Brain activity pattern differences in ADHD

1 Approaches to Predict ADHD Diagnosis and Sex

1.1 Multimodal Neural Networks (Deep Learning)

Rationale: Combines categorical features (socio-demographic, emotional data) and continuous data (fMRI), integrating both pathways to make predictions.
Benefits and Drawbacks: This approach captures complex relationships between structured and high-dimensional data, improving prediction accuracy for ADHD and sex. However, it is computationally expensive, requires large labeled datasets and the interpretability of the models can be challenging.

1.2 CNN for fMRI + MLP for Categorical Data

Rationale: CNNs extract spatial features from fMRI data, while MLPs handle categorical features for socio-demographic data.
Benefits and Drawbacks: CNNs are effective at feature extraction and MLPs efficiently handle categorical data, making it suitable for combined brain connectivity and socio-demographic analysis. However, this method has a high computational cost and involves the complex integration of CNN and MLP components.

1.3 Random Forest or Gradient Boosting (Ensemble Learning)

Rationale: Robust to overfitting and handles both categorical and continuous data, making it suitable for predicting ADHD and sex.
Benefits and Drawbacks: This method handles missing data well and provides feature importance, aiding interpretability. However, it may underperform on high-dimensional fMRI data and is less effective for spatial or sequential data.

1.4 XAI-Based Neural Networks (SHAP or LIME with Deep Models)

Rationale: SHAP or LIME explain the contributions of features (e.g., brain regions, socio-demographics) to predictions, improving interpretability.
Benefits and Drawbacks: SHAP or LIME provide transparency, helping clinicians understand model predictions. However, they are computationally expensive, especially with large models and can be difficult to interpret for non-experts.

2 Evaluation Metrics

- **ROC Curve & AUC:** Measures class distinction. Higher AUC indicates better discrimination, crucial for minimizing misdiagnosis in clinical applications.
- **Cohen's Kappa:** Evaluates inter-rater agreement adjusted for chance, ensuring model predictions align with expert clinical assessments.
- **MCC:** A balanced measure for imbalanced datasets like ADHD, providing a more informative evaluation than accuracy by considering all outcomes (TP, FP, TN, FN).
- **Hamming Loss:** Measures misclassified labels in multi-label tasks, important for evaluating both ADHD and sex prediction accuracy.
- **PRF (Precision, Recall, F1):** Key classification metrics:
  - **Precision:** True positives among predicted positives. High precision minimizes false positives in ADHD diagnosis.
  - **Recall:** True positives among actual positives. High recall reduces missed ADHD cases.
  - **F1-Score:** Harmonic mean of precision and recall, balancing false positives and false negatives.