

基于多模态机器学习模型的东南沿海地区翼状胬肉相关危险因素

叶霞¹, 周喆媛², 王施鲲¹, 赵娜¹, 刘震¹

引用: 叶霞, 周喆媛, 王施鲲, 等. 基于多模态机器学习模型的东南沿海地区翼状胬肉相关危险因素. 国际眼科杂志, 2026, 26(9): 1638-1643.

基金项目: 舟山市医药卫生科技计划军地共建项目 (No.2022JYB02)

作者单位: ¹(316000) 中国浙江省舟山市, 东部战区海军医院眼科; ²(22202) 美国弗吉尼亚州阿灵顿市, 亚马逊科技公司

作者简介: 叶霞, 女, 本科, 主任医师, 研究方向: 白内障及眼表疾病。

通讯作者: 叶霞. yex413@126.com

收稿日期: 2026-03-20 修回日期: 2026-07-24

摘要

目的: 调查东南沿海地区翼状胬肉的患病率, 分析相关危险因素并建立多模态机器学习预测模型。

方法: 横断面流行病学调查。对舟山群岛居民群体随机抽样, 通过现场眼部检查及流行病学调查, 用单因素及多因素分析翼状胬肉相关的危险因素, 并采用人工智能机器学习多模态模型对结构化与非结构化数据进行综合分析建立预测模型。

结果: 本研究共调查 1 842 人, 其中男 926 人 (50.27%), 女 916 人 (49.73%); 平均年龄 48.62±10.35 岁。患翼状胬肉共 644 例 (34.96%), 男 323 例 (50.15%), 女 321 例 (49.84%); 其中单眼翼状胬肉者 502 例 (77.95%), 双眼翼状胬肉者 142 例 (22.05%)。无翼状胬肉者 1 198 例 (65.04%), 其中男 603 例 (50.33%), 女 595 例 (49.67%)。职业、户外日照时长、睑板腺开口堵塞、年龄、吸烟、饮酒是引起翼状胬肉的危险因素, 其中职业 ($OR = 6.125, P < 0.001$) 与户外日照时长 ($OR = 5.348, P < 0.001$) 是最危险因素; 其次是年龄 (>40 岁) ($OR = 5.295, P < 0.001$) 及睑板腺开口堵塞 ($OR = 5.248, P < 0.001$)。多模态机器学习模型预测效能良好, ROC 曲线、PR 曲线、混淆矩阵均具备较高预测价值, 其中 $AUC = 0.86, AP = 0.77$ 。

结论: 40 岁以上, 日照时长 >4 h 的渔民、农民是东南沿海地区患翼状胬肉的高发人群, 睑板腺功能障碍与翼状胬肉的角膜侵入相关, 可能促进其发展, 应积极宣教与干预。多模态机器学习模型对翼状胬肉患病风险具有精准可靠的预测效能, 有助于东南沿海人群翼状胬肉的风险分层与早期预警。

关键词: 东南沿海; 翼状胬肉; 流行病学; 危险因素; 机器学习

DOI: 10.3980/j.issn.1672-5123.2026.9.22

Multimodal machine learning model – based study on pterygium – related risk factors in southeast coastal areas

Ye Xia¹, Zhou Zheyuan², Wang Shikun¹, Zhao Na¹, Liu Zhen¹

Foundation item: Military – Civilian Joint Project of the Zhoushan Medical and Health Science and Technology Plan (No.2022JYB02)

¹Department of Ophthalmology, Eastern Theater Navy Hospital, Zhoushan 316000, Zhejiang Province, China; ²Amazon Web Services, Arlington, VA 22202, USA

Correspondence to: Ye Xia. Department of Ophthalmology, Eastern Theater Navy Hospital, Zhoushan 316000, Zhejiang Province, China. yex413@126.com

Received: 2026-03-20 Accepted: 2026-07-24

Abstract

• **AIM:** To investigate the prevalence of pterygium in the southeast coastal area, analyze associated risk factors, and develop a multimodal machine learning predictive model.

• **METHODS:** A cross – sectional epidemiological study. A random sampling of the resident population of the Zhoushan Islands was performed. Data were collected through on – site ophthalmic examinations and epidemiological surveys. Univariate and multivariate analyses were performed to identify risk factors associated with pterygium. A multimodal artificial intelligence machine learning model was constructed to integrate and analyze both structured and unstructured data for predictive modeling.

• **RESULTS:** A total of 1 842 participants were enrolled in this study, including 926 males (50.27%) and 916 females (49.73%), with a mean age of 48.62±10.35 y. A total of 644 patients (34.96%), including 323 males (50.15%) and 321 females (49.84%) were diagnosed with pterygium, including 502 cases (77.95%) were unilateral, and 142 cases (22.05%) were bilateral. There were 1 198 cases without pterygium, including 603 males (50.33%) and 595 females (49.67%). Occupation, daily outdoor sunlight exposure duration, meibomian gland orifice obstruction, age, smoking, and alcohol consumption were identified as significant risk factors for pterygium. Occupation and daily outdoor sunlight exposure duration were the most potent risk factors ($OR = 6.125, P < 0.001$; $OR = 5.348, P <$

0.001, respectively), followed by age >40 y ($OR=5.295$, $P<0.001$) and meibomian gland orifice obstruction ($OR=5.248$, $P<0.001$). The multimodal machine learning model demonstrated good predictive performance, as evidenced by the receiver operating characteristic (ROC) curve, precision-recall (PR) curve, and confusion matrix, all indicating high predictive value, $AUC=0.86$, $AP=0.77$.

• **CONCLUSION:** Fishermen and farmers aged over 40 y with more than 4 h of daily sunlight exposure are at high risk for pterygium in the southeast coastal area. Meibomian gland dysfunction is correlated with corneal invasion of pterygium and may facilitate its progression; therefore, active health education and targeted intervention are recommended. The multimodal machine learning model provides accurate and reliable predictive capability for pterygium risk, which may facilitate risk stratification and early warning for pterygium in southeast coastal populations.

• **KEYWORDS:** southeast coastal area; pterygium; epidemiology; risk factors; machine learning

Citation: Ye X, Zhou ZY, Wang SK, et al. Multimodal machine learning model-based study on pterygium-related risk factors in southeast coastal areas. *Guoji Yanke Zazhi (Int Eye Sci)*, 2026, 26(9):1638-1643.

0 引言

翼状胬肉是眼科常见的眼表疾病,除影响眼球美观外,侵犯角膜后可引起眼红、眼干、角膜散光,严重者甚至遮盖瞳孔,损害视力^[1]。其发病机制尚未明确,目前认为可能与长时间的紫外线暴露以及结膜慢性炎症有关^[2]。既往研究显示低纬度地区翼状胬肉患病率明显高于高纬度地区,尤其在长时间户外工作人员中显著增加^[3],日光中紫外线 (ultravioletray, UVR) 的反射率因地面环境不同而有所差异,草地对 UVR 的反射率为 5%-15%,沙地为 30%,雪地为 70%,而水面和冰面高达 85%^[4]。一项研究显示,印度沿海、平原和丘陵地区的翼状胬肉患病率分别为 28.3%、11.2% 和 9.1%,而翼状胬肉患病率与居住于沿海地区和一生中累积日照时间呈正相关^[4]。东南沿海地区纬度为 20-32 度,环海,具有特殊的地理环境,如海洋气候,紫外线强度等,目前少有对东南沿海地区翼状胬肉患病率及其影响因素的研究。因此,我们通过对东南沿海舟山群岛居民翼状胬肉患病率进行研究,并采用人工智能机器学习多模态模型,对义诊场景下采集的原始自由文本进行语义编码,融合结构化与非结构化数据建立预测模型,明确影响翼状胬肉患病的危险因素,为该病的预防和治疗提供参考依据。本研究是舟山市卫生健康委员会科研项目中舟山地区人群翼状胬肉流行病学调查的一部分,由东部战区海军医院联合舟山市红十字会合作实施。

1 对象和方法

1.1 对象 横断面流行病学调查。于 2023 年 1 月至 2024 年 12 月依托舟山市红十字会组织的“下海岛医疗队”,采取多阶段随机抽样方式,在舟山地区定海、普陀、岱山、嵊泗、嵊山、枸杞、六横七个小岛义诊期间,随机抽取前来义诊的居民作为研究对象。调查对象纳入标准:(1) 在本岛

长期居住 (≥ 5 a) 的原居民;(2) 资料完整。排除标准:(1) 患有感染性眼病如急性结膜炎、角膜炎;(2) 严重睑内翻倒睫;(3) 长期滴用滴眼液或配戴隐形眼镜者;(4) 有慢性变应性结膜炎病史;(5) 患有全身风湿免疫疾病以及结缔组织病者;(6) 睑裂斑患者。翼状胬肉诊断标准:(1) 依据《眼科学》第三版^[5] 标准,单眼或双眼有明确侵入角膜缘内的球结膜增生组织者;(2) 曾有翼状胬肉病史并已行翼状胬肉切除术者。本研究严格按照随机抽样原则,受检者均充分了解研究内容及可能后果后自愿参与并签署知情同意书。本研究符合《赫尔辛基宣言》原则,并获得东部战区海军医院伦理委员会的批准。

1.2 方法

1.2.1 问卷调查 问卷调查包括可能影响翼状胬肉相关因素调查问卷。翼状胬肉相关因素调查问卷包括:性别、年龄、吸烟史和饮酒史(详细记录每日剂量及年限,如“1 包/日,40 a”,如“1 两/天,30 a”),职业、文化程度、每日户外日照时间(全天累计时长)、每日阅读及看视频终端时长、眼部疾病及眼部手术史。

1.2.2 眼部检查 眼部检查采用手持式裂隙灯检查。检查内容包括有无翼状胬肉、睑板腺开口堵塞情况 ($>1/3$ 睑板腺开口堵塞者纳入睑板腺开口堵塞)。

1.2.3 质量控制 所有检查均由同一专业的眼科医生和护士配合完成,数据记录采用双人核对。问卷调查通过面对面询问方式由医护人员录入。

1.2.4 多模态机器学习模型建立 在统计学分析的基础上,为进一步探索翼状胬肉患病风险的预测方法,本研究针对舟山群岛居民翼状胬肉的患病情况进行预测分析,提出了一种基于自然语言处理 (natural language processing, NLP)^[6] 的逻辑回归模型构建方法,通过构建多模态机器学习模型 (此处“多模态”指数据表示形式的异质性:自由文本数据需经语义编码转化为特征向量,而数值数据可直接作为特征输入,二者属于不同的数据模态,需通过不同的特征提取管道后进行特征级融合),融合两类数据模态预测舟山群岛居民翼状胬肉的患病风险:(1) 非结构化文本模态,包括职业描述、吸烟史 (如“1 包/日,40 a”)、饮酒史 (如“1 两/天,50 a”)、户外日照时长及文化程度等问卷中以自由文本形式记录的字段;(2) 结构化数值模态,包括年龄、性别等可直接量化的指标。由于上述自由文本字段表述方式多样、难以直接量化,本研究应用深度学习模型 Sentence Transformer^[7] 对其进行语义编码,将异质性文本转化为统一的固定维度特征向量。随后,采用数据特征表示矩阵的形式进行逻辑回归模型^[8] 的构建与应用。将特征向量数据转换为数据特征表示矩阵,并以该矩阵作为模型训练、评估及测试的输入数据。在训练阶段,基于现有特征表示矩阵的数据,构建了类别平衡补正 (class balance correction)^[9] 的回归模型,在此基础上,构建带有类别平衡校正的逻辑回归分类器,经 500 轮迭代训练后,在测试集上评估模型效能,预测舟山群岛居民翼状胬肉风险的融合大语言模型文本编码与数值特征的逻辑回归框架见图 1。

1.2.5 最优多模态机器学习模型及选择方法

1.2.5.1 最优模型及其维度选择方法 本研究以 F1 Score、AUC^[10]、Precision^[11] 及 Recall^[12] 四项常用评价指标对模型进行性能衡量。各指标的数值越高,通常表明模型的预测能力越强。

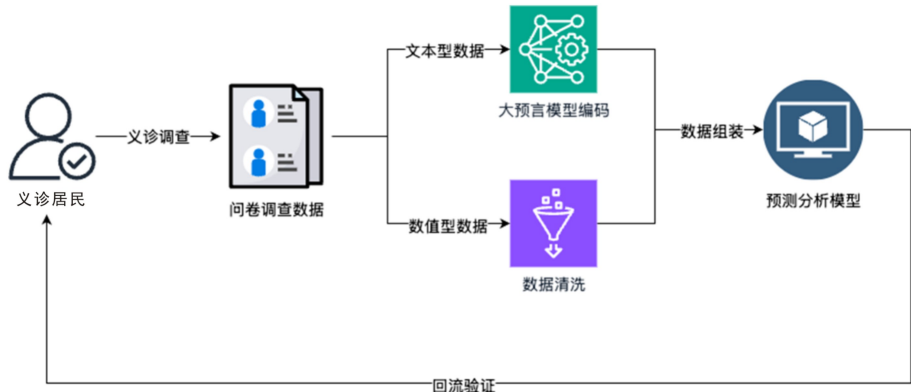


图1 预测舟山群岛义诊居民翼状胬肉风险的融合大语言模型文本编码与数值特征的逻辑回归框架。

1.2.5.2 特征降维的动机 为解决高维特征在小样本下易过拟合的问题,对特征向量进行主成分分析降维。通过对编码向量使用主成分分析(PCA)^[13]进行降维,分别保留以下维度的特征: $d \in \{2, 4, 8, 16, 32, 64, 128, 256, 384\}$,最后确定最优维度。结果显示,维度为8的模型在所有指标上均出现“拐点”(elbow point),即在保持较高判别能力的同时显著抑制了过拟合。该维度选择策略基于肘部法则(elbow method),是机器学习中选择最优超参数的标准方法,以下游任务性能而非方差贡献率作为维度选择依据,更适用于以预测为目标的建模任务,见图2。基于上述定量结果,我们选取8维特征向量作为最终的输入特征,并在全体训练数据上重新训练回归模型。用于后续解释性分析与风险因素挖掘。

统计学分析:采用 SPSS27.0 进行统计学分析。计数资料以例(%)描述,单因素组间比较采用 χ^2 检验;将单因素分析中 $P<0.05$ 的变量纳入多因素 Logistic 回归模型,筛选翼状胬肉患病的危险因素。以 $P<0.05$ 为差异有统计学意义。后与多模态机器学习模型比较,为多模态模型基线。

2 结果

2.1 研究对象一般资料 本次研究共录入样本 1 842 人,其中男 926 人(50.27%),女 916 人(49.73%);平均年龄 48.62 ± 10.35 岁,年龄 ≤ 40 岁 386 人(20.96%), >40 岁 1 456 人(79.04%);渔民职业 708 人(38.44%),农民 513 人(27.85%),非渔民、农民 621 人(33.71%);户外日照时长 >4 h/d 者 1 194 人(64.82%), ≤ 4 h/d 648 人(35.18%);有抽烟史 645 人(35.02%),无抽烟史 1 197 人(64.98%);有饮酒史 682 人(37.02%),无饮酒史 1 160 人(62.98%);受教育程度方面,文盲 267 人(14.50%),小学 901 人(48.91%),中学 522 人(28.34%),大学 152 人(8.25%);看视频终端时长 <2 h/d 者 866 人(47.01%), $2\sim 4$ h/d 者 293 人(15.91%), ≥ 4 h/d 者 683 人(37.08%);睑板腺开口堵塞者 756 例(41.04%),睑板腺开口未堵塞者 1 086 例(58.96%)。

2.2 样本人群翼状胬肉患病率及复发率 样本人群 1 842 人中患翼状胬肉者共 644 例,总患病率 34.96%。644 例翼状胬肉患者中单眼患病 502 例(77.95%),双眼患病 142 例(22.05%),曾行翼状胬肉手术者 135 例,占翼状胬肉患者总数的 20.96%。术后复发者 24 例,复发率 17.78%。

2.3 影响翼状胬肉患病的危险因素分析

2.3.1 翼状胬肉患病单因素分析 对所有可能引起翼状胬肉的影响因素,包括年龄、性别、职业、文化程度、吸烟史、

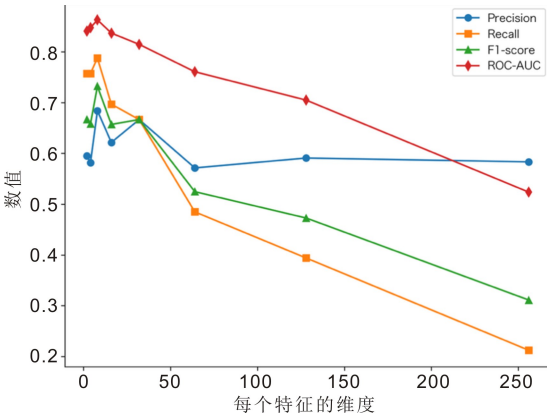


图2 PCA 特征维度选择对模型表现的影响。

饮酒史、户外日照时长、看视频终端时长、睑板腺开口堵塞情况,进行单因素分析,结果显示,无翼状胬肉者与有翼状胬肉患者的年龄、职业、户外日照时长、睑板腺开口堵塞差异、吸烟、饮酒史差异具有统计学意义(均 $P<0.001$),见表1。

2.3.2 翼状胬肉患病多因素分析 Logistic 回归分析显示,职业、户外日照时长 >4 h、年龄(>40 岁)、睑板腺开口堵塞、吸烟史、饮酒史是影响翼状胬肉患病率的危险因素(均 $P<0.05$)。其中职业是翼状胬肉最危险因素($OR=6.125$);其次是户外日照时长 >4 h($OR=5.348$)、年龄 >40 岁($OR=5.295$)、睑板腺开口堵塞($OR=5.248$)、性别中女性患病风险低于男性($OR=0.520, P<0.001$),见表2。

2.4 预测模型的评估 对在特征向量维度为8的逻辑回归模型进行系统评估,所有指标均在独立的测试集($n=97$)上计算得到。ROC 曲线(图3A)展示了模型在不同判定阈值下真正例率与假正例率的权衡。曲线下面积(AUC)是衡量模型整体判别能力的标量指标,数值越接近1表示模型区分正负样本的能力越强。本研究模型的AUC达到0.86,说明在所有可能阈值中,模型能够以86%的概率正确排序正负样本,具备显著的预测价值。PR (Precision-Recall) 曲线(图3B)直接刻画了精确率与召回率随阈值变化的关系,尤其适用于正负样本比例失衡的情形。PR 曲线下的面积(average precision, AP)^[13]概括了在不同召回水平下模型保持的精确率水平。本研究模型的AP为0.77,表明在较宽的召回范围内,模型能够维持约77%的精确率,进一步验证了模型在不平衡数据集上的稳健表现。混淆矩阵^[14]分析显示,模型整体准确率为

表 1 影响翼状胬肉患病的单因素分析结果

n(%)

指标		n	无翼状胬肉组(n=1198)	有翼状胬肉组(n=644)	χ^2	P
年龄(岁)	<40岁	386	362(93.78)	24(6.22)	33.82	<0.001
	≥40岁	1 456	836(57.42)	620(42.58)		
户外日照时间(h/d)	>4	1 194	669(56.03)	525(43.97)	95.583	<0.001
	≤4	648	529(81.64)	119(18.36)		
性别	男	926	603(65.12)	323(34.88)	3.798	0.056
	女	916	595(64.96)	321(35.04)		
吸烟史	无	1 197	840(70.18)	357(29.82)	29.68	<0.001
	有	645	358(55.50)	287(44.50)		
饮酒史	无	1 160	826(71.21)	334(28.79)	22.56	<0.001
	有	682	372(54.55)	310(45.45)		
职业	渔民、农民	1 221	698(57.17)	523(42.83)	98.86	<0.001
	非渔民、农民	621	500(80.52)	121(19.48)		
文化程度	文盲	267	170(63.67)	97(36.33)	6.64	0.084
	小学	901	582(64.59)	319(35.41)		
	中学	522	334(63.98)	188(36.02)		
	大学	152	112(73.68)	40(26.32)		
看视频终端时长(h/d)	≤2	866	576(66.51)	290(33.49)	1.603	0.449
	3-<4	293	188(64.16)	105(35.84)		
	≥4	683	434(63.54)	249(36.46)		
睑板腺开口堵塞	是	756	394(52.12)	362(47.88)	89.70	<0.001
	否	1 086	804(74.03)	282(25.97)		

表 2 翼状胬肉影响因素的多因素 Logistic 回归分析

影响因素	B	SE	Wald χ^2	P	OR	95%CI
年龄(>40岁)	1.983	0.412	23.157	<0.001	5.295	2.098,11.25
户外日照时长>4 h	1.677	0.469	12.811	<0.001	5.348	2.060,12.88
职业	1.827	0.426	18.423	<0.001	6.125	4.008,10.21
吸烟史	0.673	0.275	5.993	0.015	1.980	1.135,3.546
饮酒史	0.839	0.307	7.472	0.025	2.295	1.109,4.691
性别(女)	-0.683	0.185	13.627	<0.001	0.520	0.354,0.721
睑板腺开口堵塞	0.974	0.408	5.701	<0.001	5.248	1.108,11.75
常量	-2.361	0.472	25.016	<0.001		

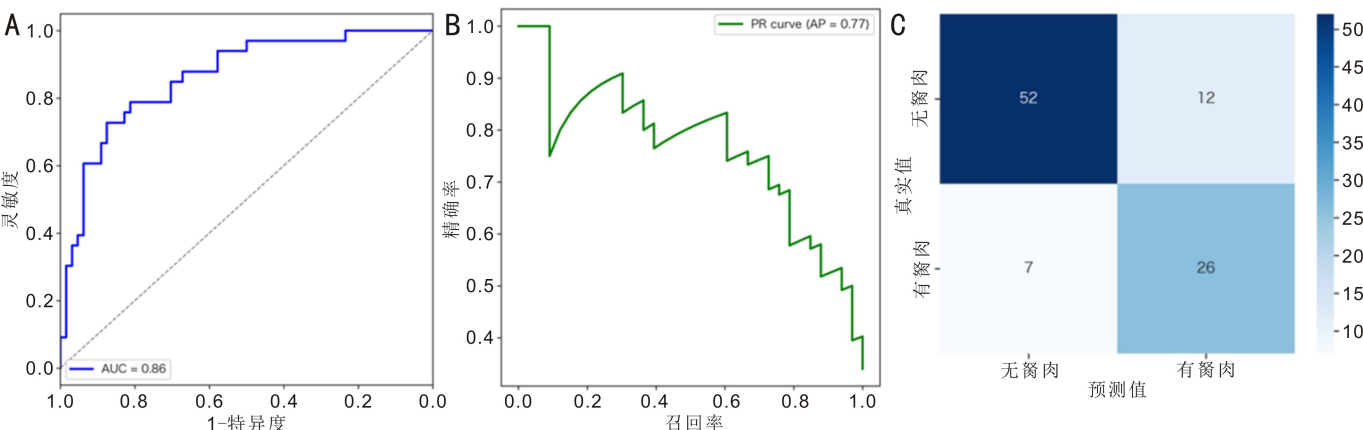


图 3 模型性能评估 A:ROC 曲线;B: PR 曲线; C: 混淆矩阵。

80.41%,假阳性 12 例,假阴性 7 例,提示未来可通过阈值优化进一步降低漏检率,见图 3C。

2.5 单因素权重分析 在多模态模型训练完成后,为揭示翼状胬肉的关键驱动因素并与统计学分析结果相比较,本

研究对各调研因素进行 PCA 降维^[15](保留前 8 主成分),基于逻辑回归系数的绝对值求和计算各因素的归一化贡献权重。权重排序结果以水平条形图展示,直观呈现风险因素的重要性层级,见图 4。PCA 权重分析显示,睑板腺

堵塞情况、年龄、户外日照时长、职业等因素的贡献权重较高,与表2多因素 Logistic 回归中 OR 值较大的因素基本吻合,表明多模态模型与传统统计学分析在识别翼状胬肉关键危险因素方面具有良好的一致性。

2.6 模型可视化与聚类分析 采用 t-SNE 将 8 维特征降维投影至二维空间进行可视化。结果显示,翼状胬肉患者与非患者在特征空间中呈现明显的聚类分布,分类边界清晰,少数错分样本集中于簇边缘,提示可能存在特征交叉或噪声干扰。该可视化结果有助于临床医生直观理解模型的风险分层逻辑,见图 5。

3 讨论

本研究通过对舟山群岛 1 842 名居民的横断面调查,发现该地区翼状胬肉总体患病率为 34.96%,远高于全球 12% 的患病率^[16],也高于同属浙江省的杭州等非沿海城市(18%)^[17],与云南省的大理白族(39%)^[18],文山州的苗族、壮族(35.27%)^[19]等高紫外线暴露地区报道值接近。

多因素 Logistic 回归分析显示,职业类型是影响翼状胬肉患病的最强危险因素,渔民农民职业的 OR 达 6.125;日均户外日照时长≥4 h 者的 OR 值为 5.348。该结果与张晓英^[19]在云南文山州的研究趋势一致,但本研究中渔农民的 OR 值更高,可能源于舟山群岛海域水面反射增强紫外线辐照剂量,且高盐湿海风持续刺激眼表上皮,形成协同损伤效应^[20-21]。机制层面,紫外线可通过诱导结膜上皮细胞凋亡抑制蛋白上调,增强翼状胬肉上皮细胞的抗凋亡能力及基质侵袭力^[22-23];同时,长期风尘与盐粒刺激可加剧局部慢性炎症,二者叠加可能是本地区患病率高于内陆强日照地区的重要病理基础^[24-25]。上述结果提示,东南沿海特定地理环境可能构成翼状胬肉高发的主要因素。

本研究中年龄>40 岁被证实为危险因素($OR=5.295, P<0.001$),与邢懿等^[26]、张晓英^[19]研究结果一致。但其发生的内在机制无法明确推断,我们推测与年龄增加导致机体生理状况发生改变,以及年龄越大,其紫外线照射总时长越长有关。

巴西研究曾报道 88% 的翼状胬肉患者伴有睑板腺形态学改变^[27],本研究进一步验证了睑板腺开口阻塞与翼状胬肉患病的关联性,644 例患者中合并开口阻塞者 426 例(66.1%),与柯红琴等^[28]及 Pan 等^[29]研究结论相符。推测其病理生理联系为:睑板腺开口阻塞导致睑脂分泌质与量异常,泪膜脂质层缺陷引发泪液蒸发过强、泪膜稳定性下降,进而导致眼表高渗透压状态及炎症因子释放增加,形成促纤维血管增殖的微环境^[27-28]。但当前证据仅限于横断面关联,尚不能确立因果关系;两者之间是否存在双向促进或仅为伴随现象,有待纵向队列研究进一步佐证。

本研究中男性患病风险高于女性($OR=0.520, P<0.001$),吸烟与饮酒亦与患病呈现正向关联,与既往多数文献一致^[30]。但目前尚无确凿证据支持性别本身独立致病。本研究推测该结果可能与义诊人群中男性多为渔农民有关,其户外暴露强度及吸烟饮酒频率亦相应增加,性别差异在调整职业因素后可能被削弱。而文化程度、看视频终端时长等在本次研究中均非导致翼状胬肉的危险因

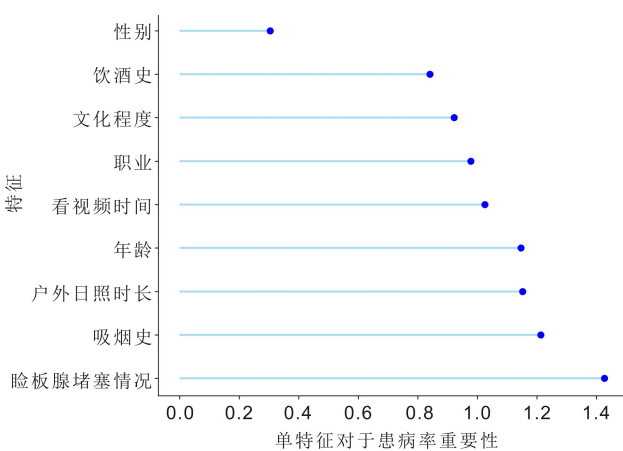


图4 翼状胬肉患病率单因素预测特征重要性。

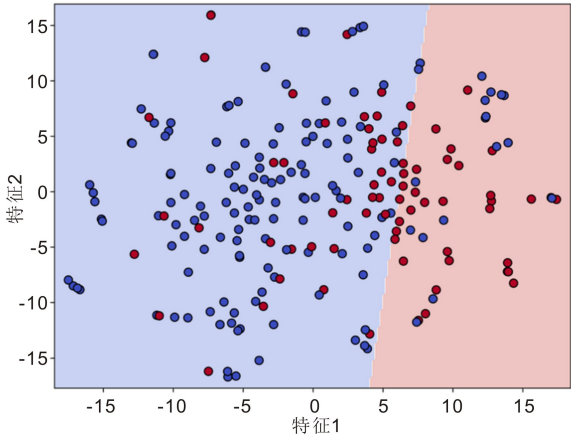


图5 翼状胬肉降维可视化聚类及决策边界。

素。但是否真正未无关因素,还需进一步扩大样本研究。此外,本研究对象来源于义诊就诊人群,该人群可能存在较高的眼部健康问题,与社区自然人群存在差异,可能导致患病率的高估,未来需采用更严格的社区随机抽样加以验证。

本研究将基于 Sentence Transformer 的语义编码技术引入东南沿海地区翼状胬肉风险预测模型,突破了既往研究中对义诊自由文本(如“1 包/日,40 a”“织网”“三轮车工”等)进行人工离散化或简单二分类的传统范式。通过预训练语言模型将非结构化的异质性文本映射至高维语义空间,在保留吸烟量、持续时间、职业特征等原始粒度信息的同时,融合结构化临床指标构建逻辑回归模型。内部验证显示,该模型 AUC 为 0.86,AP 值为 0.77,测试集整体准确率 80.41%,且特征权重排序与临床回归分析结果一致,证实其具备良好区分度与稳健性。该技术路径为基层义诊及大规模流行病学筛查中非标准化数据的深度挖掘提供了可复用的解决方案,同时模型输出的个体化风险概率可辅助基层医师精准识别高危亚群,为早期干预提供量化决策支持。

然而,本研究对象源自义诊人群,且为单中心数据,可能导致患病率等数据存在偏差与局限性。希望未来可以更广泛、更详细、东南沿海不同地域的进一步研究。东南沿海舟山地区年龄>40 岁,户外日照时长超过 4 h/d 的渔农民是翼状胬肉患者高发群体,需重点关注,加强眼部健康宣教,减少紫外线眼部辐射,定期检查睑板腺功能,适当给予人工泪液,是减少该人群翼状胬肉高发的必要措施。

本研究的预测模型可对翼状胬肉进行准确预测,为临床医师早期干预疾病提供了理论依据。

利益冲突声明:本文不存在利益冲突。

作者贡献声明:叶霞论文选题,初稿撰写,审稿及修改;周喆媛机器模型建立,数据分析及模型部分初稿撰写,审稿及修改;刘震文献检索;王施鲲、赵娜数据分析。所有作者阅读并同意最终的文本。

参考文献

[1] Li XM, Ge HM, Yao J, et al. Genome-wide identification of circular RNAs as a novel class of putative biomarkers for an ocular surface disease. *Cell Physiol Biochem*, 2018,47(4):1630-1642.

[2] 潘周娴,单广良,王雪娇,等. 甘肃省汉族和裕固族的翼状胬肉患病率及相关危险因素研究. *中华眼科杂志*, 2020,56(8):600-607.

[3] Alsarhani W, Alshahrani S, Showail M, et al. Characteristics and recurrence of pterygium in Saudi Arabia; a single center study with a long follow-up. *BMC Ophthalmol*, 2021,21(1):207.

[4] Hung KH, Lin C, Roan J, et al. Application of a deep learning system in pterygium grading and further prediction of recurrence with slit lamp photographs. *Diagnostics (Basel)*, 2022,12(4):888.

[5] 黎晓新,王宁利. *眼科学*. 北京:人民卫生出版社,2016:180-181.

[6] Jurafsky D, Martin JH. *Speech and language processing (3rd ed.)*. Prentice Hall, 2023.

[7] Reimers N, Gurevych I. Sentence-BERT: sentence embeddings using Siamese BERT-networks. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Hong Kong, China Association for Computational Linguistics, 2019:3982-3992.

[8] Cox DR. The regression analysis of binary sequences. *J R Stat Soc Ser B Stat Methodol*, 1958,20(2):215-232.

[9] VanRijsbergen CJ. *Information Retrieval*. 2nd ed. London; Boston; Butterworths, 1979: 6.

[10] Manning CD, Raghavan P, Schütze H. *Introduction to Information Retrieval*. Cambridge; Cambridge University Press, 2008: 42.

[11] Pearson K. LIII. On lines and planes of closest fit to systems of points in space. *Lond Edinb Dublin Philos Mag J Sci*, 1901,2(11):559-572.

[12] Davis J, Goadrich M. The relationship between precision-recall and ROC curves. *New York ACM*, 2006: 233-240.

[13] Powers DMW. Evaluation: from precision, recall and f-measure to roc, informedness, markedness & correlation. *J Mach Learn Technolog*, 2011,2(1):37-63.

[14] Cui Y, Jia ML, Lin TY, et al. Class-balanced loss based on effective number of samples. *IEEE*, 2019: 9260-9269.

[15] Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 1982,143(1):29-36.

[16] Rezvan F, Khabazkhoob M, Hooshmand E, et al. Prevalence and risk factors of pterygium: a systematic review and meta-analysis. *Surv Ophthalmol*, 2018,63(5):719-735.

[17] 陈昌凤,倪琴,宣光城. 翼状胬肉切除术后复发的危险因素分析. *国际眼科杂志*, 2025,25(2):311-314.

[18] 张方琳,赵新荣,万慧娟,等. P53 和 mTOR 在翼状胬肉的表达和相关性分析. *国际眼科杂志*, 2024,24(9):1381-1386.

[19] 张晓英. 高原地区翼状胬肉相关危险因素分析及预测模型的建立. *国际眼科杂志*, 2022,22(7):1215-1219.

[20] Lai CC, Tseng SH, Hsu SM, et al. Conjunctival expression of toll-like receptor 3 plays a pathogenic role in the formation of ultraviolet light-induced pterygium. *Invest Ophthalmol Vis Sci*, 2021,62(10):6.

[21] 陈杨,魏超群,弥禹,等. 翼状胬肉相关致病因子的研究进展. *国际眼科杂志*, 2022,22(7):1133-1136.

[22] Park M, Mazalo J, Di Girolamo N. Insulin-like growth factor binding protein-7: a marker of conjunctivalization in an animal model of limbal stem cell deficiency. *Ocul Surf*, 2019,17(3):447-457.

[23] 董彦平,王洁敏,王玉瑾,等. 翼状胬肉术后复发 Nomogram 预测模型的构建与评估. *实用临床医药杂志*, 2022,26(7):52-56.

[24] 吴静,田静静,孟捷. 翼状胬肉切除联合角膜缘干细胞移植术后感染危险因素分析及预测模型构建. *中国临床研究*, 2025,38(8):1260-1264.

[25] 赵殊艺,崔金鹏,刘晓晨. 日光辐射与翼状胬肉发生的相关性研究进展. *国际眼科杂志*, 2025,25(11):1802-1807.

[26] 邢懿,江方方,姜波,等. 合肥地区 116 例老年患者翼状胬肉流行病学表现及危险因素分析. *中华全科医学*, 2020,18(8):1279-1282.

[27] Wanzeler ACV, Barbosa IAF, Duarte B, et al. Impact of pterygium on the ocular surface and meibomian glands. *PLoS One*, 2019,14(9):e0213956.

[28] 柯红琴,张文佳,刘海,等. 云南省文山州 40 岁及以上壮族和苗族人群翼状胬肉患病率及其影响因素. *国际眼科杂志*, 2022,22(2):347-351.

[29] Pan ZX, Cui JT, Shan GL, et al. Prevalence and risk factors for pterygium: a cross-sectional study in Han and Manchu ethnic populations in Hebei, China. *BMJ Open*, 2019,9(2):e025725.

[30] Qadi R, AlAmri A, Elnashar M, et al. Prevalence of pterygium and associated risk factors in the high-altitude area of Ta'if City, Saudi Arabia. *Cureus*, 2021,13(1):e12638.