Neurodiagnostic Advancements: Deep Learning-Driven MRI Processing for ADHD detection

by

Md. Shajjat Hossain Shahat
24141244
Fardin Hassan Tamim
23341111
Al-Amin Sagor
24141225
Simran Zaman Mayabi
24141242
Smita Biswas
24241240

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science

Department of Computer Science and Engineering
Brac University
January 2026.

Declaration

It is hereby declared that

1. The thesis submitted is my/our own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Shahat

Md. Shajjat Hossain Shahat
24141244

Tamim

Fardin Hassan Tamim
23341111

Al-Amin Sagor

Al-Amin Sagor
24141225

Simran

Simran Zaman Mayabi
24141242

Smita

Smita Biswas
24241240

Approval

The thesis/project titled “Neurodiagnostic Advancements: Deep Learning-Driven MRI Processing for ADHD detection” submitted by

1. Md. Shajjat Hossain Shahat (24141244)
2. Fardin Hassan Tamim (23341111)
3. Al-Amin Sagor (24141225)
4. Simran Zaman Mayabi (24141242)
5. Smita Biswas (24241240)

of Fall, 2025 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science in January 31, 2026.

Examining Committee:

Supervisor:
(Member)

Tasnim

Tasnim Ahsan Prome
Lecturer
Department of Computer Science and Engineering
Brac University

Co-Supervisor:
(Member)

Anika

Anika Tasnim
Adjunct Lecturer
Department of Computer Science and Engineering
Brac University

Thesis Coordinator:
(Member)

Dr. Md. Golam Rabiul Alam
Professor
Department of Computer Science and Engineering
Brac University

Head of Department:
(Chair)

Dr. Sadia Hamid Kazi, PhD
Chairperson and Associate Professor
Department of Computer Science and Engineering
Brac University

Abstract

The vast advancement of neurodiagnostic technology offers an opportunity to transform mental health assessment with more precise, objective evaluations. Which previously was reliant on subjective matters like individual perspectives (Doctors in this case), interpretations, and such. With the advancement of technology in the field of medical science, there are so many resources that can be used and worked on to find efficient and better solutions for any disease or health issue. Assessing the state of one's mental health will have an impact on unleashing one's potential fully. This research solely focuses on the application of deep learning-driven MRI processing in neurodiagnostic frameworks to improve mental health assessment. In this research, our main focus was on Attention Deficit Hyperactivity Disorder (ADHD), which is a very common and concerning mental health issue, especially amongst young children. To understand its mechanism, it's crucial to investigate how functional connectivity patterns between various brain regions portray the neurobiological processes of the condition. With the help of the Graph Attention Network with Phenotype fusion (GAT-PhoNet) method, a deep learning model, we analyze the brain activity pattern through graphs, which is created from the ADHD-200 dataset resting-state fMRI data and phenotypic data that include age, gender, and IQ information. The model aimed to identify the brain activity patterns at the network level while also obtaining the crucial clinical covariates. Our GAT-PhoNet model achieved a classification accuracy of 72.09% along with an AUC-ROC score of 0.7517 and F1-scores of 0.74 for healthy controls and for ADHD subjects 0.68. This proved to be an understandable and powerful technique, which helps with ADHD detection through objective methods. The study demonstrates how multimodal graph-based deep learning systems can enhance neurodiagnostic procedures by providing more accurate quantitative analysis results.

Keywords: mental health, deep learning, ADHD, Graph Neural Network, MRI

Acknowledgement

The thesis represents our own original work while pursuing degree at the Computer Science and Engineering Department, BRAC University. We extend our deepest gratitude to the Department of Computer Science and Engineering at BRAC University for providing us with the opportunity and resources to pursue this research. We are deeply grateful to our supervisor, Ms. Tasnim Ahsan Prome and co-supervisor, Ms. Anika Tasnim for their invaluable guidance, continuous support and insightful feedback throughout the process. All sources of information are referenced appropriately and the views expressed in the paper are solely of the authors. As far as our knowledge, this work does not contain any material of any previously published or submitted works of elsewhere except where explicitly cited with proper references.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

Attention Deficit Hyperactivity Disorder (ADHD) represents a common neurodevelopmental disorder, which is characterized by persistent symptoms of inattention and hyperactivity along with impulsive behavior that disrupts an individual's ability to learn, think, and interact socially. As well as that, ADHD affects approximately 5–10% of children and 2.5% of adults worldwide because it creates long-lasting effects that shrink their educational success, job performance, and most importantly, mental health[15]. The symptoms of ADHD present in various ways, and because of this, researchers face difficulties in diagnosing ADHD. On top of that, the symptoms closely resemble other neuropsychiatric conditions. In recent times, Doctors diagnose patients through three main methods, including clinical interviews, behavioral assessments, and rating scales completed by patients or their caregivers. Diagnosing an accurate assessment is difficult with the technology used nowadays for testing because clinicians face two major problems, which are both rater bias and developmental changes. Also, cultural differences create assessment challenges as well, which result in diagnosis delays and misdiagnosis rates ranging up to 30% in certain clinical settings [18]. Due to these major diagnostic flaws, it is necessary to develop new diagnostic technologies that utilize neurobiological evidence to augment long-established evaluations.

Advances in neuroimaging, particularly resting-state fMRI, have developed into an effective technique for studying the brain structures that cause ADHD. Previous research has found that brain regions supporting executive function, attention control, and impulse control exhibit consistent but subtle neurofunctional alterations, affecting the prefrontal cortex, basal ganglia, cerebellum, and default mode network. The anatomical differences between the two groups reveal minor variations that affect multiple brain areas, rather than concentrating on a single specific area of the brain. The complex High-Dimensional resting-state fMRI data patterns remain hidden from common statistical analysis methods and area-based analysis methods because these methods fail to handle the data's intricate structure. The development of effective computational models faces challenges because researchers have access to only small sets of neuroimaging data that are properly annotated for ADHD research.

Deep learning has shown its ability to analyze medical images by developing methods that extract features from unprocessed imaging data since 2015 [10]. The use of Convolutional Neural Networks (CNNs) has become common in fMRI classification work because these networks can learn to identify different brain region structures through the hierarchical spatial mapping capacity. The CNN-based methods concentrate their efforts on analyzing features at the local voxel level and region levels, which hinders their capacity to understand how different brain regions that share anatomical or functional ties work together [13]. The understanding of ADHD now reveals that the disorder operates as a network-level condition, which causes problems in distributed neural systems that work together instead of showing only structural defects. The use of CNN architectures may create incomplete models that fail to capture all the biological processes that cause ADHD because they depend on these particular architectural designs [16]. Consequently, the necessity for an architectural system that can master distinct features from marginal data while simultaneously revealing brain network patterns and contextual connections, considering the existing challenges.

In this research paper, a graph-based deep learning framework is used to identify ADHD, which uses Functional Magnetic Resonance Imaging (fMRI). The proposed architecture designs brain regions as graph edges, with nodes replicating functional connectivity strength computed from resting state fMRI time series correlations using the AAL atlas with 90 cerebral regions [4]. The GAT (Graph Attention Network) model determines which brain regions are most crucial because it weighs their importance based on their connection strength to nearby areas that have an impact on ADHD classification. Additionally, the MLP (Multi Layer Perceptron) examines subject-level phenotypic features, which include clinical and demographic attributes that can affect ADHD classification as well. As a result, the combination of GAT and MLP generates the ultimate classification result.

Therefore, this study aims to determine an automated ADHD classification method that uses both neuroanatomy and phenotypic information as its foundation. Detecting ADHD symptoms using a moderate sample size of data with the proposed GAT-Pheno framework addresses attention-based graph learning along with phenotype-aware modeling to achieve better discrimination results. Accordingly, this research employs regularization methods, including dropout and weight decays, which prevent overfitting while maintaining consistent learning outcomes. In brief, this thesis proposes a prototype system for ADHD neurodiaonistics, which demonstrates how graph attention techniques combined with additional phenotype information improve deep learning analysis of resting fMRI data.

1.1 Problem Statement

The study of Attention-Deficit/Hyperactivity Disorder (ADHD) has benefited from present-day neuroimaging techniques, but researchers still encounter difficulties when establishing dependable automated diagnostic systems. The standard technique of machine learning develops brain region analysis as separate units, while convolutional neural networks rely on Euclidean space distribution, which creates difficulties for studying brain networks that have essential graph-structured characteristics. On the other hand, Graph Neural Networks (GNNs) use Graph Attention Networks (GATs) as a sturdy method to create brain region functional connectivity models. The existing GNN-based ADHD classification studies fail to use clinical phenotypic information, while they face difficulties with class imbalance problems that exist in clinical datasets. Therefore, this research fills existing gaps through the creation of a multimodal GNN framework that combines functional connectivity data with phenotypic information to enhance ADHD classification performance.

1.2 Research Objective

Primary Objective

The thesis study aims to create and assess an ADHD classification system that applies a Graph Attention Network (GAT) model, which combines functional connectivity data from resting-state fMRI with clinical phenotypic data to resolve core challenges present in current neuroimaging-based diagnostic systems.

Specific Objectives

- **The thesis aims to create a GAT-based classification system** that uses brain connectivity graphs derived from resting-state fMRI data to identify vital connections between brain regions through its attention mechanisms.
- **The study aims to implement a multimodal fusion method** that combines multimodal phenotypic information with brain connectivity data through feature concatenation.
- **The research aims to build functional connectivity graphs through Pearson correlation matrices**, which establish a standardized threshold ($|r| > 0.6$) that defines edge creation limits for creating consistent brain network models.
- **The research aims to improve model performance on clinical datasets** with unequal distribution of classes through class-weight adjustment, which they will apply to the binary cross-entropy loss function.
- **The study will assess model performance through standard train/ /validation/test splits** that will evaluate using complete metrics that include accuracy, AUC-ROC, F1-score, precision, and recall.
- **The GAT layer attention mechanisms provide the model with basic interpretability** that will enable this research to trace essential brain region connections used for ADHD classification.

1.3 Methodology in Brief

The GAT-Pheno model functions through its predefined set of processes. In this research, Pearson correlation matrices are used to establish functional connectivity graphs from the resting-state fMRI time-series data which is converted using a fixed correlation threshold that requires $|r|$ to exceed 0.6. The graph nodes represent brain regions from the AAL atlas that contain 90 nodes, while the node features display the correlation profile of each region.

Additionally, the model uses a Graph Attention Network, which contains two layers to examine brain connectivity graphs because it adapts how brain regions connect through its attention mechanisms. Besides, a parallel multilayer perceptron processes phenotypic attributes such as age, gender, and IQ to create complementary feature embeddings. The technique combines graph-based brain representations along with phenotype-based clinical representations through simple concatenation followed by a classification layer. Lastly, the training process utilizes class weighting in the loss function to solve the problem of class imbalance.

1.4 Scopes and Challenges

This research presents a system that evaluates adolescent ADHD through fMRI brain network analysis and phenotypic data assessment. The developed framework achieves its first operational milestone through multimodal ADHD assessment, yet restricts its capabilities to binary results without supporting ADHD subtype distinction or progress tracking.

Challenges

- Neuroimaging datasets face two major obstacles, which are addressed through a class weighting method to handle data scarcity and class imbalance issues.
- The brain networks require non-Euclidean data representation, which is solved through the graph-based modeling method.
- Integrated two types of data, which are neuroimaging results and clinical data.

Chapter 2

Literature Review

2.1 Preliminaries

Over the past few years the interest in applying machine learning and deep learning methods in neuroimaging data has been growing at a rapid speed for the objective diagnosis of neurological and neuro development disorders in patients. Traditional methods for diagnosing such conditions like attention deficit hyperactivity disorder (ADHD), cognitive impairment (MCI), dyslexia, depression and other mental health disorders mostly depend on behavioural evaluation and clinical assessments that can often be subjective and drawn to variability. To overcome these challenges scientists have been exploring brain imaging modalities with structural MRI, functional MRI (fMRI), resting state MRI (rs fMRI), EEG and multimodal datasets along with advanced computational frameworks. Researches that were conducted between the years from 2016 to 2025 have illustrated that deep learning frameworks such as convolutional neural networks (CNN), generative adversarial networks (GAN), recurrent models and graph based approaches are able to extract complex and high dimensional patterns from neuroimaging datasets automatically. They often outperform traditional machine learning techniques that rely on handcrafted features. Many other reviews and survey papers that were published during this particular period emphasises the rapid evolution of this field. Additionally it highlights both the promising diagnostic performance of the mentioned models and the persistent obstacles that are related to data scarcity, interpretability and clinical adaptability. This literature review that includes findings from previous researches, encourages research efforts before proceeding to work, mentions common methodological trends and limitations faced by other scientists and creates motivation for development of more robust, explainable and generalizable neuroimaging based diagnostic architecture.

2.2 Review of Existing Research

The study titled "Brain MR Image Classification for ADHD Diagnosis Using Deep Neural Networks" [5] is a research conducted on usage of deep learning methods to categorize Brain MRI scans for the accurate detection of Attention Deficit Hyperactivity Disorder (ADHD) which is a neurodevelopmental condition and the chances of prevalence in ADHD cases are increasing with advanced technologies and time. According to the researchers, the key obstacles in computerized diagnosis of ADHD, which are the complex and high dimension features of MRI datasets. They are solving this issue with a balance of deep neural networks or DNN based framework by extracting and interpreting the minor structural details of brain scans and it differentiates ADHD from regular development. This method achieved an accuracy rate of 90% which outperformed traditional ML approaches. The authors monitored and designed a categorized framework which operates MRI data of the brain through layered neural architectures that are capable enough to learn hierarchical representations from the input raw images. This process is different from typical machine learning methods because it doesn't require manual engineered features as deep learning networks can learn potential features relevant for ADHD detection by itself and simultaneously enhancing objectivity and efficiency of diagnosis. This follows the neuroimaging research trend that captures complex spatial patterns present in brain MRI for neurological classifications by the convolutional and deep neural models. This paper also includes similar techniques such as the usage of 3D convoluted neural networks for detecting ADHD classifications that highlights both possibilities and obstacles which are small samples of the database, complexity of neural models and finally the importance of robust evaluation on greater more detailed datasets.

In this study, author Cao et al. (2023) [12] focused on brain MRI scans and deep learning techniques to make diagnosis easy for ADHD patients. The identification of this brain disorder normally depends on behavioral tests and clinical judgements but these methods have a chance of being subjective. For this risk, authors have been exploring several automatic and objective methods using medical imaging. The researchers stated that previously used machine learning methods such as SVM, traditional neural networks for classifying ADHD had shown some success, yet they mostly depend on hand crafted features and their accuracy rates are also limited while they deal with complex datasets of complex brain MRIs. To tackle these challenges, researchers recently have shifted their focus on deep learning models, more specifically convolutional neural networks which is able to learn useful features from MRI scans on its own. This research work used a deep learning based process to examine MRI images and label them as ADHD or non-ADHD. The outcome shows that deep neural networks is capable of capturing significant brain patterns which traditional methods are unable to detect. The proposed model here gives more accurate results with reliability. Besides, this paper sheds light on the challenges faced in the study such as small datasets, overfitting, and lack of model interpretability. However, these drawbacks are overshadowed by increasing demands of deep learning techniques in brain image analysis.

This research paper by Lohani et al. (2025) [20] was written after reviewing a total of 87 published papers (from 2016 to 2022) that used MRI scans and electroencephalography data with machine learning/deep learning models for ADHD diagnosis. Here, they observed a scientific research method which is called PRISMA guidelines to make sure the research and diagnosis were unbiased and well organised. The paper shows that scientists used various datasets, ML models with feature extraction techniques and evaluation methods. The dataset that was most popular at that time, the ADHD-200 dataset. Scientists also observed huge differences between the studies regarding how the data were fetched and how the ML models were trained and tested for diagnosis of ADHD. This made it hard to compare the outcome directly. This research also identifies regular challenges faced in this field such as dataset size, irregularity in datasets and hard to understand for common people. The writers focused more on the necessity of these work in the upcoming years and for that the models need to be more general, easily interpretable and reliable for all so everyone can use this in real clinics.

This research conducted by Taspinar and Ozkurt (2024) [19] observed the machine learning techniques that are used with the resting-state functional MRI (rsfMRI) data to identify ADHD in patients. Traditional ADHD detection methods mostly rely on behavioral observations and that can be subjective and sometimes inaccurate. That is the reason why we need more objective and reliable tools. The researchers described that rsfMRI is a non-invasive imaging process which helps in studying brain activity while a person is at rest. Many scientists have used rsfMRI data in inclusion to ML models to detect brain patterns similar to ADHD. This research paper is relevant because this mainly focuses on ADHD diagnosis using rsfMRI and machine learning. Additionally, this study provides detailed information on commonly used fMRI databases that are widely used by researchers all over the world. It also studied all major stages of ADHD diagnosis along with data preprocessing and brain networks with atlas selection, feature extraction, selection and classification. It also compared separate ML models such as SVM, random forest, graph based models, neural networks and deep learning models which in accuracy rates in the range between 65% to over 90% that also depends on various preprocessing choices. This paper highlights the fact that the choices made before classification, like network and feature selection, greatly affected the performance of the models. The authors also mentioned the challenges and strengths of several researches that help with the future research directions.

This paper by Samyuktha S. and Anandaraju M. B. [2024] [17] which is titled "Classification of ADHD and its Sub-Types using Machine Learning Models" is a significant work in the field of ADHD identification with the use of neuroimaging and deep learning. Highlights of this study are how usage of Convolutional Neural Networks (CNN) is able to effectively identify ADHD and its other different types from functional MRI datasets. ADHD is such a neurodevelopmental condition that can be often misdiagnosed due to overlapping symptoms with other mental health conditions. Previous studies on this topic mainly concentrated on hand crafted features and classical classifiers like SVM, texture based techniques or EEG derived features. Here an author, Gülay Çiçek et al. applied texture analysis on MRI scan images and another author named Ahmed Salman et al. has explored functional connectivity measures from fMRI datasets. These approaches reached reasonable classification accuracy but their performance was narrowed down by manual feature selection and poor generalization throughout the datasets. This particular research takes a deep learning approach using CNNs and it automatically learns hierarchical features from high dimensional neuroimaging datasets. The presented model was estimated using 5-fold cross validation and it earned an average classification accuracy rate of 91.68% which outperformed the VGG-16 baseline model that had 89.33% accuracy. Subtype wise accuracies were also high: for combined ADHD it was 98.9%, 97.7% for hyperactive ADHD and lastly 95.8% accuracy for inattentive type ADHD. The normal subjects were classified with 99.8% accuracy rate. From the results, it can be demonstrated that CNN based framework remarkably enhanced ADHD subtype classification. As usual, this research also comes with few challenges like small dataset, less interpretability and generalization issues. Future researchers may consider the issues and conduct a more explainable deep learning model that can support clinical adoption.

In this study titled "Fusion of deep learning models of MRI scans, Mini–Mental State Examination, and logical memory test enhances diagnosis of mild cognitive impairment", the author Qiu et al. (2018) [3] investigated the potentiality of integrating multimodal data which are mainly MRI scans, mini mental state examination (MMSE) and lastly Logical memory (LM) test scores with the help of deep learning models targeting the improvement in the diagnosis of mild cognitive impairment (MCI). They combined both image data (MRI scans) and non image data (memory test) in order to address the urgency needed in the clinical sector for the early and specific detection of MCI which is also an early indicator of Alzheimer's disease. Although brain images are able to capture physical changes in the brain that are correlated to MCI, this also seems to have limits if we fully depend on it. The researchers took initiative to improve the results with deep learning models and for that purpose they collected data from 386 people. They used the method named "majority voting" to combine deep learning models for which they had to build three models which are based on MRI images, MMSE scores, and LM test results. Here the ultimate prognosis was based on what was agreed on by most models. In the test the MRI model precisely identified 83.1% cases, the MMSE model guessed 84.3% cases accurately and finally the LM mode's accuracy was 89.1%. However, by combining the three outcomes, the accuracy rate jumped up to 90.9% which indicates that by using multiple types of data we can scale up the results in a better way rather than just using one particular test. The MRI and MMSE models did not miss any MCI cases which showed higher recall where the LM model had the highest precision, meaning it correctly identified patients with MCI. Combining these results, the researchers balanced the strengths of each model which resulted in a more reliable diagnosis. Another technique applied in the research was transfer learning. At first it allows the DL models to learn patterns from other large datasets which helps the MRI models with more precise prediction even with smaller number of brain scans. The research tools of this article can help doctors to identify early stages of MCI which will greatly impact the treatments of affected patients.

Among many neurological disorders, dyslexia is the condition that affects one's reading and language skills. Researchers all around the world have explored different ways of diagnosing dyslexia as traditional methods like observing individual's reading behaviors or using reading tests can be subjective and can miss out on narrow signs. For the advancement of the process, researchers decided to depend on brain imaging tools such as MRI scans that track blood flow in the brain and EEG that records electrical activity of the brain to get a better understanding of the functionality of dyslexic brains (Alkhurayyif & Sait, 2024) [14]. There were researches conducted that used DL models in detection of dyslexia with brain imaging data. Such as, some models that used children's handwriting and eye movements to detect dyslexia. On the other hand there were models that are based on convolutional neural networks (CNNs) which processed fMRI images or used LSTM models to examine the EEG signals progressively. According to the research some scientists have tried combining multiple sources of multi-modal data in order to improve the rate of accurate detection of the condition. Techniques like transfer learning, attention mechanisms and generative adversarial networks are applied to these models to boost the results and work with limited resources. For example, models that rely on attention can target main patterns connected to dyslexia and at the same time GANs are able to produce extra information that can enhance model training. Although the current models struggle with few limitations despite having all the advanced techniques. High computational needs, challenges in interpreting results, and issues with generalizing to a new dataset are few of the concerns in this case. Due to the patient's individual traits, analyzing the results of handwriting and eye movement data can not be relied upon fully. Besides, many models require huge datasets and impactful hardware that makes the process harder for everyday uses in schools or small health clinics. All these obstacles inspired the researchers to come up with a more accurate model of dyslexia detection with the use of deep learning models and multi-modal data. The purpose was to invent an easy to understand and precise system that processes brain images and EEG data in dyslexia detection that is reliable and requires minimal amount of resources.

Researchers presented a new deep learning approach in the paper titled "Functional network connectivity (FNC)-based generative adversarial network (GAN) and its applications in classification of mental disorders" which classifies different mental health conditions using brain imaging datasets (Zhao et al., 2020) [7]. Scientists took help from resting-state fMRI data that shows the brain activity of a person when he is not doing any specific task. By processing the data they created Functional Network Connectivity (FNC) maps like essential charts which show the different interactions of different parts of the brain. Later these maps were used in a particular deep learning model as inputs and the model is known as Generative Adversarial Network (GAN). For this study, researchers used the GAN in two parts: a generator that creates fake FNCs and the other part is a discriminator that tries to differentiate real FNCs from the fake ones. For effective learning the model trained with both parts together. A different approach called feature matching was also used to enhance the quality of the generated datasets that helped the model with better learning distinctions between the affected and non affected persons. The model was tested on two separate mental disorders by the scientists. The first one is major depressive disorder (MDD) and the second is schizophrenia. The accuracy rate of the model for MDD was 70.1% on average and 82.1% for schizophrenia after using 10 fold cross validation. This performed better than six other classification methods. The models were tested across various hospitals and the results were good as expected which showed it is generalized for new datasets too. The research identified the key connections of the brain which were the most crucial parts for setting off between the patients if they are healthy or suffering from any mental disorder. All these sums up the high accuracy and interpretability of the model which is generally not found in deep learning.

In a study, authors (Zhang and Zhang, 2025) [21] talked about depressive symptoms as depression has become one of the most common mental health problems these days and it is quite frequent among the young adults which can impact them in the long run. Identifying depression in early stages is crucial for prevention of the worsening condition and it helps with the reduction of personal and social burdens.

To detect depression-like mental health issues, researchers used many traditional methods previously such as self-reported queries that were helpful but they had some limitations which could not fully uncover the underlying brain issues. To resolve these matters, scientists turned to biological indicators such as brain imaging (MRI scans) with a goal to identify early signs of depression. Researchers showed most of the past studies that were done using MRI scans, gave confusing results and were mainly focused on patients who were already suffering from depression. The goal of this study titled "Predicting depression in healthy young adults: A machine learning approach using longitudinal neuroimaging data" is to recognize symptoms of depression in young adults with the help of brain imaging and deep learning models. Structural MRI and resting-state functional MRI scans were collected from a large number of students by the scientists. Then they applied feature selection methods like LASSO, Boruta, VSURF and constructed DL models using Support Vector Machine (SVM) and Random Forest (RF) to detect depression. Researchers divided the human brain into eight important regions which includes the orbital, superior frontal and parahippocampal gyrus, the parts that helped to predict future depression traits. According to the research, the best model gave 85% accurate test results and 0.80 AUC that showed exceptional prognostic outcome. Regarding early detection of the brain condition, functional MRI data performed better than structural MRI. The outcome of this research revealed that combining brain imaging with machine learning can be helpful for recognizing early symptoms of depression in young adults, in some cases even before clinical diagnosis. The research assists the future development process of biological tools regarding mental health disorders and its prevention.

In this paper titled "A Survey on Deep Learning for Neuroimaging-Based Brain Disorder Analysis", researchers provided a glimpse of how deep learning techniques are used to analyze brain disorders with the help of neuroimaging (Zhang et al., 2020) [6]. Deep learning models examine brain images like MRI, fMRI, and PET scans which helps to identify mental disorders such as Alzheimer's, Parkinson's, Autism Spectrum Disorder (ASD), and Schizophrenia. Authors of this paper introduced various DL models which include CNN, RNN, GAN and even autoencoders. Instead of manual help from experts, these methods act helpful while extracting crucial patterns generated from brain scans which makes the whole process more swift and precise. Researchers analyzed how the models were applied to find certain brain disorders. Results showed that CNNs and autoencoders gave impressive signs in identifying Alzheimer's and Parkinson's. On the other hand GCN, RNN models performed better in diagnosis of Alzheimer's and Parkinson's as they analyze critical brain connectivities thoroughly. Authors also discussed the limitations of the recent deep learning models that are being used frequently. E.g lack of interpretability, small datasets etc. The upcoming researchers should aim on merging different types of datasets and enhancing transparency of the DL models.

2.3 Summary of Key Findings

From the research timeline of 2016 to 2025 all day deep learning based techniques have shown strong potential for identifying neuro developmental and mental health disorders using brain imaging data. Previously the researchers relied mostly on traditional machine learning classifiers like support vector machines (SVM) and random forest merged with handcrafted features taken from MRI, fMRI and EEG data. Although these techniques have achieved moderate success, their precision rates were limited by manual feature engineering and failure to fully capture the complex functional patterns that exist in high dimensional neuroimaging data.

Several studies of recent times showed a clear improvement toward deep neural networks, particularly the convolutional neural networks that are able to learn hierarchical features automatically straight from the raw brain image inputs. CNN based deep learning models that are applied to train structural MRI and functional MRI data are outperforming regular ML approaches successfully, with classification precision rates frequently crossing 85 to 90% for brain disorders such as ADHD and MCI detection. These models are mainly effective in catching minor brain structure flaws, variations and functional connectivity patterns that are hard to examine manually.

Another remarkable thing that we noticed across the literature reviews is the efficiency of multimodal learning. Researches integrating MRI data with cognitive analysis or EEG signals gained higher diagnostic position compared to the single modality techniques. Multimodal fusion techniques come in to help cover up for the challenges faced in individual data resources that result in more robust and clinically appropriate predictions. Also the techniques such as transfer learning and data augmentation mainly use GANs and are hopeful in reducing small data set issues which has been a continuous obstacle in neuroimaging research.

Even after all these advances the literature keeps highlighting the key limitations. Several models are trained on small and highly controlled data sets that reduce their ability to generalise to real world medical settings. Besides the insufficiency of interpretability in deep learning architectures is still considered a major obstacle to clinical adaptation as a result all the recent surveys and reviews take into consideration the need for an explainable standardised and generalisable model which is reliable and can be applied across numerous data sets and brain imaging resources.

Chapter 3

Requirements, Impacts and Constraints

3.1 Final Specifications and Requirements

The primary objective of our proposed system is to develop a deep learning based architecture that is capable of supporting the diagnosis of neurological or neuro development disorders with the proper use of neuroimaging data such as MRI scans. The system needs to process pre-processed brain image scans such as structural MRI scans, functional fMRI scans accurately and generate specific classification outputs that can distinguish between affected and non-affected patients. High classification accuracy rate robustness and proper consistency across various validation falls are crucial functional requirements. The system has to support methodological feature learning from high dimensional brain imaging data using convolutional neural networks or deep learning architectures similar to that. The model has to be trained and estimated using standardised performance matrices such as precision accuracy, recall, F1 score and area under the ROC curve. It must follow some non-functional requirements that includes computational efficiency of the model, reproducibility of the test result, ability to handle increased data volume without degradation of performance and compatibility with recurrently used deep learning frameworks (Prisma-based review, 2016–2022) [20]. Obstructions related to limited size of data set availability of hardware and training times are recognised and mentioned through methods such as data augmentation and transfer learning.

3.2 Societal Impact

This presented system has the potential of creating a positive societal impact through contribution to earlier and more objective diagnostic processes of neurological and mental health disorders. Many similar conditions including ADHD and depression are often traditionally diagnosed via behavioural evaluation which might be subjective and consistent. By showing data driven decision support the system will be able to assist medical professionals in providing more and far diagnostic outcomes that can potentially reduce mass diagnosis and delayed treatment. Moreover, early and precise identification of brain disorders helps and improvement of patient outcomes reduces long-term healthcare costs and improves living quality. This type of automated diagnostic tools can help bridging the gaps in healthcare availability in places where access to specialized neurological experts are limited. Although the motivation behind the invention of this system is to support the medical decision making process and not to replace medical professionals.

3.3 Environmental Impact

This project has a relatively limited environmental impact as it is mainly a software oriented project and it doesn't include the manufacturing of any physical structures. Still training deep learning models specifically on a large neuroimaging data set can be computationally challenging and energy demanding. Recent researches have found that deep learning in neuroimaging contributes to the increasing energy consumption because of its extensive model training and hyper parameter optimization (Frontiers in Neuroscience, 2020) [6]. To lighten these effects, our proposed idea adapts efficient training techniques with optimised model architectures, usage of pre-trained networks through transfer learning and conduct experiments keeping in control instead of exhaustive searches. As the system is outlined to work on existing computing infrastructure that is why no other secondary and viral mental Lord related to hardware procedures is presented.

3.4 Ethical Issues

Consideration of ethical issues play a significant role in neuroimaging based deep learning research. The use of brain imaging data causes concerns about the privacy of the patient's informed consent and security of the personal data. Along with other established researches, the proposed idea makes proper use of masked datasets and avoids processing personally recognisable information hence sticking to ethical standards for uses of medical data. One more major ethical distress detected in existing research is lack of transparency and algorithmic bias. Though deep learning models exhibit biased performance when they are trained on non-representative datasets gradually leading to an unfair or inaccurate diagnostic outcome over different populations (Frontiers in Neuroscience, 2020) [6]. To tackle this issue our proposed system makes sure to emphasise diligent evaluation and clearly specifies its role as a clinically used decision support aid.

3.5 Standards

There is no single international standard that inspects deep learning based academic diagnostic research but our proposed system adjusts with best practices that are widely accepted in medical neuroimaging and deep learning research. According to systematic reviews and service studies, things that need to be made sure of are transparent reporting of all the experimental strategies, usage of standard evaluation matrix and the reproducibility of the test results (Prisma-based ADHD review, 2016–2022) [20]. Additionally, principles of uRTH data handling consistent with clinical research rules such as keeping patient's personal information protected and restricted data access are followed. Software development sticks to general coding standards to confirm clarity and sustainability of the proposed deep learning model.

3.6 Project Management Plan

This research follows an organised development plan then begins with an extensive literature review to find out all the gaps and form a methodological foundation. This step follows a few necessary steps which are data at selection pre-processing the data set and analytical analysis. All of these are critical steps that highlight prior neuroimaging research due to the strong influence they have on model performance (Taşpınar & Özkurt, 2023)[19]. All of these are critical steps that highlight prior neuroimaging research due to the strong influence they have on model performance.

3.7 Risk Management

There exists many risks that are associated with the learning based neuro imaging techniques that have been found in the literature. One consistent risk that has been found in most of the researches is overfitting caused by the limited size of the data search that significantly reduces the generalisation capability (Abdolmaleki-Abadeh et al., 2020) [5]. This risk can be addressed through a technique called cross validation regularisation and data augmentation. Besides overfitting there are a few additional risks such as computational constraints, data imbalance and lack of variability across imaging sites. Also ethical risks like misinterpretation of automated predictions are managed by maintaining human supervision.

3.8 Economic Analysis

Considering the economic aspect, the proposed research is a cost efficient research because of its dependency on the openly available learning architectures and publicly available brain imaging data sets. Development costs are mainly linked with computational resources instead of physical infrastructure. After properly being trained, the system can be Utilised with minimal cost, which makes it scalable for research and medical evaluation purposes. From the existing studies we get the suggestion that early and precise diagnosis conducted by machine learning techniques can decrease long-term health care losses by minimising delay treatment and other unnecessary procedures. Even though this research is conducted as an academic research, its low operational cost and dependency on existing model architectures signals towards a strong prospective for future extensive research into medical research setting.

Chapter 4

Proposed Methodology

4.1 Design Process or Methodology Overview

This paper suggests a deep learning-based architecture of ADHD detection based on rs-fMRI signals. The pipeline in general combines neuroimaging preprocessing, functional connectivity modeling, graph-based representation learning, and multi-modal deep learning to jointly use the information of brain networks and phenotypic variables.

The dataset followed Brain Imaging Data Structure (BIDS) format. Subject-specific imaging data were automatically identified via recursive directory scan and site-specific TSV files were used to extract site-specific phenotypic data, such as age, gender, iq, and diagnostic labels. Data that were not available or were inadequate were not subjected to further analysis.

The Automated Anatomical Labeling (AAL) atlas was used in the preprocessing of neuroimaging to design brain regions of interest (ROI). The loading and application of the atlas locally as well as the extraction of regional time series through preprocessed fMRI volumes were done using a NiftiLabelsMasker. The standard pre-processing steps were used, such as spatial smoothing (Gaussian kernel with full-width at half-maximum of 6 mm), temporal band-pass filtering (0.009–0.08 Hz), linear detrending and z-score normalization. All time series were cut to a set length of 172 time points and only the first 90 AAL regions were used to facilitate consistency in model input as previously described in the literature.

Functional brain networks were built with Pearson correlation coefficients between regional time series, which produces subject-specific functional connectivity matrices of dimension 90×90. Graph representations were defined using these matrices with the nodes being brain regions and the edges being functional connections that surpass an absolute correlation threshold. The feature vectors were given to each node (representing the correlation profile of that node, both local and global connectivity patterns of the brain network), thereby representing the local and global connectivity patterns of that node.

Resultant graphs were represented by the Graph Attention Network (GAT) architecture that allows the adaptive weighting of the inter-regional links according to the learned significance. Simultaneously, the phenotypic variables (age, gender, and IQ) were also normalized and run through a multilayer perceptron (MLP) branch. The GAT-learned graph-level embedding and MLP-learned phenotype embedding were then combined and fed to a final classifier to generate binary prediction of ADHD. A weighted binary cross-entropy loss was used to solve the issue of class imbalance, and dropout regularization, early stopping and learning rate scheduling were used to stabilize training.

In order to compare the effectiveness of various deep learning paradigms in a detailed way, this GAT-based multimodal framework was compared to other architectures, such as BrainNetCNN to perform convolutional learning on connectivity matrices, a Connectome Transformer model to identify long-range dependency patterns within functional networks, and TabNet to perform phenotype-driven classification. A uniform split of the data was used to train and test all models using stratified train, validation, and test sets. Accuracy, ROC-AUC, F1-score, average precision and the use of a confusion matrix were used to determine performance.

In general, the presented methodology offers an integrated and scalable platform of ADHD detection that brings together neurobiological interpretability and the latest deep learning models to offer solid analysis of brain connectivity patterns and their interplay with subject-specific phenotypic attributes.

4.2 Dataset Preprocessing

Data in the form of functional MRI (fMMRI) adopted in the current research were in the ADHD-200 Consortium repository[1] , which is a collection of multi-site resting-state fMRI scans, as well as phenotypic data on Attention-Deficit/Hyperactivity Disorder and typical developing controls. Multi-site scans were provided (e.g. NYU scans, NeuroIMAGE scans, Peking scans, Pittsburgh scans) and 637 subjects were left after quality control, comprising 362 healthy controls and 275 patients with ADHD. We retrieved age, gender, and IQ of each of the subjects in the site-specific phenotypic files to construct the phenotype branch of the model. Data were splitted into training, validation, and test sets, maintaining the content of ADHD/control ratios in each subdivision.

4.2.1 Data Cleaning

The presence of both phenotypic and imaging records were scanned in all the sites, and the subjects lacking both fMRI volume and missing or invalid diagnosis labels were eliminated. The diagnosis text fields of various sites were standardized through matching typical-control descriptors (ex: typically developing, control) to classify 0 and any category related to ADHD as 1. Questionable categories were disqualified. The gender variables were both mapped to a binary variable (male = 0, female =1), whereas age and IQ were updated to numerical values with invalid values being set as default placeholders and then treated with the help of normalization.

[1]ADHD200. (n.d.). https://fcon_1000.projects.nitrc.org/indi/adhd200/

4.2.2 Data Transformation

Each subject had their 4D rs fMRI extracted atlas based time series using the AAL atlas, downloaded the ROIs (116 regions) and processed with a NiftiLabelsMasker with z score standardization, spatial smoothing (6 mm) and spectral band-pass filtering (0.009–0.08 Hz) and detrending. In order to have a fixed temporal dimension among subjects, every time series was cut off at 172 time points; any subjects that had fewer time points were discarded. Owing to the fact that the proposed model is interested in 90 cortical and subcortical regions, the atlas-based time series were cut to keep only the first 90 regions, which imposes a final tensor size of 172×90 per subject.

4.2.3 Data Integration

On each subject, we pooled the imaging and phenotypic data into one structured sample that included:

1. a 172×90 regional time-series matrix,
2. a three dimensional phenotype vector (age, gender, IQ),
3. a binary label and
4. a site identifier.

Phenotypic variables were piled onto all the subjects and standardized using a global Standard Scaler, and the normalized features of each subject were added as an additional attribute into the graph data structure. Valid samples were all put into one NumPy file which is used as the integrated input when it comes to the graph construction and model training.

4.2.4 Data Reduction

Subjects with no phenotypic file, no functional scan, no BOLD run were located, time-series length under 172, or on less than 90 usable atlas regions and all subjects failing these checks were dropped. Following preprocessing and reduction, 637 subjects were left out of an original number of 1,036 potential scans, which is a trade-off between data completeness and temporally atlas consistent results. Re calculation of class distribution and site distribution was done at this point to ensure that adequate ADHD and control samples are available at each center.

4.2.5 Summary of Preprocessed Data

The final processed data contains 637 subjects (172×90 regional time series each), a normalized 3 dimensional phenotype, and a binary ADHD/control tag.

Site Name	Subjects
NYU	201
NeuroIMAGE	57
Peking_1	187
Peking_2	65
Peking_3	42
Pittsburgh	85

Table 4.1: Site Distribution Table

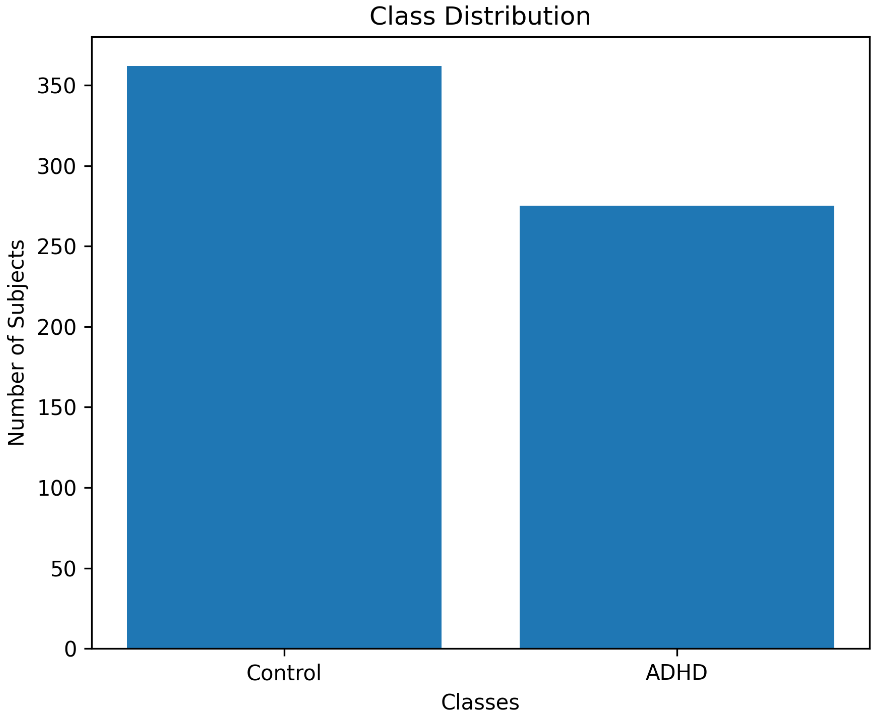

Figure 4.1: Class Distribution after Preprocessing

The number of controls (n = 362) is a bit more than the number of ADHD subjects (n = 275), and the data collection covers several acquisition locations, the biggest subsets being those of NYU and Peking.

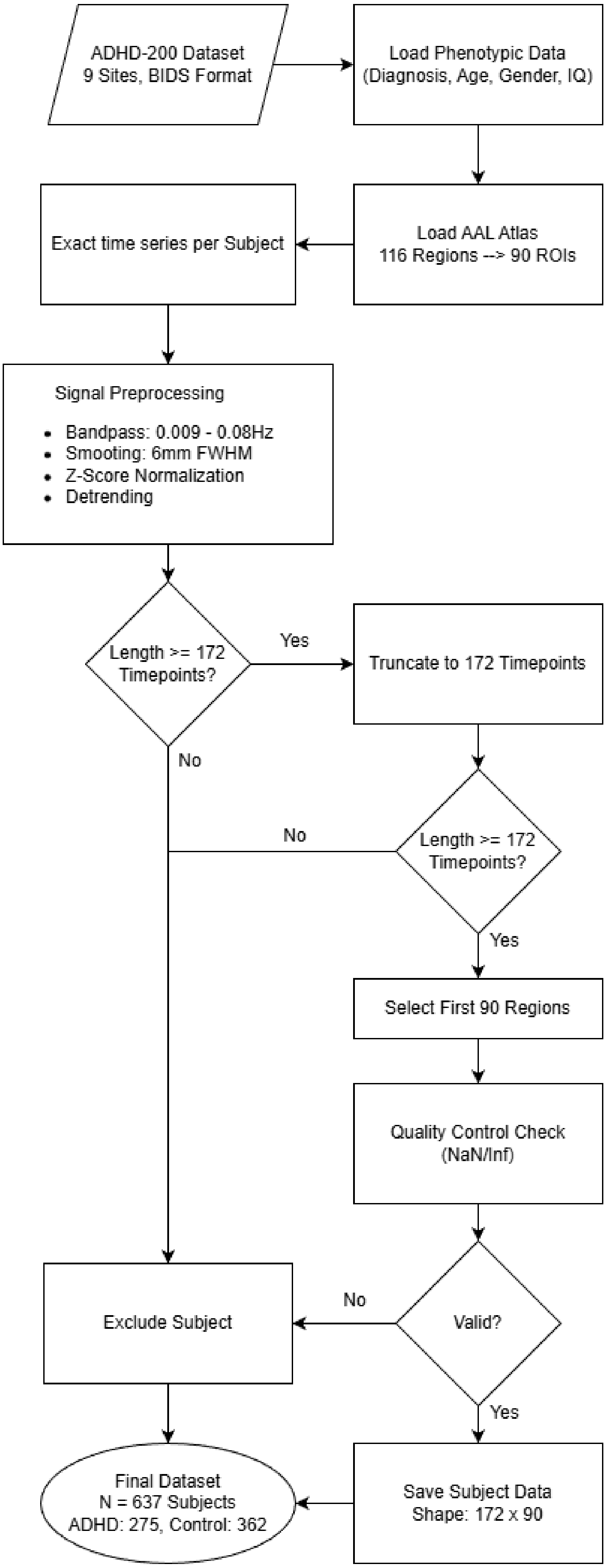

Figure 4.2: Pre-processing line of the raw multi-site fMRI and phenotypic files to the single NumPy data set to be used in graph construction

4.3 Implementation of Selected Design

4.3.1 GAT-PhoNet (Graph Attention and Phenotype Fusion)

The proposed model can be treated as a derived AAL functional connectivity graph that is represented as the fully attributed graph representing each subject and the result is a graph attention branch and a phenotype MLP branch. Experiments involving all deep learning models were run in Python with PyTorch and PyTorch Geometric, and were trained on a single GPU(Google Colab T4 GPU) and stratified train/validation/test subsets controlled by a global random seed to ensure reproducibility.

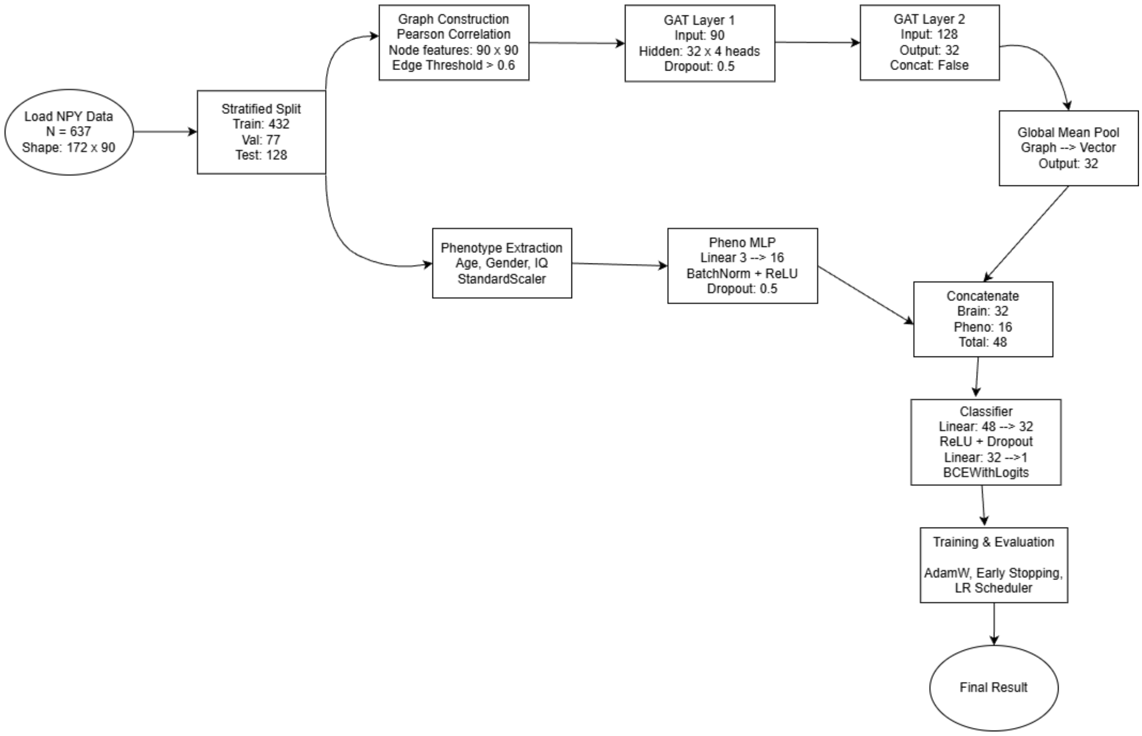

Figure 4.3: Flowchart of entire work pipeline of GAT-PhoNet

Graph Construction

Individual subject time series had their 172 × 90 matrix transposed to 90 × 172, and transformed into a 90 × 90 Pearson correlation between regions, where NaNs were set to zero. The node-feature matrix was taken to be a correlation matrix directly (the feature vector of each node is its correlation profile with all other regions), and an edge was formed between two nodes when their absolute correlation was at least a fixed threshold. The sparse form of edge indices and edge weights were stored and the resulting graph along with the phenotype vector and label wrapped as a PyTorch Geometric Data object.

Model Architecture

The brain branch is represented by two graph attention layers with a stack, each layer is a multi-head Graph Attention Convolutional layer (GATConv) that takes 90 features and H hidden channels, and the other layer sums the heads producing an H dimensional node embedding.

$$h_i' = \sigma\left(\sum_{k=1}^{K} \sum_{j \in \mathcal{N}(i)} \alpha_{ij}^k \, W^k h_j\right) \tag{4.1}$$

Following graph convolution, a dropout regularized global mean pooling will then aggregate the node embeddings into a single H dimensional graph representation.

$$r_G = \frac{1}{|\mathcal{V}|} \sum_{i \in \mathcal{V}} h_i' \tag{4.2}$$

The phenotype branch is a feed forward MLP containing one hidden layer: 3 to 16 unit linear projection, then batch normalization, ReLU activation, and dropout. The fused representation is simply achieved by concatenating the H dimensional graph embedding with the 16 dimensional phenotype embedding and running it through a two layer classifier (Linear ReLU Dropout Linear) to give a single logit by which to predict binary ADHD.

$$z = W_2\Big(Dropout\big(ReLU\big(W_1[\, r_G \,\|\, h_p \,] + b_1\big)\big)\Big) + b_2 \tag{4.3}$$

Training Configuration

Each model was trained on top of AdamW optimizer and learning rate of 0.001, weight 5×10^{-4} and dropout 0.5 both in the branches and the following classifier. The loss function was binary cross entropy with logits, where the positive class weight is the ratio of negative samples in the training example to the positive samples.

$$\mathcal{L} = -\left[w_{pos}\, y \cdot \log(\sigma(z)) + (1-y) \cdot \log(1-\sigma(z))\right], \quad \text{where } w_{pos} = \frac{N_{neg}}{N_{pos}} \tag{4.4}$$

A simple early stopping policy was followed that kept track of loss of validation with a patience of 10 epochs and the model snapshot with the best performance was stored and subsequently used to evaluate and report the final test results.

4.3.2 Graph Convolutional Network

Model Architecture

In the suggested framework, the graph convoluted neural network (GCN) is used to simulate the functional brain relatedness with the aim of automatic classification of ADHD. Resting-state fMRI data of each subject is presented as a graph giving the network the opportunity to learn topological and functional relationships between brain regions as well as consider phenotypic information.

Graph Construction and Input Representation

All the subjects are represented as undirected graphs (G = (V, E)) with nodes (V) representing 90 brain regions that are labeled using the Automated Anatomical Labeling (AAL) atlas. The features of nodes are obtained using resting-state fMRI time series, and each node is described as a 172-dimensional feature of the time signal of the region. Edges are functional connectivity between brain regions and are calculated using Pearson correlation coefficient between node time series. The top 20 percent of strongest absolute correlation are kept to decrease noise and impose sparseness, which leave only a small and biologically-realistic graph structure.

$$W_{ij} = \frac{\sum_{t=1}^{T} (x_{i,t} - \bar{x}_i)(x_{j,t} - \bar{x}_j)}{\sqrt{\sum_{t=1}^{T} (x_{i,t} - \bar{x}_i)^2 \sum_{t=1}^{T} (x_{j,t} - \bar{x}_j)^2}} A_{ij} = \begin{cases} 1, & \text{if } |W_{ij}| \in \text{top } 20\% \\ 0, & \text{otherwise} \end{cases} \tag{4.5}$$

Besides neuroimaging data, every subject is related to phenotypic variables (age, gender, and IQ), which are introduced at a later point to the model.

$$Y_{pred} = Softmax(MLP(Concat(h_{GCN}, x_{pheno}))) \tag{4.6}$$

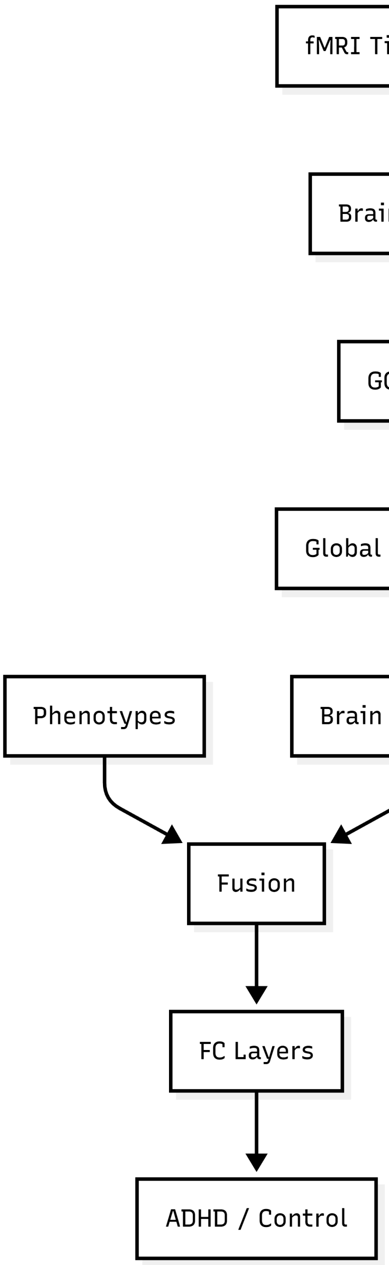

Figure 4.4: Flowchart of Graph Convolutional Network

The main aspect of the model is a sequence of three stacked graph convolutional layers, which are based on the Graph Convolutional Network (GCN) operator. The layers recursively learn node representations through the combination of information on neighboring nodes, enabling the model to learn both local and higher-order functional interactions between brain regions. The model is able to learn hierarchical representations, which captures regional activity and global network organization, by stacking several graph convolutional layers.

$$H^{(l+1)} = \sigma\left(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}H^{(l)}W^{(l)}\right) \tag{4.7}$$

To get a representation of each node in subject level, global mean pooling is implemented on all the nodes of each graph. It is an operation that compiles node features into a fixed-length vector, generating a succinct description of the overall pattern of brain connectivity of the subject but being independent of node ordering. A fully connected layer is then used to refine the pooled representation giving a 64-dimensional embedding that encodes salient brain network features useful to the classification of ADHD.

Multimodal Feature Fusion and Classification

The proposed model has explicit graph structure and differentiable architecture, which is why it supports post-hoc interpretability. Integrated Gradients are used in order to approximate the value of individual brain parts to the model prediction by giving node properties significance marks. Such scores may be summed up over time points to establish the parts of the brain that may serve as biomarkers of ADHD.

The findings reveal a balanced accuracy in the two classes, thus indicating that the model does not show a high degree of bias, either in making predictions of ADHD or control. Nevertheless, the recall of the ADHD class is significantly lower than the recall of the control class, showing the natural challenge in identifying neurobiological patterns of ADHD-related issues because of its heterogeneity and similarity to normal development.

The model is more specific to the control class, indicated by a rather small number of false positives. On the other hand, the fact that the number of false negatives in the ADHD group is higher means that subtle patterns of connectivity related to ADHD might not always be salient enough above controls, especially in multi-site data.

4.3.3 RoI Transformer

Model Architecture

This model is a transformer-based architecture that can be used to learn the global spatial dependencies of the fMri data. In contrast to conventional methods where the fMRI data is considered as time-sequences, our model considers the brain regions as unique tokens and their time-series activity as feature embeddings. This enables the self-attention mechanism to learn a flexible functional connection pattern between regions, which is an effective optimisation of the connectome in the brain to classify.

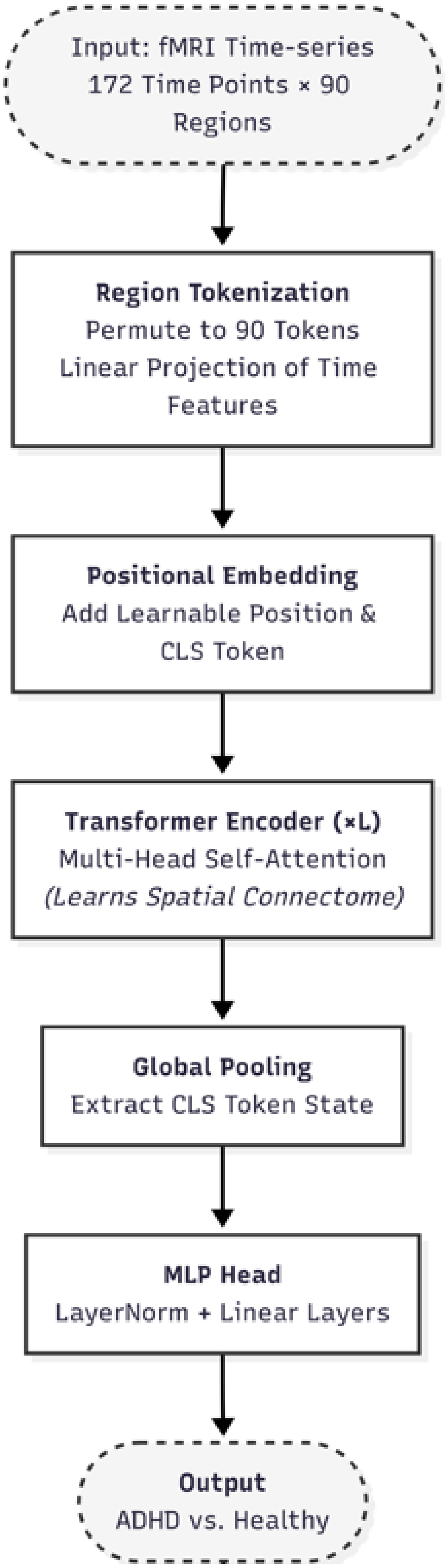

Figure 4.5: Flowchart of the entire pipeline of RoI Transformer

Data Representation and Embedding

Consider the fMRI scan of a single subject as input to be represented as a matrix $X \in \mathbb{R}^{T \times R}$ of Z where $T = 172$ is time points, and $R = 90$ are Regions of Interest (RoIs) defined in this case by the AAL atlas. In order to exploit the Transformer architecture to compute spatial correlation, we transpose the input to $X' \in \mathbb{R}^{R \times T}$, where r_i in this case is considered a token and the time dynamics is used as a base feature embedding. We use linear projection $W_p \in \mathbb{R}^{T \times D}$ where we project the time-series features to a latent dimension D.

It is marginally prepended with a learnable classification token z_{cls} and learnable positional embeddings E_{pos} to store the information of region identity are added:

$$Z_0 = [Z_{cls} \,;\, X'_1 W_p \,;\, \ldots \,;\, X'_R W_p] + E_{pos} \tag{4.8}$$

Multi-Head Self-Attention

The Multi-Head Self-Attention layer is the main element, and it calculates the association among the brain parts. Attention mechanism of a given head is defined as:

$$Attention(Q, K, V) = softmax\left(\frac{QK^\top}{\sqrt{d_k}}\right)V \tag{4.9}$$

Here Q, K, V are linear projections of the features of the region. The SoftMax itself is a useful term that essentially is a learnable soft-maximum-entering-soft-maximum function connectivity matrix and serves to define the strength of attention of one brain region to another during the task.

Encoder and Classification

The model is made up of L encoder layers. The layers utilize Layer Normalization (LN) and a Multi-Layer Perceptron (MLP) with residual links.

$$Z'_l = \mathrm{MSA}(\mathrm{LN}(Z_{l-1})) + Z_{l-1}, \tag{4.10}$$

$$Z_l = \mathrm{MLP}(\mathrm{LN}(Z'_l)) + Z'_l \tag{4.11}$$

Lastly, a class token state of Z is obtained at the final layer and processed through a classification head (LayerNorm, Linear, GELU, Linear) to obtain the probability of ADHD:

$$y = Softmax\big(Classifier\big(z_{cls}^{L}\big)\big) \tag{4.12}$$

4.3.4 AttentionTabNet

Data Preprocessing and Feature Engineering

In order to model the functional connectivity of the brain we converted raw fMRI time-series data into fixed size feature vectors. $T \in \mathbb{R}^{T \times R}$ is denoted as time-series data of one subject, and R is the count of regions of interest (ROIs), and t is the count of time points.

To compute the functional connectivity matrix C, we used Pearson correlation coefficient. Then for computing the connectivity strength between two regions i and j, using:

$$r_{ij} = \frac{\sum_{k=1}^{T} (x_{i,k} - \bar{x}_i)(x_{j,k} - \bar{x}_j)}{\sqrt{\sum_{k=1}^{T} (x_{i,k} - \bar{x}_i)^2} \sqrt{\sum_{k=1}^{T} (x_{j,k} - \bar{x}_j)^2}} \tag{4.13}$$

Where $x_{i,k}$ represents the signal intensity of the region i at time k, and $\bar{x}_i$ is the mean signal of the region.

In order to solve redundancy, we removed the upper triangle of the correlation matrix (excluding the diagonal), and flattened it to a connectivity vector V_{conn}. This vector was concatenated with phenotypic information V_{pheno} to make the final input feature vector X of every subject:

$$X = [V_{\text{conn}} \oplus V_{\text{pheno}}] \tag{4.14}$$

Model Architecture

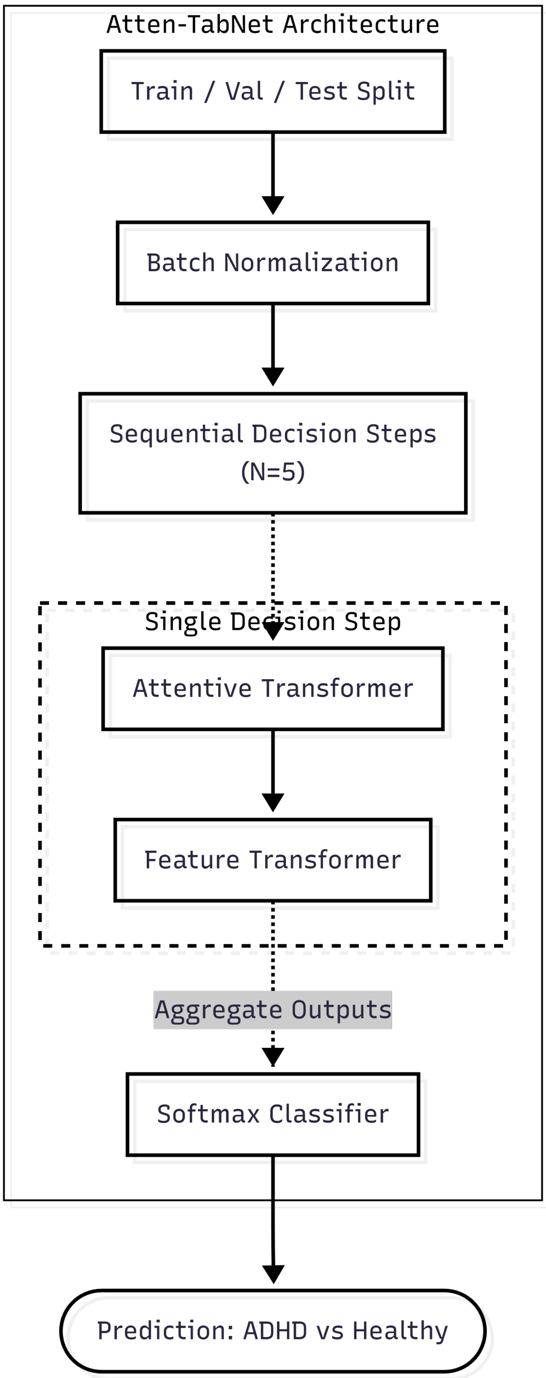

Figure 4.6: Flowchart of AttentionTabNet

We applied TabNet which is a deep tabular architecture of learning a network which integrates the interpretability of decision trees and the representation learning of deep neural networks. TabNet has a sequential multi-step architecture in which each decision step takes a set of particular features with a learnable mask.

The model also executes the input features $f \in \mathbb{R}^D$ using N steps of decision making. In every step i, the model uses an Attentive Transformer to compute a learnable mask $M[i]$, which defines the significance of features of that particular step:

$$M[i] = entmax(P[i-1] \cdot h_i) \tag{4.15}$$

In this case, $P[i-1]$ denotes the prior scale (reliance applied in the preceding stages), h_i denotes the processed representation and **entmax** is a sparse normalization function that permits soft feature selection.

The chosen features are further submitted to a Feature Transformer which computes them to produce two outputs: $d[i]$ (to be used in the final prediction) and $a[i]$ (to be used in the next decision step). The outcome of the linear combination of the production of each decision step is taken as the final classification decision and then transformed by the softmax:

$$\hat{y} = Softmax\left(\sum_{i=1}^{N_{steps}} ReLU(d[i])\right) \tag{4.16}$$

A sparsity regularization term L sparse is introduced to the total loss in order to promote parameter efficiency and interpretability, and the entropy of the mask values.

Chapter 5

Result Analysis

5.1 Performance of the Proposed GAT-PhoNet Model

The test accuracy of the proposed GAT-PhoNet architecture was 72.09% and the ROC-AUC = 0.7517, which is moderate but encouraging, indicated a discriminative ability of the proposed GAT-PhoNet architecture between the ADHD and healthy control groups. The model was well balanced in classes depending on the F1-score of 0.74 in Healthy controls and 0.68 in ADHD subjects as indicated in the classification report (Table 5.1).

5.1.1 Classification Report Analysis of GAT-PhoNet

The macro average F1-score (0.71) suggests that the treatment of both classes is fairly fair in spite of the inbuilt class disparity in the dataset (73 healthy and 55 ADHD subjects included in the test sample).

Class	Precision	Recall	F1-Score	Support
Control	0.76	0.71	0.74	73
ADHD	0.65	0.71	0.68	55
Accuracy			0.72	128
Macro Average	0.71	0.71	0.71	128
Weighted Average	0.72	0.71	0.71	128

Table 5.1: Classification performance metrics

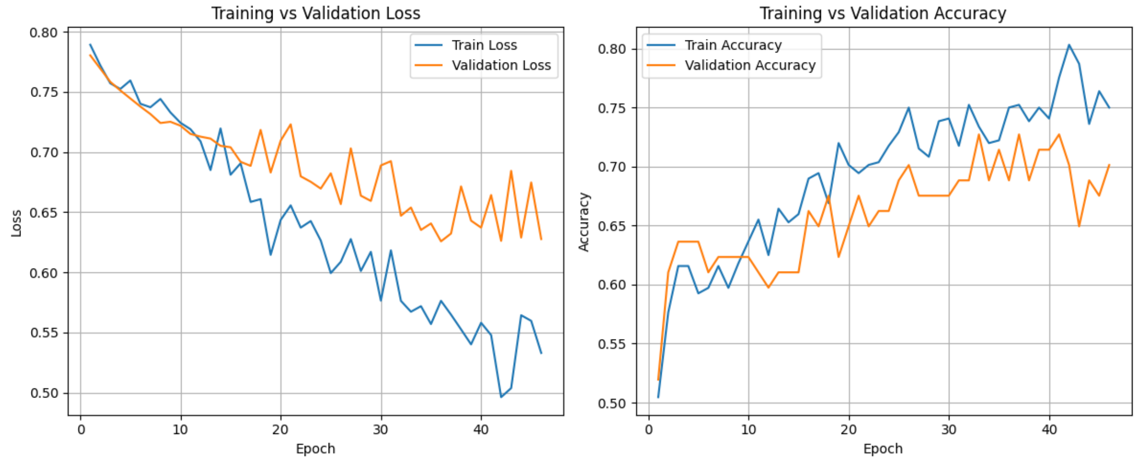

Figure 5.1: Learning Curves (Training/validation loss and accuracy)

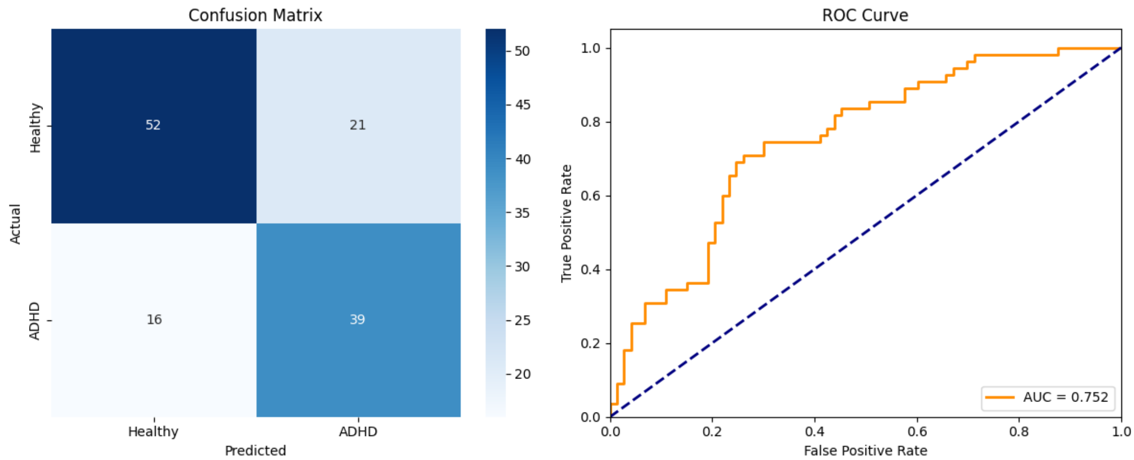

Figure 5.2: Confusion Matrix, ROC curve

The fact that the model is a little more precise in case of Healthy controls (0.76) than in ADHD (0.65) indicates a conservative bias to labeling ambiguous cases as healthy, a trend in line with the difficulty in finding heterogeneous neurobiological responses to ADHD within variation in typical development.[8], [9]

5.1.2 Ablation Study: Contribution of Multimodal Components

The ablation study (Table 5.2) provides critical insights into the value of multimodal integration in our framework:

Model Variant	Accuracy	AUC-ROC	F1-Score
Combined Model (GAT + Phenotype)	0.6953	0.7579	0.6549
Neuroimaging Component Only (GAT)	0.6250	0.6057	0.6190
Phenotypic Data Only (MLP)	0.6562	0.7098	0.6667

Table 5.2: Performance comparison of different model variants

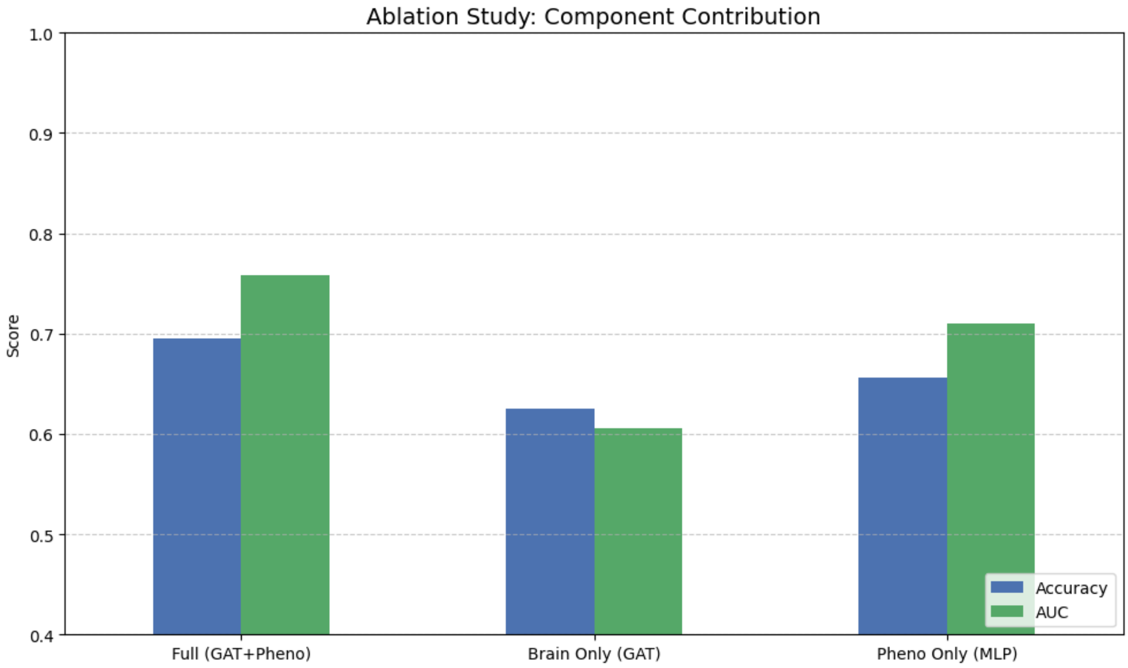

Figure 5.3: Bar chart comparison scores across ablation variants

The full hybrid model has always performed better in all the measures as compared to the unimodal variants with the worst result belonging to the Brain-Only GAT variant (62.50% accuracy). It is important to mention that the Phenotype-Only MLP with no neuroimaging data obtained demonstrated a competitive result in the terms of F1-score (0.6667), which supports the diagnostic significance of demographic factors (age, gender, IQ) in ADHD classification[11]. Nevertheless, the 7.03 percent of the improvement in accuracy caused by multimodal fusion indicates that brain connectivity patterns can help in furnishing the information that is not as effectively captured by phenotypic variables alone, especially in understanding the neurobiological cause of ADHD.

5.2 Comparative Analysis with Baseline Architectures

When benchmarked against alternative deep learning approaches (Table 5.3), GAT-PhoNet demonstrated superior performance:

Model	Accuracy	F1-Score	Key Observation
GAT-PhoNet (proposed)	0.7209	0.6783	Demonstrated the best overall performance, attributed to multimodal fusion.
GCN	0.69	0.67	Exhibited strong specificity for the Healthy class (recall = 0.82) but limited performance in recalling the ADHD class (recall = 0.51).
RoI-Transformer	0.58	0.57	Performed significantly below expectations, potentially due to insufficient training data for the Transformer architecture.
AttentionTabNet	0.64	0.63	Achieved moderate performance, characterized by robust recall for the Healthy class (0.75) but weak recall for the ADHD class (0.49).

Table 5.3: Performance comparison and qualitative observations across different models

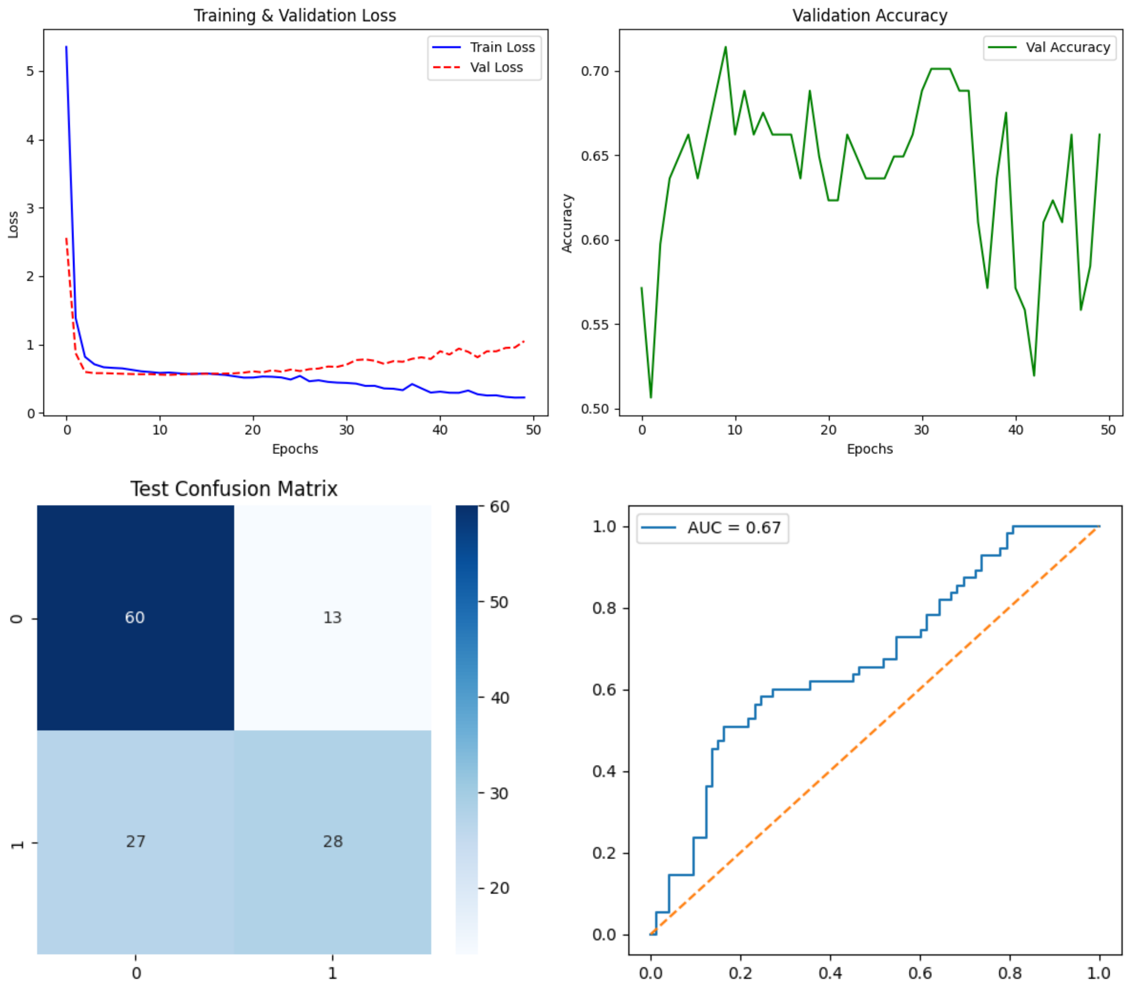

Figure 5.4: GCN learning curves, confusion matrix

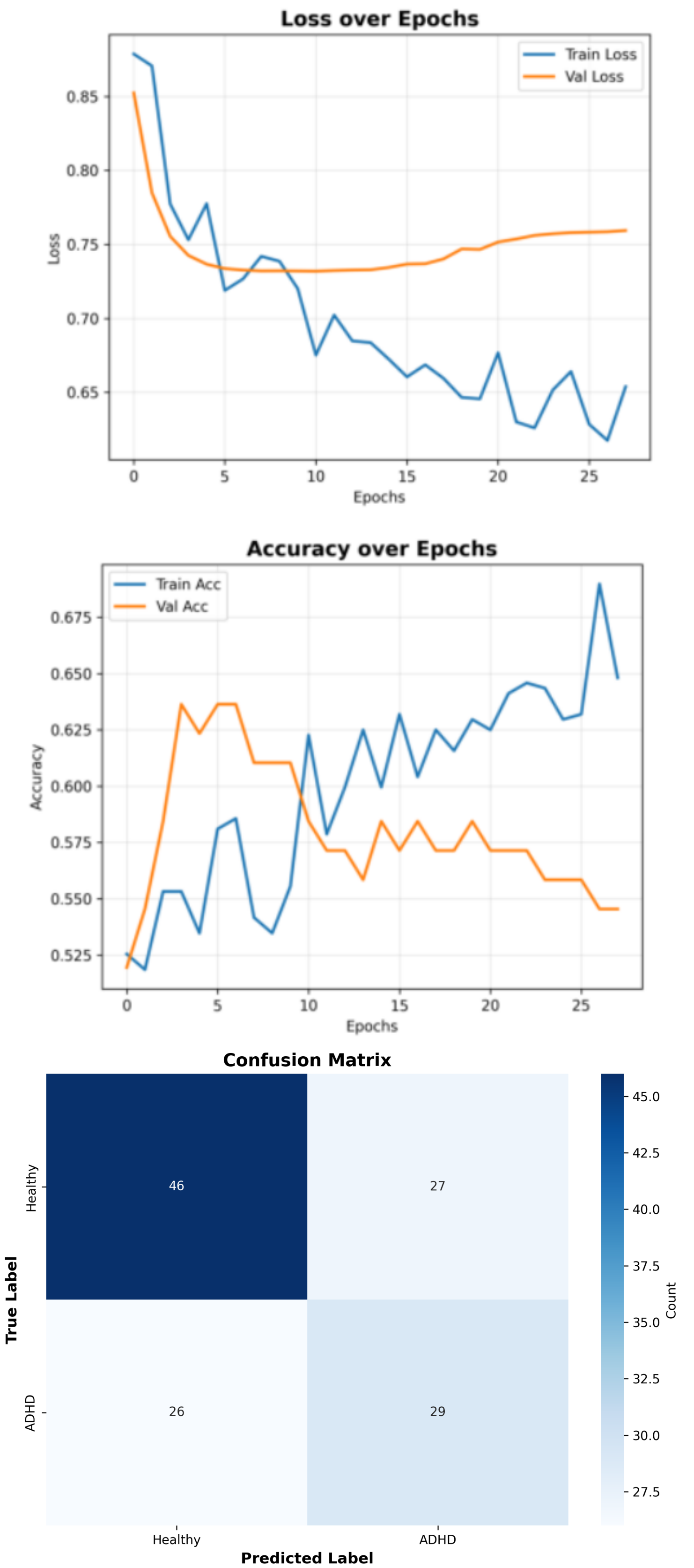

Figure 5.5: RoI-Transformer model learning curves, confusion matrix

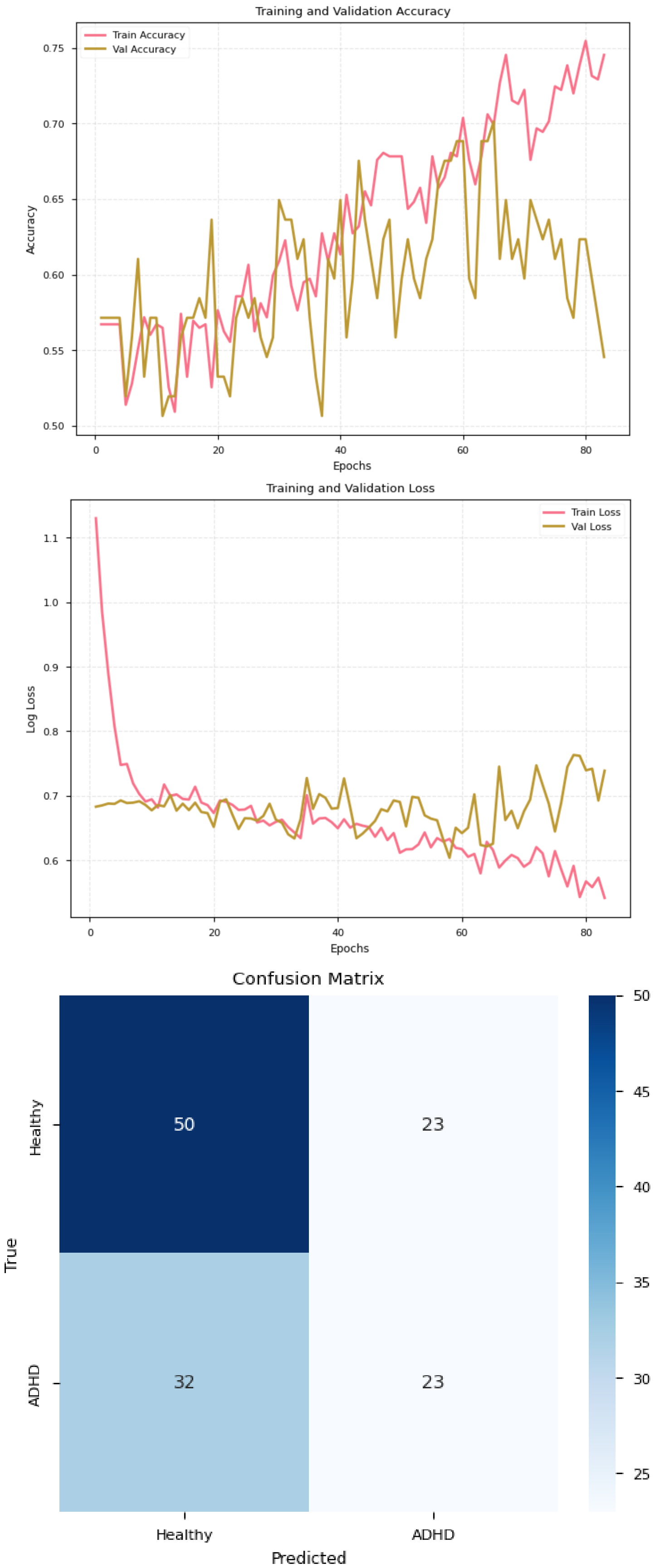

Figure 5.6: Attention TabNet model learning curves, confusion matrix

The GCN model presented an interesting performance profile as it was considerably specific to healthy controls (82% recall) but considerably smaller in terms of sensitivity to ADHD cases (51% recall) showing a bias towards the majority class. In a similar manner, AttentionTabNet maintained a good healthy-class recall (75%) and critically low ADHD recall (49%), which indicates the ongoing problem of detecting positive cases in imbalanced neuroimaging datasets. The poor performance of the RoI-Transformer (58% accuracy) indicates that potentially much more data will need to be gathered by transformer architecture to capture long-range interaction in functional connectivity networks.

5.3 Discussions

5.3.1 Model Performance Interpretation

The average performance of GAT-PhoNet (72.09) is consistent with the state-of-the-art in the classification of single subjects with ADHD using the information provided by rsfMRI. As the model is not the most accurate diagnostic element, it allows to combine complementary streams of information: graph-based data on functional connectivity can induce spatial organization of brain networks, whereas phenotypic variables can be adopted as essential demographic information. The ablation experiment supports the fact that either modality is not efficient enough to yield the best performance- neuroimaging is more specific in a biological context whereas phenotypes are specific in a population context. The continuous problem of lower ADHD recall than precision in all models is the heterogeneous aspect of ADHD neurobiology. In contrast to the disorders with focal lesions, the ADHD can be characterized by subtle, distributed changes in functional connectivity overlapping with normal developmental variation to a significant degree- especially in multi-site datasets where the protocols of acquisition may not be the same.

5.3.2 Limitations

Dataset Constraints and Generalizability: The ADHD-200 dataset, despite being useful in its multi-site composition form, is associated with several challenges such as site-specific differences of acquisition, variable preprocessing pipelines across contributing centers and high heterogeneity of the ADHD diagnostic category (it combines inattentive, hyperactive-impulsive and combined subtypes). These contribute to noise which essentially reduces the possible classification accuracy.

5.3.3 ADHD-200 Performance Ceiling Competition

Our findings need to be located in the context of the larger research frame. Even in the ADHD-200 Global Competition (2011) that challenged researchers around the world to create automated ADHD classifiers, the winning accuracy was an insignificant 63%, or just a bit better than that of chance, based on the distribution of classes in the dataset [1], [2]. Later findings with ADHD-200 have always recorded the range of accuracy at 60 to 75% implying that there is a limit to the performance of the data set not necessarily due to flaws in the methodology. The accuracy of our GAT-PhoNet is therefore 72.09% accuracy, which is a competitive result in this limited ecosystem.

5.3.4 Overfitting concerns

There is a slight disparity between the optimal validation accuracy (69%), and the final test accuracy (72.09) indicating evidence of possible overfitting despite the use of regularization techniques (dropout, early stopping). This could be an indication of the small sample size that was left following intense preprocessing (637 subjects who survived 1,036) of samples, restricting the model to generalization outside of training distribution. It is also characterized by the relatively small size of the test set (128 subjects) that also leads to variation in the estimation of performance.

5.3.5 Class Imbalance

Though it used the class-weighted loss functions, all models showed asymmetrical results in performance, with a better recall in healthy controls as compared to the ADHD subjects. This is a symptom of the inherent imbalance of the data (362 pairs of controls, versus 275 groups of ADHD subjects after preprocessing) as well as the neurobiological truth that ADHD has no single, strong imaging biomarker.[11]

Chapter 6

Conclusion

6.1 Summary of Findings

This study indicates that deep learning schemes based on multimodal representations that combine graph-based modeling of connectivity of brain functions with phenotypic response variables can moderate but significant accuracy in the classification of ADHD. Our hypothesis, that complementary streams of information improve discriminative power, was confirmed when the proposed GAT-PhoNet architecture with the use of Graph Attention Networks to adaptively weight functional connections and a phenotype processing branch was compared to unimodal variants and alternative architectures (GCN, Transformer, TabNet).

Although the absolute accuracy still falls short of the requirements to deploy it to clinical settings, our efforts would help advance the current knowledge on ADHD neurobiology with interpretable graphs. The salient attentional mechanisms of GAT layers provide avenues toward the detection of salient brain regions and connections with ADHD, a move in the direction of biologically based diagnostic biomarkers.

6.2 Contributions to the Field

The primary contribution of this work is the proposal of GAT-PhoNet, a multimodal architecture that optimizes ADHD classification by leveraging both the spatial topology of the brain and clinical phenotypic data. The specific contributions of this research are summarized as follows:

- Multimodal Architecture Design: We developed a dual-stream framework that fuses graph-attentional neural features with an MLP-based phenotypic branch (Age, IQ, Gender), achieving a peak accuracy of 72.09% on the ADHD-200 dataset.
- Graph Topological Refinement: We established a pipeline for transforming 4D fMRI signals into 90-node graphs using the AAL atlas. By implementing a 0.6 Pearson correlation threshold, we effectively removed spurious connections while preserving the functional 'small-world' network properties essential for neuroimaging analysis.

- Dynamic Connectivity Weighting: We used Multi-Head Graph Attention (GAT) to determine non-linear relations between regions of interest (ROIs). This way it allows to assign pathologically significant edges with learnable weights in the model which helps to avoid over-smoothing like the traditional models from the GNN family.
- Architectural Validation: We provided an empirical benchmarking study against GCN, TabNet, and ROI Transformers. Our results demonstrate that GAT-PhoNet's local attention mechanism offers superior noise-robustness and generalization over 'spatial-blind' tabular models and data-intensive global attention architectures.

6.3 Recommendations for Future Work

In the future, our focus will be:

1. Combining larger harmonized multi-site datasets using standardized acquisition procedures.
2. Including other data modalities (structural MRI, diffusion tensor imaging, cognitive measures).
3. Building subtype-specific classifiers that recognize the heterogeneity of ADHD.
4. External validation in independent cohorts. Finally, these methods can change neuroimaging to a research instrument to a clinical support instrument of objective, yet biology-informed mental health assessment — supplementing but not replacing extensive clinical assessment.

Bibliography

[1] T. Yarkoni, *Brain-based prediction of adhd—now with 100% fewer brains!* https://talyarkoni.org/blog/2011/10/12/brain-based-prediction-of-adhd-now-with-100-fewer-brains/, Accessed: 2026-01-28, 2011.

[2] ADHD-200 Consortium, *The ADHD-200 global competition: Results*, https://fcon_1000.projects.nitrc.org/indi/adhd200/results.html, Accessed: 2026-01-28, 2012.

[3] S. Qiu, G. H. Chang, M. Panagia, D. M. Gopal, R. Au, and V. B. Kolachalama, "Fusion of deep learning models of mri scans, mini–mental state examination, and logical memory test enhances diagnosis of mild cognitive impairment," *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, vol. 10, pp. 737–749, 2018. DOI: 10.1016/j.dadm.2018.08.013

[4] E. Jabason, M. O. Ahmad, and M. N. S. Swamy, "Multimodal neuroimaging fusion in nonsubsampled shearlet domain using location-scale distribution by maximizing the high-frequency subband energy," *IEEE Access*, vol. 7, pp. 97 865–97 886, 2019. DOI: 10.1109/ACCESS.2019.2930225 [Online]. Available: https://doi.org/10.1109/ACCESS.2019.2930225

[5] S. Abdolmaleki and M. S. Abadeh, "Brain mr image classification for adhd diagnosis using deep neural networks," in *Proc. 2020 International Conference on Machine Vision and Image Processing (MVIP)*, IEEE, 2020, pp. 1–5. DOI: 10.1109/MVIP49855.2020.9116877

[6] L. Zhang, M. Wang, M. Liu, and D. Zhang, "A survey on deep learning for neuroimaging-based brain disorder analysis," *Frontiers in Neuroscience*, vol. 14, p. 779, 2020. DOI: 10.3389/fnins.2020.00779 [Online]. Available: https://www.frontiersin.org/journals/neuroscience/articles/10.3389/fnins.2020.00779/full

[7] J. Zhao et al., "Functional network connectivity (fnc)-based generative adversarial network (gan) and its applications in classification of mental disorders," *Journal of Neuroscience Methods*, vol. 341, p. 108 756, 2020. DOI: 10.1016/j.jneumeth.2020.108756

[8] T. Li et al., "Characterizing neuroanatomic heterogeneity in people with and without adhd based on subcortical brain volumes," *Journal of Child Psychology and Psychiatry*, vol. 62, no. 9, pp. 1140–1149, 2021. DOI: 10.1111/jcpp.13384

[9] K. E. Nett et al., "Infralimbic cortex functioning across motivated behaviors: Can the differences be reconciled?" *Neuroscience & Biobehavioral Reviews*, vol. 131, pp. 704–721, 2021. DOI: 10.1016/j.neubiorev.2021.10.002

[10] X. Chen et al., "Recent advances and clinical applications of deep learning in medical image analysis," *Medical Image Analysis*, vol. 79, 2022, Art. no. 102444. DOI: 10.1016/j.media.2022.102444 [Online]. Available: https://doi.org/10.1016/j.media.2022.102444

[11] F. Dehghani Firouzabadi, S. Ramezanpour, M. Dehghani Firouzabadi, I. J. Yousem, N. A. J. Puts, and D. M. Yousem, "Neuroimaging in attention-deficit/hyperactivity disorder: Recent advances," *AJR American Journal of Roentgenology*, vol. 218, no. 2, pp. 321–332, 2022. DOI: 10.2214/AJR.21.26316

[12] M. Cao, E. Martin, and X. Li, "Machine learning in attention-deficit/hyperactivity disorder: New approaches toward understanding the neural mechanisms," *Translational Psychiatry*, vol. 13, p. 236, 2023. DOI: 10.1038/s41398-023-02536-w

[13] Z. Ye, Y. Qu, Z. Liang, M. Wang, and Q. Liu, "Explainable fmri-based brain decoding via spatial temporal-pyramid graph convolutional network," *Human Brain Mapping*, vol. 44, no. 7, pp. 2921–2935, 2023. DOI: 10.1002/hbm.26255 [Online]. Available: https://doi.org/10.1002/hbm.26255

[14] Y. Alkhurayyif and A. R. Wahab Sait, "Deep learning-driven dyslexia detection model using multi-modality data," *PeerJ Computer Science*, vol. 10, e2077, 2024. DOI: 10.7717/peerj-cs.2077 [Online]. Available: https://peerj.com/articles/cs-2077/

[15] B. French et al., "The impacts associated with having adhd: An umbrella review," *Frontiers in Psychiatry*, vol. 15, 2024, Art. no. 1343314. DOI: 10.3389/fpsyt.2024.1343314 [Online]. Available: https://doi.org/10.3389/fpsyt.2024.1343314

[16] N. Huynh et al., "The use of generative adversarial networks and graph convolution networks for neuroimaging-based diagnostic classification," *Brain Sciences*, vol. 14, no. 5, 2024, Art. no. 456. DOI: 10.3390/brainsci14050456 [Online]. Available: https://doi.org/10.3390/brainsci14050456

[17] S. S. and M. B. Anandarajuuthor, "Classification of adhd and its sub-types using machine learning models," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 12, no. 3, pp. 1256–1262, 2024. [Online]. Available: https://ijisae.org/index.php/IJISAE/article/view/5406

[18] C. Shore, S. M. P. Norris, C. Berkower, N. Ontjes, and K. R. Babik, "Impact of misdiagnosis, bias, and stigma," in *Adult Attention-Deficit/Hyperactivity Disorder: Diagnosis, Treatment, and Implications for Drug Development: Proceedings of a Workshop*, National Academies of Sciences, Engineering, and Medicine, Health and Medicine Division, Board on Health Sciences Policy, Forum on Neuroscience and Nervous System Disorders, and Forum on Drug Discovery, Development, and Translation, Eds., Washington (DC): National Academies Press (US), Aug. 14, 2024. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK606330/

[19] G. Taspinar and N. Ozkurt, "A review of adhd detection studies with machine learning methods using rsfmri data," *NMR in Biomedicine*, vol. 37, no. 8, e5138, 2024. DOI: 10.1002/nbm.5138

[20] D. C. Lohani, V. Chawla, and B. Rana, "A systematic literature review of machine learning techniques for the detection of attention-deficit/hyperactivity disorder using mri and/or eeg data," *Neuroscience*, vol. 570, pp. 110–131, 2025. DOI: 10.1016/j.neuroscience.2025.02.019

[21] A. Zhang and H. Zhang, "Predicting depression in healthy young adults: A machine learning approach using longitudinal neuroimaging data," *NeuroImage*, vol. 315, p. 121 285, 2025. DOI: 10.1016/j.neuroimage.2025.121285