This repository contains python and R implementation of the algorithms proposed in "Diagnostic classification of cancers using XGBoost and multi-omics data".

The data used in this research are collected from The Cancer Genome Atlas (TCGA) project and that are publicly available at <https://portal.gdc.cancer.gov>.

The analysis is divided into two sections saved in this repository.

1. Optimizing different classifier, which include Extreme gradient boosting (XGBoost), support vector machine (SVM), random forest (RF), Elastic Net (EN),

Naive Bayes (NB), K-Nearest Neighbor (KNN).

XGBoost-methy-optimation.py: Optimizing the hyper-parameters of XGBoost model SVM-methy-optimation.py: Optimizing the hyper-parameters of SVM model

RF-methy-optimation.py: Optimizing the hyper-parameters of RF model

EN-methy-optimation.py: Optimizing the hyper-parameters of Elastic Net model

KNN-methy-optimation.py: Optimizing the hyper-parameters of KNN model

NB-methy-optimation.py: Optimizing the hyper-parameters of NB model

1. Test procedure for different classification model.

XGBoost-methy-test.py: Test procedure for XGBoost

SVM-methy-test.py: Test procedure for SVM

RF-methy- test.py: Test procedure for RF

EN-methy- test.py: Test procedure for XGBoost

KNN-methy-test.py: Test procedure for KNN

NB-methy-test.py: Test procedure for NB

The files in the repository are examples, which construct different predictive models to classify the stage (early or late) of KIRP patients. For the other three types of cancer, you can find the parameters used by the different machine learning methods in obtaining the best classification performance in 'parameters \_table’.