Identification of Autism Spectrum Disorder Using AI Techniques

Sabrine Bendimerad
Master's Student in Computational Neuroscience and Neuroengineering
Supervisors: Sabir Jacquir and Arthur Tenenhaus
CentraleSupélec
université PARIS-SACLAY

1.Introduction and Motivation

- Autism Spectrum Disorder (ASD) affects social, communication, and behavioral skills, with 1 in 100 children impacted globally.
- Current diagnostic methods rely on subjective assessments, leading to variability and delays.
- Functional MRI (fMRI) combined with Machine Learning (ML) and Deep Learning (DL) offers a promising approach for efficient ASD diagnosis[1].
- Challenges include small datasets, site variability, underutilized phenotypic data, and limited ability to capture global brain context[2].
- **This project establishes baselines with existing models and aims to explore generative AI for data augmentation and Vision Transformers for classification as advanced approaches during the internship.**

2.Data

- Utilized the autism brain imaging data exchange (ABIDE)dataset, consisting of 1,035 fMRI scans from 17 international sites.

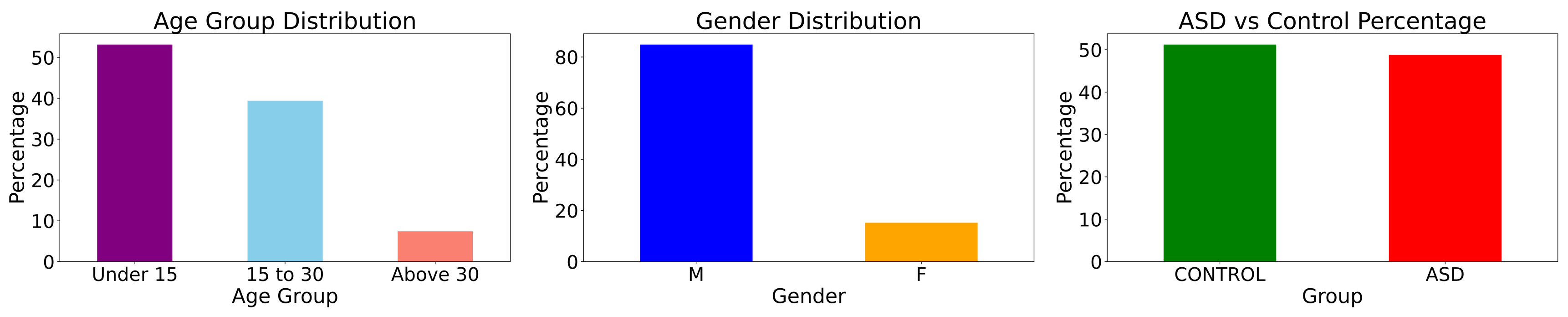

Figure 1: Age, Gender and Asd VS Control Distributions

- fMRI data is provided in 1D files representing BOLD signal time-series for multiple brain regions defined by six atlases: CC200, AAL, HO, EZ, TT, and Dosenbach160.

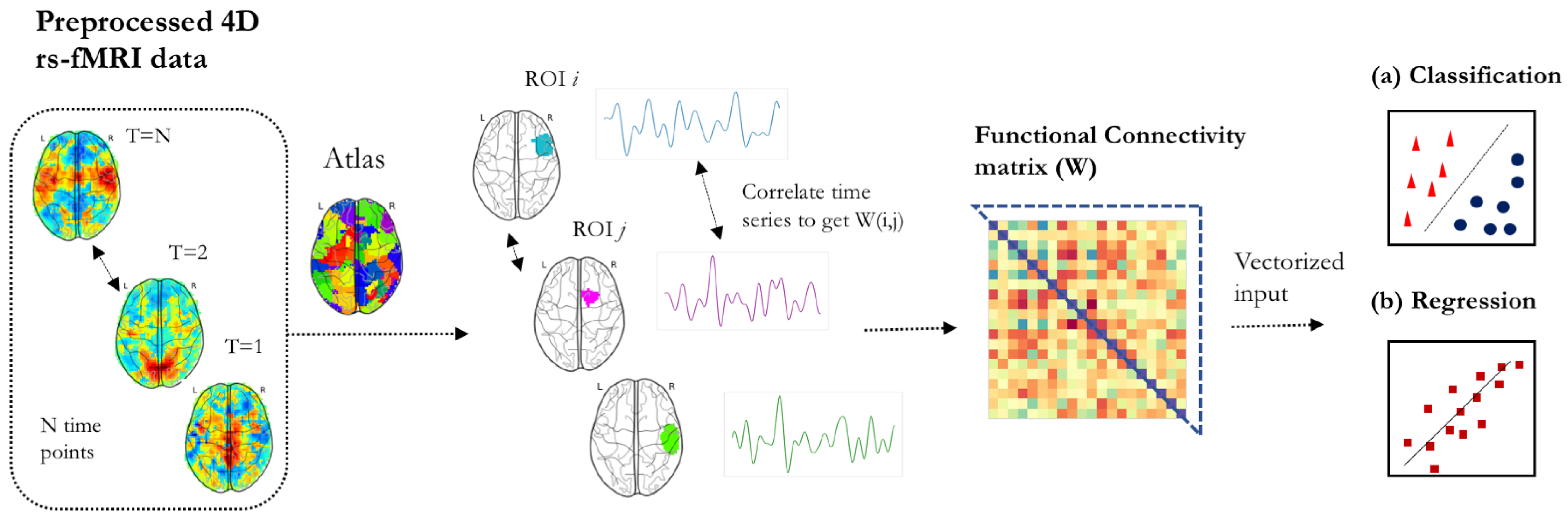

Figure 2: Workflow for Functional Connectivity Matrix Generation and examples of ML process

3.Methodology

- Generated connectivity matrices to represent functional connectivity between brain regions.
- Normalized phenotypic data (age, sex, IQ).
- Conducted exploratory data analysis to address missing data and assess class distributions.
- Split data into training, validation, and test sets, ensuring balanced representation across sites.
- Built Machine Learning models including Logistic Regression, Random Forest, and XGBoost.
- Developed Deep Learning models inclusing Multi Layer PErceptron(MLP), COnvolution Neural Networks (CNNs), and Autoencoders.
- Optimized hyperparameters using grid search and cross-validation.
- Evaluated the best-performing model on unseen test data.

4.Results

Machine Learning Results (Accuracy %)				
Setup	**+Phenotypic**	**-Phenotypic**	**Male**	**NYU**
Logistic Regression	68	65	67	66
Random Forest	74	70	72	73
XGBoost	78	72	75	76

Table 1: Machine Learning Results

Deep Learning Results (Accuracy %)				
Setup	**+Phenotypic**	**-Phenotypic**	**Male**	**NYU**
CNN	74	70	74	74
MLP	76	73	77	78
Autoencoder	82	78	83	84

Table 2: Deep Learning Results

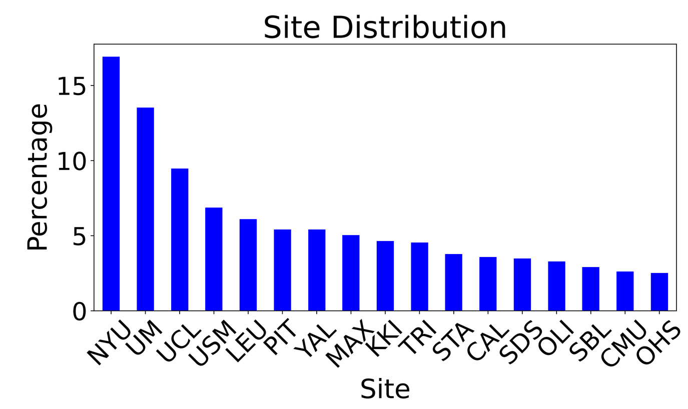

Figure 3: Site-wise sample distribution in ABIDE

- Including phenotypic data improved accuracy, with the highest gains observed in Autoencoders (+4-6%).
- Deep learning models consistently outperformed machine learning methods, demonstrating the ability to capture complex fMRI patterns.
- Site-specific experiments with New York University (NYU), the most representative site in the dataset, showed slightly better performance.
- Male-only datasets had marginally reduced accuracy, suggesting potential bias or missing diversity in features.
- Random Forest and XGBoost were strong traditional ML performers, but they plateaued compared to CNNs and Autoencoders.
- Autoencoders excelled in capturing non-linear and high-dimensional patterns in connectivity matrices.

5.Limitations and Future Work

Limitations:

- Connectivity matrices simplify data but risk losing critical spatial patterns.
- Dataset heterogeneity and small size limit the potential for achieving higher accuracy.
- Existing models struggle to effectively capture both temporal and spatial information[3].

Future Work:

- Leverage Generative Adversarial networks (GANs) for neuroimaging-specific data augmentation to enhance dataset diversity.
- Implement Vision Transformers (ViTs) and hybrid CNN-ViT models for more comprehensive feature extraction.
- Combine sMRI and fMRI data for a multi-modal approach.

6.References

1. Huo, Y., et al. "3D Whole Brain Segmentation Using Spatially Localized Atlas Network Tiles." *NeuroImage*, 194 (2019): 105-119.
2. Nguyen, H.-D., et al. "Towards Better Interpretable and Generalizable AD Detection Using Collective Artificial Intelligence." *Computerized Medical Imaging and Graphics*, 104 (2023): 102171.
3. Liu, X., et al. "MADE-for-ASD: A Multi-Atlas Deep Ensemble Network for Diagnosing Autism Spectrum Disorder." *Computers in Biology and Medicine*, 164 (2024): 106036.