Graduate School of Science and Technology Master's Thesis Abstract

Laboratory name (Supervisor)	Computational Systems Biologoy (Shigehiko Kanaya (Professor))		
Student ID	2311419	Submission date	2025 / 7 / 24
Name	MUHAMMAD HENDRICK SEDAYU		
Thesis title	Extraction of Molecules Feature for Classification Metabolic Pathway of Natural Compound		

Abstract

This study developed an Augmented Attention Graph Convolutional Network (AA-GCN) as a high-performance molecular feature extractor. This hybrid architecture was designed through multidimensional optimization: integrating molecular graph representation with MACCS chemical descriptors, implementing a sigmoid-based attention mechanism to focus on critical atoms, and balancing model complexity through batch normalization.
In computational biology evaluations, AA-GCN demonstrated superior feature extraction capabilities with an end-to-end validation accuracy of 98.11% and a generalization rate of 98.16% on test data—achievements reflecting early learning stabilization through early stopping at epoch 44. The representation optimization was further evident when the combined feature embedding (graph embedding + MACCS) improved Random Forest classification accuracy by 1.81% compared to single graph extraction, demonstrating the synergy of structural-descriptive representation.
Generalizability was tested on the KEGG dataset using paradigm transfer learning, where the pre-trained AA-GCN extractor yielded a ROC-AUC of 0.9857 (XGBoost) and macrometric consistency (F1 > 0.79)—an indicator of model robustness to novel compounds. These findings not only validate the efficacy of AA-GCN as an interpretable feature extractor but also open up potential applications in virtual screening of bioactive compounds and deep learning-based metabolic network modeling.

Keywords:
Attention Layer, Metabolic Pathway Prediction, Molecular Feature Extraction, Graph Convolutional Network