

中文引用格式:胡啸峰,黄玲. 基于机器学习的户外作业警务人员热应激预测模型[J]. 中国安全科学学报,2024,34(11):220-228.

英文引用格式:HU Xiaofeng, HUANG Ling. Heat stress prediction model for outdoor policeman based on machine learning[J]. China Safety Science Journal, 2024,34(11):220-228.

基于机器学习的户外作业警务人员热应激 预测模型*

胡啸峰^{1,2,3}副教授,黄玲¹

(1 中国人民公安大学 信息网络安全学院,北京 100038; 2 安全防范技术与风险评估公安部重点实验室,北京 100038; 3 中国矿业大学(北京) 应急救援工效与防护研究院,北京 100083)

中图分类号:X968

文献标志码:A

DOI: 10.16265/j.cnki.issn1003-3033.2024.11.0171

基金项目:国家自然科学基金资助(72174203)。

【摘要】 为解决高温环境下户外作业警务人员热应激风险预测问题,构建不同环境工况、劳动强度和服装等条件下的警务人员核心温度监测试验数据集,提取身高、体质量、年龄、性别、体脂率、身体活动比率(PAR)、服装热阻(CI)、环境温度、相对湿度为特征,基于K近邻(KNN)、随机森林(RF)、梯度提升决策树(GBDT)等多种机器学习方法,建立户外作业警务人员核心温度及热应激风险预测模型,并进行验证。结果表明:对于高温环境下户外作业警务人员的核心温度预测模型,KNN、RF和GBDT的拟合优度 R^2 均超过0.9,在误差方面,KNN模型的预测误差最小,均方根误差(RMSE)为 0.053°C ;对于高温环境下户外作业警务人员热应激预测模型,RF、GBDT和KNN模型的预测性能明显优于其他模型。

【关键词】 机器学习; 户外作业; 警务人员; 热应激; 核心温度; 高温环境

Heat stress prediction model for outdoor policeman based on machine learning

HU Xiaofeng^{1,2,3}, HUANG Ling¹

(1 School of Information and Network Security, People's Public Security University of China, Beijing 100038, China; 2 Key Laboratory of Security Prevention and Risk Assessment, Beijing 100038, China; 3 Institute for Emergency Rescue Ergonomics and Protection, China University of Mining & Technology-Beijing, Beijing 100083, China)

Abstract: To address the issue of predicting heat stress risks for police officers engaged in outdoor operations under high-temperature conditions, a test dataset for monitoring core temperature of police officers under different environmental working conditions, levels of physical exertion and clothing scenarios was constructed. First, features such as height, weight, age, gender, body fat percentage, physical activity ratio (PAR), clothing insulation (CI), environmental temperature and relative humidity were extracted. Then, machine learning methods, including K-nearest neighbors (KNN), random forest (RF) and gradient boosting decision trees (GBDT), were used to establish predictive models of core temperature and heat stress risk for outdoor police officers. These models were subsequently validated. The results indicate that for the predictive model of core temperature for outdoor police officers working in high-

temperature environments, the goodness-of-fit measure R^2 exceeds 0.9 for KNN, RF and GBDT. In terms of error, the KNN model has the smallest prediction error, with a root mean square error (RMSE) of 0.053 °C. For the heat stress prediction model for police officers engaged in outdoor operations under high-temperature conditions, the predictive performance of RF, GBDT and KNN models is significantly better than that of other models.

Keywords: machine learning; outdoor operations; police officers; heat stress; core temperature; high temperature environment

0 引言

高温是威胁人类健康的职业危害因素之一^[1]。为保障城市治安与交通安全,大量警务人员需长时间暴露于户外高温环境中,加之许多公安一线民警长期承受超负荷工作,无法得到适时休息,极易造成人体产热和散热不平衡,导致体温上升,进而出现热应激现象,引发严重后果^[2]。因此,研究高温环境中户外作业警务人员热应激风险预测,符合保障警务人员身体健康和生命安全的现实要求。

当前,对热应激风险的预测主要基于人体热生理模型,如 MALCHAIRE 等^[3]提出的预测热应激模型综合考虑了外界环境因素、人体生理参数和服装热阻(Clothing Insulation, CI)、湿阻,可预测人体核心温度和出汗量,并基于上述变量,提出热应激指标,进行风险预测。基于热生理模型的预测模型被广泛应用于特殊职业人群(如煤矿工人、电网工人等^[4-6])的热应激风险预测。然而,在使用热生理模型进行预测时,简化人体结构、生理参数及热反应过程极易导致预测误差。随着人工智能技术的发展,部分研究开始使用机器学习技术来预测人体的热舒适性,将人体特征和环境特征作为基本特征来预测人体的热舒适度,进而预测热应激风险。人体热舒适性是人对热环境满意度的主观感受,受物理、心理及生理多重因素的影响,其中,核心温度是判断热舒适性的一个重要指标,也是热应激现象的重要判断依据^[7],世界卫生组织和美国国家职业安全卫生研究所建议,在高温、高湿环境下作业的工人,核心温度不能超过 38 °C,这一极限值受到热应激研究领域的普遍认可^[8]。部分学者采用开源数据构建机器学习模型,预测人体的热舒适水平,如 HAN Xinge 等^[9]将热舒适水平设置为标签,训练了一系列基于环境和人体特征的监督学习模型,通过比较发现深度森林模型的性能最优;CHAUDHURI 等^[10]使用公开数据库,以环境变量和个体生理参数作为输入特征来预测人体的热舒适水平,预测准确率为 73.1%

~81.2%。此外,部分学者基于人工气候模拟试验数据构建机器学习模型,预测人体的热舒适性水平。如 MORRESI 等^[11]基于试验数据,考虑生理参数和环境参数,基于机器学习算法预测人体热舒适性,平均准确率为 92.2%;LIU Kuixing 等^[12]采用局部皮肤温度和热负荷作为输入特征,建立基于支持向量机(Support Vector Machine, SVM)的室外环境热舒适评价模型,并通过比较不同输入组合的预测精度,确定预测室外热舒适水平的关键特征;MORISHIMA 等^[13]提出基于机器学习的皮肤温度预测方法,预测模型的均方根误差(Root Mean Square Error, RMSE)为 0.315 °C,百分比 RMSE 为 0.9%。

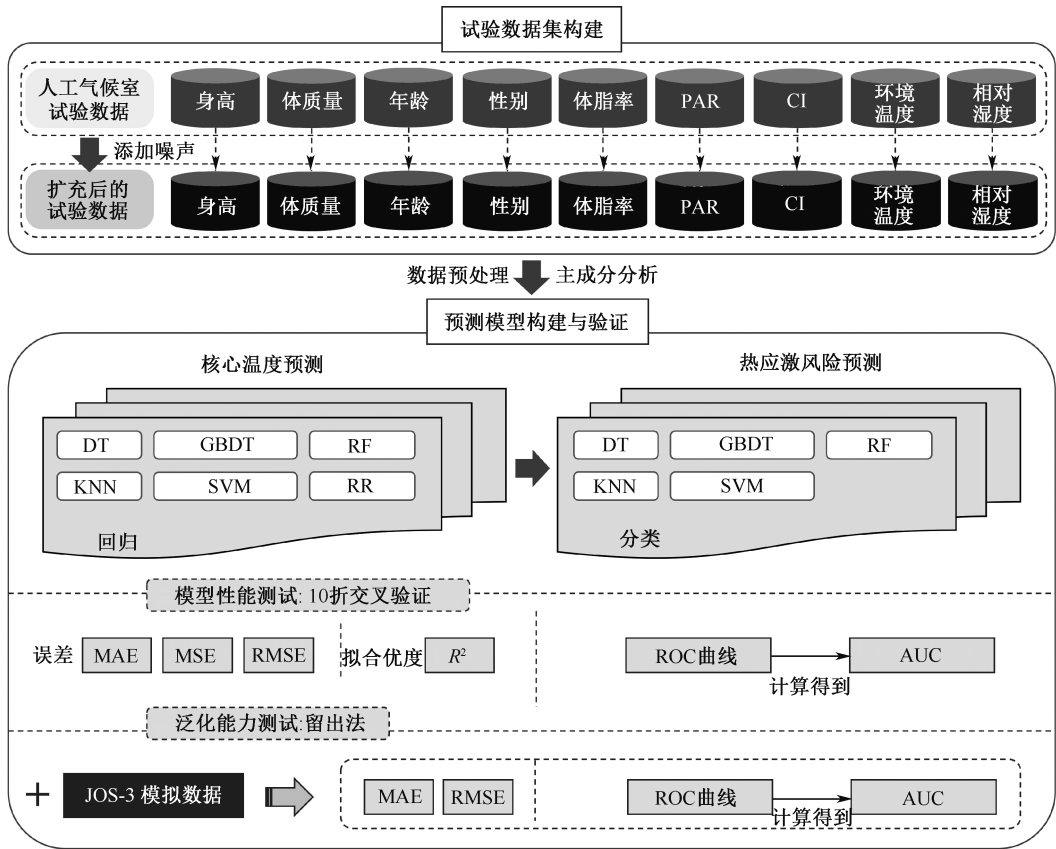
当前,大部分学者主要致力于改进热生理模型,仅部分学者采用基于机器学习的方法预测人体的热舒适水平。热舒适性是一种人体主观感觉,难以准确代表人体机制反应;而热应激是人体受到热负荷引起的生理反应,更能代表人体受到的热风险,通常通过试验获得实际数据,更适合采用机器学习方法进行预测。鉴于此,笔者拟通过人工气候室模拟高温环境,构建专门针对户外作业警务人员的核心温度试验数据集,综合考虑不同环境工况、劳动强度和服装等条件,运用多种机器学习算法建立户外作业警务人员热应激预测模型,以期为暴露在高温环境中作业的警务人员提供预警技术支持,从而降低警务人员的热应激风险和事故发生率。

1 热应激预测模型构建

图 1 为高温环境下户外作业警务人员热应激预测模型的技术路线,主要包括试验数据集构建、机器学习模型构建及验证 2 部分。

1.1 热应激预测模型试验数据集构建

热环境下不同个体的生理参数存在一定的差异性^[14]。KARJALAINEN 等^[15-16]通过试验研究发现,男性和女性存在明显的热舒适性差异。FERRARO 等^[17]通过调查和试验,研究了年龄和性别对人体热舒适的潜在影响,指出性别和年龄是评估热舒适性



注:联合系统体温调节 (Joint System Thermoregulation, JOS-3); 身体活动比率 (Physical Activity Ratio, PAR); 决策树 (Decision Trees, DT); 梯度提升决策树 (Gradient Boosting Decision Trees, GBDT); 随机森林 (Random Forest, RF); K 近邻 (K-Nearest Neighbors, KNN); 岭回归 (Ridge Regression, RR); 受试者工作特征 (Receiver Operating Characteristic, ROC); 曲线下的面积 (Area Under Curve, AUC); 均方误差 (Mean Square Error, MSE); 平均绝对误差 (Mean Absolute Error, MAE)

图 1 高温环境下户外作业警务人员热应激预测模型

Fig. 1 Framework of heat stress prediction model for outdoor police officers in high temperature environment

时必须考虑的因素。在人体热生理模型建立过程中,普遍考虑身高、体质量、性别、脂肪含量以及体型等因素^[18-20]。文中选择身高、体质量、年龄、性别、体脂率、PAR、CI、环境温度、相对湿度共 9 个指标,作为核心温度及热应激预测的特征。

针对上述特征,利用人工气候室模拟不同环境温度(34、36、38、40 °C)和相对湿度(30%、60%),共设置 20 组试验,工况见表 1。人工气候室的尺寸为 5 m×3 m×2.9 m(长×宽×高)。通过控制系统,调节温度、湿度等环境参数,模拟多种高温环境工况。其中,该人工气候室的温度调节区间为-25~50 °C,控制稳定区间在±1 °C 范围内,湿度调节区间为 20%~90%。参照国际标准《热环境人体工程学:代谢率的测定》(ISO 8996—2021),对跑步机设定不同的步行速度,令被试于跑步机上完成步行运动,从而分别模拟轻度(0 km/h)、中度(3.5 km/h)和重度(5.5 km/h)作业强度^[21]。

表 1 试验工况

Table 1 Experimental conditions

组别	温度/°C	相对湿度/%	作业强度	服装
1	34	60	轻度	警服
2	36	60	轻度	警服
3	38	60	轻度	警服
4	40	60	轻度	警服
5	34	60	中度	警服
6	36	60	中度	警服
7	38	60	中度	警服
8	40	60	中度	警服
9	34	60	重度	警服
10	36	60	重度	警服
11	38	60	重度	警服
12	40	60	重度	警服
13	38	60	轻度	便服
14	34	60	中度	便服
15	36	60	中度	便服
16	38	60	中度	便服
17	40	60	中度	便服

续表 1

组别	温度/℃	相对湿度/%	作业强度	服装
18	38	60	重度	便服
19	36	30	轻度	警服
20	38	30	轻度	警服

选取无脑部及与心脏有关的疾病、心理疾病及其他严重疾病的 20~28 周岁的 9 名男性和 6 名女性警务人员作为研究对象,每位被试人员统一穿着夏季警务执勤服(该服装为户外警务作业人员标准着装),上衣热阻为 $0.039 \text{ m}^2 \cdot \text{K/W}$ 、长裤热阻为

表 2 数据样例

Table 2 Data sample

特征	性别	身高/m	体质量/kg	体脂率	年龄/岁	相对湿度/%	温度/℃	PAR	CI/ ($\text{m}^2 \cdot \text{K} \cdot \text{W}^{-1}$)					
	男	1.8	82	25.3	20	60	34	1.4	0.077 5					
标签	核心温度/℃													热应激 风险
	N_1	N_2	N_3	N_4	N_5	N_6	N_7	N_8	N_9	N_{10}	N_{11}	N_{12}	N_{13}	
	37.59	37.59	37.54	37.50	37.44	37.37	37.33	37.29	37.28	37.27	37.27	37.30	37.32	

利用 python3.6 中 sklearn 模型库的 StandardScaler 模块标准化处理特征及标签值^[22],经过处理后数据符合标准正态分布,即均值为 0,标准差为 1,如下式:

$$X' = \frac{x - \mu}{\sigma} \quad (1)$$

式中: X' 为标准化后的数据; x 为原始数据; μ 为每种特征的均值; σ 为标准差。

对于性别这一特征,男性标记为 1,女性标记为 0。各节点($N_1 \sim N_{13}$)核心温度数据(表 2)作为标签值,用于构建核心温度预测模型;热应激状态作为标签值,用于建立热应激预测模型,参考 13 个时间节点的全部核心温度数值,当全部低于 $38 \text{ }^\circ\text{C}$ 时,热应激状态标记为 0,当 13 个时间节点的核心温度数值中有一个超过阈值($38 \text{ }^\circ\text{C}$)时,则标记为 1。

为充分训练机器学习模型,提高模型的鲁棒性和泛化能力,利用 python3.6 中 NumPy 库中的 np.random.normal() 函数定义了一个均值为 0、标准差为 0.01 的高斯随机分布噪声,添加至原始样本,扩充原始数据集,共得到 1 440 条数据,用于机器学习预测模型的训练和测试。

1.2 热应激预测模型的构建与验证

首先构建高温环境下户外作业警务人员的核心温度(回归)预测模型(图 1)。以选定的 9 个指标作为特征,以 13 个时间节点测量得到的核心温度值作为标签,分别建立基于 SVM^[23]、DT^[24]、GBDT^[25]、

$0.037 \text{ m}^2 \cdot \text{K/W}$, 对照组人员穿着统一规格的短袖便衣和短裤,热阻均为 $0.012 \text{ m}^2 \cdot \text{K/W}$ 。

试验期间,被试每运动 20 min 休息 10 min,如此循环 2 个周期,总时长为 60 min。试验过程中持续测量人体核心温度,每 5 min 采集一次核心温度数据(每个工况下各采集 13 个时间节点的核心温度数据)。通过试验共得到 144 条原始数据,每条数据包含上述 9 个特征以及对应的共 13 个时间节点的核心温度(标签)数据。数据样例见表 2。

RF^[26]、KNN^[27] 和 RR^[28] 共 6 种回归预测模型,预测核心温度。从预测的起点开始,每间隔 5 min 设置一个时间节点,最终核心温度预测结果由 13 个节点的预测值构成标签向量。预测值(标签)与特征之间的函数关系如下:

$$t_c(i) = f(z_s, z_w, z_h, z_f, z_{age}, z_{ta}, z_{PAR}, z_{CI}) \quad (2)$$

式中: t_c 为核心温度值,℃; i 为时间节点编号($i=0, 1, \dots, 13$); z_s 为性别; z_w 为体质量,kg; z_h 为身高,m; z_f 为体脂率; z_{age} 为年龄; z_{ta} 为环境温度,℃; z_{PAR} 为 PAR; z_{CI} 为 CI, $\text{m}^2 \cdot \text{K/W}$ 。

使用 10 折交叉验证,将全部样本划分成 10 个大小相等的样本子集,依次遍历这 10 个子集,每次把当前子集作为测试集(10%),其余所有样本作为训练集(90%),进行模型的训练和测试。并将 10 个测试集对应的评价指标的平均值作为最终的模型评价结果。

采用 MSE、RMSE、MAE 评价核心温度预测模型的误差水平;采用拟合优度 R^2 计算上述回归模型中因变量可由自变量解释的百分比。具体公式如下:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \quad (3)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (4)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (5)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (6)$$

式中: n 为样本数量; y_i 为实际值; \hat{y}_i 为预测值。

在准确预测核心温度的前提下,构建高温环境下户外作业警务人员热应激预测模型(图1)。表2中,设置热应激状态标签,以核心温度 $38\text{ }^\circ\text{C}$ 为阈值,超过 $38\text{ }^\circ\text{C}$ 的状态标记为1,代表热应激风险较高;低于 $38\text{ }^\circ\text{C}$ 的状态标记为0,代表热应激风险较低。设置该标签时,参考13个时间节点的全部核心温度数值,当全部低于 $38\text{ }^\circ\text{C}$ 时,将该标签标记为0;13个时间节点的核心温度数值中有一个超过阈值($38\text{ }^\circ\text{C}$)时,则标记为1。该模型采用的特征值与核心温度回归预测模型所使用的9个特征相同。该模型采用的特征值与核心温度回归预测模型所使用的9个特征相同。分别采用DT^[24]、GBDT^[25]、RF^[26]、KNN^[27]和SVM^[23,29]算法,建立热应激状态(分类)预测模型。

采用ROC曲线^[30]评价热应激风险预测模型的性能,ROC空间以假正例率(False Positive Rate, FPR)作为横轴,真阳性率(True Positive Rate, TPR)作为纵轴,通过计算ROC的AUC,评估各个所选分类模型对于样本的分类能力。FPR、TPR、AUC的计算方法如下。

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}} \quad (7)$$

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

$$\text{AUC} = \frac{1}{2} \left(1 + \frac{\text{TP}}{\text{TP} + \text{FN}} \right) - \frac{\text{FP}}{\text{FP} + \text{TN}} \quad (9)$$

式中:TP为正类样本预测结果为“正”的个数;FP为负类样本预测结果为“正”的个数;TN为负类样本预测结果为“负”的个数;FN为正类样本预测结果为“负”的个数。AUC越接近1,表示分类器的性能越好。

1.3 基于模拟数据与试验数据的混合模型构建

试验所得数据量为144条,在添加噪声的数据集上测试得到的模型性能无法表征其泛化能力。为解决上述问题,基于JOS-3模型^[20],随机生成与试验数据特征一致的648条模拟数据。JOS-3模型由83个节点组成,采用后向差分法计算人体生理反应和体温。JOS-3的基本物理性质由Stolwijk模型^[31]推导,在JOS-2的基础上安装棕色脂肪组织活动、衰老效应和短波太阳辐射对皮肤的热增益,考虑到瞬态和非均匀热环境中的个人特征。随机抽取上述648条JOS-3模拟数据中的576条数据,并随机抽取50%人工气候室试验数据(72条)共同训练得到一

个基于混合数据的热应激预测模型(命名为J-1),以未被抽取的50%的人工气候室试验数据(72条)作为测试集,训练集与测试集的比例为9:1。作为对照,利用全部648条JOS-3模拟数据单独训练一个模型(命名为J-2),并使用与J-1同样的测试集进行测试。上述随机抽取试验数据的过程反复执行10次,即得到10组误差值(RMSE与MAE),取其均值作为最终的泛化能力评价结果。而对于热应激风险预测模型(即分类模型),采用同样方法得到平均的ROC曲线,用于评价混合模型性能,以测试热应激风险预测模型的泛化能力。

2 热应激预测模型评价与结果分析

2.1 主成分分析

主成分分析选择身高、体质量、年龄、性别、体脂率、PAR、CI、环境温度、相对湿度9个特征^[32],以累计贡献率(反映主成分对原有信息的保留情况)作为主成分个数选择的依据。选择80%累计贡献率作为主成分个数的判断标准,当主成分个数为4时,累计贡献率约为78%,接近于80%,而主成分个数为5时,累计贡献率则接近90%,为实现数据降维,选择主成分数量为4,在此基础上训练各个机器学习模型。

2.2 模型性能评价

2.2.1 核心温度预测模型性能评价

10折交叉验证下各机器学习模型的性能评价结果见表3。其中,每一个指标的评价结果数值均为13个时间节点的核心温度预测平均值。结果显示,KNN、RF、GBDT和DT模型的拟合优度 R^2 均超过0.9,主成分分析方法降维输入数据特征后,模型预测精度会有较小的提升, R^2 上升幅度为0.01~0.05;误差方面,KNN的预测误差最小,MSE、RMSE和MAE分别为0.003、0.053、0.010。KNN模型对各时间节点下的核心温度预测性能评价结果见表4。 R^2 的数值均达到0.9以上,RMSE的数值在各节点均低于0.1,可见:该模型对不同节点的预测性能达到较高的水平。KNN模型对各时间节点核心温度的预测值与实际值的对比结果如图2所示。

2.2.2 热应激风险预测模型性能评价

以FPR作为横轴,TPR作为纵轴,计算ROC的AUC,以量化评价热应激风险分类预测模型的性能,见表5。结果显示,RF、GBDT和KNN模型预测性能明显优于其他模型,其中,随机森林模型AUC最高,为1。

表 3 核心温度预测模型性能评价结果

Table 3 Evaluation results of performance for core temperature prediction model

模型	MSE	RMSE	MAE	R ²
KNN	0.003	0.053	0.010	0.986
RR	0.134	0.402	0.304	0.573
SVM	0.081	0.238	0.152	0.847
DT	0.025	0.145	0.034	0.918
GBDT	0.023	0.143	0.088	0.925
RF	0.013	0.105	0.042	0.954

表 4 KNN 模型对各时间节点核心温度预测性能评价结果

Table 4 Evaluation results of KNN model on core temperature prediction performance of each time node

因变量	MSE	RMSE	MAE	R ²	
核心温度	节点 1	0.003	0.058	0.010	0.999
	节点 2	0.004	0.065	0.013	0.995
	节点 3	0.000	0.014	0.003	1.000
	节点 4	0.002	0.039	0.008	0.997
	节点 5	0.002	0.045	0.009	0.996
	节点 6	0.009	0.093	0.013	0.977
	节点 7	0.003	0.055	0.011	0.981
	节点 8	0.003	0.054	0.010	0.978
	节点 9	0.002	0.048	0.009	0.978
	节点 10	0.002	0.045	0.009	0.979
	节点 11	0.003	0.053	0.009	0.975
	节点 12	0.004	0.063	0.013	0.975
	节点 13	0.003	0.056	0.012	0.982

表 5 热应激风险模型的 AUC 值

Table 5 AUC values for heat stress risk model

模型	AUC
KNN	0.97
SVM	0.90
DT	0.93
GBDT	0.99
RF	1.00

表 6 核心温度预测模型、热应激风险预测模型特征值重要性得分

Table 6 Importance scores of eigenvalues for core temperature prediction model and heat stress risk prediction model

核心温度预测模型				热应激风险预测模型			
RF		GBDT		RF		GBDT	
特征	重要性得分	特征	重要性得分	特征	重要性得分	特征	重要性得分
环境温度	0.29	环境温度	0.3	环境温度	0.27	环境温度	0.31
PAR	0.23	PAR	0.22	PAR	0.24	PAR	0.27
相对湿度	0.18	年龄	0.19	CI	0.11	体脂率	0.12
年龄	0.13	相对湿度	0.17	体脂率	0.1	体质量	0.098
体脂率	0.074	体脂率	0.058	体质量	0.097	CI	0.083

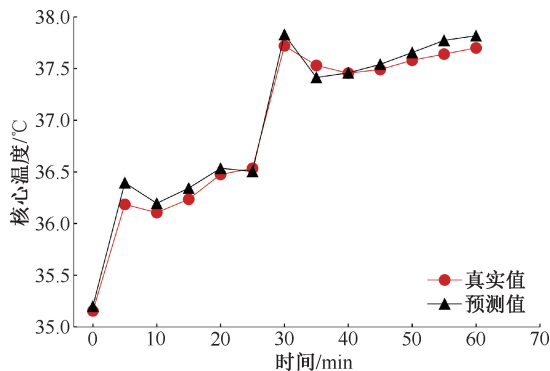


图 2 KNN 模型对各时间节点核心温度的预测值与实际值的对比

Fig. 2 Comparison of predicted and actual values of core temperature of each time node by KNN model

2.3 特征重要性排序

基于平均不纯度减少^[33]算法,利用随机森林和梯度提升决策树模型对所选的 9 个特征进行重要性排序(使用未经主成分分析的特征构建模型并进行特征重要性排序)。核心温度预测模型和热应激风险预测模型的特征值重要性排序和得分见表 6。上述 2 组模型得出的特征重要性排序中,环境温度、身体活动比率均排在前 2 位,可见:环境温度和劳动强度是对人体热应激产生影响最大的 2 个因素,因为这 2 个因素直接影响人体的热平衡和代谢活动。其中,对于核心温度预测模型,环境温度、身体活动比率、相对湿度、年龄是最重要的 4 个指标。对于热应激风险预测模型,环境温度、身体活动比率、服装热阻、体脂率则是最重要的 4 个指标,说明在考虑热应激风险时,服装热阻和体脂率更具有预测能力,服装热阻可影响人体的散热效率,而体脂率则与身体内部热量储存和释放有关。因此,对于高温作业下的户外作业警务人员,除需要考虑警务人员自身的身体健康状况(如体脂率),也需要重点关注其劳动强度和作业着装的影响,从而采取必要措施降低热应激风险及相关事故概率。

续表 6

核心温度预测模型				热应激风险预测模型			
RF		GBDT		RF		GBDT	
特征	重要性得分	特征	重要性得分	特征	重要性得分	特征	重要性得分
CI	0.039	CI	0.026	身高	0.069	身高	0.067
体质量	0.03	体质量	0.016	年龄	0.052	年龄	0.036
身高	0.027	身高	0.012	相对湿度	0.036	相对湿度	0.02
性别	0.008 6	性别	0.008 1	性别	0.029	性别	0.002 9

2.4 泛化能力测试结果

基于混合数据的核心温度预测模型的泛化能力评价结果见表 7,其中,每一个指标的评价结果数值均为 13 个时间节点的核心温度预测平均值。结果显示, J-1 模型的误差明显小于 J-2 模型,其中, RMSE 均小于 0.5, MAE 均小于 0.4,且各机器学习模型间的误差差距不大,说明基于混合数据的核心温度预测模型相比于单纯基于 JOS-3 数据训练而得的模型(J-2),具有更高的预测精度。

表 7 核心温度混合预测模型性能评价结果

Table 7 Evaluation results of performance for core temperature hybrid prediction model

模型	J-1		J-2	
	RMSE	MAE	RMSE	MAE
KNN	0.44	0.33	1.03	0.86
RR	0.41	0.30	1.06	0.89
SVM	0.42	0.29	0.99	0.84
DT	0.50	0.39	1.02	0.86
GBDT	0.44	0.32	1.02	0.87
RF	0.42	0.31	1.02	0.86

基于混合数据的热应激风险预测模型的平均 ROC 曲线见表 8。可见:基于不同机器学习算法的 J-1 模型的 AUC 均高于 J-2 模型。其中,基于 RF 算法的 J-1 模型的 AUC 最高,为 0.74。以上结果表明:基于混合数据的热应激风险预测模型相比于单纯基于 JOS-3 数据训练而得的模型具有更好的预测效果。

作为比较,将测试集的特征数据输入 JOS-3 模型中,对比计算得到的核心温度与试验得到的核心

温度,得到 RMSE 为 0.88, MAE 为 0.79, AUC 为 0.7,其误差高于 J-2 模型,但低于 J-1 模型,说明 J-1 在测试集上性能也优于 JOS-3。

表 8 热应激风险混合预测模型性能评价结果

Table 8 Evaluation results of the performance for heat stress risk hybrid prediction model

模型	AUC	
	J-1	J-2
KNN	0.70	0.67
SVM	0.66	0.73
DT	0.61	0.61
GBDT	0.68	0.48
RF	0.74	0.61

3 结 论

1) 基于机器学习的户外作业警务人员热应激预测模型在添加高斯噪声的数据集上具有较好的预测性能,拟合优度 R^2 普遍超过 0.9,误差 RMSE、MAE 较小。

2) 在泛化能力测试过程中发现,混合试验数据与 JOS-3 数据的热应激风险预测模型相比单纯基于 JOS-3 数据训练而得的模型具有更好的预测效果,这说明人工气候室试验数据对于提升热生理模型的预测能力具有显著作用,未来应持续通过上述试验方式积累基于不同工况和研究对象的试验数据。

3) 基于 RF 和 GBDT 模型的特征重要性排序发现,环境温度、PAR 均排在前 2,对预测结果的刻画能力最强,在未来的研究工作中应重视这 2 个指标的研究与应用。

参 考 文 献

[1] POTTER A W, GONZALEZ J A, KARIS A J, et al. Biophysical assessment and predicted thermophysiological effects of body armor[J]. Plos One, 2015,10(7): DOI: 10.1371/journal.pone.0132698.

[2] DU Chenqiu, LI Baizhan, LI Yongqiang, et al. Modification of the predicted heat strain (PHS) model in predicting human thermal responses for Chinese workers in hot environments[J]. Building and Environment, 2019, 165: DOI:

10. 1016/j. buildenv. 2019. 106349.
- [3] MALCHAIRE J, PIETTE A, KAMPMANN B, et al. Development and validation of the predicted heat strain model[J]. *The Annals of Occupational Hygiene*, 2001, 45(2): 123–135.
- [4] 安启启, 徐刚, 杨杰, 等. 高温高湿环境下人体热生理模型的检验及应用[J]. *西安科技大学学报*, 2021, 41(2): 253–261.
- AN Qiqi, XU Gang, YANG Jie, et al. Examination and application of human thermo physiological model in high temperature and humidity environment[J]. *Journal of Xi'an University of Science and Technology*, 2021, 41(2): 253–261.
- [5] 李春艺, 李川, 闫肖岳, 等. 户外高温环境建筑工人热应激预测分析评价[J]. *劳动保护*, 2022(6): 78–82.
- [6] 吴建松, 李乐天, 韩新阳, 等. 高温户外环境电网作业人员热应激预测评价[J]. *中国安全生产科学技术*, 2021, 17(12): 169–175.
- WU Jiansong, LI Letian, HAN Xinyang, et al. Prediction and evaluation on heat strain of power grid workers in high-temperature outdoor environment[J]. *Journal of Safety Science and Technology*, 2021, 17(12): 169–175.
- [7] 聂兴信, 王廷宇, 孙锋刚, 等. 高温矿井热湿环境对人体机能的影响[J]. *金属矿山*, 2020(4): 186–193.
- NIE Xingxin, WANG Tingyu, SUN Fenggang, et al. Influence of heat and humidity environment on function of human body in high temperature mine[J]. *Metal Mine*, 2020(4): 186–193.
- [8] 柯莹, 周文. 个体降温服降温效应评价指标及方法[J]. *服装学报*, 2021, 6(1): 1–7.
- KE Ying, ZHOU Wen. Evaluation indicators and methods of cooling effects for personal cooling clothing[J]. *Journal of Clothing Research*, 2021, 6(1): 1–7.
- [9] HAN Xinge, HU Zhuqiang, LI Chuan, et al. Prediction of human thermal comfort preference based on supervised learning[J]. *Journal of Thermal Biology*, 2023, 112: DOI:10. 1016/j. jtherbio. 2023. 103484.
- [10] CHAUDHURI T, SOH Y C, LI H, et al. Machine learning based prediction of thermal comfort in buildings of equatorial Singapore[C]. 2017 IEEE International Conference on Smart Grid and Smart Cities (ICSGSC). IEEE, 2017: 72–77.
- [11] MORRESI N, CASACCIA S, SORCINELLI M, et al. Sensing physiological and environmental quantities to measure human thermal comfort through machine learning techniques[J]. *IEEE Sensors Journal*, 2021, 21(10): 12 322–12 337.
- [12] LIU Kuixing, NIE Ting, LIU Wei, et al. A machine learning approach to predict outdoor thermal comfort using local skin temperatures[J]. *Sustainable Cities and Society*, 2020, 59: DOI:10. 1016/j. scs. 2020. 102216.
- [13] MORISHIMA S, XU Yingjie, URASHIMA A, et al. Human body skin temperature prediction based on machine learning[J]. *Artificial Life and Robotics*, 2021, 26: 103–108.
- [14] YANG Jie, WANG Lijuan, YIN Haiguo, et al. Thermal responses of young males of three thermal preference groups in summer indoor environments[J]. *Building and Environment*, 2021, 194(11): DOI:10. 1016/j. buildenv. 2021. 107705.
- [15] KARJALAINEN S. Gender differences in thermal comfort and use of thermostats in everyday thermal environments[J]. *Building and Environment*, 2007, 42(4): 1 594–1 603.
- [16] KARJALAINEN S. Thermal comfort and gender: a literature review[J]. *Indoor Air*, 2012, 22(2): 96–109.
- [17] FERRARO S, IAVICOLI S, RUSSO S, et al. A field study on thermal comfort in an Italian hospital considering differences in gender and age[J]. *Applied Ergonomics*, 2015, 50: 177–184.
- [18] HAVENITH G. Individualized model of human thermoregulation for the simulation of heat stress response[J]. *Journal of Applied Physiology*, 2001, 90(5): 1 943–1 954.
- [19] HUIZENGA C, ZHANG H, THOMAS D. An improved multimode model of human physiology and thermal comfort[J]. *Proceedings of Building Simulation*, 1999, 35(1): 353–359.
- [20] TAKAHASHI Y, NOMOTO A, YODA S, et al. Thermoregulation model JOS-3 with new open source code[J]. *Energy and Buildings*, 2021, 231: DOI:10. 1016/j. enbuild. 2020. 110575.
- [21] ISO 8996–2021, Ergonomics of the thermal environment: determination of metabolic rate [S].
- [22] 范淼. 机器学习及实践: 从零开始通往 Kaggel 竞赛之路[M]. 北京: 清华大学出版社, 2016: 183.
- [23] HEARST M A, DUMAIS S T, OSUNA E, et al. Support vector machines[J]. *IEEE Intelligent Systems and Their Applications*, 1998, 13(4): 18–28.
- [24] FRIEDL M A, BRODLEY C E. Decision tree classification of land cover from remotely sensed data[J]. *Remote Sensing of Environment*, 1997, 61(3): 399–409.

[25] KE Guolin, MENG Qi, FINLEY T, et al. LightGBM: a highly efficient gradient boosting decision tree[C]. 31st International Conference on Neural Information Processing Systems, 2017:3 149–3 157.

[26] RIGATTI S J. Random forest[J]. Journal of Insurance Medicine, 2017,47(1):31–39.

[27] BUZA K, NANOPOULOS A, NAGY G. Nearest neighbor regression in the presence of bad hubs[J]. Knowledge-Based Systems, 2015,86:250–260.

[28] MCDONALD G C. Ridge regression[J]. Wiley Interdisciplinary Reviews: Computational Statistics, 2009, 1(1): 93–100.

[29] AWAD M, KHANNA R. Support vector regression[J]. Neural Information Processing Letters & Reviews, 2007, 11(10):203–224.

[30] FAWCETT T. An introduction to ROC analysis[J]. Pattern Recognition Letters, 2006,27(8):861–874.

[31] STOLWIJK J A. A mathematical model of physiological temperature regulation in man[R]. Nasa Langley, 1971.

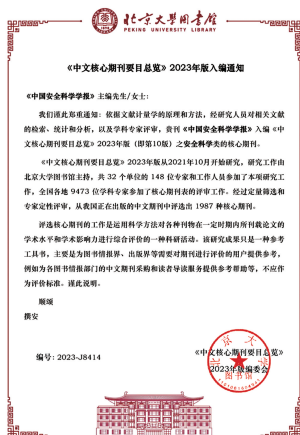
[32] WOLD S, ESBENSEN K, GELADI P. Principal component analysis[J]. Chemometrics and Intelligent Laboratory Systems, 1987,2(1/2/3):37–52.

[33] LOUPPE G, WEHENKEL L, SUTERA A, et al. Understanding variable importances in Forests of randomized trees[J]. Advances in Neural Information Processing Systems, 2013,26:431–439.

作者简介：胡啸峰（1986—），男，河北唐山人，工学博士，副教授，主要从事风险评估与预测预警技术方面的研究。E-mail: huxiaofeng@ppsuc.edu.cn。



《中国安全科学学报》再次入编“北大中文核心期刊”



依据文献计量学的原理和方法,经研究人员对相关文献的检索、统计和分析,以及学科专家评审,《中国安全科学学报》入编《中文核心期刊要目总览》2023年版(即第10版)之“安全科学”类的核心期刊。