Multi-Omics enrichment pathway score analysis

Jaemin Jeon[1], Inuk Jung[1*]
[1]School of Computer Science and Engineering, Kyungpook National University, Daegu

Background

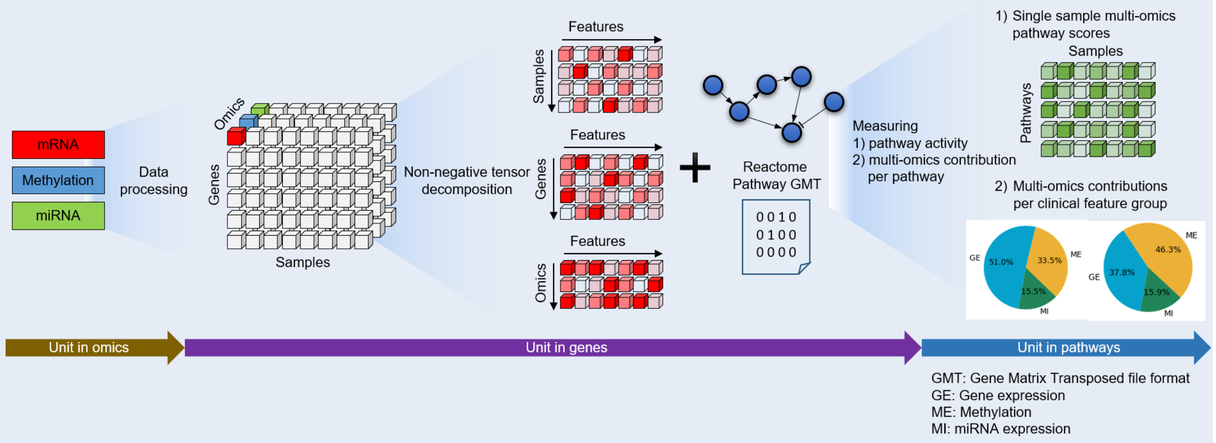
Pathway analysis is used in summarizing individual biological mechanisms in interpretable pathway unit. Pathway is group of genes that are related to each biological mechanism. In human disease, difference of gene expression is measured and various omics are involved. Because of large scale of biological system, single omics is limited to explain complex system. This study helps to interpret individual patients' diseases in pathway units with multi-omics.

Limitation of current methods

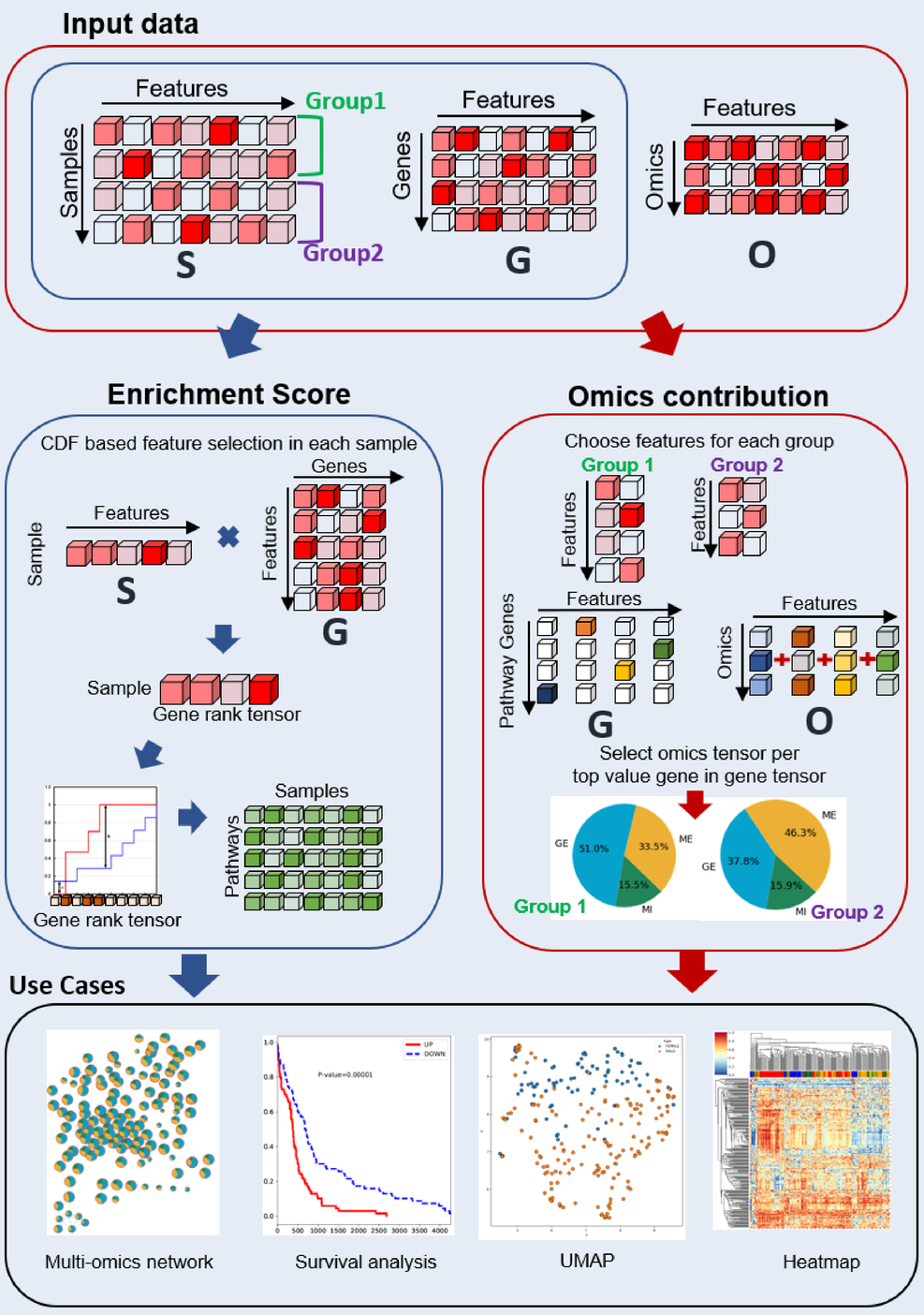
Most of pathway analysis is limited in using single omics. There are not many methods to see a single sample , and there are few cases where clinical data is used in pathway analysis.

Proposed Method

We used three omics(mRNA, methylation level, miRNA). To mix up multi- omics data we used tensor decomposition. Because of different number of each omics, we changed all of omics to gene-centric to overcome dimension difference.



Using decomposed sample, gene, omics matrix, we calculated Enrichment score and Omics contribution with multiplying. In Enrichment score , K-S random statistic is used. We can make multi-omics network or survival analysis with these.

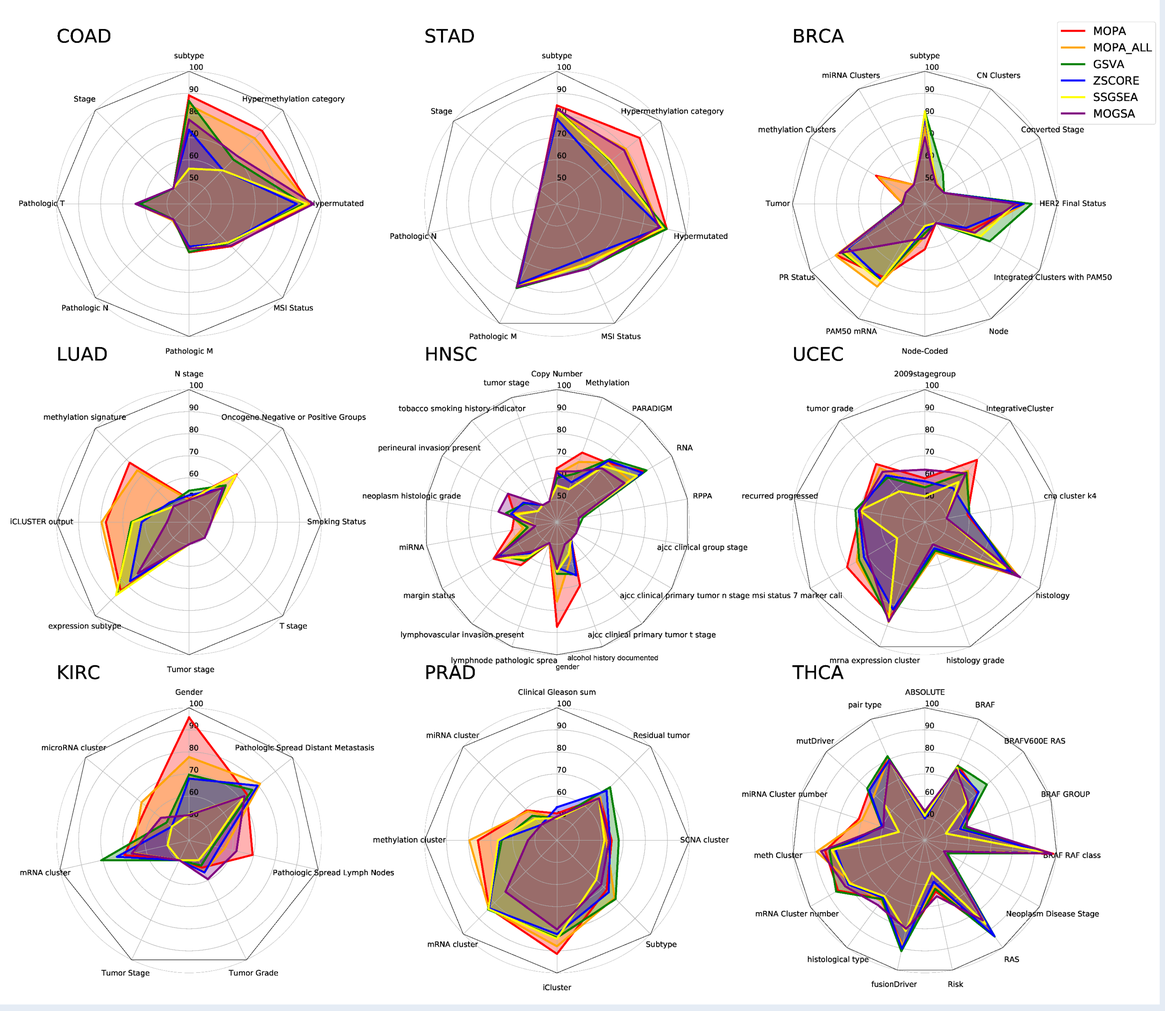


Data set

Among the TCGA data, 94 clinical data according to nine cancer types.
(COAD,STAD,LUAD,BRCA,UCEC,HNSC,PRAD,THCA,KIRC).
For pathway data, we used KEGG database(340 pathways)
14,464 genes are used and mRNA, miRNA, methylation-level omics are used

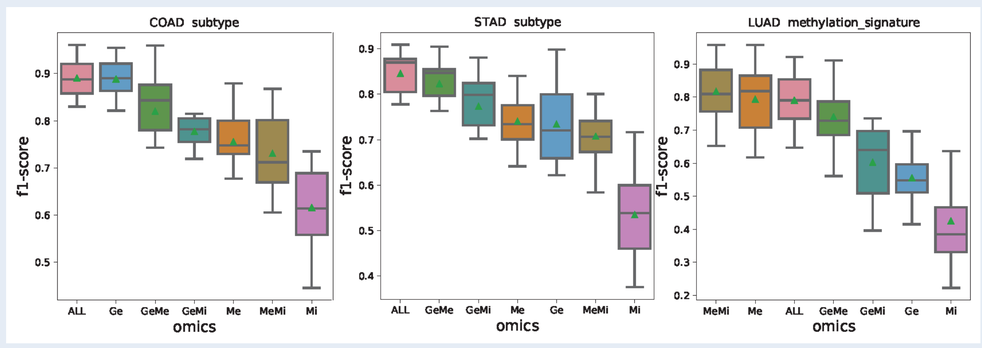
Classification result

F1-score comparison with other method.



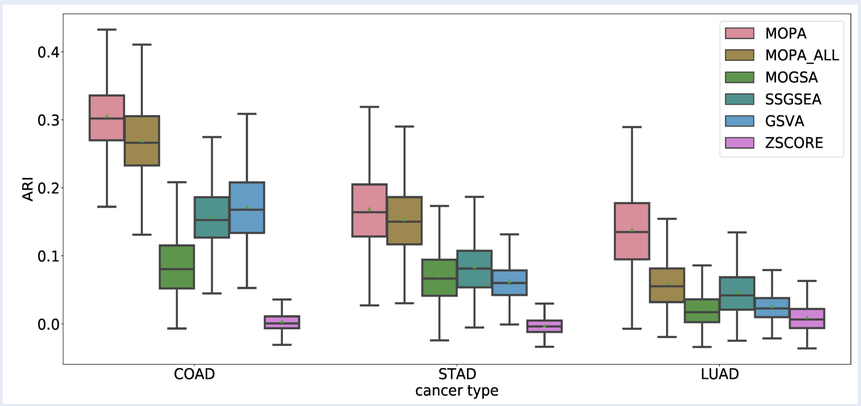
Omics combination performance

F1-score according to the combination of omics used. To show multi-omics is better than using single omics for classification we compared with combination of omics.
As below figure, we can see using all of omics made better performance than using 2 or 1 omics. In LUAD, however using methylation and gene expression showed better performance. It is because, clinical data was classified by methylation signature. Also using 3 omics didn't made big performance drop



Clustering result

Adjusted rand index comparison with other methods. Bootstrapping 10% samples with 1000 iterations. MOPA_ALL method is not using labeled data which means unsupervised method. We considered when we don't have labeled data and we must use unclassified data. However as we can see MOPA_ALL method had similar performance when we compare with MOPA. We can use in unlabeled data.



This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2020M3C9A5085604).