

基于深度森林的下肢股直肌疲劳检测算法

卢小龙¹, 郭旭东^{1*}, 郝又国^{2*}, 高体杰¹, 王 鹏¹, 周绣华¹

¹上海理工大学健康科学与工程学院, 上海

²上海市普陀区人民医院, 上海

收稿日期: 2022年1月5日; 录用日期: 2022年2月25日; 发布日期: 2022年3月7日

摘 要

肌肉疲劳评估是当今社会探索与研究的热点之一, 与人们日常生活健康密切相关。目前常用的肌疲劳分类方法有支持向量机(Support Vector Machine, SVM)、随机森林(Random Forest, RF)和逻辑回归等浅层机器学习方法。为了进一步提高运动过程中肌肉疲劳识别的精度, 因此, 本文通过设计硬件采集装置进行下肢股直肌部位肌肉疲劳实验采集数据, 构建深度森林(Deep Forest, DF)模型将预处理后的表面肌电(sEMG)数据分成三类: 正常(normal)、疲劳(fatigue)和极度疲劳(extreme fatigue), 并与传统方法SVM和RF模型对比。实验结果表明: 深度森林模型的疲劳识别效果最好, 整体准确率为92%, 所提出的方法在各评价指标上均优于传统方法, 基于深度森林的下肢股直肌疲劳检测算法, 可以作为肌肉疲劳识别的方法, 为人类健康提供参考与帮助。

关键词

肌疲劳, 股直肌, sEMG, 深度森林

Fatigue Detection Algorithm of Lower Limb Rectus Femoris Muscle Based on Deep Forest

Xiaolong Lu¹, Xudong Guo^{1*}, Youguo Hao^{2*}, Tijie Gao¹, Peng Wang¹, Xiuhua Zhou¹

¹School of Health Science and Engineering, University of Shanghai for Science and Technology, Shanghai

²Shanghai Putuo District People's Hospital, Shanghai

Received: Jan. 5th, 2022; accepted: Feb. 25th, 2022; published: Mar. 7th, 2022

Abstract

Muscle fatigue assessment is one of the most popular research filed recently, and it is closely re-
*通讯作者。

文章引用: 卢小龙, 郭旭东, 郝又国, 高体杰, 王鹏, 周绣华. 基于深度森林的下肢股直肌疲劳检测算法[J]. 建模与仿真, 2022, 11(2): 297-305. DOI: 10.12677/mos.2022.112027

lated to people's daily health. At present, the commonly used classification methods of muscle fatigue include Support Vector Machine (SVM), Random Forest (RF), logistic regression and other shallow machine learning methods. In order to further improve the accuracy of muscle fatigue recognition, this paper designs a hardware acquisition device to collect data from the muscle fatigue experiment of the rectus femoris of the lower limbs, and constructs a deep forest model to divide the preprocessed surface electromyography (sEMG) data into three categories: normal, fatigue and extreme fatigue, compared with the traditional methods of SVM and RF models. The experimental results show that the deep forest model has the best fatigue recognition effect, with an overall accuracy rate of 92%. The proposed method is superior to traditional methods in all evaluation indicators. The deep forest-based fatigue detection algorithm for the rectus femoris of the lower limbs can be used as a method of muscle fatigue recognition to provide reference and help for human health.

Keywords

Muscle Fatigue, Rectus Femoris, Surface Electromyography (sEMG), Deep Forest

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

表面肌电信号(sEMG)是人体在运动过程中,由运动神经元产生的动作电位序列叠加的生物电信号,其变化规律反应了肌肉部位的生理状态和动作模式。目前对肌电信号研究主要分为肌肉疲劳识别和肌肉运动模式识别,王宏鹏等[1]通过收集并分析驾驶员颈部肌电数据,发现随着驾驶时间的增长,驾驶员颈部肌电特征时域特征值呈上升趋势,频域特征值呈下降趋势,驾驶员行驶约1小时后会进入疲劳状态;隋修武等[2]提出了基于非负矩阵分解与支持向量机联合模型来识别上肢sEMG手势,通过样本分离实验表明准确率高达91.7%以上;刘光达等[3]采用回归分析和皮尔逊相关系数分析的方法证实均方根(RMS)与中值频率(MF)相结合作为评估手臂肌肉疲劳的联合参数,更加合理准确;盛鑫军等[4]设计了一款集成多模态信号采集系统,能够同时采集并处理电生理信号sEMG、近红外光谱(NIRS)血氧变化以及肌电图(MMG)变化情况,使肌肉疲劳检测更加精确。

本文设计硬件采集设备实现高精度采集sEMG信号,通过下肢股直肌肌疲劳实验收集受试者的肌电数据,对sEMG数据预处理后计算各特征值,构建深度森林模型对数据进行正常(normal)、疲劳(fatigue)和极度疲劳(extreme fatigue)三分类,其分类效果优于传统的机器学习算法。本文的总体算法流程如图1所示,首先自主设计出一款肌电信号的硬件采集设备,接着采集大量的肌电信号数据,通过一系列预处理算法进行去噪以及去除连续微弱的无用信号,再通过大小滑动窗口进行信号特征的提取,最后将提取到的特征用深度森林算法进行预测,可以达到对肌电信号的疲劳等级自动分类的目的。

2. 数据采集与处理

2.1. 硬件部分

sEMG信号是十分微弱的电信号之一,容易受到外界电磁、静电的干扰,其幅值为0~5mv,频率范围在10~500Hz之间。因此需要设计硬件电路采集该信号,采集设备的总体框图如图2所示,肌电

信号经电极进入 ADS1292R 可编程放大器模块, 实现信号放大和滤波, ADS1292R 是两通道的同步采样 24 位模数转换器, 内置的可编程增益放大器(PGA)、内部基准和板载振荡器; 通过微处理器 stm32 编译采样频率 2000 Hz, 放大倍数 12 倍; 再进入后级放大电路将信号放大 84 倍, 后级放大电路选用仪表放大器 AD8221 芯片, 该芯片具有低输入失调电压, 低压摆率和低噪声等特性。通过无线串口发送模块将信号发送至 PC 上位机端, 无线模块使用 433 M 的无线串口 HC-12, 采用 SI4438 芯片, 其最大发射功率为 100 mW, 5000 bps 空中波特率下接收灵敏度为-116 dBm, 实现远距离之间的数据传输。之后将采集的 sEMG 数据传输至上位机进一步处理分析, 通过基于深度森林的下肢股直肌疲劳检测算法对肌肉状态精准分类。

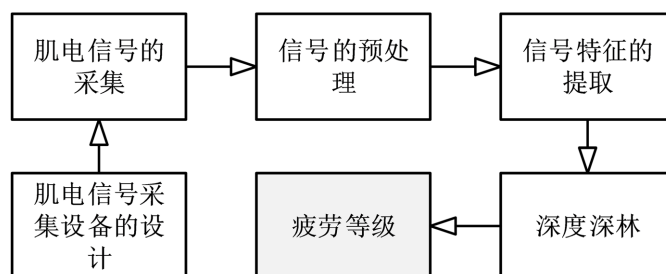


Figure 1. Overall flow chart

图 1. 总体流程图

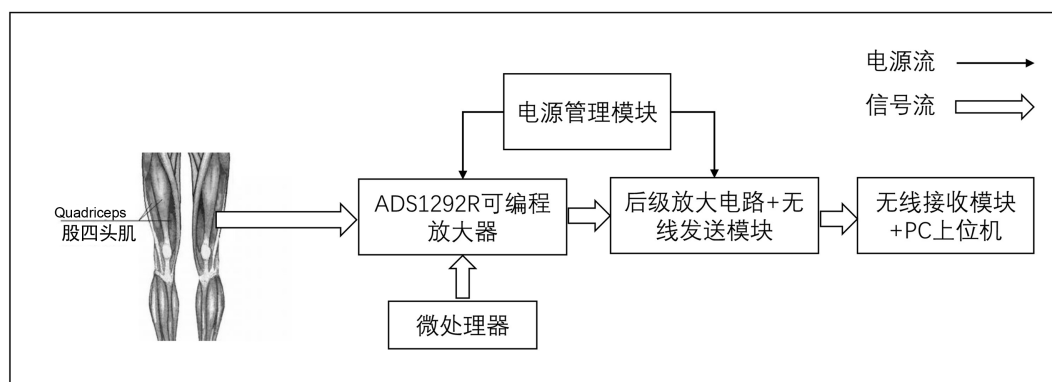


Figure 2. Overall block diagram of the hardware acquisition device

图 2. 硬件采集设备总体框图

2.2. 下肢股直肌疲劳实验

受试者选择年龄 21~24 岁、身高 165~180 cm、体重 55~70 kg 且 24 h 内无剧烈运动的 23 名大学生进行下肢肌疲劳实验。其中男性成员 15 名, 女性成员 8 名。股四头肌是维持下肢屈伸、站立的重要肌肉之一, 主要包括股直肌、股外肌、股中肌和股内侧肌[5]。考虑到运动过程中疲劳实验数据的方便采集, 本文实验的电极放置部位为大腿股直肌处, 采集电极选择氯化银贴片电极, 输入电极采用三点式差分布局方式。采样频率设置为 2000 Hz, 疲劳实验采集前先用酒精擦拭受试者采集部位, 去除皮肤表面油脂和皮屑。本次实验采用模拟骑行阻尼自行车的方式进行肌电数据采集, 具体实验过程如图 3 所示, 开始前受试者需保持站立状态, 采集受试者在非疲劳状态下的肌电数据; 然后调节阻尼自行车参数, 让受试者全程保持匀速骑行, 根据受试者主观感受并参考主观疲劳量表(RPE) [6] [7], 当受试者股直肌部位肌肉开始感到疲劳时, 继续骑行半分钟, 确保此时的肌肉状态为疲劳状态, 采集该状态股直肌部位的肌电数据,

定义为疲劳状态下肌电数据；之后让受试者继续骑行阻尼自行车致力竭无法坚持，采集该状态下股直肌部位的肌电数据，并且定义为极度疲劳状态下的肌电数据，将所有数据保存待下一步处理。

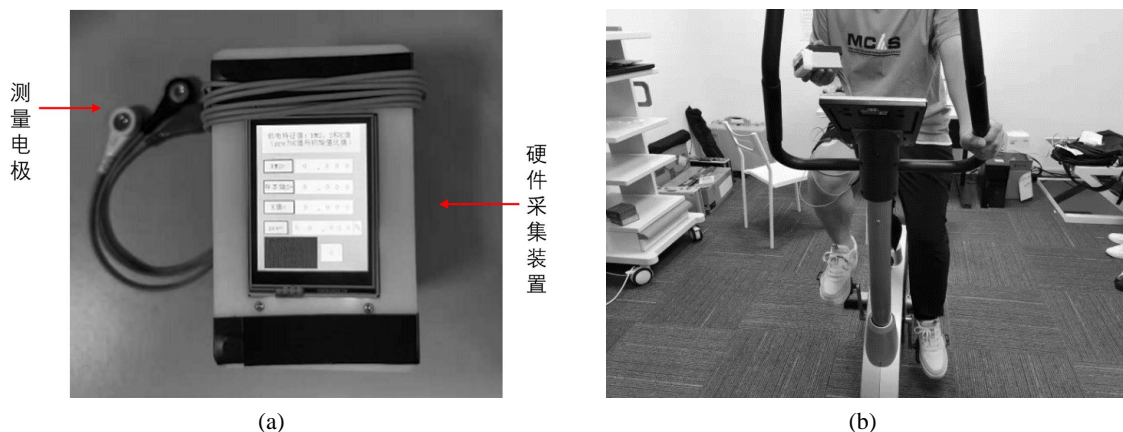


Figure 3. Lower extremity muscle fatigue experiment. (a) sEMG acquisition equipment; (b) Simulated cycling damping bicycle muscle fatigue experiment

图 3. 下肢肌疲劳实验。(a) 肌电采集设备；(b) 模拟骑行阻尼自行车肌疲劳实验

2.3. 数据预处理

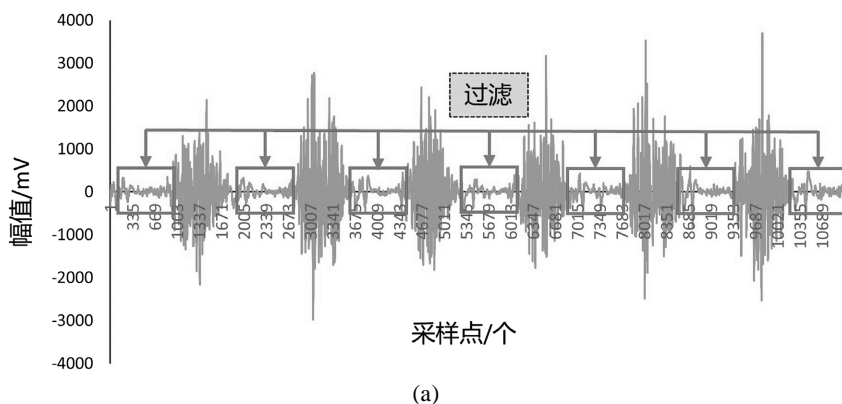
经过硬件采集得到的 sEMG 信号包含基线漂移、高斯白噪声。需进一步去噪处理，因此使用 10~500 Hz 的 64 阶巴特沃斯带通滤波和小波阈值去噪算法滤除噪声干扰，其中，小波阈值去噪采用 Stein 的无偏似然估计原理获得自适应阈值，选择风险最小的阈值，算法公式具体如下：

$$\eta_\lambda = \frac{\sum_{\lambda=1}^N |x_\lambda|}{(0.6745 \cdot N)} \sqrt{x_\lambda^2}, (\lambda=1, \dots, N) \tag{1}$$

$$\text{risk}(\lambda) = \frac{N - 2\lambda + \sum_{i=1}^{\lambda} x_i^2 + (N - \lambda) \cdot x_\lambda^2}{N} \tag{2}$$

其中， N 为肌电序列长度， x_i 为肌电数据， η_λ 为第 λ 元素的平方根作为阈值， $\text{risk}(\lambda)$ 为第 λ 个元素阈值对应的风险值，最终根据风险函数曲线得到最小风险对应的 η_λ 作为阈值。

同时通过局部滑动窗计算方差值对肌电信号进行不等长动作分割，消除平稳无用信息的干扰，起到对数据降维的效果，得到序列长度为 5000 个采样点的 sEMG 数据。其去噪效果如图 4 所示。



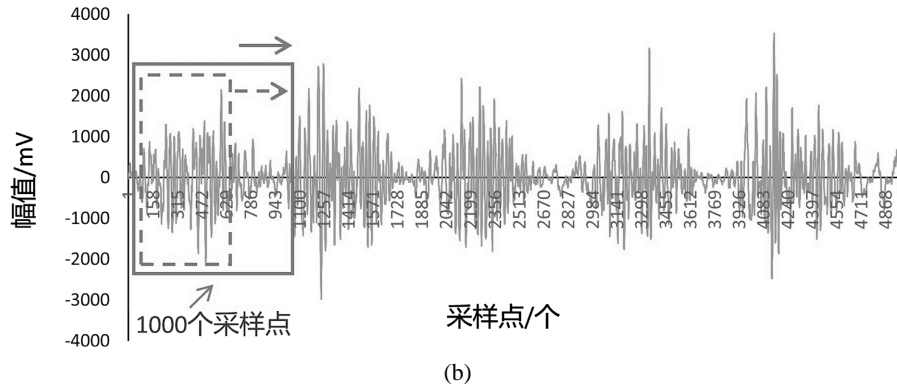


Figure 4. Preprocessing of sEMG signal. (a) Raw EMG signal; (b) Signal after denoising and motion segmentation
图 4. sEMG 信号预处理。(a) 原始肌电信号；(b) 去噪与动作分割后信号

sEMG 信号的时域特征参数有肌电积分值(IEMG)和均方根值(RMS)，主要反映了肌电信号幅值、有效值变化，与肌肉部位的肌张力和肌肉质量相关[8]。具体计算公式如下：

$$IEMG = \frac{1}{N} \sum_{i=0}^{N-1} |x_i| \quad (3)$$

$$RMS = \sqrt{\frac{\sum_{i=0}^{N-1} (x_i - x')^2}{N}} \quad (4)$$

其中 x' 为肌电信号序列的平均值， N 为序列长度， x_i 代表每一个肌电采样点。

sEMG 信号的频域特征参数有中值频率(MF)和平均功率频率(MPF)，主要反映了 sEMG 信号的功率谱能量和频率变化情况，进而分析疲劳过程中运动单元的募集情况[9]。计算公式如下：

$$MF = \frac{1}{2} \int_0^{\infty} P(f) df \quad (5)$$

$$MPF = \frac{\int_0^{\infty} fP(f) df}{\int_0^{\infty} P(f) df} \quad (6)$$

其中 $P(f)$ 为肌电序列的功率谱密度函数。

sEMG 信号的非线性特征参数有近似熵(ApEn)和样本熵(sEn)，主要反映 sEMG 信号复杂程度。

根据图 3(b)可知肌电信号的一个动作起始点为 1000 个采样点，因此采用大小窗口技术，其中大窗口包含 1000 个采样点，小窗口包含 500 个采样点，滑动的步长 f 为 500，以重叠率 μ 为 50% 的方式滑动小窗口处理数据，这样一个大窗口包含了 3 个小窗口。对小窗口数据进行特征参数计算，得到 IEMG、RMS、MF、MPF 和 sEn 的 5 个特征值。由于一个大窗口包含了 sEMG 信号的一个动作电位的基本特征，对该肌肉部位的状态具有表征意义，将其划分为一个数据集 Φ ，包含的特征个数 ϕ 计算公式如下：

$$\phi = \left(\frac{1000}{f * \mu} - 1 \right) * 5 \quad (7)$$

经过计算 ϕ 等于 15，这样不仅对数据进行了降维，同时扩增了数据集。最后将所有受试者正常(normal)、疲劳(fatigue)和极度疲劳(extreme fatigue)肌肉状态的数据分布绘制成柱状图如图 5，可知正常状态下的数据容易识别，疲劳与极度疲劳的特征比较相近，存在一定的分类难度。

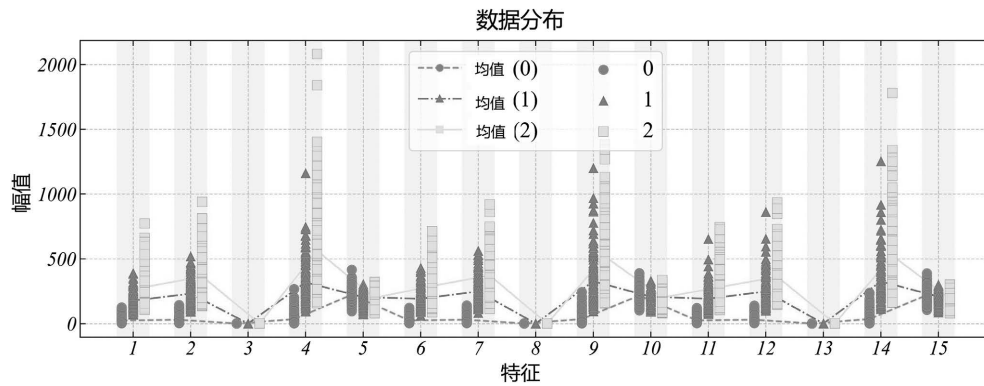


Figure 5. Data distribution of the three muscle states of all subject
 图 5. 所有受试者三种肌肉状态的数据分布情况

3. 分析方法

3.1. 随机森林

随机森林(Random Forest, RF)是一种将 bagging 集成学习理论与随机子空间方法相结合的集成学习模型[10]。其具体实现过程如图 6 所示，首先通过 bootstrap 方法对原始数据集 C 进行有放回的 m 次随机采样，构成 N 个含不同样本数 m 的采样集 $C_k (k=1,2,\dots,N)$ ，根据抽样的原理，采样集 C_k 每次抽样过程中，任何一个样本未被选中的概率为： $p=1-1/N$ ，抽样 N 次后任何一个样本进入 C_k 的概率为 $1-p^N$ ，当 N 趋于无穷时，这个概率为 63.2%，因此得到的每个决策树的采样集都不完全相同，避免了数据过拟合问题。最后由 N 颗决策树投票决定分类的结果。

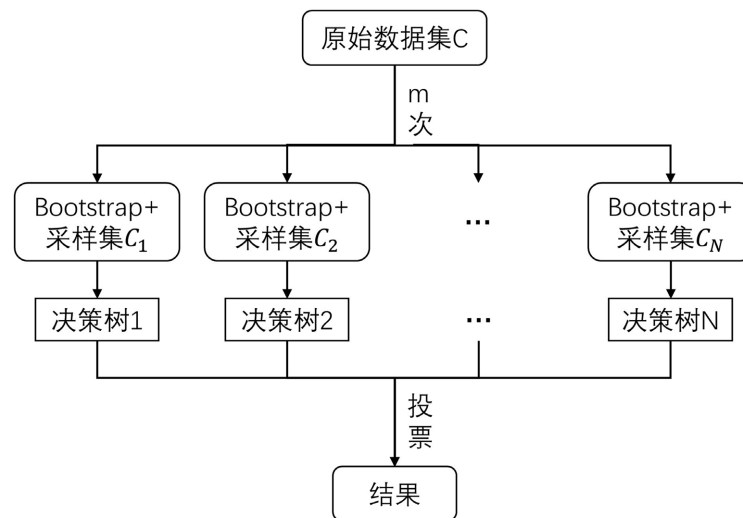


Figure 6. RF algorithm implementation process
 图 6. RF 算法实现流程

3.2. 深度森林

深度随机森林(Deep Forest, DF)是 2017 年提出的深度学习新算法，在多种复杂分类数据处理上有良好的效果。训练速度与模型鲁棒性极大提升，并且能够适用于不同规模的训练集数据分类，解决了小样本数据的分类问题。

深度森林的结构主要包括级联森林(Cascade Forest)和多粒度扫描(Multi-Grained Scanning) [11]。本文基于深度森林的下肢肌疲劳检测算法具体实现流程如图 7 所示，其中多粒度扫描是为了增强级联森林，对输入特征进行多尺度扫描，获得更多的、更丰富的特征关系向量。肌电信号的输入特征向量分别为 IEMG、RMS、MF、MPF 与样本熵 sEn，将肌电信号每三组特征向量，即 15 个特征作为输入，分别使用大小为 2 和 3 的窗口滑动扫描，可以得到 13 个 3 维特征向量与 14 个 3 维特征向量，将得到的不同尺度的特征进行拼接，最终获得 162 维多尺度特征向量。之后进入级联森林，它是在随机森林的基础上进一步集成学习，每一层由大量随机森林构成，每个随机森林又有大量的决策树构成，从而保证了模型的多样性；最终每一个随机森林决策出结果，取均值后并选择其中的最大值，其对应的类别作为模型的预测结果，级联森林的迭代终止条件为决策效果不能提升则结束。

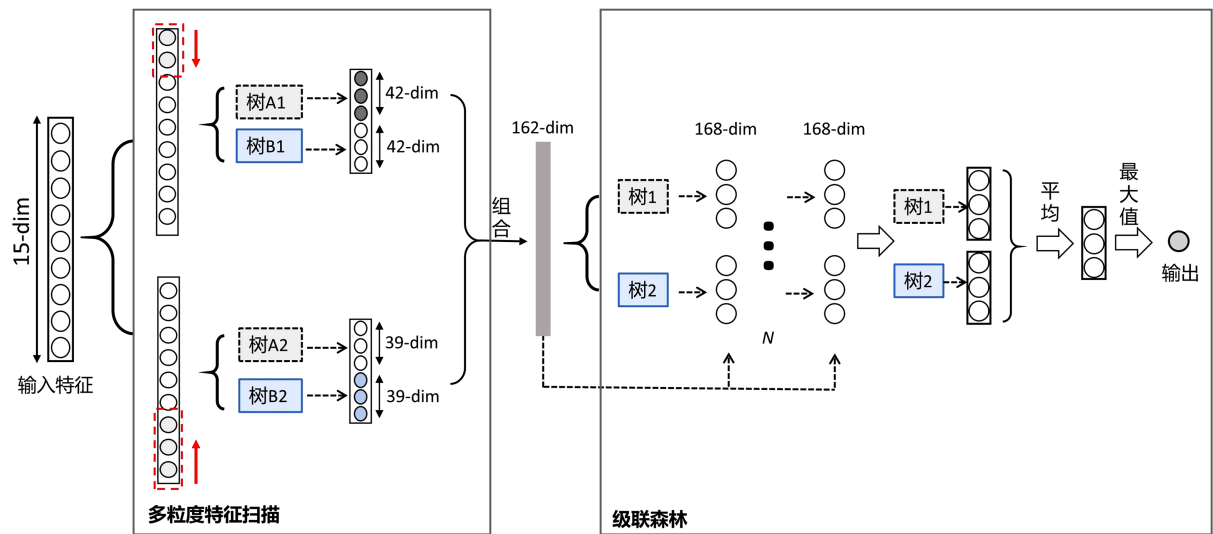


Figure 7. Deep forest implementation process
图 7. 深度森林实现流程

3.3. 模型评价指标

本文将原始 sEMG 数据去噪后经窗口函数滑动处理分别计算了 IEMG、RMS、MF、MPF 和样本熵 sEn 五个特征参数作为模型输入参数，模型分类结果为正常、疲劳和极度疲劳三类。

为了评估本文模型对下肢肌肉状态的分类能力，采用准确率(Accuracy)、精度(Precision)、召回率(Recall)和 F1 分数(F1-score)、混淆矩阵作为模型性能评估指标。其中 Accuracy、Precision、Recall 和 F1-score 指标的计算公式如下：

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + FP + TN + FN)} \times 100\% \quad (8)$$

$$\text{Precision} = \frac{TP}{(TP + FP)} \times 100\% \quad (9)$$

$$\text{Recall} = \frac{TP}{(TP + FN)} \times 100\% \quad (10)$$

$$\text{F1-score} = \frac{(2 \times \text{Precision} \times \text{Recall})}{\text{Precision} + \text{Recall}} \times 100\% \quad (11)$$

TP 代表真阳性样本数；FP 代表假阳性样本数；FN 代表假阴性样本数；TN 代表真阴性样本数[12]。混淆矩阵通过可视化算法直观展示模型分类性能，矩阵列代表预测值，行代表的是实际分类值，在预测分析时，从表格对角线结构中可以清楚地观察分类准确性。

4. 结果

为了验证本文所提方法的有效性，将预处理后的数据按照 8:2 划分为训练集与测试集。测试集实验结果如表 1 所示，三个模型对肌肉状态的识别都达到了 80% 以上的准确度，实现了肌肉不同状态的自动识别。三种模型都能很好地区分肌肉正常状态，其 F1-score 均为 100%，但其中 DF 模型对于识别疲劳与极度疲劳状态的准确度明显高于其他两种模型，分别为 87% 与 88%，总体准确度也高达 92%，这表明基于深度森林的下肢股直肌疲劳检测算法在各项指标上呈现一定的优势。

模型测试结果的混淆矩阵如图 8，可以直观地看出 DF 模型在识别疲劳和极度疲劳的误判率最低，识别精度明显优于其他模型。

Table 1. Classification results of lower limb muscle state
表 1. 下肢肌肉状态分类结果

SVM	Precision	Recall	F1 score	Accuracy
Normal	100%	100%	100%	
fatigue	88%	56%	0.77	
extreme fatigue	69%	93%	79%	
Total				84%
RF	Precision	Recall	F1 score	Accuracy
Normal	100%	100%	100%	
fatigue	83%	70%	76%	
extreme fatigue	76%	86%	81%	
Total				86%
DF	Precision	Recall	F1 score	Accuracy
Normal	100%	100%	100%	
fatigue	89%	85%	87%	
extreme fatigue	86%	90%	88%	
Total				92%

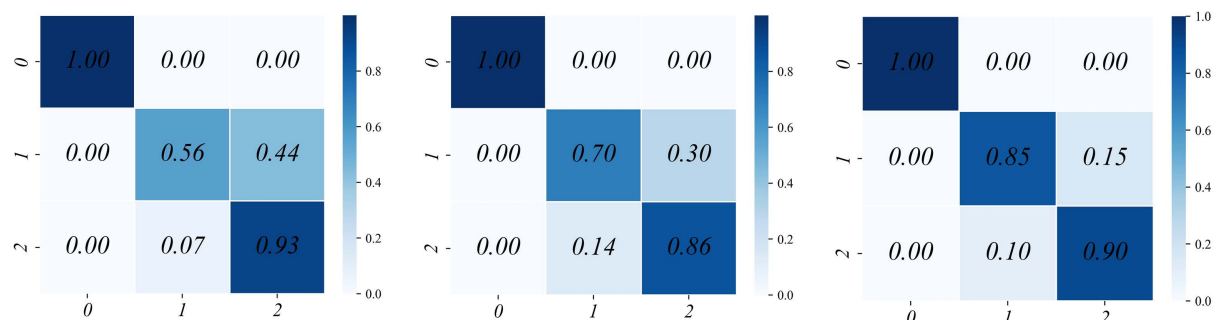


Figure 8. Comparison of the confusion matrix results of the network
图 8. 模型测试混淆矩阵结果的对比

5. 结论

本文针对下肢股直肌部位肌肉疲劳状态识别, 自行设计硬件采集装置, 进行模拟骑行阻尼自行车疲劳实验, 采集相应 sEMG 数据, 并对数据预处理后将肌电信号各特征参数作为 DF 模型的输入; 同时为了验证 DF 模型对肌电数据集分类的精度, 对比了支持向量机与随机森林两种机器学习方法, DF 模型的准确率高达 92%, 明显优于其他两种模型。通过 DF 模型可以对下肢肌肉状态进行识别, 提高肌肉疲劳状态的识别分类精度, 将深度森林与肌肉状态分类结合是肌疲劳识别研究领域的一项重要探索。在未来可通过增加不同运动类型下的疲劳实验, 扩充 sEMG 数据集, 进一步探究 DF 模型对肌肉状态的分类。

致 谢

感谢我的导师、同学以及基金项目包括上海市自然科学基金面上项目(20ZR1437700)和上海市产业协同创新项目(2021-cyxt1-kj07)的支持与帮助。同时, 本文实验数据是在本校同学之中进行实验采集的, 非常感谢大家的参与。

基金项目

上海市自然科学基金面上项目(20ZR1437700); 上海市产业协同创新项目(2021-cyxt1-kj07)。

参考文献

- [1] 王宏鹏, 艾力·斯木吐拉. 基于表面肌电信号的高原公路驾驶员颈部疲劳试验分析[J]. 科学技术与工程, 2016, 16(26): 323-328.
- [2] 隋修武, 牛佳宝, 李昊天. 基于 NMF-SVM 模型的上肢 sEMG 手势识别方法[J]. 计算机工程与应用, 2020, 56(17): 161-166.
- [3] 刘光达, 董梦坤, 许蓝予. 手臂疲劳时表面肌电信号特征[J]. 科学技术与工程, 2021(25): 10690-10696.
- [4] Sheng, X., Ding, X., Guo, W., *et al.* (2020) Toward an Integrate Multi-Modal sEMG/MMG/NIRS Sensing System for Human-Machine Interface Robust to Muscular Fatigue. *IEEE Sensors Journal*, **PP**, 1. <https://doi.org/10.1109/JSEN.2020.3023742>
- [5] 陈佳旭, 谢志明, 蒋拥军. 肌肉电刺激训练对伸膝迟滞患者股内侧肌表面肌电的影响[J]. 中国实用医药, 2021, 16(22): 96-98.
- [6] Stggl, T. and Born, D.P. (2021) Near Infrared Spectroscopy for Muscle Specific Analysis of Intensity and Fatigue during Cross-Country Skiing Competition—A Case Report. *Sensors*, **21**, 2535. <https://doi.org/10.3390/s21072535>
- [7] 程宁波, 吴志明, 柯思成. 服装压力对男子骑行运动中肌肉疲劳的影响[J]. 纺织学报, 2019, 40(8): 130-135.
- [8] Waris, M.A., Rehman, M.Z.U., Niazi, I.K., *et al.* (2020) A Multiday Evaluation of Real-Time Intramuscular EMG Usability with ANN. *Sensors*, **20**, 3385. <https://doi.org/10.3390/s20123385>
- [9] 杨铮, 王立玲, 马东. 基于自回归模型表面肌电信号检测肌肉疲劳研究[J]. 中国生物医学工程学报, 2018, 37(6): 673-679.
- [10] 王鑫, 汪晋宽, 刘志刚. 基于随机森林的认知模型频谱感知算法[J]. 仪器仪表学报, 2013, 34(11): 2471-2477.
- [11] 李志远, 黄亦翔, 刘成良. 基于经验模态分解与深度森林的液压泵健康评估[J]. 机械与电子, 2020, 38(5): 3-8.
- [12] 吴青洋, 程旭, 邓程鹏. 基于聚类和奖惩用户模型的协算法[J]. 计算机系统应用, 2020, 29(8): 135-143.