

基于数据驱动的烟机主轴箱机械故障预测研究

赵致铭, 黎 阳

红塔烟草(集团)有限责任公司玉溪卷烟厂, 云南 玉溪

收稿日期: 2026年3月12日; 录用日期: 2026年5月13日; 发布日期: 2026年6月15日

摘 要

烟草加工行业中, 主轴箱作为核心传动部件, 其故障易导致生产中断、成本增加, 而传统维护模式难以适配现代化生产的精准运维需求。文章以烟机主轴箱为研究对象, 结合其运行工况与故障特点, 构建数据驱动的故障预测体系, 通过信号预处理、时频特征提取及特征降维优化数据质量, 设计传统机器学习与混合深度学习预测模型并进行对比优选, 最终通过实验验证模型有效性。研究表明, 优选的混合深度学习模型预测精度高、鲁棒性强, 可实现故障提前预警, 为烟机主轴箱这一特定关键设备提供并验证了一套完整且高效的数据驱动故障预测方案, 为烟机主轴箱精准运维提供技术支持, 助力烟草生产连续稳定运行。

关键词

烟机主轴箱, 数据驱动, 故障预测, 特征提取

Research on Data-Driven Mechanical Fault Prediction of Tobacco Machine Spindle Box

Zhiming Zhao, Yang Li

Yuxi Cigarette Factory, Hongta Tobacco (Group) Co., Ltd., Yuxi Yunnan

Received: March 12, 2026; accepted: May 13, 2026; published: June 15, 2026

Abstract

In the tobacco processing industry, the spindle box is a core transmission component. Its failure can easily lead to production interruptions and increased costs. Traditional maintenance methods are difficult to adapt to the precise operation and maintenance needs of modern production. This paper takes the tobacco machine spindle box as the research object, and constructs a data-driven fault prediction system based on its operating conditions and fault characteristics. Through signal pre-processing, time-frequency feature extraction, and feature dimensionality reduction to optimize data quality, traditional machine learning and hybrid deep learning prediction models are designed,

compared, and optimized. Finally, the effectiveness of the model is verified through experiments. The study shows that the optimized hybrid deep learning model has high prediction accuracy and strong robustness, and can realize early warning of faults. It provides and verifies a complete and efficient data-driven fault prediction scheme for the tobacco machine spindle box, a specific key piece of equipment, providing technical support for the precise operation and maintenance of the tobacco machine spindle box and helping to ensure the continuous and stable operation of tobacco production.

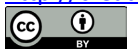
Keywords

Tobacco Machine Spindle Box, Data-Driven, Fault Prediction, Feature Extraction

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

烟机主轴箱作为烟草加工机械的核心传动部件, 承担动力传递与转速调节功能, 其运行稳定性直接决定生产效率与产品质量。长期高频运行中, 主轴箱易因轴承磨损、润滑失效、装配误差及油烟污染等引发故障, 导致停机检修, 造成经济损失且影响生产连续性[1]。当前传统维护模式依赖人工巡检与经验判断, 存在预警滞后、误判率高的弊端, 难以满足现代化生产的精准运维需求[2]。因此, 开展基于数据驱动的烟机主轴箱故障预测研究, 实现故障提前预警与精准预判, 对降低故障发生率、减少停机损失、保障烟草生产连续稳定运行具有重要现实意义与应用价值。

2. 烟机主轴箱机械结构与故障机理分析

2.1. 主轴箱机械结构及传动原理

烟机主轴箱是烟草加工设备的动力核心, 结构紧凑且集成度高, 主要由主轴、轴承、齿轮组、箱体、润滑系统及定位机构组成。主轴作为核心承载部件, 通过滚动轴承与箱体连接, 承担切削或传动载荷; 齿轮组采用多级减速设计, 实现电机动力的分级传递与转速调节, 满足烟机加工的转速需求; 润滑系统采用强制润滑方式, 减少各运动副的摩擦损耗; 定位机构保证主轴与齿轮组的同轴度, 避免运行偏差。

如图 1 所示, 其传动原理为: 电机输出动力经联轴器传递至主轴箱输入轴, 通过多级齿轮啮合实现转速与扭矩的调节, 最终由主轴将动力传递至执行部件, 完成烟草加工的各项工序, 整个传动过程需保证各部件配合精度, 确保动力传递平稳高效。

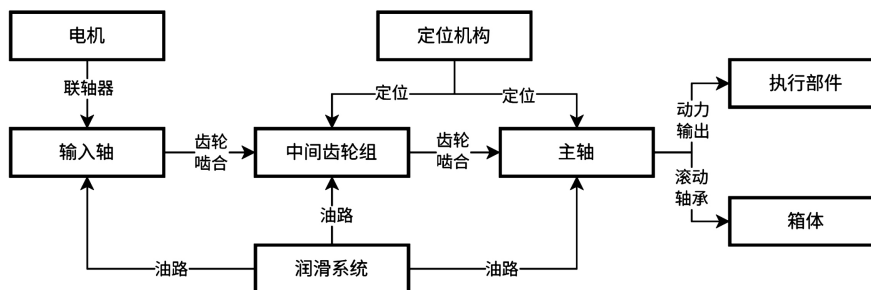


Figure 1. Schematic diagram of mechanical transmission principle of the main shaft box of the smoke machine
图 1. 烟机主轴箱机械传动原理图

2.2. 典型机械故障模式与失效机理

烟机主轴箱长期处于高频、重载及油烟污染的工况下, 典型机械故障主要包括轴承失效、齿轮损坏、主轴弯曲及润滑失效四类[3]。轴承失效是最常见故障, 因长期摩擦、润滑不足或油烟杂质侵入, 导致滚珠磨损、滚道剥落, 进而引发振动加剧; 齿轮损坏多由齿面磨损、齿根疲劳裂纹导致, 主要源于载荷波动、啮合精度不足及润滑不良, 严重时会出现断齿现象; 主轴弯曲主要因装配误差、长期重载或温度应力不均, 导致主轴轴线偏移, 影响传动精度; 润滑失效则因润滑油老化、油路堵塞, 使运动副摩擦阻力增大, 加速部件磨损, 诱发连锁故障, 各类故障均会影响主轴箱运行稳定性, 最终导致设备停机[4]。

3. 故障信号预处理与时频特征提取

3.1. 信号预处理方法

烟机主轴箱运行过程中采集的故障信号, 容易受到设备振动、油烟干扰、传感器误差以及环境噪声的影响, 存在信号失真、信噪比低等问题, 这些问题会直接影响后续特征提取与模型预测的准确性, 因此信号预处理是故障预测工作的关键前置环节[5]。

本文结合烟机主轴箱的实际运行工况, 采用去噪平稳化标准化的三级预处理流程。其中小波阈值去噪的关键参数设置如下: 选取 db4 小波作为小波基函数, 分解层数设为 5 层, 阈值选择采用改进型启发式阈值法(公式为 $\lambda = \sigma\sqrt{2\ln N}$, 其中 σ 为噪声标准差, N 为信号长度), 通过该参数设置滤除高频环境噪声与电磁干扰, 保留与故障特征相关的有效信号。其次采用差分法消除信号趋势项, 解决主轴箱长期运行中因温度升高、部件磨损导致的信号漂移问题, 使信号趋于平稳。最后通过 Z-score 标准化处理, 将信号归一化至同一数量级, 消除不同传感器、不同运行工况下的信号尺度差异, 为后续时频特征提取提供标准化、高可信度的信号数据, 保障特征提取的一致性与准确性。

3.2. 时域特征提取

时域特征提取基于故障信号的时间域波形, 提取能够反映主轴箱故障状态的统计特征, 具有计算简便、物理意义明确的优势, 适用于烟机主轴箱故障的初步识别。

结合烟机主轴箱的故障特性, 选取 12 个核心时域特征, 涵盖幅值、波形形状及能量三类指标。幅值类特征包含峰值、峰峰值、有效值, 能够直接反映信号的振动强度, 故障发生时幅值会出现显著增大。波形形状类特征包含偏度、峭度、波形因子, 可表征信号波形的不对称性与陡峭程度, 轴承磨损时峭度值会出现明显升高。能量类特征包含平均绝对值、能量值、脉冲因子等, 能够反映信号的能量分布, 齿轮裂纹等故障会导致能量分布出现异常。通过计算各时域特征的数值, 构建时域特征向量, 实现对主轴箱运行状态的初步量化描述, 为故障分类与预测提供基础特征支撑。

3.3. 频域特征提取

频域特征提取通过傅里叶变换将预处理后的时域信号转换至频率域, 挖掘信号在不同频率段的分布规律, 能够有效识别主轴箱故障的特征频率, 弥补时域特征难以捕捉高频故障信息的不足。烟机主轴箱的典型故障比如轴承失效、齿轮损坏都对应特定的特征频率, 轴承滚珠磨损的特征频率与滚珠数量、滚动体直径相关, 齿轮损坏的特征频率与齿轮齿数、转速相关。

本文采用快速傅里叶变换将时域信号转换为频域信号, 提取核心频域特征。这些特征包括基频、谐波频率、峰值频率、频率方差、频域能量占比等, 基频与谐波频率可直接对应故障类型, 峰值频率反映故障信号的主要能量集中频段, 频域能量占比可表征故障的严重程度。通过频域特征提取, 能够精准捕

捉故障的频率特征, 进一步丰富故障特征向量, 提升后续故障预测的准确性。

3.4. 时频域特征提取

烟机主轴箱的故障信号多为非平稳、非线性信号, 单一的时域或频域特征难以全面捕捉故障信息, 而时频域特征提取可同时兼顾信号的时间特性与频率特性, 精准刻画故障发展过程中的信号变化规律。

本文采用小波包变换实现时频域特征提取, 该方法能够对信号进行多尺度分解, 将信号分解至不同的时频窗口, 既保留时域信号的时间局部性, 又具备频域信号的频率分辨能力。小波包变换的关键参数设置选用 db4 小波作为基函数, 分解层数与小波去噪保持一致, 为 5 层, 具体流程为对预处理后的信号进行小波包分解, 选取合适的分解层数与小波基函数, 计算各分解节点的小波包系数。提取各节点的能量、熵值、方差等特征, 包括小波包能量熵、节点能量占比、系数方差等, 能量熵可反映信号的复杂度, 故障发生时信号复杂度升高, 能量熵会出现显著变化。将各时频域特征整合, 构建完整的时频域特征向量, 弥补单一域特征的局限性, 为故障预测模型提供更全面、更精准的特征支撑。

3.5. 特征选择与降维

通过时域、频域及时频域提取的特征向量中, 存在冗余特征与无关特征, 这些特征不仅会增加模型的计算量、延长训练时间, 还可能引入噪声干扰, 降低故障预测的精度, 因此需要进行特征选择与降维处理。

本文采用特征筛选维度压缩的两步法。首先运用互信息法筛选特征, 计算各特征与故障标签之间的互信息值, 剔除互信息值低于阈值(阈值设为 0.3)的无关特征, 保留与故障状态高度相关的核心特征, 减少特征冗余。其次采用主成分分析(PCA)进行维度压缩, PCA 降维的关键参数与结果: 设置累计贡献率阈值为 95%, 经降维处理后, 特征维度从原始的 32 维降至 8 维, 通过线性变换将高维特征向量映射至低维空间, 在保留原始特征主要信息的前提下, 降低特征维度, 简化模型结构。通过特征选择与降维, 既保留了能够有效表征主轴箱故障状态的关键特征, 又消除了冗余信息与噪声干扰, 提升模型的训练效率与预测精度, 为后续数据驱动预测模型的构建奠定基础。

4. 数据驱动的烟机主轴箱故障预测模型

4.1. 数据驱动预测框架

数据驱动的烟机主轴箱故障预测框架以故障数据为核心, 整合数据采集、预处理、特征提取、模型训练与预测推理等关键环节, 形成闭环式预测体系, 可实现主轴箱故障的精准预警与状态评估。该框架依托烟机主轴箱运行过程中的振动、温度等多源监测数据, 经过前文所述的预处理与时频特征提取流程, 得到标准化的核心特征向量(PCA 降维后 8 维), 作为模型输入数据。

框架核心分为数据层、特征层与模型层, 数据层负责多源监测数据的采集与存储, 保障数据的完整性与实时性; 特征层完成特征提取与降维, 为模型提供高质量输入; 模型层通过训练学习特征与故障状态的映射关系, 实现故障预测与状态识别。整个框架无需依赖主轴箱复杂的物理模型, 仅通过数据挖掘即可捕捉故障发展规律, 适配烟机主轴箱复杂多变的运行工况, 为后续各类预测模型的构建提供统一且规范的技术支撑。

4.2. 传统机器学习预测模型

传统机器学习预测模型凭借结构简单、训练效率高、可解释性强的优势, 在烟机主轴箱故障预测中具有较强的适用性。本文选取三类经典模型用于故障预测, 各模型的关键超参数设置如表 1 所示:

Table 1. Key hyperparameter settings for each model**表 1.** 各模型的关键超参数设置表

预测模型	关键超参数	参数取值
支持向量机(SVM)	核函数、惩罚系数 C、gamma 值	核函数为 RBF, C = 10, gamma = 0.1
随机森林	决策树数量、最大深度、最小样本分裂数	决策树数量 = 100, 最大深度 = 10, 最小样本分裂数 = 5
K 近邻算法(KNN)	邻居数 K、距离度量方式	K = 5, 距离度量为欧氏距离

支持向量机通过寻找最优分类超平面, 实现不同故障类型与健康状态的精准区分, 可有效处理高维特征数据, 适配经过降维后的特征向量, 对小样本数据具有较好的预测效果。随机森林基于多棵决策树集成学习, 通过投票机制输出预测结果, 能够有效缓解过拟合问题, 对噪声数据具有较强的鲁棒性, 可捕捉特征与故障状态之间的复杂非线性关系。K 近邻算法通过计算待预测样本与训练样本的相似度进行分类预测, 无需复杂的模型训练过程, 实时性较强, 可快速响应主轴箱的状态变化。

三类模型均经过特征向量训练与参数优化, 适配烟机主轴箱的故障特征, 为后续与深度学习模型对比奠定基础。

4.3. 深度学习预测模型

深度学习模型具备强大的特征学习与非线性拟合能力, 可自动挖掘特征向量中的深层关联信息, 弥补传统机器学习模型在复杂故障预测中的局限性, 适用于烟机主轴箱多故障、非平稳信号的预测需求。

本文选取卷积神经网络(CNN)与长短期记忆网络(LSTM)构建混合深度学习预测模型(CNN-LSTM), 其完整网络结构与超参数设置如下。

4.3.1. 网络结构图

输入层(8 维特征向量)→卷积层 1 (卷积核数量 32, 卷积核大小 3×1 , 步长 1, 激活函数 ReLU)→池化层 1 (最大池化, 池化核大小 2×1 , 步长 1)→卷积层 2 (卷积核数量 64, 卷积核大小 3×1 , 步长 1, 激活函数 ReLU)→池化层 2 (最大池化, 池化核大小 2×1 , 步长 1)→扁平化层(将二维特征映射为一维向量)→LSTM 层(隐藏层神经元数量 128, dropout = 0.3, 激活函数 tanh)→全连接层 1 (神经元数量 64, 激活函数 ReLU)→全连接层 2 (神经元数量 4, 对应 4 种故障/健康状态, 激活函数 Softmax)→输出层(输出故障类型预测结果)。

4.3.2. 关键超参数

学习率 = 0.001, 批次大小(batch size) = 32, 训练轮次(epochs) = 100, dropout 系数 = 0.3 (防止过拟合), 优化器采用 Adam, 损失函数为交叉熵损失函数, 正则化系数 = 0.0001。

卷积神经网络通过卷积层、池化层自动提取特征向量中的深层特征, 无需人工干预, 可有效捕捉故障特征的细微变化, 提升故障识别的精准度。长短期记忆网络擅长处理时序数据, 能够捕捉故障特征随时间的变化规律, 适配烟机主轴箱故障渐进式发展的特点, 可实现故障的提前预警。将两种模型结合, 充分发挥卷积神经网络的特征提取优势与长短期记忆网络的时序建模能力, 进一步提升对主轴箱复杂故障的预测性能, 满足精准运维的实际需求。

4.4. 模型对比与优选

为筛选出适配烟机主轴箱故障预测的最优模型, 从预测精度、训练效率、鲁棒性三个核心指标出发,

对传统机器学习模型与深度学习模型进行系统性对比。

选取相同的特征向量作为输入数据, 统一训练集与测试集的划分比例(7:3), 确保对比的公平性。在预测精度方面, 深度学习模型尤其是混合模型的预测准确率明显高于传统机器学习模型, 能够更好地识别细微故障与早期故障。在训练效率方面, 传统机器学习模型训练速度更快, 无需大量算力支撑, 适合快速部署与实时监测。在鲁棒性方面, 深度学习模型对噪声数据的适应能力更强, 而传统机器学习模型在数据存在异常时预测精度下降较为明显。

结合烟机主轴箱故障预测的实际需求, 优先考虑预测精度与鲁棒性, 最终优选混合深度学习模型作为最终预测模型, 同时保留传统机器学习模型作为备用, 提升预测系统的可靠性。

5. 实验验证与结果分析

为验证所构建数据驱动烟机主轴箱故障预测模型的有效性与实用性, 开展针对性实验验证, 实验以某烟草加工厂在用烟机主轴箱为研究对象, 该设备型号为 YJ17, 长期运行于高频重载工况, 常见故障包括轴承失效、齿轮损坏、主轴弯曲及润滑失效, 与前文分析的典型故障一致。

实验搭建完整的故障监测与数据采集系统, 具体采集过程如下:

1) 传感器布置: 在主轴箱轴承座、齿轮箱外壳及主轴端部各布置 1 个加速度振动传感器(型号: PCB 352C33)和 1 个温度传感器(型号: PT100), 振动传感器采样方向为垂直于主轴轴线方向, 温度传感器紧贴部件表面以确保测温准确性;

2) 运行工况设置: 实验期间设备运行工况与实际生产保持一致, 具体运行参数为: 主轴转速 1480 r/min, 负载 5.5 kW, 运行环境温度 $25^{\circ}\text{C} \pm 3^{\circ}\text{C}$, 相对湿度 $60\% \pm 5\%$, 油烟浓度控制在 $0.5 \text{ mg}/\text{m}^3$ 以内;

3) 数据采集参数: 采样频率设置为 10 kHz, 采样时长为 0.1024 s, 每组样本包含 1024 个数据点;

4) 样本采集: 采集设备健康状态、早期故障、中期故障及晚期故障四种状态下的运行数据, 每种状态采集 100 组样本, 共 400 组实验数据, 其中早期故障样本通过人工模拟轻微轴承磨损、齿轮齿面划痕获得, 中期故障为中度磨损/裂纹, 晚期故障为严重磨损/断齿、主轴明显弯曲等。

实验流程分为数据采集、预处理、特征提取、模型训练与预测验证五个环节, 严格遵循前文提出的预处理与时频特征提取方法, 确保实验数据的可靠性与实验过程的规范性。

实验采用 Python 编程语言搭建预测模型, 依托 Scikit-learn 库实现传统机器学习模型, 基于 TensorFlow 框架构建混合深度学习模型, 实验硬件配置为 Intel Core i7-12700H 处理器、32 GB 内存、NVIDIA RTX 3060 显卡, 保障模型训练与预测的效率。实验选取预测准确率、召回率、F1 分数及训练时间作为评价指标, 其中预测准确率反映模型整体预测能力, 召回率体现模型对故障样本的识别能力, F1 分数综合衡量准确率与召回率, 训练时间反映模型的部署适用性。对前文优选的混合深度学习模型与三类传统机器学习模型进行对比验证, 实验结果如表 2 所示。

Table 2. Experimental results data table

表 2. 实验结果数据表

预测模型	预测准确率	召回率	F1 分数	训练时间
支持向量机	88.5%	86.2%	87.3%	12.3 s
随机森林	91.3%	89.7%	90.5%	15.7 s
K 近邻算法	85.7%	83.1%	84.4%	8.9 s
混合深度学习模型	97.8%	97.2%	97.5%	89.5 s

由实验结果可知, 混合深度学习模型在各项评价指标上均优于传统机器学习模型, 预测准确率达到 97.8%, 召回率与 F1 分数均超过 97%, 能够精准识别主轴箱各类故障及不同故障阶段, 尤其是对早期故障的识别能力突出, 可实现故障提前预警。传统机器学习模型中, 随机森林表现最优, 预测准确率达到 91.3%, 但在早期故障识别中存在一定误判; K 近邻算法训练速度最快但预测精度最低; 混合深度学习模型训练时间较长, 主要由于其结构复杂、需挖掘特征深层关联, 但结合烟机主轴箱精准运维的实际需求, 训练时间的增加可通过硬件升级弥补, 其高预测精度与强鲁棒性更符合实际应用场景。

6. 结语

本文围绕烟机主轴箱机械故障预测展开系统性研究, 针对其故障隐蔽性强、传统维护效率低的问题, 构建了一套完整的数据驱动预测方案, 涵盖信号预处理、特征提取、模型构建与实验验证等关键环节。通过对比不同预测模型的性能, 证实混合深度学习模型在故障识别精度与鲁棒性上的优势, 有效解决了主轴箱早期故障难以识别的痛点。但研究仍存在一定局限, 后续可进一步优化特征提取方法、简化模型结构以提升部署效率, 同时结合实际生产数据拓展模型适用场景, 为烟草加工设备的智能化运维提供更完善的理论与实践参考。

参考文献

- [1] 吴俊成, 张耀清, 杨忠宝. 卷烟机主动轴位置闭环驱动复合补偿控制方法分析[J]. 中国机械, 2025(31): 94-97.
- [2] 胡学军, 罗红梅, 曾政乡, 等. ZJ118 型卷烟机无油劈刀总成的设计[J]. 机电工程技术, 2025, 54(12): 199-203+215.
- [3] 周吉文, 刘立强, 马生栋, 等. 卷烟机牌子箱和副传动机构拆装工装的设计与应用[J]. 设备管理与维修, 2024(23): 61-64.
- [4] 白朝鑫, 洪家升, 杨凯. ZJ17 卷接机上胶辊传动装置研究与设计[J]. 设备管理与维修, 2024(23): 117-119.
- [5] 曹炳强, 周靖博. PASSIM 卷烟机烟支定点搓转装置的研发[J]. 设备管理与维修, 2022(1): 79-82.