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

Qingsong Zhang[1] , Fei Liu[1,*], Xin Lai[2,3,*]

[1]School of Software Engineering, South China University of Technology, Guangzhou, China

[2]Systems and Network Medicine Lab, Biomedicine Unit, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland

[3]Friedrich-Alexander-Universität Erlangen-Nürnberg and Universitätklinikum Erlangen, Erlangen, Germany

[*]Corresponding author: South China University of Technology, Guangzhou, China. Email: feiliu@scut.edu.cn (FL). Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland. Email: and xin.lai@tuni.fi (XL).

This is the file library for HallmarkGraph

The current purpose of this file library is to provide reviewers with reproducible experimental results.

The repository contains the following:

code file :

06_BioGCN_load_model_and_test_pred.py

cleaned_data_and_target file :

Training and testing data : cleaned_data_out.csv

Training and testing label : cleaned_data_test_12_7.xlsx

adjacency_matrix file:

ten Hallmark related adjacency matrices

saved model file :

train : test = 0.6 : 0.4 : save_model(0.4)_18

analysis file :

Analysis file of experimental results : analysis

supplmentary materials file :

Supplementary documents for manuscript submission

If you wish to repeat our experiment, please refer to the following configuration :

tensorflow=2.8.2=gpu_py39hc0c9373_0
keras=2.8.0=py39h06a4308_0
shap=0.45.1
scikit-learn=1.4.1.post1
matplotlib=3.9.2

We only use torch to clean up our GPU devices (our device storage space is insufficient), so this file is not limited by the torch version

how to get result:

1、The following files need to be downloaded from the repository: (1) code file, (2) clean data and target file, (3) adjacency matrix file, (4) saved model file

2、Modify the file path of the following tags in the *06_BioGCN_load_model_and_test_pred.py* file:

data_path: training and testing data in ***cleaned_data_out.csv***

graph_path: The adjacency matrix constructed based on the biological prior knowledge required for BioGCN in ***adjacency_matrix***

target_path: training and testing label in ***cleaned_data_test_12_7.xlsx***

best_model_path: Import the saved model in ***save_model(0.4)_18***

Modify the file path of the cleaned_data_and_target file in the *06_BioGCN_load_model_and_test_pred.py* file, and then run it. If you want to predict hard samples, please set ***Whether_to_predict_hard_stamples = TRUE***, If you want to calculate the shap, please set ***Whether_to_calculate_the_shap = TRUE***