

• 论 著 • DOI:10.3969/j.issn.1672-9455.2023.07.001

数据挖掘技术联合血浆 CEA、CYFRA21-1、AOC3、CLEC3B 及流行病学资料构建的模型对肺癌的诊断价值*

马明芮¹,倪静¹,何倩¹,安欣萍¹,玄路晗¹,朱晓彤¹,黄云霄¹,吴拥军²,段书音^{1△}

1. 山东第一医科大学公共卫生与健康管理学院, 山东济南 250001; 2. 郑州大学公共卫生学院, 河南郑州 450001

摘要:目的 分析基于数据挖掘技术联合血浆癌胚抗原(CEA)、细胞角蛋白 19 片段抗原 21-1(CYFRA21-1)、血管粘连蛋白 1(AOC3)、四连接素蛋白(CLEC3B)及流行病学资料构建的模型对肺癌的诊断价值。**方法** 经匹配年龄、性别、吸烟史、饮酒史、肿瘤家族史与肺癌家族史,该研究共收集肺癌患者(肺癌组)与健康对照组(健康对照组)各 68 例。采用 ELISA 检测血浆 CEA、CYFRA21-1、AOC3 与 CLEC3B 的表达水平;采用数据挖掘技术中的决策树 C5.0 模型、人工神经网络(ANN)模型、支持向量机(SVM)模型与传统 Fisher 判别分析模型联合 4 种蛋白及流行病学资料构建肺癌诊断模型。**结果** 肺癌组血浆 CEA 与 AOC3 表达水平高于健康对照组($P < 0.05$),CLEC3B 表达水平低于健康对照组($P < 0.05$),CYFRA21-1 表达水平在两组之间差异无统计学意义($P > 0.05$)。决策树 C5.0 模型、ANN 模型、SVM 模型与 Fisher 判别分析模型诊断肺癌的准确率分别为 72.09%、90.70%、81.40%与 76.74%,灵敏度分别为 50.00%、87.50%、70.83%与 62.50%,特异度分别为 100.00%、94.74%、94.74%与 94.74%,曲线下面积分别为 0.750、0.911、0.828 与 0.786;经对比,ANN 模型对肺癌的诊断效能最好。**结论** 基于血浆 CEA、CYFRA21-1、AOC3、CLEC3B 与流行病学资料构建的 ANN 模型对肺癌的诊断效能较高,具有潜在的临床应用价值。

关键词:肺癌; 数据挖掘; 人工神经网络模型; 支持向量机模型; 决策树 C5.0 模型

中图分类号:R734.2

文献标志码:A

文章编号:1672-9455(2023)07-0865-06

Diagnostic value of model constructed by data mining technology combined with plasma CEA, CYFRA21-1, AOC3, CLEC3B and epidemiological data in lung cancer*

MA Mingrui¹, NI Jing¹, HE Qian¹, AN Xinping¹, XUAN Luhan¹, ZHU Xiaotong¹, HUANG Yunxiao¹, WU Yongjun², DUAN Shuyin^{1△}

1. College of Public Health and Health Management, Shandong First Medical University, Jinan, Shandong 250001, China; 2. School of Public Health, Zhengzhou University, Zhengzhou, Henan 450001, China

Abstract: Objective To investigate the diagnostic value of the model constructed by the data mining technology combined with plasma carcinoembryonic antigen (CEA), cytokeratin 19 fragment antigen 21-1 (CYFRA21-1), amine oxidasecopper containing 3 (AOC3), tetranectin (CLEC3B) and epidemiological data in lung cancer. **Methods** Sixty-eight patients with lung cancer (lung cancer group) and 68 healthy controls (healthy control group) were collected in this study after matching age, sex, smoking history, drinking history, family history of tumor and family history of lung cancer. The expression levels of CEA, CYFRA21-1, AOC3 and CLEC3B in plasma were detected by ELISA. The decision tree C5.0 model, artificial neural network (ANN) model, support vector machine (SVM) model and traditional Fisher discriminant analysis model were adopted to combined with 4 kinds of proteins and epidemiological data to construct the lung cancer diagnosis model. **Results** The expression levels of plasma CEA and AOC3 in the lung cancer group were higher than those in the healthy control group ($P < 0.05$), the expression level of CLEC3B was lower than that in the healthy control group ($P < 0.05$), and the expression level of CYFRA21-1 had no statistically significant difference between the two groups ($P > 0.05$). The accuracies of decision tree C5.0 model, ANN model, SVM model and Fisher discriminant model in lung cancer were 72.09%, 90.70%, 81.40% and 76.74% respectively, and the sensitivities were 50.00%, 87.50%, 70.83% and 62.50% respectively. The specificities were 100.00%,

* 基金项目:大学生创新创业训练计划项目(S202110439051);山东省泰安市科技创新发展项目(政策引导类,2021NS269)。

作者简介:马明芮,女,在读本科,主要从事肺癌早期筛查模型构建方面的研究。△ 通信作者,E-mail:shuyinduan@126.com。

网络首发 [http://kns.cnki.net/kcms/detail//50.1167.R.20230213.1624.004.html\(2023-02-14\)](http://kns.cnki.net/kcms/detail//50.1167.R.20230213.1624.004.html(2023-02-14))

94.74%, 94.74% and 94.74% respectively, and the areas under the curve (AUCs) were 0.750, 0.911, 0.828 and 0.786 respectively. After the comparison, the ANN model had the best diagnostic efficiency for lung cancer. **Conclusion** The ANN model constructed by plasma CEA, CYFRA21-1, AOC3, CLEC3B and epidemiological data has the better diagnostic efficiency for lung cancer, and has the potential clinical application value.

Key words: lung cancer; data mining; artificial neural network model; support vector machine model; decision tree C5.0 model

肺癌是全球范围内发病率和病死率均较高的恶性肿瘤^[1]。研究显示,肺癌的5年生存率为4%~17%^[2-3],若能采用有效的筛查手段发现早期肺癌并规范治疗,可将肺癌5年生存率提高至70%。因此,研发肺癌筛查工具,制订肺癌早期检测策略,对降低肺癌的疾病负担尤为重要。常见的肺癌诊断方法包括组织和细胞学检查、影像学检查与肿瘤标志物检测等^[4-5]。组织和细胞学检查有创,且具有侵入性,不宜用于肺癌早期筛查^[6]。低剂量螺旋CT(LDCT)是国际上推荐使用的肺癌筛查手段,但仍存在假阳性率过高、放射性危害与过度诊断等局限性^[7-8]。目前,寻找高效灵敏的肺癌生物标志物仍然是研发肺癌筛查工具的重要任务,也是实现肺癌早发现、早诊断和早治疗的重要前提。数据挖掘技术能从大量不完整、模糊、有噪声、随机的数据中提取出未知的、潜在有用的信息,在肺癌的智能诊断方面具有传统统计方法无法比拟的优势^[9-10]。目前,常见的数据挖掘技术主要有神经网络(ANN)、支持向量机(SVM)与决策树等^[11-12]。本研究拟采用数据挖掘技术,联合传统肿瘤标志物癌胚抗原(CEA)、细胞角蛋白19片段抗原21-1(CYFRA21-1)与血管粘连蛋白1(AOC3)、四连接素蛋白(CLEC3B)及流行病学资料,构建智能型肺癌诊断模型,并与传统的Fisher判别分析模型对比,为肺癌筛查模型的研发提供参考依据。

1 资料与方法

1.1 一般资料 选取某医院呼吸内科与体检中心为临床研究现场,收集确诊为肺癌的患者68例纳入肺癌组。患者中肺腺癌36例,肺鳞癌10例,肺腺鳞癌2例,小细胞肺癌8例,未明确病理分型的肺癌12例。肺癌患者入组标准:(1)病理检查确诊的原发性肺癌;(2)具有完善的临床资料;(3)患者未经手术治疗、药物治疗与放化疗;(4)未合并其他器官的肿瘤且不具有其他器官的肿瘤史;(5)同意入组、依从性好。同时,收集该院体检中心的健康人群68例纳入健康对照组。经知情同意后,收集各组研究对象的性别、年龄、吸烟史、饮酒史、肿瘤家族史、肺癌家族史等基本资料。参照世界卫生组织(WHO)的建议,吸烟定义为每天抽烟大于或等于1支,一生中连续或累积抽烟6个月以上;饮酒定义为每周饮酒大于2次,连续饮酒6个月以上。本研究通过医院伦理委员会审核。

1.2 标本处理与分子检测 采集研究对象晨起空腹状态静脉血5 mL放置于含乙二胺四乙酸(EDTA)抗

凝剂的真空管中,轻微颠倒5次,于低温离心机中4℃条件下以1 500 r/min离心5 min,抽取其上层血浆置于冻存管,-80℃保存备用。采用紫外分光光度计(美国,BioTek有限公司)联合ELISA试剂盒(武汉,华美生物工程有限公司)检测肺癌患者与健康对照者血浆标本中CEA、CYFRA21-1、AOC3与CLEC3B的表达水平,具体操作依照说明书进行。

1.3 基于数据挖掘技术的肺癌诊断模型构建

1.3.1 数据归一化处理 本研究采用极差法对4种蛋白CEA、CYFRA21-1、AOC3与CLEC3B表达水平进行归一化处理,使输入变量分布在0~1,公式: $X_{\text{norm}} = (X - X_{\text{min}}) / (X_{\text{max}} - X_{\text{min}})$,X表示原始数据, X_{norm} 表示归一化后的数据, X_{max} 和 X_{min} 分别表示原始数据的最大值和最小值。

1.3.2 模型构建 在SPSS Clementine 12.0软件上建立决策树C5.0模型、ANN模型、SVM模型和Fisher判别分析模型。通过SPSS File节点输入数据,输入变量为流行病学资料和4种蛋白表达水平,其中流行病学资料包括吸烟史、饮酒史、肿瘤家族史和肺癌家族史,而4种蛋白为CEA、CYFRA21-1、AOC3和CLEC3B,共8个变量;采用Type节点对变量类型进行说明,吸烟史、饮酒史、肿瘤家族史与肺癌家族史均为二分类变量,4种蛋白表达水平为连续型变量;利用Partition节点对数据的分割比例进行设置,将样本按大致2:1的比例分为训练集和预测集,随机种子数设为1111111。

1.3.3 4种模型参数设置 经过多次训练和优化,决策树C5.0模型参数设置如下。Mode name: Auto; Use partitioned data: No; Output type: Decision tree; Group symbolics: No; Use boosting: Yes; Number of trials: 10; Cross-validate: No; Mode: Expert; Pruning severity: 75; Minimum records per child branch: 2; Use global pruning: Yes; Winnow attributes: No; Use misclassification costs: No; Model Evaluation: Calculate variable importance。

ANN模型参数设置如下。Mode name: Auto; Use partitioned data: Yes; Method: Exhaustive prune; Sample: 50%; Set random seed: No; Stop on: Time(mins) 1 min; Optimize: Memory; Continue training existing model: No; Use binary set encoding: Yes; Show feedback graph: Yes; Model selection: Use best network; Mode: Expert; Model Evaluation: Calculate variable

importance.

SVM 模型参数设置如下。Mode name: Auto; Use partitioned data: No; Mode: Expert; Stopping criteria: 1. 0E-3; Regularization parameter (C): 10; Regression precision (epsilon): 0. 1; Kernel type: Polynomial; Bias: 0; Gamma: 1; Degree: 3; Model Evaluation: Calculate variable importance.

Fisher 判别分析模型参数设置如下。Mode name: Auto; Use partitioned data: No; Method: Enter; Mode: Expert; Prior probabilities: All groups equal; Use covariance matrix: Within-groups; Model Evaluation: Calculate variable importance.

1.3.4 模型评估 模型采用筛检试验的评价指标进行评价, 主要指标包括准确度、灵敏度、特异度、阳性预测值、阴性预测值、曲线下面积(AUC)。

1.4 统计学处理 采用 SPSS21.0 软件进行统计分析。符合正态分布的计量资料采用 $\bar{x} \pm s$ 表示, 两组间比较采用 *t* 检验; 不符合正态分布的计量资料采用 $M(P_{25}, P_{75})$ 表示, 两组间比较采用 Mann-Whitney *U* 检验; 计数资料采用例数或百分率表示, 组间比较采用 χ^2 检验或 Fisher 确切概率法; 采用受试者工作特征(ROC)曲线分析模型的诊断效能。以 $P < 0.05$ 为差异有统计学意义。

2 结果

2.1 研究对象的基本特征 本研究共纳入研究对象

136 例, 其中健康对照组 68 例, 肺癌患者 68 例。健康对照组的平均年龄为 (70.63 ± 5.33) 岁, 肺癌组的平均年龄为 (70.19 ± 6.15) 岁, 两组人群年龄之间的差异无统计学意义 ($P > 0.05$); 两组人群的性别构成、吸烟史、饮酒史、肿瘤家族史与肺癌家族史比较, 差异均无统计学意义 ($P > 0.05$)。见表 1。

表 1 基本特征在两组间比较 [$\bar{x} \pm s$ 或 $n(\%)$]

项目	健康对照组 (<i>n</i> =68)	肺癌组 (<i>n</i> =68)	χ^2/t	<i>P</i>
年龄(岁)	70.63 ± 5.33	70.19 ± 6.15	0.477	0.656
性别			3.609	0.057
男	32(47.06)	44(64.71)		
女	36(52.94)	24(35.29)		
吸烟史	14(20.59)	11(16.18)	0.196	0.658
饮酒史	11(16.18)	8(11.76)	0.245	0.621
肿瘤家族史	0(0.00)	5(7.35)	—	0.058*
肺癌家族史	0(0.00)	3(4.41)	—	0.244*

注: —表示无数据; *为 Fisher 确切概率法的 *P* 值。

2.2 肺癌组与健康对照组血浆 4 种蛋白的表达水平比较 肺癌组血浆 CEA 与 AOC3 的表达水平高于健康对照组 ($P < 0.05$), CLEC3B 表达水平低于健康对照组 ($P < 0.05$); 然而, CYFRA21-1 表达水平在健康对照组与肺癌组之间的差异无统计学意义 ($P > 0.05$)。见表 2。

表 2 肺癌组与健康对照组血浆 4 种蛋白表达水平比较 [$M(P_{25}, P_{75})$]

组别	<i>n</i>	CEA(ng/mL)	CYFRA21-1(ng/mL)	AOC3(ng/mL)	CLEC3B(μg/mL)
健康对照组	68	1.47(1.26, 1.84)	0.91(0.76, 1.22)	7.04(5.02, 9.75)	1.06(0.75, 1.43)
肺癌组	68	2.99(1.55, 21.86)	0.96(0.75, 1.36)	9.69(7.51, 14.03)	0.95(0.69, 1.31)
<i>Z</i>		5.227	0.516	4.509	2.800
<i>P</i>		<0.001	0.606	<0.001	<0.05

2.3 模型构建

2.3.1 决策树 C5.0 模型 基于 4 种流行病学资料与 4 种血浆蛋白表达水平构建决策树 C5.0 模型。结果显示, 决策树 C5.0 模型在训练集的准确率为 80.65% (75/93), 其中对照全部被正确判别, 18 例肺癌被误判为对照; 其在预测集的准确率为 72.09% (31/43), 其中对照全部被正确判别, 12 例肺癌被误判为对照。见表 3。

表 3 决策树 C5.0 模型分析结果 (*n*)

组别	C5.0 训练集 (<i>n</i> =93)			C5.0 预测集 (<i>n</i> =43)		
	对照	肺癌	合计	对照	肺癌	合计
健康对照组	49	0	49	19	0	19
肺癌组	18	26	44	12	12	24
合计	67	26	93	31	12	43

2.3.2 ANN 模型 基于 4 种流行病学资料与 4 种血浆蛋白表达水平构建 ANN 模型。结果显示, ANN 模型在训练集的准确率为 86.02% (80/93), 其中对照全部被正确判别, 14 例肺癌被误判为对照; 其在预测集的准确率为 90.70% (39/43), 其中 1 例对照被误判为肺癌, 3 例肺癌被误判为对照。见表 4。

表 4 ANN 模型分析结果 (*n*)

组别	ANN 训练集 (<i>n</i> =93)		合计	ANN 预测集 (<i>n</i> =43)		合计
	对照	肺癌		对照	肺癌	
健康对照组	49	0	49	18	1	19
肺癌组	14	31	44	3	21	24
合计	63	31	93	21	22	43

2.3.3 SVM 模型 基于 4 种流行病学资料与 4 种血浆蛋白表达水平构建 SVM 模型。结果显示, SVM

模型在训练集的准确率为 86.02%(80/93),其中 2 例对照被误判为肺癌,11 例肺癌被误判为对照;其在预测集的准确率为 81.40%(35/43),其中 7 例肺癌被误判为对照,1 例对照被误判为肺癌。见表 5。

表 5 SVM 模型分析结果(n)

组别	SVM 训练集(n=93)		合计	SVM 预测集(n=43)		合计
	对照	肺癌		对照	肺癌	
健康对照组	47	2	49	18	1	19
肺癌组	11	33	44	7	17	24
合计	58	35	93	25	18	43

2.3.4 Fisher 判别分析模型 基于 4 种流行病学资料与 4 种血浆蛋白表达水平构建 Fisher 判别分析模型。结果显示,Fisher 判别分析模型在训练集的准确率为 74.19%(69/93),其中 5 例对照被误判为肺癌,19 例肺癌被误判为对照;其在预测集的准确率为 76.74%(33/43),其中 1 例对照被误判为肺癌,9 例肺癌被误判为对照。见表 6。

表 6 Fisher 判别分析模型分析结果(n)

组别	Fisher 训练集(n=93)		合计	Fisher 预测集(n=43)		合计
	对照	肺癌		对照	肺癌	
健康对照组	44	5	49	18	1	19
肺癌组	19	25	44	9	15	24
合计	63	30	93	27	16	43

表 7 4 种模型对预测集的分类效果评估

模型	准确率(%)	灵敏度(%)	特异度(%)	阳性预测值(%)	阴性预测值(%)	AUC(95%CI)
决策树 C5.0	72.09	50.00	100.0	100.0	61.29	0.750(0.59~0.869)
ANN	90.70	87.50	94.74	95.46	85.71	0.911(0.784~0.976)
SVM	81.40	70.83	94.74	94.44	72.00	0.828(0.682~0.926)
Fisher 判别分析	76.74	62.50	94.74	93.75	66.67	0.786(0.635~0.896)

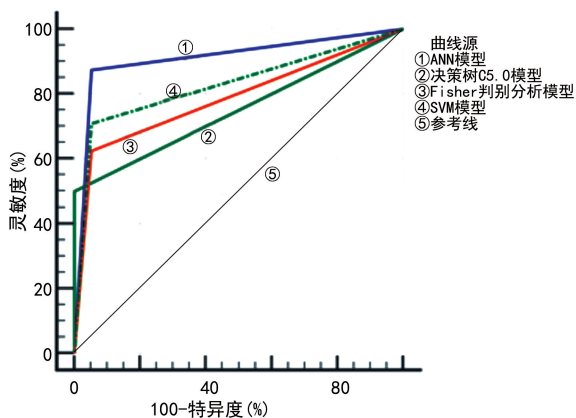


图 1 4 种模型对预测集分类的 ROC 曲线

3 讨论

全球癌症数据显示,2020 年肺癌新发病例 220 万,仅次于乳腺癌;肺癌死亡病例 180 万,居于所有癌症之首^[1]。鉴于较高的发病率与高死亡率^[13],肺癌已成为威胁人类健康的一大公共卫生问题。2018 年美国癌症报告显示:肺癌的 5 年生存率在 I 期为 55.1%,到 IV 期逐渐下降至 4.2%,而 I、II 期肺癌的诊断率仅为 25.3%,III、IV 期肺癌的诊断率为 66.9%^[14]。究其原因,肺癌发生早期无明显症状,不易发现,当出现明显症状就诊时多属晚期^[15-16]。研究

2.4 4 种模型预测效果评估 4 种模型预测集的 AUC 均在 0.75 及以上,提示模型具有较好的分类效果。经对比,ANN 模型的 AUC 高于决策树 C5.0 模型($Z=2.832, P<0.05$),但其与 SVM 模型、Fisher 判别分析模型之间的差异无统计学意义($P>0.05$)。然而,ANN 模型的准确率(90.70%)、灵敏度(87.50%)、阳性预测值(95.46%)与阴性预测值(85.71%)均高于 SVM 模型与 Fisher 判别分析模型,且 3 个模型的特异度相同,均为 94.74%。此外,ANN 模型的 AUC 高达 0.911,说明 ANN 模型对健康对照者与肺癌患者具有较好的区分能力。见表 7、图 1。

显示,在高危人群中实施肺癌筛查,及早发现肺癌并采取相应的干预措施,可以明显降低肺癌的病死率^[17]。然而,缺乏针对肺癌的有效早期检测手段成为制约肺癌筛查的重要原因^[18]。组织病理学检查是临床上诊断肺癌的金标准,但其取材困难,且具有创伤性,不宜用于肺癌早期检测,尤其是肺癌的筛查^[6]。影像学中的 LDCT 是目前较为认可的肺癌筛查手段,鉴于其存在假阳性率高、放射性危害等不足^[7,19-20],尚无法在普通人群中开展。

肿瘤标志物是指肿瘤的发生和发展过程中由肿瘤细胞或肿瘤微环境或机体对肿瘤细胞反应产生,可在血液、尿液、肿瘤组织等检测到的一类活性物质。CEA 与 CYFRA21-1 是目前临床上常用的肺癌分子标志物^[21-23]。CEA 是一种糖蛋白,在肺癌患者体内表达水平升高,尤其是肺腺癌,其诊断肺癌的灵敏度和特异度分别为 69%和 68%^[24]。CYFRA21-1 是肺鳞癌的标志物,在肺鳞癌患者血清中明显升高,其诊断非小细胞肺癌的灵敏度和特异度分别为 47%和 80%^[25-26]。本课题组前期采用数据挖掘技术联合肿瘤标志物(CEA、神经元特异性烯醇化酶、CYFRA21-1、血管内皮生长因子与胃泌素释放肽前体)与流行病学资料及临床症状构建的模型中 ANN 模型效果较

好,其 AUC 高达 0.889,灵敏度为 84.38%,特异度为 93.48%^[27]。虽然该模型诊断效果较好,但模型共纳入 6 项流行病学资料,8 项临床症状与 5 种分子标志物,变量多达 19 个,操作相对复杂,过多耗费人力、物力、财力。有研究显示,将传统肺癌生物标志物 CEA、CYFRA21-1、CA125 仅与一种新型分子标志物纽约食管鳞状细胞癌-1 联合构建早期肺癌诊断模型,其灵敏度可达 71%,特异度可达 83%,AUC 可达 0.820^[25]。

为了优化模型,本研究在保留传统肺癌分子标志物 CEA 与 CYFRA21-1 的基础上,纳入了两种新型蛋白分子 AOC3 与 CLEC3B。AOC3 主要参与淋巴细胞在肝脏和其他器官内皮细胞间的黏附与转运^[28]。据报道,AOC3 可以通过调控肿瘤周围淋巴细胞的转运与黏附影响癌症发生进程^[29]。CLEC3B 主要编码四连接素,可以通过激活纤溶酶原介导肿瘤的浸润与迁移^[30]。有研究显示,CLEC3B 在早期肺癌组织中表达下调,具有较高的诊断价值,可能成为肺癌免疫治疗的新靶标^[30]。本研究采用数据挖掘技术,基于以上 4 种蛋白分子,同时联合 4 种流行病学资料(吸烟史、饮酒史、肿瘤家族史、肺癌家族史)构建了 4 种肺癌诊断模型,分别为决策树 C5.0、ANN、SVM 模型与 Fisher 判别分析模型。通过 AUC 对比发现,ANN 模型的 AUC 高于决策树 C5.0 模型($P < 0.05$),但其与 SVM 模型、Fisher 判别分析模型之间的差异无统计学意义($P > 0.05$)。然而,ANN 模型的准确率(90.70%)、灵敏度(87.50%)、阳性预测值(95.46%)与阴性预测值(85.71%)均高于 SVM 模型与 Fisher 判别分析模型,且 3 个模型的特异度相同,均为 94.74%。此外,ANN 模型的 AUC 达 0.9 以上,说明 ANN 模型对肺癌具有较好的诊断效能。

综上所述,本研究采用数据挖掘技术联合分子标志物 CEA、CYFRA21-1、AOC3、CLEC3B 及流行病学资料成功构建了肺癌诊断模型,其中 ANN 模型诊断效能最好,为肺癌筛查工具的研发提供了参考依据。本研究对健康对照组和肺癌组进行了良好的匹配,使两组人群的性别、年龄、吸烟史、饮酒史、肿瘤家族史、肺癌家族史等均衡分布,增强了研究结果的可信性;然而,样本的匹配也为样本的收集增加了挑战,致使本研究样本数有限。在下一步研究中,一方面要收集更多的样本,扩大样本量进一步验证模型的性能;另一方面,将构建的 ANN 模型应用到高危人群中以验证模型对早期肺癌的检测效果。

参考文献

[1] SUNG H, FERLAY J, SIEGEL R L, et al. Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries[J]. CA

Cancer J Clin, 2021, 71(3):209-249.

[2] ZHANG H, JIN Z, CHENG L, et al. Integrative analysis of methylation and gene expression in lung adenocarcinoma and squamous cell lung carcinoma[J]. Front Bioeng Biotechnol, 2020, 8:3.

[3] MANGIAMELI G, CIOFFI U, TESTORI A. Lung cancer treatment: from tradition to innovation[J]. Front Oncol, 2022, 12:858242.

[4] LEE J H, HWANG E J, KIM H, et al. A narrative review of deep learning applications in lung cancer research: from screening to prognostication[J]. Transl Lung Cancer Res, 2022, 11(6):1217-1229.

[5] LI C, WANG H, JIANG Y, et al. Advances in lung cancer screening and early detection[J]. Cancer Biol Med, 2022, 19(5):591-608.

[6] 刘海涛, 闫卫东. 周围型肺癌 CT 征象与病理学的对照研究[J]. 现代医用影像学, 2016, 25(4):735-736.

[7] KIM H, KIM H Y, GOO J M, et al. Lung cancer CT screening and lung-RADS in a tuberculosis-endemic country: the Korean Lung Cancer Screening Project (K-LUCAS)[J]. Radiology, 2020, 296(1):181-188.

[8] VAN MEERBEECK J P, O'DOWD E, WARD B, et al. Lung cancer screening: new perspective and challenges in Europe[J]. Cancers, 2022, 14(9):2343-2449.

[9] ZHANG K, LIU X, JIANG J, et al. Prediction of postoperative complications of pediatric cataract patients using data mining[J]. J Transl Med, 2019, 17(1):2-10.

[10] KAUR I, DOJA M N, AHMAD T. Data mining and machine learning in cancer survival research: an overview and future recommendations[J]. J Biomed Inform, 2022, 128:104026.

[11] WANG W, DING M, DUAN X, et al. Diagnostic value of plasma microRNAs for lung cancer using support vector machine model[J]. J Cancer, 2019, 10(21):5090-5098.

[12] BENNING L, PEINTNER A, PEINTNER L. Advances in and the applicability of machine learning-based screening and early detection approaches for cancer: a primer [J]. Cancers (Basel), 2022, 14(3):623.

[13] MITHOOWANI H, FEBBRARO M. Non-small-cell lung cancer in 2022: a review for general practitioners in oncology[J]. Curr Oncol, 2022, 29(3):1828-1839.

[14] CRONIN K A, LAKE A J, SCOTT S, et al. Annual report to the nation on the status of cancer, part I: national cancer statistics [J]. J Nation Cancer Instit, 2018, 124(13):2785-2800.

[15] 贾万理, 杨阳, 晋婕. CTCs、ProGRP、NSE 在小细胞肺癌诊断及预后评估中的临床价值[J]. 检验医学与临床, 2022, 19(8):1054-1056.

[16] GUO Q, LIU L, CHEN Z, et al. Current treatments for non-small cell lung cancer [J]. Front Oncol, 2022, 12:945102.

- hypertension, dementia, and aging[J]. *Int J Mol Sci*, 2021, 22(24):13178.
- [7] 杨思进, 张文利, 苏琦, 等. 新疆昌吉回族自治州原发性高血压患者汉族与回族、哈族高血压患者危险因素和临床特征的对比研究[J]. *中国社区医师*, 2021, 37(2):182-183.
- [8] 员战民, 刘晓蕊, 陈浩. 多民族人群同社区居住环境下慢性疾病的治疗与管理[J]. *当代医学*, 2022, 28(3):184-186.
- [9] 汪晓洲, 边惠萍, 杨蕾, 等. 高原地区原发性高血压与高原高血压患者的临床特征比较[J]. *中华高血压杂志*, 2022, 30(1):51-57.
- [10] 易天骄, 赵友云. CYP3A5、CYP2C9、AGTR1 基因多态性对单药治疗原发性高血压疗效的影响研究[J]. *国际检验医学杂志*, 2022, 43(13):1607-1611.
- [11] 李星辉, 侯彦军, 马光荣, 等. 血管紧张素原基因 M235T、T174M 多态性与东乡族人群原发性高血压的相关性[J]. *兰州大学学报(医学版)*, 2021, 47(4):18-23.
- [12] 史嘉翊, 王晓雪, 赵若晗, 等. 同型半胱氨酸及 5,10-亚甲基四氢叶酸还原酶 C677T 多态性与朝鲜族原发性高血压相关性研究[J]. *中国卫生检验杂志*, 2021, 31(13):1545-1548.
- [13] KHURANA V, GOSWAMI B. Angiotensin converting enzyme (ACE)[J]. *Clin Chim Acta*, 2022, 524:113-122.
- [14] BÂNHEGYI V, ENYEDI A, FÜLÖP G Á, et al. Human tissue angiotensin converting enzyme(ace) activity is regulated by genetic polymorphisms, posttranslational modifications, endogenous inhibitors and secretion in the serum, lungs and heart[J]. *Cells*, 2021, 10(7):1708.
- [15] 刘雪丽, 甘家丽, 孙林, 等. 中医药调节肾素-血管紧张素-醛固酮系统防治慢性心力衰竭研究进展[J]. *上海中医药杂志*, 2021, 55(12):100-104.
- [16] 马文霞, 刘芙蓉, 杨丹蓉, 等. 肾素-血管紧张素-醛固酮系统基因多态性与原发性高血压的研究进展[J]. *国外医学(医学地理分册)*, 2017, 38(2):207-210.
- [17] 焦晓璐, 秦彦文. 血管紧张素转换酶 2 与心血管疾病研究进展[J]. *基础医学与临床*, 2020, 40(5):701-706.
- [18] LIU J, FENG D, KAN X, et al. Polymorphisms in the CYP3A5 gene significantly affect the pharmacokinetics of sirolimus after kidney transplantation[J]. *Pharmacogenomics*, 2021, 22(14):903-912.
- [19] FAN Z, ZHENG D, WEN X, et al. CYP3A5 * 3 polymorphism and age affect tacrolimus blood trough concentration in myasthenia gravis patients[J]. *J Neuroimmunol*, 2021, 355:577571.
- [20] LIANG H, ZHANG X, MA Z, et al. Association of CYP3A5 gene polymorphisms and amlodipine-induced peripheral edema in Chinese han patients with essential hypertension[J]. *Pharmg Pers Med*, 2021, 14:189-197.
- [21] GANESAPANDIAN M, RAMASAMY K, ADITHAN S, et al. Influence of cytochrome P450 3A5 (CYP3A5) genetic polymorphism on dose-adjusted plasma levels of carbamazepine in epileptic patients in South Indian population[J]. *Indian J Pharmacol*, 2019, 51(6):384-388.
- (收稿日期:2022-08-21 修回日期:2022-12-23)
-
- (上接第 869 页)
- [17] 赫捷, 李霓, 陈万青, 等. 中国肺癌筛查与早诊早治指南(2021, 北京)[J]. *中国肿瘤*, 2021, 30(2):81-111.
- [18] 肖佳龙, 郑莹. 全球肺癌的流行及预防进展[J]. *中国癌症杂志*, 2020, 30(10):721-725.
- [19] 何敏, 黄燕. 肺癌早期诊断方法的研究进展[J]. *现代预防医学*, 2020, 47(23):4381-4384.
- [20] MALLER B, TANVETYANON T. Emerging approaches to complement low-dose computerized tomography for lung cancer screening: a narrative review [J]. *Cureus*, 2022, 14(7):e27309.
- [21] 周孟杰, 王军文, 杨传松, 等. 非肿瘤人群中血清癌胚抗原水平升高的研究进展[J]. *检验医学与临床*, 2022, 19(11):1569-1572.
- [22] 马丽娟, 王蕴, 李殿明. 肺癌患者血清 CEA、CYFRA21-1、ALP、ALB 与 KPS 评分的相关性及诊断价值分析[J]. *检验医学与临床*, 2022, 19(4):454-458.
- [23] HUANG H, YANG Y, ZHU Y, et al. Blood protein biomarkers in lung cancer[J]. *Cancer Lett*, 2022, 551:215886.
- [24] ZAMAY T N, ZAMAY G S, KOLOVSKAYA O S, et al. Current and prospective protein biomarkers of lung cancer[J]. *Cancers (Basel)*, 2017, 9(11):155-161.
- [25] DOSEEVA V, COLPITTS T, GAO G, et al. Performance of a multiplexed dual analyte immunoassay for the early detection of non-small cell lung cancer[J]. *J Transl Med*, 2015, 13(1):55-63.
- [26] YANG G, XIAO Z, TANG C, et al. Recent advances in biosensor for detection of lung cancer biomarkers[J]. *Biosens Bioelectron*, 2019, 141(1):111416.
- [27] DUAN S, CAO H, LIU H, et al. Development of a machine learning-based multimode diagnosis system for lung cancer[J]. *Aging*, 2020, 12(10):9840-9854.
- [28] SALMI M, JALKANEN S. Vascular adhesion protein-1: a cell surface amine oxidase in translation[J]. *Antio Redox Signal*, 2019, 30(3):314-332.
- [29] CHEN J, XI J, TIAN Y, et al. Identification, prioritization, and evaluation of glycoproteins for aggressive prostate cancer using quantitative glycoproteomics and antibody-based assays on tissue specimens[J]. *Proteomics*, 2013, 13(15):2268-2277.
- [30] SUN J, XIE T, JAMAL M, et al. CLEC3B as a potential diagnostic and prognostic biomarker in lung cancer and association with the immune microenvironment[J]. *Cancer Cell Int*, 2020, 20:106.
- (收稿日期:2022-06-11 修回日期:2022-11-09)