Computer Science and Application 计算机科学与应用, 2026, 16(11), 169-183
Published Online November 2026 in Hans. https://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2026.1611334

Hans汉斯

# 面向预测性维护的CNC刀具磨损状态识别

## ——机器学习与深度学习对比、多轴信号表征及SHAP可解释分析

**黄扬庭[1]，颜轲越[1*]，岳子焕[2]，李　莹[2]**

[1]广东金融学院大数据与人工智能学院，广东 清远
[2]北京理工大学(珠海)国际教育与培训学院，广东 珠海

收稿日期：2026年10月7日；录用日期：2026年11月6日；发布日期：2026年11月13日

## 摘　要

**针对机床时序数据中独立加工实验数量少、相邻窗口相关性强及模型判别依据难以解释的问题，本文构建了结合多轴统计变量、加工实验分组验证与SHAP解释的刀具磨损状态识别方法。本文基于18次计算机数字控制(CNC)蜡铣削实验，从电流、主轴功率、实际速度和速度误差等10个信号通道提取40项统计变量，并采用外层留一加工实验与内层分层四折验证，比较11种机器学习分类器和6种时序深度学习模型；预处理和变量筛选均在训练折内完成。模型比较结果表明，使用全部40项变量的AdaBoost、逻辑回归与GBDT准确率均为77.78%，其中AdaBoost的平衡准确率为77.50%，磨损与未磨损召回率分别为80.00%和75.00%。在6种时序深度学习模型中，1D-CNN的准确率和平衡准确率最高，分别为66.67%和67.50%。变量筛选结果表明，SHAP筛选提高了部分分类器的准确率；在由内层验证从5种分类器中选择模型的实验中，保留SHAP排名前15项变量与使用全部40项变量的准确率均为72.22%。进一步对全变量AdaBoost进行SHAP解释发现，X、Y轴电流水平、主轴电流与速度波动以及Y轴速度误差是模型识别刀具磨损状态的主要信号依据。本文将识别性能比较与模型判别依据相结合，为小样本刀具磨损识别和维护人员复核预测结果提供了方法参考。**

## 关键词

**刀具磨损，多轴信号，分组交叉验证，AdaBoost，SHAP，预测性维护**

# CNC Tool Wear State Identification for Predictive Maintenance

## —Comparison of Machine Learning and Deep Learning, Multi-Axis Signal Representation and SHAP Interpretable Analysis

**Yangting Huang[1], Keyue Yan[1*], Zihuan Yue[2], Ying Li[2]**

[*]通讯作者。

文章引用: 黄扬庭, 颜轲越, 岳子焕, 李莹. 面向预测性维护的 CNC 刀具磨损状态识别[J]. 计算机科学与应用, 2026, 16(11): 169-183. DOI: 10.12677/csa.2026.1611334

[1]School of Data Science and Artificial Intelligence, Guangdong University of Finance, Qingyuan Guangdong
[2]School of International and Continuing Education, Beijing Institute of Technology (Zhuhai), Zhuhai Guangdong

Received: October 7, 2026; accepted: November 6, 2026; published: November 13, 2026

## Abstract

**To address the small number of independent machining experiments, strong correlations between adjacent windows, and difficulties in interpreting model predictions in machine tool time-series data, this study develops a tool wear state identification method combining multi-axis statistical features, validation grouped by machining experiment, and SHAP interpretable analysis. Based on 18 Computer Numerical Control (CNC) wax-milling experiments, this study extracts 40 statistical features from ten signal channels, including current, spindle power, actual velocity, and velocity error, and compares 11 machine learning classifiers and six time-series deep learning models using outer leave-one-experiment-out validation and inner stratified four-fold validation; preprocessing and feature selection are performed within the training folds. The model comparison results show that AdaBoost, logistic regression, and GBDT using all 40 features each achieve an accuracy of 77.78%. AdaBoost achieves a balanced accuracy of 77.50%, with recalls of 80.00% and 75.00% for the worn and unworn classes, respectively. Among the six time series deep learning models, 1D-CNN achieves the highest accuracy and balanced accuracy, at 66.67% and 67.50%, respectively. The feature selection results show that SHAP based selection improves the accuracy of some classifiers; in the experiment where inner validation selects a model from five classifiers, retaining the top 15 SHAP ranked features and using all 40 features both yield an accuracy of 72.22%. Further SHAP analysis of AdaBoost using all features identifies X- and Y-axis current levels, fluctuations in spindle current and velocity, and Y-axis velocity error as the main signal evidence used by the model to identify tool wear states. By combining identification performance comparisons with explanations of model predictions, this study provides a methodological reference for tool wear identification with small samples and for maintenance personnel to review prediction results.**

## Keywords

**Tool Wear, Multi-Axis Signals, Grouped Cross-Validation, AdaBoost, SHAP, Predictive Maintenance**

Copyright © 2026 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

刀具磨损状态识别是数控加工状态监测的重要环节。磨损引起的切削负载变化、运动跟踪偏差和加工质量波动，为利用机床运行数据辅助换刀判断提供了观测依据。将这些信号转化为维护人员能够理解的识别结果，是制造企业数字化转型在设备管理中的具体需求。清远传统制造业数字化转型研究强调技术整合、组织能力与外部支持的联系[1]，相关综述指出，数据技术的价值需要结合业务流程与组织调整实现[2]。因此，面向刀具维护的研究既需评价状态识别是否准确，也需说明模型判断依据，使识别结果能够用于刀具检查与维护复核。

刀具状态监测通常包括信号采集、特征提取和状态识别三个环节。已有综述讨论了智能识别算法、多传感器融合、特征降维与工业系统集成等方法[3]-[5]。复杂工况研究进一步指出，加工参数变化会影响

信号与磨损状态之间的关系，因此状态识别与剩余寿命预测应根据标签类型分别建立任务[6]。本文采用磨损与未磨损两类标签，研究一次加工实验结束后的刀具状态识别。所用输入为该实验已记录的切削信号，识别结果用于辅助状态复核。

现有工业设备诊断研究为信号处理与模型选择提供了多种思路。时域与频域分析用于提取设备运行特征[7]，小波系数可与神经网络结合用于故障诊断[8]，机器学习方法用于齿轮箱状态分类[9]，二维卷积、多传感器融合和注意力机制用于轴承故障识别[10]-[12]。深度学习在工业设备诊断及电气设备故障预测中的应用也已有研究[13] [14]。在刀具磨损预测中，成佳闻等通过多域特征、核主成分分析与注意力堆叠长短期记忆网络实现多传感器信息融合[15]。这些研究说明，统计特征与时序深度学习模型均可用于工业信号建模；针对本文数据，需要在相同评价单位下比较两类方法，并检验变量筛选的实际效果。

本文数据的主要特点是独立加工实验数量少，而同一实验内包含大量相邻采样点。若随机划分同一加工过程的窗口，训练集与测试集可能同时包含相似的信号片段，使评价偏高。Roberts 等提出，交叉验证应保留数据的时间、层级或分组结构[16]；Varma 和 Simon 指出，用同一交叉验证结果选择参数并估计误差会产生评价偏差，嵌套验证可减小这一偏差[17]。本文据此按加工实验编号划分数据，将预处理、参数选择与变量筛选放在训练数据内完成，再用留出实验评价预测结果。

本文围绕三个问题展开：多轴信号的统计变量与时序深度学习模型分别能取得怎样的识别效果；SHAP 变量筛选能否减少输入变量或改善预测；识别效果较好的模型主要依据哪些信号作出判断。为此，从 10 个信号通道构造 40 项统计变量，比较 11 种机器学习分类器与 6 种时序深度学习模型，再开展训练集内的 SHAP 筛选及 AdaBoost 预测解释。补充实验考察阶段划分、信号去噪、新变量构造和加工条件变化对结果的影响。

## 2. 数据与多轴信号表征

### 2.1. 数据来源与识别任务

研究采用 CNC Mill Tool Wear 数据集[18]，数据记录铣床在蜡块上完成“S”形加工的过程，包含 18 个时间序列数据实验文件和一个实验标签表。经核对，原始文件共记录了 25,286 行时间序列数据，每个时间序列文件包含 48 个原始字段，标称采样间隔为 100 ms，约为 10 Hz。该记录主要反映机床控制与电机运行状态。标签表中磨损实验 10 次，未磨损实验 8 次；18 次实验包括 14 次完成加工和 4 次中止加工，涉及 10 种进给率与夹紧压力组合。

本研究令磨损状态为正类、未磨损状态为负类，预测对象为一次加工实验的刀具状态。完成加工与中止加工的实验均作为独立分组纳入评价。每个实验形成一条统计特征记录，或一组具有共同实验编号的时间序列窗口。有效评价样本数为 18，窗口数和采样点数不计作独立标签样本数。刀具状态、加工完成标志和视觉质检结果不进入模型输入；实验编号用于关联文件与分组。进给率和夹紧压力用于工况分析及构造变量的辅助对照，不进入主要模型比较所用的 40 项统计变量。

根据原始过程字段，保留 Layer 1 Up、Layer 1 Down、Layer 2 Up、Layer 2 Down、Layer 3 Up 和 Layer 3 Down 六个切削阶段，排除空行程。主要模型比较将六个阶段的有效采样点合并，在整个已观测切削过程上计算统计变量，而不是为每个阶段分别生成一套输入。对于中止加工的实验，使用中止前已记录的切削信号。因此，本文预测对应加工结束或中止后的状态评估。表 1 给出样本与输入结构。

### 2.2. 信号选择与统计变量

本文选取 10 个信号通道：X 轴、Y 轴与 S 轴的电流反馈，主轴输出功率，X 轴、Y 轴与 S 轴的实际速度，以及对应三个轴的速度误差。其中速度误差由实际速度减指令速度得到。电流与功率用于描述驱

动负载，实际速度与速度误差用于描述运动状态及跟踪偏差。特征计算保留原始字段的数值尺度，再使用训练集计算的参数进行标准化，使不同通道的输入尺度便于比较。

**Table 1.** Dataset and input structure
**表 1.** 数据集与输入结构

| 项目 | 数量或定义 |
|---|---|
| 独立加工实验 | 18 次；磨损 10 次，未磨损 8 次 |
| 时序记录 | 25,286 行；48 个原始字段 |
| 采样间隔 | 标称 100 ms，约 10 Hz |
| 切削阶段 | 第 1 至第 3 层各含 Up 与 Down 两个阶段，共六阶段 |
| 主统计输入 | 10 个通道，每通道 4 项，共 40 项变量 |
| 标签与质检字段 | 不作为输入 |
| 评价单位与预测时点 | 加工实验；加工结束或中止后 |

对于每个通道，在一次实验的切削采样点上计算中位数、四分位距、均方根及 90%与 10%分位数之差。每个通道提取 4 项，即在 10 个通道共提取 40 项变量。其中中位数表示典型水平；四分位距和分位数跨度描述变化范围；均方根同时受幅值和波动影响。四分位距为 75%分位数与 25%分位数之差，分位数跨度为 90%分位数与 10%分位数之差。均方根是将该实验有效切削采样值分别平方、求平均后再开平方所得。

采用分位数统计可以减少个别极端值对特征的支配，同时保留加工过程中负载和运动变化。40 项变量共同描述不同轴的信号水平和波动情况，不预先依据全体标签进行删减。另一组对照实验在六个切削阶段内分别计算相同的统计量，最多形成 240 项变量，用于检验进一步细分阶段的作用。

### 2.3. 补充实验的变量构造

除了主要模型比较使用的 40 项统计变量外，我们还构造 50 项新变量，检验工况校正与窗口变化是否提供额外识别信息。设置两组使用相同实验划分的对照：一组保留原有 40 项变量，另一组在此基础上增加 50 项新变量。新变量包括负载残差分布、相对四分位距、相邻窗口负载变化以及速度误差的窗口统计与变化。窗口长度为 32，步长为 16，仅从同一切削阶段且原始采样序号连续的片段中提取；8 至 31 点的短片段形成一个窗口，少于 8 点的片段不生成窗口。

负载残差为观测负载与相似工况下预期负载之差。预期负载模型采用正则化系数为 10 的岭回归，以阶段、进给率、夹紧压力与速度信息为输入。对训练集中的每个实验，先用其余训练实验拟合负载模型，再计算该实验的残差；验证和测试实验则使用对应训练集拟合的负载模型计算残差。该过程在每个训练折内重新执行，避免被评价实验参与自身负载基线的拟合。

其余补充方案包括：在六个阶段内分别计算统计变量；采用 db2 小波进行一层软阈值去噪后提取阶段统计量，去噪阈值仅从对应训练实验的连续片段估计；仅用进给率与夹紧压力作为输入；以及分别留出加工条件组合或只评价完成加工的实验。各方案的结果见第 4.5 节。

## 3. 模型与分组评价方法

实验流程如图 1 所示。首先，以加工实验为单位划分训练集、内层验证集和外层测试集；其次，分别比较机器学习分类器与时序深度学习模型，并检验 SHAP 筛选后的预测效果；最后，解释全变量

AdaBoost 对留出实验的预测。由此将模型性能比较、变量筛选效果和信号解释联系起来。

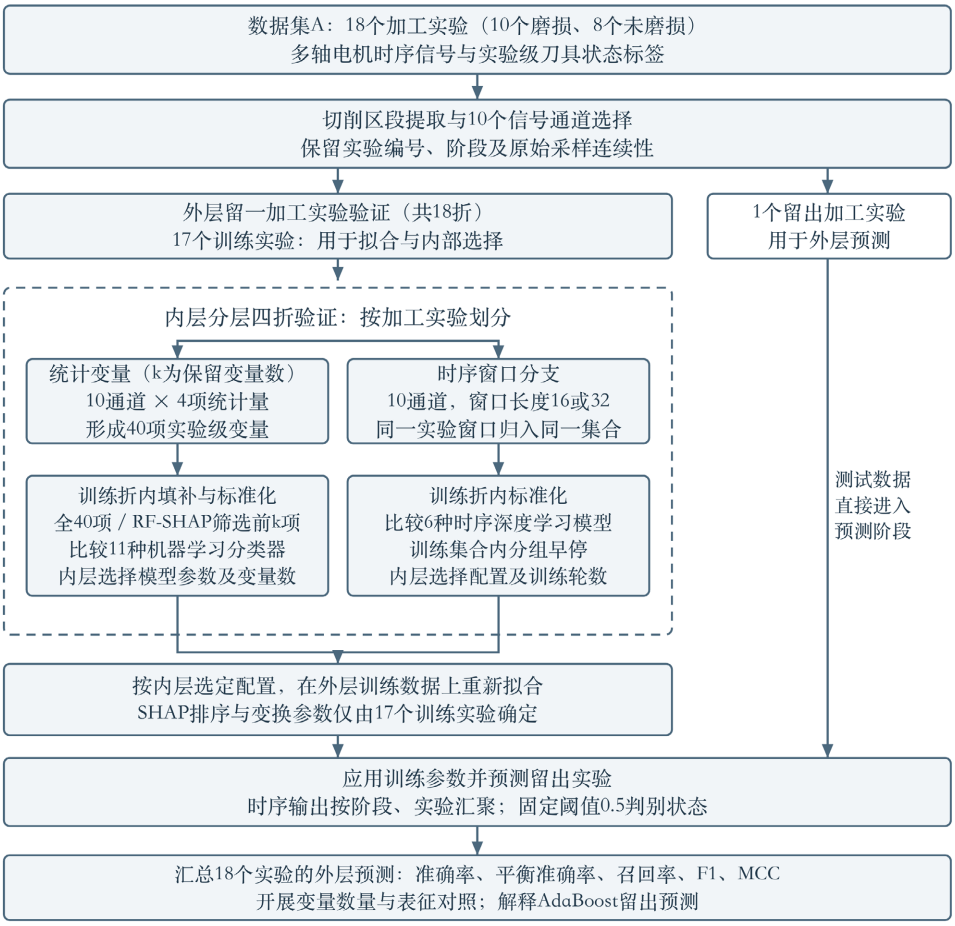



**Figure 1.** Experimental workflow for tool wear identification and SHAP analysis
**图 1.** 刀具磨损识别与 SHAP 分析的实验流程

### 3.1. 按加工实验分组的嵌套交叉验证

外层采用留一加工实验验证，共 18 折。每折留出一个加工实验作为测试对象，其余 17 个实验用于模型训练和内部选择。传统模型内层采用分层四折验证，在实验级标签记录上划分；时序深度学习模型也先划分实验编号，再将窗口分配到对应集合。同一实验的所有窗口均属于同一集合，切窗过程中不跨越阶段边界或原始采样间断。

统计变量模型在各训练折内剔除常量或全缺失变量，以训练折中位数填补缺失值并拟合标准化参数。测试记录应用训练参数变换。SHAP 变量排序仅在内层训练集上计算，模型也仅在该训练集上拟合。内层验证集用于比较保留不同数量变量与采用不同模型参数时的识别效果，从中确定变量数量和模型参数；外层测试标签仅用于最终评价。全部外层预测完成后，汇总 18 个实验的准确率、平衡准确率、类别召回率、F1 值及 Matthews 相关系数(Matthews Correlation Coefficient, MCC)。

第 4.1 至 4.4 节的主要模型比较均优先选择内层验证准确率最高的方案。准确率相同时，依次比较平衡准确率、MCC 和保留变量数量；仍相同时，按预先固定的模型与参数顺序选择。补充实验的模型选择与验证设置见第 4.5 节。

### 3.2. 传统机器学习分类器与参数设置

传统机器学习分类器包括逻辑回归(Logistic Regression, LR)、支持向量机(Support Vector Machine, SVM)、随机森林(Random Forest, RF)、决策树(Decision Tree, DT)、K 近邻(K-Nearest Neighbors, KNN)、高斯朴素贝叶斯(Gaussian Naive Bayes, GNB)、自适应提升(AdaBoost)、梯度提升决策树(Gradient Boosting Decision Tree, GBDT)、极端梯度提升(XGBoost)、轻量梯度提升机(LightGBM)和多层感知机(Multilayer Perceptron, MLP)。每种模型比较两组预先设定的参数，具体取值见表 2。

**Table 2.** Prespecified parameter settings of classical classifiers
**表 2.** 传统分类器的预设参数

| 模型 | 两套配置的主要参数 |
|---|---|
| LR | C 取 0.1 或 1；最大迭代 2000 次 |
| SVM | RBF 核；C 取 1 或 10；gamma 采用 scale |
| RF | 150 棵树；深度 2 或不限；叶节点最少 2 样本 |
| DT | 深度 2 或 3；叶节点最少 2 样本 |
| KNN | 邻居数 3 或 5；距离加权 |
| GNB | 平滑参数取十亿分之一或千分之一 |
| AdaBoost | 100 个树桩；SAMME；学习率取 0.1 或 1.0 |
| GBDT | 80 棵树；深度 1 或 2；学习率 0.05；叶节点最少 2 样本 |
| XGBoost | 80 棵树；深度 1 或 2；学习率 0.05；正则化系数 10 |
| LightGBM | 80 棵树；叶子数 2 或 4；学习率 0.05；正则化系数 10 |
| MLP | 隐藏层为 8 个节点或 16 与 8 个节点；正则化系数取 1.0 或 0.1；tanh；L-BFGS |

传统机器学习分类器比较分为两个层次。第一层次是分别评价 11 种分类器：每次固定分类器种类，由内层验证选择参数，再预测外层留出的实验。第二层次是评价自动选择模型的效果：每个外层训练集均通过内层验证从多种模型中选择模型及其参数，再预测留出实验。第一层次得出各分类器的识别效果，第二层次得出答由训练数据选择模型后能够取得的预测效果。两者分别报告。

每种分类器均比较使用全部 40 项变量与 SHAP 筛选变量的结果。AdaBoost 采用 SAMME 算法，以 100 个决策树桩为弱分类器，学习率在 0.1 和 1.0 中选择，通过迭代调整训练样本权重形成最终分类结果。各模型的随机种子固定为 20261001。

### 3.3. 时序深度学习模型与加工实验预测结果的计算

时序深度学习模型包括了循环神经网络(Recurrent Neural Network, RNN)、长短期记忆网络(Long Short-Term Memory, LSTM)、双向长短期记忆网络(BiLSTM)、门控循环单元(Gated Recurrent Unit, GRU)、一维卷积网络(1D-CNN)和二维卷积网络(2D-CNN)。所有模型均输入 10 个信号通道。两套配置分别采用窗口长度 16、隐藏宽度 8、学习率 0.003，以及窗口长度 32、隐藏宽度 16、学习率 0.001。窗口在阶段内连续且互不重叠，仅保留长度满足对应配置且各通道采样值均为有限值的窗口；2D-CNN 处理时间与按固定顺序排列的通道矩阵，未将信号转换为频谱图。

RNN、LSTM、BiLSTM 和 GRU 均采用单层循环结构。RNN、LSTM 和 GRU 取最终隐藏状态，BiLSTM 拼接两个方向的最终隐藏状态。1D-CNN 和 2D-CNN 均包含两层卷积及 ReLU 激活，随后进行全局平均池化；卷积核大小分别为 3 和 $3 \times 3$。循环层隐藏维数与卷积层输出通道数均按对应配置取 8 或

16。各网络通过 dropout 和全连接层输出类别得分，预测时经 sigmoid 转换为磨损概率。训练采用 Adam 优化器，梯度范数上限为 1。

网络最多训练 100 个 epoch，早停参数 patience 为 12，即早停集合上的损失连续 12 个 epoch 未改善时停止训练；dropout 为 0.2，权重衰减为 0.001。每个内层训练集另保留一个磨损实验和一个未磨损实验作为早停集合，其窗口不参与模型权重更新。外层重新训练时，采用各内层折最佳 epoch 数的中位数，使用全部 17 个外层训练实验拟合模型。

时序深度学习模型的标准化参数只从参与模型拟合的训练实验中计算。先计算各实验每个通道的均值与平方值均值，再对实验等权平均，得到总体均值和标准差，使每个实验对标准化参数的贡献相同，避免长实验因窗口更多而占据较大权重。每个 epoch 从各训练实验抽取 8 个窗口，抽样时先等概率选择切削阶段，再从该阶段抽取窗口。

损失函数采用带正类权重的二元交叉熵，权重由当前拟合集合中未磨损与磨损实验数量之比确定。测试时先平均同一阶段的窗口预测概率，再等权平均该实验已观测阶段的概率，使用固定 0.5 阈值判断状态。通过上述平均计算，将多个窗口的预测转换为一次加工实验的预测状态，使时序深度学习模型与机器学习分类器均按加工实验评价性能。

### 3.4. SHAP 变量筛选与预测解释

SHAP (Shapley Additive Explanations)将模型输出相对于背景的变化分配给输入变量，以各变量贡献之和解释预测结果[19]。本文将 SHAP 用于两项不同工作：在训练集内排序并筛选变量，以及解释模型对留出实验的预测。变量筛选采用训练折内拟合的随机森林，设置 200 棵树、最大深度 3、叶节点最少 2 个样本，按磨损类平均绝对 TreeSHAP 值排序[20]。以 k 表示保留的变量数量，比较 k 为 3、5、8、10、15、20 和 40 的方案。每一折均重新计算排序；固定 k 的实验只选择模型和参数，自动确定 k 的实验同时选择变量数量、模型和参数。

预测解释针对使用全部 40 项变量的 AdaBoost 开展。对每个留出实验，使用实际生成其预测的模型，以对应 17 个训练实验作为背景，采用 Permutation SHAP 解释磨损类概率输出。使用 10 组正向与反向排列近似，并核对贡献加和与模型输出的一致性。对 18 个测试实验的绝对贡献取平均值，我们可以得到变量排名。该排名用于说明预测依据，不再反馈到变量筛选。SHAP 值为正或为负，分别表示相对于背景推动预测向磨损或未磨损方向，表示模型贡献而非物理因果关系。

## 4. 实验结果与分析

本节依次讨论模型识别效果、SHAP 变量筛选效果及 AdaBoost 判断依据。补充实验另行比较变量构造和加工条件的影响，以说明不同结果分别对应的实验设置。

本研究以一次加工实验为评价单位，汇总全部外层留出预测后计算评价指标。将磨损定义为正类、未磨损定义为负类：TP 为正确识别的磨损实验数，TN 为正确识别的未磨损实验数，FP 为将未磨损误判为磨损的实验数，FN 为将磨损误判为未磨损的实验数。各指标计算如下：

$$\text{准确率} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$\text{磨损召回率} = \frac{TP}{TP + FN} \tag{2}$$

$$\text{未磨损召回率} = \frac{TN}{TN + FP} \tag{3}$$

$$\text{平衡准确率} = \frac{\text{磨损召回率} + \text{未磨损召回率}}{2} \tag{4}$$

$$\text{F1值} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}} \tag{5}$$

$$\text{MCC} = \frac{\text{TP} \times \text{TN} - \text{FP} \times \text{FN}}{\sqrt{(\text{TP} + \text{FP})(\text{TP} + \text{FN})(\text{TN} + \text{FP})(\text{TN} + \text{FN})}} \tag{6}$$

准确率反映总体正确识别比例；磨损召回率反映磨损检出能力，未磨损召回率反映未磨损状态的正确识别能力。平衡准确率取两类召回率的平均值，使两类获得同等权重[21]。F1 值综合磨损类的精确率与召回率，其中精确率指被预测为磨损的实验中实际磨损的比例。准确率、召回率和平衡准确率以百分数报告，F1 值和 MCC 保留小数。

MCC 综合考虑 TP、TN、FP 和 FN，衡量预测类别与真实类别的一致程度，取值范围为−1 至 1，越接近 1 表示一致程度越高，0 表示无相关性，负值表示反向相关。本研究使用 MCC 补充评价分类效果；在第 4.1 至 4.4 节的主要模型比较中，内层验证的准确率和平衡准确率均相同时，优先选择 MCC 较高的模型配置。当 F1 或 MCC 的计算分母为 0 时，按实验代码约定记为 0。

### 4.1. 传统机器学习分类器的识别性能

表 3 和图 2 展示 11 种机器学习分类器在两种变量输入方案下的准确率。使用全部 40 项变量时，AdaBoost、逻辑回归和 GBDT 均正确识别 14 个实验，准确率为 77.78%。其中，AdaBoost 的平衡准确率为 77.50%，高于 LR 和 GBDT 的 76.25%；其磨损召回率为 80.00%，未磨损召回率为 75.00%。逻辑回归和 GBDT 的磨损召回率为 90.00%，未磨损召回率为 62.50%，表明不同模型对磨损漏报与未磨损误报的表现有所差异。

进一步比较类别指标，AdaBoost 的磨损类 F1 值为 0.800，MCC 为 0.550；逻辑回归与 GBDT 的 F1 值为 0.818，MCC 约为 0.553。三种模型的准确率相同，但 AdaBoost 对磨损与未磨损的召回率更接近，适合作为同时重视漏报与误报的分析对象。本文因此选取使用全部 40 项变量的 AdaBoost 开展后续预测解释；逻辑回归与 GBDT 则提供更高磨损召回率的对照。

**Table 3.** Classical classifier performance with all and SHAP-selected features
**表 3.** 传统模型使用全部变量与 SHAP 筛选变量的识别性能

| 模型 | 全变量<br>准确率/% | 全变量<br>平衡准确率/% | SHAP 筛选<br>准确率/% | SHAP 筛选<br>平衡准确率/% |
|---|---|---|---|---|
| LR | 77.78 | 76.25 | 66.67 | 65.00 |
| SVM | 72.22 | 70.00 | 66.67 | 66.25 |
| RF | 61.11 | 60.00 | 66.67 | 66.25 |
| DT | 44.44 | 43.75 | 66.67 | 66.25 |
| KNN | 55.56 | 51.25 | 66.67 | 66.25 |
| GNB | 72.22 | 70.00 | 72.22 | 70.00 |
| AdaBoost | 77.78 | 77.50 | 72.22 | 70.00 |
| GBDT | 77.78 | 76.25 | 61.11 | 58.75 |
| XGBoost | 55.56 | 50.00 | 55.56 | 50.00 |
| LightGBM | 50.00 | 47.50 | 61.11 | 58.75 |
| MLP | 72.22 | 70.00 | 72.22 | 71.25 |

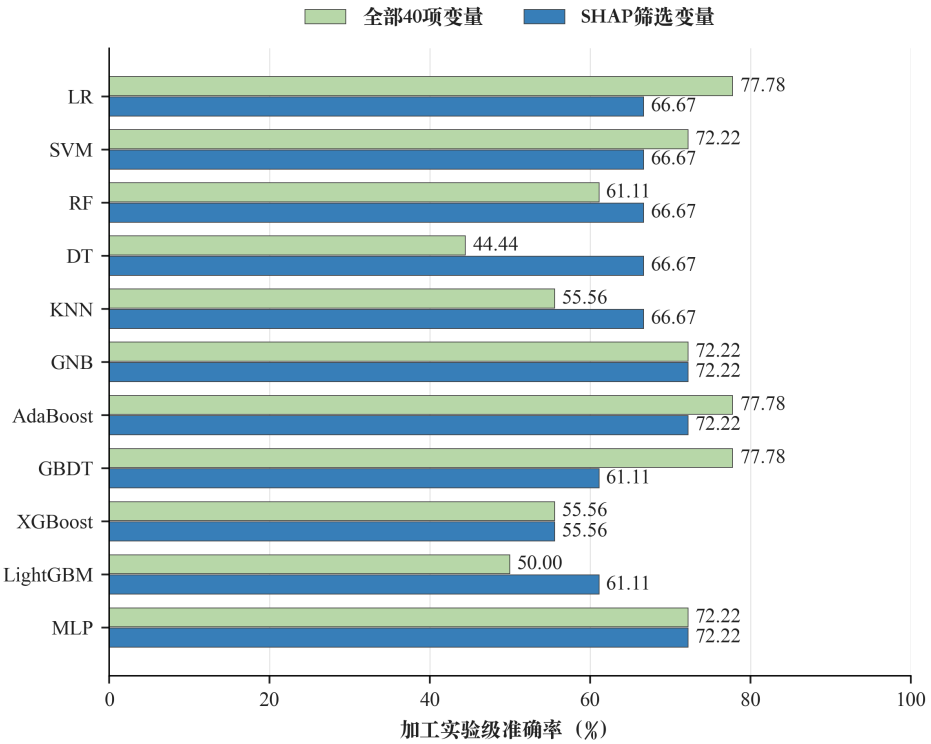



**Figure 2.** Accuracy of classical models with all and SHAP-selected features
**图 2.** 传统模型全变量与 SHAP 筛选变量的准确率对照

表 3 同时显示，SHAP 筛选的作用随分类器而变化：决策树准确率由 44.44%变为 66.67%，KNN 由 55.56%变为 66.67%，LightGBM 由 50.00%变为 61.11%，随机森林由 61.11%变为 66.67%；AdaBoost 筛选后的准确率为 72.22%。这些结果说明，变量筛选是否有益需结合具体分类器判断。第 4.3 节进一步比较保留不同变量数量及自动选择分类器时的结果。

## 4.2. 时序深度学习模型的性能比较

时序深度学习模型的结果如表 4 和图 3 所示。1D-CNN 的准确率为 66.67%，平衡准确率为 67.50%，磨损与未磨损召回率分别为 60.00%和 75.00%；2D-CNN 的准确率为 61.11%，平衡准确率为 61.25%。GRU 的准确率为 50.00%，其余循环网络在当前配置下为 22.22%–38.89%。在所比较的 6 种时序深度学习模型中，1D-CNN 的准确率和平衡准确率均最高。

**Table 4.** Run-level performance of sequence networks
**表 4.** 时序深度学习模型的加工实验级识别性能

| 模型 | 准确率/% | 平衡准确率/% | 磨损召回率/% | 未磨损召回率/% |
|---|---|---|---|---|
| RNN | 38.89 | 43.75 | 0.00 | 87.50 |
| LSTM | 22.22 | 21.25 | 30.00 | 12.50 |
| BiLSTM | 27.78 | 25.00 | 50.00 | 0.00 |
| GRU | 50.00 | 52.50 | 30.00 | 75.00 |
| 1D-CNN | 66.67 | 67.50 | 60.00 | 75.00 |
| 2D-CNN | 61.11 | 61.25 | 60.00 | 62.50 |

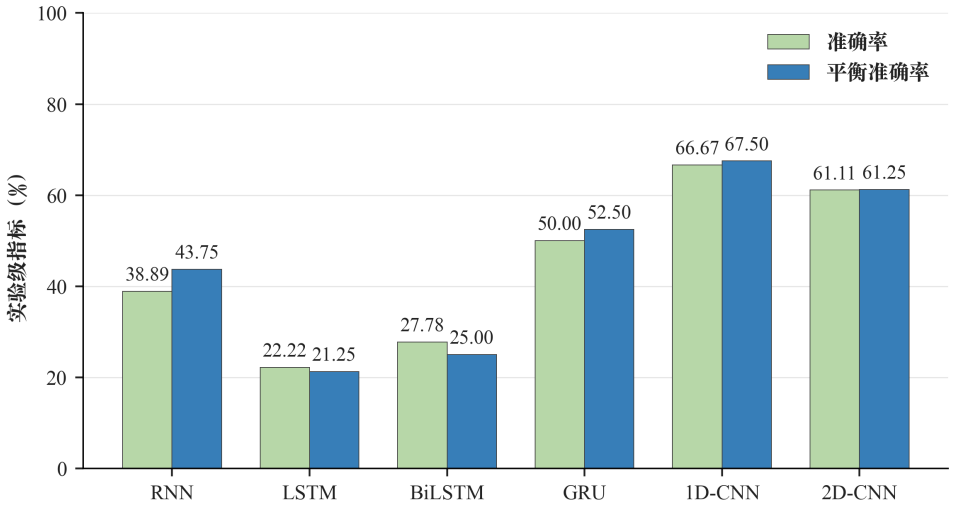



**Figure 3.** Accuracy and balanced accuracy of sequence models
**图 3.** 时序深度学习模型的准确率与平衡准确率

在相同的加工实验评价单位下，使用 40 项统计变量的 AdaBoost、逻辑回归和 GBDT 均达到 77.78% 的准确率，高于本次时序深度学习模型中的最高值 66.67%。统计变量概括整个切削过程的信号水平与波动，时序深度学习模型则从局部窗口学习信号变化。尽管窗口数量较多，独立的实验标签仍为 18 个，增加窗口并未增加独立加工实验数量。因此，在本文数据和参数设置下，统计变量与传统机器学习分类器的组合取得了较高的识别效果。

### 4.3. SHAP 变量筛选对识别效果的影响

固定变量数量实验在 LR、SVM、RF、XGBoost 和 LightGBM 分类器中通过内层验证选择模型和参数。变量排名在每个训练折内重新计算；保留前 15 项时，固定的是变量数量，具体变量清单可随折次变化。图 4 显示，保留 SHAP 排名前 15 项与全部 40 项变量的准确率均为 72.22%，平衡准确率分别为 71.25% 和 70.00%。保留前 15 项使输入变量减少 62.5%，同时保持正确识别 13 个实验。

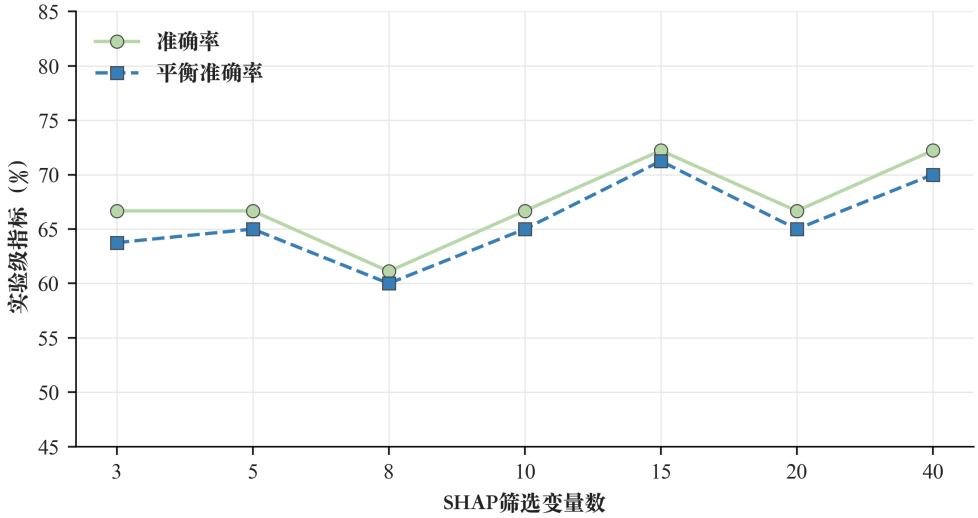



**Figure 4.** Performance at different SHAP feature counts
**图 4.** 不同 SHAP 变量数量下的预测性能

**Table 5.** SHAP feature selection under different model selection settings
**表 5.** 不同模型选择设置下的 SHAP 变量筛选结果

| 实验设置 | 分类器范围 | 全变量准确率/% | 筛选变量准确率/% | 变量数量确定方式 |
|---|---|---|---|---|
| 固定保留前 15 项 | 5 种 | 72.22 | 72.22 | 每折固定保留 15 项 |
| 自动选择变量数量 | 5 种 | 72.22 | 66.67 | 内层验证确定 |
| 自动选择变量数量 | 11 种 | 66.67 | 72.22 | 内层验证确定 |

表 5 比较了不同候选分类器范围下，固定保留 SHAP 排名前 15 项变量与自动确定变量数量两种策略的分类准确率。在 5 种分类器范围内，固定保留 15 项与使用全部 40 项变量的准确率均为 72.22%，输入变量减少 62.5%；同一范围内，自动确定变量数量的准确率为 66.67%。将候选范围扩展至 11 种分类器后，自动筛选的准确率达到 72.22%，较对应的全变量对照(66.67%)提高 5.56 个百分点。

在 11 种分类器的自动选择设置中，SHAP 筛选使磨损召回率保持 90.00%，未磨损召回率由 37.50% 提高至 50.00%，即多正确识别一个未磨损实验。结合固定 15 项的结果，SHAP 筛选在本实验中既支持以较少变量保持分类准确率，也改善了特定设置下的未磨损状态识别。不同设置的结果说明，变量数量与候选分类器范围需共同考虑，不能单凭保留变量的多少判断筛选效果。

## 4.4. AdaBoost 的类别错误与信号解释

AdaBoost 的 18 个留出预测中，10 个磨损实验有 8 个被正确识别，8 个未磨损实验有 6 个被正确识别，漏报与误报各 2 次。作为简单参照，若始终预测为样本较多的磨损类，准确率为 55.56%，平衡准确率为 50.00%；AdaBoost 分别高出 22.22 和 27.50 个百分点。这说明其预测能够区分两类状态。在此基础上，使用 SHAP 分析模型判断所依据的信号。

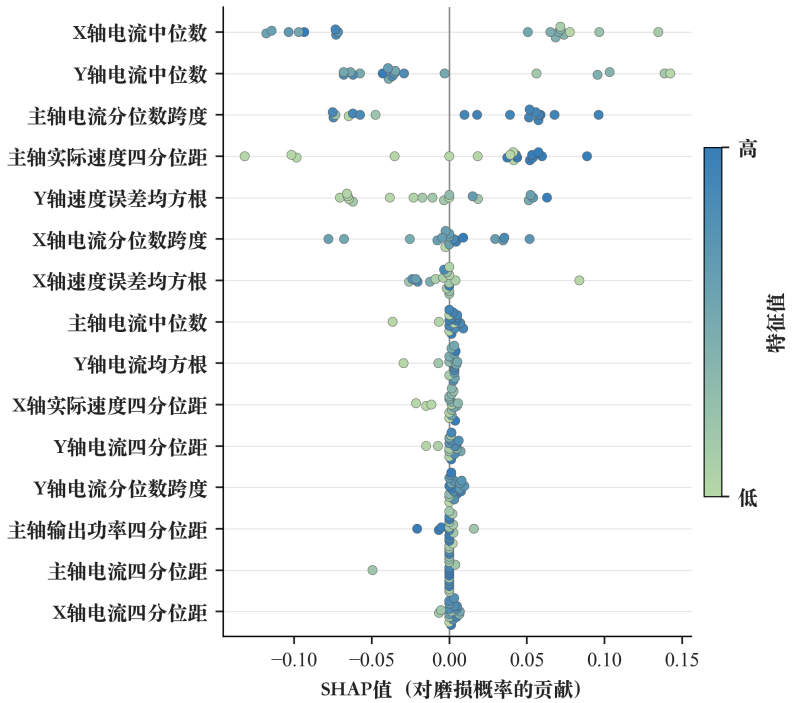



**Figure 5.** Held-out SHAP explanations of AdaBoost
**图 5.** 使用全部 40 项变量的 AdaBoost 对测试实验的 SHAP 解释

图 5 展示 SHAP 对 AdaBoost 测试实验预测结果的解释。平均绝对贡献前五项依次为 X 轴电流中位数、Y 轴电流中位数、主轴电流分位数跨度、主轴实际速度四分位距和 Y 轴速度误差均方根，其平均绝对贡献约为 0.0847、0.0620、0.0566、0.0554 和 0.0375，单位为磨损概率输出的贡献量。前两项对应典型驱动电流水平，第三项对应主轴电流变化范围，第四项对应主轴实际速度的波动范围，第五项对应 Y 轴运动跟踪偏差幅值。

这些变量对应驱动负载水平、负载变化和运动跟踪偏差，为维护人员复核预测提供了具体的信号方向。蜂群图显示，同一变量在不同实验中的贡献方向和大小并不相同，因此应结合该实验的其他信号及加工条件理解预测，不能由全局排名直接设定统一电流阈值。前五项变量用于说明主要判断依据；实际预测仍使用全部 40 项变量。

### 4.5. 变量构造与加工条件的补充实验

补充实验以平衡准确率为模型选择依据，从 LR、SVM、RF、XGBoost 和 LightGBM 的 60 组预设配置中选择分类器、参数及变量处理方式；变量处理采用训练集内的方差分析筛选或 PCA。留一加工实验的对照在内层采用分层四折验证并重复三次；留一工况组合的对照采用内层三折分组验证并重复三次。表 6 汇总各方案的结果。该组选择依据与参数范围不同于第 4.1 节，各方案在对应设置内比较。

**Table 6.** Supplementary results under different feature and evaluation settings
**表 6.** 不同变量方案与评价设置下的补充实验结果

| 方案 | 评价范围与外层划分 | 准确率/% | 平衡准确率/% |
|---|---|---|---|
| 整段切削统计(40 项) | 18 次实验；留一加工实验 | 61.11 | 60.00 |
| 分阶段统计(最多 240 项) | 18 次实验；留一加工实验 | 44.44 | 43.75 |
| 小波去噪后分阶段统计 | 18 次实验；留一加工实验 | 61.11 | 58.75 |
| 仅进给率与夹紧压力 | 18 次实验；留一加工实验 | 33.33 | 30.00 |
| 原 40 项增加 50 项新变量 | 18 次实验；留一加工实验 | 55.56 | 51.25 |
| 分阶段统计，按工况留出 | 18 次实验；留一工况组合 | 61.11 | 57.50 |
| 分阶段统计，完成加工子样本 | 14 次实验；留一加工实验 | 71.43 | 70.83 |

(1) 整段切削统计：确定比较基准。为了判断改变变量构造方式是否有帮助，先将整个切削过程汇总为 40 项统计变量，作为本组实验的比较基准。在 18 次实验的留一加工实验评价中，准确率和平衡准确率分别为 61.11%和 60.00%。分阶段统计、去噪和新增变量方案均与这一结果对照，以判断改动后能否提高识别效果。

(2) 分阶段统计：检验细分加工阶段是否有帮助。不同切削阶段的负载和运动状态可能不同，因此在六个阶段分别计算统计量，形成最多 240 项变量。该方案的准确率为 44.44%，平衡准确率为 43.75%，均低于整段切削统计方案。这说明，在本实验中，按阶段细分并增加变量数量，并没有比整段统计取得更好的识别结果。

(3) 小波去噪：检验减少信号噪声能否改善阶段统计结果。在计算各阶段的统计量前，先采用 db2 小波去噪，去噪阈值由训练实验估计。准确率和平衡准确率分别为 61.11%和 58.75%，高于未去噪方案的 44.44%和 43.75%。这说明，去噪有助于改善本组分阶段统计方案的结果。不过，其准确率与整段 40 项变量方案相同，平衡准确率略低，因此主要模型比较仍采用整段统计变量。

(4) 工况变量：检验只用加工设置能否识别磨损。为判断加工设置本身能提供多少识别信息，仅以进给率与夹紧压力作为输入。准确率为 33.33%，平衡准确率为 30.00%。这说明，在本数据和评价方式下，

仅凭这两项加工设置还不能有效区分磨损与未磨损状态。因此，本研究采用多轴运行信号提供的电流、速度等信息进行识别。

(5) 新增变量：检验加入更多信号信息是否有帮助。为进一步利用不同工况下的负载差异和短时变化信息，在原有 40 项变量上增加 50 项负载残差、相对波动及窗口变化变量。准确率为 55.56%，平衡准确率为 51.25%，低于原 40 项变量方案的 61.11%和 60.00%。磨损召回率由 70.00%提高至 90.00%，未磨损召回率则由 50.00%降至 12.50%。也就是说，新增变量使更多磨损实验被正确检出，但未磨损实验也更容易被误判为磨损。因此，主要模型比较保留原有 40 项变量。

(6) 按工况留出：检验模型对训练中未见工况的识别效果。将进给率与夹紧压力组成的 10 种工况逐一用于测试：每次留出一种工况，用其余工况训练模型，输入仍为分阶段统计变量。汇总 18 次实验的预测后，准确率为 61.11%，平衡准确率为 57.50%。与逐次留出加工实验相比，这种划分要求测试工况不出现在训练集中，结果用于说明模型面对新工况时的识别表现。

(7) 完成加工子样本：查看完整加工记录上的识别效果。为了解只使用完成加工记录时的结果，从 18 次实验中选出 14 次完成加工的记录，采用分阶段统计变量进行留一加工实验评价。准确率和平衡准确率分别为 71.43%和 70.83%。这一结果对应完成加工的 14 次实验。由于使用的实验记录和训练数据都与 18 次实验的方案不同，两组结果分别报告，有助于说明识别结果适用于哪些加工记录。

## 5. 面向预测性维护的应用讨论

### 5.1. 识别结果在刀具维护中的应用

本文的识别结果可用于加工结束后的刀具状态复核。维护人员可先查看磨损状态预测，再结合关键变量贡献检查电流水平、主轴波动和运动跟踪偏差，并对照加工条件、刀具使用记录及人工检查结果判断是否需要进一步检查。SHAP 在这一过程中提供可检查的信号依据，帮助维护人员理解模型输出。

维护辅助需要将运行信号转化为可供检查的状态信息。已有研究采用聚类和组合模型分析设备故障[22] [23]，通过状态监测及信号分解与分类方法的结合识别设备异常[24] [25]，并将集成模型与工业数据采集架构用于预测性维护和生产应用[26] [27]。对本文的刀具状态识别而言，模型提供状态预测，SHAP 指出贡献较大的信号；维护人员再结合刀具使用记录与现场检查作出判断。企业可在这一基础上积累预测、复核和处理记录，使模型结果与实际维护工作相衔接。

### 5.2. 评价范围与后续验证

本文结论对应 18 次 CNC 蜡铣削实验及预设模型参数范围。不同方法均使用相同的加工实验，文中分别报告单个模型的性能与内层验证自动选择模型后的性能，以区分两类评价结果。按实验划分训练与测试避免了同一实验的数据交叉使用，但这些结果仍需由新加工实验检验；目前数据没有可核实的物理刀具或设备批次标识，尚不能据此评价跨刀具或跨设备的识别效果。

后续可采集记录具体刀具、设备和加工时间的新批次数据，并在评价前固定模型、变量、参数选择方法与分类阈值。同步记录刀具检查、误报复核和换刀处理结果，可进一步分析状态识别对维护工作的作用。本文的概率输出用于分类与 SHAP 解释；若进一步估计故障发生概率，需进行独立概率校准；若预测剩余寿命，需采集磨损量、退化轨迹及失效时间；若评价维护效益，需补充停机时间、加工质量与维护成本等记录。

## 6. 结论

本文构建了结合多轴统计变量、加工实验分组验证与 SHAP 解释的 CNC 刀具磨损状态识别方法。以

加工实验为分组与评价单位，将预处理、参数选择及变量筛选限定在训练数据内。模型比较表明，在本文参数设置下，使用全部 40 项变量的 AdaBoost、逻辑回归与 GBDT 准确率均为 77.78%。其中 AdaBoost 的平衡准确率为 77.50%，磨损与未磨损召回率分别为 80.00%和 75.00%，两类识别较为均衡。6 种时序深度学习模型中，1D-CNN 的准确率和平衡准确率最高，分别为 66.67%和 67.50%。

SHAP 变量筛选提高了 DT、KNN、LightGBM 和 RF 的准确率；在 5 分类器选择实验中，保留前 15 项变量与使用全部 40 项变量均取得 72.22%的准确率，输入变量减少 62.5%。对全变量 AdaBoost 的预测解释进一步表明，X、Y 轴电流水平、主轴电流与速度波动以及 Y 轴速度误差是主要判断依据。研究将识别性能、变量筛选效果与信号解释联系起来，为加工结束后的刀具状态复核及维护辅助提供方法参考。

## 基金项目

本研究由 2026 年度清远市科技智库专项课题(QYZKZZ-2026-32)，广东金融学院基金项目(0000-KC2026004004004)，广东金融学院科研启动基金(0000-KC2026014001126)资助。

## 参考文献

[1] 张媛媛, 杨奕东, 陈晓丽. 清远传统制造业数字化转型问题与对策研究[J]. 商场现代化, 2026(1): 5-8.

[2] Vial, G. (2019) Understanding Digital Transformation: A Review and a Research Agenda. *The Journal of Strategic Information Systems*, **28**, 118-144. https://doi.org/10.1016/j.jsis.2019.01.003

[3] 田佩彬, 王宝金, 沈锦桃. 刀具磨损状态智能监测技术研究现状与展望[J]. 林业机械与木工设备, 2021, 49(12): 4-8, 20.

[4] Munaro, R., Attanasio, A. and Del Prete, A. (2023) Tool Wear Monitoring with Artificial Intelligence Methods: A Review. *Journal of Manufacturing and Materials Processing*, **7**, Article 129. https://doi.org/10.3390/jmmp7040129

[5] Mohamed, A., Hassan, M., M'Saoubi, R. and Attia, H. (2022) Tool Condition Monitoring for High-Performance Machining Systems—A Review. *Sensors*, **22**, Article 2206. https://doi.org/10.3390/s22062206

[6] 王民, 车昌家, 高相胜, 咎涛, 高鹏, 张云飞. 复杂工况条件下刀具磨损状态监测与剩余寿命预测研究进展[J]. 机械工程学报, 2025, 61(19): 299-326.

[7] 钟荣锋. 基于时频融合与故障频率特征提取的智能诊断技术应用[J]. 传感器技术与应用, 2025, 13(6): 951-962.

[8] 杨好, 李军成, 杨磊. 基于小波分析与神经网络的电子设备故障诊断[J]. 仪器与设备, 2024, 12(2): 230-238.

[9] 张嵘, 朱寅, 李帅. 基于机器学习的齿轮箱故障诊断[J]. 建模与仿真, 2023, 12(2): 1415-1427.

[10] 海天恒. 基于二维卷积神经网络的轴承故障智能诊断[J]. 建模与仿真, 2024, 13(2): 1333-1345.

[11] 吴英豪, 姚丹妍, 李蓉, 曹正豪, 肖思凯. 基于多传感器融合与深度学习的工业轴承源域故障诊断[J]. 图像与信号处理, 2026, 15(2): 221-234.

[12] 高程远. 基于融合注意力的卷积神经网络滚动轴承故障诊断模型[J]. 建模与仿真, 2024, 13(5): 5503-5512.

[13] 范慧鹏, 李瑞华, 李福林, 房哲续, 彭六保. 深度学习算法在工业设备故障诊断应用研究[J]. 传感器技术与应用, 2021, 9(4): 193-200.

[14] 胡惠媛, 赵祖玲. 基于深度学习的电气自动化设备故障诊断与预测技术研究[J]. 传感器技术与应用, 2025, 13(3): 447-458.

[15] 成佳闻, 赛希亚拉图, 张超勇, 罗敏. 基于注意力机制堆叠 LSTM 的多传感器信息融合刀具磨损预测[J]. 工业工程, 2024, 27(3): 64-77, 86.

[16] Roberts, D.R., Bahn, V., Ciuti, S., Boyce, M.S., Elith, J., Guillera-Arroita, G., *et al*. (2017) Cross-Validation Strategies for Data with Temporal, Spatial, Hierarchical, or Phylogenetic Structure. *Ecography*, **40**, 913-929. https://doi.org/10.1111/ecog.02881

[17] Varma, S. and Simon, R. (2006) Bias in Error Estimation When Using Cross-Validation for Model Selection. *BMC Bioinformatics*, **7**, Article No. 91. https://doi.org/10.1186/1471-2105-7-91

[18] Kaggle (2026) CNC Mill Tool Wear. https://www.kaggle.com/datasets/shasun/tool-wear-detection-in-cnc-mill

[19] Lundberg, S.M. and Lee, S.-I. (2017) A Unified Approach to Interpreting Model Predictions. *Proceedings of the* 31*st International Conference on Neural Information Processing Systems*, Long Beach, 4-9 December 2017, 4768-4777.

[20] Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., *et al*. (2020) From Local Explanations to Global Understanding with Explainable AI for Trees. *Nature Machine Intelligence*, **2**, 56-67. https://doi.org/10.1038/s42256-019-0138-9

[21] Brodersen, K.H., Ong, C.S., Stephan, K.E. and Buhmann, J.M. (2010) The Balanced Accuracy and Its Posterior Distribution. 2010 20*th International Conference on Pattern Recognition*, Istanbul, 23-26 August 2010, 3121-3124. https://doi.org/10.1109/icpr.2010.764

[22] 李思崎, 关博. 基于 K 均值算法的设备故障分析与应对策略[J]. 应用数学进展, 2023, 12(12): 5050-5056.

[23] 李涛, 陈丽, 高松. 基于组合分类器的 MSF 发动机综合故障诊断研究[J]. 运筹与模糊学, 2019, 9(2): 170-176.

[24] 王强, 许超, 许春梅. 电力工程中电力设备状态监测与故障诊断技术的最新进展[J]. 传感器技术与应用, 2025, 13(4): 615-631.

[25] 许志鹏, 陈虎. 基于 CNN-SVM 融合算法的矿用充电机功率模块早期故障识别[J]. 建模与仿真, 2026, 15(8): 51-60.

[26] 尹红珊, 庞进, 李静, 刘安, 王景初, 刘孟奇, 许伟民. 基于机器学习集成模型的配电网预测性维护方法研究[J]. 智能电网, 2026, 16(4): 113-121.

[27] 李嘉钰. AI 技术工业生产中的应用研究[J]. 软件工程与应用, 2020, 9(4): 288-293.