# Project 2 Report: GNN for AD Classification using Brain Connectivity

## 1. Abstract

This project implements a Graph Convolutional Network (GCN) for binary classification of Alzheimer’s Disease (AD) versus Cognitively Normal (CN) subjects using 150×150 brain connectivity matrices. We treat structural connectivity (SC) as the graph topology and functional connectivity (FC) as node features. Our objective prioritizes sound methodology, preprocessing, and evaluation with cross-validation over raw accuracy.

## 2. Dataset

We use 20 subjects total (10 AD, 10 CN). Each subject directory contains FunctionalConnectivity.txt and StructuralConnectivity.txt matrices (150×150). Stratified K-fold ensures balanced labels per fold.

## 3. Preprocessing & Graph Construction

Structural connectivity (SC):

- log1p compression to reduce heavy tails

- optional unit scaling to [0, 1]

- add self-loops and compute GCN symmetric normalization Â = D^{-1/2}(A+I)D^{-1/2}

Functional connectivity (FC):

- optional Fisher z-transform of correlations

- standardization either globally per subject or row-wise per node

Design choice: SC → adjacency; FC → features.

## 4. Model Architecture

A compact two-layer GCN with batch normalization and dropout. Node features from FC are linearly projected to a hidden dimension, followed by two GCN layers that perform message passing with the pre-normalized SC adjacency. Global mean pooling aggregates node embeddings into a graph-level representation, then a small MLP head produces logits.

## 5. Training Procedure

- Loss: CrossEntropyLoss

- Optimizer: AdamW with weight decay

- Regularization: Dropout and gradient clipping

- Early stopping on validation ROC-AUC (patience)

- Training loss printed every N epochs; confusion matrix plotted per fold

## 6. Cross-Validation & Metrics

We use StratifiedKFold (default 5 folds). Metrics: Accuracy, Balanced Accuracy, F1, ROC-AUC. Per-fold confusion matrices are saved as images (cm\_fold\_k.png).

## 7. Experimental Rationale from Literature

Connectomics literature supports using SC as a stable anatomical scaffold and FC as a functional signal. Standard GCN normalization D^{-1/2}(A+I)D^{-1/2} promotes numerically stable message passing. Given the small sample size, a lightweight two-layer GCN is preferred over deeper or heavier variants. We conducted brief paper reconnaissance to guide these choices and avoided over-complex multimodal fusion.

## 8. Results (Cross-Validation)

We report mean metrics across folds printed by the training script (Accuracy, Balanced Accuracy, F1, ROC-AUC). Confusion matrices are visualized for each fold. Exact numbers depend on random seeds and splits; see console logs and saved figs for your run.

## 9. Ablations & Extensions

- FC as adjacency (|z|) vs FC as features

- Fisher z-transform on/off; global vs row-wise standardization

- Hidden size and number of GCN layers

## 10. Reproducibility

We fix random seeds and set deterministic flags where possible.

## 11. Repository Structure

project\_root/  
├── AD\*/ CN\*/ # subject folders with .txt matrices  
├── main.py # training + evaluation entry point  
├── requirements.txt  
├── cm\_fold\_\*.png # generated confusion matrices  
└── README.md

## 12. References

- Kipf & Welling (2017). Semi-Supervised Classification with Graph Convolutional Networks.

- General SC/FC connectomics practices in neuroimaging GCN literature.