
**UNIVERSITY OF PADOVA**

Department of General Psychology

**Master's degree in Cognitive Neuroscience & Clinical Neuropsychology**

*Final dissertation*

# Depression Classification from rs-fMRI: A Local-to-Global Graph Neural Network Approach

Supervisor: Dr. Alberto Testolin

Co-supervisor: Dr. Francesco Prinzi

Candidate: Leoni-Stavroula Christakou

Student ID: 2037908

October 2024

***Depression Classification from rs-fMRI: A Local-to-Global Graph Neural Network Approach,*** October 2024

Author:

Leoni-Stavroula Christakou [1]

Supervisor:

Dr. Alberto Testolin [1]

Co-supervisor:

Dr. Francesco Prinzi [2]

Instututes:

1 Universita Degli Studi di Padova

2 Universita Degli Studi di Palermo | Dipartimento di Biomedicina Neuroscienze e Diagnostica Avanzata

# ACKNOWL EDGMENTS

I am profoundly grateful to Francesco Prinzi and Professor Vitabile for their acceptance, encouragement, and unwavering support which were instrumental to my progress and personal growth. I extend my deepest appreciation to professor Testolin whose understanding was pivotal throughout my research journey.

A special thanks to Dimitri, who has been by my side throughout this journey, never faltering in his support. I am equally thankful to Valeria, whose presence and love have been a constant source of strength.

I must express my gratitude to my uncle Ted, for his kindness and support, and to Elias, without whom I could not have accomplished what I did.

Most importantly, I am forever indebted to my loving parents, Eleni and Theodosis. They have given me everything, expecting nothing in return, accepted and supported me along with my successes and failures, and provided the freedom to make my own decisions, allowing me to flourish. Their unconditional love and support have shaped the person I am today. Σας αγαπώ πολύ !

# TABLE OF CONTENTS

# FIGURES & TABLES

## FIGURES



## TABLES

# LIST OF ABBRAVATIONS

| Abbreviation | Definition |
|---|---|
| MDD | Major Depressive Disorder |
| Rs-fMRI | Resting-state functional Magnetic Resonance Imaging |
| ROI | Region of Interest |
| LG-GNN | Local-to-Global Graph Neural Network |
| GNN | Graph Neural Network |
| GCN | Graph Convolutional Network |
| GAT | Graph Attention Network |
| GIN | Graph Isomorphism Network |
| SABP | Self-Attention Based Pooling |
| fMRI | Functional Magnetic Resonance Imaging |
| RSN | Resting State Networks |
| BOLD | blood-oxygen-level-dependent |
| DMN | Default Mode Network |
| SN | Salience Network |
| CEN | Central Executive Network |
| ML | Machine Learning |
| SVM | Support Vector Machine |
| DL | Deep Learning |
| ANN | Artificial Neural Network |
| CNN | Convolutional neural network |
| AD | Alzheimer's disease |
| HC | Healthy controls |
| ASD | Autism Spectrum Disorder |
| HAMD | Hamilton Depression Rating Scale |
| HAMA | Hamilton Anxiety Rating Scale |
| MNI | Montreal Neurological Institute |
| GC | Graph Convolution layer |
| AWAB | Adaptive Weight Aggregation Block |
| ReLU | Rectified Linear Unit |

# ABSTRACT

Graph Convolutional Networks (GCNs) have revolutionized the analysis of complex diseases by mapping the intricate brain connectivity data inherent in brain disorders. Building upon these foundations, this research explores the novel application of the Local-to-Global Graph Neural Network (LG-GNN) framework to Major Depressive Disorder (MDD), leveraging resting-state functional magnetic resonance imaging (rs-fMRI) data. This adaptation refines a proven model to uncover both localized and systemic disruptions characteristic of MDD. The approach integrates a local region of interest (ROI) based GCN, which identifies critical biomarkers and extracts pivotal features from distinct brain regions, with a global GNN that synthesizes these data points with supplementary non-imaging information to articulate a comprehensive view of network pathologies. This refined application enhances MDD classification by capitalizing on both the relational and attribute-based information of brain networks. By extending the use of sophisticated feature learning and comprehensive network analysis to MDD, this research not only broadens the model's applicative spectrum but also significantly enhances the classification accuracy of MDD. This improvement underscores the model's capability to delineate depression's underlying mechanisms more clearly, potentially facilitating earlier and more precise diagnostic and therapeutic intervention.

# 1 INTRODUCTION

Major depressive disorder (MDD) is considered a priority public health concern, affecting over 264 million individuals in the world and making huge contributions to a global disease burden. MDD is a complex mental health disorder that results in excessive rates of chronic sadness, reduced interest in everyday activities, elements of fatigue, deficits in cognition, as well as a variety of somatic symptoms that impair social and vocational functioning (American Psychiatric Association, 2013). Moreover, the presence of MDD is significantly associated with increasing morbidity and suicide risk, which seriously underlines the need for timely and accurate diagnosis (Orsolini et al., 2020).

The diagnosis of MDD depends exceedingly on symptomatology based on DSM-5 criteria, where at least five of the following must have continuously occurred during a period of at least two weeks: depressed mood, loss of interest, changes in appetite or sleep, fatigue, and feelings of worthlessness, among others (American Psychiatric Association, 2013). Nevertheless, despite the well-set diagnostic criteria, the diagnosis of MDD has remained exceptionally elusive. The great heterogeneity in symptom presentation between individuals impresses complexity on the diagnostic process, whereby patients may present symptom constellations that vary over the course of a day both quantitatively and qualitatively (Cusack et al., 2024).

Comorbid disorders can also mask symptoms of depression or accentuate them. For example, anxiety and psychotic disorders (Thaipisuttikul et al., 2014) may present with similar symptoms or, or substance use disorders (Calarco & Lobo, 2020). may complicate the symptomatology and, thus, make correct recognition by a clinician of the main disorder not an easy task. Such subjectiveness in symptom reporting and clinical interview is increasing variability even more and one of the causes for delayed or missed diagnoses.

With these limitations, the need for more objective, biologically-based diagnostic tools that would help bring about more reliable and consistent assessment of MDD has increased. Neuroimaging technologies, such as functional Magnetic Resonance Imaging (fMRI), have recently emerged as promising methods that try to explain the

neurobiological underpinnings of MDD and may provide important knowledge that can be useful in diagnosis.

Neuroimaging techniques hold much promise in unravelling the neurobiological substrates underlying the processes of MDD. Rs-fMRI provides a way to investigate large-scale brain networks which are abnormally connected in MDD patients. However, the nature of brain data is often high-dimensional and complex, rendering most of the traditional methods of analysis difficult to apply. This is the complexity that has grown interest in the advanced machine learning (ML) and deep learning (DL) techniques for the classification of psychiatric disorders. GNNs represent one of the state-of-the-art approaches that find wide acceptance because they can handle non-Euclidean structure and hence fit for brain networks. GNN models, in a very efficient way, complex interactions of various brain networks and have, therefore, grown as an important tool to test the connectivity patterns associated with MDD.

While GNNs promise much, there are still issues to be fully improved, such as model interpretability, ensuring performance across diverse datasets, and integrating more types of data for a better understanding of MDD neurobiology. Though some works have been done in this respect, few studies have explored multi-scale models which can simultaneously capture both local and global patterns of the brain; thus, ample scope for further research and improvement exists.

To address these challenges, in this work, we enhanced the LG-GNN framework for classifying MDD from rs-fMRI data. The contributions in this article are threefold, summarized below.

- Application to MDD Classification: For the first time, the LG-GNN framework that integrates both the local information of each brain region and the global relationship between subjects was applied to the MDD classification task.

- Variants of Models: We have tried different GNN layer architectures for both the Local-GNN and the Global-GNN. Within those different settings, we investigated which one yields the best capability in improving the model's ability in learning local and global patterns associated with MDD.

- Fine Tuning of Local-GNN: We fine-tuned the self-attention-based pooling module within the Local-GNN to retain the most important features about brain regions.

This tuning is done in such a way that the model performance of classification gets improved through focusing on those features which are most relevant in predicting MDD.

- Biomarker Extraction: Employing the use of a Local-GNN helped in identifying some potential biomarkers indicated by features in certain regions of the brain which were highly associated with MDD. Such findings enhanced the understanding of the disorder's local abnormalities within the brain.

- Non-imaging data incorporation: We integrated the non-imaging data into the Global-GNN. This further enhanced the model capability of capturing relationships among subjects, hence enhancing the classification accuracy. Specifically, GNN studies using such non-imaging data for depression classification remain sparse, and the current study has represented an essential contribution in light of integrating more comprehensive subject-level information into the predictive model.

The performance achieved was very good on most metrics and far outperforms many other studies on great performances in achieving state-of-the-art results. This clearly illustrates the strength of the LG-GNN framework in the modelling of both local and global brain network patterns for the two classes in MDD classification and therefore setting a new benchmark in the field of research.

# 2 THEORETICAL BACKGROUND

## 2.1 ADVANCES IN NEUROIMAGING: FUNCTIONAL MRI

A groundbreaking tool in investigations on Neuroscience, fMRI provides a unique insight into the in vivo activity of the brain. Under both task-oriented and resting-state situations, fMRI to capture spatial and temporal dynamics, permitting researchers to investigate functional connections between several regions of the brain.

### 2.1.1 Overview of fMRI

fMRI measures brain activity by detecting changes in blood flow and oxygenation levels, which occur in response to neuronal activation. Unlike structural MRI, which focuses on the brain's anatomy, fMRI identifies variations in neuronal activity that reflects changes in the concentration of oxygenated and deoxygenated blood, thus focusing on functional connectivity between several brain regions (Ogawa et al., 1990).

The most important advantages of fMRI are non-invasiveness, increased availability, quite high spatiotemporal resolution, and the potential to display the whole network of brain areas active (Logothetis, 2008). It has been particularly useful in psychiatric research because it allows for the investigation of brain network dysfunctions that may be related to mood, cognition, and emotional regulation.

#### *2.1.1.1 Signal Acquisition*

A very strong magnetic field is applied in fMRI to align the hydrogen protons in water molecules throughout the brain. The MRI scanner briefly perturbs this alignment by broadcasting radiofrequency pulses that stimulate the protons to spin synchronously with each other. As protons return to their usual positions, they emit energy, and the scanner detects this energy as an MRI signal. In more general terms, an MR image represents a snapshot of the transverse magnetization at some Time to Echo following the Radiofrequency Pulse which initially created the transverse magnetization (Buxton, 2013).

The scanner picks up a time series of 3-D images in which the change in the BOLD signal is measured over time to provide a representation of the localized activity of the

brain via voxels. Voxels are spatial units that capture small regions of neuronal activation.

In rs-fMRI, participants are asked to lie still without performing any cognitive or motor tasks. The aim is to capture the brain's intrinsic functional connectivity. In this state, low-frequency oscillations in the BOLD signal reflect spontaneous fluctuations in brain activity (Biswal et al., 1995).

### 2.1.2 The BOLD Signal

The BOLD signal is an indirect measure of neural activity, capturing metabolic changes that follow neuronal firing (Bandettini, 2012). It is based on the principle that brain activity increases oxygen consumption in localized areas of the brain, prompting an influx of oxygenated blood. This inflow, known as neurovascular coupling, temporarily increases the concentration of oxygenated haemoglobin relative to deoxygenated haemoglobin (Ogawa et al., 1990). Oxygenated haemoglobin has different magnetic properties than deoxygenated haemoglobin and it produces local changes in the magnetic field. These changes can be measured by the scanner and can further be converted into the units of activity of the brain (Bandettini, 2012). The overall pathway is seen in Figure 1.

In rs-fMRI, the BOLD signal reflects spontaneous fluctuations in brain activity while subjects are not engaged in any explicit task. These fluctuations form the basis of functional connectivity measurements, which quantify the degree of synchrony between different brain regions over time. By capturing this baseline activity resting-state networks (RSNs) that are active even in the absence of tasks can be identified (Fox & Raichle, 2007).

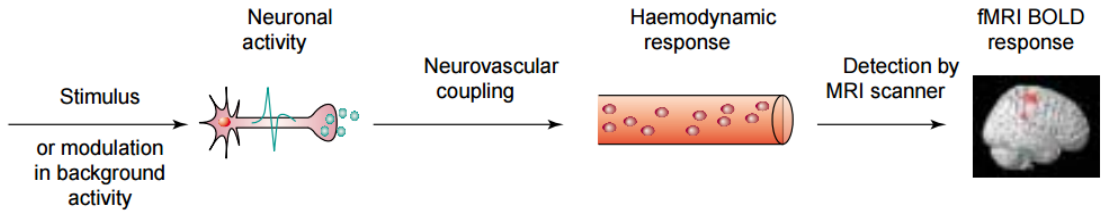


***Figure 1*** *The pathway from neural activity to the fMRI activity map. Reproduced from Arthurs and Boniface (2002)*

### 2.1.3 Resting-State Networks

The discovery of RSNs in the mid-1990s significantly advanced our understanding of intrinsic brain activity. In a seminal study, Biswal et al. (1995) demonstrated that low-frequency fluctuations in the BOLD signal during rest show significant correlations between spatially distant brain regions, notably within the motor cortex. This finding revealed that the brain remains highly active even in the absence of tasks, with synchronized activity between functionally related regions (Biswal et al., 1995). This intrinsic connectivity approach is essential for understanding how large-scale brain networks function in psychiatric disorders.

Several major RSNs have since been identified, each playing a role in maintaining the brain's functional architecture during rest (Figure 2):

- Default Mode Network (DMN): The DMN is active during periods of self-directed thought or introspection Anatomically, the DMN includes the anterior medial prefrontal cortex, posterior cingulate cortex, and angular gyrus (Gillespie et al., 2020). Altered activity in the DMN, particularly hyperconnectivity, has been associated with rumination and excessive internal focus seen in depression (Sheline et al., 2009).
- Salience Network (SN): The anterior cingulate and ventral anterior insular cortices are the cortical hubs of an ensemble of brain areas known as SN. The amygdala, hypothalamus, ventral striatum, thalamus, and particular brainstem nuclei are also part of this network (Seeley, 2019). It plays a role in a wide range of cognitive functions, including executive control and salience filtering, as well as the interoception of reward-related emotions (Pace-Schott & Picchioni, 2016). It also helps sustain alertness and arousal (Fasiello et al., 2021). Emotional dysregulation in depression has been associated with disturbances in the SN (Menon, 2011).
- Central Executive Network (CEN): Primary components of the CEN include the anterior inferior parietal lobule, the rostral lateral and dorsolateral prefrontal cortex (particularly the middle frontal gyrus). Other areas include the head of the caudate nucleus, posterior inferior temporal lobe, dorsomedial thalamus, middle cingulate gyrus, and possibly the dorsal precuneus (Uddin et al., 2019). It plays a role in executive function and cognitively taxing, goal-oriented tasks.

It is essential for actively preserving and modifying information in working memory, solving problems based on rules, and making decisions within the framework of goal-directed behavior (Menon, 2011). According to DeMaster et al. (2022), this network exhibits aberrant activity in MDD, which contributes to cognitive processing issues.

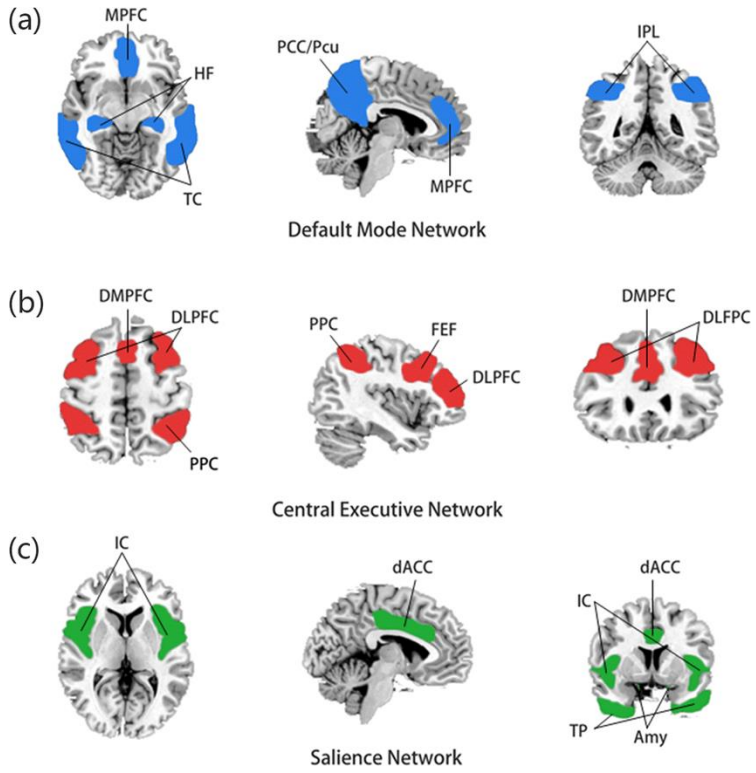


***Figure 2*** *Components of RSs. Reproduced from Dai et al. (2019)*

## 2.2 Predictive Modelling

### 2.2.1 Machine Learning

ML has now become an important tool in neuroimaging analysis, considering the big and complex data output this technique produces. ML algorithms-facilitated especially through supervised learning-are immune to detecting subtle nonlinear patterns unattainable by conventional techniques; thus, they offer a potent approach which can tell effectively whether one is a sufferer of MDD or not. More than a tool for handling large data sets, ML has implications for a paradigm shift in psychiatry itself (Breiman, 2001)-from a focus on hypothesis testing to one of prediction and from fixed to flexible

handling of multiple variables simultaneously (Chekroud et al., 2021). Such a shift enables the identification of more complex relationships between brain activity and clinical symptoms and could thereby offer deeper insights into the underlying mechanisms of psychiatric disorders.

#### *2.2.1.1 Fundamentals of ML*

ML, a term introduced by Arthur Samuel in 1959, refers to the study that enables computers to learn autonomously without being explicitly programmed (Samuel, 1969). It emerged from the fields of computational learning theory and pattern recognition, becoming a key method in data analytics for making predictions by applying models and algorithms (Angra & Ahuja, 2017).

The learning process in a simple ML model is typically divided into two phases: training and testing. During the training phase, samples from the training data are used as input, where a learning algorithm, or learner, extracts features and builds the predictive model (Sandhya & Charanjeet, 2016). In the testing phase, the trained model utilizes an execution engine to make predictions on test or production data, ultimately producing data that represents the final prediction (Nasteski, 2017).

These ML techniques automatically construct computational models of complex relationships by processing data to optimize problem-specific performance criteria. This training process adjusts the model's parameters, revealing how input variables relate to outputs and enabling the model to predict outcomes for new, unseen data (Baştanlar & Özuysal, 2013). Conceptually, this can be viewed as a learning problem, where ML algorithms explore a vast space of possible solutions, guided by training experience, to identify the model that best enhances a specific performance measure (Jordan & Mitchell, 2015). The goal of a ML algorithm is to produce accurate predictions for data points not included in the training set. The model's capacity to generate correct outputs for new, unseen samples after being trained is known as generalization (Baştanlar & Özuysal, 2013).

#### *2.2.1.2 Supervised learning*

Supervised learning is a ML task where a function maps input data to corresponding output values based on example input-output pairs. This approach requires each training sample to include both input and output values, which are used by the algorithm to train a model. If the output variables are continuous, the model is referred to as a "regression function," such as in predicting air temperature. When the output variables are discrete, the model is called a "classifier" (Baştanlar & Özuysal, 2013).

In supervised learning, the input dataset is divided into a training set and a testing set. The pipeline is seen in Figure 3. The training set includes labelled data, where the output variable is already known, and the algorithm learns patterns from this data. These learned patterns are then applied to the test set for prediction or classification. Supervised algorithms require external assistance and operate by comparing their predicted outputs with actual outputs, adjusting the model based on any detected errors (Mahesh, 2020).

One important aspect of supervised learning is that it typically leaves the probability of input values undefined unless the expected output is known. The model receives features (inputs) and correct outputs (labels) to learn the relationship between them. If input values are missing, the model cannot infer the outputs, as the estimator relies on complete input data for accurate predictions (Nasteski, 2017).

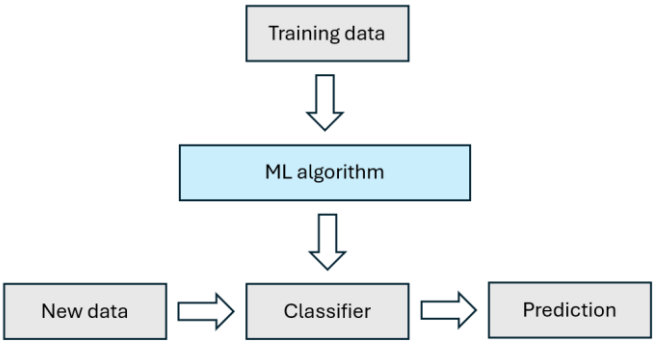


***Figure 3*** *Supervised learning process. Adapted from Nasteski (2017)*

#### *2.2.1.3 Binary classification*

ML classifiers are frequently designed for binary classification problems, which simplifies both their design and analysis. During the training process, these classifiers learn to establish a decision boundary in the feature space, which effectively separates data points into two classes as seen in Figure 4. The goal is to optimize this boundary to ensure the best possible separation between the classes. Once the training is complete, the classifier can predict the class of a new data point by determining its position relative to the learned decision boundary. This process allows the model to make accurate predictions even for unseen data (Baştanlar & Özuysal, 2013).

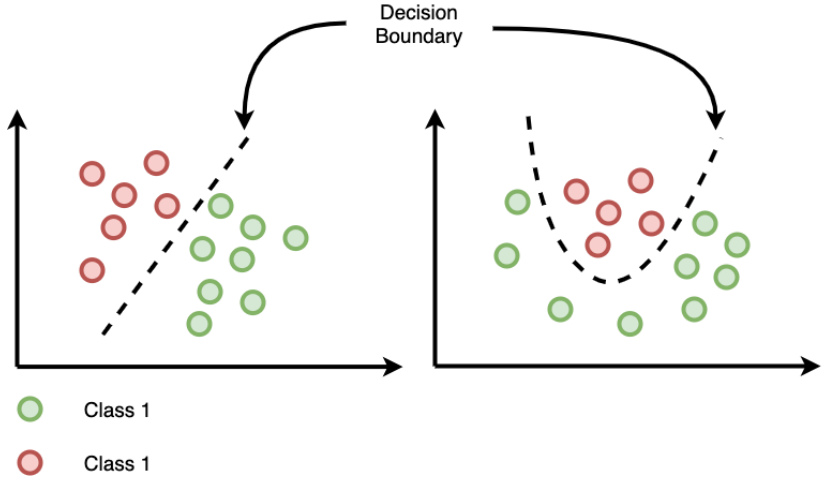


***Figure 4*** *Binary classification. Reproduced from Wickramarachchi (2020)*

#### *2.2.1.4 ML Algorithms*

##### 2.2.1.4.1 Ridge Classifier

The Ridge Classifier is a linear model that applies L2 regularization to prevent overfitting and manage multicollinearity in the data. This approach is useful when dealing with high-dimensional or noisy datasets, as the Ridge regularization penalizes large coefficients, helping the model generalize better. This technique has been widely adopted in various ML tasks (Hoerl & Kennard, 1970).

##### 2.2.1.4.2 Random Forest

The Random Forest Classifier, as proposed by Breiman (2001), is an ensemble method that builds multiple decision trees and combines their outputs to make a final prediction. Each tree is trained on different subsets of the data, which reduces overfitting and

improves accuracy. In this implementation, the Random Forest uses 100 trees, which helps it handle complex classification tasks effectively while minimizing overfitting risks.

#### 2.2.1.4.3 Support Vector Machine (SVM)

The SVM classifier, developed by Cortes and Vapnik (1995), is designed to find the optimal hyperplane that separates data points from different classes. This model works particularly well in high-dimensional feature spaces and is effective for binary classification tasks. In this implementation, the SVM is configured to provide probability estimates, allowing the model to output class probabilities alongside predictions.

### 2.2.2 Neural Networks

More recent approaches have moved towards DL, which automatically learns hierarchical feature representations from raw neuroimaging data. DL has been described as a potentially much more powerful approach when compared to traditional shallow ML, given its ability to learn very intricate and abstract patterns of data, especially in the case of brain-based disorders where subtle and diffuse alterations are common (Plis et al., 2014). In fact, Vieira et al. reviewed studies that applied DL in neuroimaging and found evidence that DL tends to outperform shallow ML methods. This is most especially true for studies that combine imaging and non-imaging data. This would suggest that any relationship between the abnormalities of the brain and symptoms may exist at deeper, more abstract levels where DL methods are better positioned to capture.

#### *2.2.2.1 Fundamentals of Artificial Neural Networks*

An Artificial Neural Network (ANN) is a computational model inspired by the structure and function of the human brain, consisting of interconnected neurons that process data in layers. ANNs have a specific architecture that mimics biological nervous systems, with neurons arranged in a complex and nonlinear form (Malekian & Chitsaz, 2021). Typically, ANNs include three main layers: the input layer, which receives raw data; the hidden layers, where data is processed through a series of weighted connections;

and the output layer, which generates the final prediction or classification (Haykin, 1998). The neurons in the network are connected by weighted links, and during the training process, these weights are adjusted to minimize the error in the network's predictions. This weight adjustment is achieved through a process called backpropagation, where the error is calculated and propagated back through the network to update the weights iteratively (Rumelhart, Hinton, & Williams, 1986). The outline of an ANN is seen in Figure 5. ANNs are highly adaptable, allowing them to adjust to changing inputs without needing to modify the output criteria. In supervised learning, the network's predicted output is compared to the actual output, and the error is used to improve the model. ANNs are widely used in various fields such as image recognition, speech processing, and medical diagnostics due to their ability to learn complex patterns in data (Mahesh, 2020).

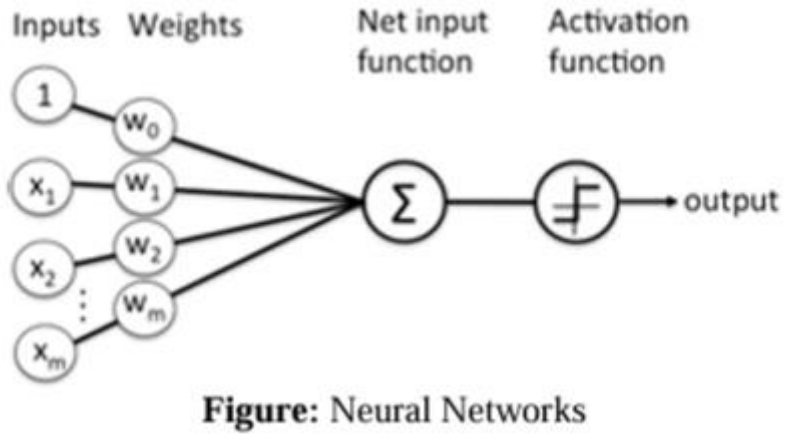


***Figure 5*** *Neural Network outline. Reproduced from Mahesh (2020)*

#### *2.2.2.2 Deep Learning*

DL works by building multiple levels of representation, where each layer transforms the input data into something slightly more abstract. Starting with raw input, each layer builds upon the previous one, capturing more complex patterns. This process, made possible through the combination of simple yet non-linear modules, allows DL models to learn very complex functions over time. In tasks like classification, deeper layers of the network focus on enhancing the important parts of the data needed to distinguish between categories, while ignoring irrelevant details (LeCun et al., 2015).

DL models, which are essentially advanced forms of ANNs, consist of multiple layers connected by weights that help pass information from one layer to the next (Wang et al., 2020). These extra layers allow the models to understand more complex relationships in the data, making them particularly effective at handling large and intricate datasets (Wang et al., 2020). For example, in an image recognition task, the early layers might pick up on simple features like edges, while the deeper layers are able to detect more complex structures, such as entire objects or faces (Chang, 2015).

### 2.2.3 Graph Neural Networks

While the training performance looks promising, existing DL approaches have faced some challenges in analysing the brain connectome. For example, Zhao et al. (2022) pointed out that Convolutional Neural Networks (CNNs) are normally for image feature learning and hence are ideally suitable for data whose structure contains well-defined local neighbourhood in the Euclidean space. It is comparatively difficult to extract representative features from brain network graphs since the structure of these graphs is irregular and non-Euclidean (Niepert et al., 2016).

To overcome this challenge, there is an increased interest in the use of GNNs that represent the brain as a graph with brain regions as nodes and functional or structural connections as edges. GNNs can capture complex relationships between brain regions more naturally compared to other models; hence, GNNs are showing great promise for classifying disorders in the brain as they have the advantage of exploiting the non-Euclidean structure present in the brain connectivity data. This would place them in a position to capture patterns that may elude more traditional models. (Parisot et al., 2018).

#### *2.2.3.1 Fundamentals of Graph Theory*

Graph theory is a branch of mathematics concerned with the study of graphs, which are structures utilized to model the relationship between objects. A graph includes entities, usually referred to as nodes or vertices, and edges that connect them, also known as links. The nodes represent entities, while the edges express the connections or interactions between these entities as seen in Figure 6. Formally, the graph is represented as G = (V, E), where V is considered a set of nodes, and E - a set of edges.

Graphs can come in the form of being directed, with edges having direction, or undirected; they can be weighted-the edges represent the strength of the relationship-or unweighted graphs (Majeed & Rauf, 2020). In fact, because of such flexibility in representation, graph theory is a powerful tool for modeling natural and artificial systems with intricate networks (Veličković, 2023).

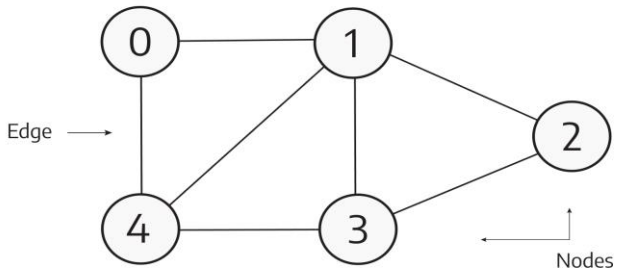


***Figure 6*** *Graph network. Reproduced from Ghazaryan (2023)*

Graph theory in neuroscience normally models the brain as a network using rs-fMRI. These methods consider nodes as ROIs; the edges are either functional or structural connections linking different regions. A functional connectome is created by constructing a matrix that captures all possible paired connections across brain regions, with the goal of quantifying how different areas of the brain interact with one another (Yang et al., 2020). The connectivity between nodes is often quantified, and a binary matrix-an adjacency matrix-created by thresholding the connectivity matrix points to whether there is connectivity between nodes or not (Shahhosseini & Miranda, 2022). A rough outline of graph networks construction is drawn in Figure 7.

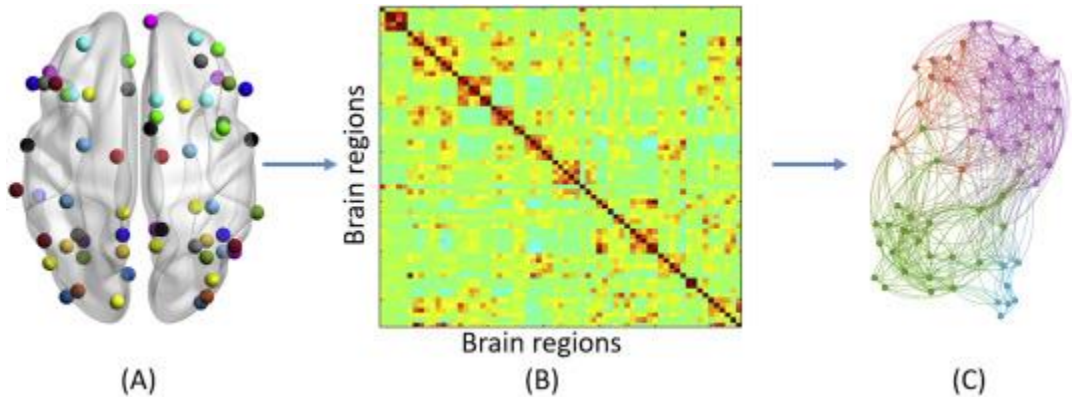


***Figure 7*** *(A) brain regions on Glass brain surface. (B) Pearson's correlation between BOLD fMRI time series of brain regions. (C) Network representation from connectivity matrix - circles represent nodes and lines represent edges. Reproduced from Yang et al. (2020)*

#### *2.2.3.2 Basic Concepts of Graph Neural Networks*

With the rapid development of DL, many researchers have shifted their attentions to DL methods for tackling graph-related tasks. Among them, GNNs have become one of the strong solutions in such a context, which allows models to deal with graph data effectively. With their usage of both hierarchical and iterative operators, GNNs are able to learn a task-oriented representation of either vertices, edges, or even an entire graph. These learned representations enable GNNs to solve a wide range of tasks on graphs using classical ML (Zhou et al., 2022).

GNNs learn to propagate and aggregate information from neighbours across multiple layers, building informative representations of nodes. First of all, it represents the input, whereby for each node in the graph, there should be a feature vector, and the graph itself must be represented in an adjacency matrix form of some sort. Later comes message passing or aggregation, which is just per-node steps that give each node a chance to collect information from each of its neighbours. It may come in many forms, which includes summation, averaging, or even the more complex operation of attention mechanisms, whereby the model learns to assign different importance to neighbours with regard to their relevance to the current node being updated (Veličković et al., 2017). That is typically done by using a non-linear function-just like in a ReLU activation-with a learnable weight matrix concatenated. These propagations and updates are repeated for several layers each of which enables the nodes to capture an increasingly larger neighbourhood in the graph. After several rounds of such message-passings, the final node representations can then serve in various downstream tasks such as node classification, graph classification, or edge prediction. (Xiao et al., 2023).

#### *2.2.3.3 Types of Graph Neural Network Layers*

##### 2.2.3.3.1 Graph Convolutional Networks (GCN)

GCNs were proposed in 2016 by Thomas Kipf and Max Welling as an extension of the idea of convolution on regular grids like images to graph-structured data. GCNs aggregate information from neighbouring nodes and update the representation of a node. This approach enables GCNs to capture the structure of the local neighbourhood while keeping the computational efficiency. GCNs work really well for node classification, graph classification, and link prediction, where the relationships between entities can be represented as graphs, due to their simplicity and scalability (Kipf & Welling, 2016).

##### 2.2.3.3.2 Graph Attention Networks (GAT)

Graph Attention Networks (GATs) were developed by Petar Veličković and his team in 2017 to introduce an attention mechanism into graph processing. GATs allow the network to weigh the importance of neighbouring nodes differently when aggregating information, rather than treating all neighbours equally. This attention mechanism makes GATs much more flexible and expressive, especially in cases when some connections are more relevant than others. The GATs showed their efficiency in such tasks as node classification and graph classification, where relationships are complex or heterogeneous (Veličković et al., 2017).

##### 2.2.3.3.3 Graph Isomorphism Network (GIN)

Graph Isomorphism Networks, or GINs, proposed by Keyulu Xu et al. in 2018, are designed to be as powerful as the Weisfeiler-Lehman test-a very powerful algorithm for distinguishing between different graph structures. By means of a certain aggregation technique, GINs guarantee that each graph has a unique representation and therefore do extremely well in tasks like graph-level classification. GINs perform particularly well when the differences in graph structure are very fine-grained, where two very similar structures can differ by large factors in properties (Xu et al., 2018).

#### 2.2.3.3.4 GraphSAGE (Sample and Aggregation)

GraphSAGE was proposed in 2017 by William Hamilton, Rex Ying, and Jure Leskovec to scale GNNs to large graphs and enable inductive learning. Unlike the traditional GNN, at training time, GraphSAGE samples a fixed-size neighbourhood of nodes and aggregates features from these neighbours to allow the model to generalize to unseen nodes and graphs. That makes GraphSAGE particularly effective for application scenarios where new data constantly emerges, say, for example networks where processing the entire graph at once is unfeasible (Hamilton, Ying, & Leskovec, 2017).

#### 2.2.3.3.5 Chebyshev Convolutional (ChebNet) Layer

ChebNet, introduced by Michaël Defferrard and colleagues in 2016, brought spectral methods into GNNs by using Chebyshev polynomials to approximate graph convolutions. This will reduce the computational cost of performing convolutions on graphs and hence will allow capturing efficiently the information about the local neighbourhoods. ChebNet turns out to be of particular use in tasks where, though the underlying graph structure may be complex, the local relationships bear much importance-for example, node classification in large-scale networks that need computational efficiency (Defferrard, Bresson, & Vandergheynst, 2016).

### 2.2.4 Related work

Several studies have ML and DL approaches to classify MDD and other neurological conditions from neuroimaging data, particularly rs-fMRI and structural MRI.

Zeng et al. (2012) used support SVM to classify MDD from rs-fMRI data, achieving high classification accuracy. However, their reliance on manually extracted features limited their ability to capture the full complexity of brain networks. In contrast, Patel et al. (2016) applied alternating decision trees to multi-feature data from rs-fMRI, effectively predicting MDD without the need for extensive manual feature extraction.

CNNs have also been used to classify brain disorders. Korolev et al. (2017) applied CNNs to structural MRI data for the classification of Alzheimer's disease (AD). Their study demonstrated that CNNs can automatically extract features related to brain atrophy and effectively distinguish between subjects with AD, mild cognitive impairment, and healthy controls (HC). Similarly, Wen et al. (2023) utilized CNNs to

classify MDD from rs-fMRI data. Their study highlighted the ability of CNNs to automatically extract both spatial and temporal features from complex neuroimaging data, significantly improving the classification accuracy of MDD.

In recent years, GNNs have emerged as a promising approach for classifying MDD using rs-fMRI data. For instance, Pitsik et al. (2021) used GNNs to analyse brain connectivity networks, finding that the topological features of these networks play a crucial role in MDD classification performance. Qin et al. (2022) implemented a GCN model and reported an accuracy of ~81.5% in a multi-site imaging study, demonstrating that GCNs can enhance classification performance while also providing insights into the neurobiological basis of MDD by identifying changes in brain network topologies.

In another study, Venkatapathy et al. (2023) proposed a GNN model for MDD classification, achieving 71.18% accuracy with upsampling and 70.24% with downsampling, reinforcing the utility of GNN-based models in clinical applications. Further advancements include the development of DepressionGraph, a two-channel GNN model designed by Xia et al. (2023). This model, which incorporates a transformer-encoder architecture, is adept at capturing complex brain connectivity patterns, outperforming traditional GNN models for MDD diagnosis. Moreover, Dai et al. (2023) introduced a novel feature extraction approach from time-series data in a GCN model, which achieved 75.8% accuracy in detecting recurrent MDD, further extending the applicability of GNNs in this domain.

Other studies have focused on subject-level GNN models, which treat individual brain connectivity networks as unique instances for classification. For example, Chen et al. (2022) applied a GNN model to classify Autism Spectrum Disorder (ASD) by leveraging both node and edge features in the brain connectivity graph. Their approach resulted in high classification accuracy, illustrating the potential of GNNs for subject-level brain disorder classification. Zhang et al. (2022) adopted a local-to-global GNN framework for the classification of AD and ASD using rs-fMRI data. This model captured both local and global brain connectivity patterns, resulting in improved classification performance and demonstrating the flexibility of GNNs in dealing with complex brain networks.

# 3 MATERIALS AND METHODS

## 3.1 DATA ACQUISITION AND PREPROCESSING

### 3.1.1 Subjects

The study was performed based on the Rest-meta-MDD consortium. The original dataset consists of 25 datasets from 17 participating hospitals in China that includes 1300 MDD patients and 1128 HC (Yan et al., 2022).

All agreed to provide diagnosis, age at scan, sex, and education. When collected systematically, measures of first-episode or recurrent MDD (if a patient's prior and current episode were diagnosed as MDD based on ICD10 or DSM-IV).

Data submitted to the consortium were fully deidentified and anonymized. The only shared information is subject ID, sex, age, education, episode status, medication status, illness duration, Hamilton Depression Rating Scale (HAMD) and Hamilton Anxiety Rating Scale (HAMA). The subject ID uploaded was reprogramed, so that it could not be traced back to the original subject ID used in the original studies. There was no face information in the MRI data, as no original T1 image was shared (Yan et al., 2019)

### 3.1.2 Image acquisition and processing

Rs-fMRI and three-dimensional structural T1-weighted MRI images were acquired from all participants at each local site. The site used in this study has 2000ms repetition time, 232 volumes were taken, and the total time was 464s (Yan et al., 2019).

The initial 10 volumes of each session were discarded to stabilize the MRI signal, followed by slice-timing correction to adjust the timing of slices acquired within each brain volume. The realignment of the time series images for each subject was achieved using a rigid body linear transformation with six parameters. Subsequently, individual T1-weighted images were co-registered to their mean functional image using a six degrees-of-freedom linear transformation, without re-sampling, and segmented into gray matter, white matter, and cerebrospinal fluid components. Transformations to standardize brain images into Montreal Neurological Institute (MNI) space were also computed (Yan et al., 2019).

To mitigate head motion artifacts, the Friston 24-parameter model was employed to regress out motion effects. Additionally, mean framewise displacement, calculated via Jenkinson's relative root mean square algorithm, was used as a covariate in group analyses to further control for motion. In validation analyses, an aggressive approach involving scrubbing—removing time points with FD exceeding 0.2mm—was applied to ensure robustness of the results. Other non-neural sources of variance, such as respiratory and cardiac effects, were regressed out from the data. A linear trend was included as a regressor to account for signal drift in the BOLD responses, and temporal bandpass filtering between 0.01 and 0.1 Hz was applied to all time series to focus on relevant low-frequency fluctuations (Yan et al., 2019)

### 3.1.3 Data selection

First, data exclusion followed standard quality control procedures based on several criteria. Site 25 was excluded due to its predominance of late-onset depression cases, primarily in patients older than 60, and those who had remitted. Subjects lacking information on sex, age, and education were also excluded, as well as those with poor imaging data and inadequate spatial normalization, as determined by visual inspection. Additionally, subjects under 18 or over 65 years of age, and those with insufficient coverage (resolution poorer than 0.2mm) were removed. To further eliminate subjects with distortions not detected through visual inspection, we excluded those with a spatial correlation below 0.6 between each participant's regional homogeneity (ReHo) map and the group mean ReHo map, a threshold set at the mean minus two standard deviations. Lastly, sites with fewer than 10 subjects in either group were removed, along with Site 4, which was identified as a duplicate (Yan et al., 2019).

For the sake of computational resources, and the fact that there were slight differences across acquisition sites, like repletion time, number of volumes and total time, only one site was selected. So, the subset used consisted of a total 479 participants, 250 depression patients and 229 controls.

## 3.2 GRAPH REPRESENTATION OF RS-FMRI DATA

This section presents the construction of local and global brain graphs from rs-fMRI data. The local graph captures functional connectivity between brain regions and produces regional brain graph embeddings for each subject, while the global graph represents relationships between subjects based on both imaging and non-imaging data. The main architecture for constructing these graphs followed the one proposed by Zhang et al. (2022). The detailed methods for constructing these graphs are explained in the following paragraphs and Figure 8.

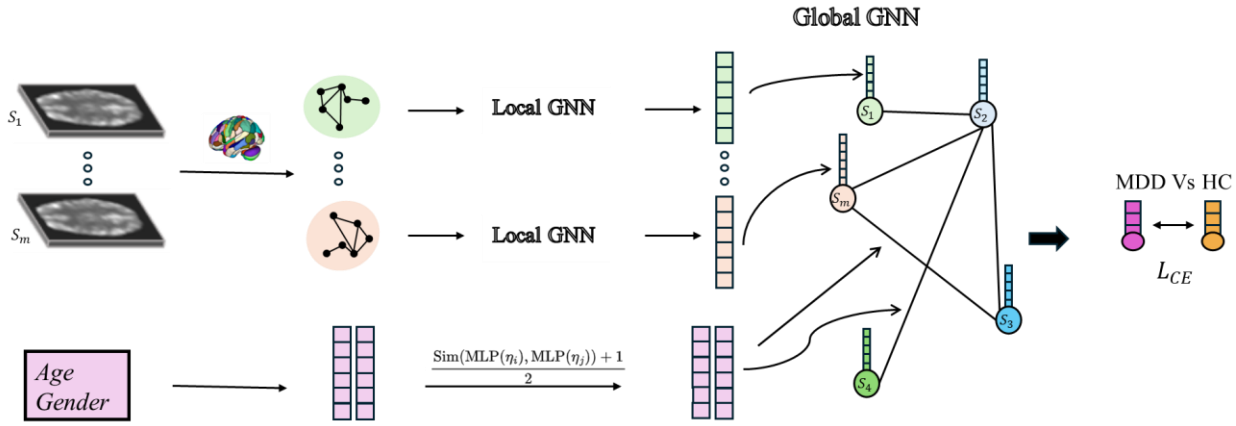


***Figure 8*** *Pipeline of the LG-GNN. Adapted from Zhang et al. (2022)*

### 3.2.1 Local graph construction

For each subject, we construct the regional brain graph based on the rs-fMRI image of the subject. First the whole brain was parcellated into 112 regions of ROIs according to Harvard-Oxford atlas (Makris et al., 2006).

Time series of BOLD signals from voxels in each ROI were extracted and averaged. Functional connectivity between each pair of ROIs was evaluated using the Pearson correlation (Pearson, 1895) coefficient of the corresponding time series.

The Pearson correlation coefficient is calculated as seen in Eq.1:

$$r_{ij} = \frac{\sum_{t=1}^{T}\left(T_{\mathrm{ROI}_i}(t) - \overline{T_{\mathrm{ROI}_i}}\right)\left(T_{\mathrm{ROI}_j}(t) - \overline{T_{\mathrm{ROI}_j}}\right)}{\sqrt{\sum_{t=1}^{T}\left(T_{\mathrm{ROI}_i}(t) - \overline{T_{\mathrm{ROI}_i}}\right)^2}\sqrt{\sum_{t=1}^{T}\left(T_{\mathrm{ROI}_j}(t) - \overline{T_{\mathrm{ROI}_j}}\right)^2}} \qquad \textit{Eq.1}$$

Where $\overline{T_{ROI_i}}$ and $\overline{T_{ROI_j}}$ are the means of the time series for $ROI_i$ and $ROI_j$respectively, and $\sum$ indicates summation over time points *t*.

The Z-score (Fischer, 1915) transformation is given by Eq.2:

$$z_{ij} = \frac{r_{ij} - \mu}{\sigma} \qquad \textit{Eq.2}$$

Where $\mu$ is the mean of the correlation coefficients, and $\sigma$ is the standard deviation.

The brain graph $G_{local} = \{V, A\}$ is constructed with each node corresponding to a ROI. The node set *V* is given by Eq.3:

$$V = \{v_1, v_2, \dots, v_n\} \qquad \textit{Eq.3}$$

Each node $v_i$ in the graph represents a specific ROI, and the corresponding node features can be described by a matrix X (Eq.4):

$$X = [x_1, x_2, \dots x_n]^T \qquad \textit{Eq.4}$$

The connections between these nodes are captured by the adjacency matrix *A*, which defines the edges of the graph. The adjacency matrix *A* is defined as follows in Eq.5:

$$A_{ij} = \frac{\mathrm{Cov}(v_i, v_j)}{\sigma_{v_i}\sigma_{v_j}} \qquad \textit{Eq.5}$$

*Where* $Cov(v_i, v_j)$ *is the cross variance between ROIs corresponding to* $v_i$*and* $v_j$.

Therefore, the weight of an edge between two is given by the value of the Pearson correlation coefficient between the features of these nodes.

**Error! Reference source not found.Error! Reference source not found.Error! Reference source not found.Error! Reference source not found.Error! Reference source not found.Error! Reference source not found.**

### 3.2.2 Global graph construction

The global graph $G_{\text{global}}$ represents the relationships between subjects based on both their brain imaging data and non-imaging data. The nodes represent subjects, and the edges, weighted by similarity scores, capture the relationships between them.

The global graph $G_{global}$ has nodes $V = \{S_1, S_2, \dots, S_m\}$ where each node corresponds to a subject $S_i$. Each node is associated with an embedding vector $Y_i$ for subject $S_i$. , which is derived from the data processed by a local GNN.

The adjacency matrix A′ defines the connections between subjects. It is given by Eq.6:

$$A' = C' \circ W \qquad \textit{Eq.6}$$

Where *C'* is a binarized connectivity matrix that captures the similarity between subjects based on both imaging and phenotypic data.

The Hadamard product ∘ (naive matrix multiplication) is used to combine C′ with a weight matrix W that is based on non-imaging data.

The elements of the weight matrix W are defined in Eq.7 based on the similarity of non-imaging data $\eta_i and\ \eta_j$ of subjects $S_i\ and\ S_j$:

$$W_{ij} = \frac{\text{Sim}\left(\text{MLP}(\eta_i), \text{MLP}(\eta_j)\right) + 1}{2} \qquad \textit{Eq.7}$$

Here, MLP is the output of a Multi-Layer Perceptron applied to the non-imaging data.

Sim denotes the cosine similarity between the outputs of the MLP for subjects $S_i$ and $S_j$.

*Similarity matrices:*

The first similarity matrix $S_1$ is created using the embedding vectors $Y_i$ from the imaging data as in Eq.8:

$$S_1(i,j) = \exp\left(-\frac{\left[\rho(Y_i, Y_j)\right]^2}{2\sigma^2}\right) \qquad \textit{Eq.8}$$

Where $\rho(Y_i, Y_j)$ is the correlation distance between the embeddings of subjects $S_i$ and $S_j$. And $\sigma$ is the mean of the squared distances $\left[\rho(Y_i, Y_j)\right]^2$.

The second similarity matrix $S_2$ is built using non-imaging information, such as age and gender, which helps capture other relevant clinical information.

The combined matrix $C'$ is created by performing the Hadamard product of $S_1$ and $S_2$ as in Eq.9:

$$C' = S_1 \circ S_2 \qquad \textit{Eq.9}$$

C' is then binarized by setting connections with values greater than 0.4 to 1, and the rest to 0.

## 3.3 Local-to-Global Graph Neural Network

The model's two primary components: the Local-GNN and the Global-GNN. The process begins by taking input from *m* subjects, incorporating both imaging and demographic data. The Local-GNN is first utilized to create embeddings of brain graphs based on regional brain networks constructed from the rs-fMRI data. Next, the Global-GNN is applied to classify the nodes, identifying the disease status for each subject, with each subject's embedding serving as a node in the graph. The following subsections provide a detailed explanation of both the Local-GNN and the Global-GNN.

### 3.3.1 Local-GNN Architecture

The Local-GNN architecture consists of a total of five graph convolution layers (GC), two of which are part of the Self-Attention Based Pooling (SABP) module. The GCs are fundamental for learning node features based on the graph's structure, while the

SABP module is responsible for selecting and retaining the most critical nodes in the graph, effectively reducing its size while preserving essential information. The overall framework can be seen in Figure 9.

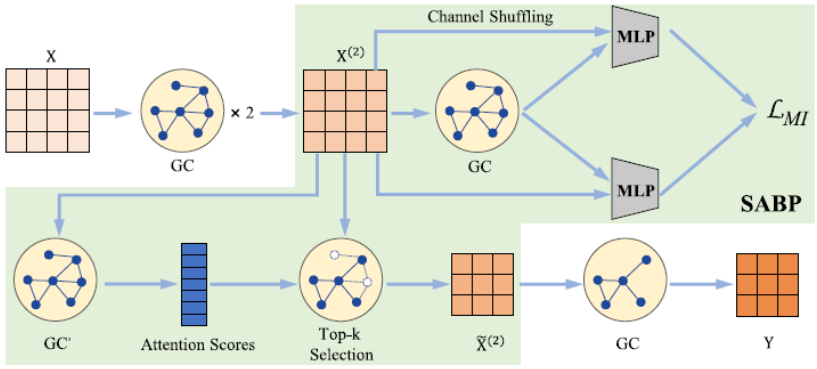


***Figure 9*** *Framework of Local-GNN component. Reproduced from Zhang et al. (2022)*

The Local-GNN starts with an initial GC layer that reduces the input node feature dimension from 112 to 64. This layer applies a ReLU activation function to introduce non-linearity. The second GC layer further reduces the feature dimension from 64 to 20, extracting more refined node-level representations.

After the second graph convolution, the SABP module is applied. The SABP module is designed to identify the most important nodes using a self-attention mechanism. Specifically, it selects 90% of the nodes based on their attention scores. This helps reduce the graph size while maintaining critical information for downstream tasks. Within the SABP module, another graph convolution is applied to learn an additional set of node embeddings. The scores for node importance are computed, and the highest-ranked nodes are selected to reduce the graph's complexity. The final node representations are further processed by pooling and attention-based filtering.

Following the SABP module, the third GC layer (outside the SABP) keeps the feature dimension at 20, refining the node features extracted from the reduced graph.

Finally, the node embeddings from the second convolution and the output of the third GC layer are concatenated and reshaped into a graph-level embedding for further processing or classification.

#### *3.3.1.1 Graph Convolution Layers*

Each GC updates the representation of a node by aggregating information from its neighbours and their connections. The core of this process lies in message passing, where each node exchanges its feature information with its neighbours, and this information is aggregated to form a new feature representation for the node (Kipf & Welling, 2016).

Mathematically, for the t-th GC layer, the node features $X^{(t)}$ are updated as follows in Eq.10:

$$X^{(t)} = ReLU\big(GC(X, A)\big) = ReLU\big(D^{-1/2}AD^{-1/2}X^{(t-1)}W^{(t)}\big) \quad \textit{Eq.10}$$

Where:

$X^{(t)}$ represents the node features at the t-th layer.

*A* is the adjacency matrix, capturing the connections between nodes.

*D* is the degree matrix, where each diagonal $D_{ii}$represents the degree of node *i*

$W^{(t)}$ is the trainable weight matrix for the t-th GC layer, which learns how to combine the information from neighboring nodes.

ReLU is an activation function that introduces non-linearity into the model, enabling it to learn more complex patterns.

*One-Hop and Two-Hop Neighbourhoods:*

During the aggregation process, each node's feature is updated based on a weighted combination of its neighbours' features. This ensures that each node's updated representation captures not only its own characteristics but also its immediate environment (Xue, Sun, & Sun, 2020).

A single GC layer primarily captures information from a node's immediate neighbours (one-hop neighbourhood), limiting their ability to explore more complex functional interactions from broader node features (Xue, Sun, & Sun, 2020).

Functional interactions in the brain may involve more distant connections (two-hop neighbourhoods). To capture this broader context, the model stacks two GC layers. The first layer captures one-hop information, and the second layer extends this to two-hop information. By stacking two or more layers, the model expands its receptive field, effectively incorporating information from increasingly distant nodes in the graph. This is particularly important in fMRI-based brain graphs, where certain brain disorders might be reflected in changes to both local and distant connectivity patterns.

The output of these two layers is a set of node representations $X^{(2)}$, where (Eq.11):

$$X^{(2)} \in R^{n \times c} \qquad \textit{Eq.11}$$

Here, *n* is the number of nodes (ROIs) in the graph, and *c* is the number of features.

#### *3.3.1.2 Self-Attention Based Pooling Module*

According to studies not all nodes in the brain graph are equally important for predicting brain disorders, as some regions of interest are more crucial than others (Li et al., 2021).

The SABP module is designed to identify these important nodes and pool them for further processing.

*Self-Attention Mechanism:*

The SABP module calculates a self-attention score as shown in Eq.12, for each node to determine its importance. This is done using a specialized GC layer, $GC'$ with one output channel:

$$z = GC'\left(X^{(2)}, A\right) \qquad \textit{Eq.12}$$

Where *z* is a vector of self-attention scores, with each element corresponding to a node in the graph. These scores indicate the relative importance of each node.

To reduce the graph's complexity, only the top-k nodes with the highest self-attention scores are retained. The indices of these top-k nodes are identified as follows in Eq.13:

$$idx = topk(z, k) \qquad \textit{Eq.13}$$

The selected nodes are then used to construct a new, smaller graph.

*Refinement of Node Features:*

The features of the selected nodes are refined by applying a tanh activation function to their self-attention scores, resulting in refined scores $\tilde{z}$ as in Eq.14:

$$\tilde{z} = tahn(z) \qquad \textit{Eq.14}$$

The final node features $\tilde{X}^{(2)}$ for the selected nodes are computed through element-wise multiplication of the original features $\widehat{X^{(2)}}$ (restricted to the selected nodes) and the refined scores $\tilde{z}$.

The original node feature matrix $X^2$ is reduced to include only the rows corresponding to the top-k selected nodes. This produces the matrix $\widehat{X^{(2)}}$ with dimensions k×c as shown in Eq.15.:

$$\hat{X}^{(2)} = X^{(2)}(idx, :) \qquad \textit{Eq.15}$$

The elementwise (Hadamard) product is performed between $\widehat{X^{(2)}}$ and the self-attention scores $\tilde{z}$. This operation (Eq.16) weights the features of the selected nodes by their corresponding self-attention scores.

$$\tilde{X}^{(2)} = \hat{X}^{(2)} \circ \tilde{z} \qquad \textit{Eq.16}$$

The adjacency matrix $A$ is also reduced to include only the rows and columns corresponding to the top-k selected nodes, maintaining the graph structure while focusing on the most relevant parts as in Eq.17.

$$\tilde{A} = A(idx, idx) \qquad \textit{Eq.17}$$

#### *3.3.1.3 Mutual Information Loss*

To ensure that the SABP module correctly identifies and retains the most important nodes, the model incorporates a Mutual Information Loss ($L_{\mathrm{MI}}$). This loss function

encourages the model to maximize the difference in importance between the selected nodes (which should have high attention scores) and the unselected nodes (which should have low scores).

The Mutual Information Loss is defined in Eq.18 as:

$$L_{\mathrm{MI}} = \frac{1}{n}\sum_{i=1}^{n} \mathrm{MLP}\left(\mathrm{Concat}\left(\mathrm{GC}\left(X^{(2)}, A\right), X^{(2)}\right)\right) - \log \frac{1}{n}\sum_{i=1}^{n} e^{\mathrm{MLP}\left(\mathrm{Concat}\left(\overline{X}, X^{(2)}\right)\right)} \quad \textit{Eq.18}$$

Where:

$\overline{X}$ is a shuffled version of the node features $X^{(2)}$.

The MLP (Multi-Layer Perceptron) is used to perform a linear transformation on the concatenated features, reducing their dimensionality.

*n* represents the number of nodes.

The goal of $L_{\mathrm{MI}}$ is to ensure that the model learns to focus on the most informative parts of the graph, preserving long-range relationships that are often missed by simpler attention mechanisms.

#### *3.3.1.4 Final Output of Local ROI-GNN*

The final output of the Local-GNN is the graph embedding *Y (shown in Eq.19)*, which represents the distilled information from the regional brain graph. This embedding is generated by applying a final GC layer to the pooled node features ($\widetilde{X^{(2)}}$) and the new adjacency matrix ($\tilde{A}$):

$$Y = GC\left(\widetilde{X^{(2)}}, \tilde{A}\right) \quad \textit{Eq.19}$$

This output ($Y = [y_1, y_2, \ldots, y_k]$) represents a compact and informative summary of the brain's regional structure, capturing the most discriminative features. It serves as the input to the subsequent Global-GNN, which performs the final classification.

### 3.3.2 Global-GNN Architecture

The Global-GNN Model, as shown in Figure 10, is designed to classify brain disorders by leveraging both imaging and non-imaging data from multiple subjects. To enhance performance and address common issues such as over-smoothing in deep GNNs, the model incorporates a multi-scale residual architecture. The architecture consists of multiple graph convolutional blocks and an adaptive weight aggregation block (AWAB), ensuring the model effectively captures both local and global patterns in brain connectivity.

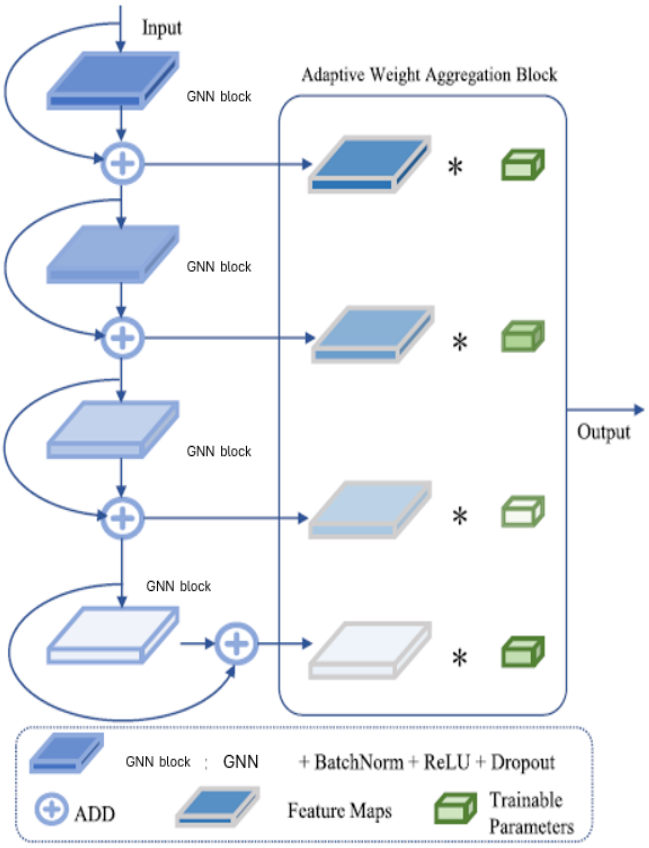


***Figure 10*** *Framework of Global-GNN component. Adapted from Zhang et al. (2022), with modifications*

#### *3.3.2.1 Graph Convolutional Blocks*

At the core of the Global-GNN architecture are four graph convolutional blocks, each designed to extract features from the graph structure. These blocks perform graph convolutions, which aggregate node features from neighbouring nodes and capture both local and global patterns. Each block processes node features of an initial dimension of

2020, reducing them to 20-dimensional feature vectors. Batch normalization is applied after each convolution layer to stabilize training, followed by a ReLU activation function to introduce non-linearity. Dropout is also applied to regularize the model and prevent overfitting.

*Graph Convolution Layer:*

The GC in each block applies a graph convolution operation that can be defined in general terms as in Eq.20:

$$H^{(l+1)} = CONV\left(H^{(l)}, A, W^{(l)}\right) \qquad Eq.20$$

where:

$H^{(l)}$ are the node features at layer l,

A is the adjacency matrix or other connectivity representation,

$W^{(l)}$ are the learnable weights for layer l,

CONV represents a general graph convolution operation, which aggregates information from a node's neighbourhood.

This layer processes the graph structure and the node features, transforming them into higher-dimensional feature spaces.

*Residual Connections:*

Residual connections are incorporated into each Graph Convolution block to mitigate the over-smoothing problem, where node features become indistinguishable after passing through multiple layers. These connections preserve the original node information by adding the input of each Graph Convolution block to its output before passing it to the next block. This ensures that deeper layers retain the initial features, facilitating the learning of relevant patterns.

*Block Structure:*

Each Graph Convolution block consists of several layers applied sequentially. First, a GC is used to apply convolution to the node features, allowing the model to capture graph structure information. This is followed by batch normalization, which stabilizes and accelerates training by normalizing the node features. Next, a ReLU activation function introduces non-linearity, enabling the model to learn more complex representations. Finally, a dropout layer is included to regularize the model by randomly dropping some node connections during training, which helps to prevent overfitting.

#### *3.3.2.2 Adaptive Weight Aggregation Block*

After passing through multiple Graph Convolution blocks, the outputs from these blocks are aggregated using the AWAB to form the final node embeddings. The AWAB dynamically assigns weights to the outputs of each Graph Convolution block based on learned parameters, allowing the model to emphasize features from different scales appropriately.

*Multi-Scale Feature Aggregation:*

The AWAB aggregates the outputs from different Graph Convolution blocks by assigning a weight $w^{(l)}$ to each block's output $H^{(l)}$, as demonstrated in Eq.21. These weights are learned during training and are normalized using a softmax function.

$$w^{(l)} = \frac{exp(r^{(l)})}{\Sigma_l exp(r^{(l)})} \qquad \textit{Eq.21}$$

where $r^{(l)}$are the trainable parameters. The final node embedding Z is computed as a weighted sum of the outputs from all Graph Convolution blocks in Eq.22:

$$Z = \sum_{l=0}^{3} w^{(l)} \cdot H^{(l)} \qquad \textit{Eq.22}$$

This aggregation approach ensures that the model can effectively combine information from different scales, capturing both local and global patterns in the graph.

#### *3.3.2.3 Final Classification Layer*

The aggregated node embeddings Z generated by the AWAB are then used for the final classification task.

The fully connected layer takes the aggregated embeddings Z and applies a linear transformation to produce the final logits. As it is a binary classification task (healthy vs. diseased), the output layer has two channels, corresponding to the two classes.

The fin

al classification logits are computed in Eq.23 as:

$$Logits = W \cdot Z + b \qquad Eq.16$$

where W and b are the weights and biases of the fully connected layer, respectively.

## 3.4 Implementation Details

The experiment was implemented on Google Colab using an A100 GPU for accelerated training. Key libraries included PyTorch, PyTorch Geometric, DGL, and Nilearn. Nilearn handled fMRI data processing, while PyTorch Geometric and DGL were used for building and training the GNNs. Additional libraries like torch-cluster, torch-sparse, and torch-scatter were installed for efficient graph computations, compatible with CUDA 12.1. Data processing relied on NumPy, Pandas, and Scipy, with NetworkX managing graph creation. This setup ensured efficient training and data manipulation, leveraging Colab's GPU resources effectively.

### 3.4.1 Experimental Setup

In this experimental setup, the Local-GNN reduces feature dimensions from 112 to 64 and then to 20, while the Global-GNN maintains a consistent feature size of 20 across all layers, incorporating batch normalization and dropout for regularization. Key

hyperparameters are finely tuned to optimize the network's performance. A learning rate of 0.001 is set to balance the speed and stability of convergence, and a weight decay of 5e-5 is used to prevent overfitting. The model is trained for a maximum of 70 epochs, with early stopping employed to avoid overfitting, using a patience of 15 epochs based on validation loss. Dropout rates are strategically applied—0.3 in the Global-GNN and 0.2 in the pooling layers of the Local-GNN—to further enhance the model's robustness against overfitting.

The Adam optimizer is selected for its ability to handle sparse gradients and its adaptive learning rate capabilities, making it ideal for DL tasks involving complex data like fMRI scans.

### 3.4.2 Loss Function

The network's loss function is a combination of categorical cross-entropy and mutual information loss. This dual-loss function plays a crucial role in enhancing both the classification accuracy and the relevance of the features extracted by the GCNs.

*Categorical Cross-Entropy*

This is the standard loss function for classification tasks. It calculates the discrepancy between the true class labels and the predicted probabilities output by the model as shown in Eq.24. In this setup, categorical cross-entropy ensures that the model focuses on minimizing the error between the ground truth and the predictions.

$$LCCE = -\sum_{i=1}^{N} y_i \, log(\widehat{y_i}) \qquad \textit{Eq.24}$$

*Mutual Information Loss:*

Based on the approach by Belghazi et al. (2018), mutual information loss is employed to maximize the dependency between the input and output, which enhances the quality of the features learned by the network. This loss encourages the model to learn representations that capture more information about the underlying structure of the input data, which is particularly beneficial when working with complex graph structures like fMRI data.

By using this dual-loss function approach, the model not only strives for accuracy in classification but also focuses on extracting meaningful and relevant features from the graph structure, ensuring that the learned representations differentiate well between depressed and non-depressed subjects.

### 3.4.3 Evaluation Metrics

To evaluate the model's performance in the context of classifying depression from fMRI data, several metrics are employed. Given the potential class imbalance, where the non-depressed class may dominate, it's essential to use metrics that capture different aspects of the model's predictions.

#### *3.4.3.1 Accuracy*

Accuracy measures the proportion of correctly predicted instances over the total number of instances (Eq.25). However, accuracy can be misleading in imbalanced datasets, as it may favor the majority class, giving a false sense of model performance (Powers, 2011).

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \qquad \textit{Eq.25}$$

#### *3.4.3.2 Precision & Recall*

Precision is the proportion of true positive predictions out of all positive predictions, indicating how well the model avoids false positives (Eq.26). This is crucial in cases, where misclassifying a non-depressed person as depressed could have significant consequences, such as unnecessary medical interventions.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \qquad \textit{Eq.26}$$

Recall, on the other hand, measures the proportion of true positives out of all actual positives, reflecting the model's ability to correctly identify true cases of depression (Eq.27). High recall is especially important in healthcare settings, where missing a true case of depression can lead to a lack of necessary treatment (Chicco & Jurman, 2020).

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \qquad \textit{Eq.27}$$

#### *3.4.3.3 AUC-ROC (Area Under the ROC Curve)*

The AUC-ROC metric (Eq.28) measures the model's ability to distinguish between depressed and non-depressed cases across various classification thresholds. This threshold-independent metric is particularly useful for imbalanced datasets because it evaluates the model's performance across all possible decision boundaries. A higher AUC value indicates better class separation (Saito & Rehmsmeier, 2015; Huang & Ling, 2005). The AUC-ROC is a common choice in medical diagnostics, where achieving a balance between sensitivity (recall) and specificity is critical.

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \qquad \textit{Eq.28}$$

$$\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{FP} + \mathrm{TN}}$$

### 3.4.4 K-Fold Cross-Validation

K-fold cross-validation is a robust technique used to evaluate model performance, especially when dealing with limited or imbalanced data. It helps to reduce both bias and variance by repeatedly training and testing the model on different subsets of the dataset, ensuring that performance estimates generalize well to unseen data.

#### *3.4.4.1 Detailed Explanation of the 5-Fold Cross-Validation Process*

In 5-fold cross-validation (Figure 11), the dataset is divided into five equally sized parts, or *folds*. The model is trained on four of these folds while the remaining fold is used as the test set. This process is repeated five times, with each fold being used as the test set exactly once. After all iterations, the performance metrics from each fold are averaged to provide a comprehensive estimate of the model's overall performance. This method ensures that the model is tested on all available data and reduces the likelihood of overfitting to a particular subset of the data (Kohavi, 1995).

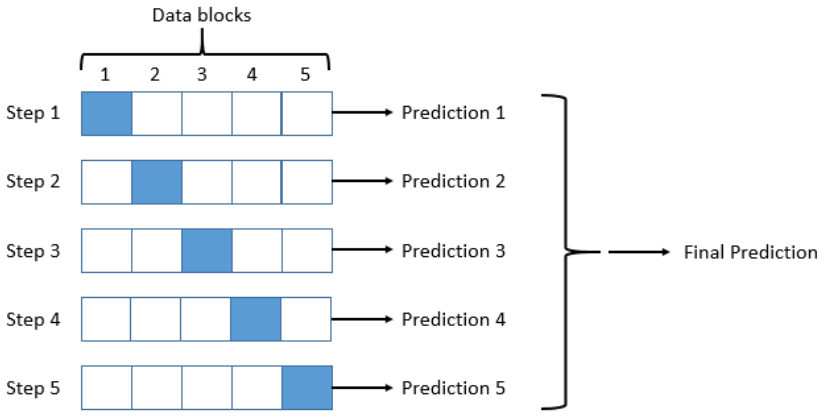


***Figure 11*** *Diagram of the 5-fold cross-validation method (blocks in blue represent the testing folds at each step) Reproduced from García et al. (2019)*

The decision to use five folds is based on the need to balance computational efficiency and performance stability. Using fewer folds, such as 2 or 3, can lead to higher variability in performance estimates, as the model might become overly sensitive to how the data is split. Conversely, using too many folds (e.g., 10 or more) increases the computational cost without a significant improvement in the quality of the performance estimate. A 5-fold approach ensures that every data point is used in both training and testing phases, allowing for a reliable estimate of model performance without requiring excessive computational resources (Arlot & Celisse, 2010).

#### *3.4.4.2 Integrating Early Stopping to Prevent Overfitting*

To further reduce the risk of overfitting, early stopping is integrated into the cross-validation process. During each training run, the model's performance on the validation set (one of the held-out folds) is monitored. If the validation loss fails to improve over a set number of consecutive epochs (in this case, 15 epochs), training is stopped. This prevents the model from overfitting to the training data by learning noise or irrelevant patterns, ensuring that it generalizes better to new, unseen data (Kohavi, 1995).

# 4 EXPERIMENTS AND RESULTS

## 4.1 OVERVIEW OF LG-GNN VARIANTS TESTED

The proposed architecture for the LG-GNN herein was evaluated by testing its Local-GNN and Global-GNN components separately. The performance of the Local-GNN in capturing region-specific brain connectivity was evaluated using various different configurations while keeping the Global-GNN fixed. Likewise, the Global-GNN integrates information coming from different regions into a global representation of the brain and is evaluated by changing its architecture and keeping the Local-GNN fixed. This method allowed them to conduct a focused analysis of how varying each component affects general classification performance.

### 4.1.1 Global-GNN Variants

Different configurations of the proposed Global-GNN architecture have been compared for performance, while keeping the Local-GNN constant. This was also done to maintain consistency in comparisons, using the same model proposed for the Local-GNN across all experiments. The only differences among the variants of the Global-GNN concerned the type of GCs: GCN, GAT, GIN, GraphSAGE, and ChebNet. All other aspects of the architecture and configuration, including pooling mechanisms and residual connections, remained identical. This setup allowed us to isolate how different configurations of Global-GNN capture global brain connectivity patterns and how such capturing affects overall performance.

### 4.1.2 Local-GNN Variants

The validity of the proposed LG-GNN architecture was tested against its variant of Local-GNN with state-of-the-art configurations. In all the experiments done, the Global-GNN architecture used was the one that gave the best performance in all scenarios consistently. Several variations in GCs—specifically GCN, GAT, GIN, and GraphSAGE—and pooling mechanisms were explored within the Local-GNN to determine the most effective approach for capturing local brain connectivity patterns. The goal was to assess how well the proposed Local-GNN configuration could encode region-specific features compared to the state-of-the-art approaches. This was achieved

by retaining the Global-GNN structure in its optimal setting so the differences in performance are only due to the modifications in the local component. Moreover, different pooling ratios were tested for the performance of the SABP module in the Local-GNN. Indeed, SABP is designed to select and retain the most informative nodes of the graph according to a learned attention mechanism that effectively reduces the size of the graph, keeping only the most crucial information. Several pooling ratios were tested, and performance versus computation efficiency trade-offs were evaluated. Various pooling ratios were tested to evaluate the trade-off between computational efficiency and performance. A higher pooling ratio preserves more detailed graph information but comes with higher computational costs, while a lower ratio speeds up computation at the risk of losing essential node features. These experiments helped to determine the most effective pooling ratio, striking a balance between graph reduction and the retention of critical information, ultimately optimizing the model's performance.

## 4.2 Overview of Baseline Models

To evaluate the performance of the proposed model, I employed several baseline models, including both graph-based and traditional ML approaches. These baselines were used to benchmark the proposed method, allowing for a robust comparison of its performance in classifying graph-structured data. Specifically, the baselines included a state-of-the-art 3-layered GCN, Ridge classifier, SVM, and Random Forest. By testing against both GCNs and classical ML models, I ensured a comprehensive evaluation of the task.

For the traditional ML models, the graph data was flattened into feature vectors, eliminating the graph structure and treating it as tabular data. In this implementation, the Random Forest uses 100 trees, which helps it handle complex classification tasks effectively while minimizing overfitting risks.

## 4.3 RESULTS

### 4.3.1 LG-GNN variants comparisons

#### *4.3.1.1 Global-GNN Variant Comparisons*

Each model was evaluated concerning how it captured the patterns of global brain connectivity that distinguish between positive and negative cases. These findings (Table 1, Figure 12) provide insight into the strengths of each architecture and their practical implications in this context.

Among all, GraphSAGE showed the best performance by correctly identifying a high number of positive cases with high accuracy 0.879, sensitivity 0.892, and specificity 0.865, yielding an AUC of 0.953. That means GraphSAGE had a very good balance between the identification of positive and negative cases, something that made the model highly reliable in most scenarios. Such a good performance demonstrates that it is really good both at low-level and global pattern capture, to which its consistent results across all metrics most likely are contributed.

GAT was particularly strong on sensitivity, 0.924, which indicates the algorithm's performance in true positives. However, its specificity was a bit lowered at 0.743, indicating a tendency to classify a lot of negative cases as positive. That goes to say this approach is very good in tasks where one needs to capture the positives with high sensitivity; however, it might run the risk of yielding more false positives. The high AUC of 0.912 shows very strong overall classification capability, but the tradeoff between high sensitivity and lower specificity is to be weighed.

GIN showed balanced performance, with sensitivity (0.812) and specificity (0.751) at relatively similar levels. This balance makes GIN a dependable option when equal attention to both positive and negative cases is necessary. However, with a relatively low accuracy (0.783) it did not reach the same level of performance as GraphSAGE or GAT, but its steadiness across metrics makes it versatile.

While GCN performed very well at 0.882 specificity, meaning that it gave more correct negatives, the sensitivity was considerably lower at 0.632, reflecting difficulty in detecting positive cases. That imbalance between both metrics hints that GCN may miss some positive cases-a problem for tasks where sensitivity is at the core of the subject.

The Cheb model was much more sensitive at 0.900, hence being highly capable of

detecting positive cases; however, its specificity at 0.682 was the lowest among models, meaning many false positives are diagnosed.

| | Acc | Sen | Spe | AUC |
|---|---|---|---|---|
| GCN | 0.751 (±0.008) | 0.632 (±0.034) | **0.882** (±0.001) | 0.817 (±0.012) |
| GAT | 0.837 (±0.003) | **0.924** (±0.003) | 0.743 (±0.018) | 0.912 (±0.002) |
| GIN | 0.783 (±0.003) | 0.812 (±0.007) | 0.751 (±0.003) | 0.869 (±0.001) |
| GraphSage | **0.879** (±0.001) | 0.892 (±0.001) | 0.865 (±0.006) | **0.953** (±0.001) |
| Cheb | 0.795 (±0.001) | 0.9 (±0.008) | 0.682 (±0.013) | 0.804 (±0.002) |

***Table 1*** *Results of Global-GNN variants comparisons*

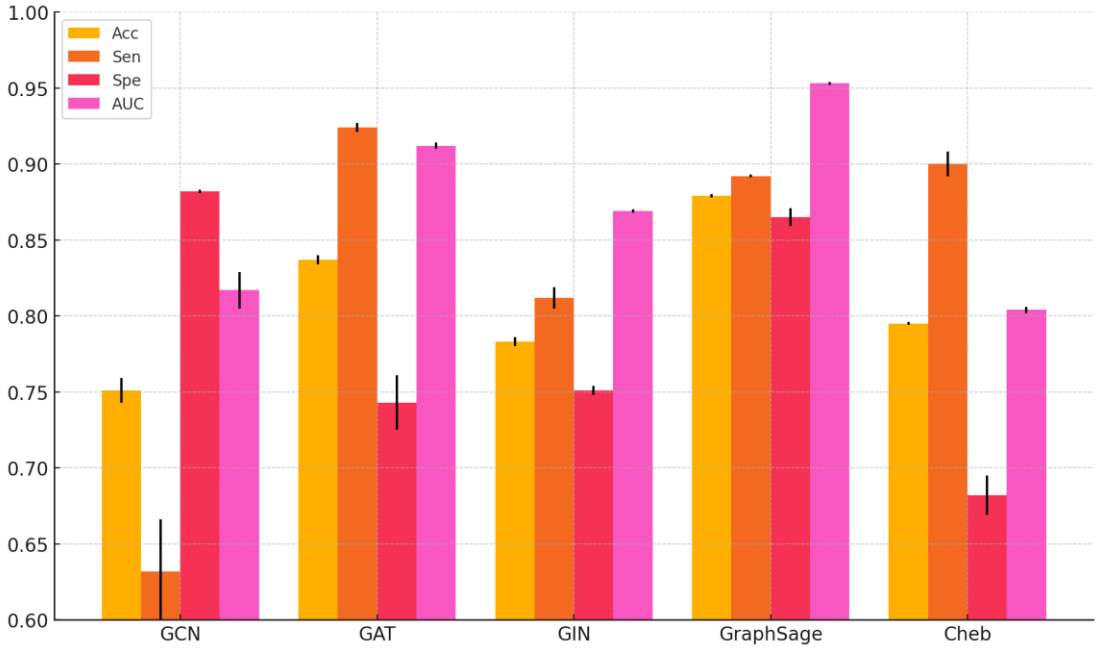


***Figure 12*** *Results of Global-GNN variants comparisons*

#### *4.3.1.2 Local-GNN Variant Comparisons*

From the comparison of the variants of Local-GNN, as shown in both Table 3 and Figure 14, LG-GCN, that is, GCN + SABP, clearly appears to be the best variant among all metrics. LG-GCN had an accuracy of 0.879, a sensitivity of 0.892, a specificity of 0.865, and an AUC of 0.953, hence demonstrating its strong capability in the classification of both positive and negative cases. The integration of the SABP module

is key to this superior performance, as SABP effectively retains the most important nodes while reducing irrelevant information, significantly boosting GCN's baseline performance. This result shows that SABP is critical for capturing meaningful local graph structures, allowing LG-GCN to outperform the other models. In particular, the pooling ratio of the SABP module is set to 0.9, which implies that 90% of the nodes input to the local-GNN are preserved. To verify the performance of SABP with respect to different pooling ratios, experiments with different ratios were carried out. The number of nodes retained in the coarsened graph significantly impacts classification performance, making the level of coarsening a critical factor in graph structure learning within brain networks. By adjusting the pooling ratio, we assess how varying the amount of retained information affects the model's ability to classify the data accurately. The classification accuracy was evaluated across varying pooling ratios as shown in Figure 13 and Table 2 and the best evaluation metrics were retained at a ratio of 0.9 pooling. This setting did the best balance; it retained the essential connections while effectively filtering out noise.

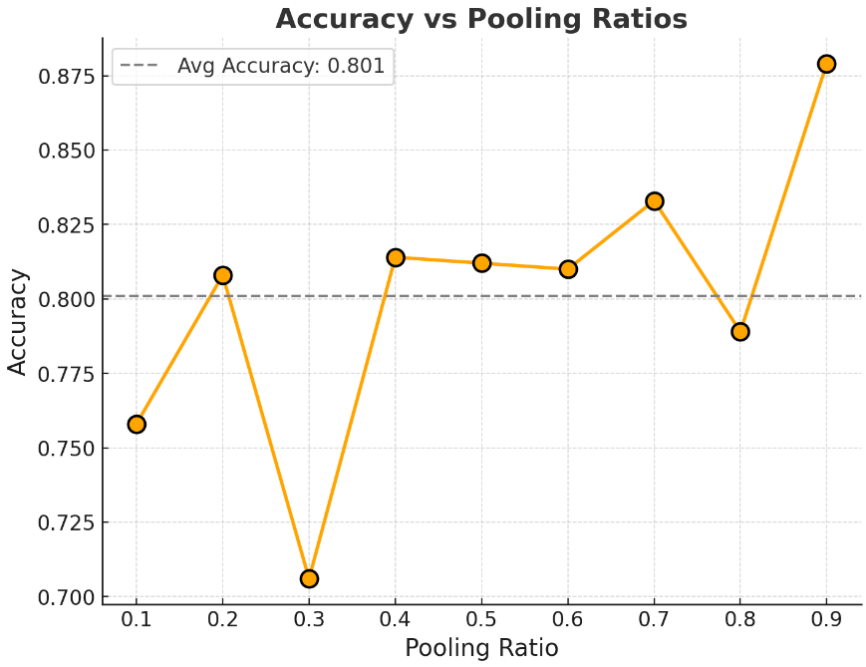


***Figure 13*** *Accuracy across different pooling ratios*

| | Acc | Sen | Spe | AUC |
|---|---|---|---|---|
| **0.1** | 0.758 (±0.0216) | 0.688 (±0.1255) | 0.835 (±0.0277) | 0.844 (±0.0155) |
| **0.2** | 0.808 (±0.0054) | 0.796 (±0.0149) | 0.821 (±0.0027) | 0.852 (±0.0068) |
| **0.3** | 0.706 (±0.0346) | 0.524 (±0.1841) | **0.904** (±0.0110) | 0.816 (±0.0108) |
| **0.4** | 0.814 (±0.0051) | 0.836 (±0.0132) | 0.791 (±0.0144) | 0.858 (±0.0046) |
| **0.5** | 0.812 (±0.0016) | 0.784 (±0.0089) | 0.843 (±0.0040) | 0.875 (±0.0015) |
| **0.6** | 0.810(±0.0028) | 0.808 (±0.0087) | 0.812 (±0.0129) | 0.847 (±0.0047) |
| **0.7** | 0.833(±0.0023) | 0.792 (±0.0043) | 0.878 (±0.0016) | 0.880 (±0.0039) |
| **0.8** | 0.789(±0.0032) | 0.840 (±0.0045) | 0.734 (±0.0064) | 0.834 (±0.0034) |
| **0.9** | **0.879** (±0.001) | **0.892** (±0.001) | 0.865 (±0.006) | **0.953** (±0.001) |

***Table 2*** *Classification metrics across different pooling ratios*

The GCN model, on which the proposed LG-GCN is based, performed reliably but modestly without the SABP module. It yielded an accuracy of 0.785, sensitivity of 0.792, and specificity of 0.778 for the GCN. Its AUC of 0.843 indicates a rather decent classification power; this model is limited with regard to holding on and being focused just on critical nodes, hence the overall performance is lower compared to LG-GCN. This really underlines how much SABP adds to enhancing the GCN architecture. GAT performs really well, especially on sensitivity, at 0.856, since it is intuitively designed into a sensitive model by the attention mechanism, giving higher importance to the weight of important nodes during feature aggregation. The accuracy is 0.833, while specificity is 0.809, hence giving the balance to GAT, though a little less sensitive towards identifying negative cases compared to the positive ones. The AUC of 0.891 underscores GAT's robust classification abilities, making it a good option for tasks where detecting positive cases is critical, though it still falls short of LG-GCN's comprehensive improvement due to SABP.
GIN delivers an overall decent performance, particularly excelling in specificity (0.822), making it effective at correctly identifying negative cases. However, its sensitivity of 0.740 suggests it struggles more with positive case identification. The accuracy of 0.779 indicates that while GIN is dependable, it doesn't perform as well as LG-GCN or GAT.

Lastly, GraphSAGE struggles the most among the models, with the lowest accuracy (0.657), sensitivity (0.588), and AUC (0.748). Its large variance in sensitivity indicates inconsistent performance, suggesting that GraphSAGE's neighbourhood sampling approach is not well-suited for tasks requiring detailed local feature aggregation. The model's poor performance highlights its limitations in handling complex local graph structures compared to more sophisticated models like LG-GCN, GAT, and even GIN.

| | Acc | Sen | Spe | AUC |
|---|---|---|---|---|
| GCN | 0.785 (±0.0035) | 0.792 (±0.0014) | 0.778 (±0.0111) | 0.843 (±0.0049) |
| GAT | 0.833 (±0.0014) | 0.856 (±0.0049) | 0.809 (±0.0216) | 0.891 (±0.0010) |
| GIN | 0.779 (±0.0047) | 0.740 (±0.0130) | 0.822 (±0.0144) | 0.846 (±0.0041) |
| GraphSage | 0.657 (±0.0169) | 0.588 (±0.1083) | 0.733 (±0.0240) | 0.748 (±0.0218) |
| LG-GCN (GCN+SABP) | **0.879** (±0.001) | **0.892** (±0.001) | **0.865** (±0.006) | **0.953** (±0.001) |

***Table 3*** *Results of Local-GNN variants comparison*

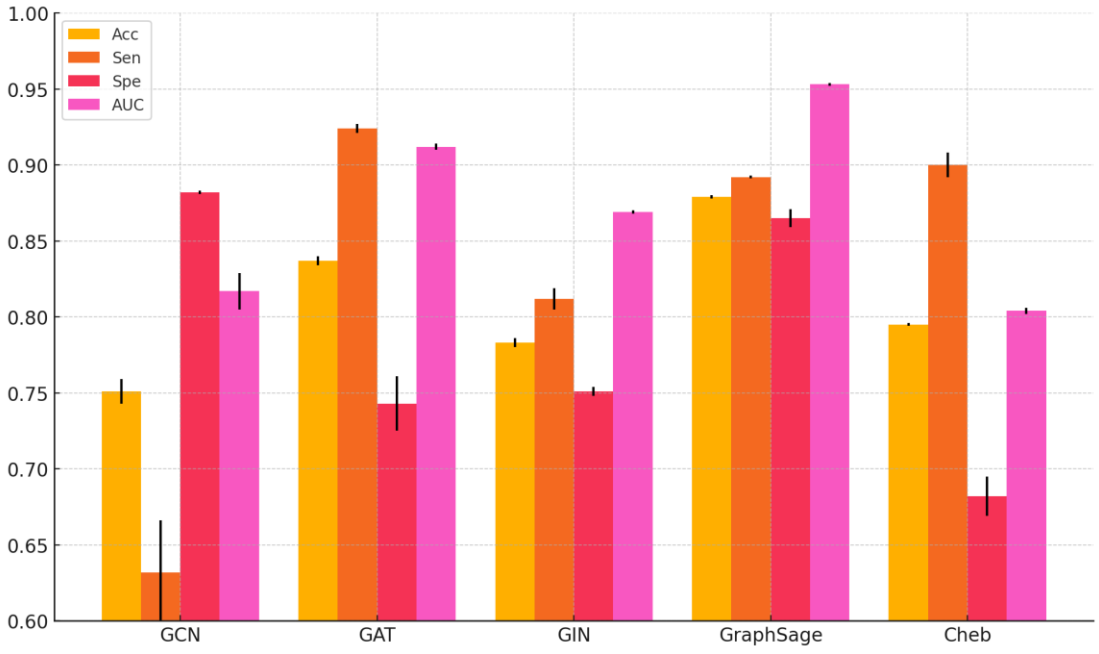


***Figure 14*** *Results of Local-GNN variants comparison*

### 4.3.2 Baseline comparisons

Compared to all the baseline methods, the proposed model of LG-GCN obviously outperforms them. It shows a very strong performance across all evaluation metrics. This will be beneficial in leveraging graph-based approaches compared to the traditional method, as illustrated in Figure 15 and Table 4.
On the other hand, the GCN-baseline model provides poor results with low accuracy (0.558) and, in particular, very low sensitivity (0.408), indicating that it has difficulty in finding the positive cases. The Random Forest classifier does better, with average accuracy 0.5825 and AUC 0.6092, although this struggles for a balanced performance, particularly on sensitivity (0.5330). The Ridge model has an accuracy of 0.6410 with an AUC of 0.7043, and a balanced performance for sensitivity and specificity, but it scores relatively lower. The SVM classifier has an accuracy of 0.6306 and AUC of 0.6736, but it shares the issue of imbalance, having a sensitivity of 0.5725 and a specificity of 0.6840. While these indeed constitute improvements, none of them reach the level of performance of LG-GCN, which significantly outperforms them on all metrics.

| | Acc | Sen | Spe | AUC |
|---|---|---|---|---|
| GCN- baseline | 0.558(±0.0114) | 0.408(±0.0519) | 0.721(±0.0381) | 0.491(±0.0203) |
| Random Forest | 0.5825(±0.0311) | 0.5330(± 0.044) | 0.6280(± 0.060) | 0.6092 (± 0.044) |
| SVM | 0.6306(± 0.0221) | 0.5725(± 0.0617) | 0.6840(± 0.0344) | 0.6736 (± 0.033) |
| Ridge | 0.6410(± 0.0315) | 0.6288(± 0.0616) | 0.6520(± 0.0873) | 0.7043(± 0.0390) |
| LG-GCN | **0.879** (±0.001) | **0.892** (±0.001) | **0.865** (±0.006) | **0.953** (±0.001) |

***Table 4*** *Results of baseline comparisons*

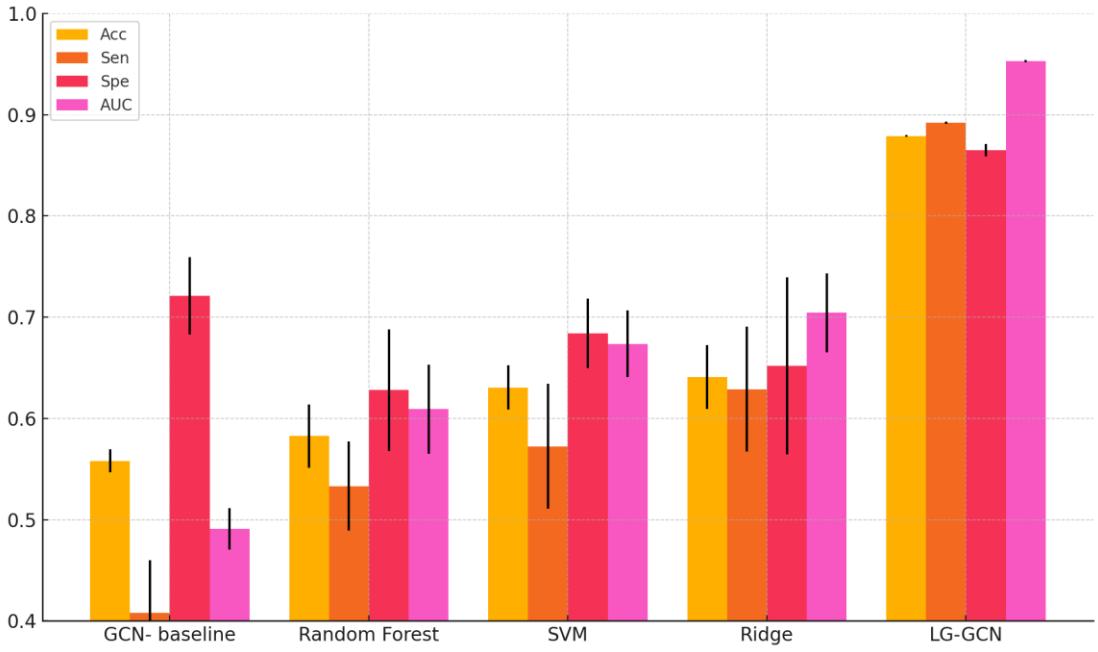


***Figure 15*** *Results of baseline comparison*

### 4.3.3 Biomarker Detection

To understand which brain regions contribute most to the classification of depression, the model outputs of the correctly classified patients from the SABP module are analysed, as these outputs highlight the discriminative features. The process for identifying the top-k important brain regions is described as follows: the weight of each node in the brain graph, representing a specific brain ROI, is obtained by passing the test set through the SABP module of the Local-GNN.

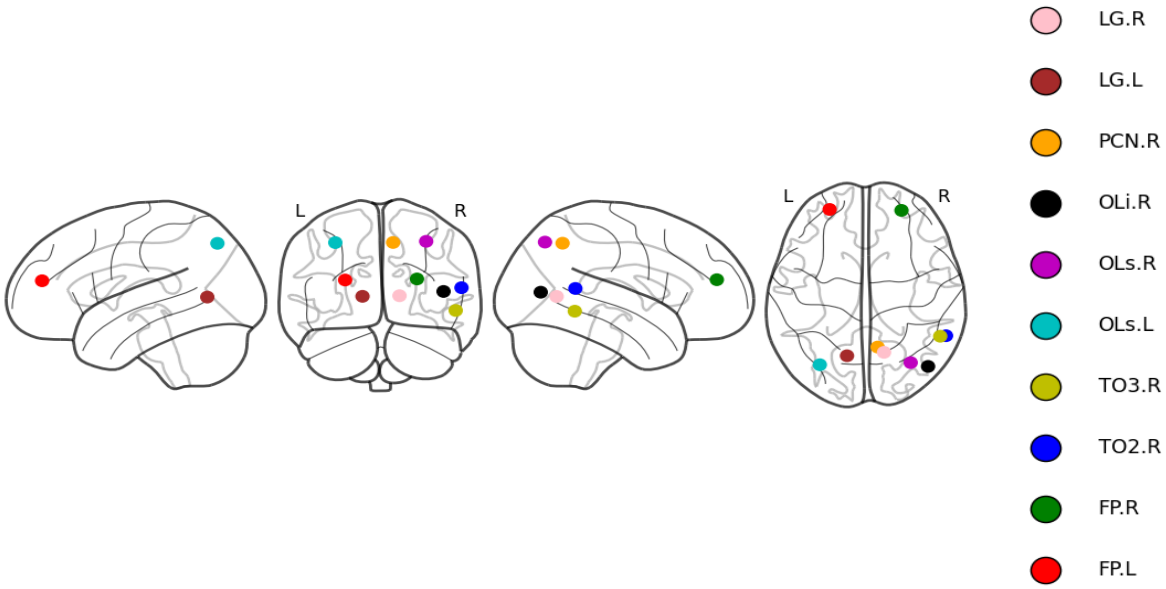


***Figure 16*** *Top-10 discriminative features for MDD classification*

A successive statistical analysis on these regions yields the top 10 based on two factors: the highest in terms of weight and the highest in terms of frequency. These selected regions can be treated as potential biomarkers. The value of k is set at 10 to ensure that all key brain regions are included in the analysis. Through this method, the brain regions most relevant to depression classification are identified, providing both accuracy and interpretability.

Biomarkers of depression revealed in this study, as shown in Figure 16, include the most prominent: the Frontal Pole (FP.L and FP.R), the right Temporal Occipital regions, especially the Middle Temporal Gyrus and Inferior Temporal Gyrus (TO2.R and TO3.R), the left and right Superior Lateral Occipital Cortex, and the right Inferior Lateral Occipital Cortex (OLs.L, OLs.R, OLi.R). Then there is the Lingual Gyrus (LG. L and LG. R) and the right Precuneus (PCN.R). These regions included the frontal, temporal, occipital, and parietal lobes, all showing significant weighted accumulations, hence making these a site of interest in the identification of neural patterns accompanying depression.

# 5 DISCUSSION

## 5.1 SUMMARY AND INTERPRETATION OF RESULTS

### 5.1.1 Model performance

The model of choice, which demonstrated the best performance, was selected based on its superior accuracy and AUC, as well as its balanced sensitivity and specificity. This model achieved better performance than the baseline models because it effectively captured both local and global aspects of brain connectivity while also incorporating demographic information. By including a local component that can provide detailed brain region patterns in each subject and a global component capturing the broader connectivity across subjects, the model employed multiscale information. Including demographic features further improved this classification performance by adding context to the brain connectivity data and yielded superior performance on all metrics. The GCN with SABP outperformed other architectures because the SABP module intelligently selected the most important nodes for classification, removing irrelevant data and focusing on the most discriminative features. This allowed the model to highlight key regions linked to depression while discarding noise, significantly boosting the GCN's baseline performance. This was further fortified in the fact that SABP can focus on relevant nodes, thereby making the already very good GCN for the capture of local relationships much more powerful for node classification at a regional brain level.

On the other hand, GraphSAGE yielded the best performance compared to other architectures on the global component due to neighbourhood sampling for efficient aggregation of information from neighbouring nodes, i.e., other subjects in the dataset. This allows the model to learn not only the local features at the node level but also global relationships between subjects, thus modelling both individual and inter-subject connectivity patterns effectively. The capability of GraphSAGE to balance local and global information, in addition to inductive nature, allowed it to generalize better on unseen nodes with handling the complexity present in a dataset, hence turning in superior performance in node classification tasks.

### 5.1.2 Biomarkers insights

The following regions were highlighted as probable imaging biomarkers for MDD in the present study: the Frontal Pole (FP.L and FP.R), the Right Temporal Occipital Regions, including the Middle and Inferior Temporal Gyri (TO2.R and TO3.R), the Left and Right Superior and Right Inferior Lateral Occipital Cortex (OLs.L, OLs.R, OLi.R), the Lingual Gyrus (LG.L and LG.R), and the Right Precuneus (PCN.R). These regions have important roles in several brain networks participating in the neural underpinnings of MDD.

Among the major subregions of the DMN commonly implicated in MDD are the frontal pole and precuneus. Abnormal functional connectivity between the two is widely regarded as related to ruminative and self-referential thinking, cognitive processes commonly disturbed in MDD sufferers (Cheng et al., 2018; Kawakami et al., 2024).A few reports have indicated increased frontal pole activity, especially during future-oriented thought, in MDD patients compared to HC (Katayama et al., 2019). More precisely, FPC has participated in several key pathological features of MDD, including reward processing, negative affect, and cognitive control (Fettes et al., 2017). On the contrary, decreased connectivity in the left lateral frontal pole indicates an impairment with attention and a reduction in the external focus of attention in patients with MDD (Veer, 2010). The frontal poles represent executive function and behavior oriented towards stimuli. Both are widely disturbed in depression (Kim et al., 2022).

Changes in MDD have also been observed in the precuneus. Lower resting-state activity within this structure is seen in depressive episodes, especially in those suffering from a recurrence of the disorder (Liu et al., 2017). Reduced activity in the right precuneus, along with impaired executive control, has also been reported in first-episode, treatment-naive patients with MDD (Luo et al., 2022). Moreover, the increased connectivity of the precuneus with the lateral orbitofrontal cortex, which takes part in non-reward processing, underlines once again the postulated involvement of the precuneus in the pathophysiology of depression (Cheng et al., 2018). Importantly, reduced connectivity between the dorsolateral prefrontal cortex and the precuneus has been shown to alleviate depressive symptoms, further highlighting its role in MDD (Taylor et al., 2022). The right precuneus has also been identified in the context of

emotional processing, while disruption in this region relates to broader brain impairments in MDD (Zhang et al., 2021).

The temporal gyrus, mainly the middle and inferior temporal gyri, has also become one of the most important regions in MDD. Asymmetry in the inferior temporal gyrus is associated with the severity of depressive symptoms, especially in medicated MDD patients (Kocsis et al., 2021). Gray matter abnormalities in the right middle temporal gyrus also occur, with altered functional connectivity within the DMN (Ma et al., 2012). There is a hypothesis suggesting that structural changes in the medial temporal lobe, including the left and right inferior temporal gyri, signal early signs of depression in adolescents suffering from early-onset MDD (Ramezani et al., 2014). Moreover, functional connectivity is enhanced between the PCC and left MTG, also involving the right middle frontal gyrus in MDD patients, at which site gray matter abnormalities have likewise been found (Zhu et al., 2022). Altered connectivity between the hypothalamus and the right MTG has also been noted in patients with depression (Liu et al., 2018). More severe depressive symptoms, including suicidal ideation, have been associated with changes in connectivity and regional homogeneity within the middle and inferior temporal gyri (Weng et al., 2019).

Other structural and functional changes that have been implicated in MDD include the occipital cortex. There was, for example, a decrease in surface area within the left superior occipital gyrus of drug-naïve MDD patients, which underlines early structural abnormalities in this region (Lee et al., 2021). Indeed, MDD patients were found with more prominent occipital bending than that of HC, further indicating the structural abnormalities within the occipital cortex (Maller et al., 2014). On one side, Scalabrini et al. (2020), Kaiser et al. (2015) refer to a reduced activity of the occipital cortex observed in MDD patients, especially in areas where there is visual processing and involvement in wider cortical networks. Moreover, changing GABA in the occipital cortex has been implicated not only in acute MDD-related fine-grained visual deficits, but such deficits were also associated with symptom severity in depression (Song et al., 2021). Thus, it is fair to claim that these data support the hypothesis that the occipital cortex might play a critical role both in the visual and psychopathological manifestations of MDD.

The structure has also been implicated in both structural and functional alterations with depression. Smaller surface area within the lingual gyrus is associated with the severity of anxiety and depression in young adults. Other areas that showed reduced dynamic rs-fMRI activity among MDD patients include the right lingual gyrus (Du et al., 2024). Among AD patients with depressive symptoms, for example, decreased functional connectivity was reported between the dorsal anterior cingulate cortex and the lingual gyrus (Liu et al., 2017). Some scholars have posited that the lingual gyrus would be of importance in helping to demarcate a depressed patient from a HC on various structural and functional features of the brain (Chiari-Correia et al., 2023). Stronger activation of the lingual gyrus is also observed in HC compared to MDD patients, again drawing much significance to this region in depression (Cooney et al., 2010). Taken together, these results collectively emphasize the network of regions within the frontal, temporal, occipital, and parietal lobes that are more or less consistently implicated in the pathophysiology of MDD, contributing to the cognitive, emotional, and sensory disruptions typical of the disorder and comprising the principal areas of interest for further investigation and targets for therapeutic intervention.

## 5.2 LIMITATIONS AND FUTURE DIRECTIONS

Despite these encouraging findings, there were several limitations in the present study. First, we did not use all the data in one hand due to computational limitations, but most importantly because RT, the number of volumes, and hence total scan time differed across subjects. Data were contributed from multiple sites using different MRI machines and/or different acquisition protocols. Although we tried normalizing data using methods such as mean/zero padding, interpolation, and truncation, they were not good enough in their own respect. In a couple of cases, these edits caused the loss of data or addition of noise to it, which could have had an impact on the result.

While necessary to evaluate performance, there do exist a few limitations in the use of random k-fold cross-validation in this study. Although we employed a subject-based k-fold methodology to ensure no data leakage occurred-meaning that no data from the same subject would be used in both training and test data sets-there is still a risk of bias. More precisely, random splits in k-fold may well result in folds that are not representative of the full diversity of data, especially if the sample size is small. This could lead to overestimation of performance when some patterns within the data happen

to be overrepresented in one fold and not in another. There is the possibility of bias with random k-fold cross-validation, but here the model has been tested on a number of different splits with reliably good results in each.

This stability was observed not only between different cross-validation splits but also often within individual folds, suggesting that the model was not overly sensitive to the way the data were partitioned. That is, the performance of the model is stable, not greatly influenced by specific data splits. Even with this consistency, there is the possibility of bias inherent in random k-fold approaches. As such, additional testing using stricter methods such as LOO cross-validation or using a completely out-of-sample dataset would be needed to effectively rule out biases and better estimate generalization capabilities for this model.

Another limitation of this dataset is its cultural specificity; the subjects all came from hospitals in China. Findings may poorly generalize to populations with other cultural or ethnic backgrounds for which symptoms and neural correlates of depression may vary.

In like manner, generalization to paediatric applications cannot be made in this study since the dataset consisted completely of adults. Depression could structurally and functionally appear different in the developing brains of children and adolescents, and this model might generalize less well when applied to younger populations. This calls for further research on applying similar modelling techniques to age-specific differences in diagnosis.

Moreover, the current study had data only on cross-sections; hence, the variance in the pattern of connectivity that informs either the development or progression of depression couldn't be estimated. This design may provide further clarification of temporal dynamics in brain networks and improves predictive power in either early detection or treatment response.

Another limitation in the present study is the static functional connectivity networks

employed. While the approach taken in this study gives insight into the overall brain connectivity, the dynamic nature of functional connectivity itself is not captured, which is very important for understanding time-varying interactions between regions of the brain. Since connectivity can change over time, future studies should be done with dynamic networks. This might offer a finer resolution in knowing exactly how brain function relates to depression. The use of multi-modal techniques which will include other imaging modalities, like structural MRI or DTI, will enhance the analyses and also make the accuracy of the classification better.

In this study, the only non-imaging data used included gender and age. Though these variables are crucial, some other important non-imaging information, which may involve genetic information, intelligence measures, or even socioeconomic factors, would go a long way in helping understand the factors for depression. The inclusion of such variables might improve model performance, ensuring more personalized and accurate predictions.

Lastly, while performance was quite high, GNNs-being the models implemented in this paper-tend to be "black-box" models. That is to say, limited understanding of why any given set of features led to a specific prediction is possible. The feature importance methods are useful in providing at least some level of interpretability-say, biomarker identification-but would be aided by either more inherently transparent models or additional posthoc interpretation techniques for enhanced clinical utility.

## 5.3 Conclusion

In this work, we extend the LG-GNN framework and apply it for MDD classification using rs-fMRI data. Although the framework of LG-GNN has been explored in other contexts, this work is dedicated to applying it to MDD by making some specific improvements that better capture the relevant pattern of brain connectivity. That would be possible with the help of the local component, including GCN with SABP, which allows further feature selection refinement in the brain region and helps the model to focus on the most critical neural patterns that can be associated with depression. It

allows finding some potential biomarkers and enhances the understanding of the brain's local abnormalities in MDD.

GraphSAGE effectively supported the capture of intersubject relationships from the Global-GNN by aggregating neighbourhood information so that the model could integrate global connectivity patterns across subjects, thus enhancing overall accuracy in classification. Besides imaging data, other non-imaging information such as gender and age further added to the level of detail at the subject level that contributed to enhancing the predictive power of the model.

We further optimized the capability of the model in capturing both local and global brain connectivity patterns relevant to MDD by exploring different GNN layer architectures and fine-tuning the SABP module. These enhancements allowed the model to perform better than baseline and other state-of-art methods, showing the potential of GNN-based frameworks for the classification of MDD.

Moreover, this study provides proof that the improved framework of LG-GNN is capable of adapting effectively to perform the classification task of MDD and thus proves its versatile usability on other neurological and psychiatric conditions. This confirms the broader applicability of the framework for brain connectivity-based classification tasks.

With that said, there were quite a few limitations that this study design overcame, and several remain. Future work should consider the employment of dynamic networks, multi-modal imaging techniques, and richer non-imaging data to achieve even better model performance and generalization. Also, extending the application to larger and more diverse datasets and methodologies with finer and more robust grains will provide further verification of its robustness across different populations and settings. This will go a long way toward the development of diagnosis and treatment in depression and other disorders by including such factors into the proposed framework of LG-GNN for more refined and diversified applications in various clinical scenarios.

# APPENDIX

## A: DATASET

### A.1. Acquisition parameters

One of the significant challenges encountered in this study was the variability in acquisition parameters across different sites, as outlined in the table below (Table A 1). Differences in parameters such as repetition time, number of volumes, and total scan time introduced heterogeneity, which could potentially confound the analysis and limit the interpretability of the results. To address this, the decision was made to retain data only from the largest site (20), ensuring a more consistent dataset across key acquisition metrics. This approach aims to minimize variability and enhance the robustness of the findings by reducing inter-site discrepancies, ultimately allowing for a more reliable analysis of the data. Retaining the largest site data helps maintain homogeneity, which is crucial for deriving meaningful insights and ensuring the validity of conclusions drawn from the study.

| Site | Sample size | | TR (ms) | Number of volumes | Total Time (s) |
|---|---|---|---|---|---|
| | MDD | HC | | | |
| 1 | 73 | 73 | 2000 | 200 | 400 |
| 2 | 16 | 14 | 2000 | 190 | 380 |
| 7 | 35 | 37 | 2000 | 174 | 348 |
| 8 | 49 | 60 | 2000 | 190 | 380 |
| 9 | 48 | 48 | 2000 | 190 | 380 |
| 10 | 45 | 26 | 2000 | 202 | 404 |
| 11 | 20 | 17 | 2000 | 190 | 380 |
| 13 | 20 | 16 | 2500 | 140 | 350 |
| 14 | 61 | 32 | 2500 | 190 | 475 |
| 15 | 30 | 37 | 2000 | 230 | 460 |
| 17 | 41 | 41 | 2000 | 230 | 460 |
| 19 | 18 | 31 | 2000 | 230 | 460 |
| 20 | 246 | 225 | 2000 | 232 | 464 |
| 21 | 79 | 65 | 2000 | 230 | 460 |
| 22 | 18 | 20 | 2000 | 240 | 480 |
| 23 | 22 | 23 | 2000 | 230 | 460 |

***Table A 1*** *Data acquisition parameters across sites*

## A.2 Demographics and disease status

The dataset used in this study comprises a diverse demographic distribution, covering variations in sex, age, and disease status. The sex distribution includes a slightly larger representation of females compared to males, as shown in the corresponding chart (Figure A.1). Age is spread across a wide range, with a noticeable concentration in the 20-30 age group, though participants span from young adults to older individuals, ensuring broad age coverage (Figure A.2). In terms of disease status, the dataset provides a balanced foundation between MDD patients and HC for examining neural patterns associated with depression (Figure A.3).


***Figure A 1*** *Bar-chart of sex distribution*

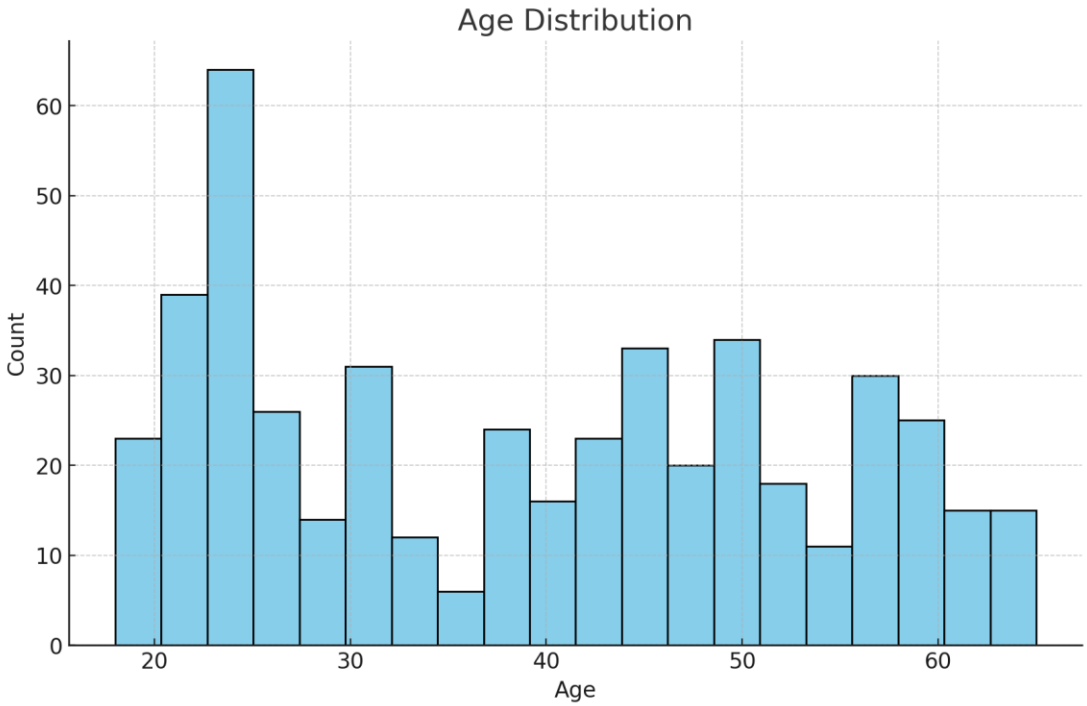


***Figure A 2*** *Histogram of age distribution*

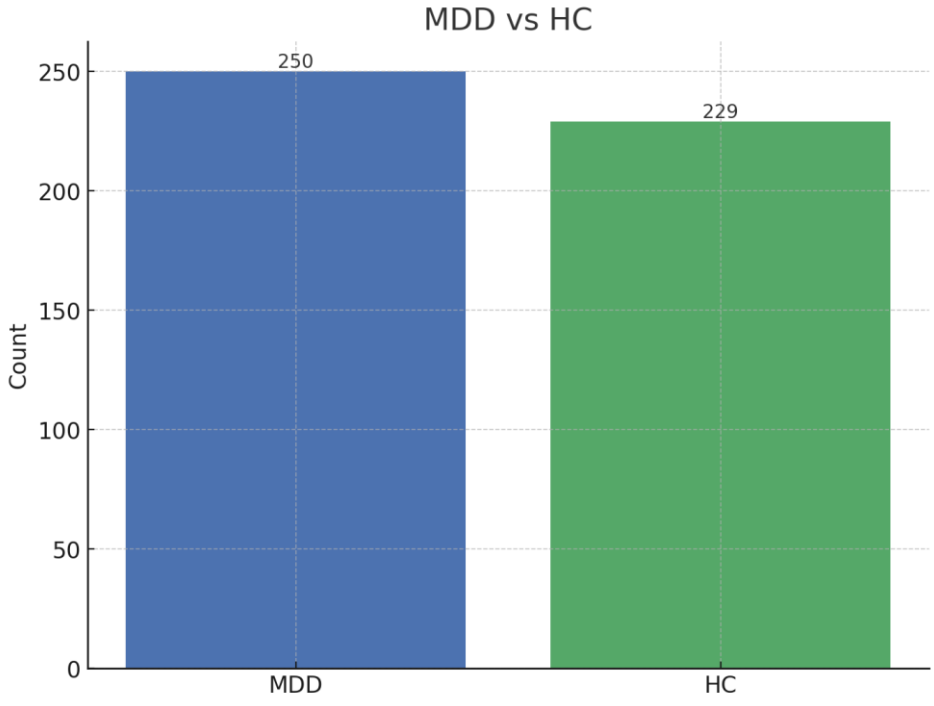


***Figure A 3*** *Bar-chart of disease status*

## B: Graph Structure and fMRI Data Representation

### B.1 Sample Connectivity Matrix

The sample connectivity matrix provides a visual representation of the Pearson correlation coefficients between various ROIs within the brain for a single subject (Figure B.1). This heatmap highlights the functional connectivity by showing the strength and direction of interactions between different brain regions, where each cell represents the correlation between two ROIs. Higher correlation values indicate stronger functional connectivity, potentially suggesting synchronized neural activity or similar functional roles between regions. By examining the patterns within this matrix, researchers can identify key areas of connectivity and explore how these patterns vary across individuals and conditions, such as in comparisons between HC and individuals with depression.

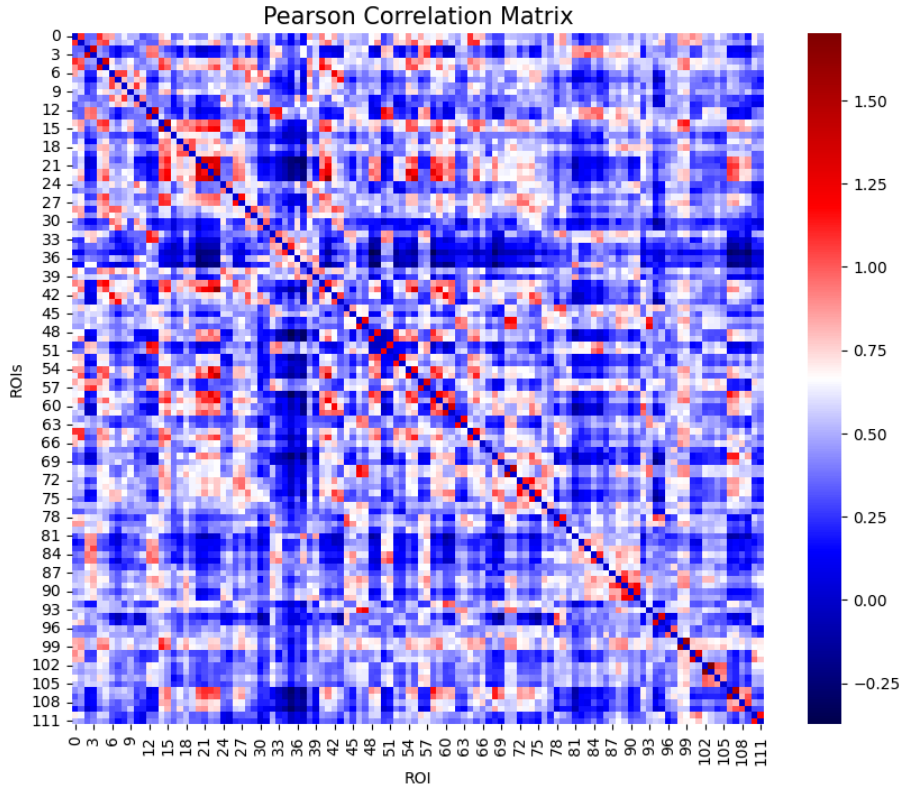


***Figure B 1*** *Sample Pearson correlation matrix*

## B.2 Visualization of global graph connectivity

The global graph connectivity visualization offers a network-based approach to understanding subject relationships in the dataset. In this graph (Figure B.2), each node represents an individual subject, while edges connecting nodes are weighted by similarity scores, which reflect the degree of similarity in connectivity profiles between subjects. This approach enables the clustering of subjects with similar connectivity patterns, potentially revealing subgroups within the population.

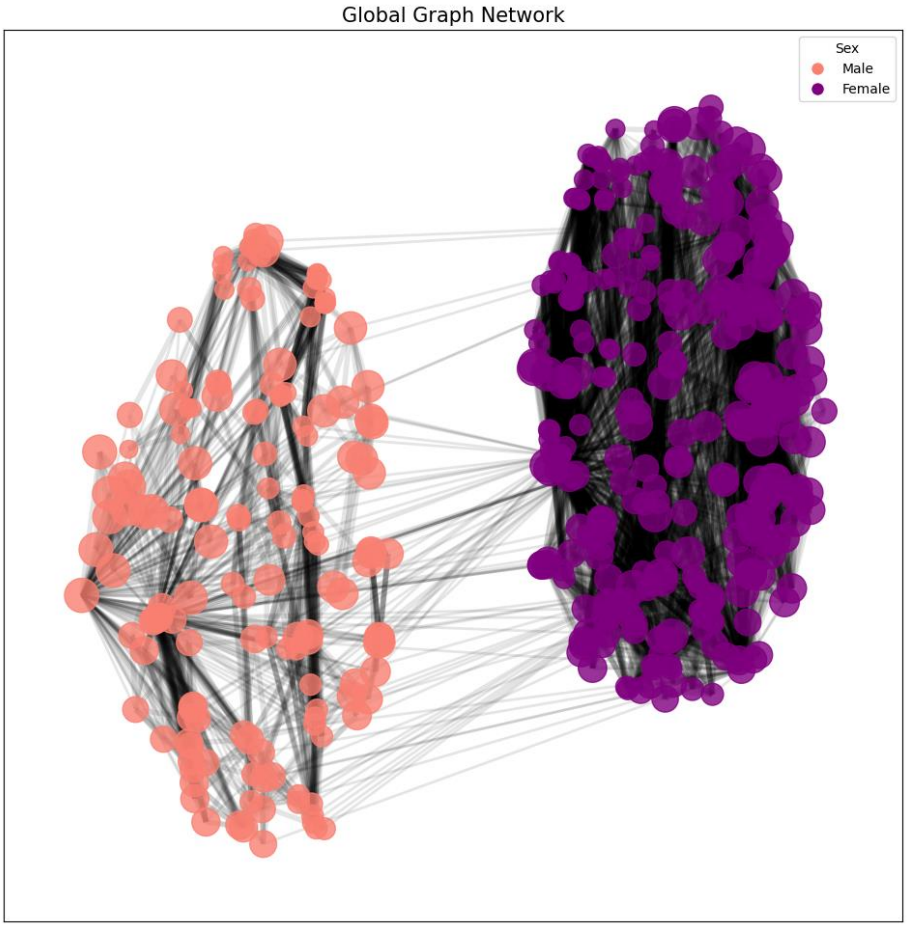


***Figure B 2*** *Graph visualization of the global-network representation*

## B.3 Visualization of local graph connectivity

The brain network connectivity graph (Figure B.3) provides a functional representation of neural activity, where nodes represent specific ROIs within the brain, and edges indicate functional connections based on observed correlations. These connections represent the strength and direction of the relationship between ROIs, forming a complex network that reflects the underlying brain architecture and dynamics.

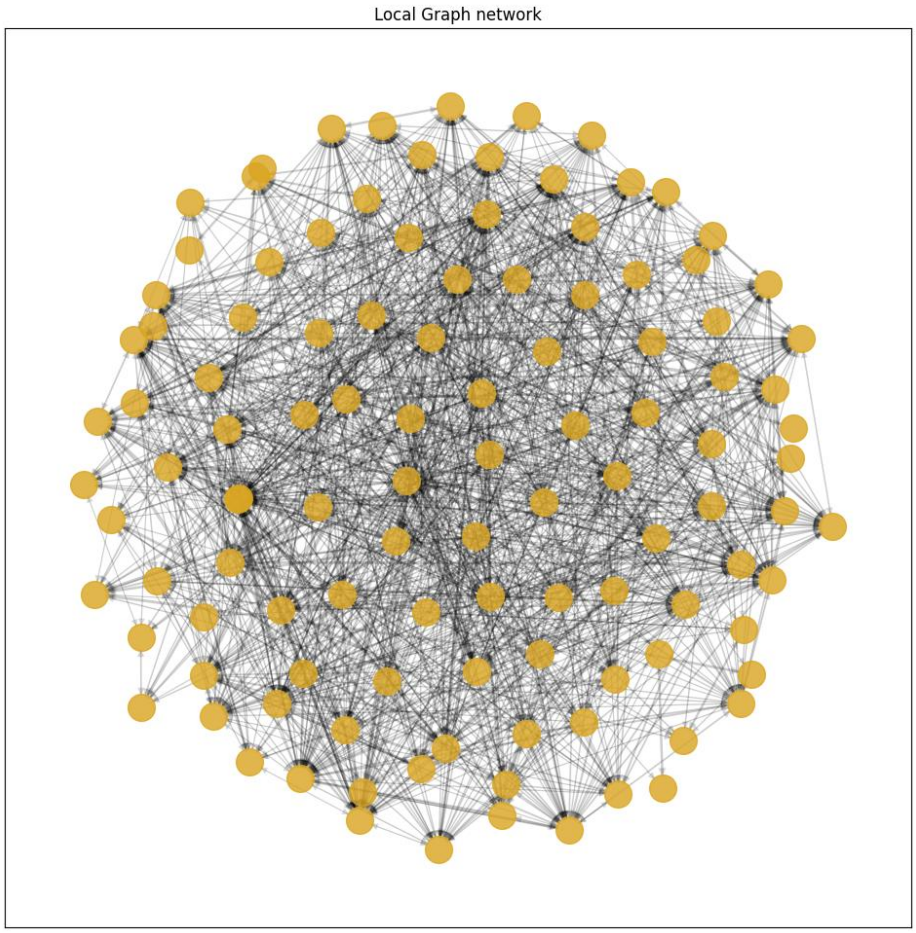


***Figure B 3*** *Graph visualization of the local-network representation*

# C: Model Performance Metrics

## C.1 Training and Validation Loss Curves

In this section we present the training and validation loss curves obtained across a 5-fold cross-validation procedure with early stopping. For each fold, the model's training loss shows a consistent downward trend, indicating effective learning, while the validation loss initially decreases before reaching a plateau or starts to increase, signaling potential overfitting. Early stopping was applied based on validation loss, halting training at the point where the validation loss stopped improving to prevent overfitting and promote generalizability. Additionally, the model was saved at the epoch where validation accuracy was highest, ensuring that the best-performing version of the model, in terms of accuracy, was retained for each fold. These curves (Figure C.1) provide insight into the training dynamics and underscore the importance of both early stopping for validation loss and selective model saving for validation accuracy to achieve an optimal balance between model complexity and generalization performance.

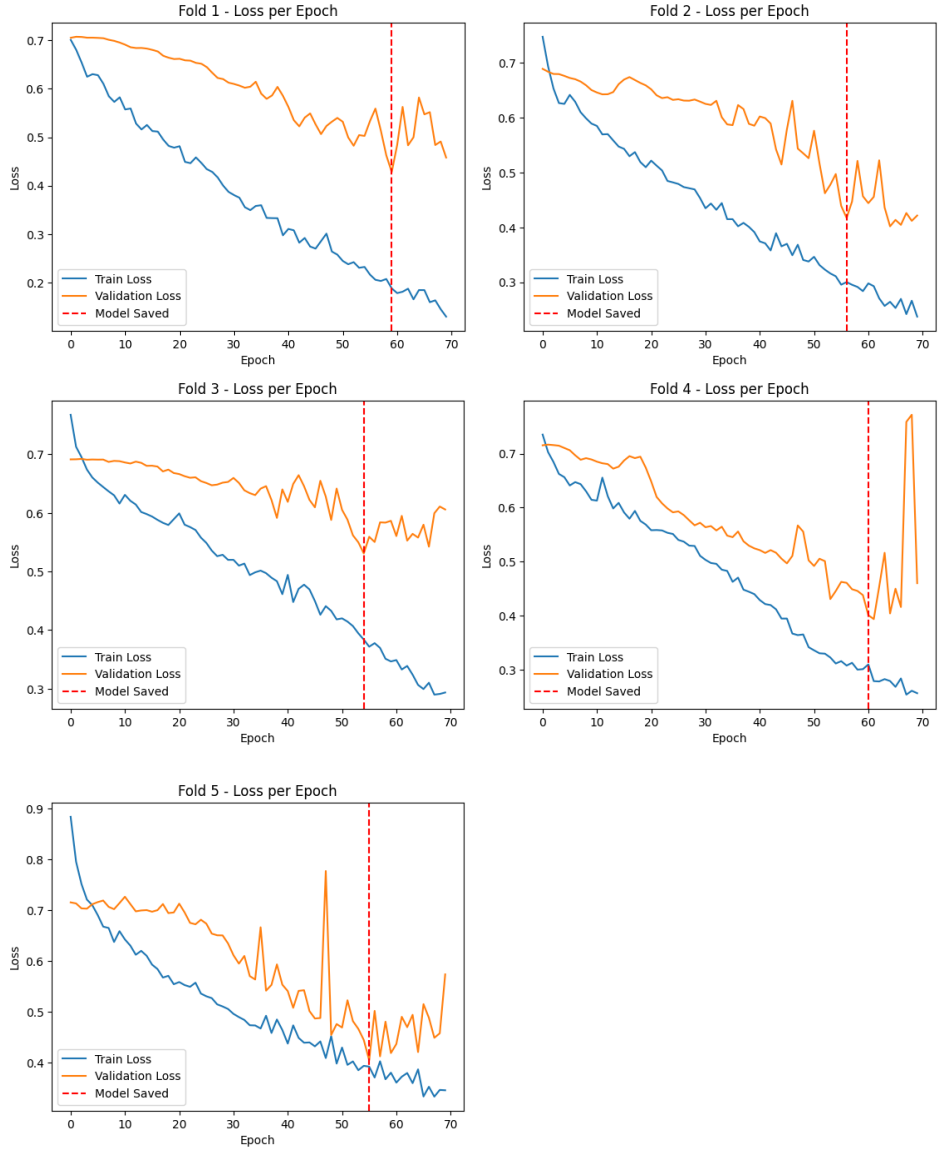


***Figure C 1*** *Training and Validation Loss Curves of proposed model*

## C.2 Confusion Matrices

In this section, we present confusion matrices that provide a detailed breakdown of the model's performance across each class. These matrices illustrate the counts of true positives, true negatives, false positives, and false negatives, allowing for a more granular assessment of the model's classification accuracy beyond overall accuracy metrics. By examining the distribution of misclassifications, we can identify specific classes where the model performs well and others where it may struggle, highlighting areas for potential improvement. The confusion matrices thus offer valuable insight into

the model's strengths and weaknesses in distinguishing between different categories, aiding in the interpretation of its predictive reliability.

### *C.2.1 Global-GNN Variants Comparisons*

The confusion matrices presented below (Figure C.2.1) for the Global-GNN model variations provide a clear visualization of classification performance across different configurations. This comparison helps highlight how adjustments to the model architecture impact its ability to correctly classify instances, providing insights into the robustness for the Global-GNN model.

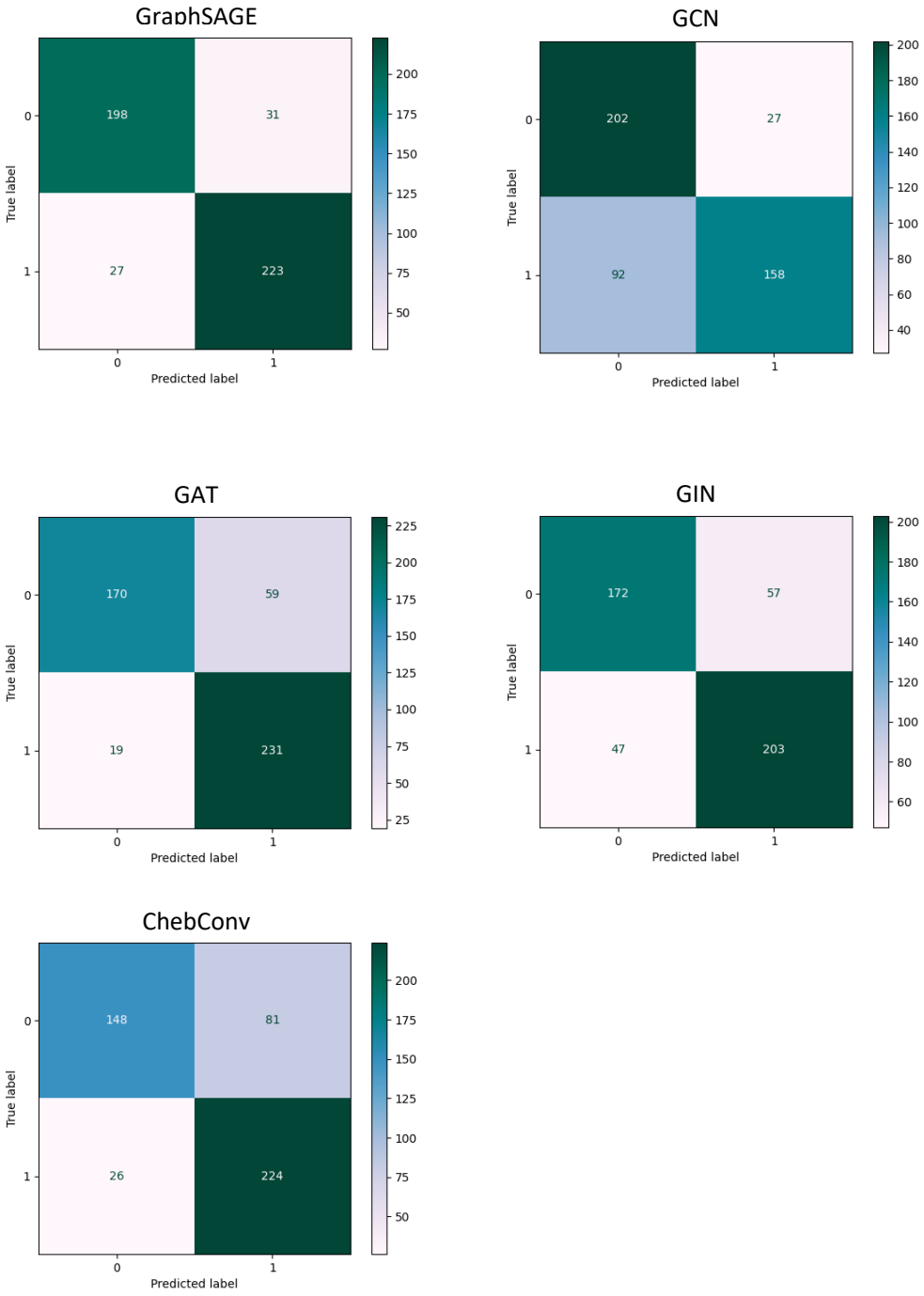


***Figure C 2 1*** *Confusion matrices of Global-GNN variations*

### *C.2.2 Local-GNN Variant Comparisons*

The confusion matrices below (Figure C.2.2) for the Local-GNN model variations offer a detailed view of classification performance across different model configurations.

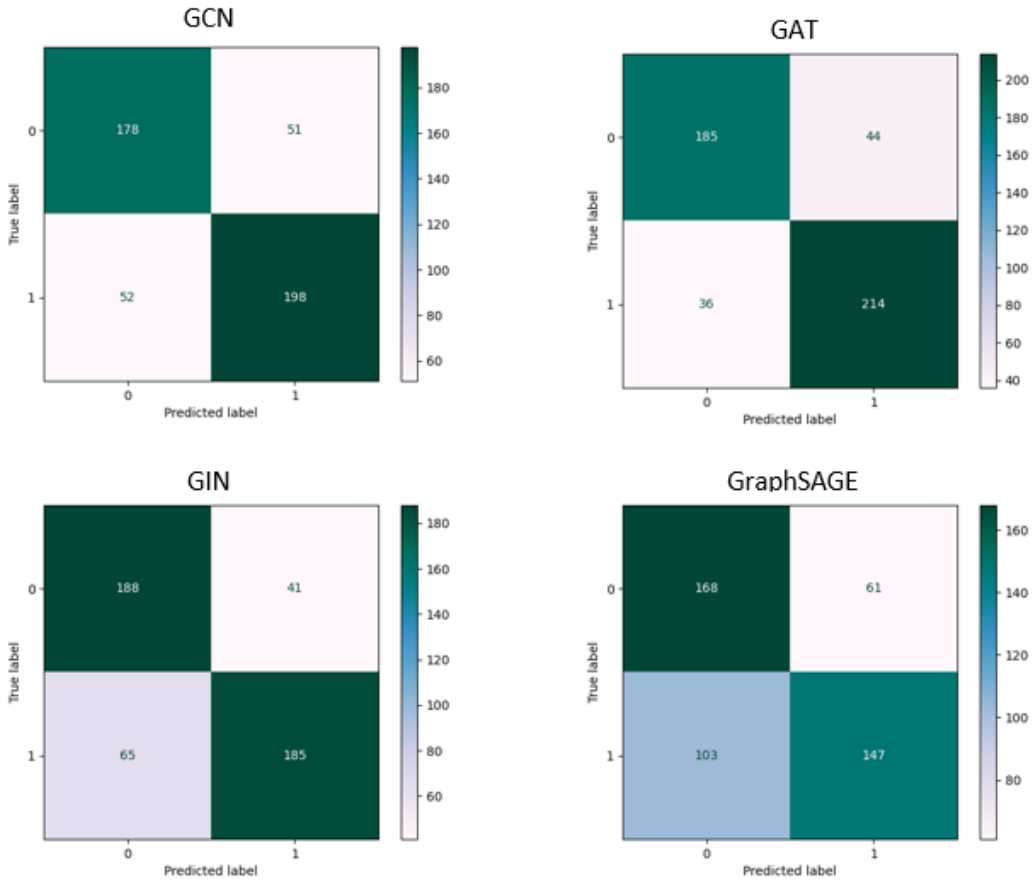


***Figure C 2 2*** *Confusion matrices of Local-GNN variations*

### *C.2.3 Baseline Comparisons*

The confusion matrices of the baseline models (Figure C.2.3) provide insight into their classification performance across each class. Comparing these matrices with those of the GNN-based models reveals how graph-based approaches may enhance classification performance, particularly in distinguishing nuanced patterns that baseline models may overlook. This comparison underscores the limitations of traditional models and highlights the strengths of GNN architectures in capturing complex, relational data.

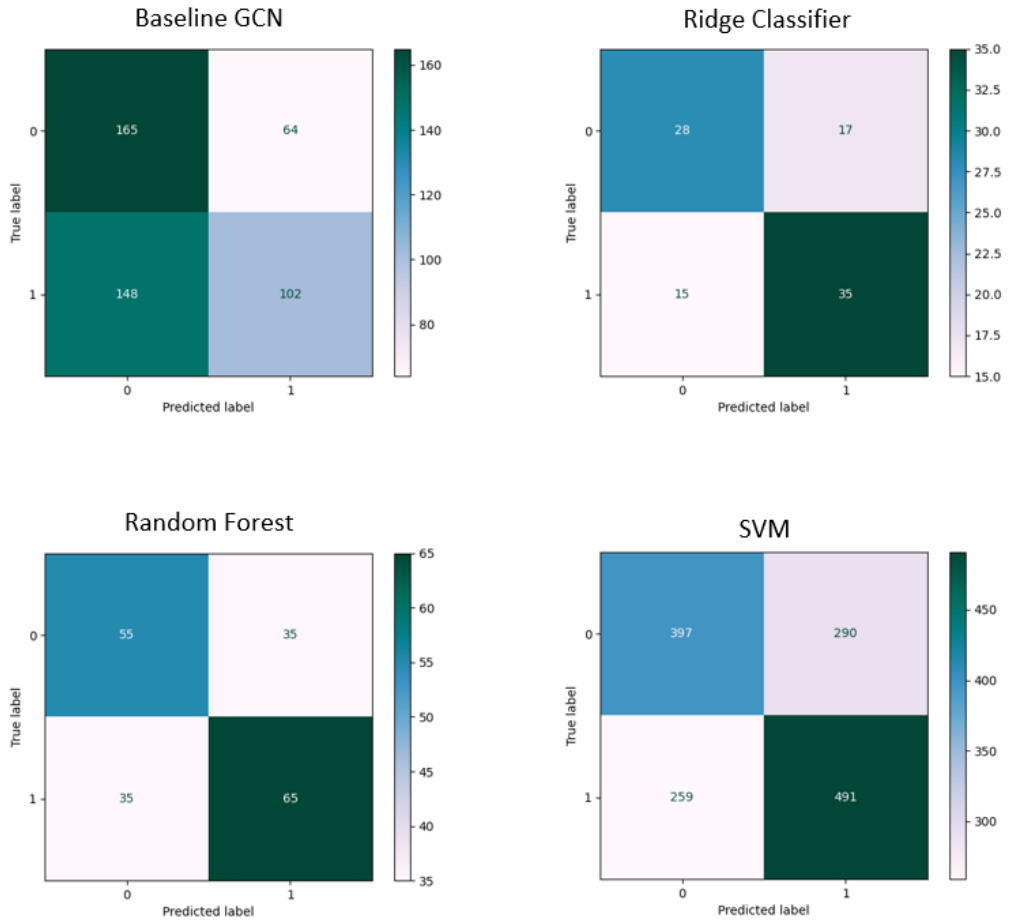


***Figure C 2 3*** *Confusion matrices of baseline models*

## C.3. ROC Curves

The ROC curves provide a comprehensive evaluation of each model's ability to distinguish between classes across various decision thresholds. By analyzing the true positive rate (sensitivity) versus the false positive rate at different thresholds, we can assess each model's classification performance in terms of both sensitivity and specificity. The area under the curve (AUC) for each model serves as a key metric, with higher AUC values indicating stronger discriminative capability. Comparing ROC curves across model variations allows us to observe how well each architecture captures the underlying patterns in the data.

### *C.3.1 Global-GNN variants Comparison*

The ROC curves for the different Global-GNN variants (Figure C.3.1) show how each configuration captures global connectivity patterns across the brain. Variants with higher AUC values demonstrate better class separation, underscoring the model's strength in identifying widespread patterns that differentiate classes.

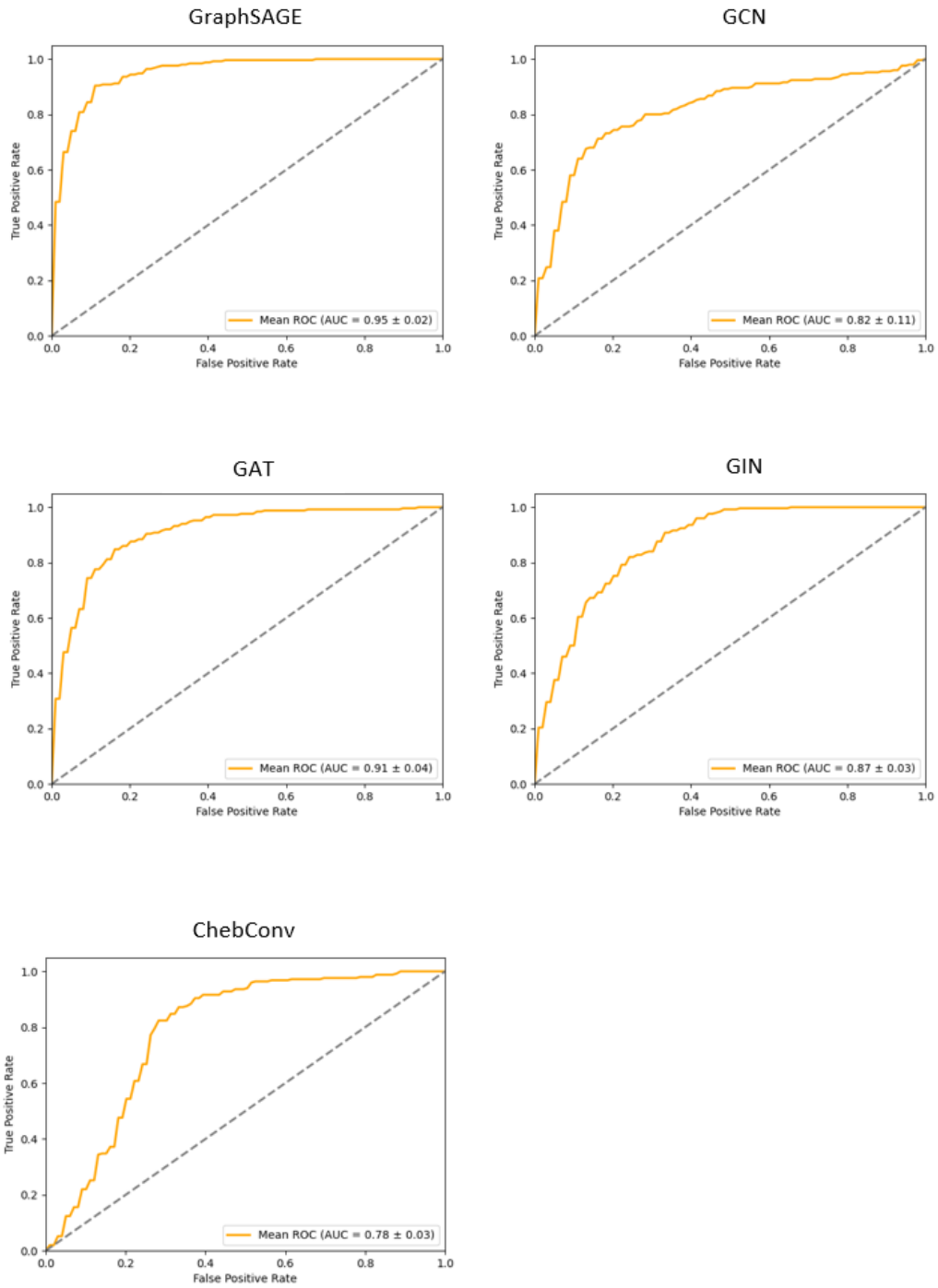


***Figure C 3 1*** *ROC curves of Global-GNN variations*

### *C.3.2 Local-GNN Variant Comparisons*

The ROC curves for the Local-GNN variants (Figure C.3.2) reflect the model's performance in capturing detailed regional connectivity information. These curves allow us to assess how each Local-GNN variant utilizes fine-grained regional data, providing insights into optimal parameter choices for enhanced local feature extraction.

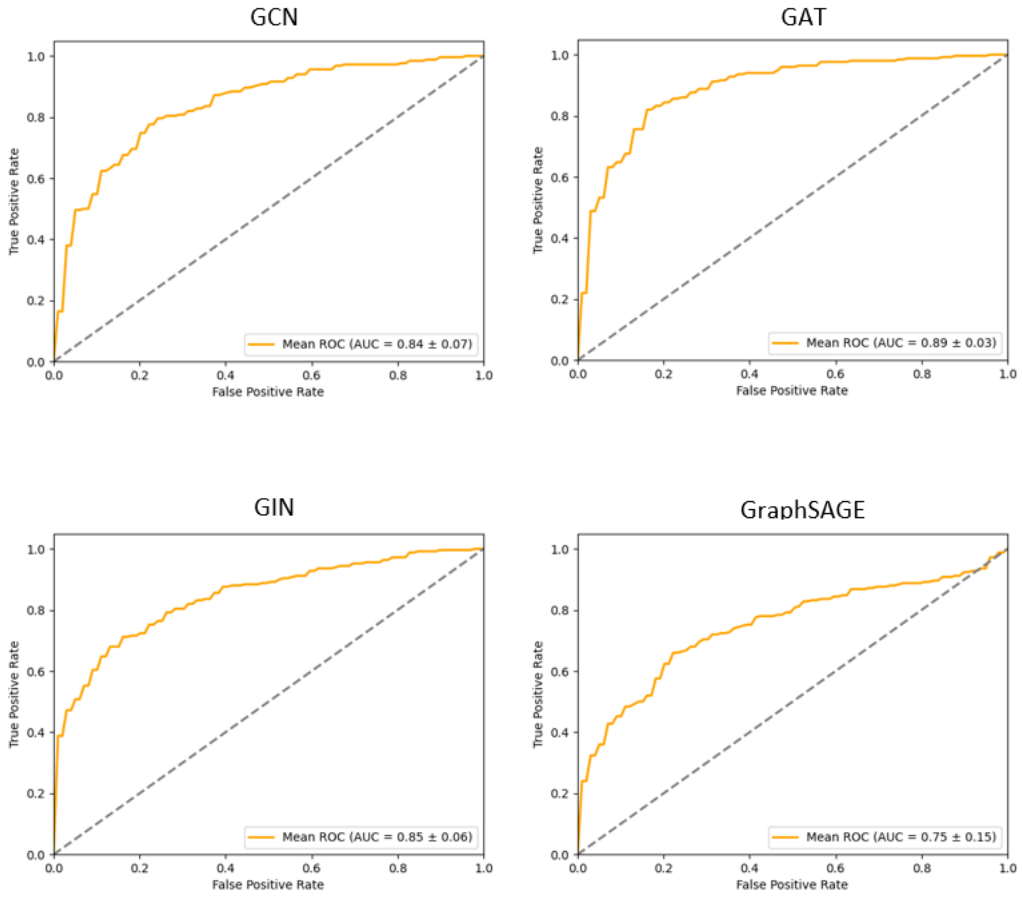


***Figure C 3 2*** *ROC curves of Local-GNN variations*

### *C.3.3 Baseline Comparisons*

The ROC curves for baseline model variants (Figure C.3.3) serve as benchmarks, with AUC values generally lower than those of the GNN-based variants. By comparing these curves, we can identify which baseline configurations perform relatively well, though they may lack the connectivity-based advantages of GNNs. This analysis highlights the limitations of traditional models in capturing complex relational information, reinforcing the added value that GNN-based approaches bring to classification performance.

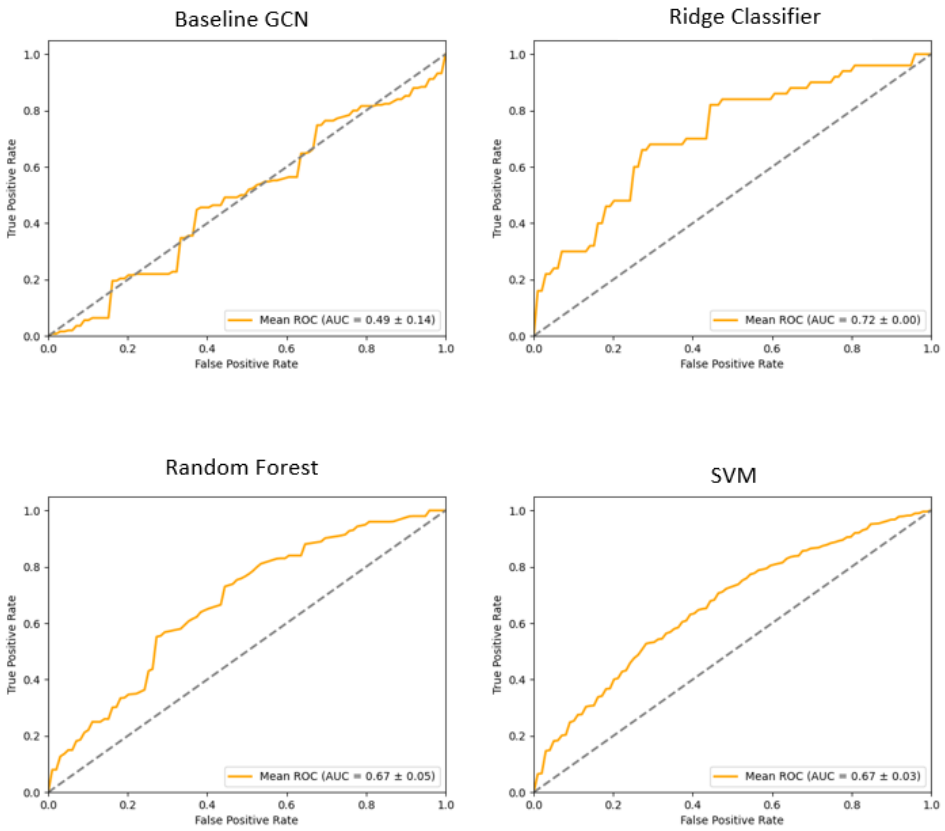


***Figure C 3 3*** *Roc curves of baseline models*

## D: BIOMARKERS

In this section, we highlight several brain regions (Table D.1) identified as prominent biomarkers for MDD classification. These include the **Frontal Pole** in both hemispheres (FP.L and FP.R), which is known for its role in executive function and emotional regulation. The **Temporal-Occipital** regions, specifically the right Middle Temporal Gyrus and Inferior Temporal Gyrus (TO2.R and TO3.R), were also significant, likely due to their involvement in visual processing and memory functions. Additionally, the **Superior Lateral Occipital Cortex** on both sides (OLs.L and OLs.R) and the **Inferior Lateral Occipital Cortex** on the right (OLi.R) were highlighted for their contributions to visual and spatial processing, areas often disrupted in MDD. The **Lingual Gyrus** (LG.L and LG.R), a region associated with visual and self-referential processing, and the **Precuneus** on the right (PCN.R), implicated in introspective thought, also emerged as key regions. Together, these regions form a network of biomarkers that support the classification of MDD, reflecting both cognitive and perceptual processing areas associated with the disorder.

| Rating | Region | Name | Lobe | Hemisphere |
|---|---|---|---|---|
| **1** | 71 | LG.R | Occipital | R |
| **2** | 0 | FP.L | Frontal | L |
| **3** | 1 | FP.R | Frontal | R |
| **4** | 31 | TO3.R | Temporal | R |
| **5** | 43 | OLs.R | Occipital | R |
| **6** | 42 | OLs.L | Occipital | L |
| **7** | 25 | TO2.R | Temporal | R |
| **8** | 70 | LG.L | Occipital | L |
| **9** | 45 | OLi.R | Occipital | R |
| **10** | 61 | PCN.R | Parietal | R |

***Table D 1*** *Most prominent brain regions that contributed to MDD classification*

# REFERENCES


Alzubi, J., Nayyar, A., & Kumar, A. (2018). Machine Learning from Theory to Algorithms: An Overview. *Journal of Physics Conference Series*, *1142*, 012012. https://doi.org/10.1088/1742-6596/1142/1/012012

American Psychiatric Association. (2013). Diagnostic and Statistical Manual of Mental Disorders (5th ed.).

Angra, S., & Ahuja, S. (2017). Machine learning and its applications: A review. *2017 International Conference on Big Data Analytics and Computational Intelligence (ICBDAC)*, 57-60. https://doi.org/10.1109/ICBDACI.2017.8070809

Arthurs, O. J., & Boniface, S. (2002). How well do we understand the neural origins of the fMRI BOLD signal? *Trends in Neurosciences*, *25*(1), 27–31. https://doi.org/10.1016/s0166-2236(00)01995-0

Bandettini, P. A. (2012). The BOLD Plot Thickens: Sign- and Layer-Dependent Hemodynamic Changes with Activation. Neuron, 76(3), 468–469. https://doi.org/10.1016/j.neuron.2012.10.026

Baştanlar, Y., & Özuysal, M. (2013). Introduction to Machine Learning. *Methods in Molecular Biology*, 105–128. https://doi.org/10.1007/978-1-62703-748-8_7

Bishop, C. M. (2016). *Pattern Recognition and Machine Learning*. Springer.

Biswal, B., Yetkin, F. Z., Haughton, V. M., & Hyde, J. S. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar mri. Magnetic Resonance in Medicine, 34(4), 537–541. https://doi.org/10.1002/mrm.1910340409

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5-32.

Breiman, L. (2001). Statistical Modeling: The Two Cultures (with comments and a rejoinder by the author). Statistical Science, 16(3). https://doi.org/10.1214/ss/1009213726

Buxton, R. B. (2013). The physics of functional magnetic resonance imaging (fMRI). Reports on Progress in Physics, 76(9), 096601. https://doi.org/10.1088/0034-4885/76/9/096601

Calarco, C. A., & Lobo, M. K. (2020). Depression and substance use disorders: Clinical comorbidity and shared neurobiology. International Review of Neurobiology, 245–309. https://doi.org/10.1016/bs.irn.2020.09.004

Caspi, A., Sugden, K., Moffitt, T. E., et al. (2003). Influence of life stress on depression: moderation by a polymorphism in the 5-HTT gene. Science, 301(5631), 386-389.

Chang, C. (2015). Deep and Shallow Architecture of Multilayer Neural Networks. *IEEE Transactions on Neural Networks and Learning Systems*, *26*(10), 2477–2486. https://doi.org/10.1109/tnnls.2014.2387439

Chao-Gan Yan, Xiao Chen, Le Li, et al. (2022) . Data of the REST-meta-MDD Project from DIRECT Consortium. V2. Science Data Bank. https://doi.org/10.57760/sciencedb.o00115.00013.

Chekroud, A. M., Bondar, J., Delgadillo, J., Doherty, G., Wasil, A., Fokkema, M., Cohen, Z., Belgrave, D., DeRubeis, R., Iniesta, R., Dwyer, D., & Choi, K. (2021). The promise of machine learning in predicting treatment outcomes in psychiatry. World Psychiatry, 20(2), 154–170. https://doi.org/10.1002/wps.20882

Chen, Y., Yan, J., Jiang, M., Zhang, T., et al. (2022). Adversarial learning based node-edge graph attention networks for autism spectrum disorder identification. IEEE Transactions on Neural Networks and Learning Systems.

Cheng, W., Rolls, E. T., Qiu, J., Yang, D., Ruan, H., Wei, D., Zhao, L., Meng, J., Xie, P., & Feng, J. (2018). Functional Connectivity of the Precuneus in Unmedicated Patients With Depression. *Biological Psychiatry Cognitive Neuroscience and Neuroimaging*, *3*(12), 1040–1049. https://doi.org/10.1016/j.bpsc.2018.07.008

Chiari-Correia, R. D., Tumas, V., Santos, A. C., & Salmon, C. E. G. (2023). Structural and functional differences in the brains of patients with MCI with and without depressive symptoms and their relations with Alzheimer's disease: an MRI study. *Deleted Journal*, *3*. https://doi.org/10.1093/psyrad/kkad008

Cooney, R. E., Joormann, J., Eugène, F., Dennis, E. L., & Gotlib, I. H. (2010). Neural correlates of rumination in depression. *Cognitive Affective & Behavioral Neuroscience*, *10*(4), 470–478. https://doi.org/10.3758/cabn.10.4.470

Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), 273-297.

Couvy-Duchesne, B., Strike, L. T., De Zubicaray, G. I., McMahon, K. L., Thompson, P. M., Hickie, I. B., Martin, N. G., & Wright, M. J. (2018). Lingual Gyrus Surface Area Is Associated with Anxiety-Depression Severity in Young Adults: A Genetic Clustering Approach. *eNeuro*, *5*(1), ENEURO.0153-17.2017. https://doi.org/10.1523/eneuro.0153-17.2017

Cusack, C. E., Ralph-Nearman, C., Christian, C., Fisher, A. J., & Levinson, C. A. (2024). Understanding heterogeneity, comorbidity, and variability in depression: Idiographic models and depression outcomes. Journal of Affective Disorders, 356, 248–256. https://doi.org/10.1016/j.jad.2024.04.034

Dai, L., Zhou, H., Xu, X., & Zuo, Z. (2019). Brain structural and functional changes in patients with major depressive disorder: a literature review. PeerJ, 7, e8170. https://doi.org/10.7717/peerj.8170

Dai, P., Lu, D., Shi, Y., Zhou, Y., Xiong, T., Zhou, X., Chen, Z., Zou, B., Tang, H., Huang, Z., & Liao, S. (2023). Classification of recurrent major depressive disorder using a new time series feature extraction method through multisite rs-fMRI data. Journal of Affective Disorders.

Davidson, R. J., Pizzagalli, D., Nitschke, J. B., & Putnam, K. (2002). Depression: Perspectives from Affective Neuroscience. Annual Review of Psychology, 53(1), 545–574.

Defferrard, M., Bresson, X., & Vandergheynst, P. (2016). Convolutional Neural Networks on Graphs with Fast Localized Spectral Filtering. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1606.09375

DeMaster, D., Godlewska, B. R., Liang, M., Vannucci, M., Bockmann, T., Cao, B., & Selvaraj, S. (2022). Effective connectivity between resting-state networks in depression. Journal of Affective Disorders, 307, 79–86. https://doi.org/10.1016/j.jad.2022.03.041

Drevets, W. C., Price, J. L., & Furey, M. L. (2008). Brain structural and functional abnormalities in mood disorders: Implications for neurocircuitry models of depression. Brain Structure and Function, 213(1-2), 93-118.

Du, Y., Luo, Q., Zou, Y., Nie, H., Li, Y., Lin, X., Shang, H., & Peng, H. (2024). Resting-State Brain Dynamics Unique to Anxiety in Major Depressive Disorder. *Depression and Anxiety*, *2024*(1). https://doi.org/10.1155/2024/4636291

Fasiello, E., Gorgoni, M., Scarpelli, S., Alfonsi, V., Strambi, L. F., & De Gennaro, L. (2021). Functional connectivity changes in insomnia disorder: A systematic review. Sleep Medicine Reviews, 61, 101569. https://doi.org/10.1016/j.smrv.2021.101569

Fettes, P. W., Moayedi, M., Dunlop, K., Mansouri, F., Vila-Rodriguez, F., Giacobbe, P., Davis, K. D., Lam, R. W., Kennedy, S. H., Daskalakis, Z. J., Blumberger, D. M., & Downar, J. (2017). Abnormal Functional Connectivity of Frontopolar Subregions in Treatment-Nonresponsive Major Depressive Disorder. *Biological Psychiatry Cognitive Neuroscience and Neuroimaging*, *3*(4), 337–347. https://doi.org/10.1016/j.bpsc.2017.12.003

Fisher, R. A. (1915). Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. Biometrika, 10(4), 507-521.

Fox, M. D., & Raichle, M. E. (2007). Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. Nature Reviews. Neuroscience, 8(9), 700–711. https://doi.org/10.1038/nrn2201

Gallo, S., El-Gazzar, A., Zhutovsky, P., & Thomas, R. M. (2023). Functional connectivity signatures of major depressive disorder: Machine learning analysis of two multicenter neuroimaging studies. Molecular Psychiatry, 28(1), 152-164.

Ghazaryan, A. (2023, July 14). What is a Graph Database? *Memgraph*. Retrieved from https://memgraph.com/blog/what-is-a-graph-database

García, V., Sánchez, J. S., & Marqués, A. I. (2019). Synergetic Application of Multi-Criteria Decision-Making Models to Credit Granting Decision Problems. *Applied Sciences*, *9*(23), 5052. https://doi.org/10.3390/app9235052

Gillespie, C. F., Szabo, S. T., & Nemeroff, C. B. (2020). Unipolar depression. In Elsevier eBooks (pp. 613–631). https://doi.org/10.1016/b978-0-12-813866-3.00036-9

Hamilton, W. L., Ying, R., & Leskovec, J. (2017). Inductive Representation Learning on Large Graphs. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1706.02216

Haykin, S. (1998). *Neural networks: A comprehensive foundation*. Prentice Hall.

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. Technometrics, 12(1), 55-67.

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, *349*(6245), 255–260. https://doi.org/10.1126/science.aaa8415

Kaiser, R. H., Andrews-Hanna, J. R., Wager, T. D., & Pizzagalli, D. A. (2015). Large-Scale Network Dysfunction in Major Depressive Disorder. *JAMA Psychiatry*, *72*(6), 603. https://doi.org/10.1001/jamapsychiatry.2015.0071

Katayama, N., Nakagawa, A., Umeda, S., Terasawa, Y., Kurata, C., Tabuchi, H., Kikuchi, T., & Mimura, M. (2019). Frontopolar cortex activation associated with pessimistic future-thinking in adults with major depressive disorder. *NeuroImage Clinical*, *23*, 101877. https://doi.org/10.1016/j.nicl.2019.101877

Kawakami, S., Okada, N., Satomura, Y., Shoji, E., Mori, S., Kiyota, M., Omileke, F., Hamamoto, Y., Morita, S., Koshiyama, D., Yamagishi, M., Sakakibara, E., Koike, S., & Kasai, K. (2024). Frontal pole–precuneus connectivity is associated with a discrepancy between self-rated and observer-rated depression severity in mood disorders: a resting-state functional magnetic resonance imaging study. *Cerebral Cortex*, *34*(7). https://doi.org/10.1093/cercor/bhae284

Kim, G. W., Farabaugh, A. H., Vetterman, R., Holmes, A., Nyer, M., Nasiriavanaki, Z., Fava, M., & Holt, D. J. (2022). Diminished frontal pole size and functional connectivity in young adults with high suicidality. *Journal of Affective Disorders, 310*, 484-492.

Kipf, T. N., & Welling, M. (2016). Semi-Supervised Classification with Graph Convolutional Networks. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1609.02907

Kocsis, K., Holczer, A., Kazinczi, C., Boross, K., Horváth, R., Németh, L. V., Klivényi, P., Kincses, Z. T., & Must, A. (2021). Voxel-based asymmetry of the regional gray matter over the inferior temporal gyrus correlates with depressive symptoms in medicated patients with major depressive disorder. Psychiatry Research Neuroimaging, 317, 111378. https://doi.org/10.1016/j.pscychresns.2021.111378

Korolev, S., Safiullin, A., Belyaev, M., & Dodonova, Y. (2017). Residual and plain convolutional neural networks for 3D brain MRI classification. International Symposium on Biomedical Imaging (ISBI), 835-838.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Lee, J. S., Kang, W., Kang, Y., Kim, A., Han, K., Tae, W., & Ham, B. (2021). Alterations in the Occipital Cortex of Drug-Naïve Adults With Major Depressive Disorder: A Surface-Based Analysis of Surface Area and Cortical Thickness. *Psychiatry Investigation*, *18*(10), 1025–1033. https://doi.org/10.30773/pi.2021.0099

Li, X., Zhou, Y., Dvornek, N., Zhang, M., Gao, S., Zhuang, J., Scheinost, D., & Duncan, J. S. (2021). **BrainGNN: Interpretable brain graph neural network for fMRI analysis.** *Medical Image Analysis*, 74, 102233. https://doi.org/10.1016/j.media.2021.102233

Liu, C., Ma, X., Yuan, Z., Song, L., Jing, B., Lu, H., Tang, L., Fan, J., Walter, M., Liu, C., Wang, L., & Wang, C. (2017). Decreased Resting-State Activity in the Precuneus Is Associated With Depressive Episodes in Recurrent Depression. *The Journal of Clinical Psychiatry*, *78*(04), e372–e382. https://doi.org/10.4088/jcp.15m10022

Liu, X., Chen, W., Hou, H., Chen, X., Zhang, J., Liu, J., Guo, Z., & Bai, G. (2017). Decreased functional connectivity between the dorsal anterior cingulate cortex and lingual gyrus in Alzheimer's disease patients with depression. *Behavioural Brain Research*, *326*, 132–138. https://doi.org/10.1016/j.bbr.2017.01.037

Liu, X., Chen, W., Tu, Y., Hou, H., Huang, X., Chen, X., Guo, Z., Bai, G., & Chen, W. (2018). The Abnormal Functional Connectivity between the Hypothalamus and the Temporal Gyrus Underlying Depression in Alzheimer's Disease Patients. *Frontiers in Aging Neuroscience*, *10*. https://doi.org/10.3389/fnagi.2018.00037

Logothetis, N. K. (2008). What we can do and what we cannot do with fMRI. Nature, 453(7197), 869-878.

Luo, L., Lei, X., Zhu, C., Wu, J., Ren, H., Zhan, J., & Qin, Y. (2022). Decreased Connectivity in Precuneus of the Ventral Attentional Network in First-Episode, Treatment-Naïve Patients With Major Depressive

Disorder: A Network Homogeneity and Independent Component Analysis. *Frontiers in Psychiatry*, *13*. https://doi.org/10.3389/fpsyt.2022.925253

Ma, C., Ding, J., Li, J., Guo, W., Long, Z., Liu, F., Gao, Q., Zeng, L., Zhao, J., & Chen, H. (2012). Resting-State Functional Connectivity Bias of Middle Temporal Gyrus and Caudate with Altered Gray Matter Volume in Major Depression. *PLoS ONE*, *7*(9), e45263. https://doi.org/10.1371/journal.pone.0045263

Mahesh, B. (2020). Machine Learning Algorithms - A Review. *International Journal of Science and Research (IJSR)*, *9*(1), 381–386. https://doi.org/10.21275/art20203995

Majeed, A., & Rauf, I. (2020). Graph Theory: A Comprehensive Survey about Graph Theory Applications in Computer Science and Social Networks. *Inventions*, *5*(1), 10. https://doi.org/10.3390/inventions5010010

Makris, N., Goldstein, J. M., Kennedy, D., Hodge, S. M., Caviness, V. S., Faraone, S. V., Tsuang, M. T., & Seidman, L. J. (2006). Decreased volume of left and total anterior insular lobule in schizophrenia. *Schizophrenia Research*, *83*(2–3), 155–171. https://doi.org/10.1016/j.schres.2005.11.020

Malekian, A., & Chitsaz, N. (2021). Concepts, procedures, and applications of artificial neural network models in streamflow forecasting. In *Elsevier eBooks* (pp. 115–147). https://doi.org/10.1016/b978-0-12-820673-7.00003-2

Maller, J. J., Thomson, R. H. S., Rosenfeld, J. V., Anderson, R., Daskalakis, Z. J., & Fitzgerald, P. B. (2014). Occipital bending in depression. *Brain*, *137*(6), 1830–1837. https://doi.org/10.1093/brain/awu072

Menon, V. (2011). Large-scale brain networks and psychopathology: a unifying triple network model. Trends in Cognitive Sciences, 15(10), 483–506. https://doi.org/10.1016/j.tics.2011.08.003

Nasteski, V. (2017). An overview of the supervised machine learning methods. *HORIZONS B*, *4*, 51–62. https://doi.org/10.20544/horizons.b.04.1.17.p05

Niepert, M., Ahmed, M., & Kutzkov, K. (2016). Learning convolutional neural networks for graphs. Proceedings of the 33rd International Conference on Machine Learning, 2016, 2014-2023.

Ogawa, S., Lee, T. M., Kay, A. R., & Tank, D. W. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. Proceedings of the National Academy of Sciences, 87(24), 9868-9872.

Orsolini, L., Latini, R., Pompili, M., Serafini, G., Volpe, U., Vellante, F., Fornaro, M., Valchera, A., Tomasetti, C., Fraticelli, S., Alessandrini, M., La Rovere, R., Trotta, S., Martinotti, G., Di Giannantonio, M., & De Berardis, D. (2020). Understanding the Complex of Suicide in Depression: from Research to Clinics. Psychiatry Investigation, 17(3), 207–221. https://doi.org/10.30773/pi.2019.0171

Pace-Schott, E. F., & Picchioni, D. (2016). Neurobiology of Dreaming. In Elsevier eBooks (pp. 529-538.e6). https://doi.org/10.1016/b978-0-323-24288-2.00051-9

Parisot, S., Ktena, S. I., Ferrante, E., Lee, M., Guerrero, R., & Glocker, B. (2018). Disease prediction using graph convolutional networks: Application to autism spectrum disorder and Alzheimer's disease. Medical Image Analysis, 48, 117-130.

Patel, M. J., Khalaf, A., Aizenstein, H. J., & Phillips, M. L. (2016). Machine learning approaches for integrating clinical and imaging features in late-life depression classification. Journal of Affective Disorders, 208, 1-12.

Pearson, K. (1895). Notes on regression and inheritance in the case of two parents. Proceedings of the Royal Society of London, 58, 240-242.

Pitsik, E. N., Maximenko, V. A., Kurkin, S. A., & Sergeev, A. P. (2023). The topology of fMRI-based networks defines the performance of a graph neural network for the classification of patients with major depressive disorder. Chaos, Solitons & Fractals, 170, 112532.

Qin, K., Lei, D., Pinaya, W. H. L., Pan, N., Li, W., Zhu, Z., Sweeney, J. A., Mechelli, A., & Gong, Q. (2022). Using graph convolutional network to characterize individuals with major depressive disorder across multiple imaging sites. EBioMedicine, 77, 103896

Rajula, H. S. R., Verlato, G., Manchia, M., Antonucci, N., & Fanos, V. (2020). Comparison of Conventional Statistical Methods with Machine Learning in Medicine: Diagnosis, Drug Development, and Treatment. Medicina, 56(9), 455. https://doi.org/10.3390/medicina56090455

Ramezani, M., Johnsrude, I., Rasoulian, A., Bosma, R., Tong, R., Hollenstein, T., Harkness, K., & Abolmaesumi, P. (2014). Temporal-lobe morphology differs between healthy adolescents and those with early-onset of depression. *NeuroImage Clinical*, *6*, 145–155. https://doi.org/10.1016/j.nicl.2014.08.007

Samuel, A. L. (1969). *Some Studies in Machine Learning Using the Game of Checkers*.

Sandhya N. dhage, Charanjeet Kaur Raina. (2016) A review on Machine Learning Techniques. In International Journal on Recent and Innovation Trends in Computing and Communication, Volume 4 Issue 3.

Scalabrini, A., Vai, B., Poletti, S., Damiani, S., Mucci, C., Colombo, C., Zanardi, R., Benedetti, F., & Northoff, G. (2020). All roads lead to the default-mode network—global source of DMN abnormalities in major depressive disorder. *Neuropsychopharmacology*, *45*(12), 2058–2069. https://doi.org/10.1038/s41386-020-0785-x

Seeley, W. W. (2019). The Salience Network: A Neural System for Perceiving and Responding to Homeostatic Demands. Journal of Neuroscience, 39(50), 9878–9882. https://doi.org/10.1523/jneurosci.1138-17.2019

Shahhosseini, Y., & Miranda, M. F. (2022). Functional Connectivity Methods and Their Applications in fMRI Data. *Entropy*, *24*(3), 390. https://doi.org/10.3390/e24030390

Sheline, Y. I., Barch, D. M., Price, J. L., Rundle, M. M., Vaishnavi, S. N., Snyder, A. Z., ... & Raichle, M. E. (2009). The default mode network and self-referential processes in depression. Proceedings of the National Academy of Sciences, 106(6), 1942-1947.

Sheline, Y. I., Price, J. L., Yan, Z., & Mintun, M. A. (2010). Resting-state functional MRI in depression unmasks increased connectivity between networks via the dorsal nexus. Proceedings of the National Academy of Sciences, 107(24), 11020-11025.

Song, X. M., Hu, X., Li, Z., Gao, Y., Ju, X., Liu, D., Wang, Q., Xue, C., Cai, Y., Bai, R., Tan, Z., & Northoff, G. (2021). Reduction of higher-order occipital GABA and impaired visual perception in acute major depressive disorder. *Molecular Psychiatry*, *26*(11), 6747–6755. https://doi.org/10.1038/s41380-021-01090-5

Sullivan, P. F., Neale, M. C., & Kendler, K. S. (2000). Genetic Epidemiology of Major Depression: Review and Meta-Analysis. American Journal of Psychiatry, 157(10), 1552-1562.

Taylor, J. E., Yamada, T., Kawashima, T., Kobayashi, Y., Yoshihara, Y., Miyata, J., Murai, T., Kawato, M., & Motegi, T. (2022). Depressive symptoms reduce when dorsolateral prefrontal cortex-precuneus connectivity normalizes after functional connectivity neurofeedback. *Scientific Reports*, *12*(1). https://doi.org/10.1038/s41598-022-05860-1

Thaipisuttikul, P., Ittasakul, P., Waleeprakhon, P., Wisajun, P., & Jullagate, S. (2014). Psychiatric comorbidities in patients with major depressive disorder. Neuropsychiatric Disease and Treatment, 2097. https://doi.org/10.2147/ndt.s72026

Uddin, L. Q., Yeo, B. T. T., & Spreng, R. N. (2019). Towards a Universal Taxonomy of Macro-scale Functional Human Brain Networks. Brain Topography, 32(6), 926–942. https://doi.org/10.1007/s10548-019-00744-6

Veer, I. M. (2010). Whole brain resting-state analysis reveals decreased functional connectivity in major depression. *Frontiers in Systems Neuroscience*, *4*. https://doi.org/10.3389/fnsys.2010.00041

Veličković, P. (2023). Everything is connected: Graph neural networks. *Current Opinion in Structural Biology*, *79*, 102538. https://doi.org/10.1016/j.sbi.2023.102538

Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., & Bengio, Y. (2017). Graph Attention Networks. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1710.10903

Venkatapathy, S., Votinov, M., Wagels, L., Kim, S., Lee, M., Habel, U., Ra, I-H., & Jo, H-G. (2023). Ensemble graph neural network model for classification of major depressive disorder using whole-brain functional connectivity. Journal of Affective Disorders.

Vieira, S., Pinaya, W. H. L., & Mechelli, A. (2017). Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: Methods and applications. Neuroscience & Biobehavioral Reviews, 74, 58-75.

Wang, X., Zhao, Y., & Pourpanah, F. (2020). Recent advances in deep learning. *International Journal of Machine Learning and Cybernetics*, *11*(4), 747–750. https://doi.org/10.1007/s13042-020-01096-5

Wen, D., Wei, Z., Li, G., Zhou, Y., Han, W., Zhang, X., & Zhang, L. (2023). Convolutional neural networks in rs-fMRI for major depressive disorder classification. Frontiers in Neuroscience, 17, 9922223.

Weng, J., Chou, Y., Tsai, Y., Lee, C., Hsieh, M., & Chen, V. (2019). Connectome Analysis of Brain Functional Network Alterations in Depressive Patients with Suicidal Attempt. *Journal of Clinical Medicine*, *8*(11), 1966. https://doi.org/10.3390/jcm8111966

Wickramarachchi, A. (2020, May 19). Logistic regression and decision boundary: Understanding logistic regression and its utility in classification. *Towards Data Science*. https://towardsdatascience.com/logistic-regression-and-decision-boundary-69e6a7a5e073

Xia, Z., Fan, Y., Li, K., Wang, Y., Huang, L., & Zhou, F. (2023). DepressionGraph: A Two-Channel Graph Neural Network for the Diagnosis of Major Depressive Disorders Using rs-fMRI. Journal of Neural Networks and Machine Learning.

Xiao, T., Chen, Z., Wang, D., & Wang, S. (2023). Learning How to Propagate Messages in Graph Neural Networks. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2310.00697

Xu, K., Hu, W., Leskovec, J., & Jegelka, S. (2018). How Powerful are Graph Neural Networks? *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1810.00826

Xue, H., Sun, X.-K., & Sun, W.-X. (2020). **Multi-hop hierarchical graph neural networks.** In *Proceedings of the IEEE International Conference on Big Data and Smart Computing (BigComp)* (pp. 82–89). https://doi.org/10.1109/BigComp.2020

Yan, C., Chen, X., Li, L., Castellanos, F. X., Bai, T., Bo, Q., Cao, J., Chen, G., Chen, N., Chen, W., Cheng, C., Cheng, Y., Cui, X., Duan, J., Fang, Y., Gong, Q., Guo, W., Hou, Z., Hu, L., . . . Zang, Y. (2019). Reduced default mode network functional connectivity in patients with recurrent major depressive disorder.

*Proceedings of the National Academy of Sciences*, *116*(18), 9078–9083. https://doi.org/10.1073/pnas.1900390116

Yang, J., Gohel, S., & Vachha, B. (2020). Current methods and new directions in resting state fMRI. *Clinical Imaging*, *65*, 47–53. https://doi.org/10.1016/j.clinimag.2020.04.004

Zeng, L. L., Shen, H., Liu, L., Wang, L., Li, B., Fang, P., ... & Hu, D. (2012). Identifying major depression using whole-brain functional connectivity: a multivariate pattern analysis. Brain, 135(5), 1493-1507.

Zhang, B., Qi, S., Liu, S., Liu, X., Wei, X., & Ming, D. (2021). Altered spontaneous neural activity in the precuneus, middle and superior frontal gyri, and hippocampus in college students with subclinical depression. *BMC Psychiatry*, *21*(1). https://doi.org/10.1186/s12888-021-03292-1

Zhang, H., Song, R., Wang, L., Zhang, L., Wang, D., Wang, C., & Zhang, W. (2022). Classification of Brain Disorders in rs-fMRI via Local-to-Global Graph Neural Networks. *IEEE Transactions on Medical Imaging*, *42*(2), 444–455. https://doi.org/10.1109/tmi.2022.3219260

Zhao, K., Duka, B., Xie, H., Oathes, D. J., Calhoun, V., & Zhang, Y. (2022). A dynamic graph convolutional neural network framework reveals new insights into connectome dysfunctions in ADHD. NeuroImage, 246, 118774. https://doi.org/10.1016/j.neuroimage.2021.118774.

Zhou, Y., Zheng, H., Huang, X., Hao, S., Li, D., & Zhao, J. (2022). Graph Neural Networks: Taxonomy, Advances, and Trends. *ACM Transactions on Intelligent Systems and Technology*, *13*(1), 1–54. https://doi.org/10.1145/3495161

Zhu, Z., Wang, Y., Lau, W. K. W., Wei, X., Liu, Y., Huang, R., & Zhang, R. (2022). Hyperconnectivity between the posterior cingulate and middle frontal and temporal gyrus in depression: Based on functional connectivity meta-analyses. *Brain Imaging and Behavior*, *16*(4), 1538–1551. https://doi.org/10.1007/s11682-022-00628-7