Late and Early Fusion Graph Neural Network Architectures for Integrative Modeling of Multimodal Brain Connectivity Graphs

Alessio Comparini[1](✉)[0009−0009−8557−1219], Léa Schmidt[1][0009−0005−5531−0895], Vanessa Siffredi[1][0000−0002−9137−0730], Damien Marie[2,3,4][0000−0001−8199−8260], Clara James[3,4∗][0000−0001−7480−0682], and Jonas Richiardi[1,6∗][0000−0002−6975−5634]

[1] Department of Radiology, Lausanne University Hospital and University of Lausanne, Lausanne, Switzerland alessio.comparini@chuv.ch
[2] University of Applied Sciences and Arts Western Switzerland HES-SO, Geneva School of Health Sciences, Geneva Musical Minds lab (GEMMI lab), Geneva, Switzerland
[3] CIBM Center for Biomedical Imaging, Cognitive and Affective Neuroimaging section, University of Geneva, Geneva, Switzerland
[4] Brain and Behaviour Laboratory, University of Geneva, Geneva, Switzerland
[5] Faculty of Psychology and Educational Sciences, University of Geneva, Geneva, Switzerland
[6] CIBM Center for Biomedical Imaging, Imaging for Precision Medicine Section, Lausanne, Switzerland
[∗] equal contribution

Abstract. The integration of structural and functional brain connectivity provides a holistic view of the brain's organization, but its application in Graph Neural Network (GNN) models for predicting "brain age" is understudied, and a systematic benchmark of optimal data fusion strategies is currently lacking. We systematically benchmark the performances of early and late fusion multimodal architectures against single-modality models for brain age prediction using structural and functional connectcomes, using five different GNN backbones on 747 healthy participants (median age, 16.3 years; IQR 13.5-18.5 years) obtained from the Philadelphia Neurodevelopmental Cohort. The late fusion architecture improved performance over the structural-only baseline in three of five models, with the GCN model achieving the highest overall score in cross-validation ($R^2 = 0.639 \pm 0.05$). The early fusion architecture showed inconsistent results and did not offer a reliable improvement over the single-modality baseline. Finally, is observed that optimal model architecture depends on the data type: structural brain graphs favors deep, narrow models to capture their hierarchy, whereas functional brain graphs requires wider, shallower models.

Keywords: machine learning · neuroimaging · connectome · multi-layer graph

1 Introduction

The human brain's capacities emerge from an intricate interplay between its physical structure and dynamic functional activity. Among the most prevalent imaging modalities used to model the brain as a network are Diffusion Weighted Imaging (DWI), which maps the white matter tracts to construct a structural connectome (SC), and functional Magnetic Resonance Imaging (fMRI), from which temporal correlations in blood-oxygen-level-dependent (BOLD) signals are calculated to derive a functional connectome (FC). While DWI provides the anatomical scaffold of potential communication pathways, it is a static representation. In contrast, fMRI captures the brain's dynamic functional organization but lacks direct information about the underlying anatomical pathways.

The integration of both modalities is a promising avenue for research, offering a more holistic view by combining the complementary strengths of each approach [25, 31, 24, 45, 44, 30]. The principle of structure-function coupling, which posits that the brain's anatomical structure constrains and shapes its functional dynamics, is central to this scientific pursuit. Seminal work has demonstrated that regions with strong structural connections often exhibit highly correlated functional activity, highlighting a direct relationship between the physical connectome and its functional expression [13]. However, this coupling is not unidirectional. Recent studies suggest that neural activity can also actively shape structural connectivity; for instance, frequently co-active regions may become more myelinated and thus form stronger, clearer anatomical connections in a process known as activity-dependent myelination [26]. Consequently, investigating this reciprocal relationship is fundamental to understand the neurobiological mechanisms that govern cognitive processes such as learning, memory, and emotion [30, 41, 8, 21, 12, 7, 28]. Furthermore, disruptions in this relationship are increasingly correlated with the pathophysiology of various neurological and psychiatric disorders, making integrative analyses critical for identifying robust imaging biomarkers [39, 38].

A powerful application of this framework is the prediction of "brain age", a composite biomarker reflecting the brain's maturational and aging trajectory. "Brain age" estimated from MRI reflects patterns of cortical atrophy, white matter loss, and network disruption, which are linked to cognitive decline and neurodegenerative risk [36, 37, 22, 32, 19]. Importantly, brain age is recognized as a significant confounding variable in studies aiming to predict cognitive scores, making its accurate estimation essential for disentangling the effects of normal development from pathological processes [29].

In recent years, Graph Neural Networks (GNNs) have emerged as a state-of-the-art method for analyzing brain connectomes. By design, GNNs can effectively model the complex, non-linear relationships inherent in graph-structured data, making them particularly well-suited for the non-Euclidean geometry of brain networks [23]. While numerous studies have successfully applied GNNs to clinical classification tasks using single-modality data [20, 16, 5], the application of multimodal GNNs to brain age prediction remains comparatively understudied. Existing multimodal studies have primarily focused on classifying brain disorders[6,

14, 42, 3, 11], often using small datasets and lacking direct comparisons with unimodal counterparts. Moreover, most of the literature in multimodal brain age prediction still relies on traditional machine learning algorithms [17, 27]. This has created significant gaps in our understanding: it is not yet clear whether combining SC and FC within a GNN framework enhances predictive performance for brain age, nor is it established which architectural or fusion strategies are optimal for this task[23]. The methodological heterogeneity in preprocessing, connectome construction, and model implementation further complicates cross-study comparisons and impedes the establishment of reliable benchmarks.

To address these gaps, this paper presents a comprehensive benchmark of GNN backbones and fusion architectures for brain age prediction using a large, multimodal and publicy available dataset. We systematically investigate the impact of combining structural and functional connectivity on predictive performance. Specifically, we explore and compare two distinct fusion methodologies—early fusion, where modalities are combined at the input level, and late fusion, where information is integrated at higher levels of the model. By keeping the GNN backbones consistent across experiments, we provide a controlled comparison of fusion techniques across a variety of GNN architectures. In this paper, we make the following contributions:

- We present a systematic comparison of different fusion methods for combining connectome data derived from DWI and fMRI. The evaluation is conducted within a Graph Neural Network (GNN) framework using a large, publicly available dataset.
- We provide a comprehensive hyperparameter analysis across various GNN backbones and data modalities (structural, functional, and multimodal). This analysis reveals novel, recurrent patterns in optimal model configurations across these different settings.

2 Materials and Methods

2.1 Data

The data used in this study are from the publicly available Philadelphia Neurodevelopmental Cohort (PNC) [33]. All PNC data used in this study is available through the database of Genotypes and Phenotypes (dbGaP), a NIH-designated repository, under the project name “Neurodevelopmental Genomics: Trajectories of Complex Phenotypes" (accession number: phs000607.v3.p2.c1).

This multi-modal dataset consists of cross-sectional magnetic resonance imaging (MRI) data of 1342 healty children and adolescents from 8 to 21 years of age. For this study, after removal of participants with a high-level of motion artefacts (see below), a subset of 747 participants were included, for whom both structural and resting-state functional neuroimaging data were available. The median age was 16.3 (IQR 13.5-18.5, min 8.2, max 23.1). Participants for the neuroimaging portion of the PNC were recruited from the greater Philadelphia area.

Image Acquisition Neuroimaging data were acquired on a single 3T Siemens TIM Trio scanner.

Structural MRI: A T1-weighted Magnetization-Prepared Rapid Gradient-Echo (MPRAGE) sequence was used to acquire high-resolution anatomical images. The parameters were as follows: $TR = 1810ms$; $TE = 3.5ms$; $TI = 1100ms$; $FA = 9°$; 160 slices; slice thickness = 1 mm; matrix size = $192x256$; field of view (FOV) = $180x240$ mm.

Diffusion-weighted imaging (DWI): DWI scans were obtained using a twice-refocused spin-echo (TRSE) single-shot EPI sequence. The sequence consisted of 64 diffusion-weighted directions with b = 1000 s/mm2 and 7 interspersed scans where b = 0 s/mm2.

Resting-State fMRI (rsfMRI): Resting-state functional connectivity data were acquired using a blood-oxygen-level-dependent (BOLD) sequence. The parameters were: $TR = 3000ms$; $TE = 32ms$; $FA = 90°$; 46 slices; slice thickness = 3 mm; matrix size = $64x64$; FOV = $192x192$ mm; 124 volumes.

2.2 Image processing

2.3 Brain Network Representation

The processed neuroimaging data for each participant is converted into a brain graph, or connectome. In this network model, the components are defined as follows:

- **Nodes**: These are the fundamental units of the network and represent 400 distinct, predefined anatomical regions of the brain according to a precomputed template called atlas.
- **Edges**: These represent the relationship or connection between any two nodes (brain regions):
  - **Structural Edges**: Derived from diffusion-weighted imaging (DWI), these edges represent the physical white matter tracts that form the brain's anatomical "wiring". A strong structural edge implies a robust physical pathway between two regions (i.e. fiber count).
  - **Functional Edges**: Derived from resting-state fMRI (rsfMRI), these edges represent the statistical synchronization of activity between two regions over time. A strong functional edge means two regions tend to activate together, suggesting they are part of a coordinated system.

Diffusion-weighted imaging Single-shell diffusion-weighted images were preprocessed using QSIprep [4]. Streamline tractography was performed using the `ss3t_csd_beta1` algorithm [9] from MRtrix to estimate fiber orientation distributions (FODs) for white matter, and cerebrospinal fluid using single shell acquisitions. The white matter FODs are used for tractography with no T1w-based anatomical constraints. Structural connectivity matrices based on the number of streamlines were extracted using the Schaefer400 atlas [34].

Resting-state fMRI The preprocessing used fMRIprep (v.23.1.4) [10]. The pipeline first generated a reference volume and corresponding brain mask. Slice-timing correction was then applied to adjust for temporal differences in slice

acquisition, followed by head motion correction. Functional data were then coregistered to each subject's T1-weighted anatomical scan, and subsequently normalized to MNI152NLin2009cAsym (adult) standard space using nonlinear registration. Participants with a mean framewise displacement (FD) greater than 0.5 mm, or a maximum FD exceeding 6 mm, were excluded due to excessive motion. Functional time-series data were band-pass filtered (0.01–0.08 Hz).

Subsequent denoising and interpolation were performed using the Nilearn Python package. Confound regressors were extracted by fMRIPrep and included the six motion parameters (translations and rotations) and their temporal derivatives. The functional connectivity matrices were calculated using Pearson's correlation on the same atlas as for the structural connectivity matrices. At the end of the processing, we obtain two conformable adjacency matrices, one from the DWI and one from the rsfMRI, of size 400×400, denoted $\mathbf{A}_S$ and $\mathbf{A}_F$.

Thresholding The matrices $\mathbf{A}_S$ and $\mathbf{A}_F$ generated from the neuroimaging pipelines were thresholded to minimize structural noise while preserving the maximum amount of information [23]. However, the selection of a specific thresholding value is arbitrary and lacks a universal optimum. Therefore, we evaluated three different proportional thresholds as part of our hyperparameter tuning: retaining the strongest 5%, 10%, and 20% of edges.

To ensure the input graph for the Graph Neural Network remains connected, which is a prerequisite for information propagation, a naive edge thresholding approach is insufficient as it can result in a disconnected graph with multiple components. To circumvent this, we first fit a Maximum Spanning Tree to the original graph. Subsequently, the remaining strongest edges are added back until the desired edge density is achieved, guaranteeing a single connected component.

2.4 Multimodal fusion strategies

As illustrated in Figure 1, we evaluated two families of fusion strategies, Early Fusion and Late Fusion, against the single-modality baselines.

Single-Modality Architecture This network takes a unique graph as input, either $\mathbf{A}_S$ or $\mathbf{A}_F$. The architecture consists of a stack of GNN layers to extract node-level features. These features are then aggregated through a mean global pooling into a single graph-level representation, which is subsequently fed into a Multi-Layer Perceptron (MLP) for regression.

Late Fusion Multimodal Architecture This model processes structural and functional connectivity graphs through two independent streams, each a distinct graph neural network with distinct features. Each stream generates a latent graph-level representation for its respective modality. These two representations are first passed through separate linear layers and then aggregated to create a final, unified multi-modal representation of the subject's brain. This final representation is then passed to an MLP for the regression task.

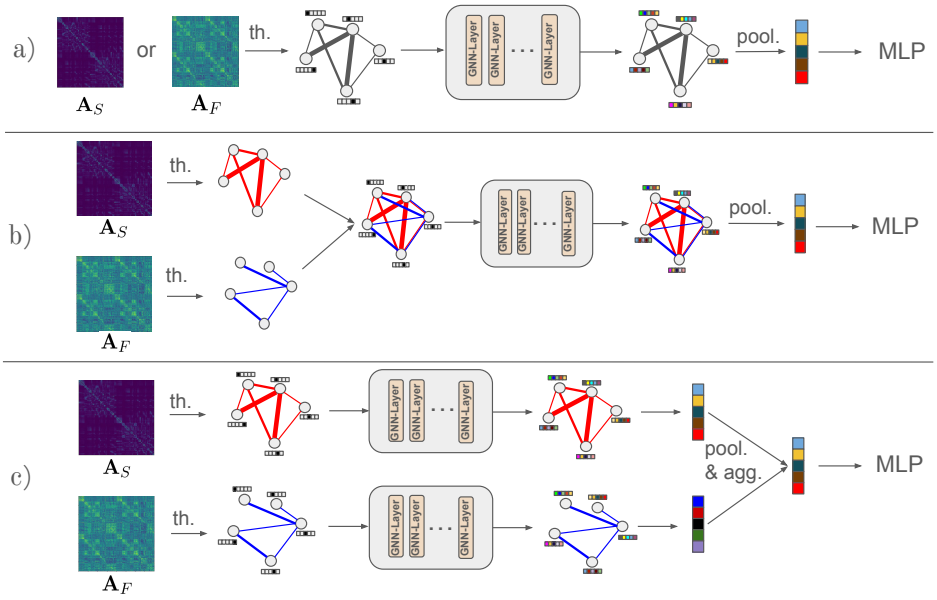

Fig. 1. Three distinct model architectures used in our study: a) Single-Modality, b) Early Fusion, and c) Late Fusion. All architectures share a common initial step where identity matrix connectivity is used for node feature initialization. The "GNN-Layer" blocks represent a generic graph neural network layer, which, in our implementation, can be instantiated as one of several types: GCN, GAT, GATv2, ARMA, or Transformer. Abbreviations: MLP: Multi-Layer-Perceptron; th.: thresholding; agg.: aggregation; pool.: pooling

Early Fusion Multimodal Architecture This approach constructs a single, unified heterogeneous graph representation for each subject, where distinct modalities are treated as different edge types connecting a common set of nodes representing brain regions - a multi-layer graph.

For a given node, the aggregation of features from its neighbors is conditioned on the type of edge connecting them. This is achieved by employing a set of distinct, scalable weight matrices for each edge type separately. Consequently, the model learns to weigh the contributions of structural and functional connectivity information differently during the feature aggregation process at each layer. The final graph-level representation, which captures the processed multimodal information, is obtained by pooling the node-level embeddings. This unified representation is then fed into a multilayer perceptron (MLP) for the final regression analysis.

2.5 Node and edge features

Node Feature Initialization A challenge in brain graph analysis is the absence of intrinsic initial features for nodes representing brain regions. These features must therefore be constructed. Building upon the findings of [5], we implemented two node initialization strategies within our framework:

- **Identity Matrix:** Each node is one-hot encoded, allowing the neural network to implicitly track individual node identities and positions.
- **Connectivity Matrix:** Nodes are initialized with their weighted degree, effectively representing their total connectivity strength within the graph. This provides the network with explicit structural information.

Edge Weight Considerations The inclusion of edge weights is treated as a tunable hyperparameter during training. When used, edge weights directly correspond to the values in the adjacency matrix, and are processed using the standard PyTorch Geometric implementations. Conversely, configurations that do not incorporate edge weights treat all edges as binary. In such cases, structural information is solely derived from the applied thresholding or, if used, from node features initialized with the connectivity matrix.

2.6 Experimental Setup

Data Partitioning To ensure a robust evaluation of our models, we use 5-fold cross-validation. For each fold, the dataset was partitioned into training, validation, and test sets, comprising 80%, 10%, and 10% of the data, respectively. To prevent data leakage, the splits were stratified by subject, ensuring that data from any given subject was contained within a single fold.

Backbones To comprehensively evaluate our proposed architectures, we benchmarked their performance using several distinct GNN models, hereafter referred to as backbones. Each backbone implements a unique mechanism for aggregating and transforming node features within the graph. The selected backbones are described below:

– **GCN (Graph Convolutional Network)** [18]: A foundational model that learns node representations by aggregating feature information from immediate neighbors using a fixed, normalized weighting scheme.
– **GAT (Graph Attention Network)** [40]: Introduces an attention mechanism to dynamically learn and assign different levels of importance to various nodes within a neighborhood during feature aggregation.
– **GATv2** [2]: An improved version of GAT that modifies the attention mechanism to be more expressive, allowing it to capture a wider range of dependencies between nodes.
– **ARMA (Auto-Regressive Moving Average)** [1]: A graph filter inspired by classical signal processing that captures information from multi-hop neighborhoods, allowing it to model longer-range dependencies more efficiently.
– **Transformer (Graph Transformer)** [35]: Adapts the successful Transformer architecture to graph data, treating all nodes as fully connected and using self-attention to dynamically learn graph structure and node relationships, often enhanced with positional or structural encodings.

We report the total number of learnable parameters for each model backbone in Table 1, using a standardized configuration for all other hyperparameters: a single-modality architecture with a 3-layer backbone with 32 hidden channels, followed by a 2-layer MLP head with 32 hidden dimensions.

Table 1. Total number of learnable parameters for each backbone under fixed hyperparameter configuration.

Backbone	Parameters
GCN	17,281
GAT	17,473
GATv2	32,417
ARMAConv	35,201
GraphTransformer	62,113

Evaluation Metric The performance of our regression models was quantified using the coefficient of determination (R-squared, R^2) on held-out data. The R^2 score indicates the proportion of the variance in the dependent variable that is predictable from the independent variable(s). It is calculated as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

where y_i is the true value, $\hat{y}_i$ is the predicted value, and $\bar{y}$ is the mean of the true values. R^2 closer to 1 indicates a better fit of the model to the data.

Training and Hyperparameter Optimization Our proposed framework was developed in PyTorch and run on an NVIDIA RTX 6000 GPU. All models were trained using the Adam optimizer for 300 epochs with early stopping after 35 epochs. A systematic hyperparameter search was conducted to identify the optimal configuration for each model architecture. This search encompassed various groups of parameters, including general training parameters, data preprocessing hyperparameters, single-modality encoder hyperparameters, and Multi-Layer Perceptron (MLP) specific parameters. Additionally, for the multimodal architectures, we explored different feature aggregation strategies. The specific hy-

perparameters and their relative search spaces are detailed in the Appendix A.1. Briefly, they include optimizer settings (e.g. learning rate), graph representation settings (e.g. edge thresholding density), backbone architecture (e.g. number of layers, residual connections), MLP architecture (e.g. activation function), as well as different aggregation functions for late and early fusion.

For the late fusion model, we allowed the hyperparameters to be independent between modalities. This approach is motivated by the inherent differences in the nature of the data from each stream. For instance, the structural modality may require a network with a different depth or a larger number of features to effectively capture its intricate spatial information. In contrast, the functional modality might benefit from a distinct set of hyperparameters tailored to its temporal or correlational characteristics. Therefore, decoupling the hyperparameter tuning process for each modality enables a more tailored and effective feature learning process, ultimately enhancing the performance of the combined model. To find the optimal hyperparameters for each model, we utilized the Bayesian search functionality of the `wandb` Python library using the validation loss as the minimizing objective, with a time budget of 10 hours for single-modality models and 20 hours for dual-modality models.

3 Results

Cross-validation results are shown in Table 2, comparing single modality networks to early and late fusion architecture in terms of the R^2 value on held-out data. Table 3 shows a summary of the hyperparameters that achived the best prediction. The complete list of all hyperparameters is in appendix A.2.

Table 2. Cross-validation results showing average and standard deviation of R^2 across folds (higher is better) and percentage change over structural baseline (N=373). TF: GraphTransformer. In **bold** are highlighted the best performing modality for each model

Model	Single modality (↑)		Fusion Architectures (↑)			
	Structural	Functional	Early	Chg.	Late	Chg.
GCN	0.583 ± 0.06	0.388 ± 0.12	0.600 ± 0.08	$+0.017$	$\mathbf{0.639 \pm 0.05}$	$\mathbf{+0.056}$
GAT	0.515 ± 0.07	0.376 ± 0.11	$\mathbf{0.526 \pm 0.04}$	$\mathbf{+0.011}$	0.506 ± 0.06	-0.009
GATv2	0.507 ± 0.05	0.335 ± 0.12	0.506 ± 0.09	-0.001	$\mathbf{0.571 \pm 0.06}$	$\mathbf{+0.064}$
ARMA	0.585 ± 0.02	0.380 ± 0.11	0.554 ± 0.04	-0.031	$\mathbf{0.608 \pm 0.06}$	$\mathbf{+0.023}$
TF	$\mathbf{0.581 \pm 0.03}$	0.314 ± 0.11	0.526 ± 0.08	-0.055	0.523 ± 0.06	-0.058
		Median Change	-0.001		**+0.023**	

3.1 Baseline Performance

The 'Structural connectivity' input modality consistently yields a better model fit than the 'Functional connectivity' modality across all tested graph neural network architectures. The GCN and ARMA models demonstrate the highest predictive performance on structural data, achieving mean R^2 scores of 0.583 ± 0.06 and 0.585 ± 0.02, respectively. Conversely, all models exhibit a significant drop in efficacy when trained exclusively on functional connectivity graphs, with the

proportion of explained variance ranging from 0.314 ± 0.11 for the Transformer to 0.388 ± 0.12 for the GCN.

3.2 Impact of Fusion Architectures

The introduction of multi-modal fusion architectures yields varied results on the models' performance. Overall four out of five models show improved performances using a multimodal architecture. Three out of the five models (GCN, GATv2, ARMA) show an improvement over their respective structural baselines when using a Late Fusion model. Notably, the GATv2 model achieves the most substantial relative gain, with its R^2 value increasing by +12.62%, from 0.507 to 0.571. The GCN model also sees a benefit, with a +9.61% improvement that results in the highest absolute score across the entire experiment ($R^2 = 0.639 \pm 0.05$).

In contrast, the Early Fusion architecture shows a less consistent and more modest impact. Only the GCN (+2.92%) and GAT (+2.14%) models register an improved score with this method. For the other three models (GATv2, ARMA, Transformer), the Early Fusion architecture results in a degradation of predictive performance compared to the structural baseline, with the most significant drop observed for the Transformer (−9.47%).

A summary of the fusion strategies is provided by the median percentage improvement in R^2. The Late Fusion architecture shows a positive impact, with a median improvement of +0.023 (+3.93% relative) across all models. Conversely, the Early Fusion architecture has a median improvement of −0.001 (−0.20%). The highest performing model-architecture combination is the GCN with Late Fusion, which explains the largest proportion of variance observed in the study.

3.3 Hyperparameters results

Table 3 shows a summary of the best hyperparameters for each backbone. The full list of hyperparameters are reported in Appendix A.1 in Tables 4, 5 and 6. Analysis of GNN hyperparameters (Table 3) reveals distinct patterns across modalities and architectures. Structural connectivity branches in Late Fusion models use no dropout in three backbones, suggesting a stable signal, while functional branches apply it in four backbones. Initial node feature choice is architecture-dependent: GCN uses identity matrices, Transformers use adjacency, and GAT, GATv2, and ARMA switch from identity for unimodal to adjacency for multimodal tasks, indicating the benefit of explicit connectivity information in complex data integration.

Architectural paradigms also differ by data type. Structural-only models favor deep, narrow configurations (5-6 layers, 16-32 channels), suggesting hierarchical feature extraction. Models incorporating functional data (unimodal or multimodal) prefer wider, shallower architectures (2-4 layers, 128-256 channels), indicating high-dimensional representation in early layers is more effective.

Finally, no single optimal fusion strategy exists. For Early Fusion, all aggregation methods were optimal for at least one backbone, and Late Fusion showed

Table 3. Core GNN Hyperparameters. This table provides an overview of best-performing hyperparamters, including aggregation and fusion methods, across all backbones and architectures.

Backbone	Architecture	Layers	Hidden Channels	Dropout	Fusion
GCN	Structural	5	32	0.2	
	Functional	2	128	0	
	Early Fusion	3	256	0.1	mul
	Late Fusion	S: 5, F: 3	S: 128, F: 128	S: 0, F: 0.1	sum
GAT	Structural	6	16	0.2	
	Functional	2	256	0.1	
	Early Fusion	2	256	0.2	min
	Late Fusion	S: 5, F: 3	S: 16, F: 32	S: 0, F: 0.1	w_sum
GATv2	Structural	6	32	0	
	Functional	2	64	0.2	
	Early Fusion	3	256	0.2	min
	Late Fusion	S: 5, F: 4	S: 128, F: 16	S: 0, F: 0.1	sum
ARMA	Structural	2	32	0	
	Functional	2	64	0.2	
	Early Fusion	3	256	0.3	mean
	Late Fusion	S: 4, F: 5	S: 128, F: 32	S: 0.1, F: 0.2	concat
Transformer	Structural	3	64	0	
	Functional	4	256	0.2	
	Early Fusion	2	32	0	max
	Late Fusion	S: 2, F: 5	S: 64, F: 32	S: 0.3, F: 0.2	sum

similar variability. This underscores that the ideal fusion mechanism is contingent on the backbone's message-passing scheme, with examples like GAT/GATv2 preferring 'min' and GCN favoring 'mul'.

4 Discussion

Our results underscore the significant potential of multimodal data fusion for improving brain age prediction using Graph Neural Networks.

Our primary finding is that the integration of structural and functional connectivity data, when managed by an appropriate fusion architecture, can yield substantial performance gains over models trained on a single modality. This is shown by the Late Fusion strategy, which produced the study's top-performing model (`GCN` with $R^2 = 0.639 \pm 0.05$), and achieved the most significant relative performance increase (a +12.62% increase for the `GATv2` model).

Our findings also reveal that the choice of fusion architecture is critical. The inconsistent and, in several cases, detrimental performance of the Early Fusion architectures highlights that simply combining modalities is not a guaranteed path to improvement.

Comparison with Prior Work Our findings align with and extend existing research in the field of brain age prediction. A recent systematic review observed that multimodal architectures consistently yielded better performance in

brain age prediction using traditional machine learning algorithms [17], although few studies in the review used connectivity data. Our work confirms that this principle holds true in the context of deep learning, specifically with Graph Neural Networks. Furthermore, our results mirror the established hierarchy of neuroimaging modalities observed in traditional machine learning approaches. Consistent with prior studies, we found that structural data (in our case DWI) serves as a more robust single-modality predictor than functional data, and that the fusion of both enhances predictive accuracy beyond what is achievable with either modality alone. To our knowledge, this is one of the first studies to systematically compare different GNN backbones and fusion strategies for multimodal brain age prediction, thus providing a valuable benchmark for future deep learning research in this area.

Interpretation of Architectural Performance A key result of our study is the clear superiority of the Late Fusion architecture over the Early Fusion one. There are several potential explanations for this difference.

First, the Late Fusion approach allows for the independent optimization of modality-specific architectures. Each branch can learn to extract the most salient features from its respective data type without interference, before a final, simpler integration step. Training two separate, well-understood architectures and then concatenating their high-level feature representations could be beneficial due to a more stable and less complex optimization landscape.

Conversely, the Early Fusion architecture, which processes different edge types within a single layer, may introduce a more complex and potentially unstable training dynamic. The direct interaction between kernels operating on structurally and functionally derived graphs could hinder effective feature learning, especially if the feature spaces are highly disparate. This may explain why three of the five models experienced a performance degradation with this method.

Limitations of the Study The findings of this study should be considered in light of several limitations. First, the models were trained and validated on a single dataset ($N = 747$). This may limit the generalizability of our findings to other populations or datasets with different demographic characteristics or acquisition parameters. Second, the present study was focused only on brain age prediction. A primary future goal is to extend the utility of these models to predict other relevant cognitive and clinical scores, which is the ultimate aim of our larger research project as detailed in our study protocol [15].

Finally, the normalization of neuroimaging data to the adult-adapted MNI1 52NLin2009cAsym standard space is suboptimal given the relatively young age of all subjects.

Significance and Future Directions Despite these limitations, our work suggests that GNN-based fusion is a promising strategy for predicting brain age, and establishes a methodological benchmark to compare fusion architectures. Furthermore, the insights derived from the hyperparameter search offer a valuable guide for future research employing GNNs on brain graphs. The superior

performance of the Late Fusion approach, in particular, offers a practical baseline for studies aiming to integrate multimodal connectomics data.

To push the field forward, future investigations should prioritize several key areas. First, it is necessary to validate these models on larger, multi-site datasets to establish their robustness and generalizability. Secondly, the integration of multiple data streams introduces a significant challenge of interpretability. Future efforts require advanced explainable AI to address opacity in multimodal models. We must isolate unimodal predictive features and deconstruct multimodal fusion interactions. This dual explainability is vital for neurobiologically plausible conclusions. Third, systematic evaluation of sophisticated fusion mechanisms is crucial. Though cross-modal attention architectures exist, their value is often shown against traditional machine learning, not strong deep learning baselines. Future work must rigorously benchmark complex architectures against well-defined baselines to clarify genuine performance benefits.

5 Conclusion

We conducted a systematic evaluation of early and late fusion Graph Neural Network architectures for brain age prediction using multimodal structural and functional brain connectivity. Our primary finding is that the choice of fusion architecture is critical for unlocking performance gains. We demonstrate that a late fusion approach, which processes each modality in a separate, optimized stream before integration, consistently outperforms both single-modality baselines and an early fusion strategy. This approach yielded the study's top-performing model, a GCN with late fusion that achieved an R^2 of 0.639. Furthermore, our hyperparameter analysis revealed distinct optimal configurations for structural and functional data, suggesting that modality-specific model design is essential for effective feature extraction.

Acknowledgments. This study was funded by the Swiss National Science Foundation under grant 214977. Support for the collection of the data for Philadelphia Neurodevelopment Cohort (PNC) was provided by grant RC2MH089983 awarded to Raquel Gur and RC2MH089924 awarded to Hakon Hakonarson. Subjects were recruited and genotyped through the Center for Applied Genomics (CAG) at The Children's Hospital in Philadelphia (CHOP). Phenotypic data collection occurred at the CAG/CHOP and at the Brain Behavior Laboratory, University of Pennsylvania.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Bianchi, F.M., Grattarola, D., Livi, L., Alippi, C.: Graph neural networks with convolutional ARMA filters. https://doi.org/10.1109/TPAMI.2021.3054830, http://arxiv.org/abs/1901.01343
2. Brody, S., Alon, U., Yahav, E.: How attentive are graph attention networks? https://doi.org/10.48550/arXiv.2105.14491, http://arxiv.org/abs/2105.14491

3. Cai, H., Zhou, Z., Yang, D., Wu, G., Chen, J.: Discovering brain network dysfunction in alzheimer's disease using brain hypergraph neural network. In: Greenspan, H., Madabhushi, A., Mousavi, P., Salcudean, S., Duncan, J., Syeda-Mahmood, T., Taylor, R. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2023. pp. 230–240. Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-43904-9_23
4. Cieslak, M., Cook, P.A., He, X., Yeh, F.C., Dhollander, T., Adebimpe, A., Aguirre, G.K., Bassett, D.S., Betzel, R.F., Bourque, J., Cabral, L.M., Davatzikos, C., Detre, J.A., Earl, E., Elliott, M.A., Fadnavis, S., Fair, D.A., Foran, W., Fotiadis, P., Garyfallidis, E., Giesbrecht, B., Gur, R.C., Gur, R.E., Kelz, M.B., Keshavan, A., Larsen, B.S., Luna, B., Mackey, A.P., Milham, M.P., Oathes, D.J., Perrone, A., Pines, A.R., Roalf, D.R., Richie-Halford, A., Rokem, A., Sydnor, V.J., Tapera, T.M., Tooley, U.A., Vettel, J.M., Yeatman, J.D., Grafton, S.T., Satterthwaite, T.D.: QSIPrep: an integrative platform for preprocessing and reconstructing diffusion MRI data **18**(7), 775–778. https://doi.org/10.1038/s41592-021-01185-5, https://www.nature.com/articles/s41592-021-01185-5, publisher: Nature Publishing Group
5. Cui, H., Dai, W., Zhu, Y., Kan, X., Gu, A.A.C., Lukemire, J., Zhan, L., He, L., Guo, Y., Yang, C.: BrainGB: A benchmark for brain network analysis with graph neural networks **42**(2), 493–506. https://doi.org/10.1109/TMI.2022.3218745, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10079627/
6. Cui, H., Dai, W., Zhu, Y., Li, X., He, L., Yang, C.: Interpretable graph neural networks for connectome-based brain disorder analysis, http://arxiv.org/abs/2207.00813
7. Dhamala, E., Jamison, K.W., Jaywant, A., Dennis, S., Kuceyeski, A.: Distinct functional and structural connections predict crystallised and fluid cognition in healthy adults **42**(10), 3102–3118. https://doi.org/10.1002/hbm.25420, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8193532/
8. Dhamala, E., Jamison, K.W., Jaywant, A., Dennis, S., Kuceyeski, A.: Integrating multimodal connectivity improves prediction of individual cognitive abilities. https://doi.org/10.1101/2020.06.25.172387, https://www.biorxiv.org/content/10.1101/2020.06.25.172387v1, pages: 2020.06.25.172387 Section: New Results
9. Dhollander, T., Raffelt, D., Connelly, A.: Unsupervised 3-tissue response function estimation from single-shell or multi-shell diffusion mr data without a co-registered t1 image (09 2016)
10. Esteban, O., Markiewicz, C.J., Blair, R.W., Moodie, C.A., Isik, A.I., Erramuzpe, A., Kent, J.D., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S.S., Wright, J., Durnez, J., Poldrack, R.A., Gorgolewski, K.J.: fMRIPrep: a robust preprocessing pipeline for functional MRI **16**(1), 111–116. https://doi.org/10.1038/s41592-018-0235-4, https://www.nature.com/articles/s41592-018-0235-4, publisher: Nature Publishing Group
11. Gamgam, G., Kabakcioglu, A., Yüksel Dal, D., Acar, B.: Disentangled attention graph neural network for alzheimer's disease diagnosis. In: Linguraru, M.G., Dou, Q., Feragen, A., Giannarou, S., Glocker, B., Lekadir, K., Schnabel, J.A. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2024. pp. 219–228. Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-72117-5_21
12. Gur, R.C., Butler, E.R., Moore, T.M., Rosen, A.F.G., Ruparel, K., Satterthwaite, T.D., Roalf, D.R., Gennatas, E.D., Bilker, W.B., Shinohara, R.T., Port, A., El-

liott, M.A., Verma, R., Davatzikos, C., Wolf, D.H., Detre, J.A., Gur, R.E.: Structural and functional brain parameters related to cognitive performance across development: Replication and extension of the parieto-frontal integration theory in a single sample **31**(3), 1444–1463. https://doi.org/10.1093/cercor/bhaa282, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7869090/

13. Honey, C.J., Sporns, O., Cammoun, L., Gigandet, X., Thiran, J.P., Meuli, R., Hagmann, P.: Predicting human resting-state functional connectivity from structural connectivity **106**(6), 2035–2040. https://doi.org/10.1073/pnas.0811168106, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2634800/
14. Jalali, P., Safayani, M.: HDGL: A hierarchical dynamic graph representation learning model for brain disorder classification. https://doi.org/10.48550/arXiv.2311.02903, http://arxiv.org/abs/2311.02903
15. James, C.E., Tingaud, M., Laera, G., Guedj, C., Zuber, S., Diambrini Palazzi, R., Vukovic, S., Richiardi, J., Kliegel, M., Marie, D.: Cognitive enrichment through art: a randomized controlled trial on the effect of music or visual arts group practice on cognitive and brain development of young children **24**(1), 141. https://doi.org/10.1186/s12906-024-04433-1, https://doi.org/10.1186/s12906-024-04433-1
16. Jiang, H., Cao, P., Xu, M., Yang, J., Zaiane, O.: Hi-GCN: A hierarchical graph convolution network for graph embedding learning of brain network and brain disorders prediction **127**, 104096. https://doi.org/10.1016/j.compbiomed.2020.104096, https://www.sciencedirect.com/science/article/pii/S0010482520304273
17. Jirsaraie, R.J., Gorelik, A.J., Gatavins, M.M., Engemann, D.A., Bogdan, R., Barch, D.M., Sotiras, A.: A systematic review of multimodal brain age studies: Uncovering a divergence between model accuracy and utility **4**(4), 100712. https://doi.org/10.1016/j.patter.2023.100712, https://linkinghub.elsevier.com/retrieve/pii/S2666389923000491
18. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. https://doi.org/10.48550/arXiv.1609.02907, http://arxiv.org/abs/1609.02907
19. Li, H., Satterthwaite, T.D., Fan, Y.: BRAIN AGE PREDICTION BASED ON RESTING-STATE FUNCTIONAL CONNECTIVITY PATTERNS USING CONVOLUTIONAL NEURAL NETWORKS **2018**, 101–104. https://doi.org/10.1109/ISBI.2018.8363532, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6074039/
20. Li, X., Zhou, Y., Dvornek, N., Zhang, M., Gao, S., Zhuang, J., Scheinost, D., Staib, L.H., Ventola, P., Duncan, J.S.: BrainGNN: Interpretable brain graph neural network for fMRI analysis **74**, 102233. https://doi.org/10.1016/j.media.2021.102233, https://www.sciencedirect.com/science/article/pii/S1361841521002784
21. Litwińczuk, M.C., Muhlert, N., Cloutman, L., Trujillo-Barreto, N., Woollams, A.: Combination of structural and functional connectivity explains unique variation in specific domains of cognitive function **262**, 119531. https://doi.org/10.1016/j.neuroimage.2022.119531, https://www.sciencedirect.com/science/article/pii/S1053811922006462
22. Lund, M.J., Alnæs, D., de Lange, A.M.G., Andreassen, O.A., Westlye, L.T., Kaufmann, T.: Brain age prediction using fMRI network coupling in youths and associations with psychiatric symptoms **33**, 102921. https://doi.org/10.1016/j.nicl.2021.102921, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8718718/

23. Luo, X., Wu, J., Yang, J., Xue, S., Beheshti, A., Sheng, Q.Z., McAlpine, D., Sowman, P., Giral, A., Yu, P.S.: Graph neural networks for brain graph learning: A survey
24. Mazumder, B., Kanyal, A., Wu, L., Calhoun, V.D., Ye, D.H.: Physics-guided multi-view graph neural network for schizophrenia classification via structural-functional coupling. vol. 15155, pp. 61–73. https://doi.org/10.1007/978-3-031-74561-4_6, http://arxiv.org/abs/2505.15135
25. Mazumder, B., Wu, L., Calhoun, V.D., Ye, D.H.: Unified cross-modal attention-mixer based structural-functional connectomics fusion for neuropsychiatric disorder diagnosis. https://doi.org/10.48550/arXiv.2505.15139, http://arxiv.org/abs/2505.15139
26. Mount, C.W., Monje, M.: Wrapped to adapt: Experience-dependent myelination **95**(4), 743–756. https://doi.org/10.1016/j.neuron.2017.07.009
27. Niu, X., Zhang, F., Kounios, J., Liang, H.: Improved prediction of brain age using multimodal neuroimaging data **41**(6), 1626–1643. https://doi.org/10.1002/hbm.24899, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7267976/
28. Park, H.J., Friston, K.: Structural and functional brain networks: From connections to cognition **342**(6158), 1238411. https://doi.org/10.1126/science.1238411, https://www.science.org/doi/10.1126/science.1238411, publisher: American Association for the Advancement of Science
29. Piccininni, M., Rohmann, J.L., Wechsung, M., Logroscino, G., Kurth, T.: Should cognitive screening tests be corrected for age and education? insights from a causal perspective **192**(1), 93–101. https://doi.org/10.1093/aje/kwac159, https://academic.oup.com/aje/article/192/1/93/6693335
30. Qu, G., Xiao, L., Hu, W., Zhang, K., Calhoun, V.D., Wang, Y.P.: Ensemble manifold based regularized multi-modal graph convolutional network for cognitive ability prediction **68**(12), 3564–3573. https://doi.org/10.1109/TBME.2021.3077875, http://arxiv.org/abs/2101.08316
31. Qu, G., Zhou, Z., Calhoun, V.D., Zhang, A., Wang, Y.P.: Integrated brain connectivity analysis with fMRI, DTI, and sMRI powered by interpretable graph neural networks **103**, 103570. https://doi.org/10.1016/j.media.2025.103570, https://www.sciencedirect.com/science/article/pii/S1361841525001173
32. Ray, B., Chen, J., Fu, Z., Suresh, P., Thapaliya, B., Calhoun, V.D., Liu, J.: Replication and refinement of brain age model for adolescent development
33. Satterthwaite, T.D., Connolly, J.J., Ruparel, K., Calkins, M.E., Jackson, C., Elliott, M.A., Roalf, D.R., Hopson, R., Prabhakaran, K., Behr, M., Qiu, H., Mentch, F.D., Chiavacci, R., Sleiman, P.M.A., Gur, R.C., Hakonarson, H., Gur, R.E.: The philadelphia neurodevelopmental cohort: A publicly available resource for the study of normal and abnormal brain development in youth **124**(0), 1115–1119. https://doi.org/10.1016/j.neuroimage.2015.03.056, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4591095/
34. Schaefer, A., Kong, R., Gordon, E.M., Laumann, T.O., Zuo, X.N., Holmes, A.J., Eickhoff, S.B., Yeo, B.T.T.: Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI **28**(9), 3095–3114. https://doi.org/10.1093/cercor/bhx179
35. Shi, Y., Huang, Z., Feng, S., Zhong, H., Wang, W., Sun, Y.: Masked label prediction: Unified message passing model for semi-supervised classification. https://doi.org/10.48550/arXiv.2009.03509, http://arxiv.org/abs/2009.03509

36. Smith, S.M., Vidaurre, D., Alfaro-Almagro, F., Nichols, T.E., Miller, K.L.: Estimation of brain age delta from brain imaging **200**, 528–539. https://doi.org/10.1016/j.neuroimage.2019.06.017, https://www.sciencedirect.com/science/article/pii/S1053811919305026
37. Soumya Kumari, L.K., Sundarrajan, R.: A review on brain age prediction models **1823**, 148668. https://doi.org/10.1016/j.brainres.2023.148668, https://www.sciencedirect.com/science/article/pii/S0006899323004390
38. Sui, J., Pearlson, G., Caprihan, A., Adali, T., Kiehl, K.A., Liu, J., Yamamoto, J., Calhoun, V.D.: Discriminating schizophrenia and bipolar disorder by fusing fMRI and DTI in a multimodal CCA+ joint ICA model **57**(3), 839–855. https://doi.org/10.1016/j.neuroimage.2011.05.055, https://linkinghub.elsevier.com/retrieve/pii/S1053811911005635
39. Uludağ, K., Roebroeck, A.: General overview on the merits of multimodal neuroimaging data fusion **102**, 3–10. https://doi.org/10.1016/j.neuroimage.2014.05.018, https://linkinghub.elsevier.com/retrieve/pii/S1053811914003838
40. Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., Bengio, Y.: Graph attention networks. https://doi.org/10.48550/arXiv.1710.10903, http://arxiv.org/abs/1710.10903
41. Wang, W., Xiao, L., Qu, G., Calhoun, V.D., Wang, Y.P., Sun, X.: Multiview hyperedge-aware hypergraph embedding learning for multisite, multiatlas fMRI based functional connectivity network analysis **94**, 103144. https://doi.org/10.1016/j.media.2024.103144, https://www.sciencedirect.com/science/article/pii/S1361841524000690
42. Xia, Z., Wang, H., Zhou, T., Jiao, Z., Lu, J.: Customized relationship graph neural network for brain disorder identification. In: Linguraru, M.G., Dou, Q., Feragen, A., Giannarou, S., Glocker, B., Lekadir, K., Schnabel, J.A. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2024. pp. 109–118. Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-72069-7_11
43. Xu, K., Li, C., Tian, Y., Sonobe, T., Kawarabayashi, K.i., Jegelka, S.: Representation learning on graphs with jumping knowledge networks, https://arxiv.org/abs/1806.03536v2
44. Zhang, X., He, L., Chen, K., Luo, Y., Zhou, J., Wang, F.: Multi-view graph convolutional network and its applications on neuroimage analysis for parkinson's disease **2018**, 1147–1156, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6371363/
45. Zhu, Y., Cui, H., He, L., Sun, L., Yang, C.: Joint embedding of structural and functional brain networks with graph neural networks for mental illness diagnosis. https://doi.org/10.48550/arXiv.2107.03220, http://arxiv.org/abs/2107.03220

A Appendix

A.1 Hyperparameter search space

The search space for the general training parameters was as follows:

- **Learning Rate:** {0.001, 0.005, 0.01}
- **Weight Decay:** {0.0, 0.001}
- **Batch Size:** {16, 32, 64}

Data hyperparameters:

- **Node feature initialization:** {Identity matrix, Connectivity matrix}
- **Edge threshold percentage:** {5, 10, 20}

Given the architectural diversity of our backbone models, we explored a comprehensive set of hyperparameters for the single-modality encoders:

- **Hidden Channels:** {16, 32, 64, 128, 256}
- **Number of Layers:** {2, 3, 4, 5, 6}
- **Dropout:** {0.0, 0.1, 0.2, 0.3}
- **Normalization:** {LayerNorm, BatchNorm1d}
- **Jumping Knowledge Strategy [43]:** {'max', 'last', 'cat'}
- **Residual Connections:** {True, False}

For the final Multi-Layer Perceptron (MLP) used for regression, the hyperparameter search space was defined as:

- **Number of Hidden Layers:** {1, 2, 3}
- **Hidden Channels:** {32, 64}
- **Dropout:** {0.0, 0.1, 0.2, 0.3}
- **MLP Activation Function:** {'ReLU', 'LeakyReLU', 'ELU'}

The early and late fusion architectures introduced an additional hyperparameter governing the feature aggregation strategy.

For the **late fusion models**, which combine features after they have been processed by modality-specific encoders, the following aggregation functions were evaluated:

- **Concatenation ('concat'):** Appends the feature vectors from each modality end-to-end, creating a single, larger vector that preserves all information.
- **Summation ('sum'):** Performs an element-wise sum of the feature vectors, requiring them to be of the same dimension. This approach assumes features occupy a similar semantic space.
- **Attention ('attention'):** Employs a learned attention mechanism to compute a weighted combination of the feature vectors, allowing the model to dynamically prioritize the most informative modalities for a given input.
- **Weighted Sum ('weighted_sum'):** Calculates a linear combination of the feature vectors, where the scalar weights for each modality's feature vector are learnable parameters.

For the **early fusion models**, which combine raw or minimally processed features at the input level, the search space for the element-wise aggregation function included:

- **Summation ('sum'):** An element-wise addition of the input feature tensors.
- **Mean ('mean'):** The element-wise average of the input feature tensors, which normalizes the combined representation.
- **Maximum ('max') / Minimum ('min'):** An element-wise selection of the maximum or minimum value across the input tensors, respectively.
- **Multiplication ('mul'):** An element-wise product of the input feature tensors, which can be used to scale features by one another.

A.2 Hyperparameter search results

Table 4. Comparison of hyperparameters for unimodal structural models versus the structural part of late fusion models across different GNN backbones. Abbreviations: Uni (Unimodal), LF (s) (Late Fusion, structural part), L. ReLU (Leaky ReLU), BN1d (BatchNorm1d), ID (Identity), Adj (Adjacency), T (True), F (False).

	GCN		**GAT**		**GATv2**		**ARMA**		**Transformer**	
Hyperparameter	**Uni**	**LF (s)**	**Uni**	**LF (s)**	**Uni**	**LF (s)**	**Uni**	**LF (s)**	**Uni**	**LF (s)**
batch_size	32	64	64	32	64	64	64	64	64	32
dropout	0.2	0	0.2	0	0	0	0	0.1	0	0.3
hidden_channels	32	128	16	16	32	128	32	128	64	64
jk	last	cat	cat	cat	last	last	max	max	last	max
learning_rate	0.001	0.001	0.005	0.01	0.001	0.001	0.001	0.005	0.001	0.005
mlp_activation	relu	elu	L. ReLU	L. ReLU	relu	L. ReLU	elu	elu	L. ReLU	relu
mlp_dropout	0.2	0.3	0.3	0.2	0.3	0	0.3	0.3	0.3	0.1
mlp_hidden_channels	32	32	32	64	64	32	32	64	64	64
mlp_hidden_layers	2	2	3	3	3	3	1	3	3	1
node_features	ID	ID	ID	Adj	Adj	Adj	ID	Adj	Adj	Adj
norm_type	BN1d	BN1d	BN1d	BN1d	BN1d	BN1d	BN1d	BN1d	BN1d	BN1d
num_layers	5	5	6	5	6	5	2	4	3	2
threshold	0.2	0.1	0.1	0.05	0.2	0.1	0.05	0.2	0.1	0.1
use_residual	F	T	F	T	T	T	T	F	F	F
use_weights	T	T	T	T	T	T	T	F	F	T
weight_decay	0.001	0.001	0.001	0	0	0	0	0	0	0

Observing Table 3, several distinctions arise between modalities and architectures. The first observation regards the regularization strategies and feature representation, particularly within multimodal frameworks. In the Late Fusion models, the structural branch consistently operates without dropout (`dropout_s` = 0), whereas the functional branch regularly applies it. This may indicate that the structural connectome provides a more stable, less noisy signal that requires minimal regularization. A notable pattern also emerges in the choice of initial node features. The GCN backbone exclusively prefers `identity` matrices, while the Transformer exclusively uses `adjacency` matrices. Interestingly, the GAT, GATv2, and ARMA backbones show a clear dependency on modality: they use `identity` features for unimodal tasks but switch to `adjacency` features for the more complex multimodal architectures. This suggests that providing the model with explicit connectivity information as node features is particularly advantageous when integrating diverse data types.

An analysis of the optimal hyperparameters also reveals distinct architectural paradigms for processing structural versus functional data. Models processing only structural connectivity (i.e., the Structural architecture) consistently benefit from deep, narrow configurations, typically employing a higher number of layers (5–6) with fewer hidden channels (16–32). This suggests that capturing the hierarchical nature of structural information is best achieved through deeper feature extraction. Conversely, models incorporating functional data, both in unimodal (Functional) and multimodal (Early Fusion) settings, exhibit a preference

Table 5. Comparison of hyperparameters for unimodal functional models versus the functional part of late fusion models across different GNN backbones. Abbreviations: Uni (Unimodal), LF (f) (Late Fusion, functional part), L. ReLU (Leaky ReLU), BN1d (BatchNorm1d), LN (LayerNorm), ID (Identity), Adj (Adjacency), T (True), F (False).

	GCN		**GAT**		**GATv2**		**ARMA**		**Transformer**	
Hyperparameter	**Uni**	**LF (f)**	**Uni**	**LF (f)**	**Uni**	**LF (f)**	**Uni**	**LF (f)**	**Uni**	**LF (f)**
batch_size	64	64	32	32	64	64	64	64	64	32
dropout	0	0.1	0.1	0.1	0.2	0.1	0.2	0.2	0.2	0.2
hidden_channels	128	128	256	32	64	16	64	32	256	32
jk	cat	last	last	cat	max	max	cat	last	cat	cat
learning_rate	0.001	0.001	0.001	0.01	0.01	0.001	0.001	0.005	0.001	0.005
mlp_activation	elu	elu	elu	L. ReLU	elu	L. ReLU	L. ReLU	elu	elu	relu
mlp_dropout	0.1	0.3	0.3	0.2	0.2	0	0	0.3	0.1	0.1
mlp_hidden_channels	32	32	64	64	32	32	64	64	32	64
mlp_hidden_layers	2	2	3	3	3	3	2	3	1	1
node_features	ID	ID	ID	ID	ID	Adj	ID	Adj	Adj	Adj
norm_type	BN1d	BN1d	BN1d	LN	BN1d	BN1d	LN	BN1d	BN1d	LN
num_layers	2	3	2	3	2	4	2	5	4	5
threshold	0.05	0.1	0.05	0.1	0.05	0.2	0.05	0.2	0.05	0.05
use_residual	F	F	T	F	F	F	T	F	F	F
use_weights	T	F	T	T	T	F	T	T	F	F
weight_decay	0.001	0.001	0.001	0	0	0	0.001	0	0	0

Table 6. Comparison of hyperparameters for Early Fusion models across different GNN backbones. Abbreviations: L. ReLU (Leaky ReLU), BN1d (BatchNorm1d), ID (Identity), Adj (Adjacency), T (True), F (False).

Hyperparameter	**GCN**	**GAT**	**GATv2**	**ARMA**	**Transformer**
batch_size	64	64	32	64	64
dropout	0.1	0.2	0.2	0.3	0
hetero_aggr	mul	min	min	mean	max
hidden_channels	256	256	256	256	32
jk	cat	cat	last	max	max
learning_rate	0.001	0.001	0.001	0.005	0.001
mlp_activation	relu	relu	elu	L. ReLU	relu
mlp_dropout	0	0.2	0.3	0.3	0.3
mlp_hidden_channels	64	32	64	64	64
mlp_hidden_layers	1	1	2	1	2
node_features	ID	Adj	Adj	Adj	Adj
norm_type	BN1d	BN1d	BN1d	BN1d	BN1d
num_layers	3	2	3	3	2
threshold_functional	0.05	0.05	0.2	0.1	0.05
threshold_structural	0.1	0.05	0.2	0.1	0.1
use_residual	T	F	F	F	T
use_weights	T	T	T	F	F
weight_decay	0	0	0.001	0	0

for wider, shallower architectures. These models achieve optimal performance with a larger number of `hidden_channels` (frequently 128 or 256) across fewer layers (2–4), implying that functional connectivity patterns are more effectively captured by high-dimensional representations in the initial layers.

Finally, our results underscore that there is no universally optimal strategy for fusing structural and functional information. For the Early Fusion architecture, every evaluated aggregation method (`mul`, `min`, `mean`, and `max`) was optimal for at least one backbone. Similarly, for Late Fusion, `sum`, `weighted_sum`, and `concat` were all chosen as the best-performing methods for different models. This highlights that the ideal fusion mechanism is highly contingent on the backbone's specific message-passing scheme. For instance, attention-based models such as GAT and GATv2 performed best with the `min` aggregator, while the simpler GCN model favored `mul`, illustrating the strong interplay between model architecture and data fusion techniques.