveal that women are either under-represented or poorly modelled in typical ASD classification workflows. Even when women subjects are included in training, the performance does not achieve same level as for men subjects. The female ASD brain connectivity cannot be adequately captured by models trained on predominantly male datasets. These findings support the hypothesis that diagnostic tools built on male dominated datasets may lead to even more cases of underdiagnostic female individuals. The results also emphasise the broader point that machine learning models in neuroscience must account for demographic subgroups rather

than assuming uniform patterns across populations.

This study also highlights directions for future research, particularly concerning sex-specific modelling for ASD: larger sex balanced datasets are essential for improving model stability and reliability; techniques such as data augmentation and graph generative models (GraphVAE, GraphGAN) could artificially increase the dataset size and produce more extensive tests.

References

Cardenas-Hernandez, N. A., Perez-Diaz, M., Batista, K. and Valdés Hernández, M. (2022). Autism spectrum disorder detection with diffusion tensor imaging and machine learning, *Available at SSRN* (5272519).

Craddock, C., Benhajali, Y., Chu, C., Chouinard, F., Evans, A., Jakab, A., Khundrakpam, B. S., Lewis, J. D., Li, Q., Milham, M., Yan, C. and Bellec, P. (2013). The neuro bureau preprocessing initiative: open sharing of preprocessed neuroimaging data and derivatives, *Neuroinformatics*, Stockholm, Sweden.

Di Martino, A., Yan, C.-G., Li, Q., Denio, E., Castellanos, F. X., Alaerts, K., Anderson, J. S., Assaf, M., Bookheimer, S. Y., Dapretto, M. et al. (2014). The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism, *Molecular Psychiatry* **19**(6): 659–667.

Harmens, M., Sedgewick, F. and Hobson, H. (2022). Autistic women's diagnostic experiences: Interactions with identity and impacts on well-being, *Women's Health* **18**: 17455057221137477.

Heo, J. S., Yang, S.-W., Lee, S., Lee, K.-S. and Ahn, K. H. (2025). Sex-based differences in prenatal and perinatal predictors of autism spectrum disorder using machine learning with national health data, *Autism Research* .

Hu, J., Cao, L., Li, T., Dong, S. and Li, P. (2021). Gat-li: a graph attention network based learning and interpreting method for functional brain network classification, *BMC Bioinformatics* **22**(1).

Jiang, X., Shou, X.-J., Zhao, Z., Chen, Y., Meng, F.-C., Le, J., Song, T.-J., Xu, X.-J., Guo, W., Ke, X. et al. (2023). A brain structural connectivity biomarker for autism spectrum disorder diagnosis in early childhood, *Psychoradiology* **3**: kkad005.

Lai, M.-C., Lerch, J. P., Floris, D. L., Ruigrok, A. N., Pohl, A., Lombardo, M. V. and Baron-Cohen, S. (2017). Imaging sex/gender and autism in the brain: Etiological implications, *Journal of Neuroscience Research* **95**(1-2): 380–397.

Lawrence, K. E., Hernandez, L. M., Bowman, H. C., Padgaonkar, N. T., Fuster, E., Jack, A., Aylward, E., Gaab, N., Van Horn, J. D., Bernier, R. A. et al. (2020). Sex differences in functional connectivity of the salience, default mode, and central executive networks in youth with asd, *Cerebral Cortex* **30**(9): 5107–5120.

Li, Y., Li, R., Wang, N., Gu, J. and Gao, J. (2023). Gender effects on autism spectrum disorder: a multi-site resting-state functional magnetic resonance imaging study of transcriptome-neuroimaging, *Frontiers in Neuroscience* **17**: 1203690.

Loomes, R., Hull, L. and Mandy, W. P. L. (2017). What is the male-to-female ratio in autism spectrum disorder? a systematic review and meta-analysis, *Journal of the American Academy of Child & Adolescent Psychiatry* **56**(6): 466–474.

McCrossin, R. (2022). Finding the true number of females with autistic spectrum disorder by estimating the biases in initial recognition and clinical diagnosis, *Children* **9**(2): 272.

Organization for Autism Research (2018). The autism dilemma for women diagnosis, `https://researchautism.org/audience/research/the-autism-dilemma-for-women-diagnosis/`. Accessed: 2025-07-30.

Pan, W., Ling, G. and Liu, F. (2025). Mgnn-bw: Multi-scale graph neural network based on biased random walk path aggregation for asd diagnosis, *IEEE Transactions on Neural Systems and Rehabilitation Engineering* .

Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L. and Petersen, S. E. (2012). Spurious but systematic correlations in functional connectivity mri networks arise from subject motion, *NeuroImage* **59**(3): 2142–2154.

Robinson, E. B., Lichtenstein, P., Anckarsäter, H., Happé, F. and Ronald, A. (2013). Examining and interpreting the female protective effect against autistic behavior, *Proceedings of the National Academy of Sciences* **110**(13): 5258–5262.

Simply Psychology (2025). Can autistic women mask without realizing it?, `https://www.simplypsychology.org/can-autistic-women-mask-without-realizing-it.html`. Accessed: 2025-07-30.

Subah, F. Z., Deb, K., Dhar, P. K. and Koshiba, T. (2021). A deep learning approach to predict autism spectrum disorder using multisite resting-state fmri, *Applied Sciences* **11**(8): 3636.

Supekar, K., de Los Angeles, C., Ryali, S., Cao, K., Ma, T. and Menon, V. (2022). Deep learning identifies robust gender differences in functional brain organization and their dissociable links to clinical symptoms in autism, *The British Journal of Psychiatry* **220**(4): 202–209.

Tang, J., Chen, J., Hu, M., Hu, Y., Zhang, Z. and Xiao, L. (2024). Diagnosis of autism spectrum disorder (asd) by dynamic functional connectivity using gnn-lstm, *Sensors* **25**(1): 156.

Trafton, A. (2022). Studies of autism tend to exclude women, researchers find, `https://news.mit.edu/2022/studies-autism-women-bias-0908`. Accessed: 2025-07-30.

Walker, L., Gozzi, M., Lenroot, R., Thurm, A., Behseta, B., Swedo, S. and Pierpaoli, C. (2012). Diffusion tensor imaging in young children with autism: biological effects and potential confounds, *Biological Psychiatry* **72**(12): 1043–1051.

Wang, Z., Xu, Y., Peng, D., Gao, J. and Lu, F. (2023). Brain functional activity-based classification of autism spectrum disorder using an attention-based graph neural network combined with gene expression, *Cerebral Cortex* **33**(10): 6407–6419.

Yang, C., Wang, P., Tan, J., Liu, Q. and Li, X. (2021). Autism spectrum disorder diagnosis using graph attention network based on spatial-constrained sparse functional brain networks, *Computers in Biology and Medicine* **139**: 104963.