

Gmade Studio 7 月份工作会报:Part of Hematology

Yanhua Zheng
CMU1H & Gmade Studio

July 4, 2023

基于基因表达谱的 AML 预后相关通路的挖掘与探索



Gmade Studio

- 生物体系实际上是一种由不同的生物化学反应通路模块组成的分子网络系统。
- 来源于系统生物学

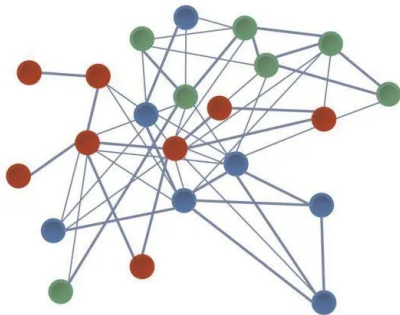


Figure 1.3 Network with nodes (circles) and edges (lines between circles). Different node colors indicate different types of connected components (e.g., proteins, mRNAs, and metabolites).

子路工作室

新思维：基于图的人工智能研究

图神经网络已经广泛应用于蛋白质设计（如 alphafold）、药物发现（如华为盘古大模型）、基因网络分析、医学影像学等医学研究任务中。

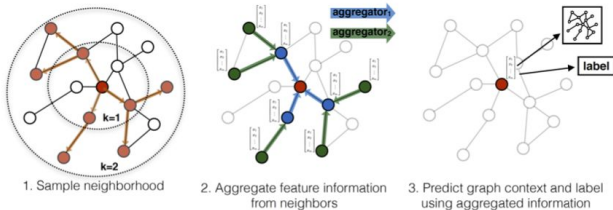


Figure: GNN

Methods: Materials

- 1 Open source codes in python computers necessary
- 2 RNA-seq within different clinical groups
- 3 At least 100 patients in total
- 4 Validation in lab remains unknown,so ...

Methods: Basic Workflow

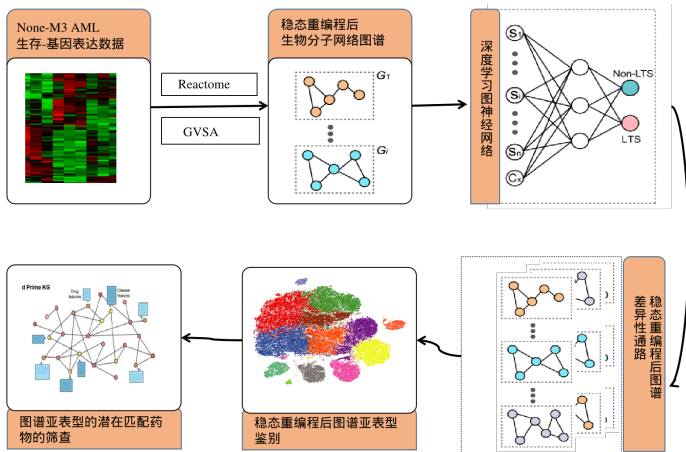


Figure: 技术路线

预实验结果



Gmade Studio

患者特征

TCGA 的 LAML 项目内共有患者 200 名，经排除后纳入分析共 132 人，其中死亡 80 人，存活 52 人，女性 60 人，男性 72 人，这些患者年龄平均 53 (23-88,min-max) 岁。

模型训练过程及验证

模型训练在 8 个周期后停止，在验证集上的准确率为 77%，模型 AUC 达 75%。

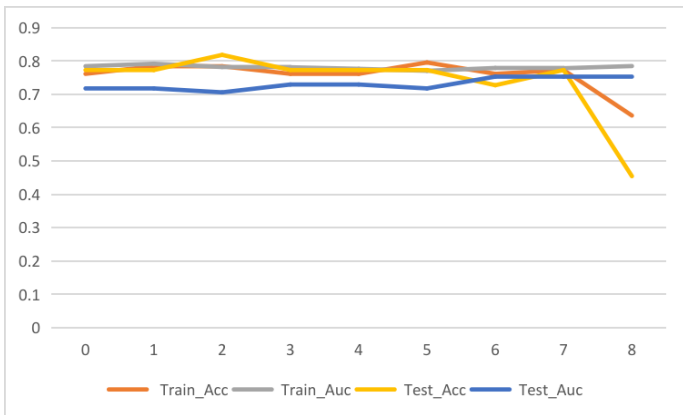


Figure: 模型训练过程可视化

模型解释与差异性通路

在 875 个通路中，寻找到 46 个对疾病演进结局具有显著差异的通路，其中前 15 个通路及其重要性评分从大到小排列见表（见表 1）。

模型解释与差异性通路

Pathway	Pvalue	Score
Incretin synthesis,secretion,and inactivation	0.033	0.452
Formation of Senescence-Associated HeterochromatinFoci(SAHF)	0.042	0.364
Signaling by NODAL	0.031	0.325
TCF dependent signaling in response to WNT	0.033	0.258
Transcriptional regulation of pluripotent stem cells	0.024	0.25
DNA Double-Strand Break Repair	0.035	0.23
Trafficking of GluR2-containing AMP Areceptors	0.049	0.209
Cellular response to chemical stress	0.045	0.208
Degradationoftheextracellularmatrix	0.022	0.207
Insulinreceptorsignallingcascade	0.026	0.189
TranscriptionalRegulationbyMECP2	0.022	0.171
XBP1(S)activateschaperonegenes	0.035	0.162
SCF(Skp2)-mediateddegradationofp27_p21	0.026	0.117
SignalingbyType1Insulin-likeGrowthFactor1Receptor(IGF1R)	0.033	0.108

Figure: 差异性通路

- 1 更高模型精度下的差异性通路探索
- 2 更优构架的 GNN 模型对比。
- 3 None-M3 AML 差异性通路图谱测绘，基于通路的亚型的聚类鉴别及药物匹配
- 4 重要通路的实验探索及验证
- 5 对匹配药物的细胞系/动物实验验证
- 6 差异性通路检验时，组间检验方式的变换？K-M 分析能否进一步突破