

## HallmarkGraph: a cancer hallmark informed graph neural network for classifying hierarchical tumor subtypes

### Authors:

*Qingsong Zhang , Fei Liu , Xin Lai*

### Keywords:

Systems Medicine, Machine Learning, Explainable AI, Molecular Tumor Subtype, Multi-Label Classification,

### Abstract

Accurate tumor subtype diagnosis is crucial for precision oncology, yet current methodologies face significant challenges. These include balancing model accuracy with interpretability and the high costs of generating multi-omics data in clinical settings. Moreover, there is a lack of validated models capable of classifying hierarchical tumor subtypes across a comprehensive pan-cancer cohort.

We present a graph neural network, HallmarkGraph, the first biologically informed model developed to classify hierarchical tumor subtypes in human cancer. Inspired by cancer hallmarks, the model's architecture integrates transcriptome profiles and gene regulatory interactions to perform multi-label classification. We evaluate the model on a comprehensive pan-cancer cohort comprising 11,476 samples from 26 primary cancers with 405 subtypes. The model demonstrates exceptional performance, achieving 5-fold cross-validation accuracy between 85% and 99% for tumor subtypes labelled with increasing details of genomic information. It also shows good generalizability on a validation dataset of 887 samples, assessed using three metrics that consider tumor subtypes at individual, combined, and sample levels. Benchmarking and ablation experiments show that hallmark-based embeddings slightly influence model performance, while the integrated multilayer perceptron plays a significant role in determining classifier accuracy. Additionally, we employ the SHAP method to link cancer hallmarks with genes, identifying key features that influence model decisions. Our findings present a biologically informed machine learning framework capable of tracking tumor transcriptomic trajectories and distinguishing inter- and intra-tumor heterogeneity in pan-cancer. This approach holds promise for enhancing cancer diagnostics.