



Project Phase 2 Report On

Autism Spectrum Disorder diagnosis through brain network analysis

*Submitted in partial fulfillment of the requirements for the
award of the degree of*

Bachelor of Technology

in

Computer Science and Engineering

By

Aleena Siby(U2003024)

Bibit Sebastian (U2003055)

Celestian Ben Mathew(U2003058)

Christo Mathew(U2003060)

Under the guidance of

Ms. Seema Safar

**Department of Computer Science and Engineering
Rajagiri School of Engineering & Technology (Autonomous)
(Parent University: APJ Abdul Kalam Technological University)**

Rajagiri Valley, Kakkanad, Kochi, 682039

April 2024

CERTIFICATE

*This is to certify that the project report entitled "**Autism Spectrum Disorder diagnosis through brain network analysis**" is a bonafide record of the work done by **Aleena Siby (U2003024)**, **Bibit Sebastian (U2003055)**, **Celestian Ben Mathew (U2003058)** and **Christo Mathew(U2003060)**, submitted to the Raja-giri School of Engineering & Technology (RSET) (Autonomous) in partial fulfillment of the requirements for the award of the degree of Bachelor of Technology (B. Tech.) in Computer Science and Engineering during the academic year 2023-2024.*

Ms. Seema Safar
Project Guide
Assistant Professor
Dept. of CSE
RSET

Dr.Sminu Izudheen
Project Coordinator
Professor
Dept. of CSE
RSET

Dr. Preetha K.G
Head of the Department
Professor
Dept. of CSE
RSET

ACKNOWLEDGMENT

We wish to express our sincere gratitude towards **Dr. P.S Sreejith**, Principal of RSET, and **Dr. Preetha K.G**, Head of the Department of Computer Science and Engineering for providing us with the opportunity to undertake our project, "Autism Spectrum Disorder diagnosis through brain network analysis".

We are highly indebted to our project coordinator, **Dr.Sminu Izudheen**, Professor, Department of Computer Science and Engineering, for their valuable support.

It is indeed our pleasure and a moment of satisfaction for us to express our sincere gratitude to our project guide **Ms.Seema Safar** for her patience and all the priceless advice and wisdom she has shared with us.

Last but not the least, we would like to express our sincere gratitude towards all other teachers and friends for their continuous support and constructive ideas.

Aleena Siby

Bibit Sebastian

Celestian Ben Mathew

Christo Mathew

Abstract

ASD, is a lifelong neurodevelopmental condition that affects social communication and behaviour. It poses challenges in early detection due to its complex nature. The current clinical diagnosis process relies on behavioural assessments, which are subjective and may be prone to misdiagnosis. Recent studies utilize neuroimaging techniques for the detection of ASD aiming for a quantitative approach. So, this project focuses on autism spectrum disorder (ASD) detection through advanced neuroimaging techniques. Functional MRI (fMRI) data utilizes a Fourier transform encoder to capture the dependencies between different brain regions. The extracted features are classified using the softmax function. This comprehensive approach aims to enhance the accuracy and sensitivity of ASD detection by using the complementary information provided by fMRI.

Magnetic Resonance Imaging (MRI) emerges as a promising method for autism identification, offering distinct advantages over other technologies. In contrast to computer vision and eye tracking systems, which encounter challenges like privacy concerns and environmental limitations, MRI provides a superior approach to autism detection. Compared to techniques like EEG and MEG, which excel in temporal resolution but suffer from poor spatial precision and the requirement for controlled conditions, MRI strikes a balance by offering excellent spatial and temporal resolution. This dual advantage provides a more comprehensive understanding of brain structure and activity dynamics related to autism. Researchers can examine both structural abnormalities and functional connectivity patterns, providing a more holistic insight into the neurological underpinnings of autism spectrum disorders.

Contents

Acknowledgment	i
Abstract	ii
List of Abbreviations	vi
List of Figures	viii
List of Tables	ix
1 Introduction	1
1.1 Background	1
1.2 Problem Definition	2
1.3 Scope and Motivation	2
1.4 Objectives	3
1.5 Challenges	3
1.6 Assumptions	3
1.7 Societal / Industrial Relevance	3
1.8 Organization of the Report	4
2 Literature Survey	5
2.1 Improving the detection of Autism Spectrum Disorder by combining structural and functional MRI information[1]	5
2.1.1 Dataset	5
2.1.2 Preprocessing	6
2.1.3 Functional data classification pipeline	6
2.1.4 Structured data classification pipeline	9
2.1.5 Combined data classification pipeline	10

2.2	A Novel Machine Learning Based Framework for Detection of Autism Spectrum Disorder (ASD)[2]	11
2.2.1	Dataset	12
2.2.2	Preprocessing	12
2.2.3	Transfer Learning based approach	12
2.2.4	Conclusion	13
2.3	CNNG: Convolutional Neural Networks With Gated Recurrent Units[3] . .	14
2.3.1	3-D Convolutional Neural Network	14
2.3.2	Gated Circuit Units	15
2.3.3	Proposed Model	17
2.3.4	Data acquisition and Preprocessing	17
2.4	FNet: Mixing Tokens with Fourier Transforms[4]	19
2.4.1	Model	19
2.5	Brain Network Transformer[5]	22
2.5.1	Problem Defintion	22
2.5.2	Multi-Head Self Attention Module	22
2.5.3	ORTHONORMAL CLUSTERING READOUT(OCRead)	23
2.5.4	Conclusion	25
2.6	Summary and Gaps Identified	26
2.6.1	Summary	26
2.6.2	Gaps identified	30
3	Requirements	31
3.1	Hardware and Software Requirements	31
3.1.1	Software Requirements	31
3.1.2	Hardware Requirements	31
4	System Architecture	32
4.1	System Overview	32
4.1.1	Data Acquisition and Preprocessing	32
4.1.2	Feature Extraction	33
4.1.3	Classification	33
4.2	Architectural Design	33

4.3	Sequence Diagram	34
4.4	Module Division	34
4.4.1	Preprocessing Module	35
4.4.2	Feature Extraction Module	37
4.5	Work Schedule - Gantt Chart	39
5	Result	41
5.1	Overview	41
5.2	Models and their result analysis	41
5.2.1	BNT	41
5.2.2	BNT-Fourier	43
5.2.3	METAFormer	45
5.2.4	METAFormer-Fourier	47
5.2.5	Comparison of models	52
6	Conclusions & Future Scope	53
References		54
Appendix A: Presentation		56
Appendix B: Vision, Mission, Programme Outcomes and Course Outcomes		74
Appendix C: CO-PO-PSO Mapping		79

List of Abbreviations

ASD - Autism Spectrum Disorder

MEG - Magnetoencephalography

EEG - Electroencephalography

fMRI - Functional Magnetic Resonance Imaging

rs-fMRI - resting-state functional MRI sMRI - Structural Magnetic Resonance Imaging

MRI - Magnetic Resonance Imaging

ML - Machine Learning

MLP - Multilayer Perceptron

PCP - Preprocessed Connectomes Project

CPAC - Configurable Pipeline for the Analysis of Connectomes

CC200 - ameron Craddock's 200 ROI

ROI - Region Of Interest

BOLD - Blood Oxygen Level Dependent

PCC - Pearson Correlation Coefficient

List of Figures

2.1	Examples of connectivity matrices from two subjects in the Caltech subset of the ABIDE I dataset constructed using the AAL atlas. (a) is from an ASD patient while (b) is from a control.[1]	7
2.2	Graphical representation of (a) the simple and (b) stacked autoencoder structures and (c) the multilayer perceptron (MLP). The colored weights of the encoding part of stacked autoencoder (b) are used as initializing weights of the MLP (c).[1]	8
2.3	Graphical representation of the functional (top) and structural (bottom) data classification pipelines, together with combined strategies, including con- catenation branch and label fusion after separate pipelines.[1]	11
2.4	VGG16 Architecture[6]	13
2.5	The structure of single-frame convolutional neural network (CNN)[3]	15
2.6	The structure of gate recurrent unit (GRU).	16
2.7	The structure of the CNNG model.	18
2.8	FNet architecture with N encoder blocks.[4]	20
2.9	Overall framework of Brain Network Transformer[5]	23
4.1	Architecture Diagram of the proposed model	33
4.2	Architecture Diagram of the proposed model	34
4.3	Pearson Correlation Coefficient heatmap	36
4.4	Fourier Encoder	37
4.5	Gantt chart	39
5.1	BNT model Evaluation Metrics	42
5.2	Training curves of BNT	42
5.3	Training curves of BNT	43
5.4	BNT-F Evaluation metrics	44
5.5	Training curves of BNT-F	44

5.6	Training curves of BNT-F	45
5.7	METAFormer model Evaluation Metrics	45
5.8	Training curves of METAFormer model	46
5.9	Training curves of METAFormer model	46
5.10	Training curves of METAFormer model	46
5.11	Training curves of METAFormer model	46
5.12	METAFormer-F model Evaluation Metrics	47
5.13	METAFormer-F result graph	47
5.14	METAFormer-F result graph	47
5.15	METAFormer-F result graph	48
5.16	METAFormer-F result graph	48
5.17	METAFormer-F result graph	48
5.18	METAFormer-F result graph	48
5.19	METAFormer-F result graph	48
5.20	METAFormer-F result graph	48
5.21	METAFormer-F result graph	49
5.22	METAFormer-F result graph	49
5.23	METAFormer-F result graph	49
5.24	METAFormer-F result graph	49
5.25	Parameters in METAFormer model	50
5.26	Parameters in METAFormer-Fourier model	50
5.27	Graphical comparison of METAFormer & METAFormer-F	50
5.28	Graphical comparison of METAFormer & METAFormer-F	51
5.29	Graphical comparison of METAFormer & METAFormer-F	51
5.30	Graphical comparison of METAFormer & METAFormer-F	51

List of Tables

2.1	Advantages and Disadvantages of Methods	30
5.1	Comparison of Evaluation metrics	52

Chapter 1

Introduction

1.1 Background

Recent studies utilize cutting-edge technologies from multiple fields for autism screening. Neuroimaging modalities, including functional magnetic resonance imaging (fMRI), magnetoencephalography (MEG), electroencephalography (EEG) and structural magnetic resonance imaging (sMRI), are essential in studying about neurodevelopmental disorders like autism. Real-time insights into brain activity and connections are provided by EEG and MEG, which provide important information concerning neural patterns linked to autism. In the meanwhile, structural and functional information is provided by fMRI and sMRI, which reveal unique brain signatures in autistic people.

Beyond medical imaging, computer vision has become a potent instrument for diagnosing autism. The non-invasive evaluation of behavioural indicators linked to autism is made possible by machine learning techniques that frequently support facial expression analysis, gaze tracking, and head pose estimation. With the use of these technologies, diagnostic procedures become more sensitive and objective, leading to a more thorough comprehension of the wide range of characteristics associated with autism.

Furthermore, autism may be predicted and classified with exceptional accuracy by machine learning models that have been trained on a variety of datasets that include behavioural, neuroimaging, and genetic data. These models lead to a more sophisticated understanding of the variety within the autism spectrum as well as early detection. The amalgamation of these heterogeneous technologies mirrors the ever-changing terrain of autism identification, typified by a fusion of state-of-the-art approaches intended to augment accuracy, efficacy, and neutrality in diagnosis.

1.2 Problem Definition

Existing ASD diagnosis heavily relies on behavioral assessments, which can be subjective and prone to variability among clinicians. This subjectivity can lead to delays in diagnosis and may hinder the accuracy of the identification process. This project focuses on utilizing advanced neuroimaging techniques, specifically Functional MRI (fMRI), for the quantitative detection of Autism Spectrum Disorder (ASD). The aim is to enhance ASD detection accuracy by using fMRI, offering a comprehensive approach to understanding the neurobiological aspects of ASD.

1.3 Scope and Motivation

Our autism detection project addresses the vital need for early diagnosis to mitigate challenges faced by undiagnosed individuals. Timely identification can prevent distress, difficult behaviours, and social isolation, fostering a better understanding of oneself. Parents and professionals gain insights for optimal support, facilitating access to tailored autism services and educational adaptations. The project aims to prevent psychiatric illnesses by providing early intervention. Additionally, it lays the foundation for potential therapeutic interventions, prescribed by child and adolescent psychiatrists, contributing to an improved quality of life for individuals with autism and their families. The motivation for undertaking the project lies in the imperative to overcome the limitations of subjective ASD diagnosis methods, emphasizing the inherent variability introduced by individual clinician experiences. By shifting the focus towards objective measures, particularly through neuroimaging, the project seeks to motivate a paradigm shift in diagnostic practices. The transformative role of machine learning, exemplified in the base study achieving an accuracy of $85.06\% \pm 3.52\%$, serves as a motivational beacon for further exploration. The urgency to improve diagnostic accuracy and consistency fuels the motivation behind investigating novel approaches, aiming to provide a more reliable foundation for diagnosing ASD and, consequently, better support for individuals affected by the disorder.

1.4 Objectives

- Examine functional MRI data to enhance understanding and classification of Autism Spectrum Disorder (ASD).
- Identify and address the limitations of current subjective ASD diagnosis methods.
- Strive towards the overarching goal of improving diagnostic accuracy and consistency in ASD identification.

1.5 Challenges

The goal is to develop a broadly applicable ASD detection model that accommodates diverse backgrounds, age groups, and manifestations of ASD without escalating model complexity. The challenge lies in achieving accurate detection while maintaining model simplicity. Additionally, acquiring high-quality data with minimal noise and interpreting it amidst alignment variations poses another significant challenge in this endeavour.

1.6 Assumptions

The neuroimaging data, functional MRI (fMRI), is assumed to be of high quality, exhibiting minimal noise or artifacts. This assumption underlies the premise that the input data is reliable and accurately reflects the neurobiological features necessary for the successful development and training of the ASD detection model.

1.7 Societal / Industrial Relevance

The societal relevance of the project lies in its potential to significantly impact the lives of individuals affected by Autism Spectrum Disorder (ASD) and the broader community. By developing advanced neuroimaging techniques for quantitative ASD detection, the project addresses a critical need in the field of neurodevelopmental disorders. Early and accurate identification of ASD can pave the way for timely interventions, personalized therapies, and support services, ultimately improving the quality of life for individuals on the autism spectrum and their families.

Moreover, the project contributes to the scientific understanding of ASD, shedding light on the intricate neurobiological markers associated with the disorder. This knowledge not only enhances diagnostic capabilities but also informs broader research efforts aimed at unraveling the complexities of neurodevelopment and neurodiversity.

On a societal level, the project has the potential to reduce the societal and economic burden associated with delayed or inaccurate ASD diagnoses. Early intervention, made possible by improved detection methods, can lead to better educational outcomes, increased social inclusion, and greater independence for individuals with ASD. This, in turn, may positively influence societal attitudes and practices towards neurodiversity, fostering a more inclusive and empathetic society.

In summary, the societal relevance of the project lies in its capacity to improve individual outcomes, advance scientific knowledge, and contribute to creating a more inclusive and supportive society for individuals with Autism Spectrum Disorder.

1.8 Organization of the Report

The report embarks on a comprehensive exploration of Autism Spectrum Disorder (ASD) classification through an innovative system architecture. The introduction sets the stage by delineating the problem, specifying the scope, outlining motivations, and establishing objectives. Following this, the literature survey delves into five seminal papers, providing a foundation for the subsequent system architecture chapter. In Chapter 3, the proposed system architecture takes center stage, introducing a novel approach to binary classification of ASD. This approach uses functional MRI data, which uses advanced techniques such as Fourier Transformer Encoder. The architecture spans data acquisition, preprocessing, feature extraction, selection, and classification. Employing CPAC for functional MRI, the system extracts critical features from connectivity matrices and utilizes innovative transformers for feature selection. The chapter concludes with a meticulous module breakdown, architectural design overview, and a Gantt chart detailing the project's timeline, ultimately aiming to enhance ASD classification accuracy while comprehensively understanding the associated neuroimaging data patterns.

Chapter 2

Literature Survey

2.1 Improving the detection of Autism Spectrum Disorder by combining structural and functional MRI information[1]

Due to its intricate nature, ASD is a complex neurological disorder that poses unique hurdles for established diagnostic approaches. This study suggests a novel method for differentiating between people with ASD and control patients using a combination of functional and structural MRI data. The main aspects are the analysis of volumetric correspondences of gray matter volumes across cortical parcels and patterns of functional connectivity among brain areas. The classification network uses a combination of supervised multilayer perceptrons and unsupervised stacked autoencoders to do this. With an ensemble of classifiers, the study achieves a notable classification accuracy of $85.06 \pm 3.52\%$ utilizing a large dataset from the Autism Brain Imaging Data Exchange I (ABIDE I). Notably, the combination of data from structural and functional sources outperforms the effectiveness of separate pipelines. This work highlights the potential of multimodal MRI data to improve classification accuracy and advances the area of neuroimaging in the diagnosis of ASD.

2.1.1 Dataset

The Autism Brain Imaging Data Exchange I (ABIDE I) dataset is used in the study; it is the result of a collaborative effort between 17 worldwide sites. There are 539 people with autism spectrum disorder (ASD) and 573 normal control participants out of 1112 cases. The dataset includes structural MRI pictures, clinical information, and resting-state functional MRI (rs-fMRI) series. The patients in the dataset range in age from 3 to 1 (median age of 14.7 years).

Direct acquisition of structural and phenotypic data was made from ABIDE I. On the other hand, the preprocessed functional dataset which included rs-fMRI data was obtained through the Configurable Pipeline for the Analysis of Connectomes (CPAC) of the Preprocessed Connectomes Project (PCP).

2.1.2 Preprocessing

The rs-fMRI information, was acquired from the Preprocessed Connectomes Project (PCP) utilizing the Configurable Pipeline for the Analysis of Connectomes (CPAC). Important processes including intensity normalization, slice time correction, motion correction, bandpass filtering (0.01 Hz - 0.1 Hz), and spatial registration to the MNI152 template space are all included in this pipeline. The functional data were subjected to two popular atlases: Cameron Craddock's 200 ROI (CC200) parcellation atlas and the Automated Anatomical Labeling (AAL) atlas. Subjects who met a threshold of 0.2 were eliminated from the evaluation of motion artifacts, which involved calculating mean framewise displacement. The final dataset consisted of 884 rs-fMRI participants, 408 of whom had an autism spectrum disorder (ASD) diagnosis, and 476 of whom were control cases.

Structural MRI information underwent cortical parcellation using the Destrieux atlas, facilitated by the Freesurfer software. Intensity normalization, skull stripping, volume registration to a shared space, segmentation, and cortical parcellation were all steps in this extensive processing pipeline. For every parcel, statistical parameters like curvature, cortical thickness, and gray matter volume were calculated. For 1014 instances (475 ASD patients and 539 control participants), the processing was successful despite the presence of artifacts in several structural MRI volumes.

2.1.3 Functional data classification pipeline

The functional data classification pipeline initiates with preprocessing using the CPAC pipeline, followed by the extraction of mean BOLD signals from diverse brain regions. Then, using either the CC200 atlas with 200 areas or the AAL atlas with 116 regions, a connectivity matrix is created for each case, showing the correlation of BOLD series between designated regions. By calculating the Pearson correlation coefficients for every pair of regions, a symmetrical matrix is created during the building of the connection matrix. Next, a 1-dimensional vector is created by flattening the upper triangle of the

matrix, omitting the main diagonal.

The Pearson correlation coefficient (r) formula is given by[1]:

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

The Pearson Correlation Coefficient (PCC) is computed to assess the linear relationship between two sets of real numbers, that is, the BOLD (Blood-Oxygen-Level-Dependent) signals from two distinct brain regions denoted as 'x' and 'y'. The numerator calculates the product of the differences between each BOLD signal and the mean BOLD signal of 'x' and 'y', while the denominator computes the product of the standard deviations of the BOLD signals of 'x' and 'y'. The ' \bar{x} ' and ' \bar{y} ' are the averages of the BOLD series for brain regions 'x' and 'y', respectively. The resulting PCC provides a numerical measure between -1 and 1, indicating the strength and direction of the linear correlation between the BOLD signals of brain regions 'x' and 'y'.

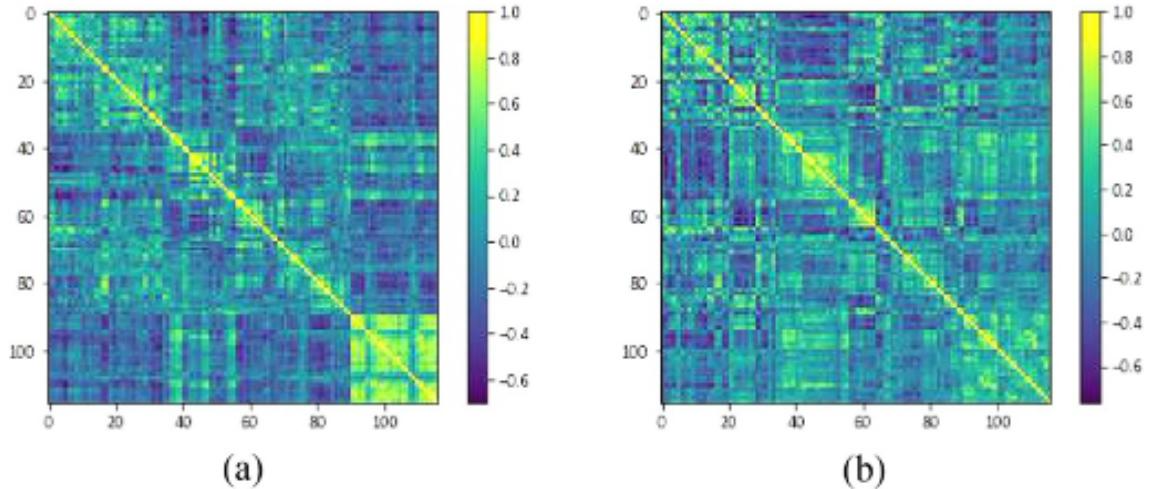


Figure 2.1: Examples of connectivity matrices from two subjects in the Caltech subset of the ABIDE I dataset constructed using the AAL atlas. (a) is from an ASD patient while (b) is from a control.[1]

To address high dimensionality, the Fisher score computation is applied, ranking fea-

tures based on distinctiveness. This technique aids in reducing overfitting and enhances the model's generalizability.

There are two stages to the categorization process. Unsupervised training of stacked autoencoders takes place in the first phase. Autoencoders are low-dimensional representations of input vectors learned by simple networks that are intended to reconstruct inputs with accuracy. The learning capabilities are improved by the stacked design, which consists of two or more autoencoders.

A multilayer perceptron (MLP) with two hidden layers and a binary output layer is used in the second stage of supervised training. The number of nodes in the stacked autoencoder's hidden layers matches the number of nodes in the encoding layers. The autoencoder's weight initialization guarantees that the MLP can efficiently learn hidden features and categorize topics.

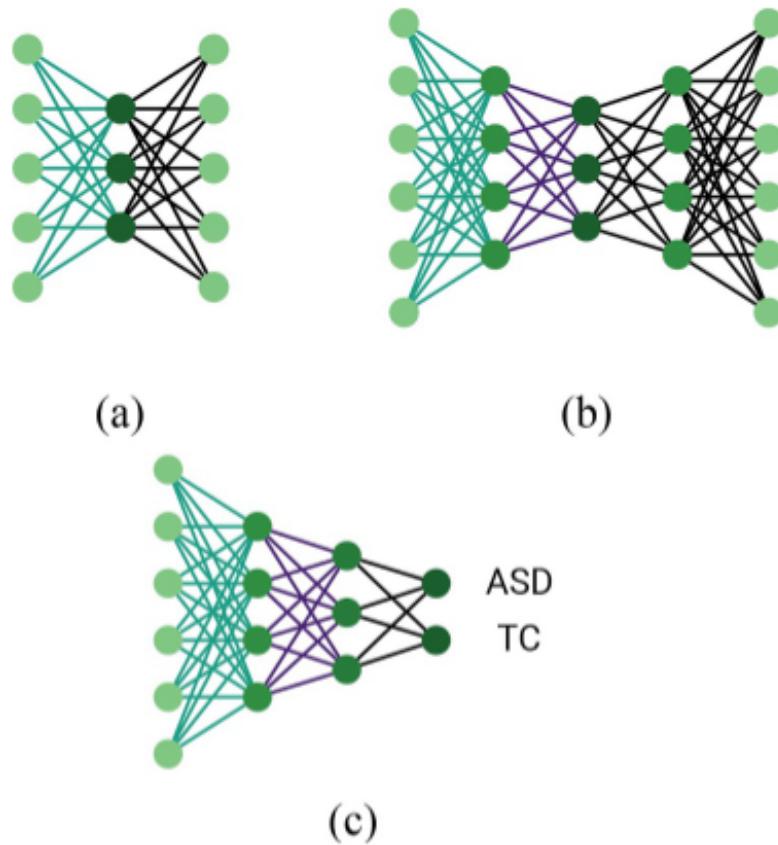


Figure 2.2: Graphical representation of (a) the simple and (b) stacked autoencoder structures and (c) the multilayer perceptron (MLP). The colored weights of the encoding part of stacked autoencoder (b) are used as initializing weights of the MLP (c).[1]

The pipeline includes batch normalization for better convergence, regularization terms,

and dropout in hidden layers to prevent overfitting. An ensemble of classifiers, consisting of 5 classifiers with different nodes in the hidden layers, is used to provide resilient results. Softmax activation probabilities are averaged to reach a final conclusion.

In conclusion, the functional pipeline consists of creating connection matrices, lowering dimensionality, training stacked autoencoders in an unsupervised manner, training an MLP in a supervised manner, and utilizing an ensemble of classifiers to increase the resilience and accuracy of classifications. Connectivity matrices are built, dimensionality is reduced, stacked autoencoders are trained unsupervised, an MLP is trained supervised, and an ensemble of classifiers is used to improve classification robustness and accuracy.

2.1.4 Structured data classification pipeline

Instead of calculating the Pearson correlation coefficient, the structural pipeline looks at the relationships between the gray matter volumes in pairs of cortical parcels that are described by the Destrieux atlas (148 regions total, 74 in each hemisphere). Although the Freesurfer pipeline elements are easily applied, the focus is on creating a matrix that represents volumetric correspondences between regions for every subject. Each element (i,j) of the matrix signifies the volumetric correspondence between two parcels (i and j), defined by[1]:

$$\frac{1}{[\mathbf{gm}(i) - \mathbf{gm}(j)]^2 + 1}$$

Here, $\text{gm}(i)$ and $\text{gm}(j)$ represent the gray matter volumes of ROIs i and j , respectively.

For every subject, a connection matrix is created by taking the lower triangular portion and flattening it into a feature vector. 10,878 characteristics can be found in the flattened vector that was taken out of the connection matrix. To make the feature vectors less dimensional, the Fisher score is used. For the classification job, the newly acquired vectors are subsequently input into an ensemble of five stacked autoencoders and an MLP.

2.1.5 Combined data classification pipeline

A key contribution of this work involves integrating the previously described functional and structural pipelines into a unified framework, aiming to enhance classification results by leveraging diverse types of information. The two pipelines, learning independent features, are merged to potentially mitigate errors.

Two integration strategies are applied to fuse functional and structural information in the classification process. The first strategy involves concatenating feature vectors after the dimensionality reduction stage. This ensures that the merger occurs before autoencoder training, allowing the autoencoder to discern informative patterns. The choice to merge after dimensionality reduction addresses potential biases introduced by variations in the sizes of original feature vectors for structural and functional data based on the atlas choice. Subsequently, classification is executed using either the obtained vector (of length 6000) or a further reduction in dimensionality using the Fisher score, resulting in an additional vector for input to the network. The classification stage maintains consistency, involving the ensemble of unsupervised stacked autoencoder training followed by the supervised training of a multilayer perceptron (MLP).

The second strategy employs separate classification pipelines for functional and structural data, resembling the previously described independent pipelines. Each pipeline encompasses ensembles of autoencoders and multilayer perceptrons trained independently using corresponding functional and structural training and validation data, resulting in 10 softmax activations. The final output is determined either by averaging the 10 softmax activation probabilities or employing a majority voting strategy.

In conclusion, this work provides a comprehensive examination of the diagnosis of Autism Spectrum Disorder (ASD) through the use of cutting-edge neuroimaging techniques. The paper presents two different pipelines that use data from magnetic resonance imaging (MRI) for classification: a functional pipeline and a structural pipeline. Whereas the structural pipeline focuses on volumetric correspondences between gray matter volumes in cortical parcels, the functional pipeline takes functional connection patterns into account. Crucially, each pipeline demonstrates strong independent classification abilities.

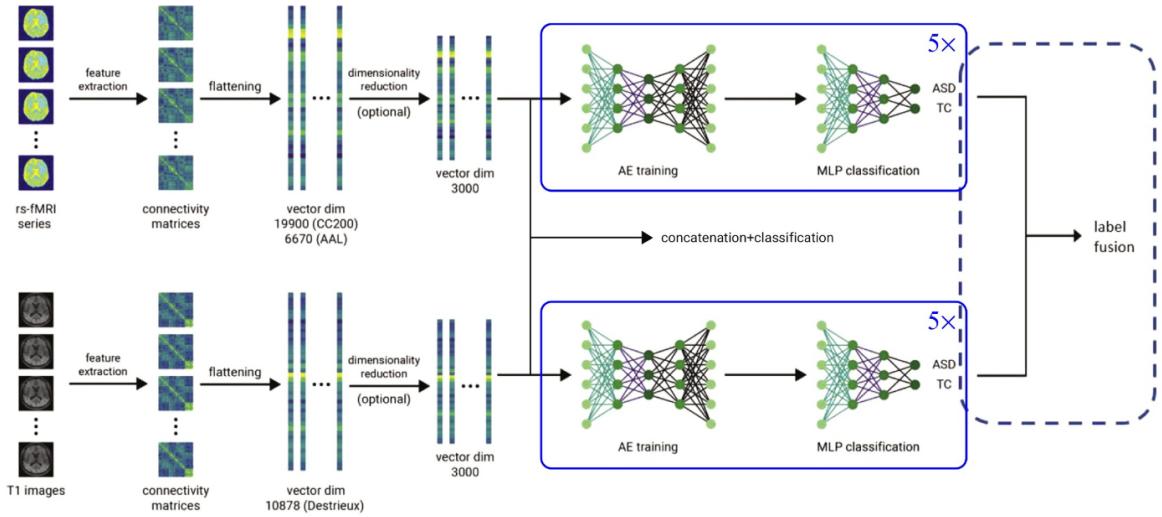


Figure 2.3: Graphical representation of the functional (top) and structural (bottom) data classification pipelines, together with combined strategies, including concatenation+classification branch and label fusion after separate pipelines.[1]

A notable contribution of this research is the integration of functional and structural information to establish a unified classification pipeline. This integration aims to improve classification results by considering diverse types of information. Two strategies for merging data are proposed: concatenating feature vectors after dimensionality reduction and employing separate pipelines with a subsequent decision-making mechanism. The results demonstrate that the combined classification pipeline surpasses individual pipelines, underscoring the effectiveness of integrating different modalities for ASD diagnosis.

2.2 A Novel Machine Learning Based Framework for Detection of Autism Spectrum Disorder (ASD)[2]

This study delves into the challenges posed by Autism Spectrum Disorder (ASD), characterized by limitations in social behavior. Despite extensive research in cognitive sciences, the biomarkers of ASD remain elusive. Drawing inspiration from neuroscience insights highlighting the relevance of corpus callosum and intracranial brain volume, the study has proposed a machine learning-based framework for automatic ASD detection using an MRI scan. This approach not only attains commendable recognition accuracy but also streamlines model training complexity by prioritizing the most discriminative fea-

tures. Additionally, we explore the potential of deep learning in neuroimaging analysis, employing the transfer learning technique with the pre-trained VGG16 model for ASD classification.

2.2.1 Dataset

This investigation utilizes structural MRI (s-MRI) scans sourced from the Autism Brain Imaging Data Exchange (ABIDE-I) dataset. ABIDE is a collaborative consortium sharing imaging data of individuals with Autism Spectrum Disorder (ASD) and control participants. It comprises data from 17 international sites, comprising a total of 1,112 subjects, including 539 cases of autism and 573 healthy control participants.

2.2.2 Preprocessing

In the preprocessing phase, various software tools were employed for accurate measurements of key brain parameters, including corpus callosum area and intracranial volume. The yuki software facilitated the segmentation of the corpus callosum area, which was further divided into subregions using the Witelson scheme. Visual inspection and manual correction, if needed, were performed using the ITK-SNAP software. Intracranial brain volume was measured using the brainwash software, which utilized the Automatic Registration Toolbox for extracting accurate intracranial regions. The voxel-voting scheme classified each voxel in the MRI scans, and visual inspections ensured precise segmentation. In cases of segmentation errors, corrective measures were implemented, including rerunning the brainwash method, manually identifying anterior and posterior commissure coordinates, and employing the "region-based snakes" feature in ITK-SNAP for minor corrections in intracranial volume segmentation. These steps collectively ensured reliable and accurate preprocessing of neuroimaging data.

2.2.3 Transfer Learning based approach

In the transfer learning approach, we harness the knowledge encoded in the pre-trained VGG16 neural network, originally trained on diverse datasets like ImageNet. Instead of training the model anew for our Autism Spectrum Disorder (ASD) detection task, we leverage the already acquired capability of VGG16 architecture shown in the figure given

below, to extract intricate features from images.

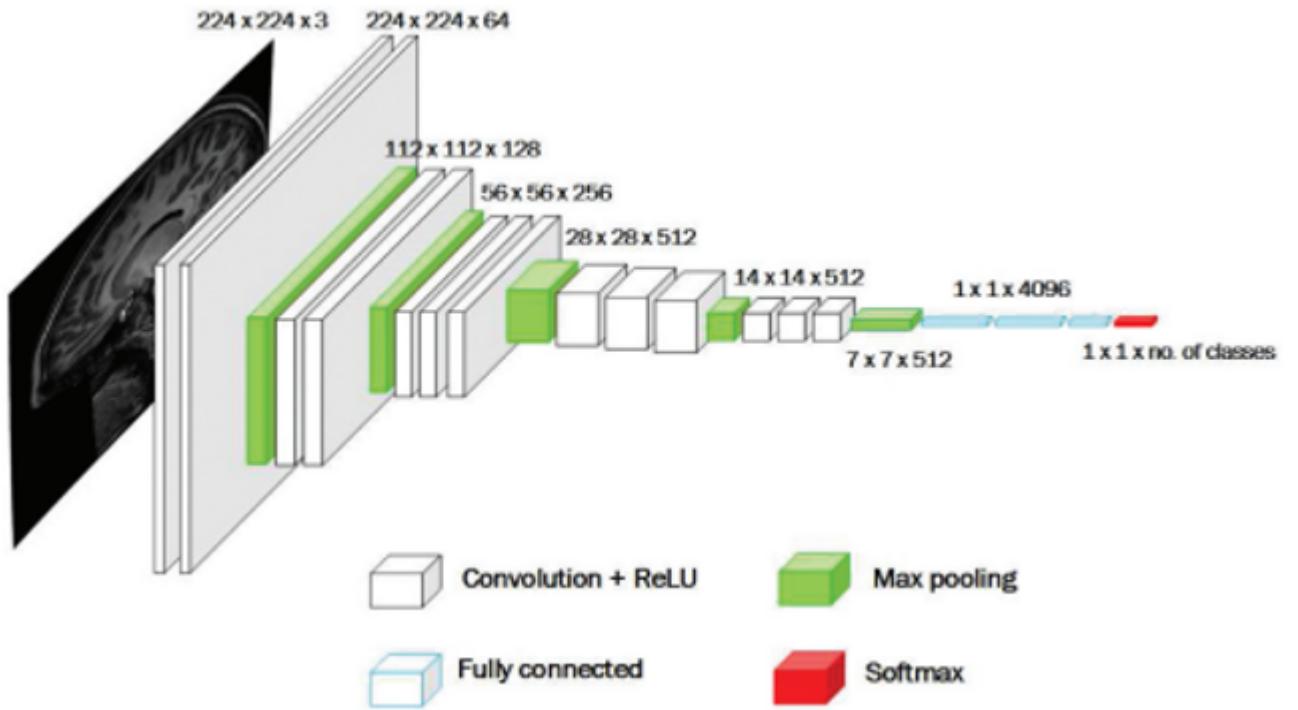


Figure 2.4: VGG16 Architecture[6]

This architecture is designed to process images of dimensions 224×224 , employing a 3×3 receptive field, a convolution stride of 1 pixel, and a padding of 1 (specifically for the 3×3 receptive field). The rectified linear unit (ReLU) serves as the activation function (Nair and Hinton 2010). Classification is executed through a softmax classification layer with x units, where x represents the number of classes to be recognized. The architecture incorporates Convolution layers, utilizing filters convolved with the input image to generate activation or feature maps. Additionally, Feature Pooling layers are strategically employed to reduce the size of the image representation, optimizing computational efficiency, and mitigating overfitting concerns. This transfer learning methodology using VGG16 enhances our model's capability to discern intricate patterns in neuroimaging data for effective Autism Spectrum Disorder (ASD) detection.

2.2.4 Conclusion

In conclusion, the application of machine learning algorithms to anatomical brain scans offers a promising avenue for the automatic detection of Autism Spectrum Disorder (ASD).

Notably, features derived from the corpus callosum and intracranial brain regions prove to be valuable discriminative markers for distinguishing individuals with ASD from the control group. The incorporation of feature selection and weighting methods not only contributes to the computational efficiency of the framework but also enhances overall classification accuracy. Additionally, we present results utilizing Convolutional Neural Networks (CNN) through a transfer learning approach, shedding light on the advantages and challenges of employing deep learning/CNN for neuroimaging data analysis. To further improve recognition outcomes, future efforts could explore the integration of a multimodal system, incorporating data from additional modalities such as EEG, speech, or kinesthetic information for enhanced ASD recognition

2.3 CNNG: Convolutional Neural Networks With Gated Recurrent Units[3]

This paper uses a classification model that is deep learning based according to the CNNG model that captures the spatio-temporal features of fMRI via a fusion scheme of 3D convolutional neural network and gated recurrent unit. The spatial aspects of brain pictures are solved by multiple 3D CNN networks sharing weights at each time point. The GRU unit then extracts the temporal information. The structure of the CNNG model is presented in Figure 2.7 [3].

2.3.1 3-D Convolutional Neural Network

Convolutional neural network is a type of deep neural network whose structure consists of local connectivity features and weights sharing. 3D convolution expands the concept of 2D convolution to area permit the convoluting the 3D kernel at 3D data during the feature extraction process.

Let the element $k_{ij}^{x_0y_0z_0}$ be value at the position (x_0, y_0, z_0) representing j -th feature map of the i -th layer, then the three-dimensional convolution can be given as:

$$k_{ij}^{x_0y_0z_0} = \partial \left(b_{ij} + \sum_c \sum_{p=0}^{P_{i-1}} \sum_{q=0}^{Q_{i-1}} \sum_{r=0}^{R_{i-1}} w_{ijc}^{pqr} (x_0 + p)(y_0 + q)(z_0 + r) \right), \quad (2.1)$$

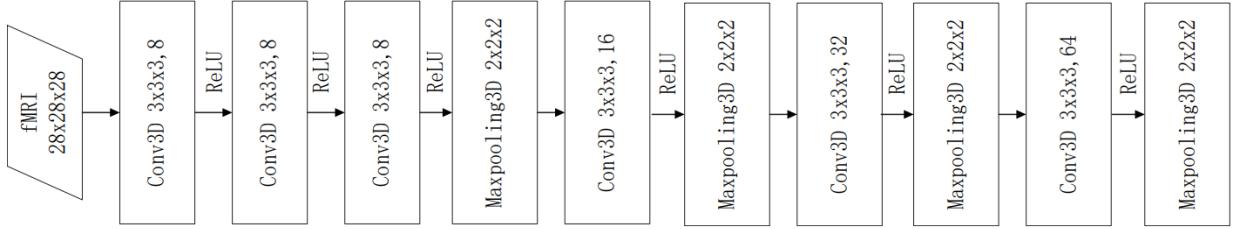


Figure 2.5: The structure of single-frame convolutional neural network (CNN)[3]

where ∂ is the activation function. P_i , Q_i , and R_i are the dimensional magnitudes of the three directions, respectively. $w^{pqr}ijc$ is the value of the convolution kernel, which connects the c -th feature map of $(i - 1)$ -th layer with the j -th feature map of the i -th layer at position (p, q, r) . bij is the bias.

Three-dimensional images are being used more and more in medical image analysis, and the 3D convolution can have taken out their spatial properties. Since fMRI uses data from three-dimensional space in the brain represented as pictures, the 3D CNN adaption is well suited for fMRI's three-dimensional spatial feature extraction.

This paper takes a 3D CNN with input size $28 \times 28 \times 28$ and consists of three convolution layers. Each of the 8 convolution kernels in three convolutional layers is of the size $3 \times 3 \times 3$, and all the convolution kernels are connected with ReLU layers. The fourth layer is the max pooling layer with a stride of 2 and kernel size of $2 \times 2 \times 2$. The use of maxpooling layer decreases that size, prevents overfitting as well as reduce running time. Also added after the pooling layer are three sets of repeated convolutional and pooling layer to extract even more advanced features. Additionally, every convolutional kernel has a dimension of $3 \times 3 \times 3$. Each convolutional kernel contains 16, 32, and 64 filters. Following each convolutional layer, there is a pooling kernel with dimensions of $2 \times 2 \times 2$ [3].

2.3.2 Gated Circuit Units

After extracting the fMRI spatial features by using 3D CNN, the spatial feautes aranged along time dimension is processed using GRU after flattening. nGRU is an modified long short-term memory (LSTM) [7], with two inputs and two outputs [8]; [9]. The

fewer parameters in GRU can aid in speeding-up the training and improving its network performance.

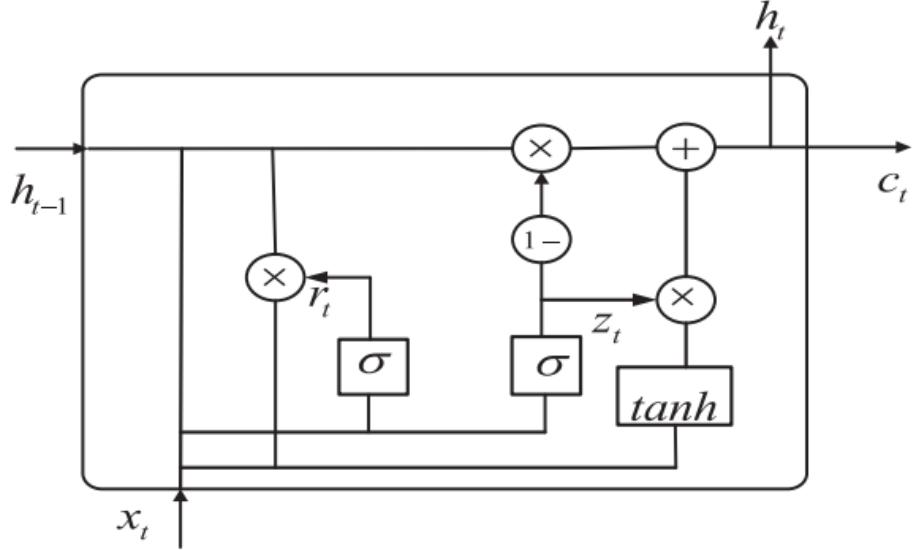


Figure 2.6: The structure of gate recurrent unit (GRU).

Let x_t be the input of the GRU, and h_t is the output of the GRU. The equations of GRU can be given as:

$$z_t = \sigma(W_z[h_{t-1}, x_t]) \quad (2.2)$$

$$r_t = \sigma(W_r[h_{t-1}, x_t]) \quad (2.3)$$

$$\tilde{h}_t = \tanh(W[r_t \cdot h_{t-1} - 1, x_t]) \quad (2.4)$$

$$h_t = (1 - z_t) \cdot h_{t-1} + z_t \cdot \tilde{h}_t \quad (2.5)$$

where z_t denotes the update gate, r_t denotes the reset gate, \tilde{h}_t denotes the hidden unit, h_t is the current moment output. W_z , W_r , and W denote weights, and σ is the activation function.

Using a hidden state, the GRU unit combines input and forgetting gates into a single update gate, controlling the quantity of data to be deleted from the hidden layer of the previous moment as well as the addition of memory data from the present moment. Additionally, it introduces a reset gate, determining whether the computation of \tilde{h}_t depends on the state h_{t-1} from the prior moment. When $r_t = 0$, \tilde{h}_t is solely linked to the

current input x_t , unrelated to the historical state. Conversely, when $r_t = 1$, \tilde{h}_t is influenced by both x_t and $ht - 1$. Compared to LSTM, GRU exhibits fewer internal "gating" mechanisms and parameters. Despite achieving comparable performance, GRU is easier to train, enhancing training efficiency. Therefore, we leverage GRU for time-dimensional feature extraction to improve ASD classification results.

2.3.3 Proposed Model

In the proposed model, the process is initiated with two consecutive three-dimensional convolutions which have a size of $3 \times 3 \times 3$ for extracting the low-level features. Later with a set of pooling and convolution operations, we identify the high-level features. In order to make the model more parametric while minimizing the parameter count, the recurrent two-layer convolution is replaced with a single $3 \times 3 \times 3$ kernel sized convolutional layer. These spatial features obtained are reshaped into 1D feature vectors or flattened out and fed to the GRU consisting of 32 neurons. Ultimately, the model yields predicted values through a fully connected layer with the Sigmoid activation function.

2.3.4 Data acquisition and Preprocessing

The rs-fMRI data used in this paper is from the international publicly available ABIDE dataset. The dataset consolidates 1,112 participants ($n=539$ ASD subjects and $n=573$ typically developing TC) from 17 different sites around the world. The database includes for each study participant rs-fMRI, structural MRI, and a detailed set of phenotypic information. In this work partial missing information was removed. After removing samples with high motion peaks, poor brain coverage, ghosting, and other scanner aberrations, a final dataset comprising 871 subjects—403 ASD patients and 468 TCs, was obtained.

Because of the large amount of noise generated during fMRI acquisition, preprocessing is necessary before use. The configurable pipeline for the analysis of connectomes (CPAC) is the preprocessing technique utilized in this paper, which include: time slice correction, head movement correction, alignment, numerical normalization, Interference signal regression, Filtering, Spatial normalization. The first ten fMRI time points were eliminated in the temporal dimension, and the eleventh frame contained 32 consecutive brain picture frames. Every image undergoes a downsampling of its spatial dimension from (61, 73,

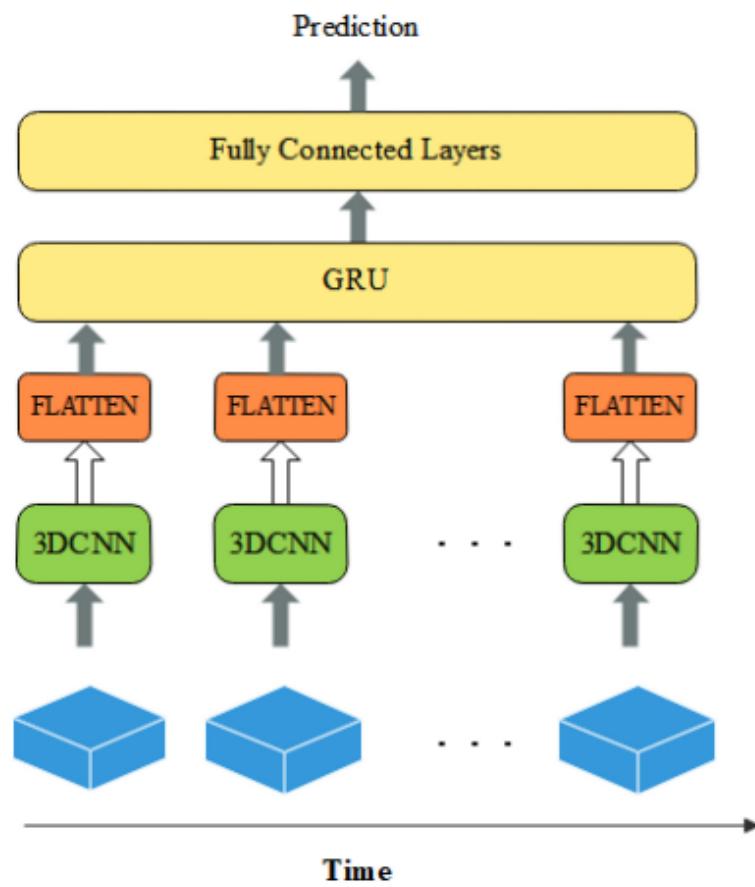


Figure 2.7: The structure of the CNNG model.

61) to (28,28, 28). Following the temporal and spatial dimensions' processing, the fMRI data's size is (28, 28, 28, 32). By doing this, the spatiotemporal properties of fMRI are preserved and the same model input is guaranteed.

2.4 FNet: Mixing Tokens with Fourier Transforms[4]

In the evolving landscape of natural language processing, this study explores the optimization of Transformer encoder architectures. By substituting self-attention sublayers with simplified linear transformations, including an unparameterized Fourier Transform, the novel FNet architecture achieves noteworthy speed enhancements. Despite being 80% faster on GPUs and 70% faster on TPUs at standard input lengths, FNet maintains an impressive accuracy of 92-97% compared to BERT on the GLUE benchmark. Notably, at extended input lengths, FNet outpaces other models on the Long Range Arena benchmark, showcasing its efficiency and superiority across various sequence lengths. This research underscores FNet's effectiveness, particularly in terms of a lighter memory footprint and superior performance at smaller model sizes, making it a compelling alternative within a fixed speed and accuracy budget compared to conventional Transformer architectures.

2.4.1 Model

Discrete Fourier Transform

The Discrete Fourier Transform (DFT) is a mathematical technique used to analyze and understand the frequency components of a sequence of values. Given a sequence $\{x_n\}$ with $n \in [0, N - 1]$, the Discrete Fourier Transform (DFT) is defined by the formula:

$$X_k = \sum_{n=0}^{N-1} x_n e^{-\frac{2\pi i}{N} nk}, \quad 0 \leq k \leq N - 1$$

For each k , the Discrete Fourier Transform (DFT) generates a new representation X_k as a sum of all of the original input tokens x_n , with so-called "twiddle factors."

One common method to compute the DFT involves the Fast Fourier Transform (FFT), which efficiently breaks down the DFT calculation, reducing its complexity to $O(N \log N)$.

An alternative approach is to utilize the DFT matrix, denoted as W , and perform matrix multiplication.

$$W_{nk} = \frac{e^{-\frac{2\pi i}{N} nk}}{\sqrt{N}}, \quad 0 \leq n, k \leq N - 1$$

The DFT matrix involves complex numbers and roots of unity, and the multiplication operation has a computational complexity of $O(N^2)$. Interestingly, despite its higher asymptotic complexity, this alternative method has proven to be faster for relatively shorter sequences when executed on Tensor Processing Units (TPUs). The choice between FFT and matrix multiplication depends on the specific characteristics and length of the sequence being analyzed, with the latter offering advantages for certain scenarios, particularly on TPUs.

FNet architecture

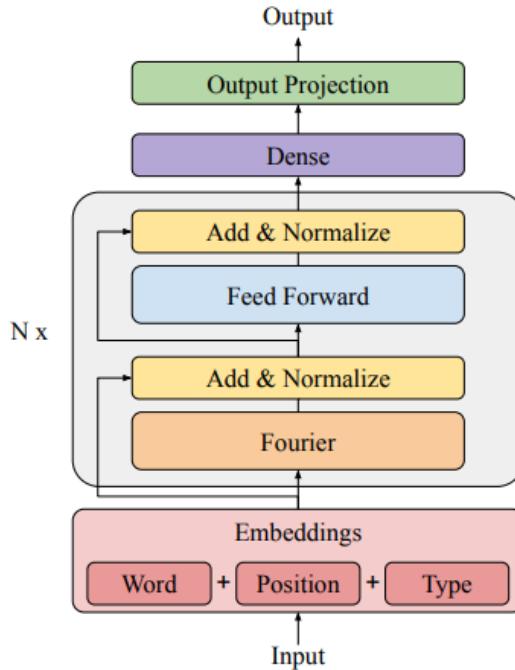


Figure 2.8: FNet architecture with N encoder blocks.[4]

FNet introduces an attention-free Transformer architecture where each layer comprises a Fourier mixing sublayer followed by a feed-forward sublayer. Departing from the conventional Transformer design, the self-attention sublayer is replaced with a Fourier sublayer,

utilizing a 2D Discrete Fourier Transform (DFT) on the embedding input. This transform involves a 1D DFT along both the sequence dimension (F_{seq}) and the hidden dimension (F_h).

Crucially, FNet retains only the real part of the total transformation, simplifying the model by eliminating the need to modify feed-forward sublayers or output layers to handle complex numbers. The real part is extracted after applying both F_{seq} and F_h . The architecture was fine-tuned through experimentation with alternative transforms, revealing the Hartley Transform as a strong contender.

The Fourier Transform is interpreted in FNet as an effective mechanism for token mixing, ensuring that feed-forward sublayers have comprehensive access to information from all tokens. The alternating encoder blocks in FNet are viewed as applying alternating Fourier and inverse Fourier Transforms, thereby transforming the input between the "time" and frequency domains.

Despite the complexity introduced by the Fourier Transform, FNet’s embedding layers remain consistent with previous work, combining word embeddings, absolute position embeddings, and type embeddings of sentences. The inclusion of position embeddings, while not strictly necessary due to the positional information encoded by the Fourier Transform, facilitates a cleaner comparison with BERT. Overall, FNet presents an attention-free architecture that leverages Fourier mixing sublayers, providing a unique approach to token interactions and model transformations.

In conclusion, FNet introduces a novel and attention-free Transformer architecture by replacing the traditional self-attention mechanism with Fourier mixing sublayers. This innovative choice offers a compelling alternative, using the Fourier Transform to capture intricate dependencies within the input sequence. The decision to extract and retain only the real part of the Fourier Transform streamlines the model without compromising its effectiveness. FNet’s attention-free approach not only simplifies the computational complexity but also demonstrates competitive performance in various natural language processing tasks. This departure from attention mechanisms introduces a fresh perspective, presenting a promising avenue for future research and exploration in the realm of deep learning for language understanding.

2.5 Brain Network Transformer[5]

This research introduces BRAINNETTF, a transformative model designed for advanced brain network analysis. Drawing from insights in effective Graph Neural Network designs, we utilize connection profiles as initial node features, avoiding computationally intensive processes. Inspired by recent advancements, our model adopts fully pairwise attention weights within Transformer-based architectures. Addressing the challenge of generating precise graph-level embeddings, we introduce ORTHONORMAL CLUSTERING READ-OUT, a novel global pooling operator. To overcome limited open-access datasets, we propose a stratified sampling method and a standardized evaluation pipeline on the ABIDE dataset. Our experiments showcase significant improvements in brain network analysis facilitated by BRAINNETTF.

2.5.1 Problem Defintion

The problem addressed in this brain network analysis is the prediction of specific properties, such as biological sex or the presence of a disease, based on the given brain network $X \in \mathbb{R}^{V \times V}$, where V represents the number of nodes (Regions of Interest-ROIs). The proposed BRAIN NETWORK TRANSFORMER comprises two main components: an L-layer attention module (MHSA) and a graph pooling operator (OCREAD). The attention module (MHSA) facilitates the learning of attention-enhanced node features Z^L through a non-linear mapping from X to $Z^L \in \mathbb{R}^{V \times V}$. Subsequently, the graph pooling operator (OCREAD) compresses the enhanced node embeddings Z^L into graph-level embeddings $Z_G \in \mathbb{R}^{K \times V}$, with K as a hyperparameter denoting the number of clusters. The graph-level embeddings Z_G are then flattened and processed through a multi-layer perceptron for making predictions at the graph level. The entire training process is supervised using the cross-entropy loss, with the goal of accurately predicting the specified properties based on the brain network data.

2.5.2 Multi-Head Self Attention Module

The Multi-Head Self-Attention (MHSA) module is a pivotal element within the BRAIN NETWORK TRANSFORMER, specifically crafted to amplify the expressiveness of node features in brain network data. Operating across various heads denoted as m , this module

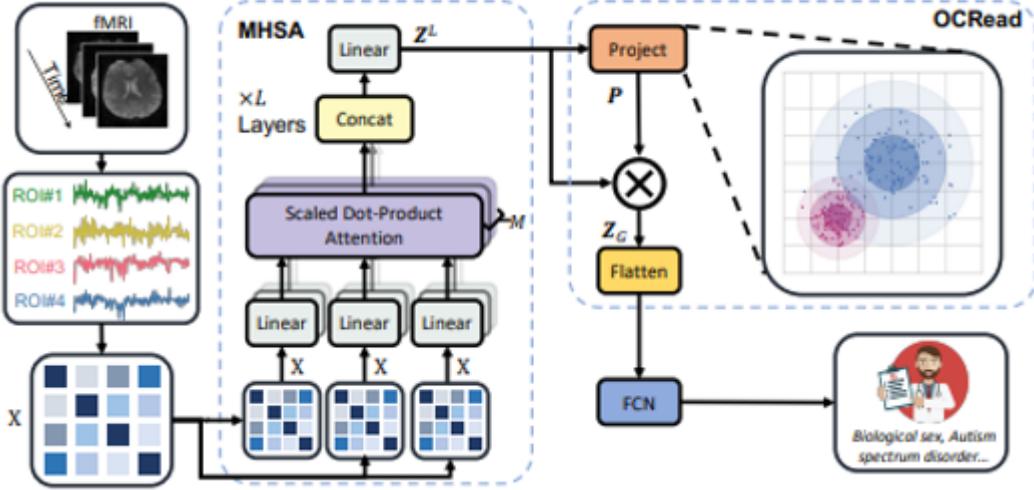


Figure 2.9: Overall framework of Brain Network Transformer[5]

orchestrates the computation of attention scores $h_{l,m}$ using Softmax.[5] These scores are derived from the dot product of query ($W_Q^{l,m}$) and key $W_K^{l,m}Z^{l-1}$ projections applied to the input node features (Z^{l-1}). The resultant attention scores guide the weighted amalgamation of value projections yielding attention-enhanced node features. The outputs from all heads are concatenated and subject to a linear transformation (W_O^l), ultimately producing the final node features (Z^l). The equation for the above module is given as,

$$Z^l = \left(\sum_{m=1}^M h_{l,m} \right) W_O^l, h_{l,m} = \text{Softmax} \left(\frac{W_Q^{l,m} Z^{l-1} \left(W_K^{l,m} Z^{l-1} \right)^T}{\sqrt{d_K^{l,m}}} \right) W_V^{l,m} Z^{l-1}$$

The MHSA module's intrinsic capability to consider diverse attention patterns across multiple heads empowers the BRAIN NETWORK TRANSFORMER[5] to adaptively focus on distinct facets of the input data. This adaptability contributes significantly to the overall expressive prowess of the model.

2.5.3 ORTHONORMAL CLUSTERING READOUT(OCRead)

The readout function is a pivotal component in brain network analysis, responsible for mapping learned node-level embeddings to a comprehensive graph-level representation. These traditional functions often overlook the inherent properties of brain networks, where nodes within the same functional modules exhibit similar behaviours and clustered representations. In response to this limitation, a novel readout function[5] is introduced which

is designed to leverage modular-level similarities among Regions of Interest (ROIs) in brain networks. Unlike conventional readouts, our function soft assigns nodes to clusters, calculating the probabilities of nodes belonging to specific clusters based on learned embeddings. This soft assignment is facilitated by a Softmax projection operator, denoted by the equation

$$P_{ik} = \frac{e^{\langle Z_i^L, E_k \rangle}}{\sum_{k'}^K e^{\langle Z_i^L, E_{k'} \rangle}},$$

where,

- $\langle \cdot, \cdot \rangle$ denotes the inner produc

- Z_i^L represents the node embedding of node i obtained from the last layer of the Multi-Head Self-Attention (MHSA) module.

- E_k . represents the cluster center embedding of cluster k , where K is the total number of clusters.

The computed softmax probability determines how likely node i belongs to cluster k based on its embedding and the cluster centres. After the soft assignment (P_{ik}) is computed using the softmax projection operator, the original learned node representations (Z_i^L) are aggregated under the guidance of the soft cluster information. The graph-level embedding (Z_G) is obtained by multiplying the transpose of P with Z_L , where Z_L is the learned set of node embeddings from the last layer of the MHSA module:

$$Z_G = P^T Z_L$$

Orthonormal initialization

The initialization of K cluster centers (E) is crucial for the unsupervised clustering process. Orthonormal initialization leverages the observation that orthonormal embeddings can improve the clustering of nodes in brain networks with respect to the functional modules underlying brain regions.

To initialize a group of orthonormal bases as cluster centers, the Xavier uniform initialization[10] is first applied to initialize K random centers, each containing V dimensions ($C \in \mathbb{R}^{K \times V}$). Xavier uniform initialization [10]is a method commonly used in neural network weight initialization. It helps maintain consistent variance in activations across layers during training, mitigating issues such as vanishing or exploding gradients.

Then, the Gram-Schmidt process is employed to obtain the orthonormal bases (E) where,

$$u_k = C_{k\cdot} - \sum_{j=1}^{k-1} \frac{\langle u_j, C_{k\cdot} \rangle}{\langle u_j, u_j \rangle} u_j, \quad E_{k\cdot} = \frac{u_k}{\|u_k\|}$$

The Gram-Schmidt process is a mathematical procedure used to transform a set of vectors into an orthonormal basis. This process enhances the quality of cluster centre initialization (E) by making vectors orthogonal and normalized. This orthonormal initialization enhances the quality of the clusters, contributing to the effectiveness of the subsequent clustering process in the brain network analysis.

2.5.4 Conclusion

In summary, the Brain Network Transformer (BRAINNETTF) introduces an innovative approach to neuroimaging analysis, leveraging a Multi-Head Self-Attention (MHSA) module and an Orthonormal Clustering Readout (OCREAD) function. The model excels in capturing intricate brain network relationships, demonstrating superior performance in tasks like autism detection and connectivity analysis. While facing challenges in interpretability, BRAINNETTF holds promise for advancing neuroimaging research and exhibits potential for broader applications beyond this study, contributing to a deeper understanding of brain function and disorders.

2.6 Summary and Gaps Identified

2.6.1 Summary

Title	Advantages	Disadvantages
Improving the detection of Autism Spectrum Disorder by combining structural and functional MRI information[1]	<ul style="list-style-type: none">• Integration of smri and fmri data provides a nuanced understanding of ASD-related brain patterns.• Use of stacked autoencoders in unsupervised learning, coupled with supervised multilayer perceptrons, enables effective feature learning and classification.• Studies functional connectivity and gray matter volumes to gain insights into both local and global brain alterations in ASD.• Classification accuracy of 85.06 ± 3.52	<ul style="list-style-type: none">• Interpreting learned features in the unsupervised phase poses challenges, hindering straightforward clinical understanding.• Dependence on diverse datasets introduces susceptibility to variations in data quality and acquisition protocols.• Head movements during MRI scans can introduce artifacts, potentially compromising signal quality and accuracy.

Title	Advantages	Disadvantages
A novel machine learning based framework for detection of autism spectrum disorder (ASD)[2]	<ul style="list-style-type: none"> • Addresses the challenge of elusive biomarkers in ASD by focusing on corpus callosum and intracranial brain volume, contributing to a deeper understanding of the disorder. • Prioritizes discriminative features, enhancing model training efficiency and reduces complexity in ASD detection 	<ul style="list-style-type: none"> • Only limited scope since study does not focus on behavioural aspects. • The effectiveness of the proposed approach heavily relies on the availability and diversity of MRI data.

Title	Advantages	Disadvantages
CNNG: A Convolutional Neural Networks With Gated Recurrent Units [3]	<ul style="list-style-type: none"> • Fully exploits the spatio-temporal information of fMRI data to extract more discriminative features of ASD. • Avoids the loss of information due to excessive dimensionality reduction caused by using brain segmentation for classification. • Prevents overfitting by using intercepting time dimension, scaling brain image size, regularization and dropout techniques 	<ul style="list-style-type: none"> • Requires a large amount of data to train the deep learning model. • Sensitive to noise and artifacts in fMRI data. • Lacks interpretability of the spatio-temporal features extracted by the model.

Title	Advantages	Disadvantages
FNet: Mixing Tokens with Fourier Transforms[4]	<ul style="list-style-type: none"> FNet introduces a novel approach to token interactions with Fourier mixing sublayers, offering a unique perspective on capturing and processing dependencies in input sequences. The replacement of self-attention with Fourier simplifies computational complexity, potentially leading to more efficient training and inference. FNet's decision to extract and retain only the real part of the Fourier Transform contributes to model simplicity. 	<ul style="list-style-type: none"> FNet's Fourier mixing sublayers may limit capturing long-range dependencies. FNet's effectiveness varies across tasks, depending on the benefits of traditional attention mechanisms in different natural language processing applications. Fourier mixing sublayers in FNet might complicate interpretability, making it more challenging to understand how the model processes information.

Title	Advantages	Disadvantages
Brain Network Transformer[5]	<ul style="list-style-type: none"> The model considers modular-level similarities between brain regions, enhancing accurate representation. Capture diverse attention patterns across multiple heads in the multi head attention module, enhancing the node features. The model's OCREAD method can be generalized to various graph learning tasks such as compressing node embeddings to subgraph embeddings. 	<ul style="list-style-type: none"> Computational complexity might increase due to complex architecture, which might affect scalability. In real-world applications, obtaining optimal initializations might be challenging, affecting the model's performance. Orthonormal clustering's limited interpretability challenges the provision of clear insights into cluster assignments, reducing overall model transparency.

Table 2.1: Advantages and Disadvantages of Methods

2.6.2 Gaps identified

- Limited consideration of diverse age groups and genders in ASD classification models.
- Inefficiencies in extracting meaningful features from both types of MRI by not using a tailored approach for each type of MRI.
- Challenges in interpretability and explanation of patterns identified by the models.
- Limited integration of findings across different technological approaches.

Chapter 3

Requirements

3.1 Hardware and Software Requirements

3.1.1 Software Requirements

- PyTorch
- Craddock 200 Atlas
- NVIDIA (8GB Memory)
- wandb
- Google colab

3.1.2 Hardware Requirements

- RAM: 8 GB or more
- Disk space: 8 GB or more
- GPU: GeForce RTX 3060 Ti

Chapter 4

System Architecture

Introduction

This chapter introduces a novel architecture using functional MRI data for precise binary classification. Our approach boosts accuracy and unravels deeper insights into the complex world of brain connectivity. Using advanced techniques in Fourier Encoder, our model aims to discern critical features.

4.1 System Overview

The system is centred around the integration of functional MRI data for binary classification, distinguishing individuals into Autism Spectrum Disorder (ASD) or typically developing categories. The emphasis is on the holistic utilization of neuroimaging modalities to enhance classification accuracy.

4.1.1 Data Acquisition and Preprocessing

- Acquired from the DPASRF preprocessing pipeline.
- Comprehensive preprocessing for extraction of mean time series from regions of interest using the CC200 functional parcellation atlas.
- The pivotal objective in feature extraction is the meticulous construction of connectivity matrices for each subject.
- For functional data, these matrices depict the intricate connections between different brain ROIs, by analyzing BOLD signals.
- After the correlation matrix's lower triangular portion is vectorized, a feature vector with length $k(k - 1)/2$, where k is the number of ROIs, is produced. This vector is

then standardized and used as an input for our models.

4.1.2 Feature Extraction

- Enhances feature extraction from the fMRI data by using a Fourier encoder that utilizes Discrete Fourier Transform (DFT) to convert functional connectivity matrices into the frequency domain, capturing unique brain connectivity patterns.
- The encoder in this architecture is composed of a Fourier mixing sublayer and a feed-forward sublayer.
- Residual connections are integrated into each sublayer for efficient information flow and gradient propagation, aiding optimization in deep networks.

4.1.3 Classification

- The extracted features are classified as autistic and non-autistic by passing through a softmax function layer.

4.2 Architectural Design

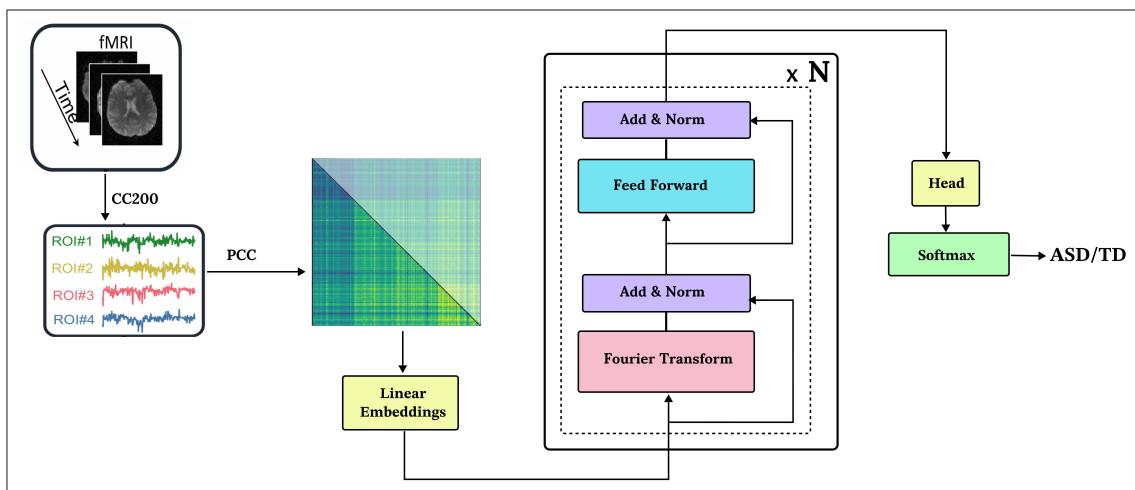


Figure 4.1: Architecture Diagram of the proposed model

4.3 Sequence Diagram

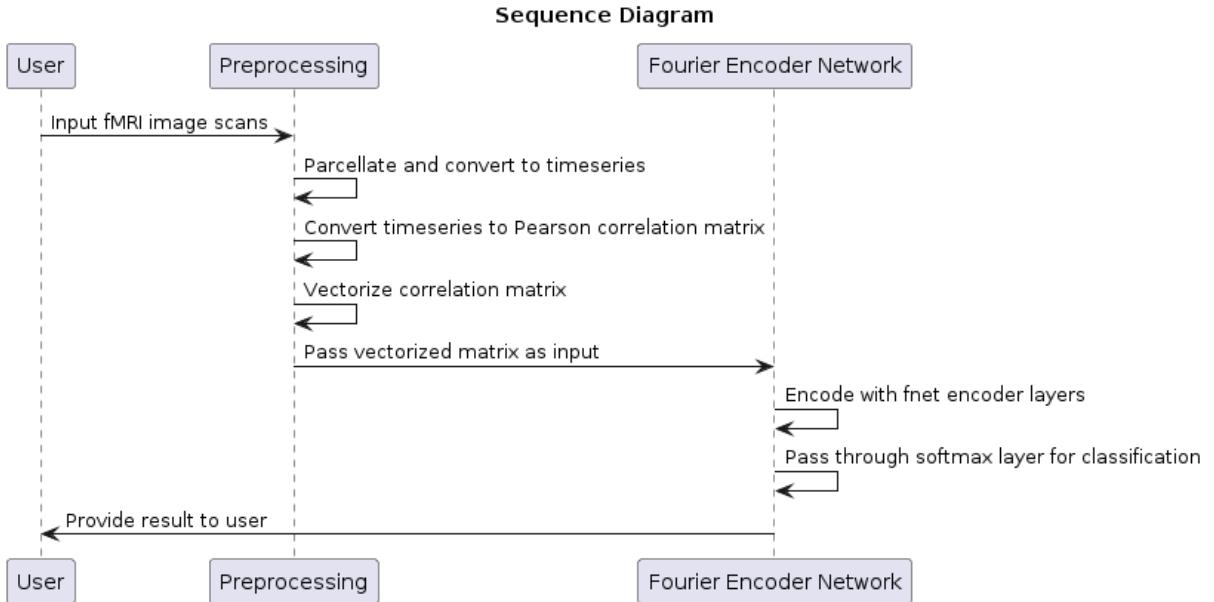


Figure 4.2: Architecture Diagram of the proposed model

4.4 Module Division

The proposed model is structured around three integral modules, each crucial in the overall processing pipeline. The preprocessing module marks the starting point, meticulously cleaning, normalizing, and transforming raw data to ensure optimal input quality for subsequent stages. This module also involves constructing functional connectivity matrices based on BOLD signals, utilizing Pearson correlation coefficients. It incorporates advanced techniques to identify and retain the most informative features.

Following this, the feature extraction module aims to distil meaningful patterns from the preprocessed data. The Fourier transformer encoder refines the feature space in the classification module by emphasizing key frequency-domain characteristics. It uses the curated features to train a robust model capable of making accurate predictions. This comprehensive approach, integrating preprocessing and feature extraction, followed by a sophisticated classification strategy, forms a holistic framework designed to unlock intricate patterns in the data and enhance the model's performance in decision-making tasks.

4.4.1 Preprocessing Module

The preprocessing module orchestrates the refinement of functional MRI data. Employing specialized pipelines like Data Processing Assistant for Resting-State fMRI (DPASRF) toolbox and atlases like CC200, this module seeks to extract meaningful insights. For functional MRI (fMRI) data, the focus lies in obtaining time series data of mean BOLD signals from diverse brain regions. This is accomplished through the application of the CC200 atlas, comprising 200 regions.

Preprocessing Steps for Functional Data

Processing Pipeline (DPARSF):

- Data Conversion: Convert raw images to compatible formats like NIfTI format.
- Slice Timing Correction: Correct for slice acquisition time differences.
- Realignment: Correct for head motion during the scan by aligning all volumes to a reference volume.
- Spatial Normalization: Standardize to a common anatomical space like MNI space.
- Spatial Smoothing: Increase signal-to-noise ratio.
- Temporal Filtering: Removes low-frequency drifts.
- Regression: Removes noise sources like motion and physiological signals.
- Detrending: Removes polynomial or linear trends from the data.
- Quality Control: Assess the quality of data through various measures.
- Functional Connectivity: Analyze temporal correlations between brain regions.

The integration phase merges these distinct processing paths. The DPARSF pipeline refines functional data, extracting BOLD signals based on the CC200 atlas.

The subsequent integration involves cross-referencing the datasets, yielding a final set of 817 cases (368 ASD + 449 control) shared between the original subsets. This consolidated dataset, enriched with both structural and functional information, serves as the foundation for the proposed classification strategy.

Functional Connectivity Matrix (BOLD signals)

For each subject, a connectivity matrix is constructed. The matrix has dimensions of 200×200 , depending on the atlas used. Each element (i, j) in the matrix represents the Pearson correlation coefficient computed for the mean BOLD series from regions i and j , using the formula:

$$r_{ij} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}$$

The correlation coefficient quantifies the degree of linear relationship between the signals. This process is performed for all pairs of regions defined by the atlas.

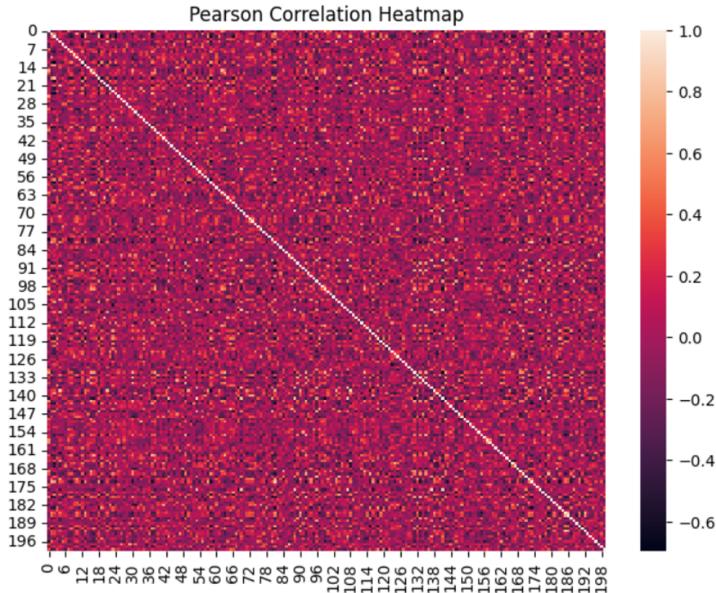


Figure 4.3: Pearson Correlation Coefficient heatmap

Functional connectivity matrices undergo preprocessing steps to prepare them for the model. Only the lower triangle of the matrix, excluding the diagonal is considered. To facilitate efficient computation and reduce redundancy, we focus on vectorizing the lower triangular part of the correlation matrix. This process results in a feature vector of length $k(k - 1)/2$, where k represents the number of ROIs. These result in features that encapsulate essential brain connectivity patterns, serving as input for further processing and classification.

4.4.2 Feature Extraction Module

The Feature extraction module is a systematic pipeline designed to extract informative features from brain network data, utilizing a Fourier transformer encoder. However, the decoder part of the transformer architecture, which is used for sequence generation, is omitted since the goal here is classification, not generation. The encoder comprises two encoder layers stacked on top of each other. Each encoder layer consists of a Fourier mixing sublayer followed by a feedforward sublayer.

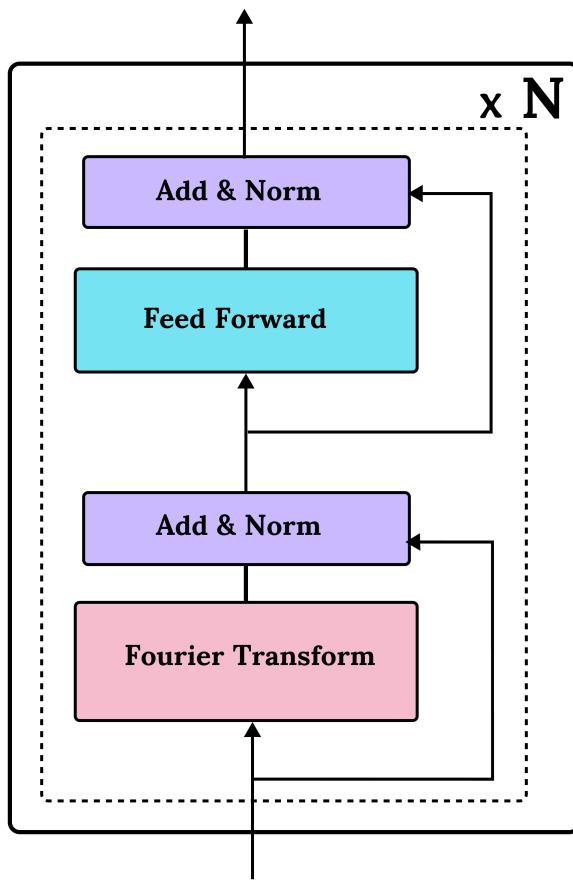


Figure 4.4: Fourier Encoder

Fourier Mixing Sublayer

The Fourier mixing sublayer is the first component within each encoder layer. It is responsible for incorporating frequency-specific features into the input data. 1D Discrete Fourier

Transform (DFT) is employed within the Fourier mixing sublayer, this transformation converts the input functional connectivity matrices into the frequency domain, generating, generating frequency-specific features. These features encapsulate unique characteristics of brain connectivity patterns, including amplitude and phase information across different frequency components. These features enrich the model’s input representation, aiding in the identification of subtle differences in neural activity relevant to Autism Spectrum Disorder (ASD).

Feed Forward Sublayer

The feedforward sublayer in the encoder comprises sequential operations aimed at processing input data to extract meaningful features. It begins with a linear transformation, $\text{Linear}(xW_1 + b_1)$, where x represents the input data, W_1 denotes the weight matrix, and b_1 represents the bias vector. This operation projects the input data into a higher-dimensional space. Following this, a Gaussian Error Linear Unit (GELU) activation function, $\text{GELU}(x) = 0.5x(1 + \tanh(2/(x + 0.044715x^3)))$, introduces non-linearity into the transformation. This smooth approximation of the rectified linear unit (ReLU) facilitates better gradient propagation during training. Subsequently, dropout regularization is applied to prevent overfitting by randomly dropping a fraction of input units. Layer normalization, LayerNorm, ensures stable activations across the feature dimension. Finally, another linear transformation, $\text{Linear}(xW_2 + b_2)$, where W_2 denotes the weight matrix and b_2 represents the bias vector, maps the data back to the original feature space, yielding the final output. Mathematically, the sequence of operations is encapsulated by the equation:

$$\text{FFN}(x) = \text{LayerNorm}(\text{GELU}(\text{Dropout}(\text{Linear}(xW_1 + b_1)))W_2 + b_2)$$

This equation represents the computational flow within the feedforward sublayer. The feedforward sublayer employs a series of operations including linear transformations, non-linear activation functions (GELU), dropout regularization, and layer normalization to process the input data and extract meaningful features facilitating the transformer’s ability to process input data and extract discriminative features effectively.

A residual connection is incorporated into each sublayer of the encoder design to promote more seamless information flow and reduce the possibility of disappearing gradients. The original input is added to the sublayer's output using this method. Within each sublayer S and input x , the output x' is computed as:

$$x' = S(x) + x$$

Remaining constant throughout training, the residual connection links the input directly to the output, allowing the network to convey gradients efficiently. This approach helps avoid performance loss in addition to assisting with optimization, especially in deeper networks with multiple layers.

The classification part involves using the extracted features from the encoder to make predictions about the class labels of the input data. Following the encoder layer, the softmax function is applied for classification. The softmax function takes the output of the encoder, which represents the learned features and converts them into probabilities across the predefined classes, ASD and TD.

4.5 Work Schedule - Gantt Chart



Figure 4.5: Gantt chart

Conclusion

In conclusion, the chapter provides a detailed exploration of the system architecture tailored for the integration of functional MRI data. The architectural design, module division, and associated methodologies underscore the depth and complexity involved in tackling the challenges of binary classification, particularly in distinguishing individuals with Autism Spectrum Disorder.

The journey unfolds with the preprocessing module, where raw data undergoes meticulous cleaning, normalization, and transformation, ensuring optimal input quality for subsequent stages. It also involves constructing functional connectivity matrices, laying the groundwork for informative features that encapsulate the essence of brain connectivity.

The feature extraction module incorporates cutting-edge techniques such as Fourier Encoder, elevating the model's ability to discern essential features from the vast array of connectivity data. Finally, the extracted features are classified using a softmax function, culminating in a robust binary classification model.

As the project progresses, the Gantt chart provides a visual representation of the work schedule, outlining key milestones and timelines. This chapter, rich in detail and methodology, sets the stage for the subsequent implementation and validation phases, promising advancements in the understanding and classification of Autism Spectrum Disorder through the lens of neuroimaging data.

Chapter 5

Result

5.1 Overview

This chapter provides a comprehensive analysis of the results obtained from our experimentation with four distinct models designed for the classification of subjects into ASD (Autism Spectrum Disorder) and TD (Typically Developing) categories. We begin by presenting detailed evaluations of each model's performance metrics, including accuracy, precision, recall, and F1 score. The comparison between the two baseline models BNT and METAFomer and their counterparts enhanced with the Fourier layer, aims to discern any performance improvements resulting from this addition. Graphical analyses showcasing the comparative results of all four models further elucidate their respective strengths and weaknesses. Through this meticulous examination, we aim to provide valuable insights into the efficacy and impact of incorporating the Fourier layer in our classification models for ASD detection and classification.

5.2 Models and their result analysis

5.2.1 BNT

A comprehensive evaluation of the model's performance across various metrics, including accuracy, precision, recall, and F1 score was conducted. By analyzing the outcomes of our experiments with the BNT model, we aim to establish a baseline understanding of its capabilities and limitations in the context of ASD detection and classification.

```
wandb: Run summary:  

wandb:           Test AUC 0.81012  

wandb:           Test Accuracy 70.9  

wandb:           Test Loss 7.70172  

wandb:           Test Sensitivity 0.69811  

wandb:           Test Specificity 0.7234  

wandb:           Train Accuracy 53.83523  

wandb:           Train Loss 3.75441  

wandb:           Val AUC 0.76483  

wandb:           micro F1 0.71  

wandb:           micro precision 0.71
```

Figure 5.1: BNT model Evaluation Metrics



Figure 5.2: Training curves of BNT

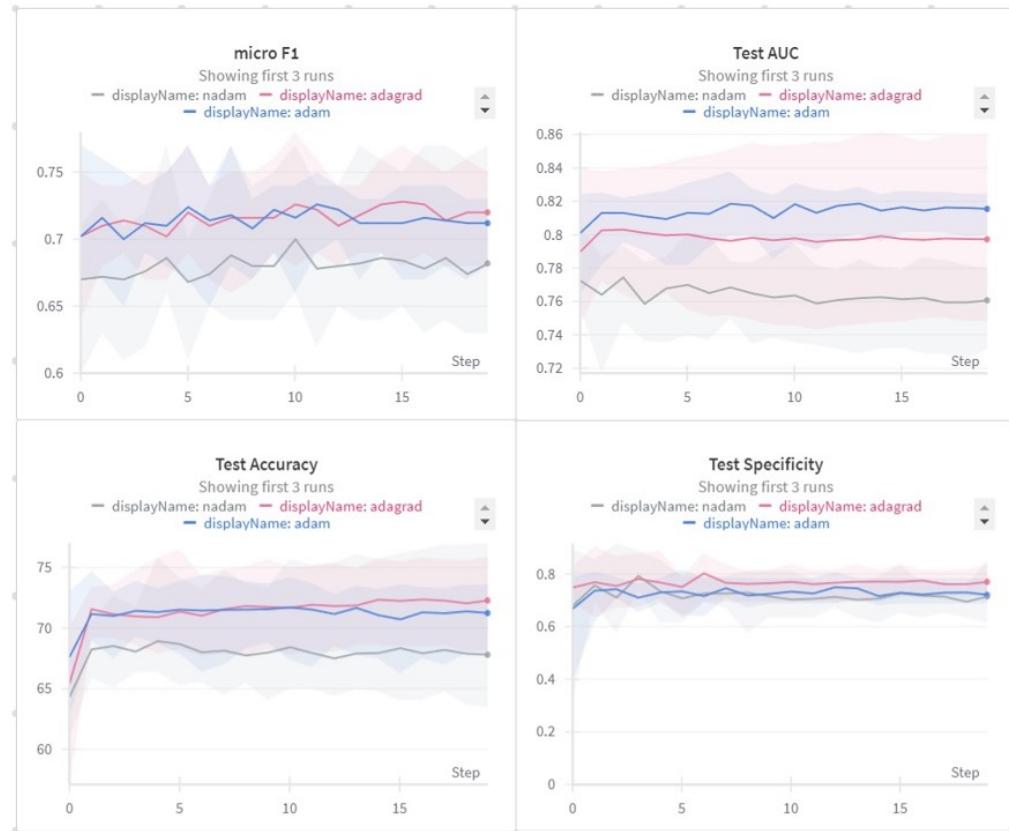


Figure 5.3: Training curves of BNT

5.2.2 BNT-Fourier

Building upon the foundational framework of the BNT model, the BNT-Fourier model integrates the Fourier layer to explore its potential impact on classification performance. Subsequently, we conduct a comprehensive evaluation of the BNT-Fourier model's performance across various metrics, including accuracy, precision, recall, and F1 score. Through this analysis, we could observe that except for the F1 score, better results were obtained in terms of other measures.

```
{
    "Train Loss": 3.8607517131688915,
    "Train Accuracy": 54.062499999999986,
    "Test Loss": 9.803627846240998,
    "Test Accuracy": 73,
    "Val AUC": 0.7721639471639471,
    "Test AUC": 0.8241185897435898,
    "Test Sensitivity": 0.6346153846153846,
    "Test Specificity": 0.8333333333333334,
    "micro F1": 0.7299999999999999,
    "micro recall": 0.73,
    "micro precision": 0.73,
    "_timestamp": 1704282324.4284894,
    "_runtime": 204.52716445922852,
    "_step": 19,
    "_wandb": {
        "runtime": 203
    }
}
```

Figure 5.4: BNT-F Evaluation metrics

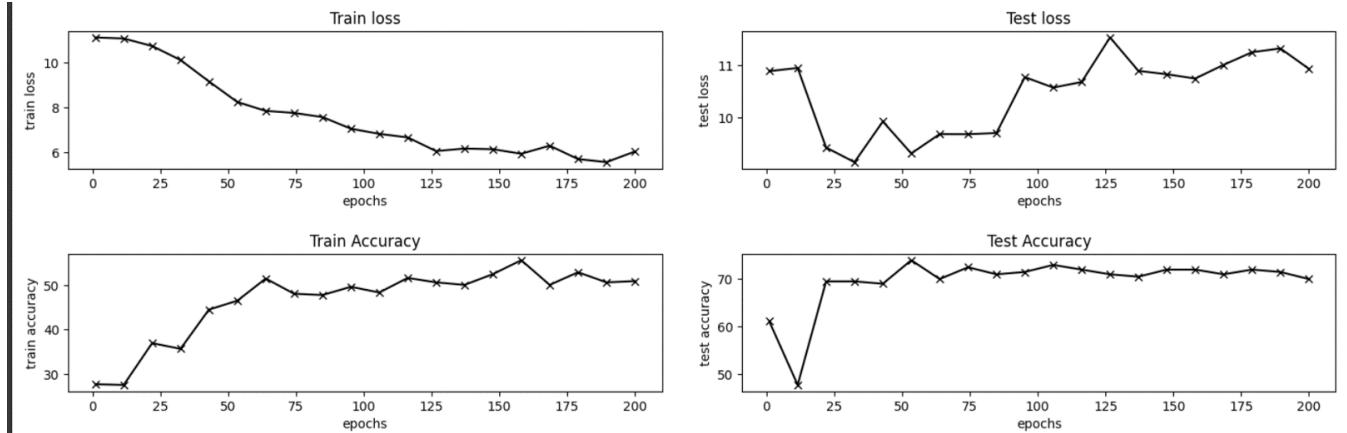


Figure 5.5: Training curves of BNT-F

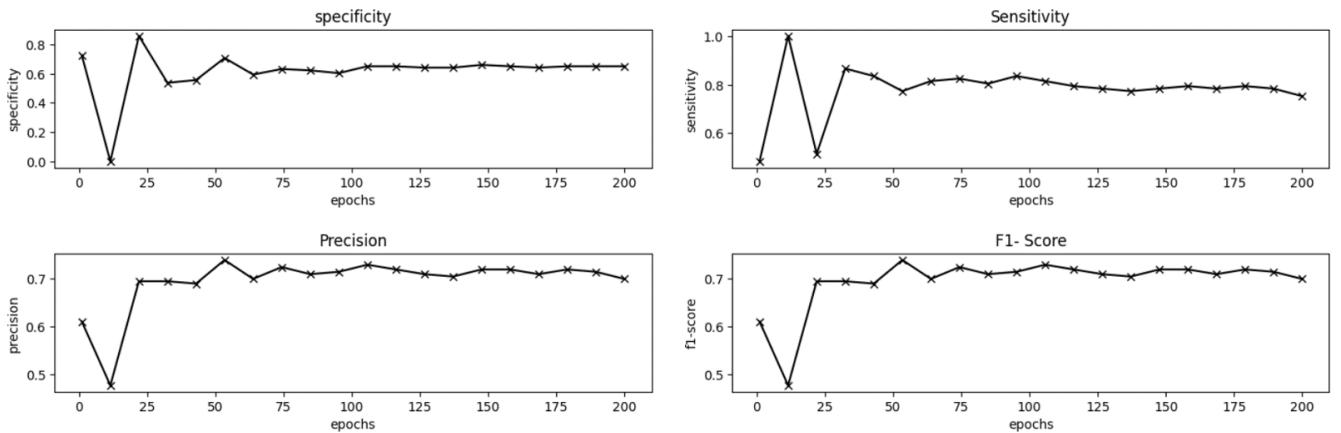


Figure 5.6: Training curves of BNT-F

5.2.3 METAFormer

METAFormer serves as another foundational framework for comparison and evaluation. We conduct a comprehensive evaluation of the METAFormer model's performance across various metrics, including accuracy, precision, recall, and F1 score. By analyzing the outcomes of our experiments with METAFormer, we aim to gain insights into its capabilities and limitations in the context of ASD detection and classification.

	Fold	Accuracy	Precision	Recall	F1	AUC	AP	\
0	0	0.644788	0.632787	0.728302	0.677193	0.642807	0.599856	
1	1	0.624758	0.619529	0.694340	0.654804	0.622963	0.586836	
		FPR	FNR	TPR	TNR			
0	0	0.442688	0.271698	0.728302	0.557312			
1	1	0.448413	0.305660	0.694340	0.551587			
<hr/>								
Mean accuracy: 0.6347729326452731								
Std accuracy: 0.010014712142371762								

Figure 5.7: METAFormer model Evaluation Metrics

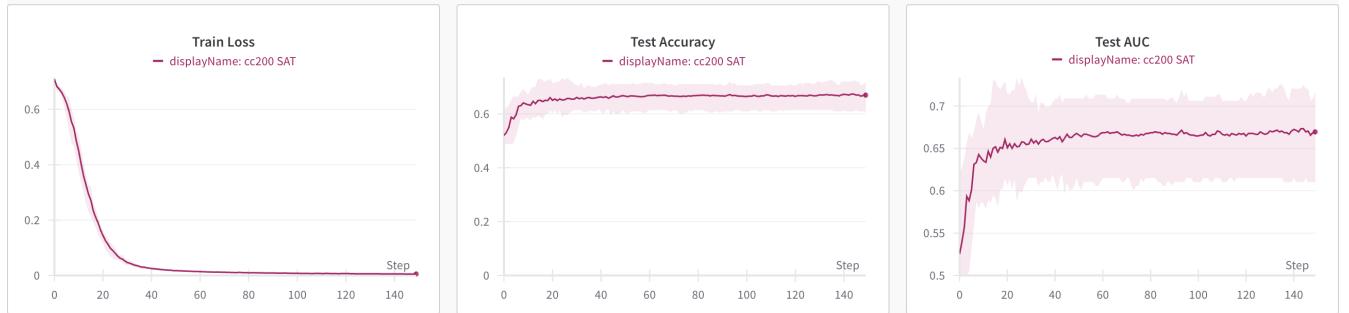


Figure 5.8: Training curves of METAFormer model

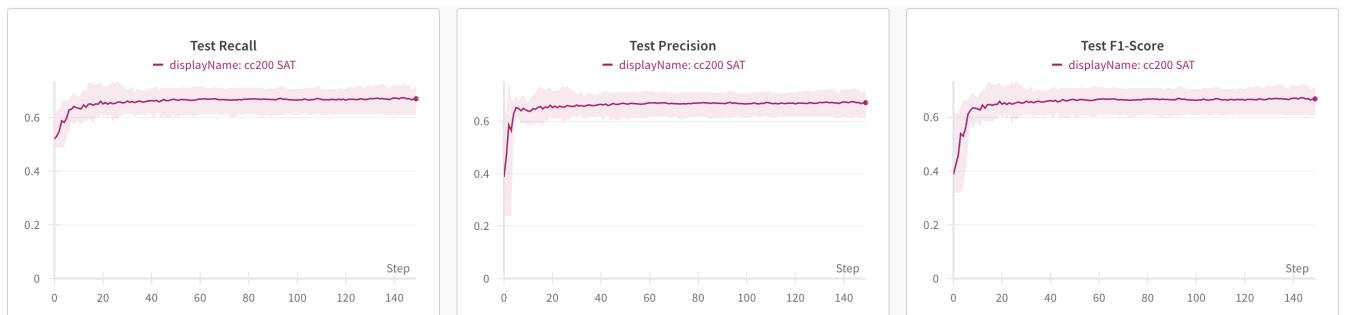


Figure 5.9: Training curves of METAFormer model

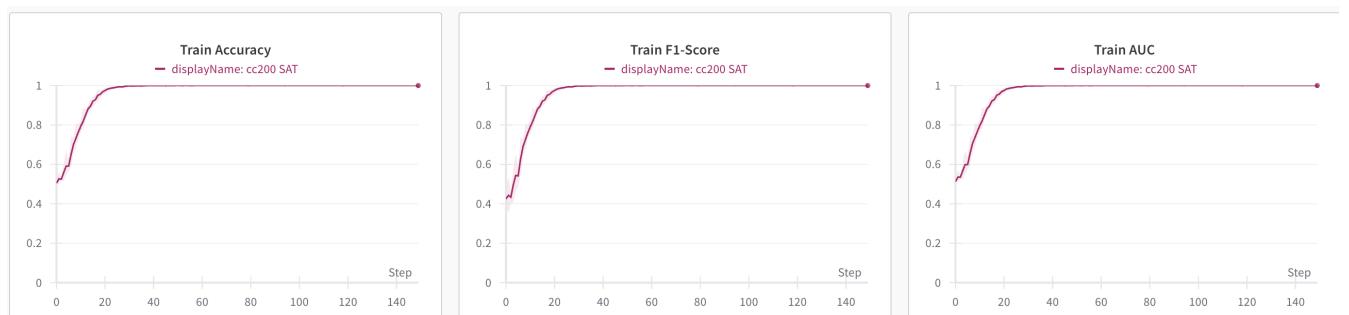


Figure 5.10: Training curves of METAFormer model

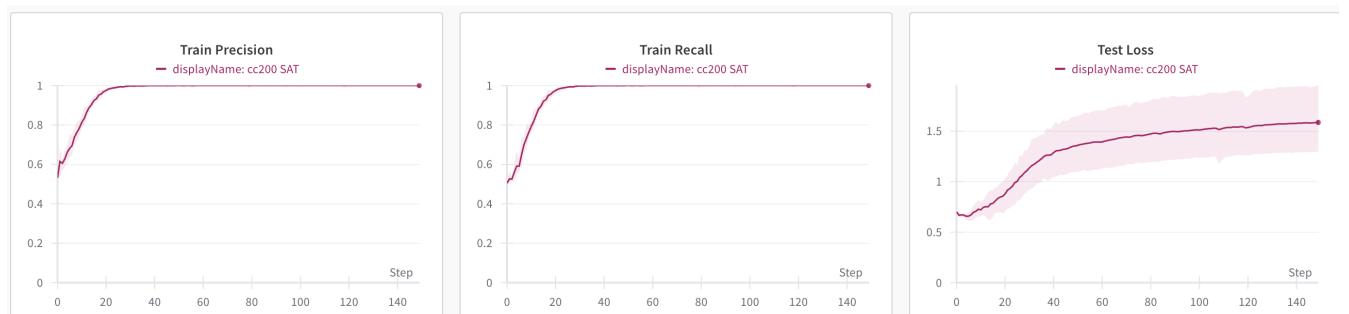


Figure 5.11: Training curves of METAFormer model

5.2.4 METAFormer-Fourier

The analysis compares the performance metrics, including accuracy, precision, recall, and F1 score, of the Metaformer model with the Fourier layer. The experiments demonstrate comparable results between the two models.

Fold	Accuracy	Precision	Recall	F1	AUC	AP	\
0	0.596525	0.595890	0.656604	0.624776	0.595100	0.566940	
1	0.638298	0.640288	0.671698	0.655617	0.637436	0.598359	
<hr/>							
FPR	FNR	TPR	TNR				
0	0.466403	0.343396	0.656604	0.533597			
1	0.396825	0.328302	0.671698	0.603175			
<hr/>							
Mean accuracy: 0.6174114844327611							
Std accuracy: 0.020886387907664505							

Figure 5.12: METAFormer-F model Evaluation Metrics



Figure 5.13: METAFormer-F result graph

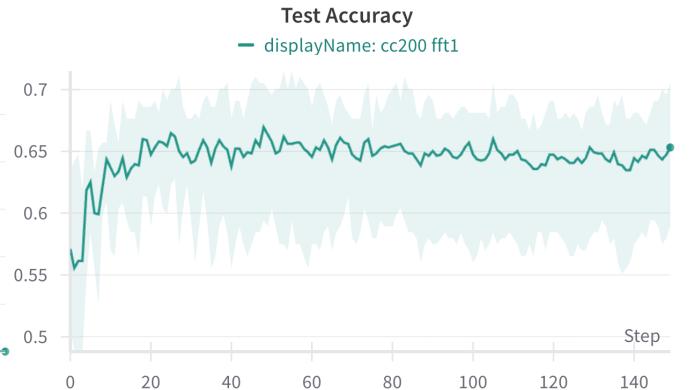


Figure 5.14: METAFormer-F result graph

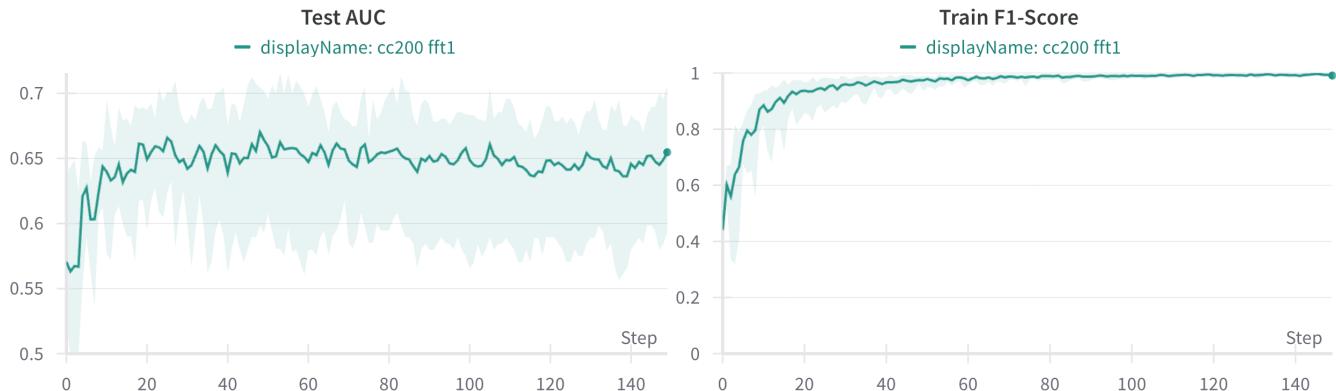


Figure 5.15: METAFormer-F result graph

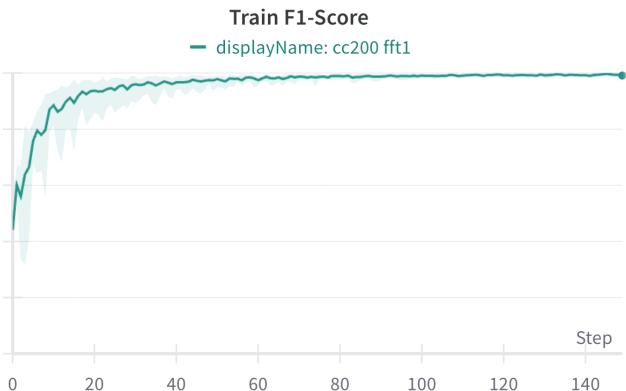


Figure 5.16: METAFormer-F result graph

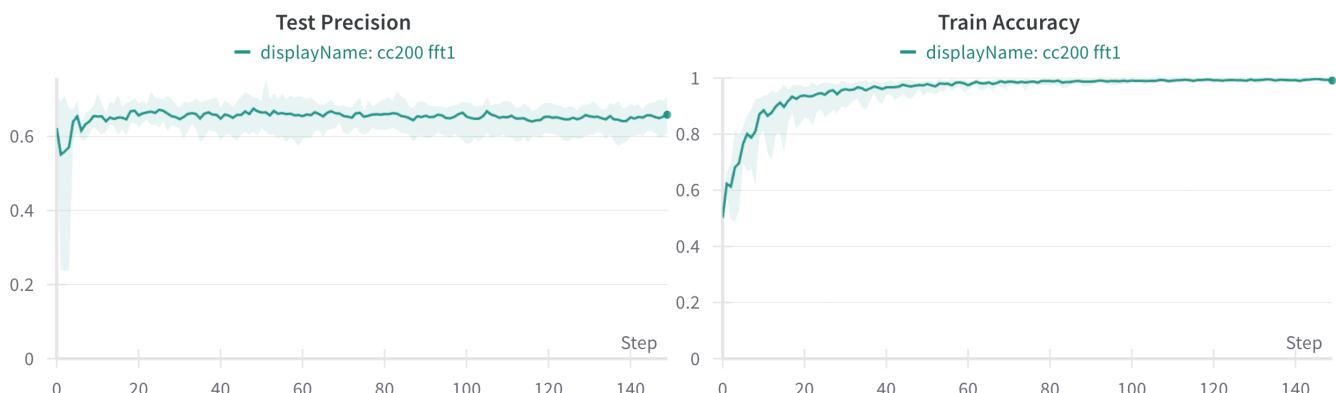


Figure 5.17: METAFormer-F result graph

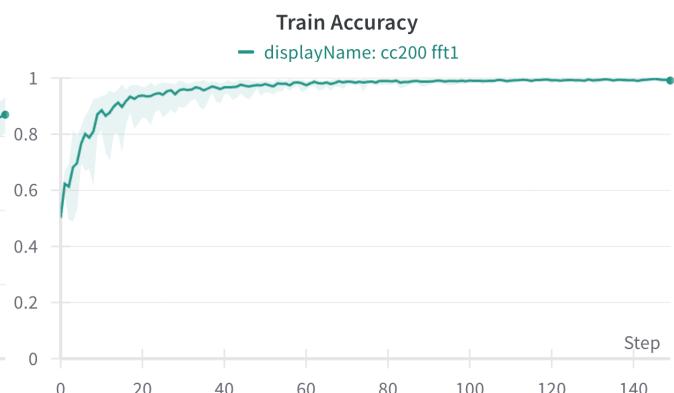


Figure 5.18: METAFormer-F result graph

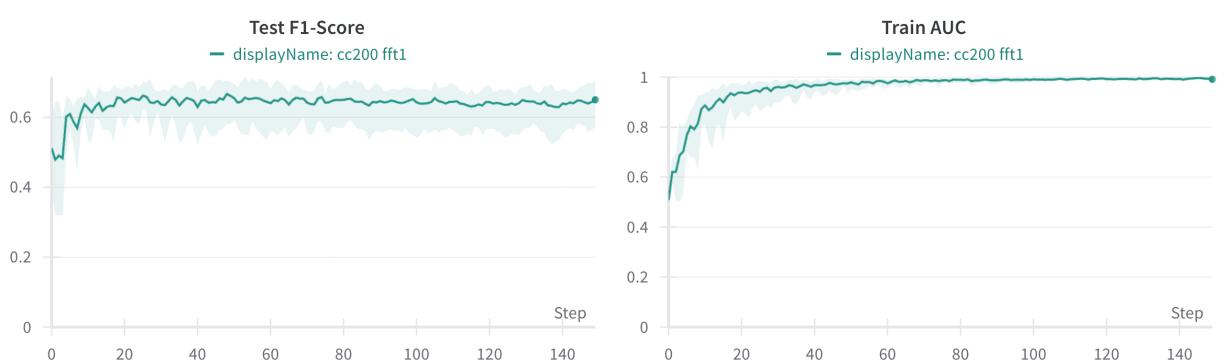


Figure 5.19: METAFormer-F result graph

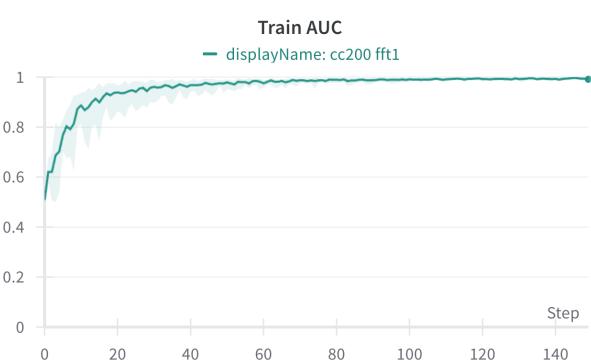


Figure 5.20: METAFormer-F result graph

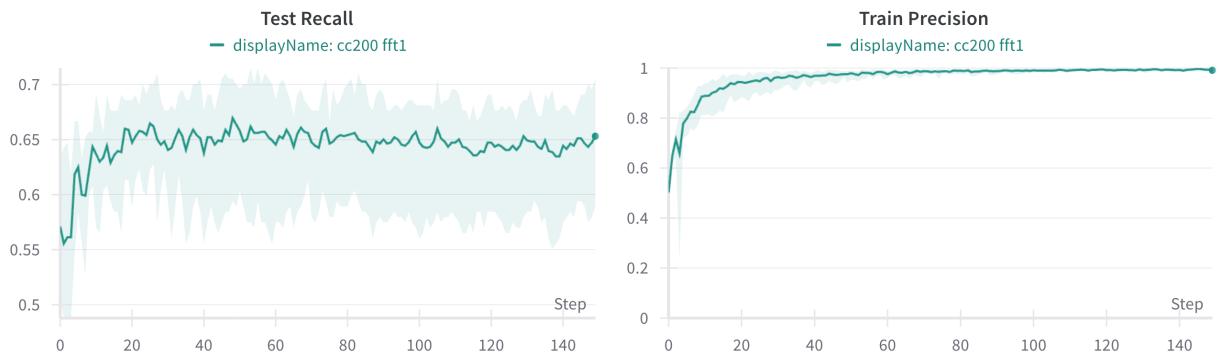


Figure 5.21: METAFormer-F result graph

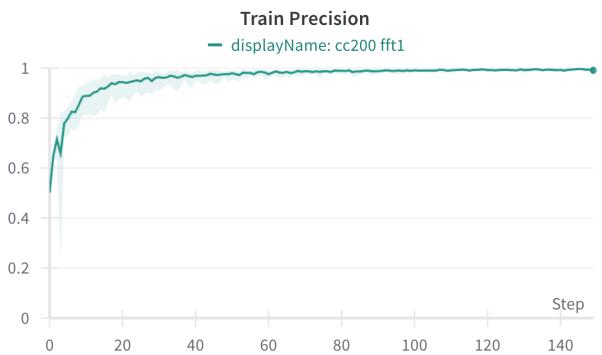


Figure 5.22: METAFormer-F result graph

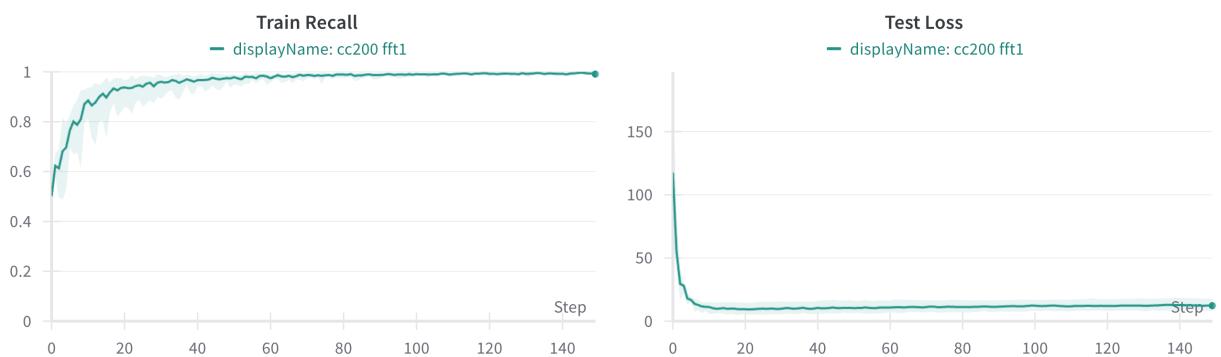


Figure 5.23: METAFormer-F result graph

Figure 5.24: METAFormer-F result graph

A comparative study between the original METAFormer model and our proposed METAFormer-Fourier model showed a reduction in the number of parameters and comparable results for our proposed model. The comparison of the parameters and evaluation metrics is shown below.

```
Metaformer model
-----
Number of parameters : 5628170
Total number of trainable parameters in model: 5628170
```

Figure 5.25: Parameters in METAFormer model

```
FNet model
-----
Number of parameters : 5103882
Total number of trainable parameters in model: 5103882
```

Figure 5.26: Parameters in METAFormer-Fourier model

The graphs shown below indicate a comparison of METAFormer and METAFormer Fourier models. The red color curves indicate metrics of METAFormer while the green indicates the metrics of METAFormer Fourier model.

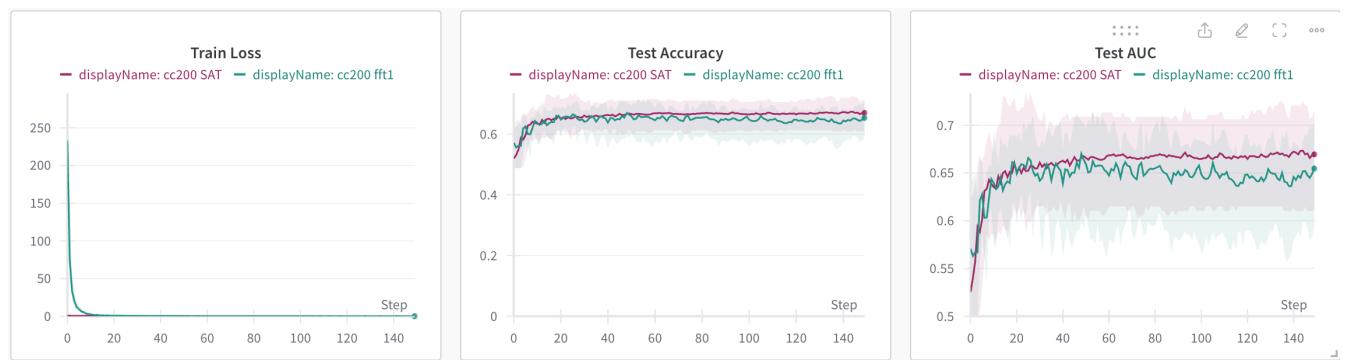


Figure 5.27: Graphical comparison of METAFormer & METAFormer-F

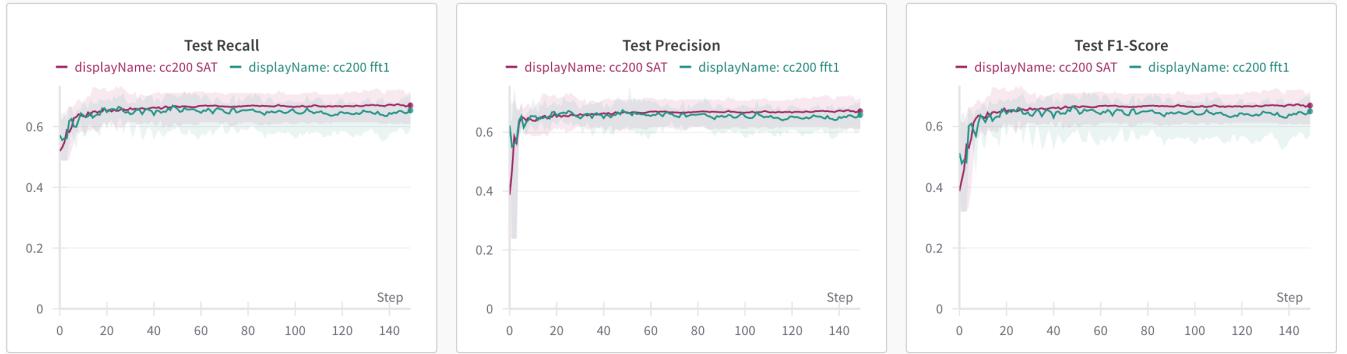


Figure 5.28: Graphical comparison of METAFormer & METAFormer-F

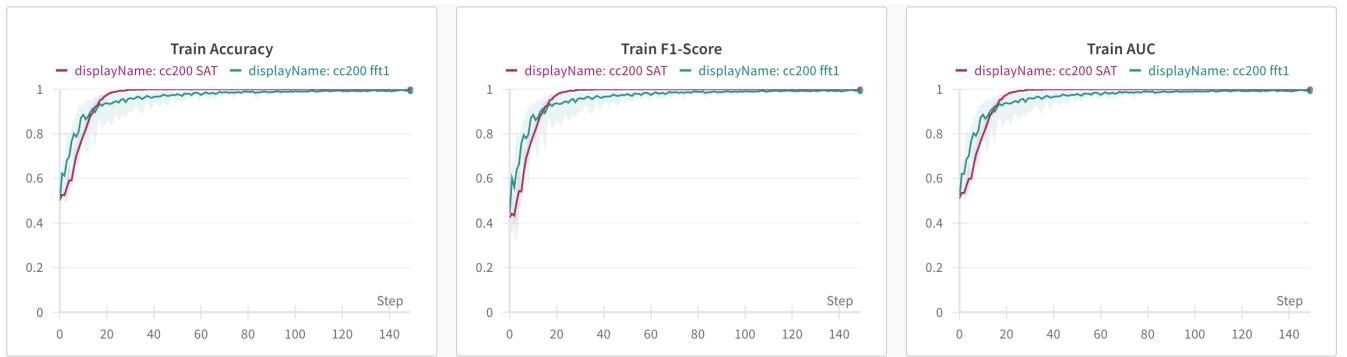


Figure 5.29: Graphical comparison of METAFormer & METAFormer-F

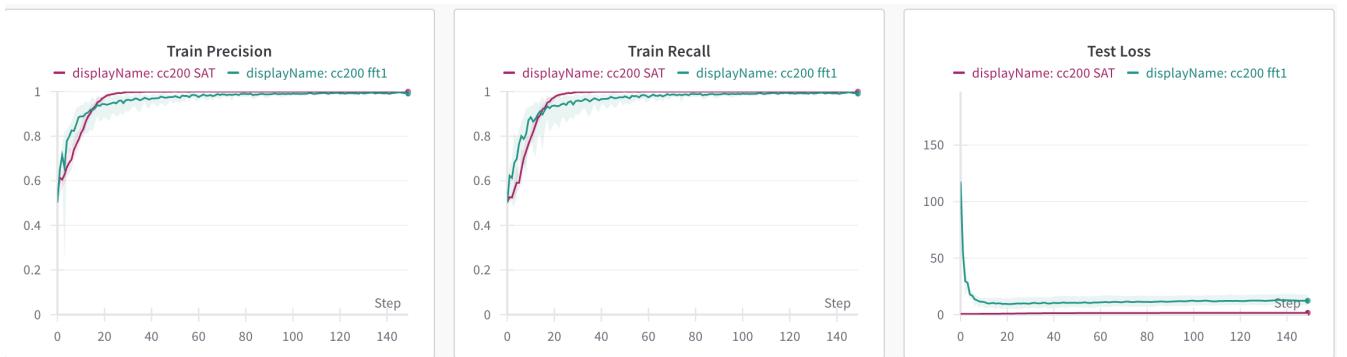


Figure 5.30: Graphical comparison of METAFormer & METAFormer-F

5.2.5 Comparison of models

The table given below outlines the comparison of the evaluation metrics of the four models, two baseline models and two Fourier-incorporated models.

MODEL	ACC	PREC	RECALL	F1	AUC	TRAIN ACC
BNT	.709	.71	.69	.71	.76	.5383
BNT-Fourier	.73	.73	.73	.72	.82	.5406
METAFormer	.6247	.6195	.6943	.6548	.6229	.975
METAFormer-Fourier	.6174	.618	.6641	.6162	.6162	.97

Table 5.1: Comparison of Evaluation metrics

Chapter 6

Conclusions & Future Scope

In conclusion, the proposed model aims to diagnose autism spectrum disorder early and accurately from fMRI images by using transformer encoder architectures. The model is structured around three modules, which are preprocessing, feature extraction and classification. Commencing with the preprocessing module, raw MRI data undergoes cleaning, normalization, and transformation through the DPARSF pipeline to ensure optimal input quality for subsequent stages. From the preprocessed data functional connectivity matrices are constructed. These matrices, leveraging BOLD signals incorporate Pearson correlation coefficient values. These matrices are flattened to obtain a 1D feature vector which acts as input to the next layer. The feature extraction module incorporates the Fourier encoder to refine the feature space with an emphasis on key frequency-domain characteristics. Ultimately, the classification module utilizes the extracted features to train a robust model capable of precise predictions.

The future scope of our autism detection project involves expanding the framework by integrating additional imaging modalities, such as fMRI or DTI, to enhance the depth of analysis. Further exploration with larger and more diverse datasets aims to improve the generalizability of findings. Continuous advancements in deep learning techniques and collaboration with multidisciplinary experts can refine the accuracy and efficiency of our classification process. Additionally, the translation of research findings into practical applications, like assistive technologies or personalized interventions, holds the potential for meaningful contributions to the field and the lives of individuals with Autism Spectrum Disorder (ASD). This ongoing trajectory ensures a dynamic and impactful future for our project.

References

- [1] M. Rakić, M. Cabezas, K. Kushibar, A. Oliver, and X. Llado, “Improving the detection of autism spectrum disorder by combining structural and functional mri information,” *NeuroImage: Clinical*, vol. 25, p. 102181, 2020.
- [2] H. Sharif and R. A. Khan, “A novel machine learning based framework for detection of autism spectrum disorder (asd),” *Applied Artificial Intelligence*, vol. 36, no. 1, p. 2004655, 2022.
- [3] W. Jiang, S. Liu, H. Zhang, X. Sun, S.-H. Wang, J. Zhao, and J. Yan, “Cnng: A convolutional neural networks with gated recurrent units for autism spectrum disorder classification,” *Frontiers in Aging Neuroscience*, vol. 14, p. 948704, 2022.
- [4] J. Lee-Thorp, J. Ainslie, I. Eckstein, and S. Ontanon, “Fnet: Mixing tokens with fourier transforms,” *arXiv preprint arXiv:2105.03824*, 2021.
- [5] X. Kan, W. Dai, H. Cui, Z. Zhang, Y. Guo, and C. Yang, “Brain network transformer,” *Advances in Neural Information Processing Systems*, vol. 35, pp. 25586–25599, 2022.
- [6] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*, 2014.
- [7] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, “Learning phrase representations using rnn encoder-decoder for statistical machine translation,” *arXiv preprint arXiv:1406.1078*, 2014.
- [8] Q. Xin, S. Hu, S. Liu, L. Zhao, and S. Wang, “Wtrpnet: an explainable graph feature convolutional neural network for epileptic eeg classification,” *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 17, no. 3s, pp. 1–18, 2021.

- [9] S. Liu, L. Zhao, J. Zhao, B. Li, and S.-H. Wang, “Attention deficit/hyperactivity disorder classification based on deep spatio-temporal features of functional magnetic resonance imaging,” *Biomedical Signal Processing and Control*, vol. 71, p. 103239, 2022.
- [10] X. Glorot and Y. Bengio, “Understanding the difficulty of training deep feedforward neural networks,” in *Proceedings of the thirteenth international conference on artificial intelligence and statistics*. JMLR Workshop and Conference Proceedings, 2010, pp. 249–256.

Appendix A: Presentation

AUTISM DETECTION USING MRI

GUIDED BY: MS SEEMA SAFAR

ALEENA SIBY | BIBIT SEBASTIAN | CELESTIAN BEN MATHEW | CHRISTO MATHEW

CONTENTS

1. Problem Definition
2. Project Objectives
3. Novelty of idea and scope of implementation
4. Challenges
5. Project Gantt Chart
6. 30% work done
7. 60% work done
8. Interim Results
9. Work to be completed(100%)
10. Task Distribution
11. Conclusion

PROBLEM DEFINITION

- ❑ Existing ASD diagnosis heavily relies on behavioral assessments, which can be subjective and prone to variability among clinicians.
- ❑ This subjectivity can lead to delays in diagnosis and may hinder the accuracy of the identification process.

NOVELTY OF IDEA

- ❑ Replacing the attention layer in multi head attention layer with a 2D Fourier transform layer to reduce complexity.

SCOPE OF IMPLEMENTATION

Healthcare and Clinical Practice:

Healthcare providers can incorporate machine learning tools into their diagnostic processes, enhancing the accuracy and efficiency of autism diagnosis.

Research Advancements:

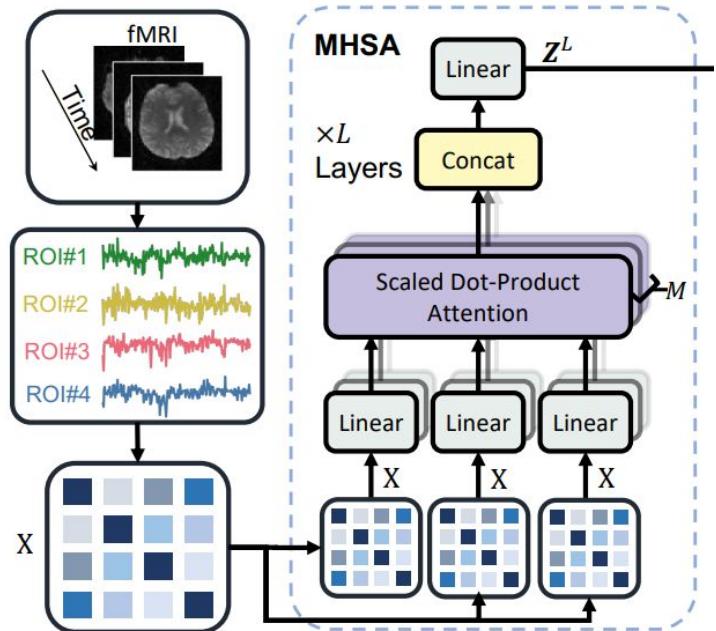
The project can provide valuable insights into the underlying factors contributing to autism, aiding ongoing research efforts.

CHALLENGES

- Lack of availability of suitable dataset
- Causes of ASD are unknown
- Lack of biomarkers.

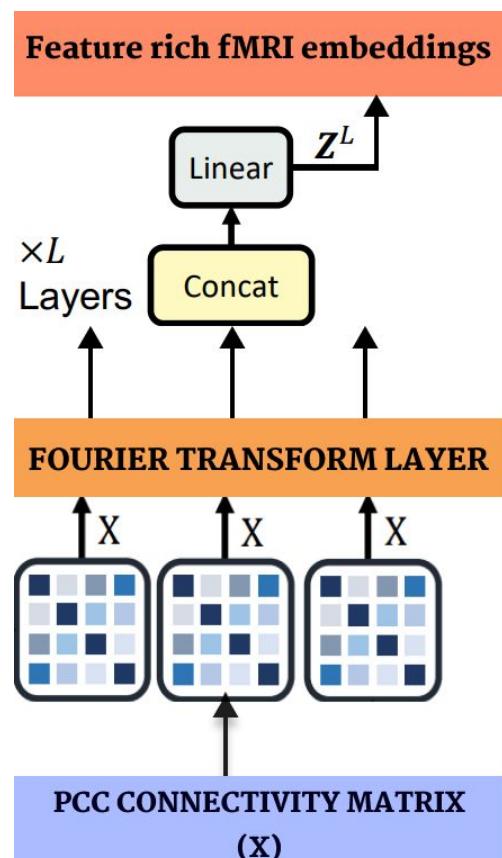
30 % WORK DONE

- Implementation of a transformer encoder that uses a multihead attention mechanism for feature learning from fMRI Time series data.



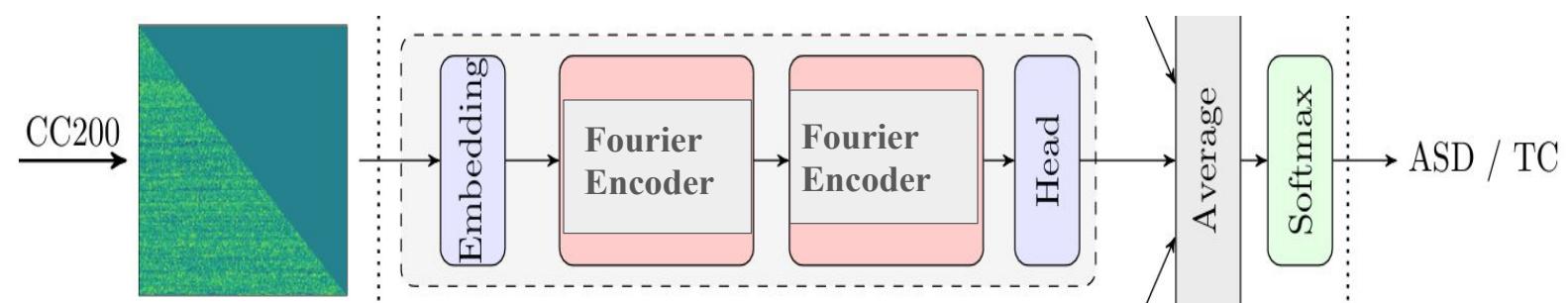
60 % WORK DONE

- Implementation of Fourier Transform Layer instead of the multihead attention layer inside the Transformer encoder.



100 % WORK DONE

- Fine Tuning the model in order to increase accuracy and decrease complexity.
- Preprocessing done on data to decrease the redundancy.
- Simplified the model by reducing the number of layers.



GANTT CHART

	October	November	December	January	February
Project Planning	Research on various input modalities	Review state of the art models	Finding Dataset		
Development			Preprocess MRI Data	Stacked Autoencoder	Attentive Fourier Transformer Encoder
Testing					Evaluate Model Performance
Deployment					Deploy Model
Documentation					Prepare Research Paper

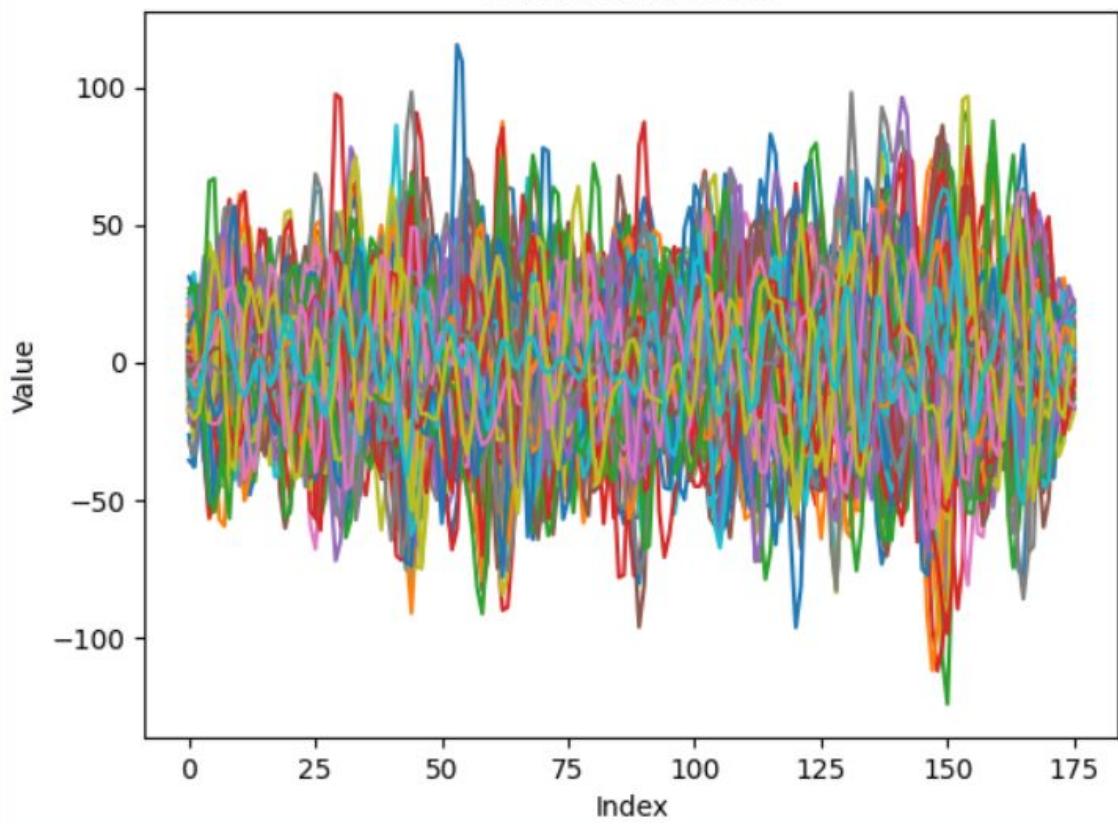
WORK BREAKDOWN AND RESPONSIBILITIES

- ❑ Christo Mathew and Celestian Ben Mathew
 - Implementation of fMRI pipeline - with Fourier transform.
 - Fine tuning.

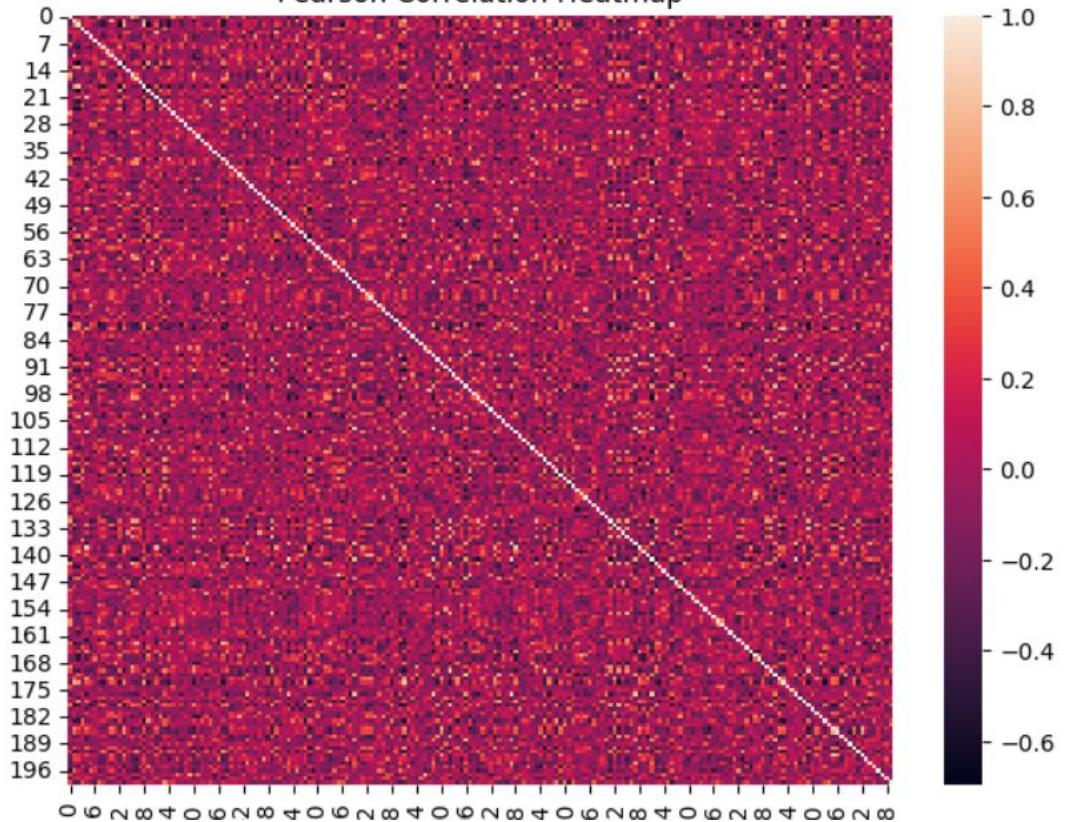
- ❑ Aleena Siby and Bibit Sebastian
 - Implementation of fMRI pipeline - with attention layer.
 - Documentation.

Result Analysis

Data from .1D file



Pearson Correlation Heatmap



BNT RESULTS

```
wandb: Run summary:  
wandb: Test AUC 0.81012  
wandb: Test Accuracy 70.9  
wandb: Test Loss 7.70172  
wandb: Test Sensitivity 0.69811  
wandb: Test Specificity 0.7234  
wandb: Train Accuracy 53.83523  
wandb: Train Loss 3.75441  
wandb: Val AUC 0.76483  
wandb: micro F1 0.71  
wandb: micro precision 0.71
```





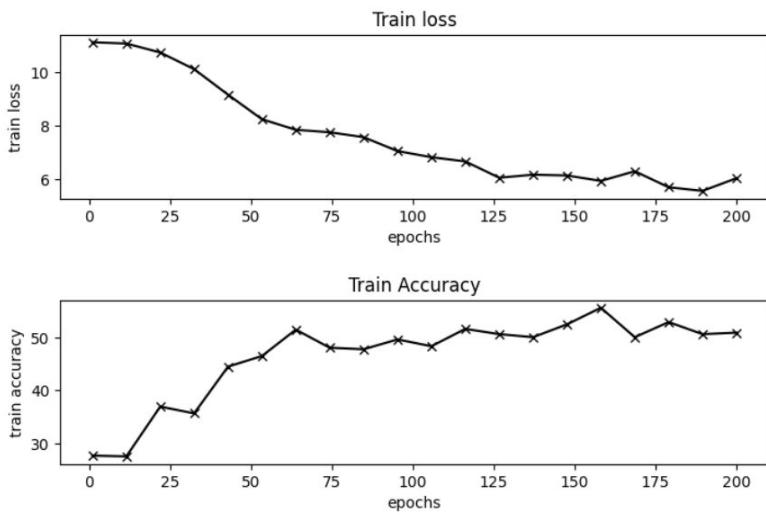
BNT-FOURIER RESULTS

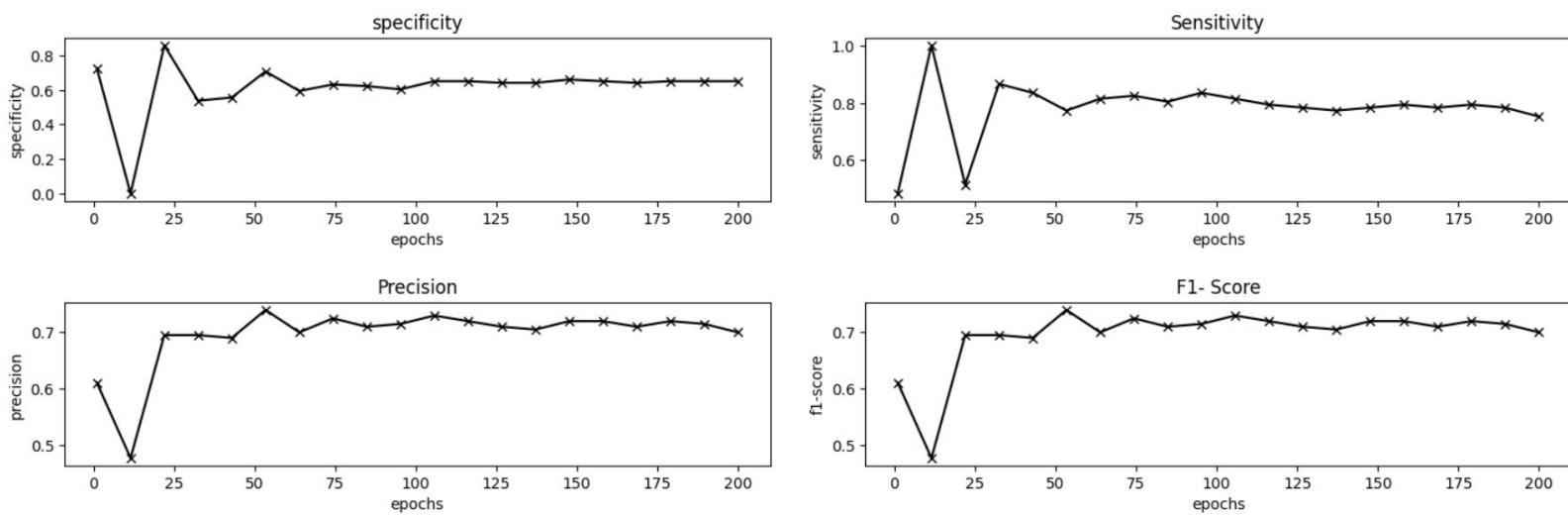
```
{
  "Train Loss": 3.8607517131688915,
  "Train Accuracy": 54.06249999999998,
  "Test Loss": 9.803627846240998,
  "Test Accuracy": 73,
  "Val AUC": 0.7721639471639471,
  "Test AUC": 0.8241185897435898,
  "Test Sensitivity": 0.6346153846153846,
  "Test Specificity": 0.8333333333333334,
  "micro F1": 0.7299999999999999,
  "micro recall": 0.73,
  "micro precision": 0.73,
  "_timestamp": 1704282324.4284894,
  "_runtime": 204.52716445922852,
  "_step": 19,
  "_wandb": {
    "runtime": 203
}
```

```

[2024-02-22 18:06:40,436][<ipython-input-44-AA1F3DE2C698>][L163][INFO] Epoch[0-9/200] | Train Loss: 11.088 | Train Accuracy: 28.608% | Test Loss: 10.944 | Test Accuracy: 49.113% | v
[2024-02-22 18:06:46,417][<ipython-input-44-AA1F3DE2C698>][L163][INFO] Epoch[10-19/200] | Train Loss: 11.081 | Train Accuracy: 28.636% | Test Loss: 10.892 | Test Accuracy: 49.951% | v
[2024-02-22 18:06:52,130][<ipython-input-44-AA1F3DE2C698>][L163][INFO] Epoch[20-29/200] | Train Loss: 11.004 | Train Accuracy: 29.673% | Test Loss: 10.644 | Test Accuracy: 53.498% | v
[2024-02-22 18:06:58,010][<ipython-input-44-AA1F3DE2C698>][L163][INFO] Epoch[30-39/200] | Train Loss: 10.322 | Train Accuracy: 35.994% | Test Loss: 10.544 | Test Accuracy: 67.094% | v
[2024-02-22 18:07:03,906][<ipython-input-44-AA1F3DE2C698>][L163][INFO] Epoch[40-49/200] | Train Loss: 9.462 | Train Accuracy: 40.909% | Test Loss: 9.466 | Test Accuracy: 71.527% | v
[2024-02-22 18:07:09,727][<ipython-input-44-AA1F3DE2C698>][L163][INFO] Epoch[50-59/200] | Train Loss: 8.747 | Train Accuracy: 43.068% | Test Loss: 9.976 | Test Accuracy: 70.443% | v
[2024-02-22 18:07:15,696][<ipython-input-44-AA1F3DE2C698>][L163][INFO] Epoch[60-69/200] | Train Loss: 8.183 | Train Accuracy: 46.463% | Test Loss: 9.875 | Test Accuracy: 70.296% | v
[2024-02-22 18:07:21,492][<ipython-input-44-AA1F3DE2C698>][L163][INFO] Epoch[70-79/200] | Train Loss: 7.752 | Train Accuracy: 47.670% | Test Loss: 9.926 | Test Accuracy: 71.527% | v
[2024-02-22 18:07:27,585][<ipython-input-44-AA1F3DE2C698>][L163][INFO] Epoch[80-89/200] | Train Loss: 7.380 | Train Accuracy: 48.324% | Test Loss: 10.300 | Test Accuracy: 71.724% | v
[2024-02-22 18:07:33,409][<ipython-input-44-AA1F3DE2C698>][L163][INFO] Epoch[90-99/200] | Train Loss: 6.968 | Train Accuracy: 49.276% | Test Loss: 10.579 | Test Accuracy: 71.379% | v
[2024-02-22 18:07:39,472][<ipython-input-44-AA1F3DE2C698>][L163][INFO] Epoch[100-109/200] | Train Loss: 6.810 | Train Accuracy: 49.517% | Test Loss: 10.705 | Test Accuracy: 71.429% | v
[2024-02-22 18:07:45,319][<ipython-input-44-AA1F3DE2C698>][L163][INFO] Epoch[110-119/200] | Train Loss: 6.493 | Train Accuracy: 50.824% | Test Loss: 10.996 | Test Accuracy: 71.724% | v
[2024-02-22 18:07:51,445][<ipython-input-44-AA1F3DE2C698>][L163][INFO] Epoch[120-129/200] | Train Loss: 6.375 | Train Accuracy: 50.710% | Test Loss: 11.151 | Test Accuracy: 71.133% | v
[2024-02-22 18:07:57,318][<ipython-input-44-AA1F3DE2C698>][L163][INFO] Epoch[130-139/200] | Train Loss: 6.221 | Train Accuracy: 51.335% | Test Loss: 11.011 | Test Accuracy: 70.542% | v
[2024-02-22 18:08:04,061][<ipython-input-44-AA1F3DE2C698>][L163][INFO] Epoch[140-149/200] | Train Loss: 6.124 | Train Accuracy: 52.074% | Test Loss: 11.084 | Test Accuracy: 70.591% | v
[2024-02-22 18:08:09,935][<ipython-input-44-AA1F3DE2C698>][L163][INFO] Epoch[150-159/200] | Train Loss: 5.919 | Train Accuracy: 52.841% | Test Loss: 10.956 | Test Accuracy: 71.338% | v
[2024-02-22 18:08:16,063][<ipython-input-44-AA1F3DE2C698>][L163][INFO] Epoch[160-169/200] | Train Loss: 5.945 | Train Accuracy: 50.966% | Test Loss: 10.974 | Test Accuracy: 71.429% | v
[2024-02-22 18:08:22,027][<ipython-input-44-AA1F3DE2C698>][L163][INFO] Epoch[170-179/200] | Train Loss: 5.847 | Train Accuracy: 52.173% | Test Loss: 11.145 | Test Accuracy: 71.626% | v
[2024-02-22 18:08:28,194][<ipython-input-44-AA1F3DE2C698>][L163][INFO] Epoch[180-189/200] | Train Loss: 5.667 | Train Accuracy: 52.827% | Test Loss: 11.230 | Test Accuracy: 71.330% | v
[2024-02-22 18:08:34,133][<ipython-input-44-AA1F3DE2C698>][L163][INFO] Epoch[190-199/200] | Train Loss: 5.873 | Train Accuracy: 51.193% | Test Loss: 11.116 | Test Accuracy: 70.788% | v
Total time taken to train: 123.18678450584412ms

```





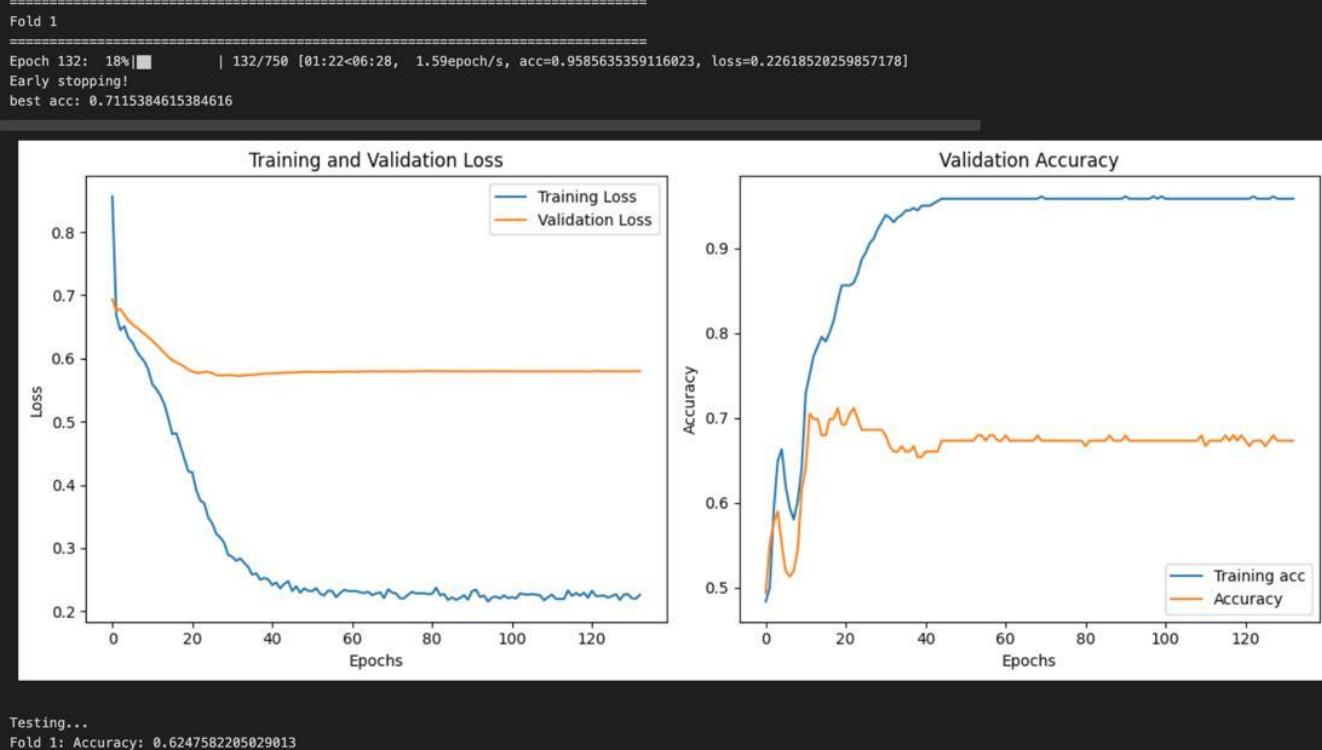
METAFORMER

	Fold	Accuracy	Precision	Recall	F1	AUC	AP	\
0	0	0.644788	0.632787	0.728302	0.677193	0.642807	0.599856	
1	1	0.624758	0.619529	0.694340	0.654804	0.622963	0.586836	

	FPR	FNR	TPR	TNR
0	0.442688	0.271698	0.728302	0.557312
1	0.448413	0.305660	0.694340	0.551587

Mean accuracy: 0.6347729326452731

Std accuracy: 0.010014712142371762



```

Testing...
Fold 1: Accuracy: 0.6247582205029013
Fold 1: Confusion matrix:
[[139 113]
 [ 81 184]]

```

PROPOSED MODEL

Fold	Accuracy	Precision	Recall	F1	AUC	AP	\
0	0 0.596525	0.595890	0.656604	0.624776	0.595100	0.566940	
1	1 0.638298	0.640288	0.671698	0.655617	0.637436	0.598359	
		FPR	FNR	TPR	TNR		
0	0.466403	0.343396	0.656604	0.533597			
1	0.396825	0.328302	0.671698	0.603175			

```

Mean accuracy: 0.6174114844327611
Std accuracy: 0.020886387907664505
=====
```

```

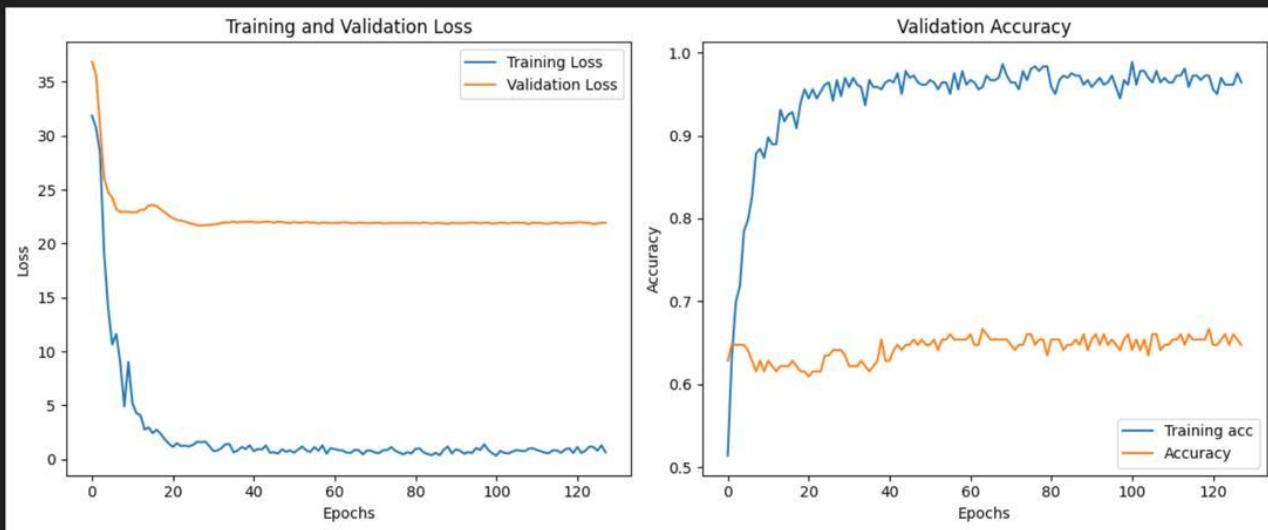
Fnet model:
SAT(
  (encoder): EncoderBlock(
    (inp_emb): Linear(in_features=19900, out_features=256, bias=True)
    (encoder): FNet(
      (layers): ModuleList(
        (0-1): 2 x ModuleList(
          (0): PreNorm(
            (norm): LayerNorm((256,), eps=1e-05, elementwise_affine=True)
            (fn): FNetBlock()
          )
          (1): PreNorm(
            (norm): LayerNorm((256,), eps=1e-05, elementwise_affine=True)
            (fn): FeedForward(
              (linear1): Linear(in_features=256, out_features=4, bias=True)
              (dropout): Dropout(p=0.0, inplace=False)
              (linear2): Linear(in_features=4, out_features=256, bias=True)
              (dropout1): Dropout(p=0.0, inplace=False)
              (dropout2): Dropout(p=0.0, inplace=False)
              (norm1): LayerNorm((256,), eps=1e-05, elementwise_affine=True)
              (norm2): LayerNorm((256,), eps=1e-05, elementwise_affine=True)
            )
          )
        )
      )
    )
  )
  (act): GELU(approximate='none')
  (do): Dropout(p=0.0, inplace=False)
  (head): Linear(in_features=256, out_features=2, bias=True)
)

```

```

Fold 1
=====
Pretraining finished...
Directory 'checkpoints/fnet_train_1' created successfully
No checkpoint files found.
Epoch 127: 25%|██████████| 127/500 [01:16<03:43, 1.67epoch/s, acc=0.9640883977900553, loss=0.6442072242498398]
Early stopping!
best acc: 0.6666666666666666

```



```

Testing...
Fold 1: Accuracy: 0.6382978723404256
Fold 1: Confusion matrix:
[[152 100]
 [ 87 178]]

```

PARAMETER COMPARISON

Metaformer model

Number of parameters : 5628170

Total number of trainable parameters in model: 5628170

FNet model

Number of parameters : 5103882

Total number of trainable parameters in model: 5103882

MODEL	ACCURACY	PRECISION	RECALL	F1 SCORE	AUC	TRAIN ACCURACY
BNT	.709	.71	.69	.71	.76	.5383
Proposed Model I	.73	.73	.73	.72	.82	.5406
METAFormer	.6247	.6195	.6943	.6548	.6229	.975
Proposed Model II	.6174	.618	.6641	.6162	.6162	.97

Status of paper publication

Exploring Machine Learning Techniques for Autism Spectrum Disorder Detection:A survey

To be communicated to the 2024 International Conference on Information Technology, Electronics and Intelligent Communication Systems (ICITEICS)

CONCLUSION

- In conclusion, the model aims to reduce complexity by incorporating a fourier transform layer by replacing the multihead attention layer of the transformer encoder.
- It aims to provide a more objective and quantitative approach moving beyond the existing subjective behavioural assessments , paving way for a generalizable model that can cater the needs of diverse populations.

REFERENCES

- [1] Osolo, R.I., Yang, Z. and Long, J., 2021. An attentive fourier-augmented image-captioning transformer. *Applied Sciences*, 11(18), p.8354.
- [2] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł. and Polosukhin, I., 2017. Attention is all you need. *Advances in neural information processing systems*, 30.
- [3] Lee-Thorp, J., Ainslie, J., Eckstein, I. and Ontanon, S., 2021. Fnet: Mixing tokens with fourier transforms. *arXiv preprint arXiv:2105.03824*.
- [4] Kan, X., Dai, W., Cui, H., Zhang, Z., Guo, Y. and Yang, C., 2022. Brain network transformer. *Advances in Neural Information Processing Systems*, 35, pp.25586-25599.
- [5] Rakić, M., Cabezas, M., Kushibar, K., Oliver, A. and Llado, X., 2020. Improving the detection of autism spectrum disorder by combining structural and functional MRI information. *NeuroImage: Clinical*, 25, p.102181.

REFERENCES

- [6] Manikantan, K. and Jaganathan, S., 2023. A Model for Diagnosing Autism Patients Using Spatial and Statistical Measures Using rs-fMRI and sMRI by Adopting Graphical Neural Networks. *Diagnostics*, 13(6), p.1143.
- [7] Bannadabhai, A., Lee, S., Deng, W., Ying, R. and Li, X., 2023, October. Community-Aware Transformer for Autism Prediction in fMRI Connectome. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 287-297). Cham: Springer Nature Switzerland.
- [8] Sharif, H. and Khan, R.A., 2022. A novel machine learning based framework for detection of autism spectrum disorder (ASD). *Applied Artificial Intelligence*, 36(1), p.2004655.
- [9] Haghhighat, H., Mirzarezaee, M., Araabi, B.N. and Khadem, A., 2022. An age-dependent Connectivity-based computer aided diagnosis system for autism spectrum disorder using resting-state fMRI. *Biomedical Signal Processing and Control*, 71, p.103108.
- [10] Yin, W., Li, L. and Wu, F.X., 2022. A semi-supervised autoencoder for autism disease diagnosis. *Neurocomputing*, 483, pp.140-147.

Thank you..!!

Appendix B: Vision, Mission, Programme Outcomes and Course Outcomes

Vision, Mission, Programme Outcomes and Course Outcomes

Institute Vision

To evolve into a premier technological institution, moulding eminent professionals with creative minds, innovative ideas and sound practical skill, and to shape a future where technology works for the enrichment of mankind.

Institute Mission

To impart state-of-the-art knowledge to individuals in various technological disciplines and to inculcate in them a high degree of social consciousness and human values, thereby enabling them to face the challenges of life with courage and conviction.

Department Vision

To become a centre of excellence in Computer Science and Engineering, moulding professionals catering to the research and professional needs of national and international organizations.

Department Mission

To inspire and nurture students, with up-to-date knowledge in Computer Science and Engineering, ethics, team spirit, leadership abilities, innovation and creativity to come out with solutions meeting societal needs.

Programme Outcomes (PO)

Engineering Graduates will be able to:

1. Engineering Knowledge: Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.

2. Problem analysis: Identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.

- 3. Design/development of solutions:** Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.
- 4. Conduct investigations of complex problems:** Use research-based knowledge including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.
- 5. Modern Tool Usage:** Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations.
- 6. The engineer and society:** Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal, and cultural issues and the consequent responsibilities relevant to the professional engineering practice.
- 7. Environment and sustainability:** Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.
- 8. Ethics:** Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.
- 9. Individual and Team work:** Function effectively as an individual, and as a member or leader in teams, and in multidisciplinary settings.
- 10. Communication:** Communicate effectively with the engineering community and with society at large. Be able to comprehend and write effective reports documentation. Make effective presentations, and give and receive clear instructions.
- 11. Project management and finance:** Demonstrate knowledge and understanding of engineering and management principles and apply these to one's own work, as a member and leader in a team. Manage projects in multidisciplinary environments.
- 12. Life-long learning:** Recognize the need for, and have the preparation and ability to engage in independent and lifelong learning in the broadest context of technological change.

Programme Specific Outcomes (PSO)

A graduate of the Computer Science and Engineering Program will demonstrate:

PSO1: Computer Science Specific Skills

The ability to identify, analyze and design solutions for complex engineering problems in multidisciplinary areas by understanding the core principles and concepts of computer science and thereby engage in national grand challenges.

PSO2: Programming and Software Development Skills

The ability to acquire programming efficiency by designing algorithms and applying standard practices in software project development to deliver quality software products meeting the demands of the industry.

PSO3: Professional Skills

The ability to apply the fundamentals of computer science in competitive research and to develop innovative products to meet the societal needs thereby evolving as an eminent researcher and entrepreneur.

Course Outcomes (CO)

Course Outcome 1: Model and solve real world problems by applying knowledge across domains (Cognitive knowledge level: Apply).

Course Outcome 2: Develop products, processes or technologies for sustainable and socially relevant applications (Cognitive knowledge level: Apply).

Course Outcome 3: Function effectively as an individual and as a leader in diverse teams and to comprehend and execute designated tasks (Cognitive knowledge level: Apply).

Course Outcome 4: Plan and execute tasks utilizing available resources within timelines, following ethical and professional norms (Cognitive knowledge level: Apply).

Course Outcome 5: Identify technology/research gaps and propose innovative/creative solutions (Cognitive knowledge level: Analyze).

Course Outcome 6: Organize and communicate technical and scientific findings effectively in written and oral forms (Cognitive knowledge level: Apply).

Appendix C: CO-PO-PSO Mapping

CO-PO AND CO-PSO MAPPING

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO 10	PO 11	PO 12	PSO 1	PSO 2	PSO 3
CO 1	2	2	2	1	2	2	2	1	1	1	1	2	3		
CO 2	2	2	2		1	3	3	1	1		1	1		2	
CO 3									3	2	2	1			3
CO 4					2				3	2	2	3	2		3
CO 5	2	3	3	1	2								1	3	
CO 6					2				2	2	3	1	1		3

3/2/1: high/medium/low

JUSTIFICATIONS FOR CO-PO MAPPING

MAPPING	LOW/MEDIUM/ HIGH	JUSTIFICATION
100003/ CS722U.1- PO1	M	Knowledge in the area of technology for project development using various tools results in better modeling.
100003/ CS722U.1- PO2	M	Knowledge acquired in the selected area of project development can be used to identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions.

100003/ CS722U.1- PO3	M	Can use the acquired knowledge in designing solutions to complex problems.
100003/ CS722U.1- PO4	M	Can use the acquired knowledge in designing solutions to complex problems.
100003/ CS722U.1- PO5	H	Students are able to interpret, improve and redefine technical aspects for design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.
100003/ CS722U.1- PO6	M	Students are able to interpret, improve and redefine technical aspects by applying contextual knowledge to assess societal, health and consequential responsibilities relevant to professional engineering practices.
100003/ CS722U.1- PO7	M	Project development based on societal and environmental context solution identification is the need for sustainable development.
100003/ CS722U.1- PO8	L	Project development should be based on professional ethics and responsibilities.
100003/ CS722U.1- PO9	L	Project development using a systematic approach based on well defined principles will result in teamwork.
100003/ CS722U.1- PO10	M	Project brings technological changes in society.

100003/ CS722U.1- PO11	H	Acquiring knowledge for project development gathers skills in design, analysis, development and implementation of algorithms.
100003/ CS722U.1- PO12	H	Knowledge for project development contributes engineering skills in computing & information gatherings.
100003/ CS722U.2- PO1	H	Knowledge acquired for project development will also include systematic planning, developing, testing and implementation in computer science solutions in various domains.
100003/ CS722U.2- PO2	H	Project design and development using a systematic approach brings knowledge in mathematics and engineering fundamentals.
100003/ CS722U.2- PO3	H	Identifying, formulating and analyzing the project results in a systematic approach.
100003/ CS722U.2- PO5	H	Systematic approach is the tip for solving complex problems in various domains.
100003/ CS722U.2- PO6	H	Systematic approach in the technical and design aspects provide valid conclusions.
100003/ CS722U.2- PO7	H	Systematic approach in the technical and design aspects demonstrate the knowledge of sustainable development.

100003/ CS722U.2- PO8	M	Identification and justification of technical aspects of project development demonstrates the need for sustainable development.
100003/ CS722U.2- PO9	H	Apply professional ethics and responsibilities in engineering practice of development.
100003/ CS722U.2- PO11	H	Systematic approach also includes effective reporting and documentation which gives clear instructions.
100003/ CS722U.2- PO12	M	Project development using a systematic approach based on well defined principles will result in better teamwork.
100003/ CS722U.3- PO9	H	Project development as a team brings the ability to engage in independent and lifelong learning.
100003/ CS722U.3- PO10	H	Identification, formulation and justification in technical aspects will be based on acquiring skills in design and development of algorithms.
100003/ CS722U.3- PO11	H	Identification, formulation and justification in technical aspects provides the betterment of life in various domains.
100003/ CS722U.3- PO12	H	Students are able to interpret, improve and redefine technical aspects with mathematics, science and engineering fundamentals for the solutions of complex problems.

100003/ CS722U.4- PO5	H	Students are able to interpret, improve and redefine technical aspects with identification formulation and analysis of complex problems.
100003/ CS722U.4- PO8	H	Students are able to interpret, improve and redefine technical aspects to meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.
100003/ CS722U.4- PO9	H	Students are able to interpret, improve and redefine technical aspects for design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.
100003/ CS722U.4- PO10	H	Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools for better products.
100003/ CS722U.4- PO11	M	Students are able to interpret, improve and redefine technical aspects by applying contextual knowledge to assess societal, health and consequential responsibilities relevant to professional engineering practices.
100003/ CS722U.4- PO12	H	Students are able to interpret, improve and redefine technical aspects for demonstrating the knowledge of, and need for sustainable development.
100003/ CS722U.5- PO1	H	Students are able to interpret, improve and redefine technical aspects, apply ethical principles and commit to

		professional ethics and responsibilities and norms of the engineering practice.
100003/ CS722U.5- PO2	M	Students are able to interpret, improve and redefine technical aspects, communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.
100003/ CS722U.5- PO3	H	Students are able to interpret, improve and redefine technical aspects to demonstrate knowledge and understanding of the engineering and management principle in multidisciplinary environments.
100003/ CS722U.5- PO4	H	Students are able to interpret, improve and redefine technical aspects, recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.
100003/ CS722U.5- PO5	M	Students are able to interpret, improve and redefine technical aspects in acquiring skills to design, analyze and develop algorithms and implement those using high-level programming languages.
100003/ CS722U.5- PO12	M	Students are able to interpret, improve and redefine technical aspects and contribute their engineering skills in computing and information engineering domains like network design and administration, database design and

		knowledge engineering.
100003/ CS722U.6- PO5	M	Students are able to interpret, improve and redefine technical aspects and develop strong skills in systematic planning, developing, testing, implementing and providing IT solutions for different domains which helps in the betterment of life.
100003/ CS722U.6- PO8	H	Students will be able to associate with a team as an effective team player for the development of technical projects by applying the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.
100003/ CS722U.6- PO9	H	Students will be able to associate with a team as an effective team player to Identify, formulate, review research literature, and analyze complex engineering problems
100003/ CS722U.6- PO10	M	Students will be able to associate with a team as an effective team player for designing solutions to complex engineering problems and design system components.
100003/ CS722U.6- PO11	M	Students will be able to associate with a team as an effective team player, use research-based knowledge and research methods including design of experiments, analysis and interpretation of data.
100003/ CS722U.6- PO12	H	Students will be able to associate with a team as an effective team player, applying ethical principles and

		commit to professional ethics and responsibilities and norms of the engineering practice.
100003/ CS722U.1- PSO1	H	Students are able to develop Computer Science Specific Skills by modeling and solving problems.
100003/ CS722U.2- PSO2	M	Developing products, processes or technologies for sustainable and socially relevant applications can promote Programming and Software Development Skills.
100003/ CS722U.3- PSO3	H	Working in a team can result in the effective development of Professional Skills.
100003/ CS722U.4- PSO3	H	Planning and scheduling can result in the effective development of Professional Skills.
100003/ CS722U.5- PSO1	H	Students are able to develop Computer Science Specific Skills by creating innovative solutions to problems.
100003/ CS722U.6- PSO3	H	Organizing and communicating technical and scientific findings can help in the effective development of Professional Skills.