



# WiDS Datathon

## ML-Powered ADHD Predictor

Sex-Specific Brain Connectivity Patterns for ADHD Prediction using Machine Learning

**Presenter:**

Kunal Prasad | BCA | CSIT |

Final Year PPT, Guru Ghasidas Vishwavidyalaya | May 2025 |

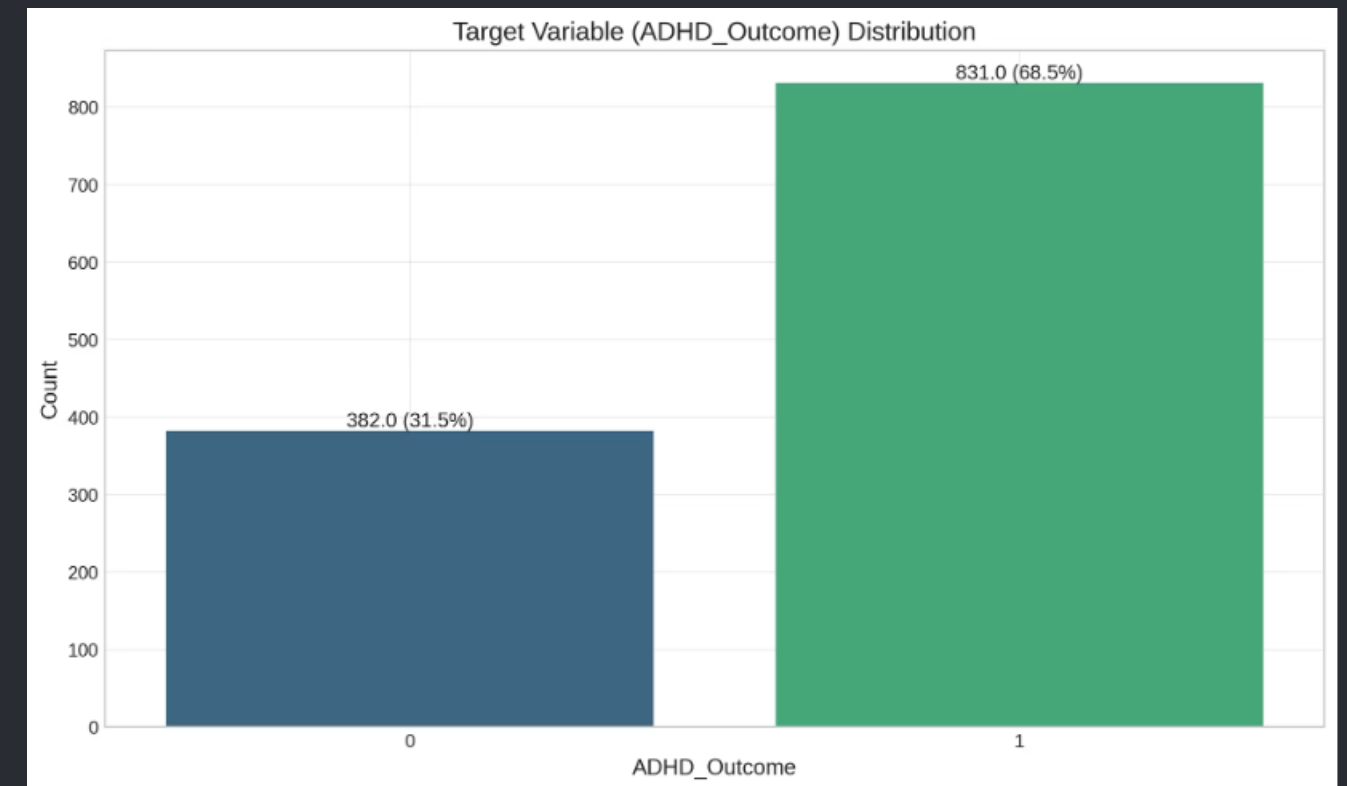
# ADHD and Research Context

## ADHD Prevalence

- 11% adolescents affected
- 14% boys, 8% girls
- Females often underdiagnosed

## WiDS Datathon 2025

Using fMRI and ML to address diagnostic gaps



# Dataset Overview

## Source

Healthy Brain Network (HBN)

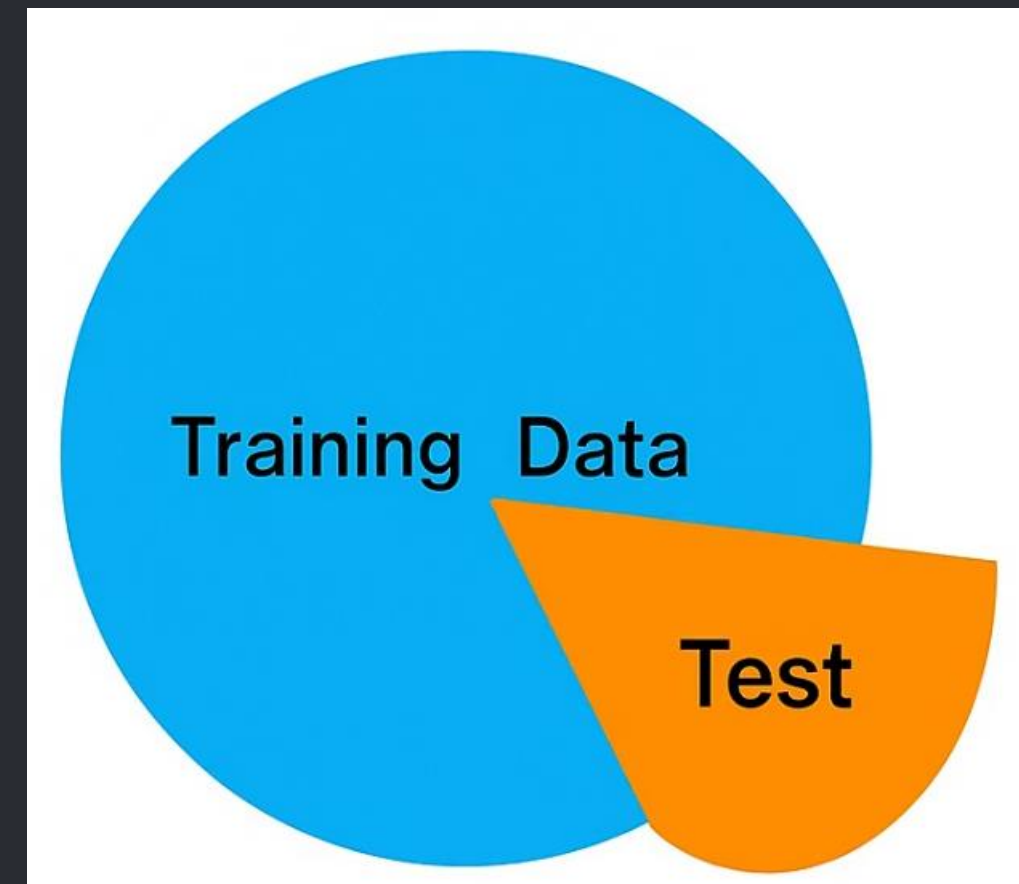
## Data

- fMRI connectomes: 36x36 matrices, 630 features
- Metadata: Age, SDQ scores, parent education

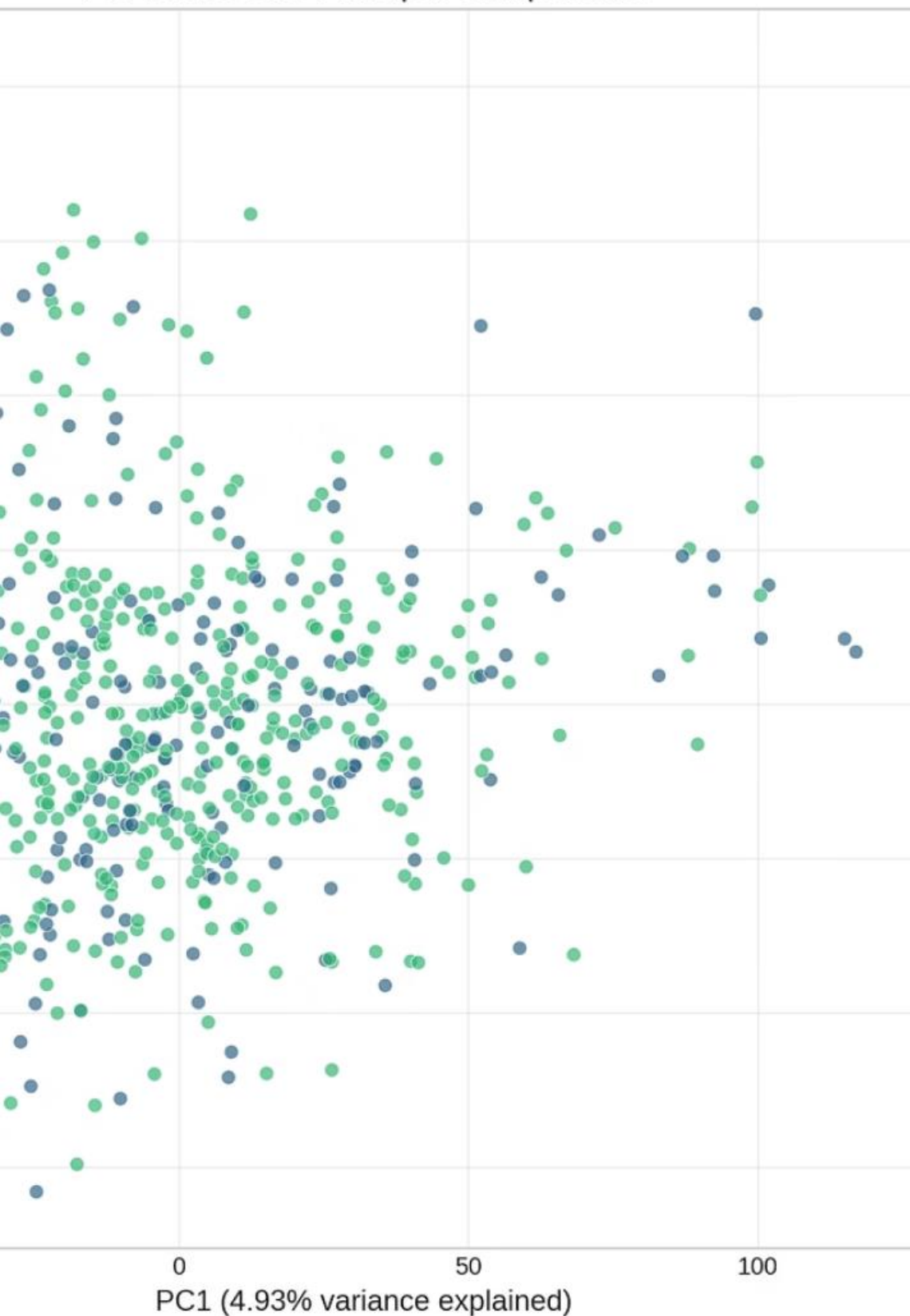
row	1213
columns	19931

## Subjects

~1,200 train; ~300 test



PCA: First Two Principal Components

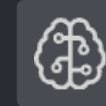


# Data Preprocessing



fMRI

PCA with ~50 components,  
95% variance



Metadata

Interaction terms, median  
imputation



Exploratory Data Analysis

14% ADHD males, 8% females; SDQ\_Hyperactivity  $r=0.42$

Data → Preprocessing → Feature Engineering → Model Training → Evaluation

# Methodology

## Models

Logistic Regression, XGBoost, Random Forest, stacking ensemble

## Training

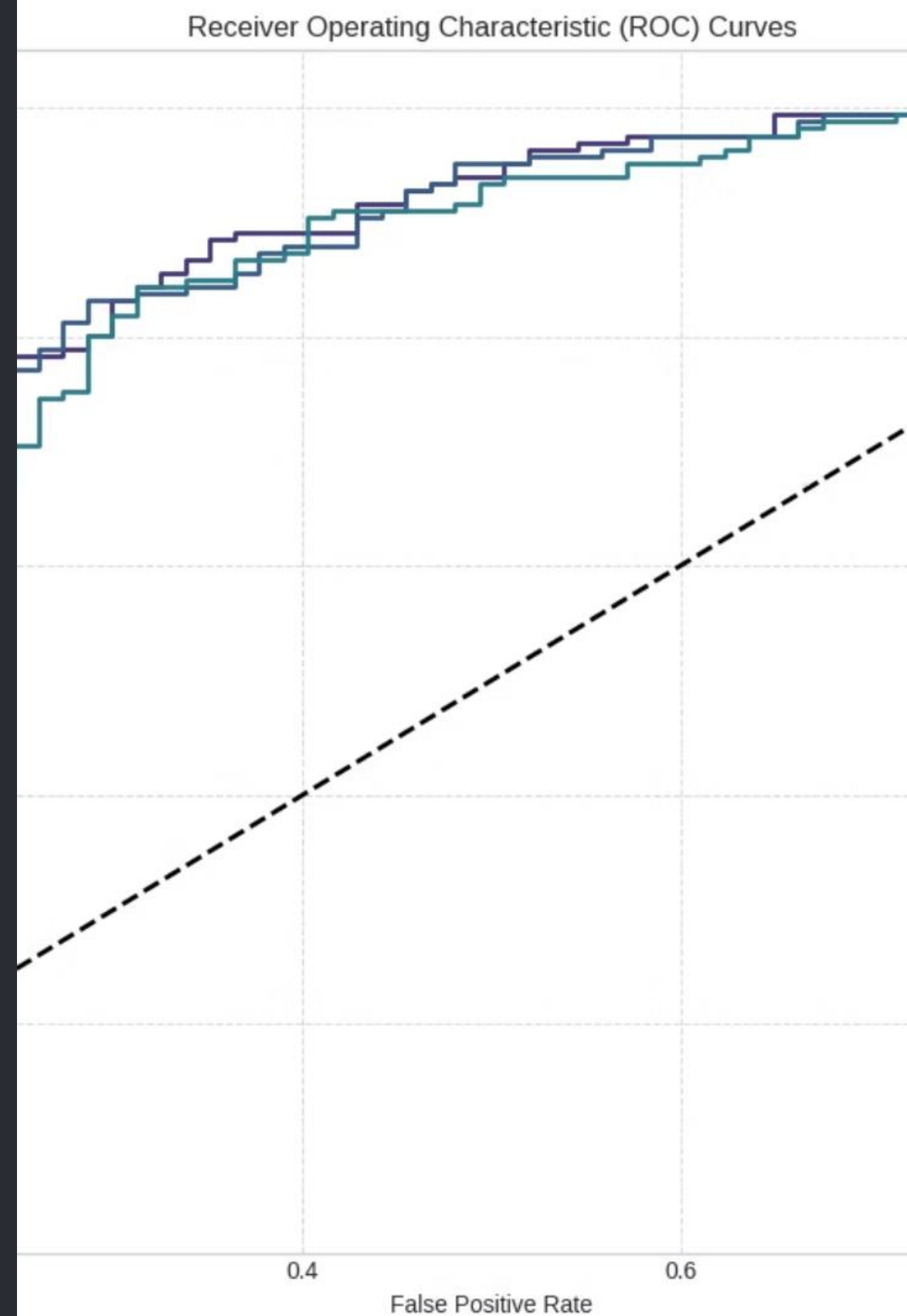
5-fold stratified CV, GridSearchCV

## Metrics

ROC-AUC, recall, precision

## Interpretability

SHAP values for feature importance



# Model Performance

Model	ROC-AUC	Recall
Logistic Regression	0.8668	0.9639
Ensemble	0.8683	0.8977

## Ensemble ROC-AUC

0.89 ± 0.02

## Logistic Regression Recall

96.4%

## Precision & Recall by Sex

- Males: 85% precision
- Females: 74% recall

# Key Findings

**DLPFC Connectivity**  
Top predictor (+0.21 SHAP)

**SDQ Hyperactivity**  
Important predictor (+0.18 SHAP)

**DMN**  
Critical for ADHD diagnosis

**Sex Differences**

- **Females: Amygdala**
- **Males: Cerebellar**



# Limitations & Future Work

## Future

- Multi-output ADHD+sex model.
- EEG diagnostics.
- Diverse datasets.

## Limitations

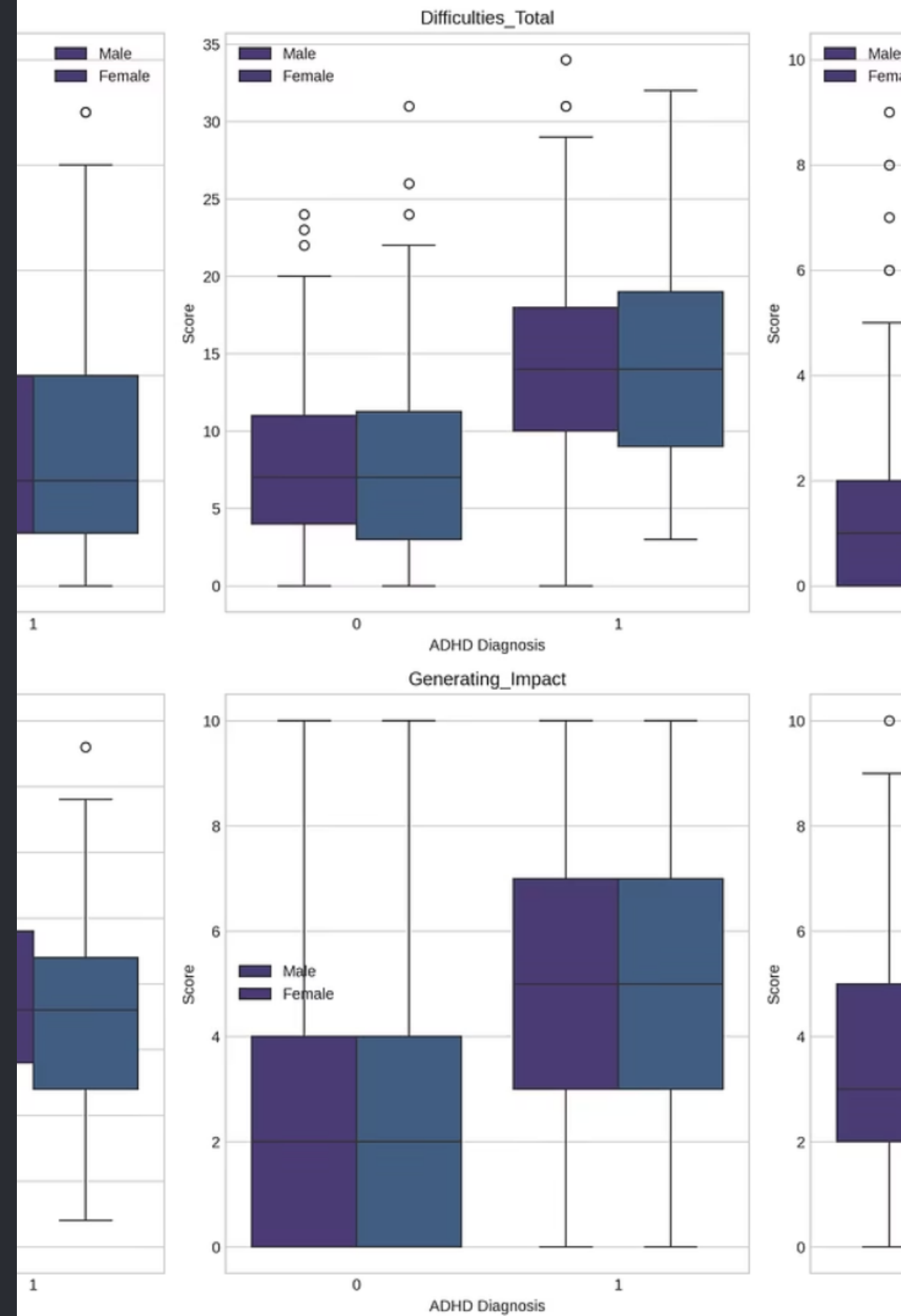
- Small sample (~1,200).
- fMRI cost (~\$500/scan).
- HBN data bias (severe cases).



# Conclusion

- Identified DLPFC, amygdala as key ADHD biomarkers.
- Improved female diagnosis via sex-specific patterns.
- Open-source pipeline advances neuroscience-ML integration.

“This work contributes to equitable and interpretable ADHD diagnostics by combining domain-specific neuroscience with ML pipelines”



Thank You



# Thank you for your attention!

Questions welcome!

- I sincerely thank | Dr. Babita Majhi, Assistant Professor | for her invaluable guidance and support throughout this project.