DOI:10.16016/j.2097-0927.202408110

# 高原脱适应症发生风险预测模型的构建与验证

丁 宇[1,3],王泽军[1,3],谢佳新[2,3],赵思雨[1,3],张 钢[1,3] 陆军军医大学(第三军医大学)高原军事医学系:[1]寒区医学教研室,[2]高原作业医学教研室,[3]极端环境医学教育部重点实验室,重庆

[摘要] **目的** 应用不同机器学习算法构建从高原返回平原人群高原脱适应症(high altitude deacclimatization syndrome,HADAS)发病的风险预测模型,并验证其预测效能。**方法** 于2020年11月至2024年2月对结束高原生活返回内地的人群实地或线上发放问卷调查。收集基本资料、慢性高原病(chronic mountain sickness,CMS)情况和脱适应症状等调查资料,经筛选最终纳入1 095例作为建模组。阳性事件定义为脱适应症状评分>5分。将建模组按7∶3随机分为训练集($n$=766)和内部测试集($n$=329),采用最小绝对收缩和选择算子(least absolute shrinkage and selection operator,LASSO)回归筛选自变量,基于多因素Logistic回归(multiple factor logistic regression,LR)、决策树(decision tree,DT)、随机森林(random forest,RF)、极端梯度提升(eXtreme gradient boosting,XGB)、支持向量机(support vector machine,SVM)、K最邻近(K-nearest neighbor,KNN)、轻度梯度提升(light gradient boosting,LGB)、朴素贝叶斯(naïve bayes,NB)8种机器学习方法构建预测模型。采用受试者工作特性曲线(receiver operating characteristic,ROC)、校准曲线和混淆矩阵对模型进行比较并进行内部测试;使用列线图或Shapley加性解释(shapley additive explanations,SHAP)图对最终模型进行展示。于2024年8月收集筛选结束高原生活返回平原132例作为外部验证组,对模型进行外部验证。**结果** 1 095例调查人群中有脱适应症者549人(50.14%)。CMS评分、年龄、高原居住时间为LASSO回归筛选出的预测因子。8种机器学习算法建立的HADAS预测模型中,以LR模型最优,ROC的曲线下面积(area under curve,AUC)训练集为0.819(95%*CI*:0.789~0.850),内部测试集为0.841(95%*CI*:0.799~0.884),F1评分内部测试集为0.801,内部测试集的AUC、F1得分在8个模型中均为最大;LR模型校准曲线的Spiegelhalter $Z$检验显示训练集$P$=0.703、内部测试集$P$=0.281;LR模型外部验证集AUC为0.867(95%*CI*:0.765~0.969)。**结论** 以CMS评分、年龄和高原居住时间为预测因子建立的LR模型在内部测试集的综合表现最好,在外部验证集中区分度好,构建的列线图便于应用。

[关键词] 预测模型;高原脱适应症;机器学习;列线图

[中图法分类号] R195.4;R363.22;R594.3 [文献标志码] A

# Construction and validation of a risk prediction model for high altitude deacclimatization syndrome

DING Yu[1,3], WANG Zejun[1,3], XIE Jiaxin[2,3], ZHAO Siyu[1,3], ZHANG Gang[1,3] [1]Department of Cold Region Medicine; [2]Department of High Altitude Operational Medicine; [3]Key Laboratory of Extreme Environmental Medicine of Ministry of Education, Faculty of High Altitude Military Medicine, Army Medical University (Third Military Medical University), Chongqing, China

[**Abstract**] **Objective** To construct risk models for predicting the occurrence of high altitude deacclimatization syndrome (HADAS) in the population returning from the plateau to the plain based on

[基金项目] 国家自然科学基金青年基金(82001991)

[通信作者] 张 钢, E-mail:zhanggang196868@163.com

2097-0927/Copyright © 2025, 陆军军医大学(第三军医大学)。这篇开放获取文章遵循 CC BY 许可协议 (https://creativecommons.org/licenses/by/4.0/)

**引用格式**:丁宇,王泽军,谢佳新,等. 高原脱适应症发生风险预测模型的构建与验证[J]. 陆军军医大学学报,2025,47(01):20-29. DOI:10.16016/j.2097-0927.202408110.

**Citation**:DING Y,WANG Z J,XIE J X,*et al*. Construction and validation of a risk prediction model for high altitude de-acclimatization syndrome[J]. J Army Med Univ,2025,47(01):20-29. DOI:10.16016/j.2097-0927.202408110.

different machine learning algorithms and validate the predicting efficiency of these models. **Methods** Field or online surveys were conducted on the individuals who had ended their high-altitude living and returned to the plain areas from November 2020 to February 2024. Basic information, chronic mountain sickness (CMS), HADAS symptoms and other data were collected. With the inclusion and exclusion criteria, totally 1 095 individuals were subjected and assigned into the modeling group. Positive events were defined as HADAS score >5. Then the modelling group was divided into a training set ($n$=766) and an internal test set ($n$=329) in a 7∶3 ratio. Least absolute shrinkage and selection operator (LASSO) regression was used to select independent variables. Risk prediction models for high-altitude adaptation symptoms were built based on 8 machine learning methods, including multiple factor logistic regression (LR), decision tree (DT), random forest (RF), eXtreme gradient boosting (XGB), support vector machine (SVM), K-nearest neighbor (KNN), light gradient boosting (LGB) and naïve bayes (NB). The models were compared and evaluated using receiver operating characteristic (ROC) curves, calibration curves and confusion matrices in the internal test set. The final model was presented using a nomogram or Shapley additive explanations (SHAP) algorithm. In August 2024, another 132 individuals who returned to the plains and met the same criteria were recruited and served as the external validation group. **Results** There were 549 individuals (50.14%) out of the 1 095 subjects having HADAS symptoms. LASSO regression identified CMS score, age and duration of high-altitude residence as significant predictors. Among the 8 machine learning algorithms, the LR model was identified as the best, with an area under the curve (AUC) value of 0.819 (95%*CI*: 0.789~0.850) and 0.841 (95%*CI*: 0.799~0.884), and an F1 score of 0.801 in the internal test set, respectively, and the AUC value and F1 score of the LR model were the largest among the 8 models in the internal test set. Spiegelhalter $Z$ test of the calibration curve of the LR model indicated that its $P$=0.703 in the training set while $P$=0.281 in the internal test set. The AUC value of the LR model was 0.867(95%*CI*: 0.765~0.969) in the external validation set. **Conclusion** The LR model constructed based on indicators including CMS score, age and duration of high-altitude residence has a good overall performance in the internal test set, and good discriminating effect in the external validation set. The constructed nomogram is convenient for application.

[**Key words**] prediction model; high altitude de-acclimatization syndrome; machine learning; nomogram

Supported by the National Natural Science Foundation of Young Scholars of China (82001991). Corresponding authors: ZHANG Gang, E-mail: zhanggang196868@163.com

随着经济发展及国防建设需要,往返高原的人越来越多。以往多关注平原人移居高原后发生高原病的情况,对返回平原后的生理、病理变化缺乏系统研究。高原世居者或已习服高原的移居人群在脱离高原环境回到平原后,机体消除对高原低氧环境的适应性重新适应平原环境的变化过程称为高原脱适应[1]。这是机体对氧含量相对增多的一种代偿性调节。研究发现,在高原脱适应过程中,有60%~90%的人群返回平原后出现神经系统、循环系统等一系列症状及心、肺、血液指标异常[2],给工作生活带来影响,严重者不得不重返高原[3]——这种全身多系统综合性疾病称为高原脱适应症(high altitude de-acclimatization syndrome,HADAS)。HADAS的发病情况与多种因素有关。以往多认为,高原居住地海拔越高、居住时间越长、返回平原时年龄越大,HADAS发生率越高[4-5],但各影响因素的重要性及量效关系仍未可知。目前国内外缺少针对HADAS的临床预测模型。通过筛选影响HADAS发生的风险因素,并以此为依据建立发病风险预测模型,筛查HADAS易感者并在其下山之前和返回平原的过程中进行早期干预,采取相应的保护措施,有助于为HADAS的防治提供科学依据。

随着人工智能在医疗领域的广泛应用,机器学习、深度学习、大模型等开始用于临床辅助诊断与决策,使预测模型的构建方法得到极大丰富[6]。本研究应用8种机器学习算法建立HADAS的发病风险预测模型,比较筛选最优模型并对其进行可视化解释,探讨影响HADAS发病的相关风险因素,有助于早预防、早发现、早干预,为HADAS的防治措施提供证据。

## 1 对象与方法

### 1.1 研究对象

选取2020年11月(返回地海拔1 060 m)、2021年9月(返回地海拔510 m)、2024年2月(返回地海拔88 m)结束高原生活返回内地的3批男性人群进行实地调查

以及对返回平原的高原干部线上调查(返回地为四川和重庆,平均海拔约为500 m)。收集研究问卷,该批数据作为建模组用于建立模型。将课题组跟踪的2024年8月返回平原的脱适应症队列人群作为前瞻性外部测试组,该群体在4 300 m海拔高原定居9个月,返回平原(返回地海拔800 m)1周后发放HADAS评分表。

### 1.2 纳入排除标准

参考周其全等[7]HADAS诊断标准,制定纳入标准:①年龄≥18岁;②高原移居汉族人群;③近期已从高原返回平原。排除标准:①有高血压、糖尿病、冠心病、肿瘤等基础疾病;②临床资料缺失率>20%;③调查时返回平原时间超过5年。于2024年8月再收集纳入132名受试者作为外部测试集。本研究经陆军军医大学第二附属医院伦理委员会审查同意研究(2022-研第505-01)。

### 1.3 调查工具

本研究采用课题组自主设计的HADAS调查表,由3部分组成:①一般资料。包括年龄、体质指数(body mass index,BMI)、文化程度、基础疾病、高原居住时间、高原居住地海拔、返回平原时间等。②高原习服状态。采用慢性高原病(chronic mountain sickness,CMS)评分表[8],参照2004年青海慢性高原病诊断标准,包括气喘和心悸、失眠、紫绀、咽部及结膜血管扩张、感觉异常、头痛、耳鸣等7个症状参与评分。③脱适应情况。采用HADAS评分表[9],头昏、嗜睡、精力不集中、失眠、头晕、疲乏、胸闷、心慌、咳喘、便秘、纳差、心前区疼痛、浮肿等13个常见症状参与计分。

### 1.4 诊断与分组

在收集HADAS相关症状后,根据周其全等[7]报道的HADAS症状评分计算方法计算、诊断,每个症状根据严重程度分别赋分为:0分(基本无症状)、1分(有症状但不影响日常工作)、2分(有症状需药物治疗但不影响工作)、3分(症状较重需休息治疗),计算所有症状总分,总分>5分者诊断为存在HADAS。根据HADAS有无分别将建模组、外部测试组人群分为有HADAS组和无脱适应症(Non-HADAS)组。

### 1.5 数据预处理与模型构建

将预测指标进行预处理,根据身高和体质量计算BMI,并根据世界卫生组织肥胖程度分类标准将其转换为等级资料。高原居住时间、高原居住地海拔、返回平原时间、文化程度以分类变量纳入,年龄、CMS评分直接以连续变量纳入。处理后的建模组数据按7∶3随机拆分为训练集和内部测试集;在高原现场收集外部验证组基本情况和CMS评分等资料,利用本研究得到的最优模型预测返回平原后是否发生HADAS,待该群体返回平原后为其发放脱适应评分表并对有HADAS的人群做出诊断,对比HADAS预测值与观测值,评估该模型的预测能力。自变量筛选、机器学习建模过程及列线图绘制使用R软件[R software package (version 4.2.1) and DCPM (V4.01, Jingding Medical Technology Co., Ltd.)]。数据集划分采用caret包;最小绝对收缩和选择算子(least absolute shrinkage and selection operator,LASSO)回归采用glmnet包;决策树(decision tree,DT)采用rpart包;随机森林(random forest,RF)采用randomForest包;极端梯度提升(eXtreme gradient boosting,XGB)采用xgboost包;支持向量机(support vector machine,SVM)、朴素贝叶斯(naïve bayes,NB)采用e1071包,K最邻近(K-nearest neighbor,KNN)采用kknn包;轻度梯度提升(light gradient boosting,LGB)采用lightgbm包;多因素Logistic回归(multiple factor logistic regression,LR)及列线图绘制采用glm包。在训练集中采用LASSO回归筛选有意义的特征变量,将筛选后的变量纳入到LR、DT、RF、XGB、SVM、NB、KNN和LGB共8个机器学习模型中。

将建立好的8个模型在内部测试集中进行独立测试,绘制ROC曲线并计算AUC,比较各模型区分度,绘制校准曲线比较各模型校准度,通过混淆矩阵对建立的各模型进行综合比较,选择合适的预测模型[10]。使用列线图或SHAP图对最优机器学习模型进行可视化展示。最后在外部验证集中对模型的预测效果进行评估。研究技术路线见图1。

### 1.6 统计学分析

采用Excel 2019、SPSS 26.0进行数据分析。符合正态分布的计量资料采用$\bar{x}\pm s$形式表示,组间比较采用独立样本 $t$ 检验;不符合正态分布的计量资料以$M(P_{25}, P_{75})$表示,组间比较采用Mann-Whitney $U$检验。分类变量以频次(%)表示,组间比较采用$\chi^2$检验。$P<0.05$为差异具有统计学意义。采用R软件中8种机器学习算法包进行模型构建、图像绘制及各模型评价指标计算。

## 2 结果

### 2.1 受试者一般资料

经严格对照纳排标准并删除缺失值,建模组最终纳入1 095例研究对象,其中有546例(49.9%)脱适应症评分>5分并被诊断为HADAS。将所有人群按7∶3随机划分为训练集($n$=766)和内部测试集($n$=329)后,得到训练集中有372例(48.56%)发生HADAS,内部测试集中有174例(52.89%)发生HADAS。HADAS

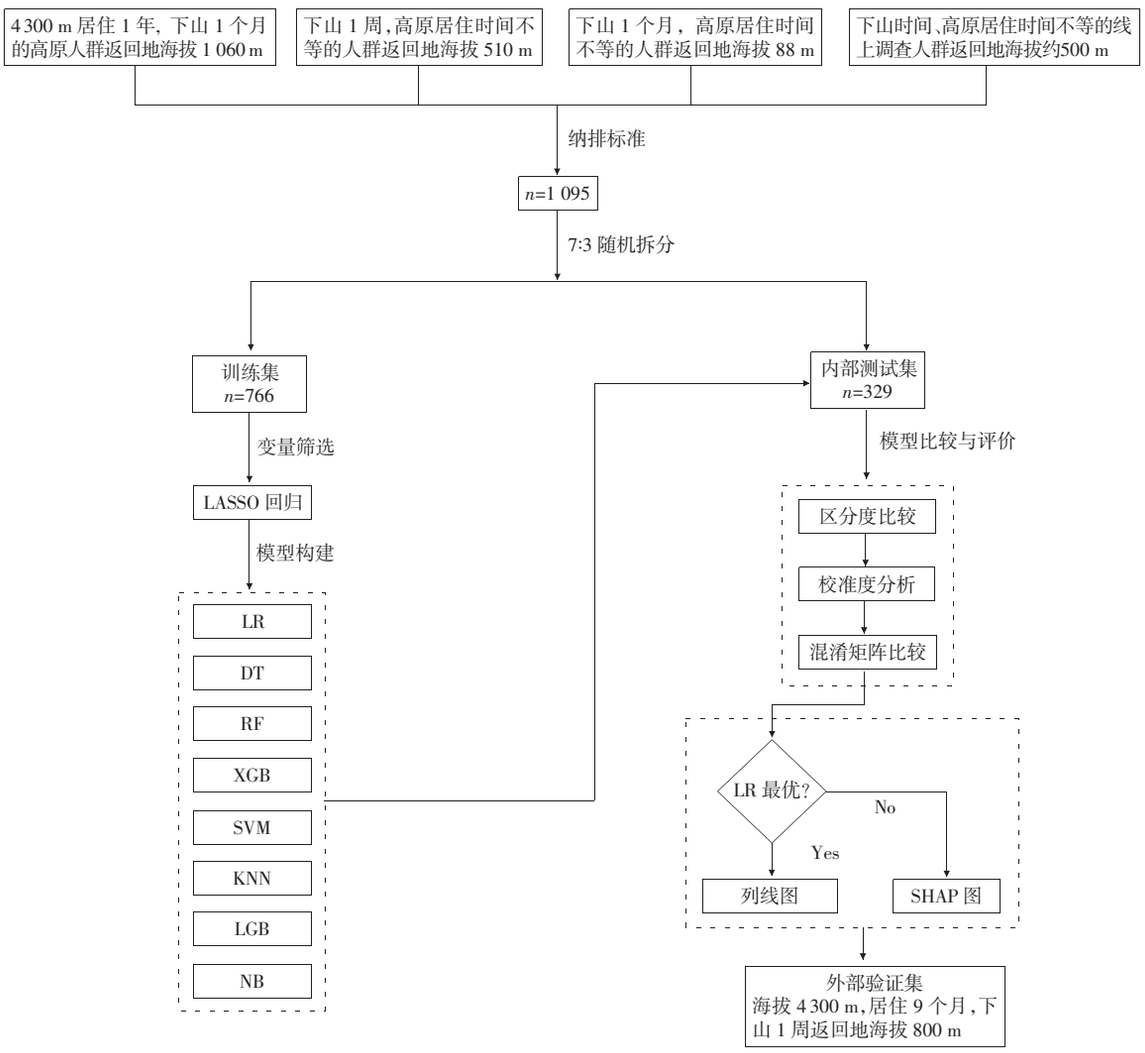



**图1 高原脱适应症风险机器学习模型构建、评价和外部测试的研究流程图**

组与Non-HADAS组比较,高原居住时间、高原居住地海拔、受教育程度、BMI、CMS评分、返回平原时长、年龄之间差异均有统计学意义;外部测试组最终纳入132例研究对象,其中有10例(7.58%)HADAS评分>5分并被诊断为HADAS(表1)。经统计学分析,训练集和内部测试集仅在CMS评分有显著性差异,其余指标差异无统计学意义(表2),表明训练集与内部测试集随机拆分较好。

### 2.2 变量筛选

采用LASSO回归对变量进行筛选,以期选出能够高效预测HADAS发生的特征变量。使用10折交叉验证选择最优Lambda参数,均方误差最小时的Lambda.min为最佳变量筛选的最优解(图2左侧虚线),未筛出任何变量;当λ=Lambda.1se时,回归结果显示CMS评分、年龄、高原居住时间为影响HADAS发生的危险因素(图2右侧虚线)。选择Lambda.1se作为最佳λ值,能够使得到的预测模型更精简。

### 2.3 模型构建

在软件中输入LASSO回归筛选出的特征变量,基于LR、NB、LGB、SVM、DT、RF、KNN、XGB共8种机器学习算法构建HADAS发生风险预测模型[11]。其中LR以赤池准则(akaike's information criterion,AIC)最小为评价准则建模,回归分析结果如表3所示,模型函数展示如下:

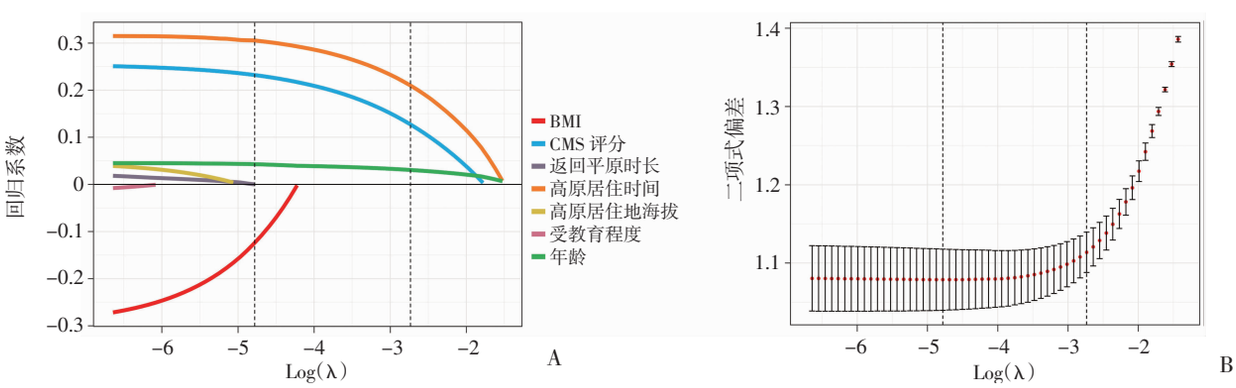
$$P(y=1|x)=\frac{1}{1+e^{-[(-2.895+0.321(高原居住时间)+0.043(年龄)+0.240(CMS评分)]}}$$

### 2.4 模型比较与评价

对8种模型的疾病预测能力进行评价。ROC曲线显示见图3A、B,在内部测试集中LR模型的AUC(0.841)最大;通过混淆矩阵评价各模型详细性能(表4),其中

**表1 建模组与外部验证组有无HADAS人群的一般资料[$M(P_{25},P_{75})$,例(%)]**

| 指标 | 建模组($n$=1 095) | | 外部验证组($n$=132) | |
|---|---|---|---|---|
| | Non-HADAS($n$=549) | HADAS($n$=546) | Non-HADAS($n$=122) | HADAS($n$=10) |
| 高原居住时间 | | | | |
| ≤6个月 | 95(17.30) | 32(5.86) | 0 | 0 |
| >6个月~2年 | 328(59.74) | 134(24.54) | 122(100.00) | 10(100.00) |
| >2~5年 | 21(3.83) | 17(3.11) | 0 | 0 |
| >5~8年 | 10(1.82) | 14(2.56) | 0 | 0 |
| >8年 | 95(17.30) | 349(63.92) | 0 | 0 |
| 高原居住地海拔/m | | | | |
| ≤2 500 | 57(10.38) | 21(3.85) | 0 | 0 |
| >2 500~3 500 | 33(6.01) | 24(4.40) | 0 | 0 |
| >3 500~4 500 | 399(72.68) | 446(81.68) | 122(100.00) | 10(100.00) |
| >4 500 m | 60(10.93) | 55(10.07) | 0 | 0 |
| 受教育程度 | | | | |
| 研究生 | 7(1.28) | 16(2.93) | 0 | 0 |
| 大学 | 355(64.66) | 444(81.32) | 91(74.59) | 6(60.00) |
| 中专或高中 | 173(31.51) | 84(15.38) | 30(24.59) | 4(40.00) |
| 初中及以下 | 14(2.55) | 2(0.37) | 1(0.82) | 0 |
| BMI/(kg/m$^2$) | | | | |
| ≤18.5 | 18(3.28) | 11(2.01) | 3(2.46) | 0 |
| >18.5~25 | 464(84.52) | 390(71.43) | 101(82.79) | 7(70.00) |
| >25~30 | 58(10.56) | 138(25.27) | 18(14.75) | 3(30.00) |
| >30 | 9(1.64) | 7(1.28) | 0 | 0 |
| CMS评分/分 | 2.00(0.00,4.00) | 4.00(3.00,7.00) | 1.00(0.00,2.00) | 5.00(3.50,5.75) |
| 返回平原时长 | | | | |
| ≤1周 | 249(45.36) | 98(17.95) | 122(100.00) | 10(100.00) |
| >1周~1个月 | 203(36.98) | 86(15.75) | 0 | 0 |
| >1个月~1年 | 4(0.73) | 6(1.10) | 0 | 0 |
| >1~5年 | 93(16.94) | 356(65.20) | 0 | 0 |
| 年龄/岁 | 23.00(21.00,26.00) | 42.02(24.00,45.02) | 26.00(24.00,30.00) | 26.00(25.00,29.50) |




A:LASSO回归变量收缩图;B:LASSO回归交叉验证曲线图

**图2 训练集内与高原脱适应症相关重要变量的LASSO回归结果**

LR的F1得分(0.801)、特异度(0.759)在8个模型中居首位,Brier评分(0.163)、阴性预测值(0.802)、约登指数(0.560)排第2位。图3C、D显示LR模型的校准曲线较其他多数曲线更接近y=x,一致性强。综合比较上述指标,LR模型预测效能最佳。对LR模型进一步评估,为验证各预测因子之间是否存在共线性,如年龄越大者高原居住时间可能也较久,分别计算了LR中各预测因子的方差膨胀系数:CMS评分1.008,高原居住时间4.811,年龄4.824,均<5,表明不存在共线性。计算LR模型校准曲线训练集和内部测试集的平均绝对误差(mean absolute error, MAE),分别为0.050、0.028,均较小。Spiegelhalter *Z*检验显示训练集*P*=0.703、内部测试集*P*=0.281,均>0.05,模型校准度可。

**表2 训练集与内部测试集的一般资料比较**

[$M(P_{25},P_{75})$,例(%)]

| 指标 | 训练集(*n*=766) | 内部测试集(*n*=329) | *P* |
|---|---|---|---|
| Non-HADAS | 394(51.44) | 155(47.11) | 0.213 |
| HADAS | 372(48.56) | 174(52.89) | |
| 高原居住时间 | | | 0.684 |
| ≤6个月 | 94(12.27) | 33(10.03) | |
| >6个月~2年 | 321(41.91) | 141(42.86) | |
| >2~5年 | 26(3.39) | 12(3.65) | |
| >5~8年 | 19(2.48) | 5(1.52) | |
| >8年 | 306(39.95) | 138(41.95) | |
| 高原居住地海拔/m | | | 0.336 |
| ≤2 500 | 57(7.44) | 21(6.38) | |
| >2 500~3 500 | 38(4.96) | 19(5.78) | |
| >3 500~4 500 | 583(76.11) | 262(79.64) | |
| >4 500 | 88(11.49) | 27(8.21) | |
| 受教育程度 | | | 0.834 |
| 研究生 | 16(2.09) | 7(2.13) | |
| 大学 | 559(72.98) | 240(72.95) | |
| 中专或高中 | 178(23.24) | 79(24.01) | |
| 初中及以下 | 13(1.70) | 3(0.91) | |
| BMI/(kg/m$^2$) | | | 0.625 |
| ≤18.5 | 18(2.35) | 11(3.34) | |
| >18.5~25 | 597(77.94) | 257(78.12) | |
| >25~30 | 138(18.02) | 58(17.63) | |
| >30 | 13(1.70) | 3(0.91) | |
| CMS评分/分 | 3.00(1.00,5.00) | 3.00(2.00,6.00) | 0.008 |
| 返回平原时长 | | | 0.990 |
| ≤1周 | 244(31.85) | 103(31.31) | |
| >1周~1个月 | 203(26.50) | 86(26.14) | |
| >1个月~1年 | 7(0.91) | 3(0.91) | |
| >1~5年 | 312(40.73) | 137(41.64) | |
| 年龄/岁 | 25.00(22.00,43.02) | 27.00(22.00,43.02) | 0.474 |

基于上述LR模型,绘制预测HADAS发生风险的列线图(图4)。上述3个指标在列线图中分别对应的分值相加,计算出总分,总分对应的概率值即为该受试者HADAS的发生概率预测值。为便于应用,在此列线图基础上开发了基于网页计算器的风险评估工具[12],网址为https://hadas.shinyapps.io/HADASRiskPridictingModel/_w_452729fa/#。

### 2.5 模型外部验证

对LR模型的预测能力进行外部验证,见图5。结果显示外部验证组AUC为0.867(95%*CI*:0.765~0.969),敏感度为0.700,特异度为0.910,表明该模型具有较好的区分度;校准曲线显示,LR模型在外部验证集的校准曲线整体位于y=x的下方,表明预测概率整体高于实际概率,该模型高估了外部验证组HADAS的发生概率。

## 3 讨论

### 3.1 HADAS影响返回平原人员的健康

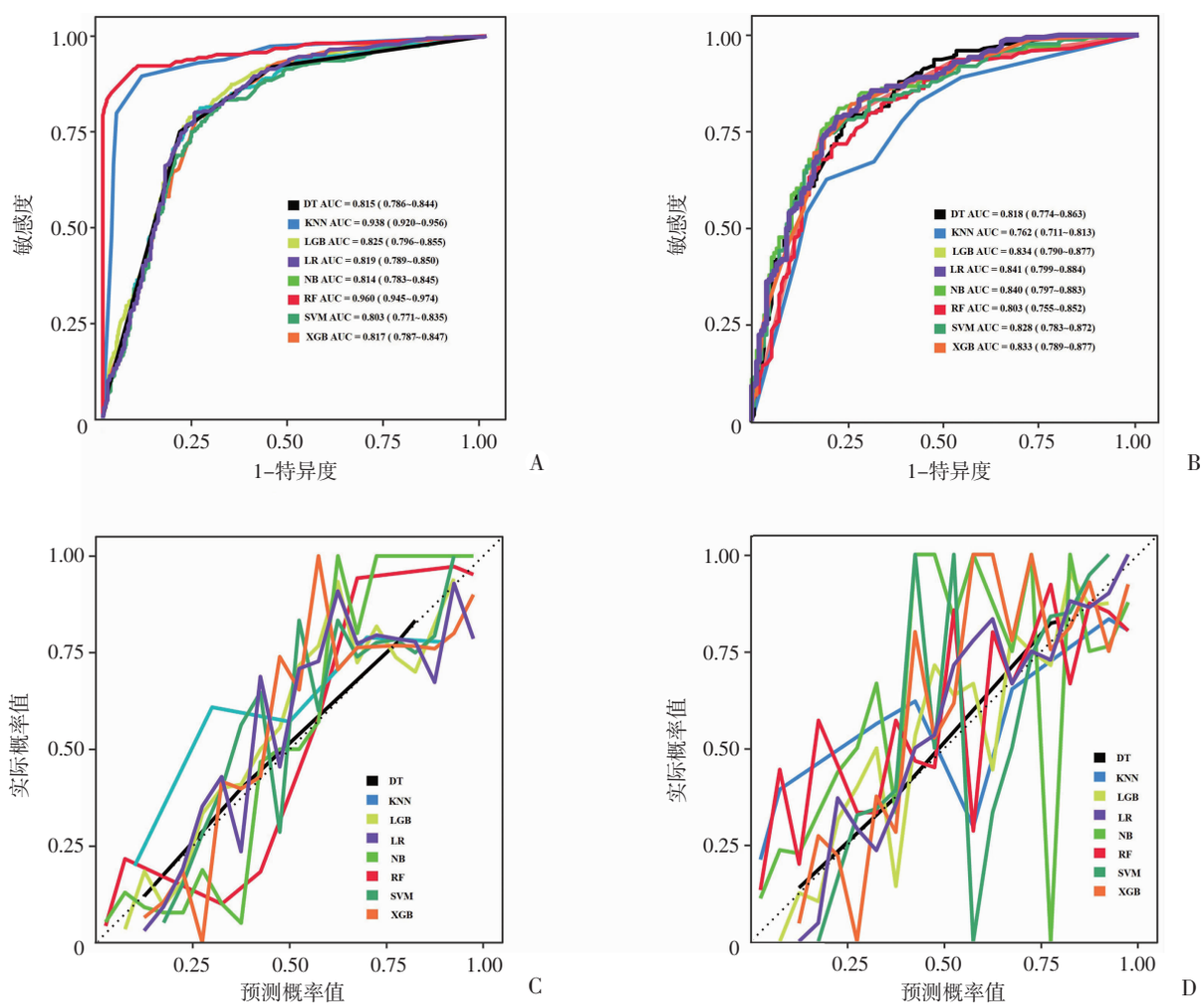
以往由于缺乏对HADAS的认识,一般认为从高原低氧环境返回平原常氧环境后,机体缺氧状态得到改善,生理功能与结构也随之恢复正常[13],但事实可能并非如此。周其全等[7]对3 011名由高原返回低海拔地区的部队官兵与铁路职工的调查研究发现,高达84.36%人员出现HADAS症状,部分人员在平原连续居住2年后仍会出现低蛋白血症心动徐缓、心功能下降、肺动脉高压等症状,返回平原20年后血小板水平才与平原对照组无显著性差异,需要相对较长时间才能恢复,严重者不得不重返高原。本研究中也发现,随返回平原时间延长,脱适应症发生率未必降低,该变量未纳入到LR模型当中。丁魁等[14]的研究表明,高原驻训士兵返回平原后HADAS与训练倦怠存在一

**表3 Logistic回归分析筛选特性的影响因素**

| 变量 | *β* | 标准误 | *OR* | 95%*CI* | *Z* | *P* |
|---|---|---|---|---|---|---|
| 年龄 | 0.043 | 0.018 | 1.044 | 1.008~1.081 | 2.424 | 0.015 |
| CMS评分 | 0.240 | 0.034 | 1.271 | 1.190~1.362 | 6.988 | <0.001 |
| 高原居住时间 | 0.321 | 0.122 | 1.379 | 1.083~1.752 | 2.631 | 0.009 |

**表4 内部测试集中8种机器学习模型的预测效能评估结果**

| 预测模型 | AUC(95%*CI*) | Brier评分(95%*CI*) | 准确率 | 灵敏度 | 特异度 | 约登指数 | 阳性预测值 | 阴性预测值 | F1得分 |
|---|---|---|---|---|---|---|---|---|---|
| LR | 0.841(0.799~0.884) | 0.163(0.143~0.187) | 0.779 | 0.801 | 0.759 | 0.560 | 0.758 | 0.802 | 0.801 |
| NB | 0.840(0.797~0.883) | 0.181(0.151~0.225) | 0.778 | 0.828 | 0.723 | 0.550 | 0.770 | 0.789 | 0.798 |
| LGB | 0.834(0.790~0.877) | 0.169(0.147~0.191) | 0.769 | 0.782 | 0.755 | 0.536 | 0.782 | 0.755 | 0.782 |
| XGB | 0.833(0.789~0.877) | 0.166(0.147~0.191) | 0.781 | 0.810 | 0.748 | 0.559 | 0.783 | 0.779 | 0.797 |
| SVM | 0.828(0.783~0.872) | 0.172(0.151~0.193) | 0.736 | 0.851 | 0.607 | 0.457 | 0.708 | 0.783 | 0.773 |
| DT | 0.818(0.774~0.863) | 0.162(0.140~0.185) | 0.784 | 0.824 | 0.747 | 0.571 | 0.753 | 0.819 | 0.787 |
| RF | 0.803(0.755~0.852) | 0.190(0.161~0.223) | 0.748 | 0.793 | 0.697 | 0.490 | 0.746 | 0.750 | 0.769 |
| KNN | 0.762(0.711~0.813) | 0.231(0.194~0.272) | 0.678 | 0.672 | 0.684 | 0.356 | 0.705 | 0.650 | 0.688 |




A:训练集中机器学习模型的ROC曲线;B:内部测试集中机器学习模型的ROC曲线;C:训练集中机器学习模型的校准曲线;D:内部测试集中机器学习模型的校准曲线

**图3 训练集和内部测试集中8种机器学习模型的ROC曲线及校准曲线图**

定程度的关联性,强调了尽早发现、诊断、干预HADAS对返回平原后训练成绩提升和战斗力维持的重要意义。随着到高原旅游、执行任务、支援西部建设的人群不断增多,这些人返回平原后的出现的健康问题也变得更需要关注。通过在下山前准确预测HADAS发病风险,有助于引起HADAS易感者对自身健康状况的重视,从而进行早期干预,减轻下山后的脱适应症状,缩短恢复时间。

### 3.2　预测模型的构建、验证及意义

近年来,机器学习已成为预测模型构建的常用手段。在高原医学领域,多种机器学习方法已应用到急性高原病的预测中,取得了较好效果[15-16]。但在HADAS方面,尚未见到相关预测模型的报道。面对高维、复杂和非线性的大数据时,机器学习模型常具有比经典模型更好的性能,但仍存在模型可解释性不强、可视化效果不佳等特点[17]。不同机器学习方法的适用场景、优缺点也不尽相同。如在本研究中,虽然RF、KNN在训练集中的AUC>0.9,Brier评分与F1得分也均较好,但在内部测试集中AUC<0.82,Brier评分和F1得分也不如LR,说明这些机器学习模型存在过拟合现象。另外,在本研究中,多数机器学习模型的校准曲线也不如LR。通过比较LR与其他7个常见机器学习模型的性能,发现仅针对本研究的人群数据而言,传统的回归模型总体优于其他机器学习模型。一些文献也相继指出,某些场景下机器学习并不优于传统模型[18]。当模型性能相近时,根据奥卡姆剃刀原则,更倾向于选择简单的模型[19]。在可解释性方面,基于LR模型构建的列线图可直接用于HADAS的预测,据此开发的网页计算器直观性和实用性更强。

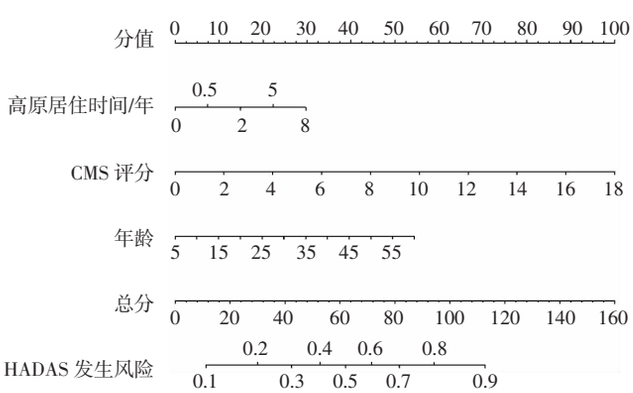



**图4　LR模型的HADAS发生风险预测列线图**

本研究通过全程跟踪返回平原的高原移居者队列,并将其作为外部验证集对前期建立的LR模型进行评价发现,该模型在外部验证集ROC的AUC>0.85,模型区分度优秀,能够很好的将有HADAS的人群与无HADAS人群区分开。但是该模型校准曲线的校准截距$a$<0,表明此模型会高估外部测试组人群的HADAS发生概率,可能是由如下因素导致:①经调查得知,该群体劳动强度较低,保障条件好,能够定期吸氧,可能是其返回平原后HADAS发病率低的部分原因;②该人群仅在高原居住9个月,高原居住时间、CMS评分、HADAS发生率与建模组相比均较低,差异具有统计学意义;研究表明,建模组与外部测试组差异越大,校准错误的可能性也越大[20],该模型在高原居住时间较短人群中应用时的外推性有所降低;③外部验证组内人群基线资料相似性高,在高原居住时间、高原居住地海拔、返回平原时间均相同的情况下,起关键作用的预测因子仅有年龄和CMS评分,可能对预测能力产生一定影响。本研究认为,该模型在不同场景下可能需要通过调整模型截距等方式进行基线

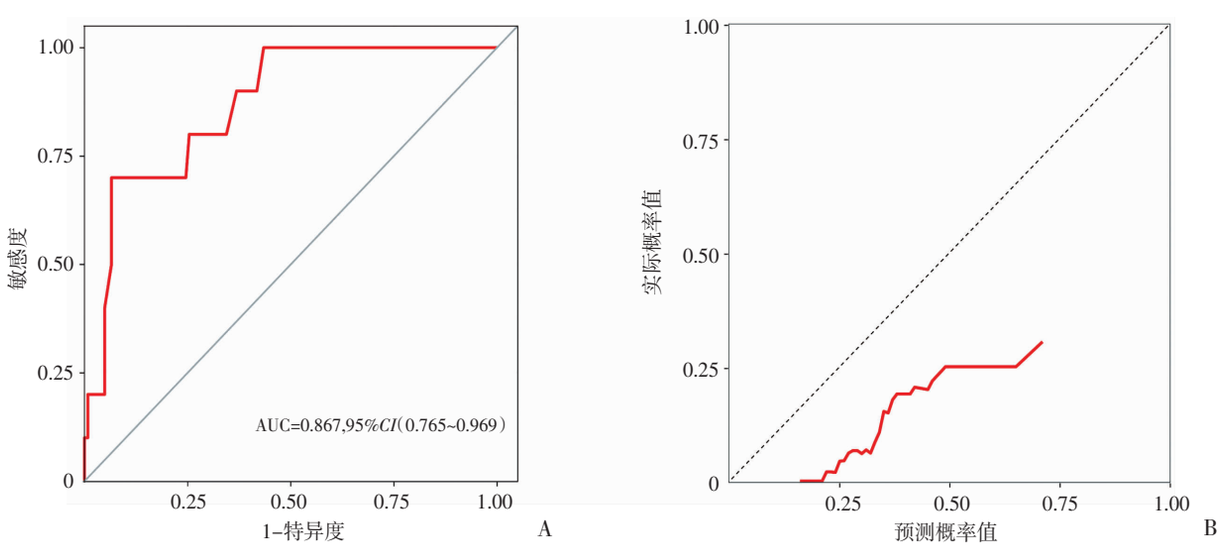



**图5　LR模型在外部验证集的ROC曲线(A)和校准曲线(B)**

校准[17]，并采用独立的外部验证对模型的可移植性进行进一步评价[19]。

本研究是第1个基于超过1 000例样本开发并基于前瞻性队列开展外部验证的HADAS发生风险预测模型，构建模型所需的预测因子数量少，指标获取方便，计算简单的同时能保证较高预测效能，可解释性强，使该模型更方便应用于个体返回平原前的发病风险自我评估，在协助HADAS早发现早诊断早治疗方面具有重要意义。

### 3.3 HADAS的风险因素

本研究发现，CMS评分、年龄、高原居住时间为脱适应症发生的有效预测变量。分析最终被纳入的3个影响因素，对可能的机制做出如下阐述。①脱适应症与高原病的关系十分密切，对高原习服不良的人群而言，返回平原后脱适应症状也更明显。其一，这与人体对环境的适应能力有关，CMS症状和评分能一定程度上反映高原习服状况，已有报道指出了CMS与脱适应症的相关性[21]。环境适应能力弱的个体，在高原时无法习服，返回平原后也难以适应富氧环境，易患CMS和HADAS；其二，这与CMS的恢复速度有关，CMS症状重的人，返回平原后需要的恢复时间也越长，CMS与HADAS相互作用，机体不适症状也更严重。②年龄是也本研究中的重要预测指标，有报道指出，老年人返回平原后比年轻人心脏损害更重[22]，脱适应症发生率更高[23]，本研究结论也支持此观点。③从列线图可以看出，高原居住时间虽然也是HADAS的危险因素，但是在3个预测因子中的重要性最低。④以往认为，高原居住地海拔越高，脱适应症发生率也越高[24]，但在本研究中，海拔并没有被纳入到预测变量中，曾以萍等[25]的流行病学调查及周发权等[2]荟萃分析研究也提示脱适应症与海拔高度无相关关系。海拔对脱适应症的影响仍需进一步研究。

本研究存在一些无法避免的局限：①建模组中一些受试者返回平原时间较久，可能对CMS症状存在回忆偏倚；②LR模型在内部测试集中效果较好，但在高原居住时间较短的外部验证集中存在高估HADAS发生概率的现象。将在后续研究中继续收集人群数据，用于丰富建模组和外部验证组数据，更新LR模型并对HADAS的预测因子做进一步探索。

综上所述，本研究通过对从高原返回平原的受试者进行HADAS调查，找到了CMS评分、高原居住时间、年龄这3个有效预测变量，通过比较多种不同机器学习方法，构建并验证了预测性能令人满意的HADAS发病风险预测模型，基于LR模型构建的列线图方便使用，可以在高原移居者返回平原前通过列线图计算高原居住者HADAS的发生风险，辅助临床医生决策，找出需要干预的患者。

**利益冲突** 本文所有作者声明不存在利益冲突

## 参考文献：

[1] 高钰琪. 高原军事医学[M]. 重庆：重庆出版社，2005.
GAO Y Q. High altitude military medicine [M]. Chongqing: Chongqing Press, 2005.

[2] 周发权，陈师，刘若鸿，等. 高原脱适应症及其临床症状检出率的荟萃分析[J]. 军事医学，2020，44(5)：331-337.
ZHOU F Q, CHEN S, LIU R H, *et al*. Detection rate of high altitude de-adaptation and clinical symptoms: a meta-analysis [J]. Mil Med Sci, 2020, 44(5): 331-337.

[3] 范勇，周其全. 高原脱适应研究进展[J]. 解放军预防医学杂志，2012，30(3)：227-230.
FAN Y, ZHOU Q Q. Research progress of altitude maladjustment[J]. China Ind Econ, 2012, 30(3): 227-230.

[4] 刘爽，龙贤哲，罗勇军. 高原脱适应症诊断及防治研究进展[J]. 人民军医，2017，60(9)：941-943.
LIU S, LONG X Z, LUO Y J. Research progress on diagnosis, prevention and treatment of plateau de-indications [J]. People Mil Surg, 2017, 60(9): 941-943.

[5] 张卫花，姜钊. 高原脱适应症的危害及防治措施研究进展[J]. 中国疗养医学，2021，30(3)：242-245. DOI：10.13517/j.cnki.ccm.2021.03.006
ZHANG W H, JIANG Z. Advances in research on the hazards and prevention and treatment of plateau de-adaptation [J]. Chin J Convalesc Med, 2021, 30 (3) : 242-245. DOI: 10.13517/j.cnki.ccm.2021.03.006

[6] 伍亚舟，陈锡程，易东. 人工智能在临床领域的研究进展及前景展望[J]. 陆军军医大学学报，2022，44(1)：89-102. DOI：10.16016/j.1000-5404.202112092.
WU Y Z, CHEN X C, YI D. Advances and perspective of artificial intelligence in clinical area [J]. J Army Med Univ, 2022,44(1):89-102.DOI:10.16016/j.1000-5404.202112092.

[7] 周其全，杨生岳，袁振才，等. 高原移居人群返回平原后高原脱适应症的诊断标准：多中心、随机对照研究[J]. 解放军医学杂志，2012，37(2)：146-155.
ZHOU Q Q, YANG S Y, YUAN Z C, *et al*. A research in diagnostic criteria of high altitude de-adaptation for plateau migrants returning to the Plains: a multicenter, randomized controlled trial[J]. Med J Chin PLA, 2012, 37(2): 146-155.

[8] 国际高原医学会慢性高原病专家小组. 第六届国际高原医学和低氧生理学术大会颁布 慢性高原病青海诊断标准[J]. 青海医学院学报，2005，26(1)：3-5.
Expert Group of Chronic High Altitude Diseases of International Society for High Altitude Medicine. The 6th International Conference on Plateau Medicine and Hypoxia Physiology promulgated the diagnostic criteria of chronic altitude sickness in Qinghai [J]. J Qinghai Med Coll, 2005, 26(1): 3-5.

[9] 阳盛洪，王引虎，杨海军，等.高压氧治疗海拔5000m以上地区移居人群脱习服的临床疗效观察[J].西北国防医学杂志，2013，34(3)：205-207. DOI：10.16021/j.cnki.1007-8622.2013.03.007.
YANG S H，WANG Y H，YANG H J，*et al*. Clinical observation of hyperbaric oxygen treatment for de-acclimatization to high altitude for people above 5000 m altitude[J]. Med J Natl Defending Forces Northwest China，2013，34(3)：205-207. DOI：10.16021/j.cnki.1007-8622.2013.03.007.

[10] COLLINS G S，REITSMA J B，ALTMAN D G，*et al*. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD)：the TRIPOD statement [J]. BJOG，2015，122(3)：434-443. DOI：10.1111/1471-0528.13244.

[11] LUO W，PHUNG D，TRAN T，*et al*. Guidelines for developing and reporting machine learning predictive models in biomedical research：a multidisciplinary view[J]. J Med Internet Res，2016，18(12)：e323. DOI：10.2196/jmir.5870.

[12] 武松.预测模型实战：基于R、SPSS和Stata(新时代·技术新未来)[M].北京：清华大学出版社，2023.
WU S. Predictive model practice with R，SPSS and Stata (New era·technology and new future) [M]. Beijing：Tsinghua University Press，2023.

[13] HEATH D，DONALD D E，EDWARDS J E. Pulmonary vascular changes in a dog after aortopulmonary anastomosis for four years [J]. Br Heart J，1959，21(2)：187-196. DOI：10.1136/hrt.21.2.187.

[14] 丁魁，李权超，王燕，等.高原驻训士兵返回平原后高原脱适应症与训练倦怠的关系研究[J].军事医学，2016，40(8)：661-664，694. DOI：10.7644/j.issn.1674-9960.2016.08.010.
DING K，LI Q C，WANG Y，*et al*. Relationships between high altitude de-adaptation syndrome and training burnout of plateau soldiers after returning to the plain[J]. Mil Med Sci，2016，40(8)：661-664，694. DOI：10.7644/j.issn.1674-9960.2016.08.010.

[15] WEI C Y，CHEN P N，LIN S S，*et al*. Using machine learning to determine the correlation between physiological and environmental parameters and the induction of acute mountain sickness [J]. BMC Bioinformatics，2022，22 (Suppl 5)：628. DOI：10.1186/s12859-022-04749-0.

[16] YANG M，WU Y，YANG X B，*et al*. Establishing a prediction model of severe acute mountain sickness using machine learning of support vector machine recursive feature elimination[J]. Sci Rep，2023，13(1)：4633. DOI：10.1038/s41598-023-31797-0.

[17] 王胜锋，王俊峰，谷鸿秋，等.临床预测模型方法与应用[M].北京：北京大学医学出版社，2024.
WANG S F，WANG J F，GU H Q，*et al*. Clinical predicting models methods and application [M]. Beijing，Peking University Press，2024.

[18] HONG C，PENCINA M J，WOJDYLA D M，*et al*. Predictive accuracy of stroke risk prediction models across black and white race，sex，and age groups [J]. JAMA，2023，329(4)：306-317. DOI：10.1001/jama.2022.24683.

[19] EFTHIMIOU O，SEO M，CHALKOU K，*et al*. Developing clinical prediction models：a step-by-step guide [J]. BMJ，2024，386：e078276. DOI：10.1136/bmj-2023-078276.

[20] RILEY R D，ARCHER L，SNELL K I E，*et al*. Evaluation of clinical prediction models (part 2)：how to undertake an external test study [J]. BMJ，2024，384：e074820. DOI：10.1136/bmj-2023-074820.[PubMed]

[21] 谢佳新，吴玉，董华平，等.高原驻训人员高原脱适应症发病特征与影响因素调查分析[J].西南国防医药，2021，31(6)：556-560.
XIE J X，WU Y，DONG H P，*et al*. Investigation and analysis on epidemiological features and influencing factors of HADAS among encamping training personnel at high altitude[J]. Med J Natl Defending Forces Southwest China，2021，31(6)：556-560.

[22] 李政波，张进，王雪.高原低氧环境短期暴露对官兵心脏的影响[J].西北国防医学杂志，2019，40(3)：174-178. DOI：10.16021/j.cnki.1007-8622.2019.03.009.
LI Z B，ZHANG J，WANG X. Effect of short-term exposure to plateau hypoxic environment on the heart of officers and soldiers[J]. Med J Natl Defending Forces Northwest China，2019，40(3)：174-178. DOI：10.16021/j.cnki.1007-8622.2019.03.009.

[23] 丰丽英，朱前勇，王保健，等.医疗分队高原驻训后脱适应症发生情况调查[J].人民军医，2017，60(2)：114-117.
FENG L Y，ZHU Q Y，WANG B J，*et al*. Investigation on the occurrence of out-of-indications of medical teams after training at high altitude[J]. People Mil Surg，2017，60(2)：114-117.

[24] 魏征华，资云海，董志伟，等.高原移居者返回平原的高原脱适应症发生率比较[J].解放军医学杂志，2012，37(12)：1151-1154.
WEI Z H，ZI Y H，DONG Z W，*et al*. Analysis on the deadaptation symptoms of plain region migrants back from different high altitude regions[J]. Med J Chin PLA，2012，37(12)：1151-1154.

[25] 曾以萍，周其全，邬云红，等.久居高原者返回平原后脱适应症状分析[J].西藏医药，2014，35(1)：46-48.
ZENG Y P，ZHOU Q Q，WU Y H，*et al*. Analysis of maladjustment symptoms of people who have lived in plateau for a long time after returning to plain[J]. Tibet Med，2014，35(1)：46-48.

(**收稿**：2024-08-26；**修回**：2024-09-13)
(编辑 栾 嘉)