華東理工大學学报 自然科学版 Natural Science Edition
JOURNAL OF EAST CHINA UNIVERSITY OF SCIENCE AND TECHNOLOGY

基于机器学习的奶牛发情识别方法研究

罗小娟 王道龙 胡鹏昊 林佳旭 管磊

Machine Learning-Based for Detecting Estrus in Dairy Cows

LUO Xiaojuan, WANG Daolong, HU Penghao, LIN Jiaxu, GUAN Lei

在线阅读 View online: http://doi.org/10.14135/j.cnki.1006-3080.20260211001

您可能感兴趣的其他文章

Articles you may be interested in

基于机器学习的酸烃界面预测与离子液体设计

Machine Learning-Based Acid-Hydrocarbon Interface Prediction and Ionic Liquids Design

华东理工大学学报（自然科学版）. 2025, 51(6): 793-803 http://doi.org/10.14135/j.cnki.1006-3080.20250121001

基于机器学习的碳酸盐微观结构与性质

Microstructure and Properties of Carbonate Salt Based on Machine Learning

华东理工大学学报（自然科学版）. 2022, 48(4): 456-466 http://doi.org/10.14135/j.cnki.1006-3080.20210331004

混合型抗机器学习攻击的强PUF电路设计

Design of Hybrid Strong PUF Circuit Against Machine Learning Attacks

华东理工大学学报（自然科学版）. 2023, 49(6): 854-861 http://doi.org/10.14135/j.cnki.1006-3080.20221009003

基于单模态生理信号无监督特征学习的驾驶压力识别

Driving Stress Detection Based on Unsupervised Feature Learning of Single Module Physiological Signal

华东理工大学学报（自然科学版）. 2021, 47(4): 475-482 http://doi.org/10.14135/j.cnki.1006-3080.20200728001

基于深度学习的农场虫情检测算法研究及实现

Research and Implementation of Farm Insect Detection Algorithm Based on Deep Learning

华东理工大学学报（自然科学版）. 2024, 50(5): 732-739 http://doi.org/10.14135/j.cnki.1006-3080.20231007001

人工神经网络和响应面法优化黑曲霉发酵产淀粉酶

Optimization of Amylase Production in *Aspergillus niger* Fermentation by Artificial Neural Network and Response Surface Method

华东理工大学学报（自然科学版）. 2024, 50(6): 840-849 http://doi.org/10.14135/j.cnki.1006-3080.20231209001

扫码关注公众号，获取更多信息！

文章编号：1006-3080(2026)05-0725-10 DOI：10.14135/j.cnki.1006-3080.20260211001

基于机器学习的奶牛发情识别方法研究

罗小娟， 王道龙， 胡鹏昊， 林佳旭， 管 磊
（华东理工大学信息科学与工程学院，上海 200237）

摘要：针对奶牛发情监测中人工观察获取标签成本高、主观性强，以及模型特征维度高导致的计算功耗大的问题，本文提出一种基于无监督聚类与可解释性特征选择的发情识别方法。该方法通过K-means对奶牛颈部佩戴的加速度计采集的数据进行聚类，生成发情标签。通过对比K近邻(K-Nearest Neighbors, KNN)、支持向量机(Support Vector Machine, SVM)与随机森林(Random Forest, RF)3种模型在10、30 s和50 s窗口下的性能，结合Friedman检验和Nemenyi检验确定最佳模型。进一步结合沙普利加性解释(Shapely Additive Explanations, SHAP)量化模型特征贡献度，筛选出最优特征子集。研究结果表明，基于10 s时间窗口的随机森林模型性能最优，经SHAP方法筛选后，特征从25维降至3维，且降维后的模型性能未出现显著性差异，其准确率、精确率、召回率和F1-score分别达到0.991 1、0.937 0、0.995 5、0.965 4。在基于奶牛三轴加速度数据的独立泛化试验中，降维后10 s窗口的RF模型的发情检出准确率达到0.909 0。本文提出的方法实现了发情标签的生成与特征的降维，且模型具有可解释性，为研发低功耗、长续航的奶牛发情监测设备提供了技术支撑。

关键词：机器学习；奶牛发情识别；加速度传感器；K-means聚类；SHAP

中图分类号：TP391, S823 **文献标志码：**A

奶牛平均发情周期为21 d左右[1]，一旦错过则需在下一个发情周期进行配种，增加了非繁殖饲养成本，故发情检测对牧场管理至关重要。传统的发情检测方法主要依靠人工进行观察，该方法存在人工检测成本高、检出效率低、准确率低等问题，容易漏检和发情时间鉴别不准，错过最佳配种时间[2]，这难以满足现代牧场需求。随着物联网与人工智能技术的发展，自动化的发情检测方法已经成为研究热点。

目前，国内外关于奶牛发情的研究主要分为基于奶牛活动量、体温等数据的直接接触方法和基于图像、视频等数据的非接触式方法[3]。Wang等[4-5]使用可佩戴颈部声学标签获取奶牛声音，并提出一种基于最优组合特征和最优时间窗的发情鉴定算法和基于最佳特征组合的双LSTM(Long Short-Term Memory)模型来检测发情，证明了声音检测可以有效检测出奶牛发情。但声学检测需要设备具备较高的采样频率，会导致原始数据量庞大，同时声音信号容易受到背景噪声的干扰。

基于图像与视频的奶牛发情检测主要围绕检测爬跨行为展开。Arıkan等[6]将增强现实技术应用于奶牛发情检测和识别，提出基于深度学习的贴装检测并使用YOLO v5识别奶牛，该模型检测爬跨行为准确率达到99%。王政等[7]提出了一种融合YOLO v5n与通道剪枝算法的轻量化奶牛发情行为识别方法，在泛化性试验中，该模型实现了84.62%的奶牛爬跨行为检出率。Tsai等[8]提出了一种针对奶牛爬跨行为的图像识别算法，对爬跨行为检测的假阳性率为0.33%。Wang等[9-10]提出了一种基于YOLO v8n的改进模型，以实现对发情奶牛的高效识别。Lodkaew等[11]从视频中获取奶牛鼻子、身体和尾根的3个键点，基

收稿日期：2026-02-11
基金项目：国家自然科学青年基金资助项目(62402182)
作者简介：罗小娟(1974—)，女，江西吉安人，副教授，博士，主要从事物联网技术、智能算法及网络优化研究。E-mail: luoxj@ecust.edu.cn
引用本文：罗小娟, 王道龙, 胡鹏昊, 等. 基于机器学习的奶牛发情识别方法研究 [J]. 华东理工大学学报(自然科学版), 2026, 52(5): 725-734.
Citation: LUO Xiaojuan, WANG Daolong, HU Penghao, *et al*. Machine Learning-Based for Detecting Estrus in Dairy Cows[J]. Journal of East China University of Science and Technology, 2026, 52(5): 725-734.

于爬跨、嗅探等行为特征分析，提出了 CowXNet 算法。但这种方法受牧场环境影响，存在背光、弱光等干扰因素，且在群体检测中易受检测距离以及奶牛之间相互遮挡的影响，难以满足全天候、全方位的精准监测需求。

基于三轴加速度计等穿戴式传感器的接触式监测方法因其抗干扰能力强、具有全天候个体行为精准捕捉的优势，是目前智慧牧场应用的重要研究方向[12]。吴东霖等[13]使用智慧项圈监测奶牛的健康情况。Shahriar 等[14]提出了一种基于 K-means 聚类的无监督方法，统计聚类标签并结合活动指数来检测奶牛发情。Schweinzer 等[15]通过奶牛耳部佩戴三轴加速度计监测数据，并运用算法和机器学习监测奶牛发情。Wang 等[16-17]通过颈部可穿戴设备，运用支持向量机(Support Vector Machine, SVM)、反向神经网络 (Back Propagation Neural Network, BPNN) 等多种机器学习方法监测奶牛发情行为。Benaissa 等[18]、Wang 等[19]和相关研究[20]通过多传感器数据融合和机器学习技术显著提高了发情监测的准确性和效率。但当前关于奶牛发情的监督学习模型中发情标签通常需人工长期观察获取，数据标注成本高昂且具有主观性；且当前研究往往提取高维特征来训练模型，但因模型黑盒而难以揭示表征发情行为的核心特征，导致特征冗余，显著增加了设备进行数据处理的功耗。近年来，可解释性分析开始应用于动物行为研究[21-22]，这为量化模型特征重要性并指导特征降维提供了思路。

针对上述接触式监测方法中标签获取困难与特征冗余导致的功耗问题，本研究提出一种基于无监督聚类与可解释性机器学习的奶牛发情识别方法。该方法通过 K-means 聚类对加速度传感器采集的数据进行无监督聚类得到发情标签，然后对比 K 近邻(K-Nearest Neighbors, KNN)、SVM、随机森林(Random Forest, RF) 3 种模型在不同时间窗口下的性能以选出最佳模型，在此基础上通过沙普利加性解释(Shapely Additive Explanations, SHAP)对模型进行全局解释性分析，量化特征贡献度，剔除冗余特征以实现模型轻量化。

1 数据采集与标签获取

1.1 数据采集与预处理

研究数据采集自张家港梁丰牧场。通过颈挂式可穿戴设备采集数据，其主要包含三轴加速度传感器、微处理器，并通过 3.7 V 的锂电池供电。设备通过尼龙项圈佩戴在奶牛颈部左侧，同时为了确保设备贴合牛体，避免因奶牛一些动作导致设备发生偏移，安装时在项圈底部增加了配重。具体安装位置如图 1 所示，确保加速度计 X 轴平行于地面，表示奶牛前后方向的运动；Y 轴垂直于地面，表示奶牛上下方向的运动；Z 轴表示左右方向的运动。研究中共采集 28 头奶牛数据，采集时间为 2024 年 12 月 15 日至 2025 年 1 月 15 日，其中 6 头用于发情标签获取与最佳模型选择，22 头用于泛化试验。



图 1 传感器安装位置示意图

Fig. 1 Sensor mounting position diagram

因传感器采集的加速度数据的重力分量占比大，需滤除重力以保留动态的线性加速度数据。重力滤波采用截止频率为 0.3 Hz、4 阶的巴特沃斯高通滤波器[23]。对滤波后的三轴加速度计算信号幅度矢量(Signal Magnitude Vector, SMV)与信号幅度面积(Signal Magnitude Area, SMA)，具体计算公式分别见式(1)、式(2)。

$$\mathrm{SMV}=\sqrt{(a_x)^2+(a_y)^2+(a_z)^2} \tag{1}$$

$$\mathrm{SMA}=|a_x|+|a_y|+|a_z| \tag{2}$$

式中，a_x、a_y、a_z 分别表示 x、y、z 三轴滤波后的加速度。

奶牛发情时的爬跨、嗅探等动作通常持续时间较短。若时间窗口过长，这些短暂的剧烈活动信号容易被日常平缓活动所平均，在提取特征时易导致发情特征不明显；若时间窗口过短，窗口内的数据量可能不足以提取有效的频域特征，故计算出 SMV 与 SMA 后，将窗口分别划分为 10、30、50 s，再对每个窗口内的三轴加速度数据和 SMV、SMA 进行特征提取，提取标准差、幅度、能量和快速傅里叶变换的前两个幅度，共 25 个特征。

1.2 基于无监督聚类的发情标签生成

已有研究表明，奶牛发情时会伴随生理指标的显著变化，如体温升高、反刍时间和次数减少，以及活动量显著增加[24]。发情初期和后期，奶牛表现为焦躁不安、行走时间增加等中等强度活动；在发情期奶

牛表现出爬跨与被爬跨这种高强度活动; 非发情期间奶牛活动量较低。因此, 可通过 K-means 聚类[25] 算法将活动量划分成 3 组: 低活动量、中活动量、高活动量。K-means 聚类的核心思想是将样本划分成 K 个不相交的簇, 其目标函数定义如式 (3) 所示。

$$J = \sum_{k=1}^{K} \sum_{x_i \in C_k} \|x_i - \mu_k\|^2 \tag{3}$$

其中, K 表示聚类簇数; C_k 表示第 k 个簇的集合; x_i 表示第 i 个数据样本; μ_k 表示第 k 个簇的质心。

试验中初始化方法采用“K-means++”以加速收敛, 最大迭代次数设定为 300, 随机种子固定为 42。K-means 聚类会给所有窗口随机分配标签, 为保证标签的一致性, 试验中选取信号幅度矢量的标准差作为核心指标, 该指标能够有效反映出奶牛颈部活动的剧烈程度。通过对聚类中心在该特征上的值进行排序, 将最低值映射为标签“0”, 表示低活动量; 最高值映射为标签“2”, 表示高活动量; 中间值为“1”, 表示中活动量。之后按照 2 h 的时间段统计中活动量与高活动量标签的数量, 并结合牧场统计的发情时间选取发情当天标签数量最多的时间段的标签打上标签“1”, 作为发情标签; 随机选取距离发情时间一周的相同时间段标签打上标签“0”, 作为非发情标签。对 3 种不同窗口的数据重复上述步骤得到 3 套独立的标签集, 最终 3 种不同窗口提取的非发情标签数量分别是 7 886、5 783、3 974, 发情标签数量分别是 1 120、1 100、1 228。

2　发情识别模型构建与特征选择

2.1　发情识别算法

2.1.1　KNN　KNN 是一种基于邻近度的分类算法[26]。其工作原理是在特征空间中寻找与新样本距离最近的 K 个已知类别的样本, 然后通过多数投票或加权投票的方式来预测新样本的类别。本文使用 KNN 方法对奶牛的发情进行预测, 距离测量函数设置为欧几里得距离, 其计算公式如式 (4) 所示。

$$D(x, x^{(i)}) = \sqrt{\sum_{j=1}^{d} (x_j - x_j^{(i)})^2} \tag{4}$$

式中, d 表示特征维数; x_j 表示第 j 个特征的取值。

在模型超参数选择方面, 研究中采用网格搜索法并结合 3 折交叉验证, 在 1~10 范围内寻优, 最终确定最优邻居数。

2.1.2　SVM　SVM 是一种由 Vapnik 等提出的高效学习分类方法, 可用于解决二分类或多分类问题[27]。SVM 在训练过程中在关注能否正确分类样本的同时, 还注重分类边界的“间隔最大化”, 从而提升模型的泛化能力, 其通过引入核函数(如线性核、多项式核或径向基核)将低维不可分问题映射到高维空间, 在高维空间中实现线性可分。研究中采用径向基函数核(RBF), 其定义如式 (5) 所示。

$$K(x_i, x_j) = \exp(-\gamma \left\| x_i - x_j \right\|^2) \tag{5}$$

式中, x_i、x_j 表示输入空间的两个样本; $\|x_i - x_j\|^2$ 表示欧氏距离的平方; γ 则依据特征数量与数据离散度自适应取值。

2.1.3　随机森林　随机森林是一种基于集成学习理论的强大监督学习算法[28]。随机森林算法的基本思想是其从原始训练集中有放回地随机抽样, 生成多个与原始数据集大小相同的自助样本集, 每一个自助样本集都被用来独立地训练一棵决策树。在构建每棵树的过程中, 随机森林在每次节点划分时并不使用全部特征, 而是从特征集中随机抽取一个子集, 其维度计算如式 (6) 所示。

$$|F_m| = \sqrt{|F|} \tag{6}$$

式中, $|F_m|$表示从全集中抽取的特征子集中的特征数量; $|F|$表示特征总数。

对于分类任务, 节点分裂采用基尼指数作为评价标准, 其定义如式 (7) 所示。

$$\mathrm{Gini}(t) = \sum_{k'=1}^{K'} p_{k'}(1 - p_{k'}) \tag{7}$$

式中, t 表示当前决策树节点; K'表示分类问题的类别总数; $p_{k'}$表示属于 k'类样本的比例。

最终预测结果通过多数投票得到, 具体计算如式 (8) 所示。

$$\hat{y} = \arg\max_{c \in C} \sum_{m=1}^{M} 1(h_m(x) = c) \tag{8}$$

式中, C 表示样本类别的集合; c 表示 C 中的一个类别; M 表示决策树的总量; $h_m(x)$ 表示第 m 棵树对样本 x 的预测结果。

2.2　模型训练与选择

研究按照 80% 与 20% 比例随机划分训练集与测试集, 训练集和测试集中发情与非发情的标签数量分别如表 1 和表 2 所示。在模型训练、特征选择阶段不采用划分的 20% 测试集, 该测试集仅用于最终评估降维前、后的模型性能。在模型训练阶段采用 80% 的训练集, 并在该数据集内部进行 5 折交叉验证。每轮使用其中 4 折作为训练集, 并仅对训练集

表1 训练集样本数量

Table 1 Number of samples in the training set

Label	Number		
	Window of 10 s	Window of 30 s	Window of 50 s
0	6308	4626	3179
1	896	880	982

表2 测试集样本数量

Table 2 Number of samples in the testing set

Label	Number		
	Window of 10 s	Window of 30 s	Window of 50 s
0	1578	1157	795
1	224	220	246

进行少量合成过采样技术[29](Synthetic Minority Oversampling Technique, SMOTE)处理以平衡样本类别;剩余1折作为验证集,保持其原始数据分布;最后,计算5折交叉验证下3种模型的平均性能指标。

为筛选出最佳模型,研究通过Friedman检验分析10、30、50 s 3种窗口下的KNN、SVM、RF 3种模型是否存在显著性差异,在此基础上进一步采用Nemenyi进行后续检验,量化模型间的性能差异,以此筛选出最佳模型。

2.3 模型性能评估

为评估模型在发情鉴别方面的性能,研究应用混淆矩阵、准确率(Accuracy, Acc)、精确率(Precision, Pr)、召回率(Recall, R)、F1-score以及ROC(Receiver Operating Characteristic)曲线和PR(Precision-Recall)曲线来评估模型性能。其中Acc、Pr、R、F1-score定义如下:

$$\mathrm{Acc}=\frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{TN}+\mathrm{FP}+\mathrm{FN}} \tag{9}$$

$$\mathrm{Pr}=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}} \tag{10}$$

$$R=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \tag{11}$$

$$\text{F1-score}=\frac{2\times\mathrm{TP}}{2\times\mathrm{TP}+\mathrm{FP}+\mathrm{FN}} \tag{12}$$

式中,TP(True Positives)表示样本的真实标签是发情,模型预测为发情;TN(True Negatives)表示样本的真实标签是非发情,模型预测为非发情;FP(False Positives)表示样本的真实标签是非发情,模型预测为发情;FN(False Negatives)表示样本的真实标签是发情,模型预测为非发情。

2.4 基于SHAP的可解释性特征降维

筛选出的最佳模型目前仍依赖于提取的25维特征,但是在实际牧场监测应用中,高维特征会增加设备的计算功耗,缩短设备续航时间。为此,本文引入SHAP[30]分析以筛选出最佳的特征子集。SHAP是一种基于合作博弈论中沙普利值的模型事后解释方法,它能够量化每个特征对预测结果的贡献,SHAP值定义如式(13)所示。

$$\phi_i(x)=\sum_{S\subseteq F\backslash\{i\}}\frac{|S|!\,(|F|-|S|-1)!}{|F|!}\left[f_x(S\cup\{i\})-f_x(S)\right] \tag{13}$$

式中,$\phi_i(x)$表示样本x中第i个特征对应的SHAP值;F表示所有特征集合;S表示特征i不在其中的任意子集;$|S|$表示子集S中包含的特征数量;$|F|$表示特征的总数量;$f_x(S)$表示模型获得特征子集S时,对样本的预测值;$f_x(S\cup\{i\})$表示模型获得子集S和特征i时,对样本的预测值。

采用平均绝对SHAP值作为衡量特征全局重要性的核心指标,并以此进行特征降维,其定义如式(14)所示。

$$I_i=\frac{1}{N}\sum_{j=1}^{N}\left|\phi_i(x_j)\right| \tag{14}$$

式中,N表示样本总数;I_i表示特征i在所有样本中对预测结果贡献的平均幅度。

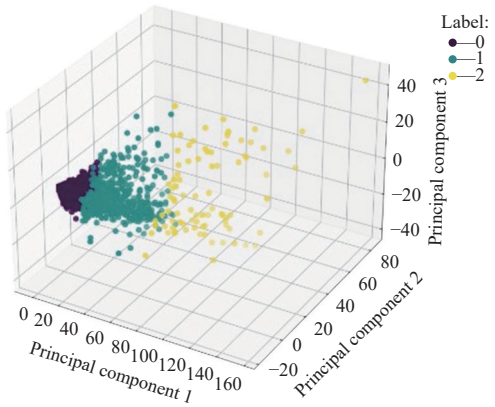
平均绝对SHAP值量化了单个特征在整个数据集上对模型输出影响的平均幅度,通过计算并排序所有特征的平均绝对SHAP值,可以选择出对模型决策贡献最大的关键特征子集。研究中还分析了蜂群图,揭示了特征值与模型贡献之间的关系,同时还反映了提取的特征是否有效,这为筛选特征提供了理论依据。

本文对初始划分的80%训练集进行SMOTE处理后训练选择出的最佳模型,并计算平均绝对SHAP值,得到特征排名。试验中依据特征重要性排序进行消融试验,逐步增加特征。在每一次新的特征组合中,对该80%训练集中进行5折交叉验证并评估模型性能,最终选择出最佳特征子集。同时,为防止样本不平衡的影响,对每一轮的训练折进行SMOTE处理。

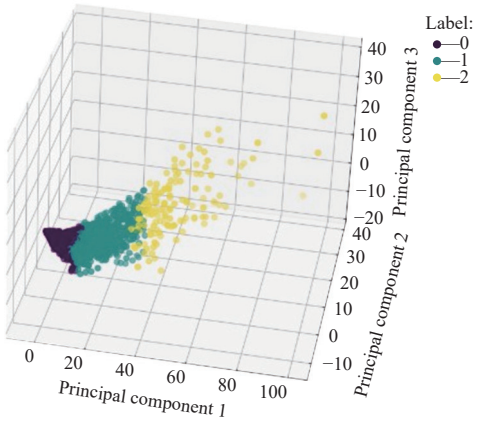
3 结果与分析

3.1 聚类结果

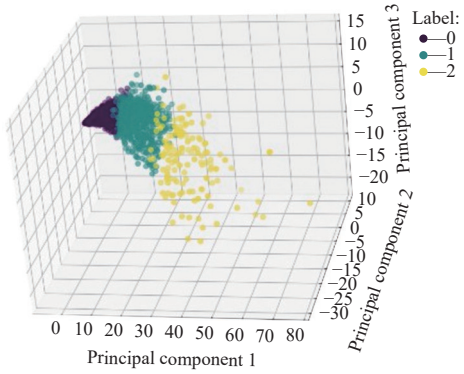
为直观地展示聚类后三类标签在特征空间中的分布情况,本文采用主成分分析将高维特征数据降至三维进行可视化。研究随机选取一头奶牛,其聚类结果如图2所示,从三维散点图中可以看出,3类标签均形成了分离的簇,这表明研究中使用的K-



(a) Clustering result with 10 s window



(b) Clustering result with 30 s window



(c) Clustering result with 50 s window

图 2　3 种窗口聚类效果

Fig. 2　Clustering performance of three different windows

means 聚类方法能够有效区分不同强度的活动量。

3.2　模型性能对比

对比 RF、KNN 及 SVM 3 种模型在 10、30 s 和 50 s 3 种时间窗口下的性能, 结果如表 3 所示。从表 3 中可以看出, 3 种模型均在 10 s 窗口下性能最佳, 其中 RF 模型的 Acc、Pr、*R* 和 F1-score 分别达到了 0.9954、0.9685、0.9955 和 0.9818; 但当窗口增加到 30 s 时, 模型的性能均出现了下滑, 其中 KNN 和 SVM 的 Pr 分别下降至 0.8212 和 0.8099, 可能是因为在较长的时间窗口内, 奶牛爬跨等剧烈活动被正常的活动所平均, 导致发情特征被弱化, 从而降低了模型的精确率。

表 3　不同模型性能对比

Table 3　Performance comparison of different models

Time of window/s	Model	Acc	Pr	*R*	F1-score
50	KNN	0.9456	0.8534	0.9297	0.8899
	SVM	0.9478	0.8358	0.9694	0.8976
	RF	0.9562	0.8795	0.9440	0.9105
30	KNN	0.9560	0.8212	0.9284	0.8712
	SVM	0.9596	0.8099	0.9784	0.8860
	RF	0.9660	0.8568	0.9454	0.8988
10	KNN	0.9940	0.9621	0.9910	0.9763
	SVM	0.9931	0.9492	0.9988	0.9733
	RF	**0.9954**	**0.9685**	**0.9955**	**0.9818**

为进一步量化模型之间的性能差异, 研究采用 Friedman 检验对 Acc、Pr、*R* 和 F1-score 四项指标进行统计分析。结果如表 4 所示, 4 项指标对应的卡方统计量分别为 36.6308、37.3333、35.2972 和 36.0533, 其 *P* 值均小于 0.001, 表明不同模型之间的性能差异在统计上具有显著性。在 Friedman 检验结果基础上, 采用 Nemenyi 事后检验对模型进行比较, 并绘制临界差异(Critical Difference, CD)图, 如图 3 所示。若 CD 图中通过水平连接线将多个模型连接在一起, 表示这些模型之间的性能差异不显著。由图中水平连线(粗横线)可看出, 尽管 10 s 窗口下的 RF 模型与部分模型之间无显著性差异, 但其在 Acc、Pr 及 F1-score 指标下均取得最低平均秩, 表明该模型在整体性能和稳定性上更优。对比不同时间窗口发现, 10 s 窗口在各类模型中普遍性能更优, 而 30 s 和 50 s 模型的整体性能相对较弱。这表明较短的 10 s 窗口能够精准且有效地捕捉爬跨、嗅探等持续时间短的高强度活动信号; 而较长的时间窗口则容易稀释关键的发情特征, 进而削弱了模型的判别能力。因此, 综合模型性能与统计检验结果, 试验选择 10 s 窗口的 RF 模型作为最优模型, 并基于此进行后续特征降维分析。

3.3　特征重要性分析

为实现特征降维以降低计算功耗, 本文采用 SHAP 方法对 10 s 窗口下的 RF 模型进行了可解释性

表 4　各模型的 Friedman 检验结果

Table 4　Friedman test results of models

Metric	χ^2_statistic	*P*-value	Significance
Acc	36.6308	***	s
Pr	37.3333	***	s
R	35.2972	***	s
F1-score	36.0533	***	s

***—p<0.001; s—Significant difference

分析，结果如图 4 所示。从图 4(a) 可见，SMA_energy、SMV_energy 的平均绝对 SHAP 值比其他特征高，表明奶牛发情时颈部活动能量是区分发情状态的关键依据。图 4(b) 的蜂群图则进一步揭示了特征值对预

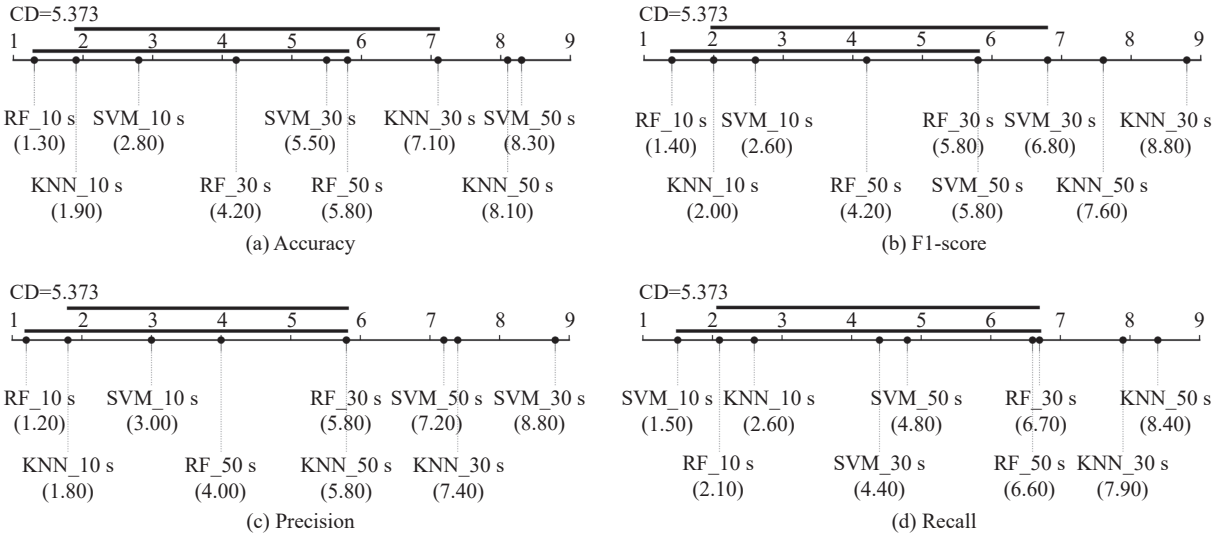
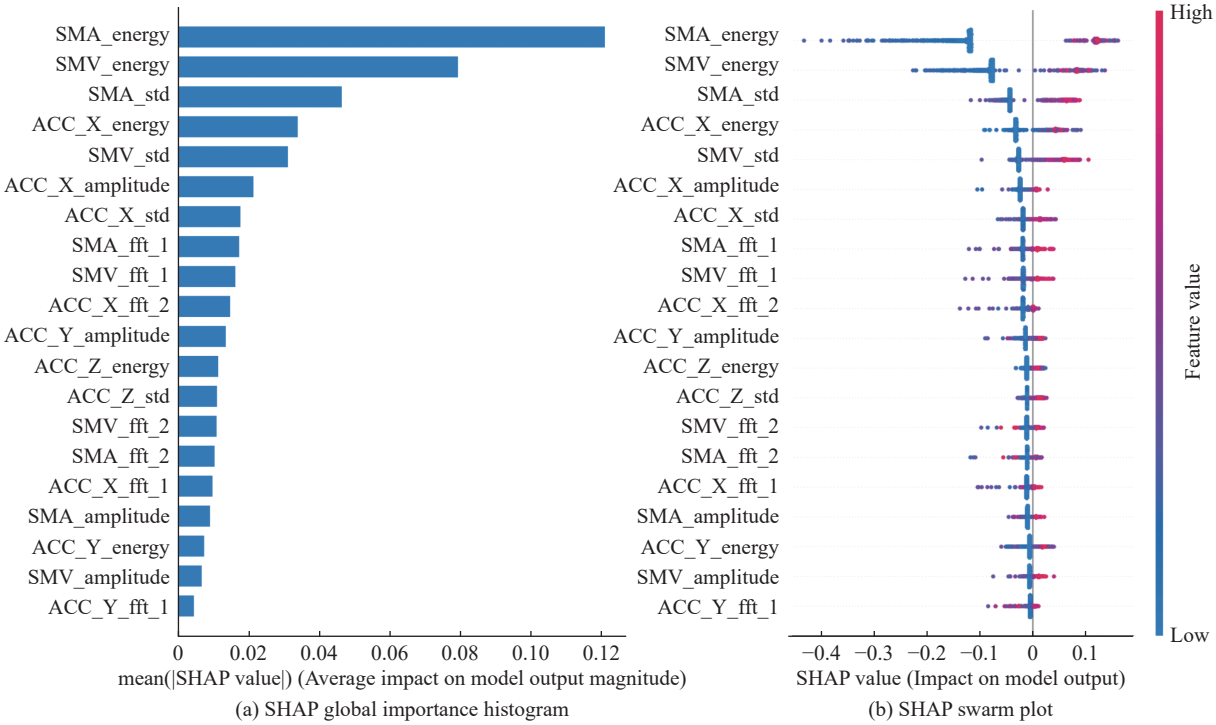


图 3　Nemenyi 检验 CD 图

Fig. 3　Critical difference results of Nemenyi test



SMA_energy—Signal magnitude area energy; SMV_energy—Signal magnitude vector energy; SMA_std—Signal magnitude area standard deviation; ACC_X_energy—Accelerometer *X*-axis energy; SMV_std—Signal magnitude vector standard deviation; ACC_X_amplitude—Accelerometer *X*-axis amplitude; ACC_X_std—Accelerometer *X*-axis standard deviation; SMA_fft_1—Signal magnitude area Top-1 fast Fourier transform magnitude; SMV_fft_1—Signal magnitude vector Top-1 fast Fourier transform magnitude; ACC_X_fft_2—Accelerometer *X*-axis Top-2 fast Fourier transform magnitude; ACC_Y_amplitude—Accelerometer *Y*-axis amplitude; ACC_Z_energy—Accelerometer *Z*-axis energy; ACC_Z_std—Accelerometer *Z*-axis standard deviation; SMV_fft_2—Signal magnitude vector Top-2 fast Fourier transform magnitude; SMA_fft_2—Signal magnitude area Top-2 fast Fourier transform magnitude; ACC_X_fft_1—Accelerometer *X*-axis Top-1 fast Fourier transform magnitude; SMA_amplitude—Signal magnitude area amplitude; ACC_Y_energy—Accelerometer *Y*-axis energy; SMV_amplitude—Signal magnitude vector amplitude; ACC_Y_fft_1—Accelerometer *Y*-axis Top-1 fast Fourier transform magnitude

图 4　基于 SHAP 的模型特征重要性与贡献度分析

Fig. 4　SHAP analysis of model feature importance and contribution

测结果的具体影响方向,可见 SMA_energy 与 SMV_enenrgy 较高的特征值对应着正向的 SHAP 值, 表示将预测推向“发情”; 而其较低的特征值则对应负向 SHAP 值, 表示将预测推向“非发情”。故当一个时间窗口的 SMA_energy 值很高时, 它会对模型将该窗口判断为发情产生正向推动作用, 可见模型的决策逻辑与奶牛发情期活动量增加的行为特征一致。

为降低算法功耗, 在边缘端设备中实现模型部署, 本文基于 SHAP 特征重要性排名进行了特征消融试验以选取最佳特征子集。从表 5 可以看出, 当特征取前 3 维时, 其 F1-score 的平均值达到 0.983 0, 与原始特征的模型性能基本一致。综合计算功耗与算法鲁棒性, 研究最终选取前 3 维作为最佳特征子集。与 25 维特征相比, 采用 3 维特征大幅减少了特征维度, 将会显著降低边缘端设备特征提取所需的计算功耗。

3.4　降维效果验证

采用初始划分的 20% 独立测试集对 10 s窗口的 RF 模型降维前后的性能进行了对比, 结果如表 6 所示。从表 6 可以看出, 当特征维度从 25 降至 3 时, 模型的准确率仅下降 0.001 1, 召回率保持不变。

对比模型在独立测试集上的混淆矩阵, 结果如图 5 所示。从混淆矩阵可以看出, 发情标签的检出数量在降维前后并未发生改变, 降维后的模型也仅将两个非发情标签误判为发情标签, 模型精确率由 0.944 9 下降至 0.937 0, 下降了0.007 9。

从图 6 的 ROC 曲线可以看出, 降维前后模型性能未发生显著衰减。由于本研究数据集中非发情样本数量远多于发情样本, ROC 曲线可能会因非发情样本数量大而呈现出较好的效果。因此, 本文进一步对比了 PR 曲线。从图 6 中的 PR 曲线可以看出降维后模型的 AP 值仍保持在 0.96 以上。

为排除单次数据集划分可能带来的误差, 本文对原始 10 s 窗口数据集进行了 5 次随机划分, 并对比了降维前后的 RF 模型性能, 结果如表 7 所示。同时, 采用 Wilcoxon 符号秩检验进行统计分析, 结果如表 8 所示。从表中可以看出, 降维前后模型性能无显著性差异, 且降维后的模型较为稳定。

为验证降维前后模型的泛化能力, 分别用降维前后的模型对奶牛原始数据提取特征, 并打上是否发情的标签, 统计每两个小时的发情标签数量, 并根据发情标签数量来判断奶牛是否发情, 最终统计模型对奶牛发情检测的准确率。以编号为 20701 的奶牛为例, 算法检测结果如图 7 所示。可以看出, 在大部分时间里, 奶牛的发情标签数量维持在一个相对较低的水平; 但在发情时, 奶牛活动量增加, 发情标签的数量增加。

研究中分别使用降维前后模型对该 22 头奶牛

表 5　基于 SHAP 排序的消融试验模型性能对比

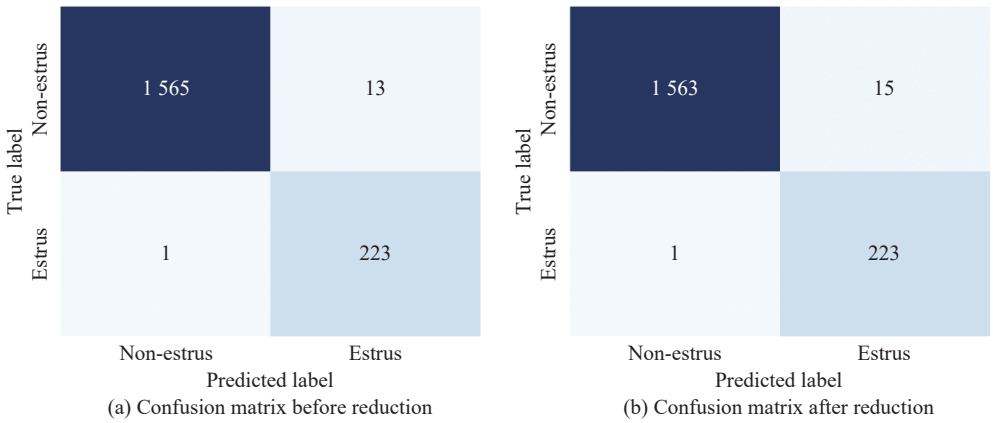
Table 5　Performance comparison of ablation testing model based on SHAP ranking

Feature dimension	Acc	Pr	*R*	F1-score
1	0.994 4	0.967 3	0.988 9	0.972 1
3	**0.995 7**	**0.967 6**	**0.998 9**	**0.983 0**
5	0.995 6	0.966 5	0.998 9	0.982 4
10	0.995 7	0.968 6	0.997 8	0.983 0
25	0.995 6	0.969 6	0.995 5	0.982 4

表 6　独立测试集中降维前后模型性能对比

Table 6　Performance comparison before and after reduction based on the independent test set

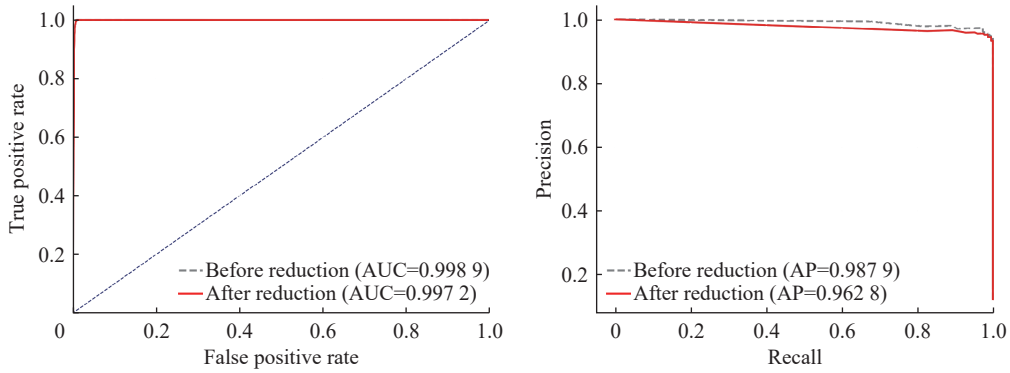
Feature dimension	Acc	Pr	*R*	F1-score
25	0.992 2	0.944 9	0.995 5	0.969 6
3	0.991 1	0.937 0	0.995 5	0.965 4



(a) Confusion matrix before reduction　　(b) Confusion matrix after reduction

图 5　独立测试集上降维前后模型的混淆矩阵对比

Fig. 5　Comparison of confusion matrix before and after reduction based on the independent test set



(a) ROC curve (b) PR curve

图 6 降维前后模型的 ROC 曲线与 PR 曲线对比

Fig. 6 Performance comparison of ROC curve and PR curve before and after reduction

表 7 5 次随机划分下降维前后 RF 模型性能

Table 7 Performance of RF models before and after feature reduction over five random splits

Feature dimension	Acc	Pr	*R*	F1-score
25	0.994 1 ± 0.001 0	0.962 7 ± 0.005 7	0.991 1 ± 0.004 0	0.976 7 ± 0.003 8
5	0.994 2 ±0.001 2	0.958 8 ± 0.008 4	0.996 4 ± 0.005 2	0.977 2 ± 0.004 9

表 8 Wilcoxon 符号秩检验结果

Table 8 Results of Wilcoxon signed-rank test

Metric	*P*-value	Result
Acc	1.0000	ns
Pr	0.2500	ns
R	0.1250	ns
F1-score	0.6250	ns

ns—Non-significant (p>0.05)

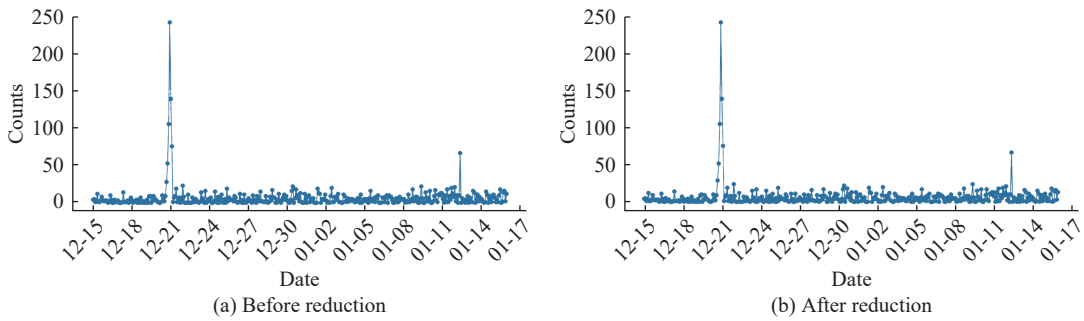
数据进行检测，发情检出准确率如图 8 所示，从图 8 的统计图可以看出，降维前后模型对奶牛发情检测的准确率相同，达到了 0.909 0。

4 讨论与结论

实现奶牛发情监测设备的低功耗、长续航是目前发展智慧牧场的迫切需求。传统的方法依赖于提取的多维特征，计算复杂度高，难以在 MCU 端进行部署。本研究通过 SHAP 分析筛选出 SMA_energy、SMV_enenrgy、SMA_std 这 3 个特征子集。基于该特征子集训练的模型在泛化性试验中保持了 0.909 0 的准确率(图 8)，且与原始 25 维模型无显著性差异。该特征维度的降低，不仅减少了数据存储所需空间，而且显著降低了计算功耗，为开发低成本、长续航的智能监测设备提供了理论与技术支撑。

但本研究中，试验数据采集自统一牧场，基于 SHAP 分析的特征重要性以及筛选出的最优特征子集亦是基于此得出的。在实际应用中，受奶牛品种、牧场环境或季节影响，奶牛发情行为特征可能存在一定的差异，未来可将基于最优特征子集训练的模型用于不同牧场奶牛的数据，检测模型性能，提高模型泛化能力。

研究中采用降维后的 RF 模型对 22 头奶牛数据



(a) Before reduction (b) After reduction

图 7 降维前后模型预测奶牛发情的结果对比

Fig. 7 Model prediction results for cow estrus before and after reduction

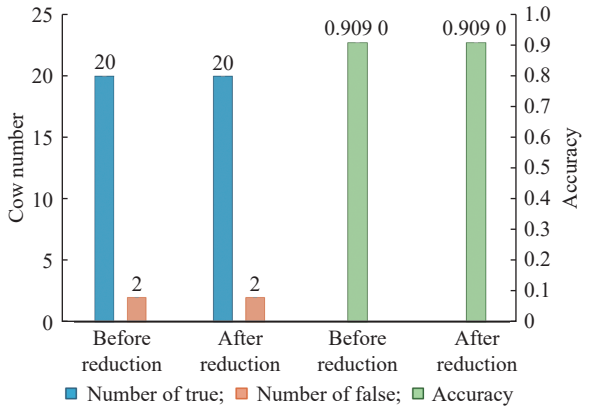


图 8　降维前后 10 s 窗口 RF 模型的发情检出准确率

Fig. 8　Estrus detection accuracy for the 10 s window RF model before and after dimensionality reduction

进行了泛化性测试, 采用统计 2 h 窗口内发情标签数量的方法进行发情判定, 其中有 2 头奶牛漏检。后续对这两头奶牛进行追溯得出, 1 头奶牛经人工判断发情后进行受精并成功怀孕, 但算法未检测出, 可能原因是该奶牛属于隐性发情或发情信号微弱, 其发情时并未出现显著的活动量增加的情况; 另 1 头奶牛经人工判断发情并进行受精之后, 其在下一个发情周期内仍出现了发情行为, 可能原因是人工观察出现了误判或者奶牛发情时的活动强度低且持续时间短暂, 导致配种时机不对而未成功受孕。未来研究中可引入多模态数据, 结合体温、反刍时间等特征, 提高发情检测精度和对隐性发情的检出。

本研究提出了一种融合无监督聚类与可解释性特征选择的奶牛发情检测方法。试验结果表明, 利用聚类标签训练的 RF 模型在 10 s 窗口下综合性能最优。研究中基于该最优模型通过 SHAP 方法对全局可解释性分析, 选出 SMA_energy、SMV_energy、SMA_std 3 个核心特征, 将特征维数降至 3 维。经 Wilcoxon 检验后得出降维前后的 10 s 窗口下的 RF 模型无显著性差异。泛化试验表明, 降维后的 RF 模型在独立数据集上的发情检出率达 0.909 0, 具有较好的鲁棒性。本研究的方法实现了发情标签生成与模型特征轻量化, 为研发低功耗、长续航的奶牛发情监测设备提供了理论依据与技术支撑。

参考文献:

[1] 罗西尔, 卢小龙, 刘庆友, 等. 肉牛智慧养殖技术研究进展 [J]. 华南农业大学学报, 2024, 45(5): 661-671

[2] 王祯, 宗哲英, 王海超, 等. 奶牛发情数字化检测方法研究现状与发展 [J]. 中国饲料, 2021, 1(21): 134-138.

[3] 王少华, 何东健, 刘冬. 基于机器视觉的奶牛发情行为自动识别方法 [J]. 农业机械学报, 2020, 51(4): 241-249.

[4] WANG J, CHEN H R, WANG J P, *et al*. Identification of oestrus cows based on vocalisation characteristics and machine learning technique using a dual-channel-equipped acoustic tag[J]. Animal, 2023, 17(6): 100811.

[5] WANG J, SI Y F, LI X X, *et al*. Discrimination strategy using machine learning technique for oestrus detection in dairy cows by a dual-channel-based acoustic tag[J]. Computers and Electronics in Agriculture, 2023, 210: 107949.

[6] ARIKAN İ, AYAV T, SEÇKIN A Ç, *et al*. Estrus detection and dairy cow identification with cascade deep learning for augmented reality-ready livestock farming[J]. Sensors, 2023, 23(24): 9795.

[7] 王政, 许兴时, 华志新, 等. 融合 YOLO v5n 与通道剪枝算法的轻量化奶牛发情行为识别 [J]. 农业工程学报, 2022, 38(23): 130-140.

[8] TSAI D M, HUANG C Y. A motion and image analysis method for automatic detection of estrus and mating behavior in cattle[J]. Computers and Electronics in Agriculture, 2014, 104: 25-31.

[9] WANG Z, HUA Z X, WEN Y C, *et al*. E-YOLO: Recognition of estrus cow based on improved YOLOv8n model[J]. Expert Systems with Applications, 2024, 238: 122212.

[10] WANG Z, DENG H, ZHANG S, *et al*. Detection and tracking of oestrus dairy cows based on improved YOLOv8n and TransT models[J]. Biosystems Engineering, 2025, 252: 61-76.

[11] LODKAEW T, PASUPA K, LOO C K. CowXNet: An automated cow estrus detection system[J]. Expert Systems with Applications, 2023, 211: 118550.

[12] LAMANNA M, BOVO M, CAVALLINI D. Wearable collar technologies for dairy cows: A systematized review of the current applications and future innovations in precision livestock farming[J]. Animals, 2025, 15(3): 458.

[13] 吴东霖, 王山山, 赵少斌, 等. 基于智慧项圈监测的奶牛围产前期采食时间对母牛和犊牛健康状况及生产性能的影响 [J]. 饲料工业, 2025, 46(9): 44-50.

[14] SHAHRIAR M S, SMITH D, RAHMAN A, *et al*. Detecting heat events in dairy cows using accelerometers and unsupervised learning[J]. Computers and Electronics in Agriculture, 2016, 128: 20-26.

[15] SCHWEINZER V, GUSTERER E, KANZ P, *et al*. Evaluation of an ear-attached accelerometer for detecting estrus events in indoor housed dairy cows[J]. Theriogenology, 2019, 130: 19-25.

[16] WANG J, ZHANG Y D, WANG J P, *et al*. Using machine-learning technique for estrus onset detection in dairy cows from acceleration and location data acquired by a neck-tag[J]. Biosystems Engineering, 2022, 214: 193-206.

[17] WANG J, ZHANG Y, BELL M, *et al*. Potential of an activity index combining acceleration and location for automated estrus detection in dairy cows[J]. Information Processing in Agriculture, 2022, 9(2): 288-299.

[18] BENAISSA S, TUYTTENS F A M, PLETS D, *et al*. Calv-

ing and estrus detection in dairy cattle using a combination of indoor localization and accelerometer sensors[J]. Computers and Electronics in Agriculture, 2020, 168: 105153.

[19] WANG R R, LI Y, TIAN F Y, *et al*. Estrus detection in dairy cows using advanced object tracking and behavioral analysis technologies[J]. Computers and Electronics in Agriculture, 2025, 235: 110331.

[20] WANG J, ZHAO Y J, LI X X, *et al*. Multimodal fusion-based detection method of estrus cows using multisource data inspired by hidden Markov model algorithms[J]. Computers and Electronics in Agriculture, 2025, 235: 110391.

[21] ZHOU X, QU Y, XU C, *et al*. Machine learning and shapely additive explanation-based interpretation for predicting mastitis in dairy cows[J]. Animals, 2026, 16(2): 204.

[22] JANA R, DIXIT S, SHARMA M, *et al*. An explainable AI based approach for monitoring animal health [EB/OL]. (2025-08-13). https://arxiv.org/abs/2508.10210.

[23] DENG M, ZHAO Z. The cow attitude recognition algorithm based on multi-sensor information fusion [C]//2019 IEEE International Conference on Artificial Intelligence and Computer Applications. Dalian, China: IEEE, 2019: 283-288.

[24] 李蓝祁, 刘江静, 陈晓丽, 等. 奶牛发情期活动量变化规律研究 [J]. 畜牧兽医学报, 2018, 49(7): 1387-1393.

[25] XIE J Y, JIANG S, XIE W X, *et al*. An efficient global K-means clustering algorithm[J]. Journal of Computers, 2011, 6(2): 271-279.

[26] COVER T, HART P. Nearest neighbor pattern classification[J]. IEEE Transactions on Information Theory, 1967, 13(1): 21-27.

[27] VAPNIK V, CORTES C. Support-vector networks[J]. Machine Laearning, 1995, 20(3): 273-297.

[28] 文家璇, 王苗, 刘济. 基于时序分解和随机森林的时间序列多步预测算法 [J]. 华东理工大学学报 (自然科学版), 2023, 49(6): 873-881.

[29] CHAWLA N V, BOWYER K W, HALL L O, *et al*. SMOTE: Synthetic minority over-sampling technique[J]. Journal of Artificial Intelligence Research, 2002, 16: 321-357.

[30] LUNDBERG S M, LEE S I. A unified approach to interpreting model predictions[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, CA, USA: Curran Associates, Inc, 2017: 4768-4777.

Machine Learning-Based for Detecting Estrus in Dairy Cows

LUO Xiaojuan, WANG Daolong, HU Penghao, LIN Jiaxu, GUAN Lei

(School of Information Science and Engineering, East China University of Science and Technology, Shanghai 200237, China)

Abstract: To address the issues of high cost and strong subjectivity associated with manual observation in dairy cow estrus detection, as well as the high computational power consumption caused by high-dimensional model features, this study proposes a dairy cow estrus detection method based on unsupervised clustering and interpretable feature selection. The method utilizes the K-means algorithm to cluster raw activity time series data collected by neck-mounted accelerometers on dairy cows to automatically generate estrus behavior labels. Three models are compared: K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Random Forest (RF). These models are evaluated under time windows of 10, 30 s, and 50 s. Finally, the optimal model is determined by the Friedman test and the Nemenyi post-hoc test. The Shapley Additive Explanations (SHAP) method is introduced to quantify feature contributions and screen for the key features. The results indicate that the RF model based on a 10 s time window achieves the best performance. Through SHAP selection, the features are reduced from 25 to 3. The performance of the dimension-reduced model shows no significant deviation, with accuracy, precision, recall, and F1 score reaching 0.991 1, 0.937 0, 0.995 5, and 0.965 4, respectively. In independent trials based on three-axis acceleration data from dairy cows, the RF model achieves an estrus detection accuracy of 0.909 0 using a 10 s window following dimensionality reduction. The proposed method enables both estrus label generation and feature dimensionality reduction, while maintaining model interpretability. This provides technical support for developing low-power, long-endurance estrus monitoring devices for dairy cows.

Key words: machine learning; cow estrus detection; accelerometer; K-means clustering; SHAP

（责任编辑：李娟）