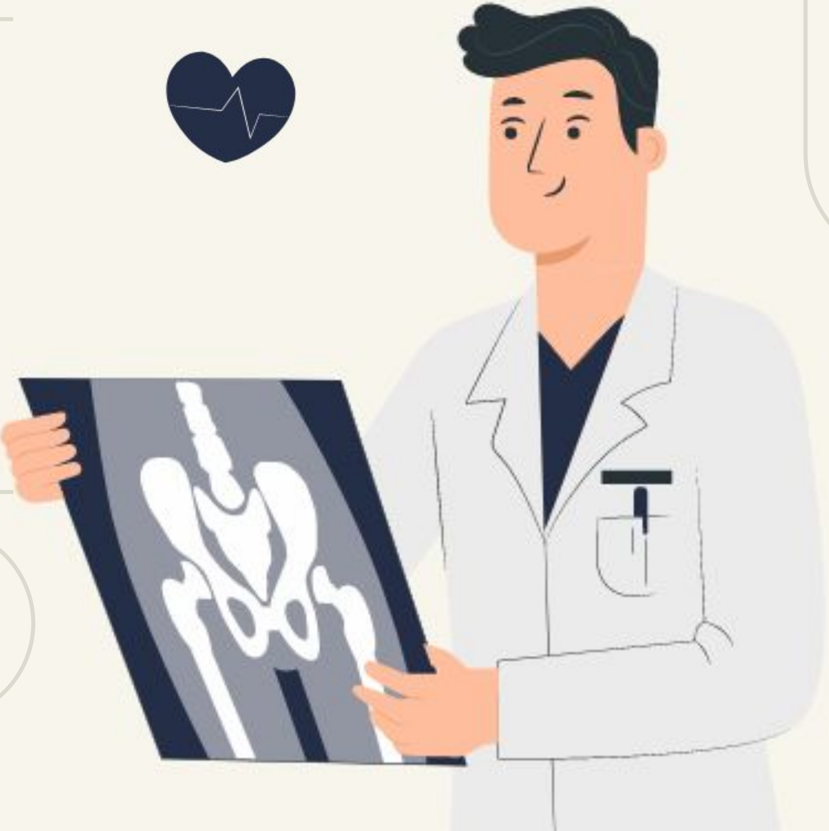
An Explainable AI Approach to ADHD and Gender Classification



Samia Rahman Misty
Registration No: 2400570
Course No: CE- 888

Presentation Demo : Zoom Link

Introduction

What is ADHD?

A neurodevelopmental disorder affecting attention, impulse control and activity levels.

How Dangerous is ADHD?

If untreated, can lead to academic, relationship issues, low self-esteem, and mental health risks.

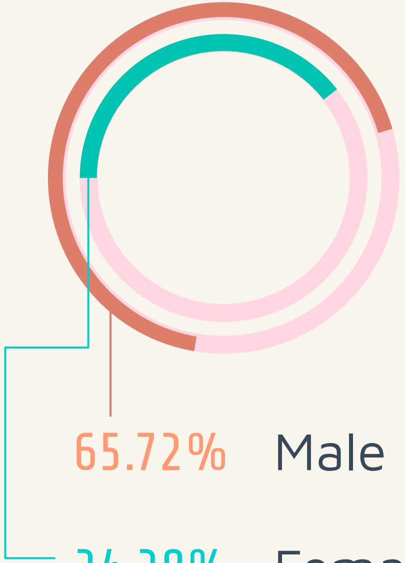
How Deep Learning Helps

- Analyzes complex data (e.g., fMRI, behavior) for accurate ADHD diagnosis.
- Example: Deep learning achieved 70% accuracy in ADHD diagnosis using fMRI data[1].

[1] Mao, Z., Su, Y., Xu, G., Wang, X., Huang, Y., Yue, W., ... & Xiong, N. (2019). Spatio-temporal deep learning method for adhd fmri classification. Information Sciences, 499, 1-11.

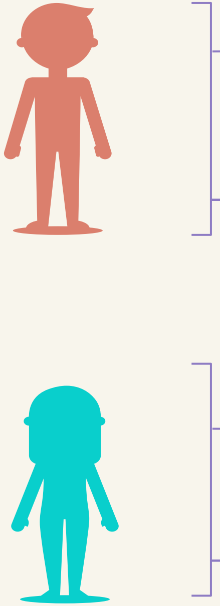
Dataset

GENDER



65.72% Male

34.28% Female



401 ADHD

157 Non-ADHD

174 ADHD

117 Non-ADHD

Objective

Is it possible to build a model to predict both an individual's sex and their ADHD diagnosis?

Is the model biased in its predictions or training data?

Can the model misdiagnose females due to bias?

Data Integrity and Optimization

Data Format Standardization

Variations in data formats and types were observed during collection. for example, the Edinburgh Handedness Questionnaire was aligned with the standard format[1].

fMRI Data Handling

The fMRI brain imaging data was large and complex. It was processed using appropriate techniques to ensure efficiency and maintain data integrity.

Missing Data

Some data entries were missing during collection, which were carefully handled through appropriate data imputation techniques.

Feature Redundancy

In the Strength and Difficulties Questionnaire, the Total Difficulties Score, Externalizing Score, and Hyperactivity Scale were highly correlated. To avoid redundancy, only one of these was retained.

[1] https://www.brainmapping.org/shared/Edinburgh.php

Selected Models

01

Logistic Regression

Its simplicity reduces the risk of errors when working with limited data.

02

CNN + MLP Hybrid Architecture

1.CNN is effective at finding patterns in numbers, especially when those numbers are arranged in a meaningful order, like images or signals.

2.MLP part is good at understanding group-based information, such as age, gender, or other categories.

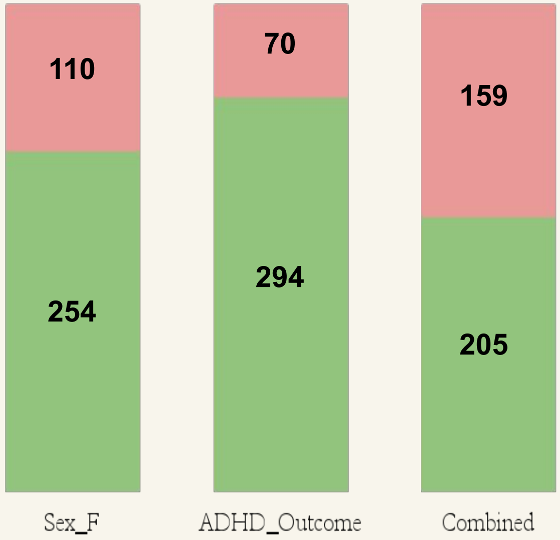
03

CNN + ANN Hybrid Architecture

1.CNNs are effective at recognizing patterns in data that have a natural order or structure, helping to make sense of complex numerical information.

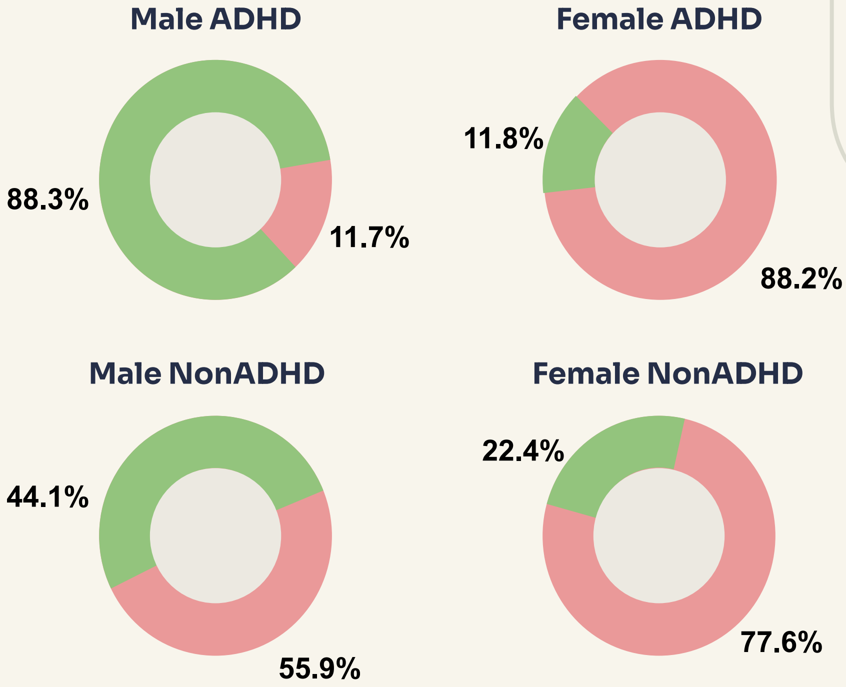
2.ANNs offer a more flexible way to understand grouped or labeled information, like age or gender, compared to simpler models.

Logistic Regression Prediction



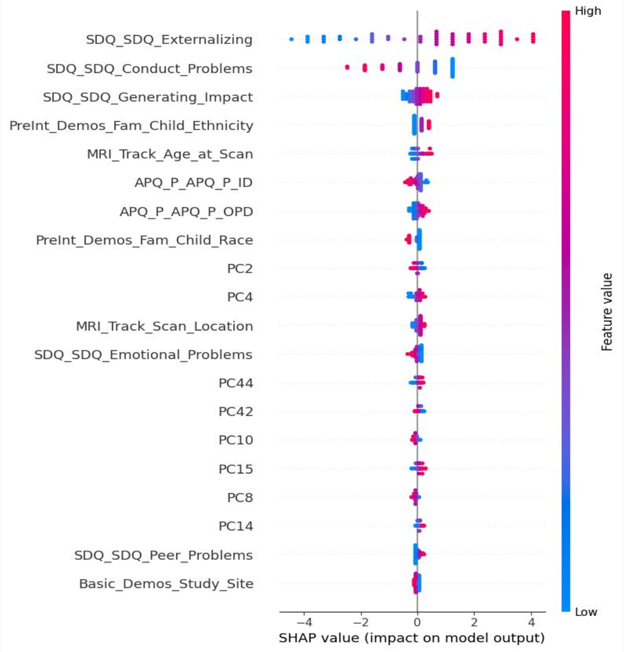
Correct vs Incorrect Predictions Per Label

🟩 Correct 🟥 Incorrect

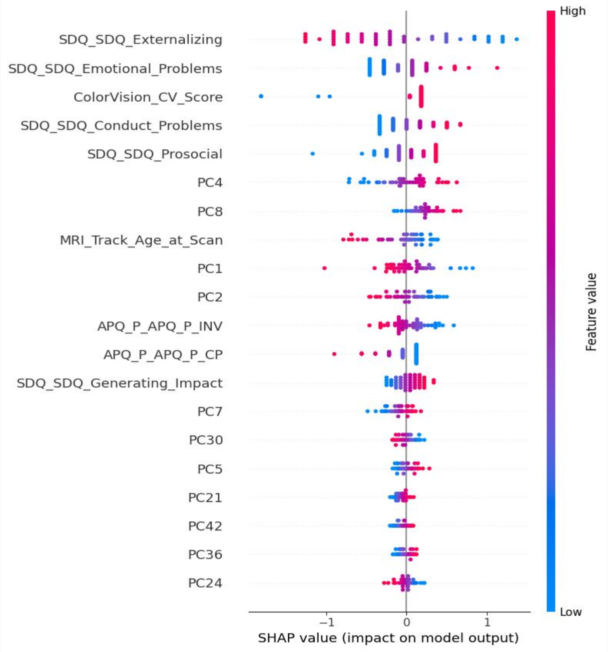


Correct vs Incorrect Predictions Per Category

Logistic Regression SHAP Analysis

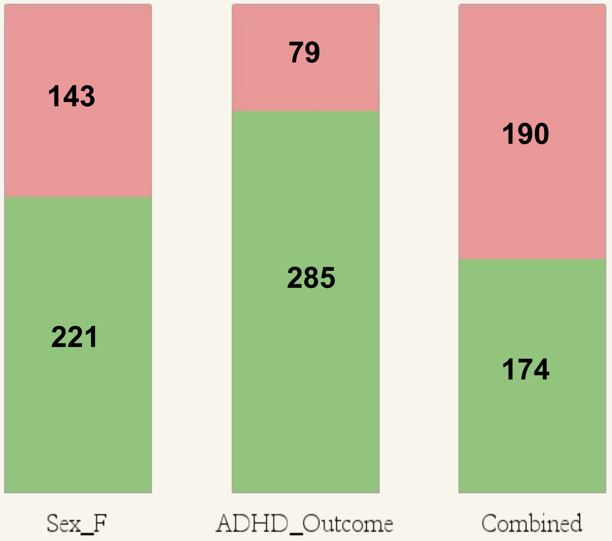


SHAP Analysis for Sex F



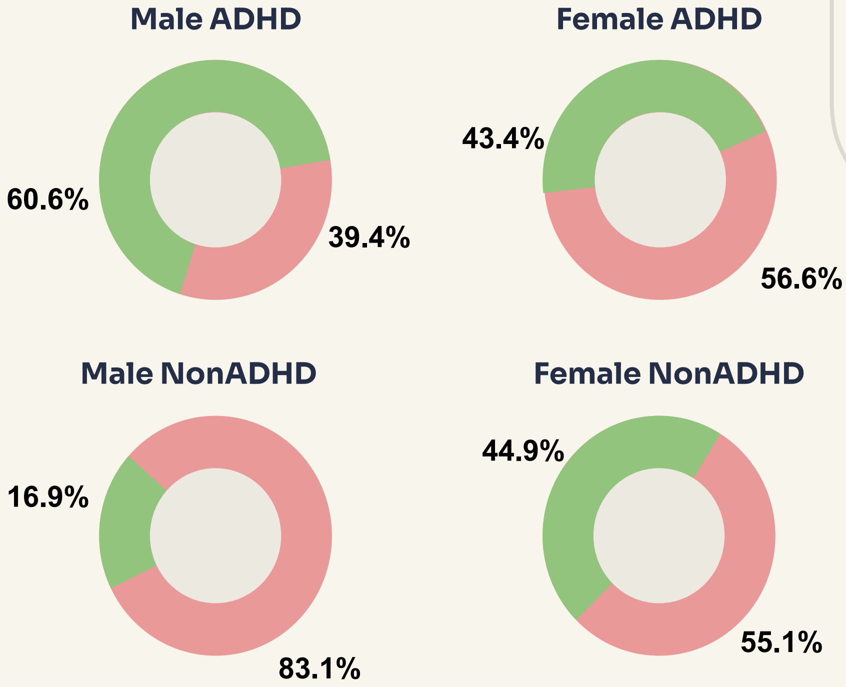
SHAP Analysis for ADHD Outcome

CNN + MLP Hybrid Architecture Prediction



Correct vs Incorrect Predictions Per Label

🟩 Correct 🟥 Incorrect


Male ADHD

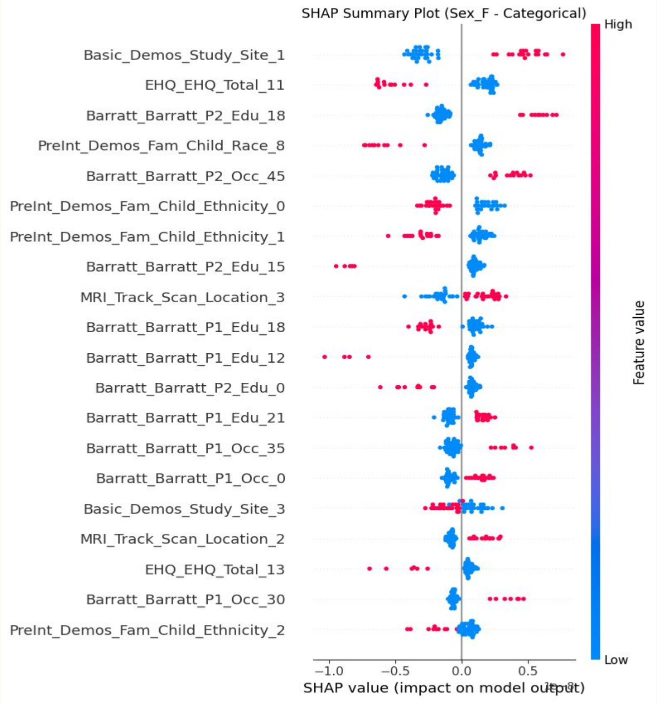
Female ADHD

Male NonADHD

Female NonADHD

Correct vs Incorrect Predictions Per Category

CNN + MLP Hybrid Architecture SHAP Analysis

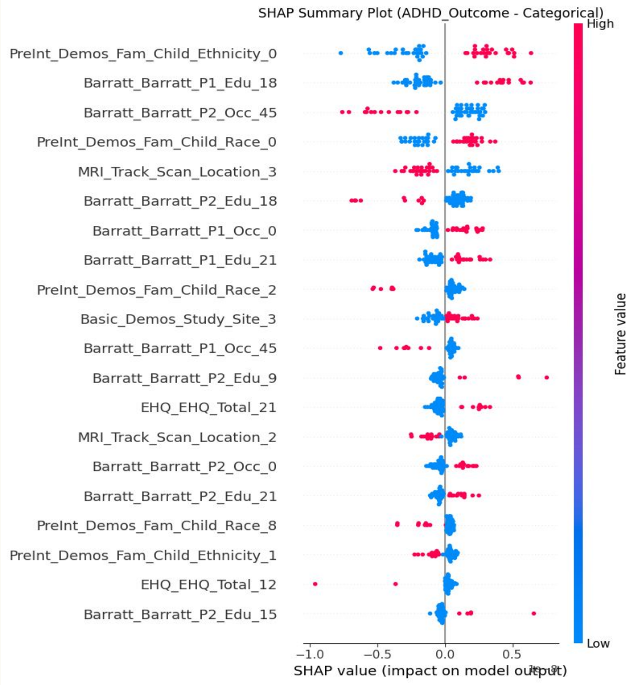


SHAP Analysis for Sex on Categorical Data

- Study site, parental education & occupation, ethnicity, and scan location impact predictions the most.

- Higher values in certain ethnicity and education features positively shift predictions, while others lower it, indicating varied demographic influence.

CNN + MLP Hybrid Architecture SHAP Analysis



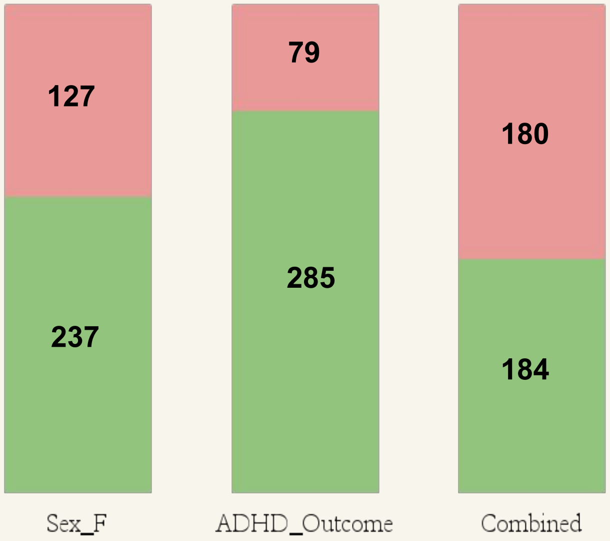
SHAP Analysis for ADHD Outcome on Categorical Data

- Top influential features include ethnicity (PreInt_Demos_Fam_Child_Ethnicity_0), parental education (Barratt_Barratt_P1_Edu_18) and occupation (Barratt_Barratt_P2_Occ_45).

- MRI scan location contributes notably, suggesting site-specific effects in prediction.

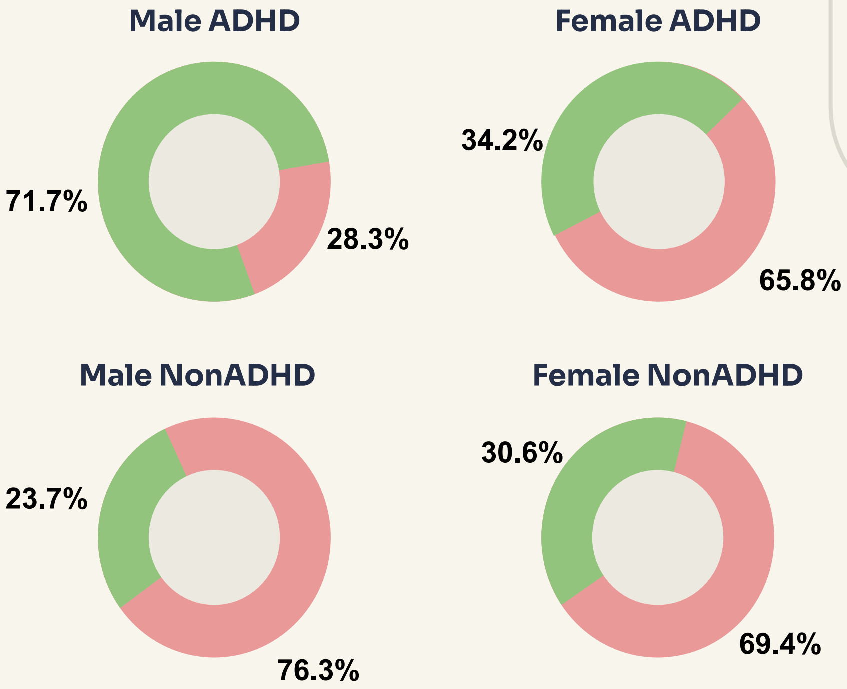
- Demographic variables show strong influence, indicating the model is sensitive to socio-demographic context.

CNN + ANN Hybrid Architecture Prediction



Correct vs Incorrect Predictions Per Label

🟩 Correct 🟥 Incorrect


Male ADHD

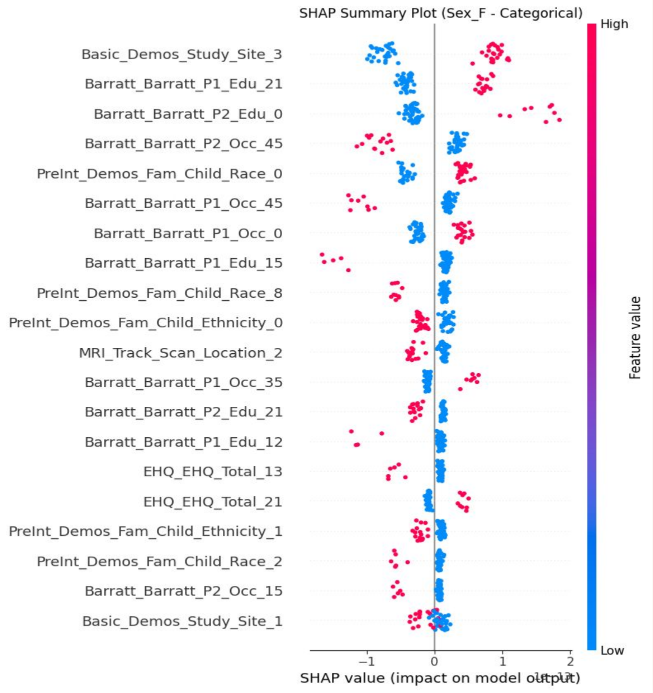
Female ADHD

Male NonADHD

Female NonADHD

Correct vs Incorrect Predictions Per Category

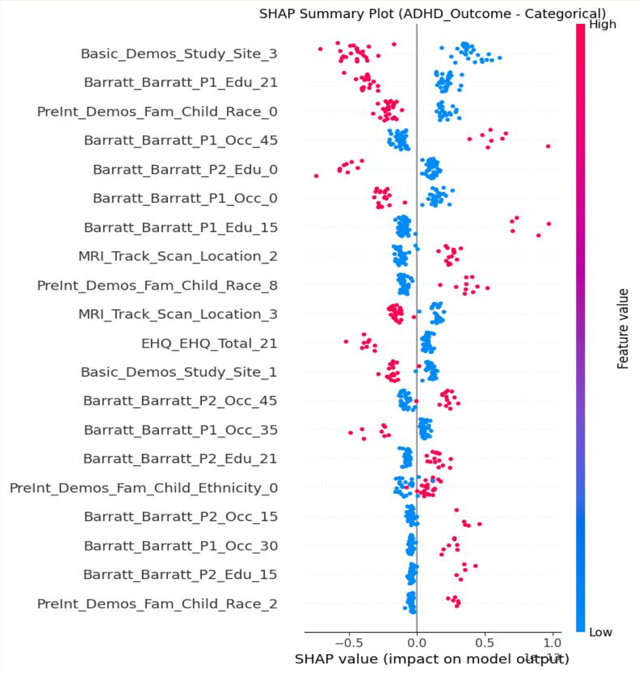
CNN + ANN Hybrid Architecture SHAP Analysis



SHAP Analysis for Sex on Categorical Data

- Study site and parental education are the most impactful categorical features for predicting female sex.
- Ethnicity and race variables also play a notable role.
- SHAP values closer to ±2 indicate stronger influence (either positive or negative) on the output.

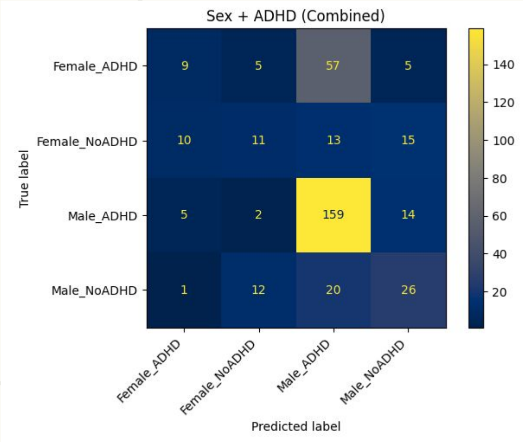
CNN + ANN Hybrid Architecture SHAP Analysis



SHAP Analysis for ADHD Outcome on Categorical Data

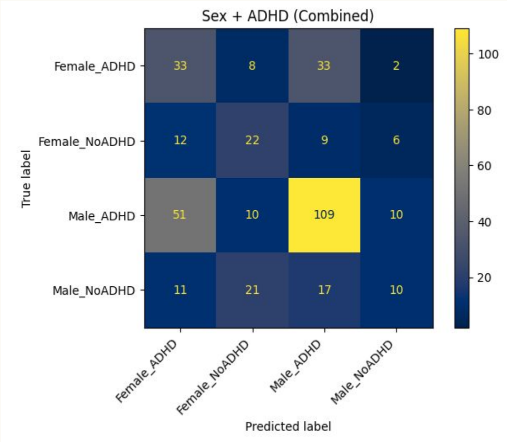
- Study Site 3 is the most influential factor in predicting ADHD.
- Parental education, especially having a graduate degree, strongly influences the model's predictions.
- The model heavily leans on socioeconomic, demographic, and site-related features to make ADHD predictions.

True vs. Predicted Labels: Model-Wise Breakdown



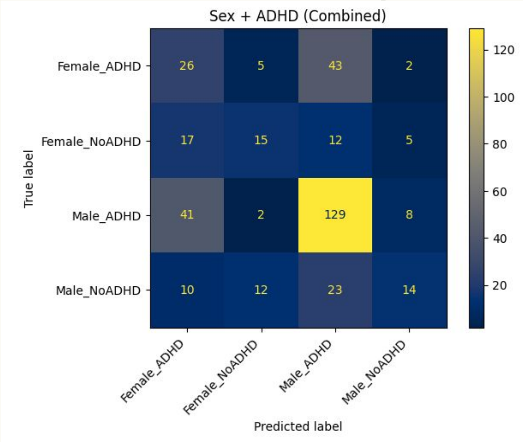
True label \ Predicted label	Female_ADHD	Female_NoADHD	Male_ADHD	Male_NoADHD
Female_ADHD	9	5	57	5
Female_NoADHD	10	11	13	15
Male_ADHD	5	2	159	14
Male_NoADHD	1	12	20	26

Confusion Matrix of Logistic Regression



True label \ Predicted label	Female_ADHD	Female_NoADHD	Male_ADHD	Male_NoADHD
Female_ADHD	33	8	33	2
Female_NoADHD	12	22	9	6
Male_ADHD	51	10	109	10
Male_NoADHD	11	21	17	10

Confusion Matrix of CNN + MLP Hybrid Architecture



True label \ Predicted label	Female_ADHD	Female_NoADHD	Male_ADHD	Male_NoADHD
Female_ADHD	26	5	43	2
Female_NoADHD	17	15	12	5
Male_ADHD	41	2	129	8
Male_NoADHD	10	12	23	14

Confusion Matrix of CNN + ANN Hybrid Architecture

Conclusions

01

Is the model suitable for the company's needs and ready for deployment?

No, the model is not ready for deployment, as it fails to perform reliably across all classes.

02

Is the model showing any signs of bias, particularly gender-related?

Yes, the model shows gender related bias, frequently misclassifying female subjects.

03

Can insights from the data support future NHS projects beyond ADHD diagnosis?

Yes, the data insights can support future NHS projects, especially those involving behavioral patterns and gender disparities in diagnosis.

Future Work

Bias Mitigation through Demographic-Specific Modeling:

- Large variation observed in parental education, occupation, and child's race.
- These factors may contribute to model bias or misclassification.
- Train separate models for specific demographic groups (e.g., based on education and occupation levels).
- Compare performance and bias metrics across groups.
- Investigate if demographic-specific training improves fairness and accuracy.

Thank You!