

基于集成学习的肿瘤药物敏感性预测研究

黄鹏杰^{1,2}, 林勇¹, 张梦欢³, 吕琳¹, 刘振浩², 裴潇倜¹, 许林锋^{1,2}, 谢鹭²

1. 上海理工大学医疗器械与食品学院, 上海 200093; 2. 上海生物信息技术研究中心, 上海 201203; 3. 中国科学院分子细胞科学卓越创新中心, 上海 200031

【摘要】肿瘤药物敏感性预测对个性化精准用药具有重要意义。本文基于GDSC数据库通过Boosting集成学习构建了面向RNA-seq基因表达和癌症药物敏感性数据的预测模型。先将183种药物集分别做归一化处理和基因特征降维,接着用AdaBoost集成SVM的方法建模,并采用十折交叉验证。实验结果表明构建的预测模型具有较高的预测精度,13种药物的AUC大于0.95,108种大于0.9,174种大于0.8。对比验证实验中,AdaBoost+SVM相比单学习器模型在整体药物集的综合评价指标中约提高4%,与其他集成模型相比提高2%。同时本文探讨了药物特异性,通过特征选择和富集分析对药物作用通路进行验证,从生物学角度提供了模型可解释性,证明其应用于临床用药指导的价值。

【关键词】集成学习;肿瘤;药物敏感性预测;AdaBoost;富集分析

【中图分类号】R318;R917

【文献标志码】A

【文章编号】1005-202X(2021)04-0511-07

Predicting anti-tumor drug sensitivity based on ensemble learning

HUANG Pengjie^{1,2}, LIN Yong¹, ZHANG Menghuan³, LÜ Lin¹, LIU Zhenhao², PEI Xiaoti¹, XU Linfeng^{1,2}, XIE Lu²

1. School of Medical Instrument and Food Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China; 2. Shanghai Center for Bioinformation Technology, Shanghai 201203, China; 3. Center for Excellence in Molecular Cell Science, Chinese Academy of Sciences, Shanghai 200031, China

Abstract: The prediction of anti-tumor drug sensitivity is of great significance for personalized and precise medication. Herein a prediction model for RNA-seq gene expression and anti-cancer drug sensitivity data is established based on GDSC database through Boosting ensemble learning. A total of 183 drug sets are normalized, and gene feature dimensionality is reduced. Then, AdaBoost+SVM is used for modeling, and 10-fold cross validation for verifying. The experimental results show that the established prediction model has a high prediction accuracy. The AUC of 13, 108 and 174 drugs are greater than 0.95, 0.90 and 0.80, respectively. AdaBoost+SVM improves the comprehensive evaluation index of the overall drug set by about 4% and 2%, compared with the models based on a learner only and other ensemble models. Meanwhile, drug specificity is also discussed; and the signal pathway of specific drug is verified through feature selection and enrichment analysis; and the interpretability of the established model is confirmed from a biological perspective. In sum, the value of the established model in clinical medication guidance is proved in the study.

Keywords: ensemble learning; tumor; drug sensitivity prediction; AdaBoost; enrichment analysis

前言

肿瘤药物敏感性预测研究是精准医学的重要分支。当前,肿瘤精准医学不断发展,基于生物标志物

和临床病理特点的抗肿瘤靶向治疗方法得到普遍应用^[1]。在合适的时间对不同癌症患者采取个性化预防、治疗和用药指导,能使某些类型肿瘤的客观缓解率和5年生存率均有较大程度提高,同时对癌症病人的副作用比传统方法小,这是精准医学在肿瘤领域应用的成功典范^[2]。

近年来,人工智能因为底层技术突破和海量数据驱动而不断发展,而生物信息领域随着第二代测序技术的发展产生大量的组学和药物反应数据,将两者结合能极大辅助肿瘤精准医学的发展。基因组学研究不断深入,大量研究表明肿瘤的产生、分化以

【收稿日期】2020-09-15

【基金项目】国家自然科学基金青年科学基金(31800700);上海市卫计委协同创新集群项目(2019CXJQ02);国家自然科学基金(31301092)

【作者简介】黄鹏杰,硕士研究生,研究方向:医学信息工程,E-mail: sl_pagelhuang@163.com

【通信作者】谢鹭,研究员,研究方向:生物信息学,E-mail: lux-iex2017@outlook.com

及药物治疗和基因密切相关^[3],通过基因表达谱来预测肿瘤药物的敏感性应答成为当下热点。Geleher等^[4]通过乳腺癌的细胞系和药物数据,建立了以岭回归(Ridge regression)算法为基础的基因表达和药物敏感性的回归模型;Menden等^[5]利用CGP数据集,通过神经网络算法来建立药物应答预测模型;基于随机森林算法,Riddick等^[6]利用NCI-60数据训练回归模型来预测药物敏感性;Gupta等^[7]研究了基于支持向量机和多种基因组类型的回归方法。

这些方法不仅促进了癌症药物基因组学的发展,也为预测药物敏感性提供了有效途径,但当前研究缺少对大批量药物和细胞系基因数据的有效处理,同时模型所采用算法不多且精准度还不够高。本文工作较上述研究具有以下特点:(1)药物种类多,且对细胞系和基因特征有效筛选用来建模;(2)采用集成学习算法,大部分药物预测准确率高,相比单模型有显著提升;(3)对特定药物进行功能分析,从生物学角度提供模型的可解释性。

本文基于GDSC2数据库中809个肿瘤细胞系样

本的RNA-seq数据和183种药物与相关细胞系反应的IC50值,经过样本筛选和基因特征降维后,结合boosting集成和4种机器学习方法来构建药物敏感性分类预测模型。实验结果表明,AdaBoost+SVM集成方法相比单模型和其他集成模型,在整体药物集上的ROC曲线下面积(AUC)、精准度(Precision)、准确率(Accuracy)、查全率(Recall)、综合评价指标(F_1)等指标中具有更好效果。同时本文以药物Nutlin-3A(-)为例进行特异性分析,通过特征排序筛选和富集分析,从生物学和统计学验证了该药物的作用靶向通路P53。综上,本文的模型能对肿瘤病人的临床精准用药起到指导和建议作用。

1 基于集成学习的肿瘤药物敏感性预测方法

对肿瘤药物敏感性预测的整体流程包括样本筛选、特征降维、机器学习建模、模型比较和药物特异性分析5个部分,如图1所示。下面对数据处理和流程实现做详细描述。

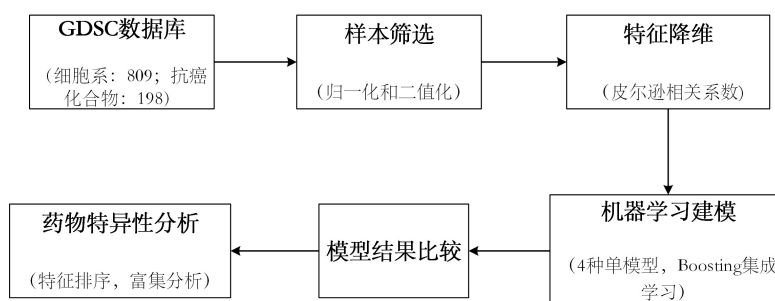


图1 分类预测方法流程图

Fig.1 Flowchart of classification prediction method

1.1 数据集和预处理

本研究数据集来源于肿瘤药敏多组学数据库(Genomics of Drug Sensitivity in Cancer, GDSC),该数据库有多于1 000种人类癌症细胞系的基因组特征和它们对不同抗癌药物的敏感性实验数据,是癌症细胞药物敏感性和药物反应分子标志物信息的最大公共资源。GDSC目前由GDSC1和GDSC2两个数据库组成,后者主要收录2015年至今的测序和试验结果,它使用代谢测定法(CellTiteGlo)来确定细胞活力,是基于改进的技术、设备和程序等得到的最新数据^[8]。

本文的数据集为GDSC2,包含有809个细胞系,198个抗癌化合物,以及135 242个药物和细胞系反应数据。其中药物反应指标为IC50值,即半抑制浓度,表示凋亡细胞与全部细胞数之比等于50%时所对应的药物浓度,反应细胞对药物的耐受程度。

IC50值越低,则说明细胞对药物越敏感。此外,为了精确匹配相关细胞系的基因表达值,笔者从ArrayExpress网站获得编号为E-MTAB-3610的数据,它包含GDSC中所有细胞系的原始基因芯片^[9]。接着,使用R包Affy读取CEL格式文件并提取为基因表达矩阵,同时利用该包中的Robust Multi-array Average (RMA)函数对原始数据进行标准化,以达到背景校正、分位数标准化、对探测强度进行log转换的目的。RMA标准化后,得到1 018个细胞系对应的17 737个基因表达值,其中去除基因名重复和空缺的数据后,还剩17 418个基因做后续分析。

1.2 细胞系样本筛选和特征基因降维

对单个药物处理来说,本文先对细胞系实验样本的IC50值进行归一化处理:先计算均值,每个IC50值减去均值后再除以方差。然后对归一化后的数据由小到大进行排序,前后20%占比的样本分别视为

有或无药物应答反应,同时去除待验证的中间样本,将预测模型从回归变为二分类问题。接着将每种药物筛选后的数据集进行8/2分,分别用来训练和验证。此外,不同药物实验对应的样本数从48到809不等,经过以上步骤处理后有15种药物的样本数小于20,样本过少不利于建模和十折交叉验证,因此这15种不在本次研究范围。

接着,通过R语言匹配细胞系样本的基因表达值作为待输入特征,但选取的特征基因数过多,会提高计算复杂度、降低预测准确性,因此对特征降维极其重要。本文将待训练样本归一化后的IC50值和单个基因的表达值进行皮尔逊相关性检验,通过R中的Cor函数选出相关系数绝对值大于0.3, P值小于0.05的基因。最后,对不同药物来说,特征基因数从17 418降维到几十至几千不等。

1.3 基于集成学习的药物敏感性预测建模

1.3.1 单模型建模 本研究使用的单模型是4种经典的机器学习模型,分别为支持向量机(Support Vector Machine, SVM)、随机森林(Random Forest, RF)、极端随机树(Extremely Randomized Trees, ET)和基于Bagging的决策树(Decision Tree Based on Bagging, BT)。其中,本研究集成的主要基分类器为SVM。

SVM是基于结构风险最小化和VC维理论的监督式学习方法,通过构造分隔面将数据进行分离,在解决非线性、高维模式识别和大样本分类中表现出独特优势^[10]。建模中,给定的肿瘤药物预测的样本集 $\{(x_i, y_i), i=1, 2, \dots, n\}$, x_i 表示预测因子基因特征向量(L维), y_i 表示某细胞系对某药物是否有应答。对于这种高维非线性预测问题, SVM通过将原始空间映射到高维特征空间来建立分类预测模型。分为以下几步:(1)建立初始超平面: $w \cdot \phi(x_i) + b = 0$, 其中, w 为权值矢量, b 为阈值, ϕ 为目标函数。(2)寻找最优超平面,其可以归纳为求解如下优化问题: $\min \phi(w, \xi_i) = \frac{1}{2} \|w\|^2 + C \sum \xi_i$; s.t. $(w \cdot \phi(x_i) + b) - 1 + \xi_i \geq 0, \xi_i \geq 0$, 其中 C 为惩罚参数, ξ_i 为非负松弛因子。(3)接着求解得到最优分类决策函数: $f(x) = \text{sgn}(\sum_{i=1}^n \alpha_i^* y_i K(x_i, x_j) + b^*)$; $b^* = y_j - \sum_{i=1}^n y_i \alpha_i^* K(x_i, x_j)$, 其中, α_i^* 为最优拉格朗日乘子, b^* 为分类阈值且为核函数, $\text{sgn}(\cdot)$ 表示返回整型变量的函数, x_j 表示不同于 x_i 的样本点。

RF、ET、BT都可看做Bagging集成的变种,本文将将其视为Boosting集成的基分类器。RF将bootstrap重抽样方法和决策树算法相结合,能对重要特征进行排序,是性能良好的用于分类或回归的有监督算

法^[11]。ET与RF相似,都由许多决策树构成,但它是完全随机得到分叉值进而对回归树分叉,不同于RF在一个随机子集内得到最佳分叉属性^[12]。此外,该方法每棵回归树都使用全部训练样本。BT是完全用Bagging分类器集成决策树的方法。

1.3.2 Boosting集成学习建模 AdaBoost是最常见的一种Boosting算法,能有效应用在两类、多类单标签等分类问题中,具有较强自适应能力、速度快和算法复杂度低等优点^[13]。其主要思想是给每个训练样本赋予一个初始权值,再根据每次训练的样本分类是否正确以及总体准确率,来自动调整权值。接着将调整后的新数据集传给下个分类器,最后对每次训练的基分类器进行加权组合得到最终决策分类器。

本研究将该方法用在肿瘤细胞系对药物敏感性的分类中。设输入的训练样本为: $\{(x_1, y_1), (x_2, y_2), \dots, (x_i, y_i), \dots, (x_n, y_n)\}$, 其中 x_i 中表示输入的带有基因表达信息的细胞系样本, $y_i \in \{0, 1\}$ 分别表示正样本和负样本,其中正样本数为1,负样本数 $m, n=1+m$ 。具体步骤如下:

(1)初始化每个细胞系样本的权重 $w_i, i \in D(i)$, 在每个弱分类器 t 中 ($t=1, \dots, T, T$ 为总个数), 其权重可以归一化为一个概率分布:

$$w_{t,i} = \frac{w_{t,i}}{\sum_{j=1}^n w_{t,j}} \quad (1)$$

(2)对每个特征训练一个弱分类器 h_t 。由于本研究为二值分类,可通过遍历得到阈值,该阈值使错分的样本最小且错误率最低。接着计算包含所有特征的弱分类器的加权错误率:

$$\varepsilon_j = \sum_{i=1}^n w_i(x_i) |h_j(x_i) \neq y_i| \quad (2)$$

(3)选取错误率 ε_t 最小的弱分类器 h_t , 并按照此最佳 h_t 调整权重:

$$w_{t+1,i} = w_{t,i} \beta_i^{1-\varepsilon_i} \quad (3)$$

其中, $\varepsilon_i=0$ 表示被正确地分类; $\varepsilon_i=1$ 表示被错误地分类。

$$\beta_t = \frac{\varepsilon_t}{1 - \varepsilon_t} \quad (4)$$

(4)加权得到最终的决策强分类器:

$$h(x) = \begin{cases} 1, & \sum_{i=1}^T \alpha_i h_i(x) \geq \frac{1}{2} \sum_{i=1}^T \alpha_i \\ 0, & \text{其他} \end{cases} \quad (5)$$

其中 α_i 的计算公式为:

$$\alpha_i = \log_2 \frac{1}{\beta_i} \quad (6)$$

经过以上一系列的迭代和集成处理后,我们可

以得到药物敏感性预测的最终强分类器,来判定肿瘤细胞系能否对该药物产生应答反应。AdaBoost默认的基分类器为决策树,本研究分别将SVM、RF、ET、BT作为基分类器进行同质集成。基分类器个数对预测精度会有影响,本文分别设置为3、10、20、50进行实验,发现个数为10时达到理想效果,即使增加到50准确率没有明显提高,反而会加大内存消耗和训练成本。因此,本文基分类器数量定为10。

1.3.3 十折交叉验证 交叉验证(Cross Validation),也称循环估计,是一种统计学上将样本切割成较小子集的使用方法,常用于评估统计分析、机器学习算法对独立于训练数据的数据集的泛化能力^[14]。

十折交叉验证(10-fold cross validation),将数据集分成10份,轮流将其中9份做训练,1份做验证,10次结果的均值作为对算法精度的估计。本研究建模过程中均使用了十折交叉验证求均值的方法,其优势在于同时重复运用随机产生的子样本进行训练和验证,以求药物敏感性预测结果更精准。

2 实验验证和结果分析

2.1 实验评估指标

为了评估模型好坏,对比4种分类模型的预测精度和泛化能力。本研究采用以下常用评价指标:AUC, Precision, Recall, F₁值, Accuracy。在二分类中,“正例(positive)”和“负例(negative)”指分类标签,而“真(true)”和“假(false)”指模型预测结果是否与测试标签相对应。可知,TP、TN表示分类结果正

确;FP、FN表示分类结果错误。各项指标的定义如下:

$$\text{Precision} = \frac{TP}{TP + FP}$$
(7)

$$\text{Recall} = \frac{TP}{TP + FN}$$
(8)

$$F_1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$
(9)

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + TN + FP + FN)}$$
(10)

其中,TP为真阳性,表示实验细胞系样本有药物应答,预测模型中也有。FP为假阳性,表示实验细胞系样本无药物应答,但预测模型中有。FN为假阴性,表示实验细胞系样本有药物应答,但预测模型中没有。ROC曲线在坐标系中越偏向左上角,即AUC越大,则表示分类器的性能越好。

2.2 实验结果与分析

2.2.1 药物模型整体比较 针对183种药物的建模结果,本文详细比较了Boosting集成方法和4种单模型的各项评估指标。由图2所有药物的AUC分布情况可知,AdaBoost+SVM集成模型有108种药物集的AUC>0.9,AUC在>0.95、0.9~0.95、0.85~0.9、0.8~0.85、0.75~0.8、<0.75等区间的药物数占比依次为7.1%、51.9%、24.6%、11.4%、3.8%和1.1%,总体效果明显优于SVM、RF、ET、BT和它们的集成模型。此外,其他3种集成模型与各自的基学习器相比在大部分药物的准确率上也有所提升。

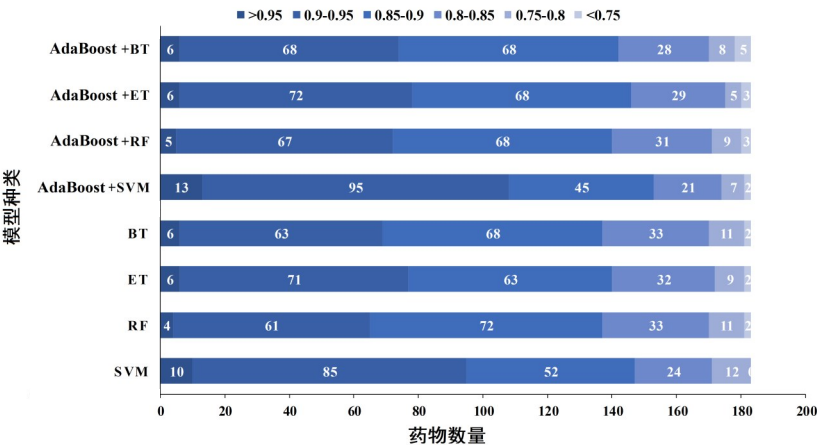


图2 183种药物在单模型和boosting集成模型的AUC分布情况
Fig.2 AUC distribution of 183 drugs in the models based on a learner only and Boosting ensemble models

得到整体分析结果后,笔者选取在所有模型中都取得较高准确率的药物做进一步分析,分别为Venetoclax、OSI-027、Nutlin-3a(-)、AZD5991、Vorinostat、Oxaliplatin、Dasatinib、Niraparib、Paclitaxel、Daporinad、Camptothecin、Cisplatin、MIRA-1。图3是这13种药物

在集成模型和SVM中的AUC对比图,可知,AdaBoost+SVM模型的预测效果最优,且AUC都大于0.95。值得注意的是,对单个药物来说,取得最优效果的不一定都是AdaBoost+SVM,其他集成方法在个别药物数据集中也可能成为最优建模方法。

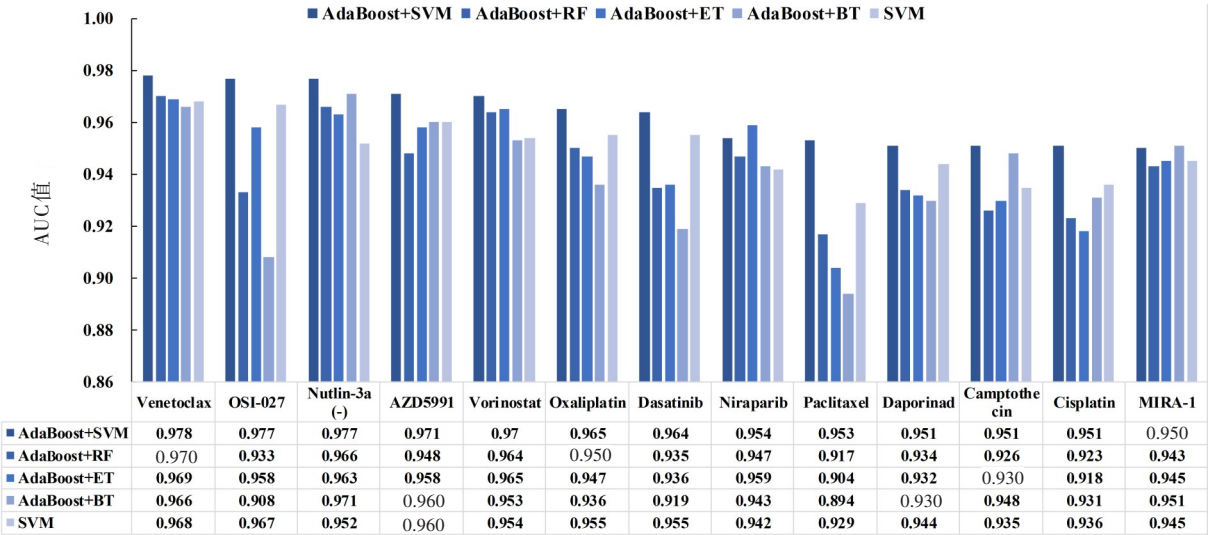


图3 高准确率药物预测模型比较
Fig.3 Comparison of high-accuracy drug prediction models

为了更详细展示每种药物在不同方法建模中的效果,笔者以药物Nutlin-3a(-)为例进行分析。表1是该药物在 Boosting 集成模型和 4 种单模型中的 AUC、Accuracy、Precision、Recall 以及 F₁ 值的具体数值和标准

差。图4为不同的模型进行十折交叉训练的ROC曲线,可以直观发现该药物在 AdaBoost+SVM 中效果最好,整体准确率相比单模型高出约 2.5%,相比其他集成模型高出约 1.5%,且各项指标的标准差相对最小。

表1 药物Nutlin-3a (-)在所有模型中的结果比较($\bar{x} \pm s$)
Tab.1 Comparison of prediction results of Nutlin-3a (-) in all models (Mean±SD)

模型	AUC	Accuracy	Precision	Recall	F ₁ 值
AdaBoost+SVM	0.977±0.017	0.925±0.043	0.945±0.063	0.908±0.08	0.922±0.048
AdaBoost+RF	0.966±0.023	0.883±0.061	0.933±0.04	0.823±0.115	0.871±0.073
AdaBoost+ET	0.963±0.028	0.879±0.067	0.917±0.068	0.836±0.123	0.869±0.08
AdaBoost+BT	0.971±0.022	0.922±0.046	0.927±0.074	0.911±0.058	0.912±0.047
SVM	0.952±0.037	0.867±0.072	0.951±0.055	0.77±0.136	0.846±0.09
RF	0.955±0.04	0.87±0.077	0.913±0.074	0.816±0.125	0.857±0.088
ET	0.956±0.031	0.889±0.069	0.935±0.057	0.835±0.124	0.877±0.084
BT	0.962±0.017	0.912±0.034	0.92±0.069	0.908±0.046	0.902±0.033

2.2.2 药物特异性分析 本文在所建模型基础上以 Nutlin-3a (-)为例进行了药物特异性分析。从 GDSC 数据库可知该抑制剂已验证的药物靶点为MDM2,作用通路为P53。如图5所示,本文通过随机森林模型选取该药数据集的重要特征基因进行富集分析,从生物学和统计学两方面验证了该药物的P53靶向通路^[15]。该生物功能分析为模型提供了可解释性和生物理论支撑。

3 总结与展望

本文针对 GDSC2 中的 183 种抗癌药物和 809 个细胞系数据,经过有效的样本筛选和基因特征降维后,首次使用 AdaBoost+SVM 集成方法建模来预测不

同细胞系对药物的敏感性。此外,通过比较 AUC、Accuracy、Precision、Recall 和 F₁ 值后,发现该模型相比 4 种单模型和其他集成方法在整体药物数据集中效果最优。与现有研究的模型相比,不仅具有较高精确度,而且通过药物特异性分析为模型提供了生物可解释性。以上研究证实了将集成学习方法应用在肿瘤药物敏感性预测的可行性,能在临床用药的筛选中发挥作用。实际应用中,可以先对病人的肿瘤组织进行基因测序,再使用模型去预测特定抗癌抑制剂对该病人是否有效,以达到提高用药疗效和减轻副作用的目的^[16]。

本研究的不足之处在于目前的模型只有基因表达数据作为输入特征,虽然基于单一组学进行疾病

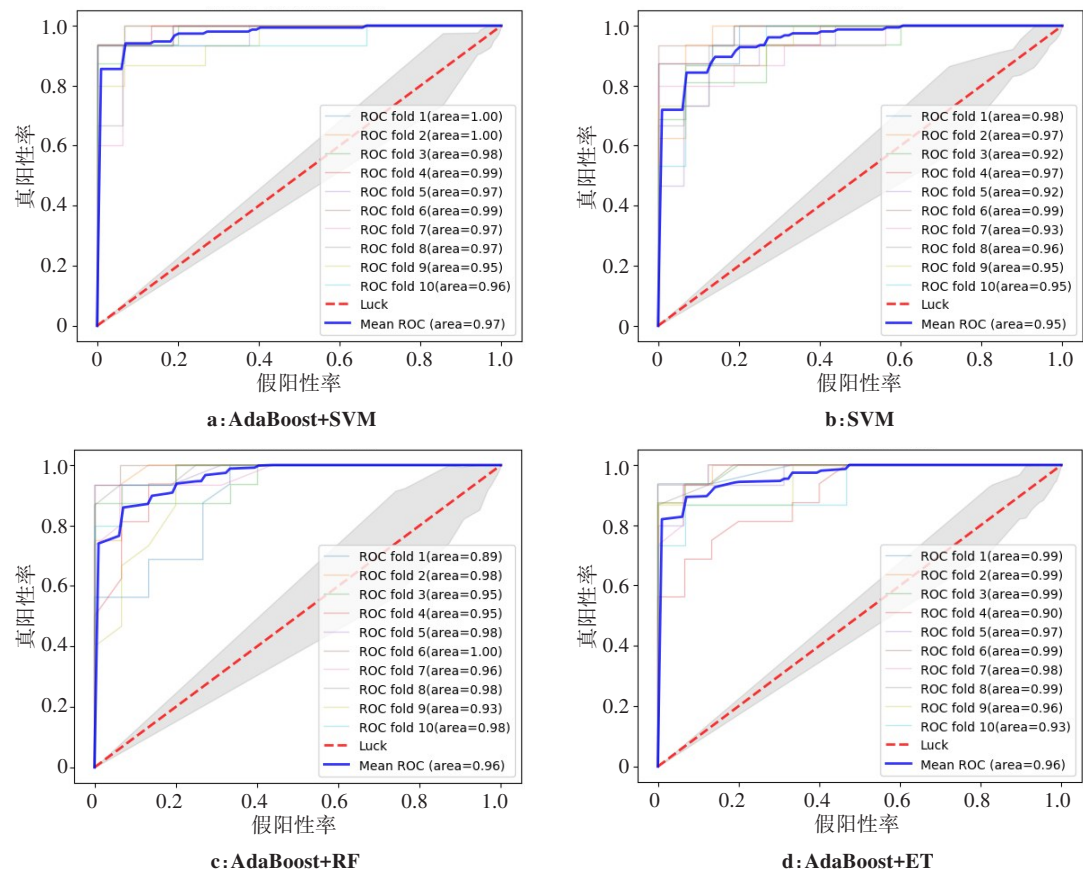


图4 Nutlin-3a (-)在 AdaBoost+SVM和其他模型中的 ROC 曲线
Fig.4 ROC curves of Nutlin-3a (-) in AdaBoost+SVM and other models

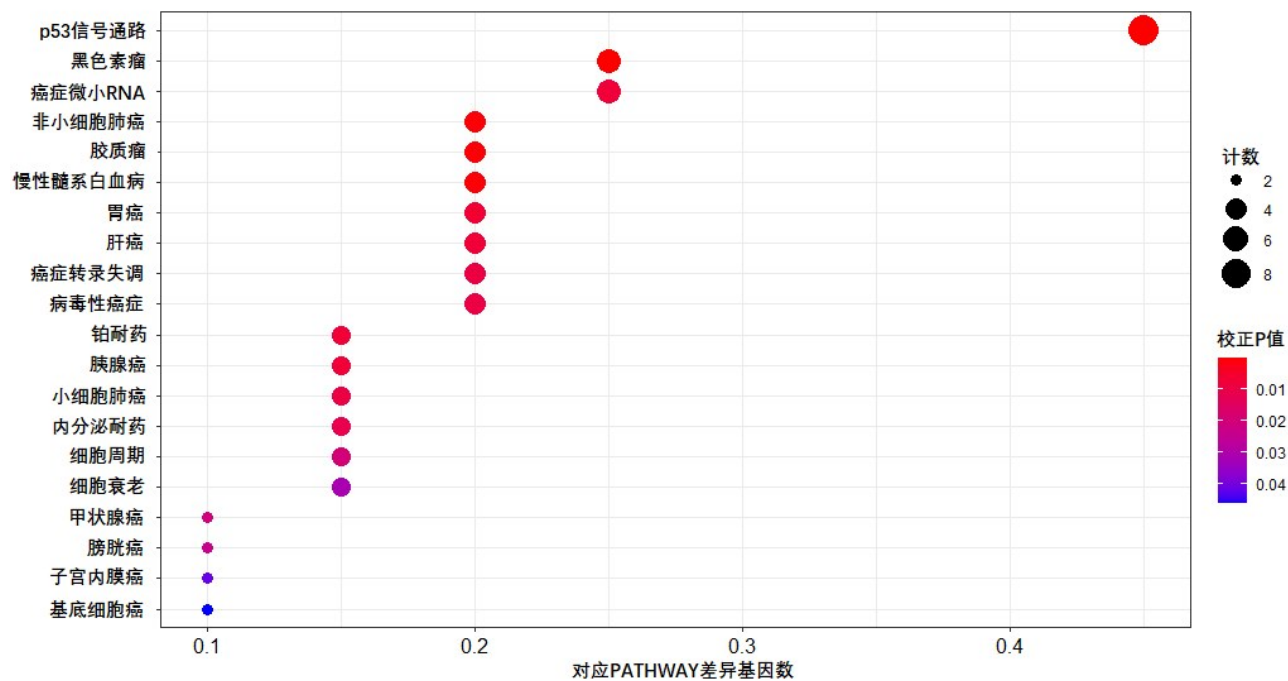


图5 Nutlin-3a (-)药物集重要特征的KEGG 富集结果
Fig.5 KEGG enrichment result of important characteristics of Nutlin-3A (-) drug set

研究已经发现了诸多新的致病因子,但疾病的发生发展是一个复杂的网络,基因拷贝数变异、体细胞突变、甲基化的异常等诸多因素都会影响生命体特征和药物反应^[17]。因此,之后的研究会处理以上多组学数据作为训练特征,同时尝试更多方法建模,以达到更好的肿瘤药物预测效果。

【参考文献】

- [1] LEE J K, LIU Z Q, SA J K, et al. Pharmacogenomic landscape of patient-derived tumor cells informs precision oncology therapy [J]. Nat Genet, 2018, 50(10): 1399-1411.
- [2] ZEHIR A, BENAYED R, SHAH R H, et al. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients[J]. Nat Med, 2017, 23(6): 703-713.
- [3] MARQUART J, CHEN E Y, PRASAD V. Estimation of the percentage of US patients with cancer who benefit from genome-driven oncology [J]. JAMA Oncol, 2018, 4(8): 1093-1098.
- [4] GEELEHER P, COX N J, HUANG R S. Clinical drug response can be predicted using baseline gene expression levels and *in vitro* drug sensitivity in cell lines[J]. Genome Biol, 2014, 15(3): R47.
- [5] MENDEN M P, IORIO F, GARNETT M, et al. Machine learning prediction of cancer cell sensitivity to drugs based on genomic and chemical properties[J]. PLoS One, 2013, 8(4): e61318.
- [6] RIDDICK G, SONG H, AHN S, et al. Predicting *in vitro* drug sensitivity using random forests [J]. Bioinformatics, 2011, 27(2): 220-224.
- [7] GUPTA S, CHAUDHARY K, KUMAR R, et al. Prioritization of anticancer drugs against a cancer using genomic features of cancer cells: a step towards personalized medicine[J]. Sci Rep, 2016, 6: 23857.
- [8] IORIO F, KNIJNENBURG T A, VIS D J, et al. A landscape of pharmacogenomic interactions in cancer [J]. Cell, 2016, 166(3): 740-754.
- [9] IRIZARRY R A, BOLSTAD B M, COLLIN F, et al. Summaries of affymetrix genechip probe level data[J]. Nucl Acids Res, 2003, 31(4): e15.
- [10] CORTES C, VAPNIK V. Support-vector networks [J]. Machine Learning, 1995, 20(3): 273-297.
- [11] BREIMAN L. Random forests[J]. Machine Learning, 2001, 45(1): 5-32.
- [12] GEURTS P, ERNST D, WEHENKEL L. Extremely randomized trees [J]. Machine Learning, 2006, 63(1): 3-42.
- [13] FREUND Y, SCHAPIRE R E. Experiments with a new boosting algorithm[C]. International Conference on Machine Learning, 1996: 148-156.
- [14] SAIN S R, BAGGERLY K A, SCOTT D W. Cross-validation of multivariate densities[J]. J Am Stat Assoc, 1994, 89: 807-817.
- [15] KENSUKE K, KONOPLEVA M, MCQUEEN T, et al. Mdm2 inhibitor Nutlin-3a induces p53-mediated apoptosis by transcription-dependent and transcription-independent mechanisms and may overcome Atm-mediated resistance to fludarabine in chronic lymphocytic leukemia [J]. Blood, 2006, 108(3): 993-1000.
- [16] MONTALBAN G, BONACHE S, MOLES-FERNÁNDEZ A, et al. Screening of BRCA1/2 deep intronic regions by targeted gene sequencing identifies the first germline BRCA1 variant causing pseudoexon activation in a patient with breast/ovarian cancer[J]. J Med Genet, 2019, 56(2): 63-74.
- [17] AMEMIYA T, GROMIHA, M M, HORIMOTO K, et al. Drug repositioning for dengue haemorrhagic fever by integrating multiple omics analyses[J]. Sci Rep, 2019(1): 523.

(编辑:薛泽玲)