

张静怡, 曹自平, 邱利康. 基于步态信息的运动疲劳识别系统研究[J]. 智能计算机与应用, 2026, 16(2): 103-111. DOI: 10.20169/j. issn. 2095-2163. 24040401

## 基于步态信息的运动疲劳识别系统研究

张静怡, 曹自平, 邱利康

(南京邮电大学 通信与信息工程学院, 南京 210003)

**摘要:** 本文设计了一种基于步态信息的疲劳识别系统。通过压力膜和惯性传感器实时采集步态信息, 并通过蓝牙将步态信号发送到上位机。上位机采用动态时间规整算法进行步态分割, 从单个步态序列中提取最大、最小、均值和标准差等统计特征, 再利用机器学习分类模型识别疲劳状态。在本文中疲劳状态被划为5种状态: 正常步行和最大耗氧量的60%、70%、80%、90%对应的运动强度。疲劳状态识别算法采用随机森林、支持向量机、梯度增强决策树等机器学习分类算法。实验结果表明, 在测试集中, 随机森林疲劳等级识别的准确率达到94.7%。经过信号筛选后, 随机森林疲劳等级识别的准确率仍可以达到91.3%。

**关键词:** 疲劳识别; 步态信息; 动态时间规整; 特征提取; 随机森林

**中图分类号:** TP18 **文献标志码:** A **文章编号:** 2095-2163(2026)02-0103-09

## Recognition system of fatigue states based on gait information

ZHANG Jingyi, CAO Ziping, QIU Likang

(School of Communications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China)

**Abstract:** In this paper, a fatigue recognition system based on gait information is designed. Gait information is collected in real time through a pressure membrane and inertial sensor, and the gait signal is sent to the upper computer by Bluetooth. The upper computer uses a dynamic time warping algorithm for gait segmentation. From a single gait sequence, the maximum, minimum, mean, standard deviation and other statistical features are extracted, and then a machine learning classification model is used to identify the fatigue state. In this paper, the fatigue state is divided into 5 types: normal walking and 60%, 70%, 80%, and 90% of the maximum oxygen consumption correspond to the exercise intensity. The fatigue state recognition algorithm uses machine learning classification algorithms such as Random Forest, Support Vector Machine, and Gradient Boosting Decision Tree. The experimental results show that in the test set, the accuracy of random forest fatigue level recognition reaches 94.7%. After signal screening, the accuracy of random forest fatigue level recognition can still reach 91.3%.

**Key words:** fatigue recognition; gait information; dynamic time warping; feature extraction; Random Forest

## 0 引言

人体运动疲劳状态一般是通过肌肉疲劳程度来反映, 肌肉疲劳被定义为肌肉因运动而产生力量的下降, 从而在自主收缩过程中无法维持可预测的最大力量<sup>[1]</sup>。对于长时间进行重复性活动的工人们来讲, 重复工作引起的肌肉疲劳会导致功能能力的丧失<sup>[2]</sup>, 而长期的肌肉疲劳则可能会导致肌肉骨骼疾病<sup>[3]</sup>。对于运动员来讲, 长期的高强度训练引发的肌肉疲劳则会导致肌肉外骨骼疾病和运动能力表

现的下降<sup>[4]</sup>。此外, 人们在生活中也会无法避免一些不当的活动, 使身体产生负担导致肌肉疲劳。

由于肌肉疲劳在生活和工作时时有发生, 因此对其检测和评价明显具有现实意义, 目前其已成为一个广受关注的研究领域。关于疲劳检测的方法, 已有众多研究学者做出了大量有价值的工作。目前用于检测人体疲劳的方法包括可穿戴传感器、光学运动捕捉系统和力板分析。对于光学捕捉系统, 其优势是能够准确地捕捉运动轨迹而进行详细的运动学分析, 但通常设备较大, 不便于携带, 往往仅适用

**作者简介:** 张静怡(1999—), 女, 硕士研究生, 主要研究方向: 可穿戴传感。Email: zjy199911212021@163.com; 曹自平(1974—), 男, 教授, 主要研究方向: 可穿戴传感; 邱利康(1996—), 男, 硕士, 主要研究方向: 可穿戴传感。

收稿日期: 2024-04-04

哈尔滨工业大学主办 ◆ 系统开发与应用

于实验室环境,同时其数据处理和分析时间较长,不能及时取得反馈。相比于光学捕捉系统,力板分析所需的计算量大大减少,其优点是能够直接测量力学参数,如脚底压力分布,可提供精确的动力学数据,但另一方面,其缺点是力板需固定于特定地点,无法用于日常监测。除此之外,其往往用于特定的运动或站立分析,而不能全面地评估疲劳<sup>[5-6]</sup>。

可穿戴传感器相对于前两者而言,通常尺寸小巧,易于佩戴,在日常生活中可以无感地监测身体生理参数,同时可穿戴传感器能够测量的生理信号多种多样,通过进一步结合机器学习方法对信号数据进行分析,能大幅提升疲劳预测的准确性。例如, Tkach 等学者<sup>[6]</sup>利用表面肌电(sEMG)信号来分析运动疲劳状态,在时域分析中同时提取频域信息和特征,较为有效地评估了肌肉疲劳。步态信息作为生理信号之一,能够反映人体的多种状态,特别是人体疲劳程度的状态。Guaitolini 等学者<sup>[7]</sup>使用佩戴在受试者下肢、骨盆和躯干的惯性传感器进行行走实验,记录受试者的步幅时间(ST)和其变化性(STV)、步幅长度(SL)和其变化性(SLV)、步速(GS)、对称指数(SI)、膝关节运动范围(ROM)和小腿角速度(AV)等数据,这些特征作为支持向量机(SVM)分类器的输入,用于区分疲劳状态和非疲劳状态,SVM分类器的准确率达到84.62%。Zhang 等学者<sup>[8]</sup>使用5个红外反射标记和2个惯性测量单元测量受试者的步态信息,计算原始信号的平均绝对值、过零次数、标志变化次数、步长和主要频率,将合成加速度和原始信号作为SVM分类器的输入,用于区分疲劳状态和非疲劳状态,分类准确率达到96%。需要指出的是,以上工作通常对疲劳划分采用的是二分法,即把肌肉运动状态粗略地分为疲劳和非疲劳,而据本文分析可知,对于不同程度疲劳的判断研究还鲜见报道。事实上,在实际当中,相关人员也同样十分关心疲劳状态的不同程度。

基于以上考虑,在本研究中,除了使用可穿戴惯性传感器来记录步态信息外,还引入压力传感器来获取更多的步态信息,在机器学习方法的辅助下探索识别5种运动疲劳状态的可行性。惯性传感器位于受试者的脚踝上,压力传感器位于鞋垫内,分别同时采集步态数据。本文还考察了Shapley值对传感器信号选择分类精度的影响,实现了对传感器信号需求量的降低,从而使疲劳判别方案更加简洁、方便、高效和经济。

## 1 材料和方法

### 1.1 受试者

在本实验中,招募10名健康成年人,均为研究生,年龄为24(±1)岁,体重为65(±12)kg。

### 1.2 实验内容

根据徐航提出的基于跑步机的心肺耐力实验方法,进行跑步机二次负荷实验。心肺耐力两级测试的公式如下:

$$FC_{\max} = \frac{EC_2 - EC_1}{HR_2 - HR_1} \times (HR_{\max} - HR_1) + EC_1 \quad (1)$$

其中,  $EC_1$  和  $EC_2$  分别表示一级运动强度和二级运动强度;  $HR_1$  和  $HR_2$  分别表示一级运动强度和二级运动强度所对应的心率;  $HR_{\max}$  表示心率的极大值,  $HR_{\max}$  的计算公式为:

$$HR_{\max} = 208 - 0.7 \times \text{Age} \quad (2)$$

其中,参数  $EC_1$ 、 $EC_2$  和  $FC_{\max}$  的单位为 MET, MET 为 3.5 ml/(kg·min)。

研究表明,一级运动强度为6 MET,跑步机速度为5.5 km/h,坡度为5%;二级运动强度为10 MET,跑步机速度8.3 km/h,坡度3%。通过上述2次负荷试验,可以计算出各试验人员的最大运动强度。

疲劳水平设为5个等级,其中0级为不运动的强度,1~4级为马淑芳提出的心肺耐力运动处方中4级增量负荷试验对应的4个运动强度。从2次负载测试来看,  $FC_{\max}$  和  $VO_{2\max}$  是可以被计算出来的。根据《ACSM运动试验及处方指南》中给出的公式:

$$FC = \frac{3.5 + 0.2 \times V + 0.9 \times V \times A}{3.5} \quad (3)$$

可以推知,当坡度为0时可以计算出60%、70%、80%、90%的  $VO_{2\max}$  对应的运动速度,并且60%、70%、80%、90%的  $VO_{2\max}$  可作为四级增量负荷的强度。

式(3)中,参数  $V$  的单位是 m/min,参数  $A$  的单位是%。测试者的运动强度设计见表1。按表1的运动强度运动4 min,记录直线行走100 m的步态信号,每次记录结束后休息3 min。

表1 疲劳程度与运动强度对应表

Table 1 Correspondence table of fatigue levels and exercise intensity

疲劳程度	运动强度
1	不运动
2	60% $VO_{2\max}$
3	70% $VO_{2\max}$
4	80% $VO_{2\max}$
5	90% $VO_{2\max}$

### 1.3 步态信号采集

步态信号采集装置由 STM32 单片机、压力膜鞋垫、惯性传感器、蓝牙和一些必要的外围电路组成,如图 1 所示。图 1 中,压力膜鞋垫采集足底压力信号,惯性传感器采集踝部运动信号,采样频率为 50 Hz。惯性传感器采集脚踝处的三轴加速度、三轴角速度和三轴角信号。压力膜鞋垫在脚底的 8 个点收集信号。

保证受试者在实验前没有剧烈运动,然后将压力薄膜插入运动鞋中,并将惯性传感器置于受试者右脚踝处。受试者佩戴传感器后,直立行走 2 min,目的是让受试者适应佩戴传感器行走。受试者适应后,采集疲劳 0 级步态信息,以正常步速在直线走廊行走 100 m,完成疲劳 0 级步态信息数据采集,休息 3 min。受试者在跑步机以 60%  $VO_{2max}$  对应的速度

跑 4 min 后,立即在走廊上直线步行 100 m,采集疲劳 1 级步态信息,休息 3 min。受试者在跑步机上以 70%、80% 和 90%  $VO_{2max}$  对应的速度,用相同的方法测量疲劳等级 2、3 和 4 的步态信息。采集的信号如图 2 所示。

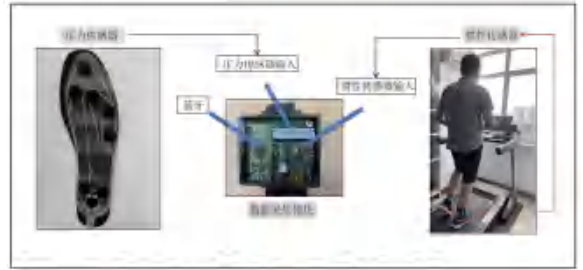
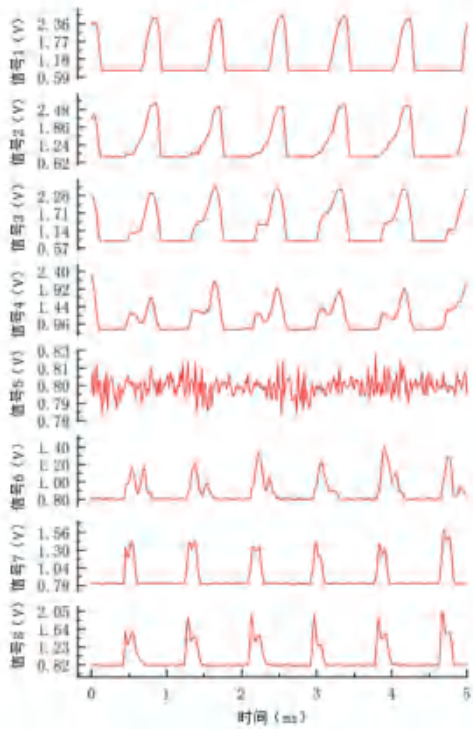
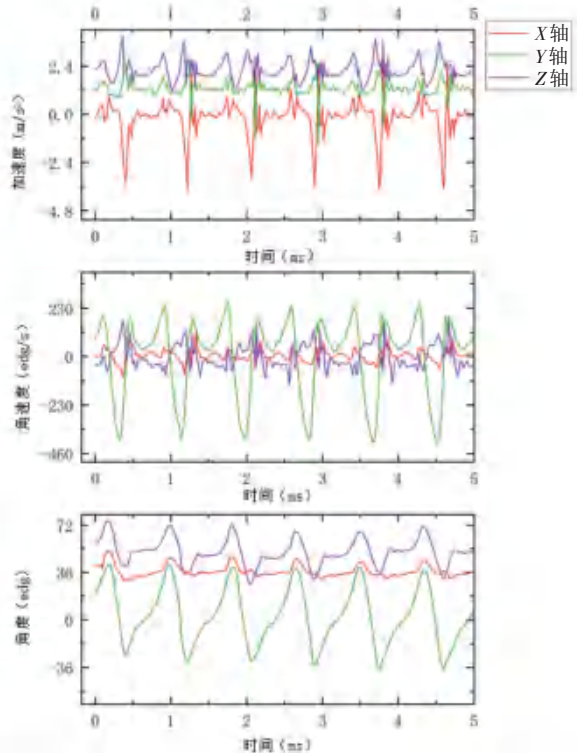


图 1 用于收集步态信息的可穿戴设备

Fig. 1 Wearable devices for collecting gait information



(a) 压力传感器阵列采集的 8 个信号



(b) 惯性传感器采集的 9 个信号

图 2 压力传感器阵列采集的 8 个信号和惯性传感器采集的 9 个信号

Fig. 2 Eight signals collected by pressure sensor array and nine signals collected by IMU

### 1.4 步态分割

动态时间规整 (DTW) 是一种常用的距离测量方法。对于 2 个时间序列,一个时间序列中的单个数据点可以匹配另一个时间序列中的多个数据点<sup>[9]</sup>。此外,DTW 可以找到相似的时间序列,尽管这些时间序列可能不对齐,长度可能不同<sup>[10]</sup>。步态分割采用动态时间翘曲算法,该算法可以计算出时间序列与模板序列最相似的子序列。这些子序列是

完成步态分割后的单步信号,如图 3 所示。该算法的时间复杂度为  $O(LN)$ ,其中  $L$  是模板长度,一般来说, $L$  是一个常数; $N$  是时间序列的长度。步态分割的主要难点在于模板的选择,因为在本文实验中有很多类型的信号。应该选择周期明显、变化简单的信号,然后使用几段来拟合成信号的一个周期。这些分段拟合的曲线就是模板。

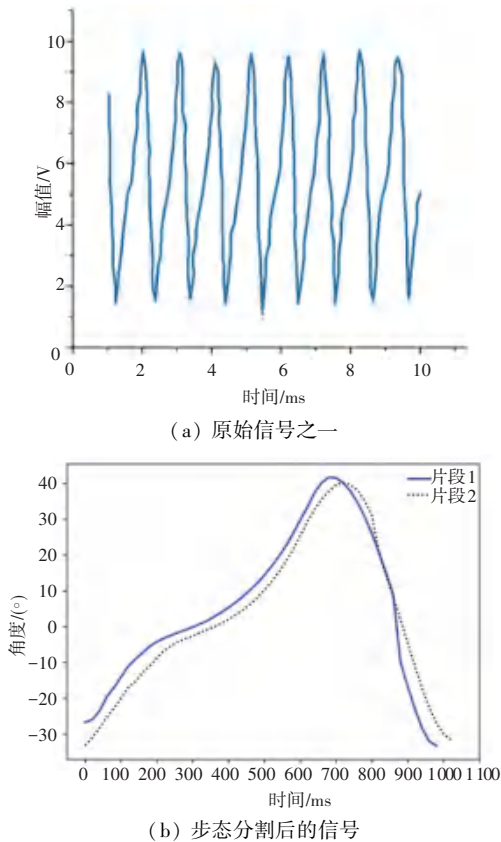


图3 步态分割后的单步信号

Fig. 3 Single-step signal after gait segmentation

## 1.5 特征提取

经过步态分割后,将单步信息从步态信号中分离出来。找到单步信号的步长<sup>[11]</sup>、最大值(Max)、最小值(Min)、范围(最大值-最小值)(Range)、均值(Mean)、均值绝对值(MeanAbsolute)、标准差(Std)和变异系数(CV)。将这些特征作为机器学习模型的输入。上述这些特征可以通过公式得到:

$$\text{Max} = \max(x_1, x_2, x_3, \dots, x_n) \quad (4)$$

$$\text{Min} = \min(x_1, x_2, x_3, \dots, x_n) \quad (5)$$

$$\text{Range} = \text{Max} - \text{Min} \quad (6)$$

$$\text{Mean} = \frac{1}{n} \sum_{i=0}^n x_i \quad (7)$$

$$\text{MeanAbsolute} = \frac{1}{n} \sum_{i=0}^n |x_i| \quad (8)$$

$$\text{Std} = \sqrt{\frac{1}{n-1} \sum_{i=0}^n (x_i - \text{Mean})^2} \quad (9)$$

$$\text{CV} = \frac{\text{Std}}{\text{Mean}} \times 100\% \quad (10)$$

## 1.6 机器学习算法和计算库

常用的机器学习分类模型包括逻辑回归(LR)、支持向量机(SVM)、决策树(DT)、随机森林(RF)、梯度增强决策树(GBDT)、AdaBoost、XGBoost和多

层感知器(MLP)。

(1)逻辑回归,也称为对数概率回归,是一种使用直线、平面或超平面来分离类别的分类器。其决策边界是输入的线性函数,现已广泛应用于数据挖掘、机器学习、计算机视觉和生物信息学等领域,并在高维稀疏数据中表现良好<sup>[12]</sup>。其优点是训练速度和预测速度都非常快,适用于高维、稀疏、线性的数据,并且易于解释。而缺点是容易欠拟合,并且非线性数据的精度不高。

(2)SVM是一种机器学习算法,可以用于分类任务。可通过训练找到最佳决策边界<sup>[13]</sup>,然后根据最佳决策边界预测新观察到的类。支持向量机也是基于线性函数的,但其对于线性不可分的情况下,将低维输入空间中线性不可分的样本通过非线性映射(核技巧)转换为高维特征空间,使其线性可分,从而使高维特征空间线性可分。在特征空间中使用线性算法可以对样本的非线性的特征进行线性分析,因此可以对线性和非线性类边界进行处理<sup>[14]</sup>。其优点是可用于线性和非线性分类和回归,易于解释,泛化误差小,计算复杂度低。缺点是对参数和核函数的选择比较敏感。

(3)决策树模型是一种简单的非参数分类器。分类与回归树(Classification and Regression Tree, CART)是Breiman提出的一种决策树算法,可用于分类或回归预测<sup>[15]</sup>。这个算法创建了一个二叉树,每个非叶子节点总是有2个分支,使用基尼指数作为损失函数,并使用剪枝方法生成一系列具有不同剪枝节点数的树。然后通过检查每棵树在不同子数据集上的分类误差,选择误差最小的树<sup>[16]</sup>。该方法的优点是易于解释,对数据尺度不敏感,不需要对数据的结构和分布做任何假设,可以捕捉变量之间的相互作用。缺点则是决策树容易对样本数据过于微调,失去稳定性和抗冲击性,决策树需要较大的样本量,容易过拟合,泛化能力差。

(4)决策树的集成模型(随机森林、梯度增强决策树、AdaBoost和XGBoost)克服了决策树的缺点,通过集成多个弱学习器(浅深度决策树)成为强学习器<sup>[17]</sup>。优点是泛化能力强,不易过拟合,不需要调整太多参数<sup>[18]</sup>。缺点是计算量大。

(5)多层感知器,是一种前馈神经网络。具体是受到生物神经网络的启发,模拟了人脑的一些智能行为<sup>[19]</sup>。同时具有自学习、自适应和非线性动态处理的特点。多层感知器由一个输入层、几个隐藏层和一个输出层组成。其优点是可以通过设置隐藏

层的大小来改变模型的复杂度,分类精度高,噪声神经网络具有较强的鲁棒性和容错性,可以充分逼近复杂的非线性关系<sup>[20]</sup>。缺点是容易过拟合,难以调整参数和梯度分散。

综上可知,分类模型的优缺点见表 2。本文使用 Scikit Learn 和 XGBoost 机器学习库构建模型。Scikit Learn 是一个免费的机器学习库,主要用

Python 语言编写。Scikit Learn 提供了多种数据预处理算法、监督学习算法和无监督学习算法<sup>[21]</sup>。其中,Numpy 计算库被广泛用于执行高性能的矩阵运算,可以反复利用计算机的多核性能<sup>[22]</sup>。XGBoost 机器学习库是一个高效、灵活、快速、优化的分布式梯度增强机器学习库<sup>[23]</sup>。XGBoost 非常适合分布式环境,可以解决大量数据的机器学习问题。

表 2 分类模型的优缺点

Table 2 Advantages and disadvantages of classification models

模型	优点	缺点
逻辑回归	计算量小;适用于高维、稀疏数据;易于解释	易欠拟合,对非线性数据精度不高
支持向量机	计算量小,易于解释,泛化能力强	敏感的选择参数和核函数
决策树	易于解释,对数据扩展不敏感	易过拟合,泛化能力差
决策树的集成模型	泛化能力强,不易过拟合	计算量大
多层感知器	模型复杂度可调,分类精度高	容易过拟合,难以调整参数,梯度分散

## 2 结果

### 2.1 分类模型的初步选择

采用 5 倍交叉验证对模型进行评估<sup>[24]</sup>。具体方法是将数据集被打乱后分成 5 部分,提取一部分作为测试集,其余部分作为训练集。这样一来,每次提取的测试集是不同的,得到了 5 个不同的数据集。这 5 个数据集用于模型训练和测试,测试集的准确率以均值和方差计算。5 重交叉验证可以有效避免数据集划分的局限性和特殊性,得到更合理、准确的模型评价。

5 重交叉验证准确率对照见表 3。模型分类精度的比较如图 4 所示。从表 3 和图 4 可以看出,随机森林、梯度增强决策树、AdaBoost 和 XGBoost 在树集成模型中的表现更好,其精度相似,标准差最小。因此,初步选择将其作为疲劳水平识别模型。

表 3 5 重交叉验证准确度对照表

Table 3 Five-fold cross-validation accuracy comparison table

模型	平均值	误差	模型	平均值	误差
逻辑回归	0.249	0.050	AdaBoost	0.949	0.007
决策树	0.796	0.031	XGBoost	0.935	0.004
随机森林	0.947	0.008	支持向量机	0.446	0.022
梯度增强	0.921	0.010	多层感知器	0.320	0.208

### 2.2 传感信号的选择

本实验同时使用压力传感器(8 个信号)和惯性传感器(9 个信号),共 17 个信号。对于如此多的信号,需要对传感器信号进行选择,减少冗余信号的数量,以简化疲劳识别数据的采集,降低成本。

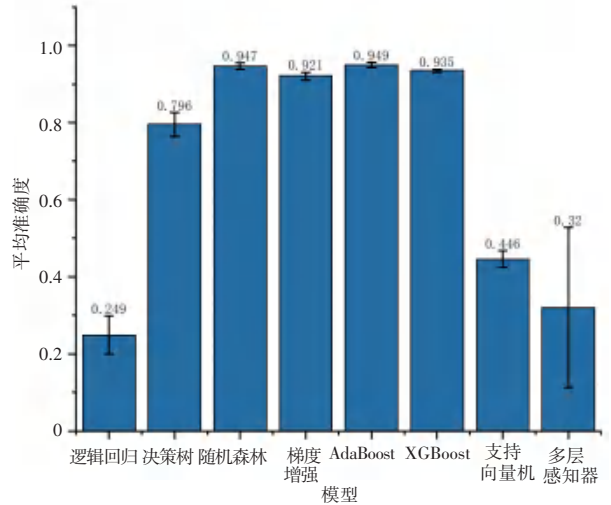


图 4 模型分类精度的比较

Fig. 4 Comparison of model classification accuracy

传感器信号选择有 3 种方案:所有信号均来自压力传感器,记为方案 1;所有信号均来自惯性传感器,记为方案 2;信号来自压力传感器和惯性传感器,记为方案 3。采用随机森林、梯度增强决策树、AdaBoost 和 XGBoost 四种模型分别识别方案 1 信号和方案 2 信号的疲劳程度。准确度结果具体见表 4、表 5。将识别结果分为 2 种:方案 1 和方案 2 的准确率接近于使用所有信号的准确率,选择准确率较高的方案;或者方案 1 和方案 2 的精度与使用所有信号的精度不同,则选择方案 3。

由表 4 和表 5 中结果可以看出,方案 2 的精度与使用全部信号的精度相差很小,随机森林的精度小于 0.024。本方案不含压力传感器,且精度较高,故采用方案 2。

表 4 5 重交叉验证准确度比较(方案 1)

Table 4 Comparison table of the accuracy of five-fold cross-validation(Scheme 1)

模型	平均值	误差	模型	平均值	误差
随机森林	0.827	0.014	AdaBoost	0.822	0.015
梯度增强	0.797	0.006	XGBoost	0.831	0.013

表 5 5 重交叉验证准确度比较(方案 2)

Table 5 Comparison table of the accuracy of five-fold cross-validation(Scheme 2)

模型	平均值	误差	模型	平均值	误差
随机森林	0.923	0.013	AdaBoost	0.917	0.009
梯度增强	0.894	0.007	XGBoost	0.914	0.008

传感器信号最初选择了惯性传感器的所有信号,但惯性传感器总共有 9 个信号。这 9 个信号极有可能存在信息冗余,因此应进一步对惯性传感器信号进行筛选。选择的依据是每个信号的 Shapley 值的平均绝对值。平均绝对值越大,表示该特征越

重要。Shapley 值可以表示特征对模型预测的贡献。Shapley 值是 Shapley 提出的一个博弈论概念<sup>[25]</sup>。让人们看到每个特征对模型预测的贡献有多大。这种方法可以有助于对模型的解释。

这些特征的重要性如图 5 所示。图 5 中,ax、ay、az、gx、gy、gz、agx、agy、agz 分别表示加速度 x 轴信号、加速度 y 轴信号、加速度 z 轴信号、角速度 x 轴信号、角速度 y 轴信号、角速度 z 轴信号、角 x 轴信号、角 y 轴信号、角 z 轴信号。排序后,传感器在随机森林预测中的重要性由大到小依次为 agx、az、agy、gx、agz、ay、ax、gz、gy; XGBoost 预测传感器重要性由大到小依次为 agx、az、agy、gx、gy、gz、ax、ay、agz; AdaBoost 预测传感器重要性的降序为 az、agx、agy、gy、ax、gx、ay、gz、agz; 梯度增强决策树预测的传感器重要性由高到低依次为 agx、az、gz、gx、agz、gy、agy、ax、ay。

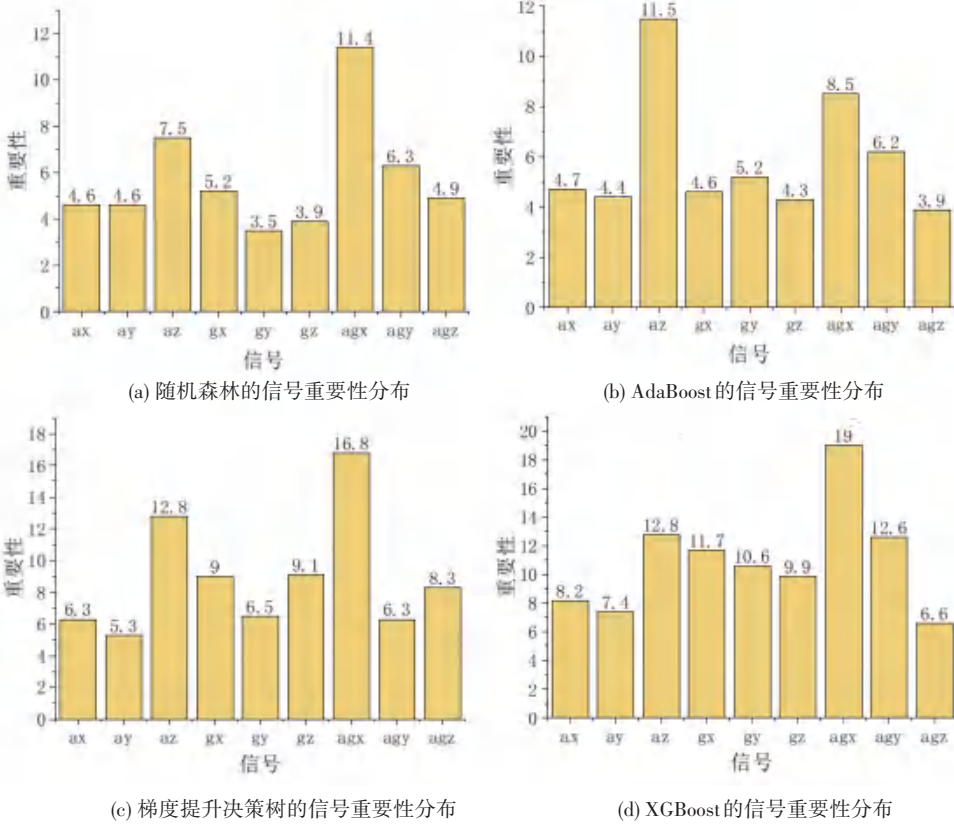


图 5 模型信号重要性分布

Fig. 5 Model signal importance distribution

将信号按重要性降序加入到模型中,进行 5 次交叉验证。每个信号相加得到一个疲劳等级识别精度,最后得到一个信号数量少、疲劳等级识别精度高的信号组合。

在此基础上,进一步得到,随机森林模型的精度

见表 6, XGBoost 模型的精度见表 7, AdaBoost 模型的精度见表 8, 梯度增强决策树的准确性见表 9。由表 6 可以看出,当信号个数大于 5 时,随机森林模型的准确率并没有明显提高。因此,随机森林模型的最佳信号数为 5,即选择 x 轴角度、z 轴加速度、y 轴

角度、 $x$  轴角速度和  $z$  轴角度信号。通过同样的方法,由表 7 可以看出,XGBoost 模型的信号个数为 4 个,分别选择了  $x$  轴角度、 $z$  轴加速度、 $y$  轴角度和  $x$  轴角速度信号。由表 8 可以看出,AdaBoost 模型信号个数为 7 个,分别选择  $z$  轴加速度、 $x$  轴角度、 $y$  轴角度、 $y$  轴角速度、 $x$  轴加速度、 $x$  轴角速度和  $y$  轴加速度信号。由表 9 可以看出,梯度提升树信号个数为 5 个,分别选择  $x$  轴角度、 $z$  轴加速度、 $z$  轴角度、 $y$  轴角度和  $y$  轴角速度信号。

表 6 随机森林模型的精度

Table 6 Accuracy of the Random Forest model

信号数量	平均值	误差	信号数量	平均值	误差
1	0.754	0.016	6	0.918	0.013
2	0.837	0.012	7	0.917	0.013
3	0.898	0.019	8	0.917	0.014
4	0.909	0.017	9	0.923	0.013
5	0.913	0.013			

表 7 XGBoost 模型的精度

Table 7 Accuracy of the XGBoost model

信号数量	平均值	误差	信号数量	平均值	误差
1	0.754	0.017	6	0.900	0.015
2	0.840	0.014	7	0.914	0.009
3	0.889	0.012	8	0.909	0.011
4	0.906	0.005	9	0.914	0.008
5	0.906	0.014			

表 8 AdaBoost 模型的精度

Table 8 Accuracy of the AdaBoost model

信号数量	平均值	误差	信号数量	平均值	误差
1	0.626	0.008	6	0.913	0.012
2	0.834	0.013	7	0.920	0.015
3	0.898	0.017	8	0.913	0.010
4	0.896	0.015	9	0.917	0.009
5	0.909	0.011			

表 9 梯度增强决策树的准确性

Table 9 Accuracy of the gradient boosting decision tree

信号数量	平均值	误差	信号数量	平均值	误差
1	0.626	0.008	6	0.915	0.011
2	0.762	0.006	7	0.913	0.013
3	0.898	0.011	8	0.914	0.010
4	0.906	0.015	9	0.917	0.009
5	0.913	0.013			

随机森林模型的最佳信号量为 5 向信号,5 倍交叉验证准确率为 91.3%;XGBoost 模型的最佳信

号为 4 通道信号,5 倍交叉验证准确率为 90.6%;AdaBoost 模型的最佳信号数为 7 个信号,5 倍交叉验证正确率为 92.0%;梯度增强树模型的最优信号数为 5 个信号,5 倍交叉验证准确率为 91.3%。

最后选择随机森林模型,该模型需要较少的信号( $x$  轴角度、 $z$  轴加速度、 $y$  轴角度、 $x$  轴角速度和  $z$  轴角度信号),并且具有较高的疲劳水平识别精度。同时又给出了 ExtraTree 准确率随样本变化如图 6 所示。

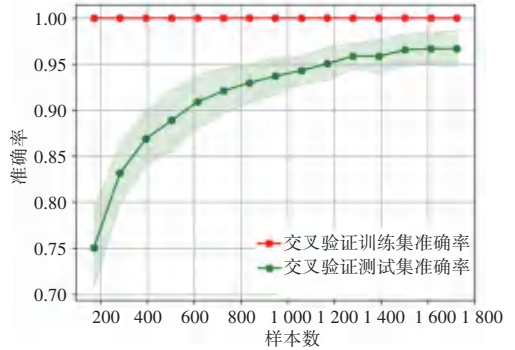


图 6 ExtraTrees 准确率随样本变化图

Fig. 6 ExtraTrees accuracy variation graph with samples

根据上述分析选择的模型和信号类型,再次对模型进行训练,得到测试集的混淆矩阵图和各类别预测结果的比例图。混淆矩阵图如图 7 所示。图 7 的横轴是预测类别,纵轴是真实类别。数值表示预测的数量。数值越大,颜色越深。每个类别中预测结果的比例如图 8 所示。图 8 中,横轴为实际类别,纵轴表示某一类别的预测结果占该类别数量的百分比。图 7 和图 8 显示,大部分预测结果是正确的。图 7 中对角线的颜色最暗,在对角线附近会出现错误预测,说明预测误差主要发生在真实类别附近,预测结果的可靠性较高。

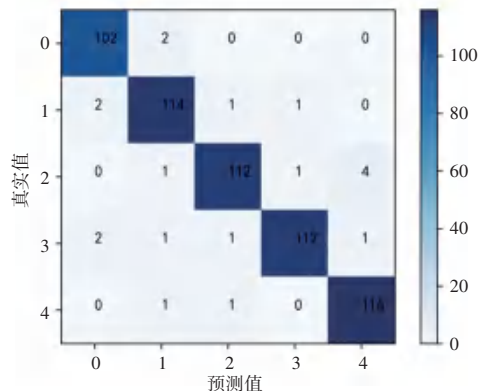


图 7 ExtraTrees 混淆矩阵

Fig. 7 Confusion matrix of ExtraTrees

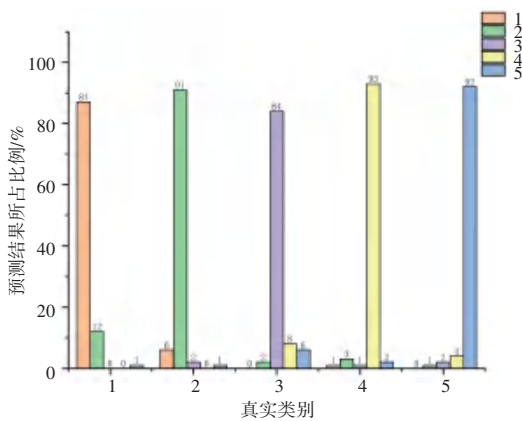


图8 预测结果在每个类别中的比例

Fig. 8 The proportion of prediction results in each category

### 3 讨论

实验结果表明,以最大耗氧量百分比作为疲劳等级的依据是合理的,可以有效降低个人因素的影响。通过采集足部信息和踝关节惯性信息,进行步态分割和特征提取,通过机器学习模型对疲劳程度进行准确识别。特别是经过信号筛选后,信号数量大大减少,疲劳识别的准确性保持在较高水平。

由于实验条件的限制,本次实验的参与者非常接近成年男性的年龄,体质也比较接近。在未来,所有年龄段的成年人都可以被邀请来进行实验。因此,实验结果证明了本文现有方法的有效性。

### 4 结束语

在本文中,研究了步态信息的疲劳检测潜力,利用压力传感器和惯性传感器信号来识别5种疲劳程度;该方法创造性地采用不运动和60%、70%、80%、90% VO所对应的运动强度作为疲劳等级,在一定程度上克服了个体差异对疲劳等级分类的影响。采用动态时间规整算法对步态数据进行分割,有效地将步态序列划分为多个单步态序列;采用简单的特征提取方法,将单个步态序列转化为最大、最小、均值和标准差等特征,降低了模型的复杂度。通过对多种分类模型的精度比较,发现随机森林模型对惯性传感器信号的疲劳程度识别精度与使用2种传感器信号的疲劳程度识别精度相差0.024,因此仅使用惯性传感器信号进行疲劳程度识别。除此之外,本文创新地采用信号Shapley值的平均绝对值来表示信号在疲劳等级识别中的重要性。最后,对惯性传感器信号进行进一步筛选,最终确定信号个数为5时,随机森林疲劳等级识别准确率为91.3%。

本文所采用的疲劳程度识别系统能够准确地实

时识别施工作业人员在行走过程中下肢5块肌肉的疲劳状态。这样不仅可以判断肌肉是否疲劳,还能准确检测出疲劳程度,适用于需要步行的高强度工作场所,如货物搬运,根据监测工人们疲劳程度进行合适的工作安排,可以提高工人的工作表现,为工人的安全提供了保证。

### 参考文献

- [1] CHOORUANG K, MANGKALAKEEREE P. Wireless heart rate monitoring system using MQTT[J]. *Procedia Computer Science*, 2016,86:160-163.
- [2] LIU Yonghuan, ZHANG Lei, WANG Yucheng. Wireless control design of intelligent window[J]. *Smart Factory*, 2016, 11: 88-90.
- [3] MA Liang, CHABLAT D, BENNIS F, et al. A new simple dynamic muscle fatigue model and its validation[J]. *International Journal of Industrial Ergonomics*, 2009,39(1): 211-220.
- [4] SILVA B A R S, MARTINEZ F G, PACHECO A M, et al. Effects of the exercise-induced muscular fatigue on the time of muscular reaction of the fibularis in healthy individuals [J]. *Revista Brasileira de Medicina do Esporte*, 2006,12: 85-89.
- [5] GERLACH K E, WHITE S C, BURTON H W, et al. Kinetic changes with fatigue and relationship to injury in female runners [J]. *Medicine and Science in Sports and Exercise*, 2005,37(4): 657-663.
- [6] TKACH D, HUANG He, KUIKEN T A. Study of stability of time-domain features for electromyographic pattern recognition [J]. *Journal of Neuroengineering and Rehabilitation*, 2010,7:21.
- [7] GUAITOLINI M, TRUPPA L, SABATINI A M, et al. Sport-induced fatigue detection in gait parameters using inertial sensors and support vector machines[C]//*Proceedings of 2020 8th IEEE RAS/EMBS International Conference for Biomedical Robotics and Biomechanics (BioRob)*. Piscataway, NJ: IEEE, 2020: 170-174.
- [8] ZHANG Jian, LOCKHART T E, SOANGRA R. Classifying lower extremity muscle fatigue during walking using machine learning and inertial sensors [J]. *Annals of Biomedical Engineering*, 2014,42: 600-612.
- [9] CHOI H R, KIM T Y. Directional dynamic time warping for gesture recognition [C]// *Proceedings of the 2nd International Conference on Multimedia Systems and Signal Processing*. New York: ACM, 2017: 22-25.
- [11] GIAO B C, ANH D T. An application of similarity search in streaming time series under DTW: Online forecasting [C]// *Proceedings of the 8th International Symposium on Information and Communication Technology*. New York: ACM, 2017: 10-17.
- [11] YONEKAWA K, YONEZAWA T, NAKAZAWA J, et al. FASH: Detecting tiredness of walking people using pressure sensors[C]// *Proceedings of the 6th Annual International Mobile and Ubiquitous Systems: Networking & Services*. Piscataway, NJ: IEEE, 2009: 1-6.
- [12] LIU Jun, CHEN Jianhui, YE Jieping. Large-scale sparse logistic regression [C]// *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York: ACM, 2009: 547-556.

- [13] AKAY M F. Support vector machines combined with feature selection for breast cancer diagnosis [J]. *Expert Systems with Applications*, 2009, 36(2): 3240–3247.
- [14] BELL J. *Machine learning for big data: Hands-on for developers and technical professionals* [M]. New York: Wiley, 2015.
- [15] BATAINEH A A. A comparative analysis of nonlinear machine learning algorithms for breast cancer detection [J]. *International Journal of Machine Learning*, 2019, 9(3): 248–254.
- [16] MITCHELL T M. *Machine Learning* [M]. Beijing: China Machine Press, 2003.
- [17] BAUMANN F, EHLERS A, ROSENHAHN B, et al. Sequential boosting for learning a random forest classifier [C]// *Proceedings of 2015 IEEE Winter Conference on Applications of Computer Vision*. Piscataway, NJ: IEEE, 2015: 442–447.
- [18] ZHAO Xiaowei, LI Xiangtao, MA Zhiqiang, et al. Prediction of lysine ubiquitylation with ensemble classifier and feature selection [J]. *International Journal of Molecular Sciences*, 2011, 12(12): 8347–8361.
- [19] GARDNER M W, DORLING S R. Artificial neural networks (the multilayer perceptron) – a review of applications in the atmospheric sciences [J]. *Atmospheric Environment*, 1998, 32(14–15): 2627–2636.
- [20] YAN Hongmei, JIANG Yingtao, ZHENG Jun, et al. A multilayer perceptron-based medical decision support system for heart disease diagnosis [J]. *Expert Systems with Applications*, 2006, 30(2): 272–281.
- [21] PEDREGOSA F, VAROQUAUX G, GRAMFORT A, et al. *Learning scikit-learn: Machine learning in python* [J]. *The Journal of Machine Learning Research*, 2011, 12: 2825–2830.
- [22] BUITINCK L, LOUPPE G, BLONDEL M, et al. API design for machine learning software: Experiences from the scikit-learn project [J]. *arXiv preprint arXiv, 1309.0238*, 2013.
- [23] CHEN Tianqi, GUESTRIN C. XGBoost: A scalable tree boosting system [J]. *arXiv preprint arXiv, 1603.02754*, 2016.
- [24] BUNCE J D, PATTERSON D E, FRANK I E. Crossvalidation, bootstrapping, and partial least squares compared with multiple regression in conventional QSAR studies [J]. *Quantitative Structure-Activity Relationships*, 1988, 7(1): 18–25.
- [25] SHORROCKS A F. Decomposition procedures for distributional analysis: A unified framework based on the Shapley value [J]. *The Journal of Economic Inequality*, 2013, 11(1): 99–126.