

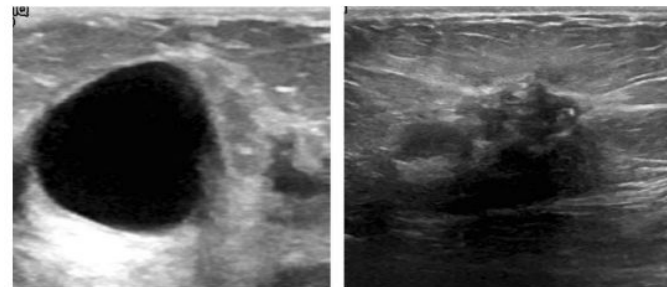
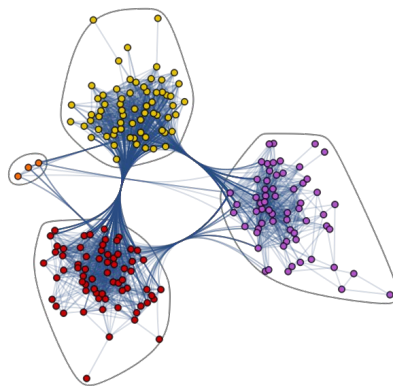
Survey on using GNNs for Breast Cancer Classification

Group number: 4

Vidhi Bhatt
vidhib@g.ucla.edu

Shruti Tyagi
shrutityagi@g.ucla.edu

Andy Polizzotto
andypolizzotto@g.ucla.edu



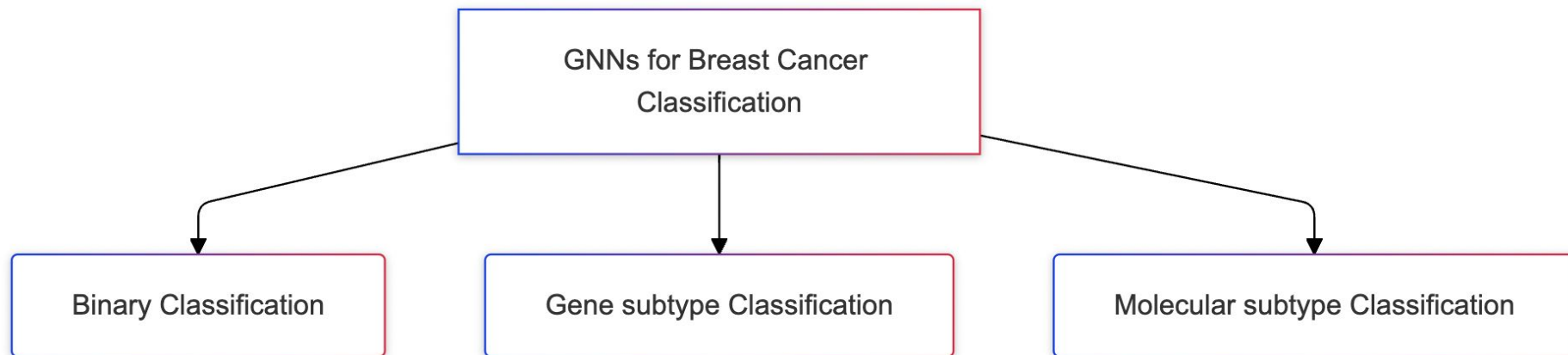
Problem Statement & Motivation

- Breast cancer is the most prevalent cancer among women globally, accounting for 2.3 million new cases annually and nearly 685,000 deaths.
- Traditional diagnostic methods like histopathology, mammography, & ultrasound, are resource-intensive, subjective & prone to variability.
- Critical need for robust, scalable & accurate models to assist radiologists in early decision-making, detection & classification of breast cancer
- Leveraging advanced GNN architectures can address challenges like dataset heterogeneity, feature interdependencies & overfitting, ultimately contributing to faster, more reliable diagnoses and better patient care.



Leading cause of cancer-related death in women.

GNNs for Breast Cancer Classification



Breast Cancer Binary Classification

Paper 1

Improved Breast Cancer Classification Through Combining Graph Convolutional Network and Convolutional Neural Network

Yu-Dong Zhang, Suresh Chandra Satapathy, David S. Guttery, Juan Manuel Górriz, Shui-Hua Wang

Paper 2

Enhancing Histopathology Breast Cancer Detection & Classification with Deep Ensemble Graph Network

Shwetha G. Krishnappa & K. R. Udaya Kumar Reddy

Paper 3

Graph neural network-based breast cancer diagnosis using ultrasound images with optimized graph construction integrating the medically significant features

Sadia Sultana Chowra, Sami Azam, Sidratul Montaha, Israt Jahan Payel, Md Rahad Islam Bhuiyan, Md. Zahid Hasan & Mirjam Jonkman

Breast Cancer Binary Classification

Papers	1: BDR-CNN-GCN	2: DEGN	3: GCN with US Images
Datasets & Input Types	Mini-MIAS (322 single-breast mammogram slices) Abnormal: 113; Normal: 209	BCSS (binary) & BACH (binary + multi) Histopathology images	BUSI (647 Ultrasound images made using high-end instruments) Benign - 467, Malignant - 180
Model	Hybrid of 8-layer CNN with BN, Dropout, & Rank-based stochastic pooling (BDR-CNN) & 2 layer-GCN	Structural Graph Module (SGM) integrated with an ensemble graph network (EGN)	Handcrafted node features + Spearman correlation-based adjacency matrix → FFN → 3 GCN layers → FFN → softmax
Data Augmentation	14-way augmentation to avoid overfitting & enhance dataset size	Multi-magnification patch-based feature extract & ensemble techniques	ROI-based handcrafted features validated statistically, compared to Histogram of Oriented Gradients

Breast Cancer Binary Classification

Papers	1: BDR-CNN-GCN	2: DEGN	3: GCN with US Images
Pre-processing Steps	<p>Original Mammogram Image $d_0(k)$</p> <p>Step 1: AN Reduction</p> <p>Step 2: MN Reduction</p> <p>Step 3: CLAHE</p> <p>Step 4: BG Removal</p> <p>Step 5: PEM Removal</p> <p>Step 6: SSBC</p> <p>Step 7: Downsampling</p> <p>Preprocessed Image $d_7(k)$</p> <p>Side flow: $d_1(k) \xrightarrow{\ln} a \xrightarrow{F} A \xrightarrow{P} B \xrightarrow{F^{-1}} b \xrightarrow{\exp} d_2(k)$</p>	<p>Image Acquisition</p> <p>Step 1. Nuclei Segmentation</p> <p>Step 2. Multi-Magnification Patch-based Feature Extraction</p> <p>Step 3. Graph Construction</p>	<p>ROI extraction of Benign and Malignant tumor</p> <p>Mask</p> <p>Original</p> <p>Bitwise AND</p> <p>ROI</p>

Breast Cancer Binary Classification

Papers	1: BDR-CNN-GCN	2: DEGN	3: GCN with US Images
Role of GNNs	Extracts spatial and structural relationships between different regions of mammogram to find malignancies which CNN's may overlook.	GNN builds structural graphs of nuclei & ensures spatial relationships are considered, enhancing long-range dependencies modelling and classification of irregular and overlapping cells.	Simplifies graph-based representation images as nodes and relationships between features as edges using Spearman correlation for graph construction enabling robust analysis
Metrics	Sensitivity, Specificity: & Accuracy: 96 %	BCSS Accuracy: 94-96% BACH Accuracy: 99% for both	Accuracy & F1-Score: 99% Precision & Recall: 100%
Uses	Promising for clinical deployment in radiology	Automate clinical histopathology image analysis; improved precision	Radiology support for early diagnosis & automation

Breast Cancer Binary Classification

Papers	1: BDR-CNN-GCN	2: DEGN	3: GCN with US Images
Strength	Effective DA to handle small datasets; robust hybrid architecture; deployable in hospitals	Enhanced granularity, Highly adaptable to dataset shifts; highly accurate segmentation & robust classification for overlapping & irregular cells with SGM	Optimized performance with simplified graph; highest accuracy among dataset studies.
Limitations	Need validation on large & heterogeneous data	Relatively smaller datasets; requires validation on real-world images	Small dataset; more generalizing test needed
Future Direction	Larger data validation; integrate with cloud for speed & cost-efficiency; Make robust against noise & heterogeneity	Optimize adaptability, graph modeling and segmentation for real-world data; explore lightweight versions & hybrid GNNs with attention mechanisms	Introduce diverse & multimodal data for richer insights & generalisation. Automate feature extract

Breast Cancer Gene Subtype Classification

Paper 1

Classifying Breast Cancer using multi view GNN based on multi omics data

Yanjiao Ren, Yimeng Gao, Wei Du, Weibo Qiao, Wei Li, Qianqian Yang, Yanchun Liang, Gaoyang Li

Paper 2

Omics GAT: Graph Attention Network for Cancer Subtype Analyses

Sudipto Baul, Khandakar Tanvir Ahmed, Joseph Filipek, Wei Zhang

Paper 3

Comparative Analysis of Multi-Omics Integration Using Advanced GNNs for Cancer Classification

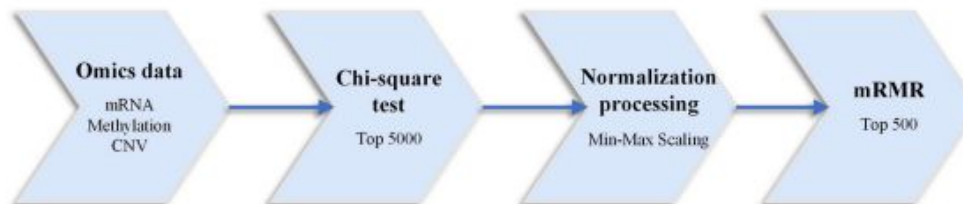
Fadi Alharbi, Aleksandar Vakanski, Boyu Zhang, Murtada K. Elbashir, Mohanad Mohammed

Breast Cancer Gene Subtype Classification

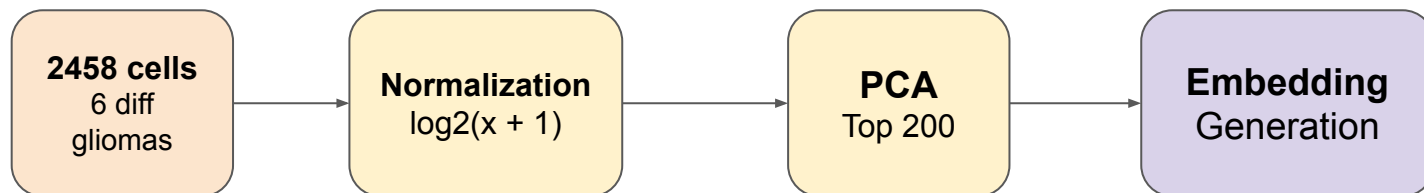
Papers	1. MVGNN	2. Omics GAT	3. Multi Omics
Primary Model	Multi-View Graph Neural Network (MVGNN).	Graph Attention Network (GAT).	GCN, GAT, and Graph Transformer Network (GTN).
Datasets/Input	TCGA-BRCA (The Cancer Genome Atlas Breast Cancer).	TCGA datasets for various cancer types.	TCGA and other public multi-omics datasets (not breast cancer-specific).
Pre - processing steps	Normalization, similarity computation, graph construction.	Normalization, feature scaling, and edge weighting.	Normalization, DEGs, Limma and Lasso regression.

Breast Cancer Gene Subtype Classification

MVGNN



Omics GAT



Multi Omics



Breast Cancer Gene Subtype Classification

Papers	1. MVGNN	2. Omics GAT	3. Multi Omics
Data Augmentation	Generates augmented similarity graphs to improve model robustness.	Limited mention of augmentation techniques.	Architectures use edge perturbations or synthetic graphs.
Feature Aggregation	Attention mechanism to integrate features across omics views.	Attention mechanism to focus on relevant neighbors.	Compared multiple architectures for aggregation methods.
Role of GNNs	Integrates multi-omics data for subtype prediction and differentiation.	Highlights neighbor-wise attention for feature extraction.	Evaluate architectural differences in multi-omics integration.
Metrics	ACC - 91.80 F1 - 95.30	ACC - 92.24 F1 - 91.04	ACC - 95.39 F1 - 93.80

Breast Cancer Gene Subtype Classification

Papers	1. MVGNN	2. Omics GAT	3. Multi Omics
Strengths	Stands out for its multi-view design, Simple, unified graph representation.	Integrates diverse omics data effectively.	Highlights tradeoffs and guides model selection.
Uses	Prognosis by identifying molecular-level insights.	Understanding of omics data interactions for precision medicine.	Provides benchmarks for selecting optimal GNN architectures.
Limitations	Requires extensive preprocessing and high computational resources.	Attention mechanisms can be computationally intensive, potentially limiting scalability to larger datasets.	Limited by dataset availability and generalizability.
Future Direction	Address scalability for large-scale datasets.	Extend to additional cancer subtypes.	Focus on benchmarking across more diverse datasets

Breast Cancer Molecular Subtype Classification

Paper 1

Hybrid Approach of Relation Network and Localized Graph Convolutional Filtering for Breast Cancer Subtype Classification

Sungmin Rhee, Seokjun Seo, Sun Kim (2018)

Paper 2

Attention-based GCN integrates multi-omics data for breast cancer subtype classification and patient-specific gene marker identification

Hui Guo, Xiang Lv, Yizhou Li and Menglong Li (2023)

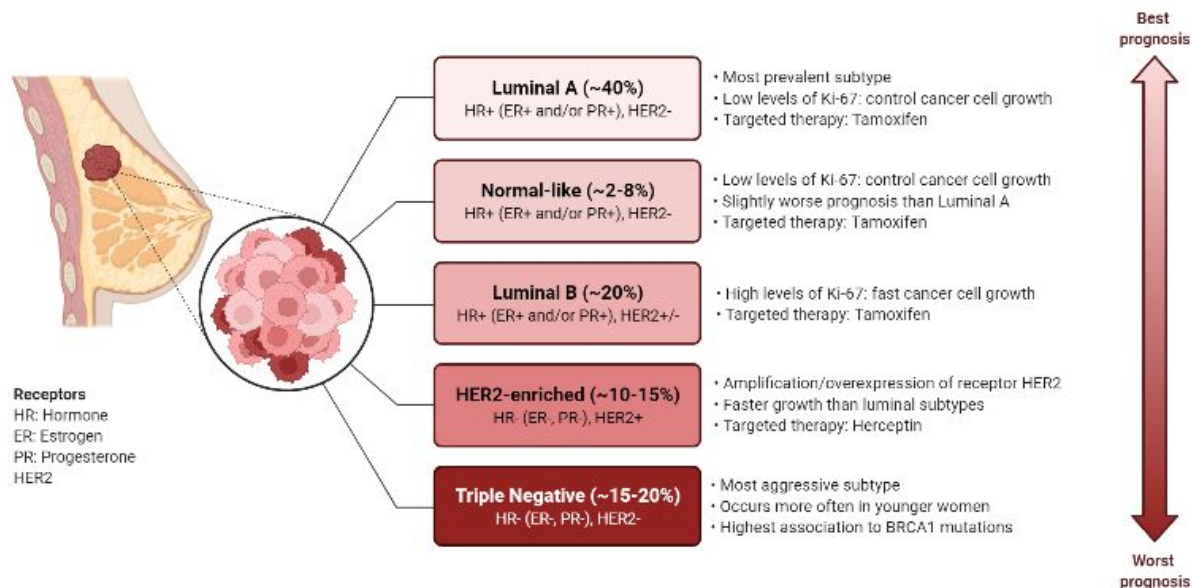
Paper 3

A Multimodal Graph Neural Network Framework for Cancer Molecular Subtype Classification

Bingjun Li and Sheida Nabavi (2024)

Breast Cancer Molecular Subtype Classification

5 Main Intrinsic or Molecular Subtypes of Breast Cancer



Breast Cancer Molecular Subtype Classification

Papers	1. GCN/RN	2. AGCN	3. GCN/GAT
GNN Structure	Combine GraphCNN with Relation Network; uses RN to capture gene-gene relations	Attention-based GCN to fuse multi-omics data with protein-protein interaction networks	Utilizes GCN for local feature extraction, GAT to capture cross-omic relationships
Feature Extraction	STRING database for gene-gene interactions; GCN to capture RNA-seq interactions	Selects highly variable genes (DNA-methyl, mRNA); constructs graph from PPI network	Cross-patient networks for omics data (mRNA, CNV, miRNA) captured locally and globally
Datasets/Use Cases	PAM50 dataset of RNA-seq profiles; localized learning of graph features	TCGA breast-invasive carcinoma (BRCA) dataset; layer-wise relevance propagation	TCGA pan-cancer and BRCA datasets

Breast Cancer Molecular Subtype Classification

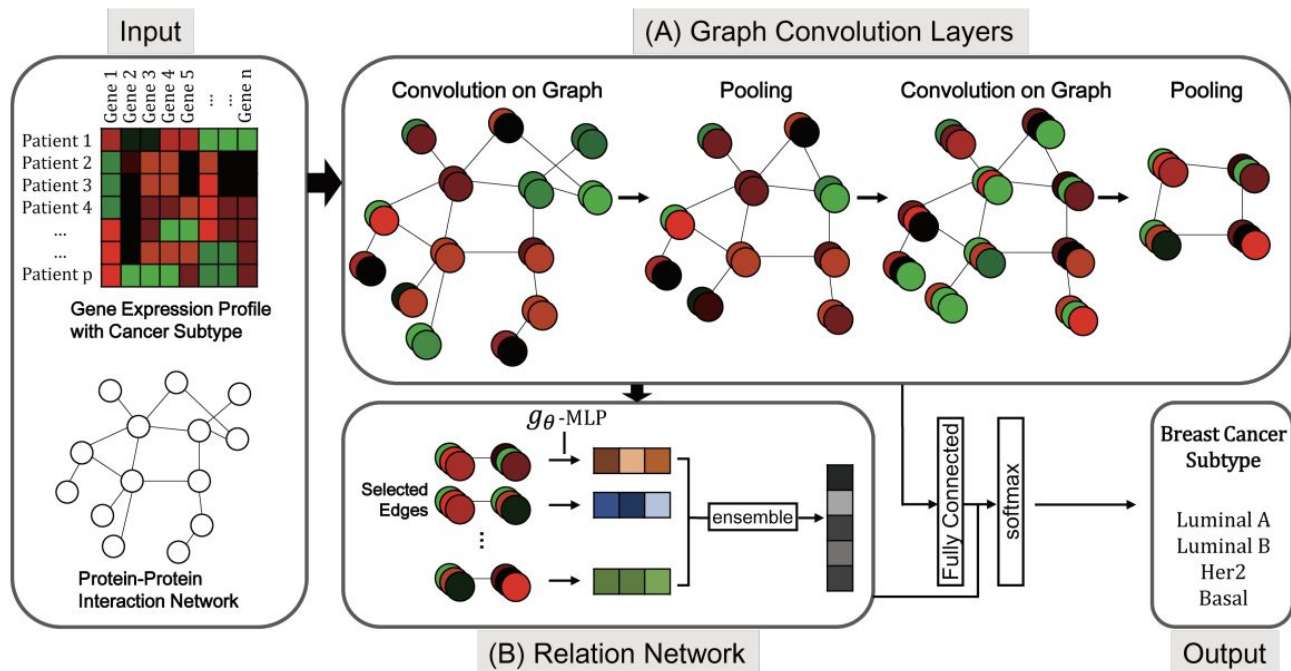
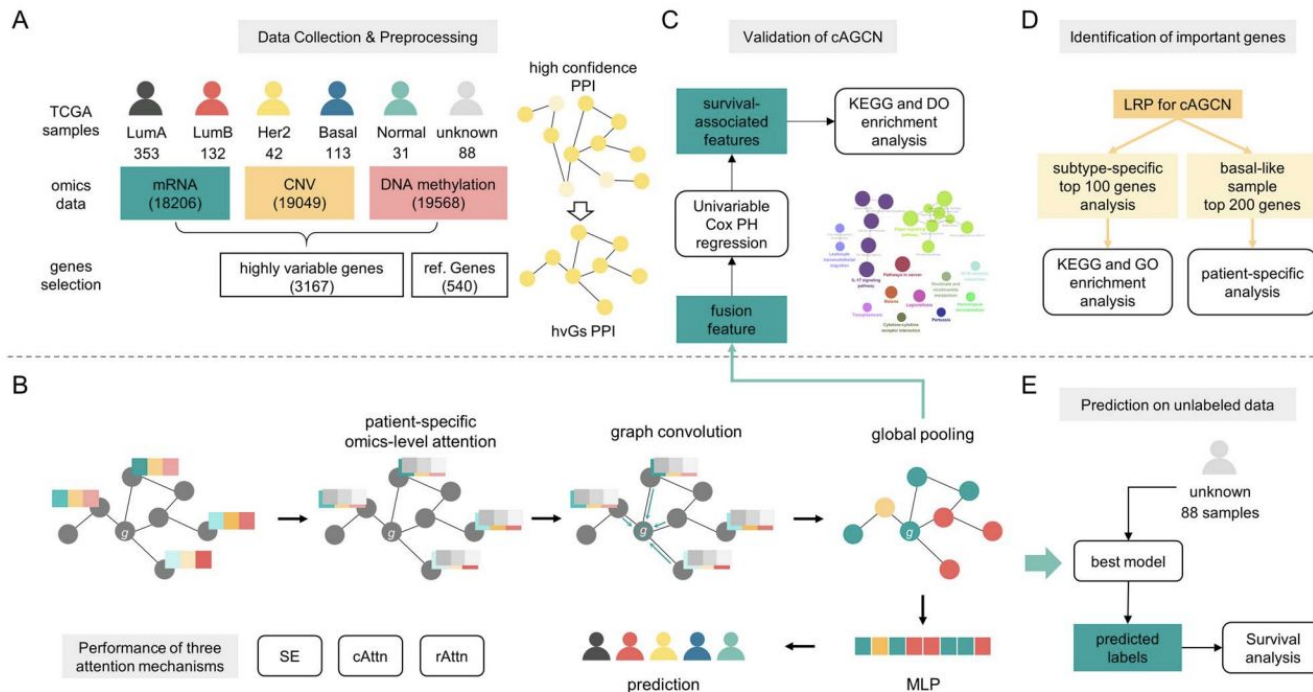
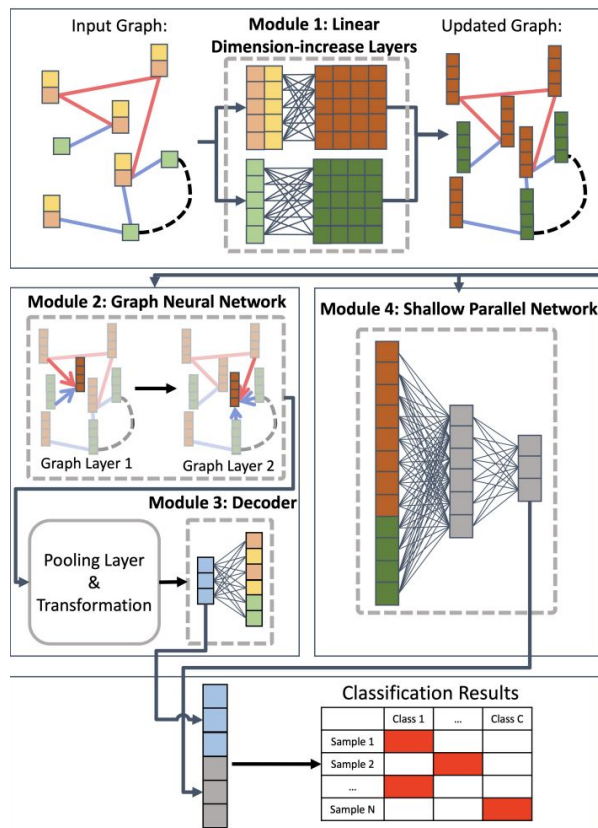


Figure 1: Overview of the proposed method

Breast Cancer Molecular Subtype Classification



Breast Cancer Molecular Subtype Classification



Breast Cancer Molecular Subtype Classification

Papers	1. GCN/RN	2. AGCN	3. GCN/GAT
Dataset	PAM50	TCGA BRCA	TCGA BRCA
Accuracy	86.29%	89.42%	86.40%
F1-Score	83.41%	-	87%
AUC	-	97.08%	-

Breast Cancer Molecular Subtype Classification

Papers	1. GCN/RN	2. AGCN	3. GCN/GAT
Strengths	Simple, intuitive, compatible with RNA-seq data	High accuracy, high explainability	Flexible, generalizable across cancers and omics
Limitations	Narrow focus; does not account for cross-omic data	Reliant on manual pre-processing. Struggles with incomplete/noisy datasets.	Computationally expensive; less fine-tuned for categorizing breast cancers
Future Work	Expand to include multi-omics data and larger datasets	Reduce reliance on manual processing; ensure compatibility with alternate datasets	Optimize scalability and incorporate domain-specific modules

References

1. <https://www.frontiersin.org/journals/genetics/articles/10.3389/fgene.2024.1363896/full>
2. <https://www.sciencedirect.com/science/article/pii/S0306457320309328?>
3. <https://link.springer.com/article/10.1007/s42979-024-02855-9>
4. <https://link.springer.com/article/10.1007/s00432-023-05464-w>
5. <https://arxiv.org/pdf/2410.05325>
6. <https://arxiv.org/abs/2302.12838>
7. <https://arxiv.org/abs/1711.05859>

Thank you