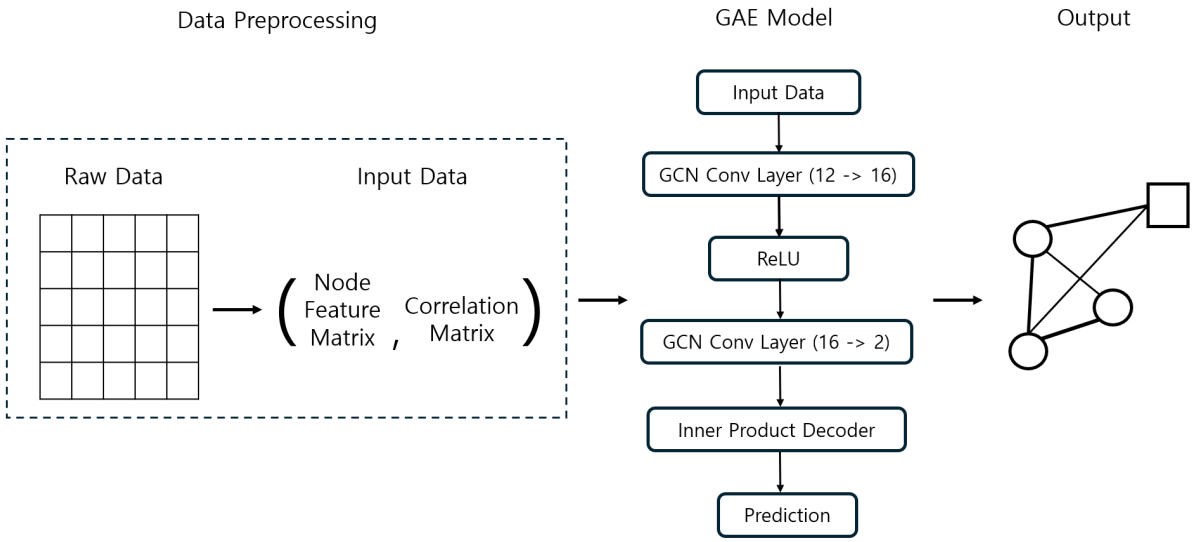
Supplement 2. Overall Workflow of Data Preprocessing and Graph Autoencoder (GAE) Modeling



This diagram illustrates the overall pipeline used in this study, from raw data preprocessing to the generation of graph-based outputs:

Data Preprocessing:

Raw Data: Monthly time-series records of air pollutant concentrations (PM_{10}, $PM_{2.5}$, SO_2, NO_2, O_3, CO) and emergency room visits for cardiovascular and cerebrovascular diseases (cardiac arrest, myocardial infarction, ischemic stroke, hemorrhagic stroke).

Input Data:

A *node feature matrix* consisting of normalized time-series or regional values for each variable.

A *correlation matrix* calculated as the Pearson correlation coefficients between node pairs, thresholded to define edges.

Graph Autoencoder (GAE) Model:

Input Layer: The combination of node features and the correlation-based adjacency matrix.

GCN Layers:

First graph convolutional layer with 12 input features transformed to 16 hidden units, followed by ReLU activation.

Second graph convolutional layer reducing the representation to a 2-dimensional latent embedding per node.

Inner Product Decoder: Reconstructs the adjacency matrix by computing similarity scores (structural similarity) between node embeddings.

Prediction: The output adjacency matrix containing predicted edge strengths (adj_pred) representing inferred structural connections.

Output:

A visual network where node positions reflect learned embeddings, and edge thickness represents predicted structural similarity.

This workflow demonstrates the study’s approach to integrating environmental and health data into a graph learning framework for structural relationship modeling and visualization.