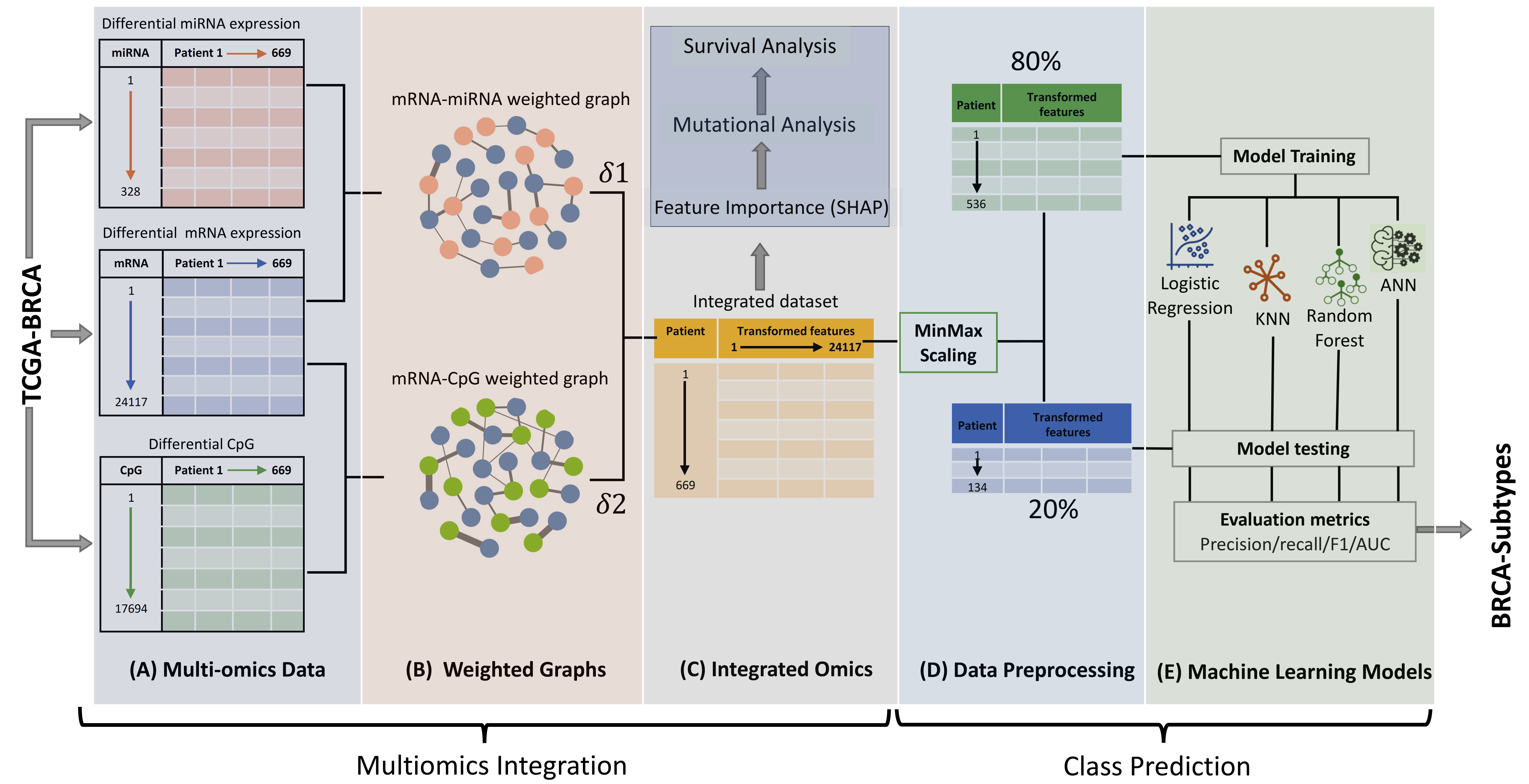
# GAIN-BRCA: A Graph-Based Explainable AI Framework for Breast Cancer Subtype Classification Based on Multi-Omics

## Overview

* GAIN-BRCA leverages multi-omics datasets to classify TCGA breast cancer subtypes based on PAM50.
* The framework integrates gene expression, DNA methylation, and microRNA data by utilizing their biological interactions.
* Specifically, it considers the regulatory relationships where gene expression is influenced by methylation and miRNA interactions.
* This graph-based explainable AI framework combines these interactions for improved multi-omics integration and prediction.

## GAIN-BRCA Workflow



## File Structure

# **GAIN-BRCA/ ├── GAIN\_BRCA.py**

# **└── dataset/**

# **└── Input/**

# **├── mRNA\_NormCount.csv**

# **├── miRNA\_NormCount.csv**

# **├── methyl\_NormBeta.csv**

# **├── miRNA\_mRNA\_interaction.csv**

# **├── CpG\_mRNA\_interaction.csv**

# **└── dependent\_variables.csv**

**├── delta\_integration.py**

# **├── GAIN\_BRCA\_ANN.py**

# **Requirements**

* Tensorflow (>= 2.9.1)
* Keras (>= 2.9.0)
* Python (>= 3.8.10)
* Scikit-learn (>=1.2.1)
* Basic Python packages: numpy, pandas, math

Ensure these libraries are installed.

**Data sources:**

1. **Clone or Download the Repository:**  
   Clone this repository or download the source code files.
2. **Prepare the Input Data:**  
   Place the breast cancer multi-omics datasets in the dataset/Input/ folder.  
   The required files are:
   * mRNA\_NormCount.csv
   * miRNA\_NormCount.csv
   * methyl\_NormBeta.csv
   * miRNA\_mRNA\_interaction.csv
   * CpG\_mRNA\_interaction.csv
   * dependent\_variables.csv

The datasets can be found here “https://zenodo.org/records/15175435”

1. **Edit File Paths (if necessary):**  
   Ensure that the directory path in GAIN\_BRCA.py matches the location of your input files.
2. **Run the Main Script:**  
   Execute the following command from the root directory:

python GAIN\_BRCA.py

The script will:

* Import and integrate the multi-omics datasets.
* Compute integrated expression values using the provided interaction data.
* Train an ANN model using stratified k-fold cross-validation.
* Output the final prediction accuracy and save prediction probabilities to a CSV file.

## Code Documentation

* **GAIN\_BRCA.py:**  
  Serves as the main execution script, orchestrating data import, integration, and model training.
* **delta\_integration.py:**  
  Contains functions to integrate genomic data:
* int\_Delta1(miRNA, mRNA, Int1): Integrates miRNA and mRNA expression data.
* int\_Delta2(methyl, mRNA, Int2): Integrates methylation data and mRNA expression data.
* **GAIN\_BRCA\_ANN.py:**  
  Implements the ANN model using TensorFlow/Keras.  
  This module handles data scaling, cross-validation, model training, evaluation, and saving prediction outputs.

# **Contact**

Please contact [japatel@unmc.edu](mailto:japatel@unmc.edu) if you have any inquiries or issues.