

doi:10.3969/j.issn.1005-3697.2026.04.005

✧智能医学研究专题✧

XGBoost-SHAP 肺结节早期识别可解释性框架构建

易付良¹, 李刚², 刘昕³, 向茹梅¹, 骆长玲¹, 邓丽春¹, 余秀莲¹, 周厚容¹, 高扬¹, 邹雪娜¹

(1. 自贡市第四人民医院公共卫生科, 四川 自贡 643000; 2. 自流井区疾病预防控制中心慢病所, 四川 自贡 643036; 3. 自贡市第一人民医院健康管理中心, 四川 自贡 643000)

【摘要】目的: 通过可解释机器学习实现肺结节早期识别与重要变量可视化解释, 助力肺癌精准防控与早诊早治。**方法:** 以肺癌高危且完成临床筛查的人群为研究对象, 提取其高危评估与影像检查结果; 依据《中国肺癌筛查标准(T/CPMA013-2020)》将受检者分为肺结节高危组与低危组; 经单因素分析筛选有意义变量作为预测变量, 以肺结节分组为因变量, 构建 XGBoost-SHAP 可解释性识别框架, 实现肺结节早期识别与结果可视化解释。**结果:** 共纳入 644 例肺癌高危受检者, 其中肺结节高危组 199 例(30.9%), XGBoost 模型识别肺结节的准确度为 0.914 6、敏感度为 0.758 7、特异度为 0.984 3、F1 值为 0.845 8、AUC 为 0.974 1。SHAP 算法显示, 吸烟量更大、暴露于同事/家人吸烟环境、做饭通风频次低、加工类食物摄入多、有石棉和氡等职业暴露、蛋白质和蔬菜水果摄入少、从事体力劳动的受检者肺结节增大风险更高。**结论:** 可解释性框架在肺结节早期识别中效果良好; 肺结节大小改变不仅与吸烟习惯、二手烟暴露、油烟暴露、石棉和氡职业暴露等传统危险因素相关, 还与受检者膳食习惯有关。

【关键词】 肺结节; 早期识别; XGBoost; SHAP; 可解释性框架

【中图分类号】 R563; R445 **【文献标志码】** A

XGBoost-SHAP-based interpretable framework for the early identification of pulmonary nodules

YI Fu-liang¹, LI Gang², LIU Xin³, XIANG Ru-mei¹, LUO Chang-ling¹, DENG Li-chun¹, YU Xiu-lian¹, ZHOU Hou-rong¹, GAO Yang¹, ZOU Xue-na¹

(1. Department of Public Health, Zigong Fourth People's Hospital, Zigong 643000; 2. Chronic Disease Institute, Ziliujing Disease Prevention and Control Center, Zigong 643036; 3. Health Management Center, Zigong First People's Hospital, Zigong 643000, Sichuan, China)

【Abstract】 Objective: To achieve early identification of pulmonary nodules and visual interpretation of key variables through interpretable machine learning, and to facilitate precise prevention, control, early diagnosis and treatment of lung cancer. **Methods:** This study enrolled individuals at high risk of lung cancer and completed clinical screening. Their high-risk assessment data and imaging results were extracted. Participants were divided into high-risk and low-risk groups for pulmonary nodules based on China's Lung Cancer Screening Standard (T/CPMA 013-2020). Variables with differences identified by univariate analysis were used as predictors, with pulmonary nodule grouping as the dependent variable, to construct an interpretable XGBoost-SHAP identification framework for early nodule detection and visual result interpretation. **Results:** A total of 644 high-risk individuals were included, with 199 (30.9%) in the high-risk pulmonary nodule group. The XGBoost model achieved an accuracy of 0.914 6, sensitivity of 0.758 7, specificity of 0.984 3, F1-score of 0.845 8, and AUC of 0.974 1 for nodule grouping. SHAP analysis revealed that higher SHAP values-and thus increased risk of nodule enlargement-were associated with greater smoking intensity, exposure to secondhand smoke from colleagues/family, infrequent kitchen ventilation during cooking, excessive intake of processed foods, occupational exposure to asbestos/radon, insufficient intake of protein, fruits and vegetables, and manual labor occupation. **Conclusion:** The constructed interpretable framework performs well in early pulmonary nodule identification. Changes in nodule size are associated not only with traditional risk factors (e.g., smoking habits, secondhand smoke exposure, cooking fume exposure, occupational asbestos/radon exposure) but also with the participants' dietary habits.

【Key words】 Pulmonary nodules; Early identification; XGBoost; SHAP; Interpretable framework

基金项目: 成都医学院教育发展基金会科研专项项目(25LHZG-12); 四川省自贡市重点科技计划项目(2024-YGY-01-04)

作者简介: 易付良(1996-), 男, 硕士, 住院医师。E-mail: 18783934308@163.com

通讯作者: 邹雪娜。E-mail: 704639005@qq.com

国家癌症中心数据显示,2022年我国肺癌发病率、死亡率均居男女性恶性肿瘤首位,新发病例106.06万(22.0%)、死亡73.33万(28.5%),高发死亡成为公共卫生重点问题^[1-2]。肺癌早期无症状难察觉,而作为肺癌早期表现的肺结节,其及时发现对提升肺癌早诊早治率至关重要^[3]。《肺结节诊治中国专家共识(2024年版)》按病灶大小将肺结节分为微小结节(≤ 5 mm)、小结节(5~10 mm),且结节大小与恶性概率相关^[4]。低剂量CT(low-dose computed tomography, LDCT)可早期发现肺部异常、改善肺癌治疗效果与生存率,在肺癌高危人群中合理筛查可降低20%肺癌死亡率^[5-6]。但传统筛查依赖医生肉眼观察CT图形与经验判断,主观因素易致假阳性率升高、过度随访。一项meta分析显示,过度随访会增加肺癌长期累积发病率($RR = 1.51$; 95% $CI: 1.06 \sim 2.14$),49%筛查检出癌症或被过度随访^[7]。人工智能(尤其机器学习)为医疗带来变革,在肿瘤学中应用广泛^[8]。相比传统回归模型,数据驱动的机器学习更易捕捉数据规律,辅助医生精准诊断。其中,XGBoost-SHAP(极限梯度提升-沙普利可解释性)联合框架的识别性能已在既往研究^[9]中验证。本研究拟依托自贡市国家城市癌症早诊早治项目肺癌筛查数据,构建XGBoost-SHAP可解释性联合框架,实现受检者肺结节大小分组精准识别,挖掘关键影响变量并可视觉解释模型结果,为肺癌精准防控与早诊早治提供科学依据与技术支撑。

1 资料与方法

1.1 一般资料

纳入2022年1月至2023年12月参加自贡市自流井区国家城市癌症早诊早治项目,被评估为肺癌高危并进一步完成免费临床筛查的人群。纳入标准:(1)年龄45~74岁;(2)户籍为自贡市自流井区或在此居住半年及以上;(3)愿意签署知情同意书并配合问卷调查与生物学检测;(4)被评估为肺癌高危者。排除标准:(1)5年内参与过本次项目;(2)既往有肿瘤病史;(3)其他严重疾病患者。本研究已通过国家项目组牵头单位,即国家癌症中心医学伦理委员会审查(15-070/997)。

1.2 研究方法

社区流调人员通过整群抽样,在自流井区舒坪街、新街、郭街、五星街、东兴寺街所辖社区招募居民,完成问卷评估(依托“中国居民癌防控制行动”公众号或小程序)与生物学检测(含幽门螺旋杆菌、便隐血、乙肝抗原检测),所有居民评估前均签署知情

同意书。评估系统以“哈佛癌症风险指数”为基础,结合我国癌症流行病学数据及专家意见开展癌症风险评估,依填写信息自动判定居民是否为癌症高危群体。高危阳性者被推荐至自贡市第四人民医院,由放射科高年资医师按《城市癌症早诊早治项目实施方案》完成免费临床筛查。随后依据《中国肺癌筛查标准(T/CPMA013-2020)》,按“实性/部分实性结节实性成分平均直径 ≥ 6.0 mm”或“非实性结节平均直径 ≥ 8.0 mm”的标准,将研究对象分为肺结节高危组与低危组^[10]。

变量筛选:涵盖受检者基本信息(性别、年龄、身高、体质量、腰围、学历、婚姻状态、职业类型);膳食习惯(蔬菜水果、加工类食物、蛋白质、碳水化合物摄入,口味偏好,食物/饮品温度);职业暴露(石棉、氡等物质接触);油烟暴露(做饭时长、做饭通风情况);吸烟习惯(是否吸烟、首次吸烟年龄、吸烟量);二手烟暴露(同居家人是否吸烟及吸烟量、是否暴露于同事吸烟环境、同事是否戒烟);疾病史(慢性疾病、结核病史、家族癌症病史)及肺结节分组。

1.3 模型构建

构建XGBoost-SHAP联合框架时,先以单因素分析有意义的变量为预测变量、结节分组为因变量,将所有预测变量输入XGBoost模型(采用5折交叉验证)并输出预测结局;对预测结局二值化后,通过准确度、敏感度、特异度、F1值及曲线下面积(AUC)等指标,对比预测结局与因变量以评价模型识别效果。随后利用SHAP输出全局解释的摘要图与依赖图,实现重要预测变量的可视化解释。

1.4 统计学分析

采用R4.3.3软件进行数据分析。对研究对象的一般资料进行描述,符合正态分布的计量资料采用 $(\bar{x} \pm s)$ 表示,组间比较采用独立样本 t 检验;计数资料采用 $[n(\%)]$ 表示,组间比较采用独立样本 χ^2 检验,等级资料组间比较采用趋势 χ^2 检验。 $P < 0.05$ 为差异有统计学意义。

2 结果

2.1 高危评估人群基本情况

研究共纳入644例受检者,其中肺结节高危组199人(30.9%)、低危组445人(69.1%)。人群特征如下:男性313人(48.60%),女性331人(51.4%);年龄 (59.82 ± 6.62) 岁,身高 (158.05 ± 7.29) cm;体质量 (59.48 ± 11.94) kg;腰围 (79.26 ± 9.75) cm;学历以小学及以下(44.25%)、初中(40.22%)为主;婚姻状态已婚者占94.25%;职业类型以体力劳动者居多(52.64%)。两组患者除了

职业差异有统计学意义 ($P < 0.05$) 外, 其余一般资料比较, 差异均无统计学意义 ($P > 0.05$)。见表 1。

表 1 两组对象一般资料比较 [$\bar{x} \pm s, n(\%)$]

资料	肺结节低危组 ($n=445$)	肺结节高危组 ($n=199$)	t/χ^2 值	P 值
性别			0.018	0.894
男	215(48.31)	98(49.25)		
女	230(51.69)	101(50.75)		
年龄(岁)	59.88±6.78	59.70±6.25	0.102	0.750
身高(cm)	158.27±7.38	157.57±7.09	1.275	0.259
体质量(kg)	59.69±12.25	59.02±11.22	0.435	0.510
腰围(cm)	79.45±9.54	78.83±10.22	0.543	0.462
学历			9.373	0.052
小学及以下	193(43.37)	92(46.23)		
初中	173(38.88)	86(43.22)		
高中/中专/技校	58(13.03)	20(10.05)		
大学专科	13(2.92)	0(0.00)		
大学本科	8(1.80)	1(0.50)		
婚姻状态			6.570	0.087
未婚	3(0.67)	0(0.00)		
已婚	414(93.03)	193(96.98)		
离异	12(2.70)	5(2.51)		
丧偶	16(3.60)	1(0.50)		
职业类型			6.664	0.036
脑力劳动	55(12.36)	6(18.09)		
体力劳动	230(51.69)	109(54.77)		
其他	160(35.96)	54(27.14)		

脑力劳动: 含国家机关、党组织、企事业单位工作人员及科研、工程技术、农业技术、飞机船舶技术、卫生专业技术、经济、金融、法律、教学、文学艺术、新闻出版、文化等领域工作人员。体力劳动: 含农、林、牧、渔、水利生产人员, 生产与运输设备操作人员及相关人员, 军人等。其他: 除上述两类外的其他职业者。

2.2 影响肺结节分组的单因素分析

膳食习惯: 受检者蔬菜水果、加工类食物、蛋白质、碳水化合物摄入及口味偏好, 在肺结节高低危组间差异有统计学意义 ($P < 0.05$); 食用食物或饮品温度两组患者无统计学差异 ($P > 0.05$)。职业与环境暴露: 职业暴露肺结节高危组石棉、氡等物质暴露率 68.84%, 高于低危组 ($P < 0.05$)。油烟暴露: 高危组做饭从不通风 (25.13%)、偶尔通风 (39.70%) 占比均高于低危组 ($P < 0.05$), 但做饭时长两组患者无统计学差异 ($P > 0.05$)。吸烟习惯: 高危组吸烟率 60.30%、吸烟量 (19.95±11.91) 支/d, 均高于低危组 ($P < 0.05$), 首次吸烟年龄两组无统计学差异 ($P > 0.05$)。二手烟暴露: 高危组同居家人吸烟率 74.87%、家人吸烟量 (20.36±13.98) 支/d、暴露于同事吸烟环境占比 88.44%, 均高于低危组 ($P < 0.05$), 同事是否戒烟两组无统计学差异 ($P > 0.05$)。高危组慢性疾病患病率 41.21%、家族癌症病史占比 73.37%, 均高于低危组 ($P < 0.05$); 结核病病史两组

患者无统计学差异 ($P > 0.05$)。见表 2。

表 2 两组对象评估条目比较 [$\bar{x} \pm s, n(\%)$]

条目	低危组($n=445$)	高危组($n=199$)	t/χ^2 值	P 值
蔬菜水果摄入			21.317	<0.001
每天	281(63.15)	155(77.89)		
每周 4~6 d	86(19.33)	34(17.09)		
每周 2~3 d	55(12.36)	9(4.52)		
每周 1 d 及以下	23(5.17)	1(0.50)		
加工类食物摄入			28.849	<0.001
每天	70(15.73)	32(16.08)		
每周 4~6 d	86(19.33)	26(13.07)		
每周 2~3 d	159(35.73)	112(56.28)		
每周 1 d 及以下	130(29.21)	29(14.57)		
蛋白质摄入			28.141	<0.001
每天	193(43.37)	98(49.25)		
每周 4~6 d	145(32.58)	87(43.72)		
每周 2~3 d	82(18.43)	8(4.02)		
每周 1 d 及以下	25(5.62)	6(3.02)		
碳水化合物摄入			8.333	0.040
每天	402(90.34)	191(95.98)		
每周 4~6 d	22(4.94)	7(3.52)		
每周 2~3 d	15(3.37)	1(0.50)		
每周 1 d 及以下	6(1.35)	0(0.00)		
口味偏好			20.830	<0.001
偏咸	205(46.07)	61(30.65)		
适中	200(44.94)	128(64.32)		
偏淡	40(8.99)	10(5.03)		
食用食物或饮品温度			5.547	0.136
低于常温	8(1.80)	3(1.51)		
常温	361(81.12)	167(83.92)		
烫	42(9.44)	23(11.56)		
滚烫	34(7.64)	6(3.02)		
石棉等物质暴露史	204(45.84)	137(68.84)	28.286	<0.001
做饭时长(年)	29.07±15.86	30.99±13.31	2.213	0.137
做饭通风情况			150.581	<0.001
从不	94(21.12)	50(25.13)		
偶尔	26(5.84)	79(39.70)		
经常	92(20.67)	44(22.11)		
总是	233(52.36)	26(13.07)		
吸烟	20(4.49)	120(60.30)	266.125	<0.001
已戒烟	90(20.22)	37(18.59)		
首次吸烟年龄(岁)	21.19±5.90	21.78±4.10	0.938	0.334
吸烟量(支/d)	4.24±8.51	19.95±11.91	361.907	<0.001
同居家人吸烟	169(37.98)	149(74.87)	73.423	<0.001
同居家人吸烟量(支/d)	7.46±10.44	20.36±13.98	168.630	<0.001
暴露在同事吸烟环境中	101(22.70)	176(88.44)	239.811	<0.001
同事目前没有戒烟	46(45.54)	88(50.00)	0.347	0.556
慢性疾病患病	134(30.11)	82(41.21)	7.102	0.008
结核病史	9(2.02)	1(0.50)	1.203	0.273
家族癌症病史	284(63.82)	146(73.37)	5.226	0.022

2.3 XGBoost 模型识别效果评价

XGBoost 模型采用 5 折交叉验证对受检者结

节分组进行识别。通过表 3 可以发现,模型的平均准确度为 0.914 6,敏感度为 0.758 7,特异度为 0.984 3,F1 值为 0.845 8,模型 AUC 为 0.974 1,各项指标均在 0.75 以上,模型识别效果表现良好。

表 3 XGBoost 模型识别效果评价

指标	第 1 折	第 2 折	第 3 折	第 4 折	第 5 折	均值
准确度	0.914 7	0.914 7	0.930 2	0.907 0	0.906 2	0.914 6
敏感度	0.725 0	0.800 0	0.775 0	0.750 0	0.743 6	0.758 7
特异度	1.000 0	0.966 3	1.000 0	0.977 5	0.977 5	0.984 3
F1 值	0.840 6	0.853 3	0.873 2	0.833 3	0.828 6	0.845 8
AUC 值	0.966 9	0.973 9	0.994 2	0.977 8	0.957 6	0.974 1

2.4 特征重要性可视化解释

XGBoost 联合 SHAP 输出摘要图(图 1)与依赖图(图 2),以解析变量对肺结节高危风险的影响。

图 1 显示,影响模型识别的前 10 个重要变量依次为:受检者每天吸烟量、暴露于同事吸烟环境、同居家人每天吸烟量、受检者是否吸烟、做饭通风情况、加工类食物摄入、石棉和氡等职业暴露、蛋白质摄入、蔬菜水果摄入、职业类型。以吸烟量为例,吸烟量越大(颜色从黄到紫),SHAP 值越高,肺结节高危可能性越大,其他变量作用方向解释类似。

图 2 进一步印证变量影响趋势:如吸烟 30 支/d 者的 SHAP 值高于 10 支/d 者,表明 SHAP 值随吸烟量增加而升高,肺结节高危风险同步上升,其他变量规律类似。

综上,受检者吸烟量更大、暴露于同事/家人吸烟环境、做饭通风频次低、加工类食物摄入多、有石棉和氡职业暴露、蛋白质及蔬菜水果摄入少、从事体力劳动,其 SHAP 值更高,肺结节高危风险也更高。

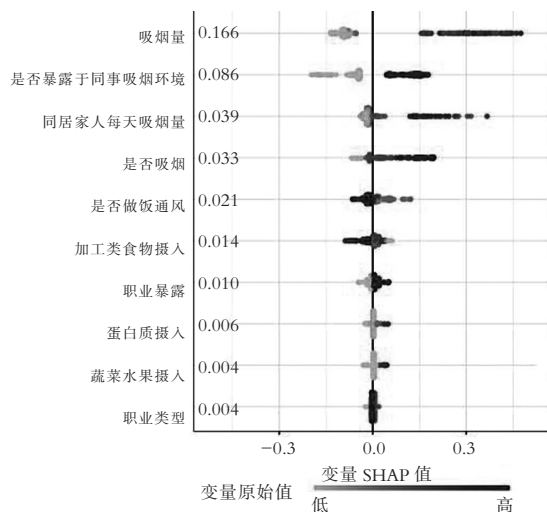


图 1 特征重要性的 SHAP 摘要图

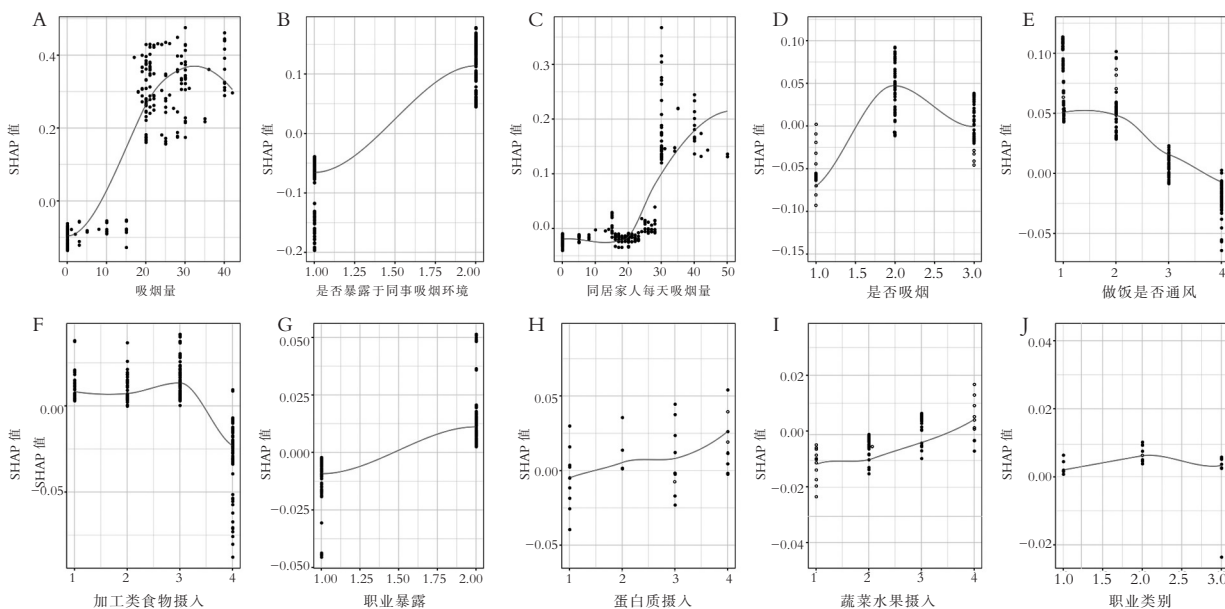


图 2 特征重要性的 SHAP 依赖图

3 讨论

本研究依托自贡市国家城市癌症早诊早治项目,整合肺癌受检者高危评估与筛查数据,构建 XGBoost-SHAP 可解释性联合框架,旨在实现该人群肺结节早期识别,并挖掘关键影响特征。

研究发现,XGBoost-SHAP 联合框架识别肺结节分组的各项评价指标均 >0.75 。传统筛查依赖医生对 CT 图形的肉眼观察与经验判断,主观因素易致假阳性率升高。数据显示,超 90% 筛查检出的肺结节最终证实为良性,既浪费医疗资源,又加重患者心理负担^[11-13];对不确定结节的误判还可能引发

额外影像学检查(增加辐射暴露)、支气管镜检查等侵入性操作,甚至导致肺癌过度诊断。相比之下,联合框架以客观量化方式评估肺结节,能更快速精准完成识别,为医生诊断决策提供有效辅助,提升诊断准确性与效率。

本研究发现,个人的吸烟习惯(是否吸烟、吸烟量),二手烟暴露状况(接触情况及暴露量),居民在家烹饪时的通风条件,及石棉和氡等职业暴露是影响受检者出现肺结节高危的重要指标。吸烟习惯对肺结节影响显著:长期吸烟会使烟草中的多环芳烃、重金属等诱导体内氧化应激与炎症反应,慢性炎症可能导致肺泡上皮细胞转化,进而促进肺结节形成与增长^[14]。本研究招募的中老年群体中,吸烟者烟草暴露时间更长,肺结节高危风险更高;且吸烟量与肺结节大小存在剂量-反应关系,与 Zeng 等^[15]研究一致。二手烟暴露危害不容忽视:近 70% 中国成年人暴露于二手烟,每年近 6 万肺癌死亡与之相关^[16]。无论家庭还是工作环境,长期接触二手烟会增加肺结节形成机会,与 WHO 报告吻合^[17];二手烟有害物质还会影响呼吸道上皮细胞渗透性及病毒受体数量与有效性,增加呼吸道病毒感染风险,进一步促进肺结节形成^[18]。家庭烹饪通风条件至关重要:通风不足(尤其使用固体燃料的厨房)会增加肺癌风险^[16],厨房油烟有害物质可直接刺激上呼吸道、加剧症状,影响肺结节发展,凸显改善烹饪环境的重要性^[19]。职业暴露风险需警惕:石棉纤维会逐步破坏肺部微观结构、引发炎症级联反应,在持续病理刺激下,肺结节更易形成且生长速度加快^[20];氡释放的放射性物质可直接作用于肺部细胞基因组,导致基因突变与染色体异常,大幅增加肺结节恶变风险^[21]。上述研究均印证了肺癌早期防控方向的正确性。

研究进一步发现,居民饮食习惯(蛋白质摄入、蔬菜水果类食物摄入及加工类食物摄入)对肺结节改变存在调节作用,为肺结节形成机制与肺癌早期防控提供了新方向。从蛋白质摄入来看,正如 Morris 等^[22]研究所述,蛋白质是构成细胞酶、受体及结构蛋白的关键成分。若机体蛋白质摄入不足,会直接导致细胞新陈代谢速率减缓,细胞修复与再生机制受阻^[23]。以肺部为例,肺泡上皮细胞的正常更新依赖充足蛋白质供应,一旦蛋白质匮乏,可能造成肺泡壁变薄、弹性减弱,使肺部局部微环境更易受到外界刺激与损伤,进而为肺结节形成创造有利条件。在蔬菜水果摄入方面,其与肺结节改变的关联虽尚未完全明确,但已有研究提示重要作用。蔬菜水果含有的维生素 C、E 及类胡萝卜素等抗氧化剂,

能有效中和体内自由基,减少氧化应激对肺部细胞的损害^[24-25]。尤其针对本研究中的中老年群体,随着年龄增长,机体自身抗氧化防御系统功能逐渐衰退,若蔬菜水果摄入不足,肺部组织将长期处于高水平自由基环境中,可能引发细胞 DNA 损伤、炎症反应加剧等问题,进而推动肺结节生长^[26]。此外,蔬菜水果中的膳食纤维还可通过“肠-肺轴”双向调节机制,间接影响肺部免疫状态与炎症水平,对肺结节发展起到潜在的抑制或促进作用^[27]。关于加工类食物摄入,其对肺结节的影响虽缺乏系统验证,但从成分特性推测关联。加工类食物通常含有大量添加剂、高盐、高脂肪及高糖等成分。如 Hunter 等^[28]研究指出,高盐饮食会引发体内钠水潴留,不仅加重心血管系统负担、影响肺部血液循环灌注,还会激活体内多种炎症信号通路,导致肺部慢性炎症状态持续存在;而高脂肪、高糖饮食则易诱发代谢综合征,造成胰岛素抵抗、血脂异常等代谢紊乱^[29]。这些代谢异常与慢性炎症会共同促使机体处于慢性低度炎症环境,刺激肺部细胞异常增殖与分化,最终导致肺结节体积增大。

综上,居民饮食习惯通过蛋白质供应、抗氧化保护、炎症调节等多重途径,对肺结节改变产生影响,这一发现为肺结节及肺癌早期防控提供了饮食干预的新思路,也为后续相关研究与实践提供了参考。

参考文献

- [1] 郑荣寿,陈茹,韩冰峰,等. 2022 年中国恶性肿瘤流行情况分析[J]. 中华肿瘤杂志, 2024, 46(3): 221-231.
- [2] Qi J, Li M, Wang L, *et al.* National and subnational trends in cancer burden in China, 2005-20: an analysis of national mortality surveillance data[J]. *The Lancet Public Health*, 2023, 8(12): e943-e955.
- [3] Chen D, Yang L, Zhang W, *et al.* Prevalence and management of pulmonary nodules: a systematic review and meta-analysis[J]. *Journal of Thoracic Disease*, 2024, 16(7): 4619-4632.
- [4] 中华医学会呼吸病学分会, 中国肺癌防治联盟专家组. 肺结节诊治中国专家共识(2024 年版)[J]. *中华结核和呼吸杂志*, 2024, 47(8): 716-729.
- [5] Caverly TJ, Zhang X, Hayward RA, *et al.* Effects of random measurement error on lung cancer screening decisions: a retrospective cohort-based microsimulation study[J]. *Chest*, 2021, 159(2): 853-861.
- [6] Becker N, Motsch E, Trotter A, *et al.* Lung cancer mortality reduction by LDCT screening-Results from the randomized German LUSI trial[J]. *International Journal of Cancer*, 2020, 146(6): 1503-1513.
- [7] Brodersen J, Voss T, Martiny F, *et al.* Overdiagnosis of lung cancer with low-dose computed tomography screening: meta-analysis of the randomised clinical trials[J]. *Breathe*, 2020, 16(1): 200013.

- [8] Wu QY, Huang Y, Wang SC, *et al.* Artificial intelligence in lung cancer screening: Detection, classification, prediction, and prognosis[J]. *Cancer Medicine*, 2024, 13(7): e7140.
- [9] Yi F, Yang H, Chen D, *et al.* XGBoost-SHAP-based interpretable diagnostic framework for Alzheimer's disease[J]. *BMC Medical Informatics and Decision Making*, 2023, 23(1): 137.
- [10] 中华预防医学会. 中国肺癌筛查标准(T/CPMA 013-2020)[J]. *中国慢性病预防与控制*, 2021, 29(1): 1-8.
- [11] He J, Wang B, Tao J, *et al.* Accurate classification of pulmonary nodules by a combined model of clinical, imaging, and cell-free DNA methylation biomarkers: a model development and external validation study[J]. *The Lancet Digital Health*, 2023, 5(10): e647-e656.
- [12] Gupta S, Jacobson FL, Kong CY, *et al.* Performance of lung nodule management algorithms for lung-RADS category 4 lesions[J]. *Academic Radiology*, 2021, 28(8): 1037-1042.
- [13] Duan XR, Ouyang ZQ, Bao SS, *et al.* Factors associated with overdiagnosis of benign pulmonary nodules as malignancy: a retrospective cohort study[J]. *BMC Pulmonary Medicine*, 2023, 23(1): 454.
- [14] Caliri AW, Tommasi S, Besaratinia A. Relationships among smoking, oxidative stress, inflammation, macromolecular damage, and cancer[J]. *Mutation Research Reviews in Mutation Research*, 2021, 787: 108365.
- [15] Zeng H, Ran X, An L, *et al.* Disparities in stage at diagnosis for five common cancers in China: a multicentre, hospital-based, observational study[J]. *The Lancet Public Health*, 2021, 6(12): e877-e887.
- [16] He S, Li H, Cao M, *et al.* Trends and risk factors of lung cancer in China[J]. *Chinese Journal of Cancer Research*, 2020, 32(6): 683-694.
- [17] Asomaning K, Miller DP, Liu G, *et al.* Second hand smoke, age of exposure and lung cancer risk[J]. *Lung Cancer*, 2008, 61(1): 13-20.
- [18] Flor LS, Anderson JA, Ahmad N, *et al.* Health effects associated with exposure to secondhand smoke: a Burden of Proof study[J]. *Nature Medicine*, 2024, 30(1): 149-167.
- [19] Fang Y, Li Z, Chen H, *et al.* Burden of lung cancer along with attributable risk factors in China from 1990 to 2019, and projections until 2030[J]. *Journal of Cancer Research and Clinical Oncology*, 2023, 149(7): 3209-3218.
- [20] Zandwijk NV, Frank AL, Reid G, *et al.* Asbestos-Related lung Cancer: an underappreciated oncological issue [J]. *Lung Cancer*, 2024, 194: 107861.
- [21] Liu Y, Xu Y, Xu W, *et al.* Radon and lung cancer: Current status and future prospects[J]. *Critical Reviews in Oncology/Hematology*, 2024, 198: 104363.
- [22] Morris R, Black KA, Stollar EJ. Uncovering protein function: from classification to complexes[J]. *Essays in Biochemistry* 2022, 66(3): 255-285.
- [23] Tomé D, Benoit S, Azzout-Marniche D. Protein metabolism and related body function: mechanistic approaches and health consequences[J]. *The Proceedings of the Nutrition Society*, 2021, 80(2): 243-251.
- [24] Bufka J, Vaňková L, Sykora J, *et al.* Exploring carotenoids: Metabolism, antioxidants, and impacts on human health[J]. *Journal of Functional Foods*, 2024, 118: 106284.
- [25] Sun Y, Li Z, Li J, *et al.* A healthy dietary pattern reduces lung cancer risk: a systematic review and meta-analysis[J]. *Nutrients*, 2016, 8(3): 134.
- [26] Wang M, Qin S, Zhang T, *et al.* The effect of fruit and vegetable intake on the development of lung cancer: a meta-analysis of 32 publications and 20,414 cases[J]. *European Journal of Clinical Nutrition*, 2015, 69(11): 1184-1192.
- [27] Zhang L, Zhang Z, Xu L, *et al.* Maintaining the balance of intestinal flora through the diet: effective prevention of illness [J]. *Foods*, 2021, 10(10): 2312.
- [28] Hunter RW, Dhaun N, Bailey MA. The impact of excessive salt intake on human health [J]. *Nature Reviews Nephrology*, 2022, 18(5): 321-335.
- [29] Tang C, Wang Y, Xu Z, *et al.* The relationships between high-fat diet and metabolic syndrome: Potential mechanisms [J]. *Food Bioscience*, 2024, 59: 104261.

(收稿日期: 2025-02-07

修回日期: 2025-10-06)