

DOI: 10.19650/j.cnki.cjsi. J2614918

基于人工智能大模型的海上风电机组故障诊断方法*

张泽辉¹, 王 昕¹, 刘 洲², 黄煜明³, 徐晓滨¹

(1. 中国-奥地利人工智能先进制造“一带一路”联合实验室 杭州 310018; 2. 江苏省国信研究院有限公司
南京 210008; 3. 金风科技股份有限公司 北京 100176)

摘 要:在“双碳”目标驱动下,海上风电已成为推动能源结构优化与绿色低碳转型的关键支撑技术。然而,海上风电机组长期工作于高盐雾、高湿度及强随机载荷等复杂海洋环境中,运行工况多变,运维难度大,亟需先进的故障诊断方法。因此,针对现有数据驱动故障诊断模型难以捕捉时序数据的深层故障特征和故障机理知识结合不紧密的问题,提出了一种基于人工智能大模型的海上风电机组故障诊断方法。首先利用聚类与贪心算法对海上风电机组数据进行预处理与特征优选,筛选出与故障演化过程密切相关的关键特征,降低数据冗余与噪声干扰;然后,以大模型作为基础架构,结合时序转换、分类识别模块,构建海上风电机组故障诊断模型。特别地,为提升故障诊断大模型识别性能与泛化能力,提出了一种融合物理约束与类别均衡的先验知识损失函数。最后,在 EDP 公开数据集和江苏某海上风电机组数据集对本文方法进行验证。实验结果表明,所提方法在故障诊断精确率、召回率以及 F1 分数指标上均优于卷积神经网络(CNN)、卷积神经网络-Transformer 组合模型(CNN-Transformer)等典型数据驱动模型,为海上风机安全运行提供了可靠的技术支撑。

关键词: 海上风电;故障诊断;工业大模型;物理信息

中图分类号: TP277 TH39 **文献标识码:** A **国家标准学科分类代码:** 460.99

A fault diagnosis method for offshore wind turbines based on artificial intelligence large model

Zhang Zehui¹, Wang Xin¹, Liu Zhou², Huang Yuming³, Xu Xiaobin¹

(1. China-Austria Belt and Road Joint Laboratory on Artificial Intelligence and Advanced Manufacturing, Hangzhou 310018, China; 2. Jiangsu Guoxin Research Institute Co., Ltd., Nanjing 210008, China; 3. Goldwind Science and Technology Co., Ltd., Beijing 100176, China)

Abstract: Driven by the dual carbon objective, the offshore wind power has become a key supporting technology for optimizing the energy structure and promoting green and low-carbon transformation. However, offshore wind turbines operate for extended periods in complex marine environments characterized by high salt spray, high humidity, and strong stochastic loads, resulting in highly variable operating conditions and significant operation and maintenance challenges that urgently require advanced fault diagnosis methods. To address the limitations of existing data-driven fault diagnosis models in capturing deep temporal fault features and in effectively integrating fault mechanism knowledge, this article proposes a large artificial intelligence model-based fault diagnosis method for offshore wind turbines. The clustering and a greedy algorithms are first employed to preprocess the turbine data and perform feature selection, thereby identifying key features closely related to fault evolution and reducing data redundancy and noise interference. Subsequently, a large model is adopted as the backbone architecture and integrated with temporal transformation and classification modules to construct the offshore wind turbine fault diagnosis framework. In particular, to enhance the recognition performance and generalization capability of the proposed model, a prior knowledge loss function incorporating physical constraints and class balancing is designed. Finally, the proposed method is evaluated on the EDP public dataset and a dataset collected from offshore wind turbines in Jiangsu Province. Experimental results show

收稿日期:2026-01-31 Received Date: 2026-01-31

* 基金项目:国家重点研发计划项目(2024YFB4207203)、国家自然科学基金项目(52401376)、浙江省尖兵领雁科技项目(2025C04005)、衢州市科技计划项目(2024K154)资助

that the proposed approach outperforms typical data-driven models, including convolutional neural networks (CNN) and convolutional neural network Transformer hybrid models (CNN-Transformer), in terms of precision, recall, and F1-Score in fault diagnosis, thereby providing reliable technical support for the safe operation of offshore wind turbines.

Keywords: offshore wind turbine; fault diagnosis; large scale industrial model; physical information

0 引言

海上风电是我国推进能源结构转型、保障能源安全的重要战略支撑^[1]。截至2023年底,我国海上风电累计装机容量超37.7 GW,占全球50%以上,居世界首位,到2030年还将增长6~8倍,保持高速发展态势^[2-5],在国家清洁能源开发中扮演愈发重要的角色。相比于陆上,海上风电机组受海浪、潮汐、盐雾、强风等影响^[6],服役环境更为恶劣,更易出现故障,并且受场地限制,海上风电运维难度也高于陆上,维修效率低。这些因素严重制约其规模化发展。如何提高海上风电机组故障监测诊断性能,已成为国内风电技术领域的研究热点之一。

海上风电机组故障诊断主要分为基于模型的故障诊断方法、基于信号分析的故障诊断方法以及基于数据驱动的故障诊断方法^[7]。基于模型的故障诊断方法通过建立部件动力学模型,将模型输出与真实设备运行数据对比,结合模型参数变化特征实现故障诊断,但海上设备工况的强非线性与多物理场耦合特性,大幅增加模型构建难度并制约诊断精度^[8];基于信号分析的诊断方法以时频域分析技术为核心,通过分析振动信号获取反映故障特征的信息来实现故障诊断^[9]。随着大数据与人工智能技术快速发展,基于数据驱动的故障诊断方法受到越来越多的关注^[10]。该方法以数据采集与监视控制系统(supervisory control and data acquisition, SCADA)为基础,通过采集设备全生命周期数据,利用人工智能技术提取关键故障特征,挖掘设备运行与故障状态的映射关系,为机组安全运行提供支撑^[11]。

尽管近年来基于数据驱动的故障诊断方法在风电机组状态监测领域取得了一定进展,但在海上风电机组复杂运行环境下仍面临显著挑战。一方面,海上风电机组运行工况具有强非线性、强随机性与多物理量耦合等特征,其SCADA数据呈现出长时间尺度依赖、跨参数隐性关联复杂的特点。现有主流数据驱动方法多依赖卷积神经网络、循环神经网络或Transformer等中小规模深度模型,其感受野和建模能力有限,难以有效捕捉长时序、多变量之间的深层故障关联特征,在复杂工况及故障早期阶段易出现诊断精度下降的问题。

另一方面,海上风电机组关键部件具有明确的物理运行机理和安全约束区间,但现有多数数据驱动诊断方法未能将设备机理知识有效融入模型训练过程,物理信

息通常仅作为特征工程或者后验分析使用,难以在模型决策过程中形成约束与引导。这种数据与机理分离的建模方式在故障样本分布不均衡、稀有故障样本不足的场景下,进一步削弱了模型诊断可靠性。

此外,受海上风电机组运行环境恶劣及运维成本高等因素影响,实际工程数据普遍存在故障样本稀缺、类别分布长尾化的问题。传统基于交叉熵或固定权重焦点损失函数(Focal Loss)的训练策略难以兼顾高频与低频故障类别,导致模型对稀有故障识别能力不足,限制了其在实际工程场景中的应用价值。

近年来,以GPT(generative pre-trained Transformer)、DeepSeek等为代表的人工智能大模型在长序列建模和隐性知识表征方面展现出显著优势,为复杂工业时序数据的深层特征挖掘提供了新的技术路径。相较于传统中小规模模型,大模型依托大规模预训练和高维参数空间,具备更强的跨时间步、跨变量关联建模能力,有望突破现有数据驱动故障诊断方法在时序特征提取与泛化能力方面的瓶颈。然而,如何将大模型有效引入海上风电机组故障诊断场景,并与设备物理机理和类别不平衡问题相结合,仍缺乏系统性研究。

针对上述问题,本文提出一种基于人工智能大模型的海上风电机组故障诊断方法,通过深度融合时序建模能力、设备物理约束与类别均衡机制,实现对复杂海上风电运行数据的高精度故障识别。本文的主要贡献如下:

1) 构建以大语言模型为核心基座的故障诊断模型,依托其在长时序序列建模中对隐性关联特征的捕捉优势,深度挖掘海上风电时序数据中的故障特征信息。

2) 提出一种融合物理约束与均衡损失的先验知识损失函数。通过动态稀有故障类别权重,结合设备物理约束,提升故障识别精度与模型整体泛化性能。

3) 为验证所提方法的有效性,采用EDP公共风电数据集(<https://www.kaggle.com/datasets/azizkasimov/wind-turbine-scada-data-for-early-fault-detection>)和江苏某海上风电数据集,相较于现有数据驱动模型,具有更高的准确率和泛化性。

1 国内外研究现状

1.1 海上风电故障诊断

1) 基于模型的故障诊断方法

该类方法主要通过数学建模或系统辨识技术建立动

态模型,描述设备的退化过程并计算零部件功能损伤程度以实现故障诊断^[12-13]。例如,Liu等^[14]针对行星齿轮齿裂故障,引入裂纹致啮合刚度与变载荷建动态模型,提出基于振动信号模型的风电机组单级齿裂诊断方法。Li等^[15]提出了一种结合迁移学习与动力学模型的行星齿轮箱故障诊断方法,实现不同域间的故障识别。Nie等^[16]建立健康齿轮箱的数学振动模型实现故障分类。

2) 基于信号分析的故障诊断方法

该类方法主要以时频域分析技术为核心,通过分析振动信号提取故障特征。例如,张健^[17]提出基于自适应完全集合经验模态分解和模糊C均值聚类的风电机组齿轮箱故障预警方法,提升振动信号的降噪效果。施耀华等^[18]通过融合电流信号和振动信号,提出变流器故障诊断方法,降低变流器开关管故障的漏检率和误检率,提高诊断的完备性和可靠性。Huang等^[19]提出时频加权谱振幅调制方法,改进谱振幅调制技术的固有缺陷,通过短时傅里叶变换与时频域差分加权实现故障特征增强与噪声抑制。

3) 基于数据驱动的故障诊断方法

该方法依托传感器采集的设备运行数据,提取设备状态特征,通过人工智能方法挖掘运行数据与故障状态间的复杂映射关系。例如,Zhang等^[20]提出改进粒子群优化算法优化深度置信网络的预警框架。Zhao等^[21]将全卷积自编码网络与注意力机制结合,提取并重建行星齿轮箱非平稳故障信号的时频特征,提高故障检测的分辨率与准确性。Sun等^[22]提出一种基于对比学习的风电机组故障诊断方法,从SCADA数据中提取时空特征以缓解数据不平衡影响。

1.2 工业大模型

近年来,以GPT、DeepSeek等系列为代表的大语言模型,依托大规模数据储备、高维度参数与强劲算力,不仅在自然语言处理、计算机视觉领域成效显著^[23],更能通过对工业运维等专业领域数据的学习,为工业运维场景提供精准的问答解析、诊断评估及决策辅助支持^[24],成为推动通用人工智能发展的标志性技术方向。工业大模型作为面向工业全生命周期的大规模深度学习模型体系,涵盖大模型基座、任务导向及行业领域等层次^[25],具体工业大模型体系框架如图1所示。

工业大模型深度融合工业数据与机理知识,凭借专业化生成、高可靠输出及跨场景自适应学习支撑核心任务^[26-28],与海上风电故障诊断对复杂数据处理、精准状态评估的需求高度契合。通过工业大模型微调技术,实现通用预训练模型与工业垂直场景深度融合,以更低的训练成本高效注入工业领域特有的设备机理、工艺规范等专业知识,适配工业场景中常见的小样本学习、时序数据处理及垂直任务需求。例如,Wu等^[29]提出参数高效微

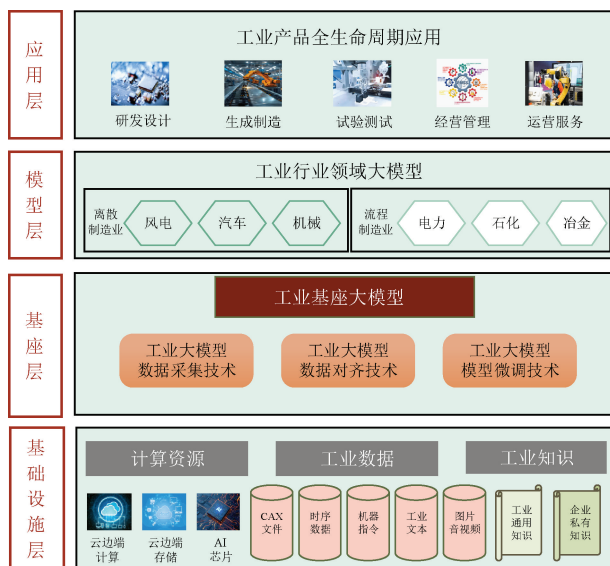


图1 工业大模型体系架构

Fig. 1 Architecture diagram of the industrial large model

调策略,采用适配器与LoRA (low-rank adaptation) 等技术手段,在模型训练过程中仅更新2%的参数,即可在17个图像分割任务中实现显著的性能提升。Zheng等^[30]将微调、编辑、融合方法融入从模型中学习,提取基础模型先验知识应用于下游任务,提升基础模型新任务性能。

2 故障诊断方法

2.1 工作流程

本文所提海上风电机组故障诊断工作流程由数据处理、特征筛选与模型训练组成。首先对SCADA系统采集的原始运行数据进行数据预处理,去除无效数据,保障后续分析;接着从处理后的正常数据中筛选关键特征,构建适配模型的最优特征子集;最后将特征数据转化为时序样本,结合所设计的融合类别均衡损失函数与物理约束的混合损失函数引导模型训练,最终实现精准故障诊断。基于人工智能大模型的海上风电故障诊断流程如图2所示。

2.2 数据预处理与特征筛选

在数据预处理环节,首先通过提取风速与有功功率特征列($\mathbf{X}_{\text{wind}}$, $\mathbf{X}_{\text{power}}$)并绘制特征关联散点图,然后采用基于网格搜索寻优的基于密度的带噪声应用空间聚类(density-based spatial clustering of applications with noise, DBSCAN)密度聚类算法,确定最优邻域距离与邻域最小样本点数阈值,按照式(1)完成聚类质量指标计算;最后依据异常数据标注结果,剔除离散异常点及有功功率小于0的无效数据。

$$Q = \frac{1}{n} \sum_{i=1}^n \max_{j \neq i} \left(\frac{\sigma_i + \sigma_j}{d(c_i, c_j)} \right) \quad (1)$$

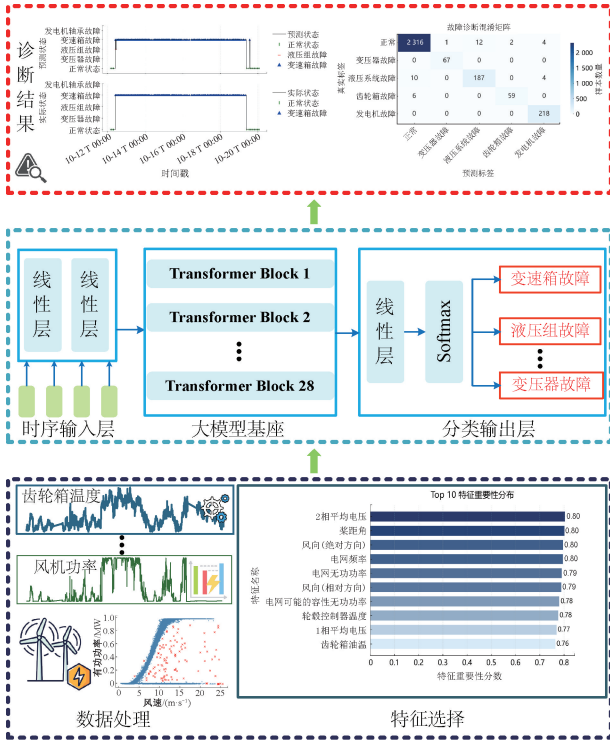


图2 基于人工智能大模型的海上风电故障诊断流程

Fig. 2 Flow chart of offshore wind power fault

diagnosis based on the artificial intelligence large model

式中: n 是簇的数量; σ_i 是簇 i 的内部距离; σ_j 是簇 j 的内部距离; $d(c_i, c_j)$ 是簇 i 和簇 j 中心之间的欧式距离; 内部

距离是指簇内所有样本到簇中心的平均距离。

根据上述清洗的正常数据提取各特征列数据 $X = \{x_1, x_2, \dots, x_n\}$, 并计算各特征列数据的均值 x_i^{mean} 与标准差 x_i^{std} 对特征列数据进行标准化处理, 然后设定初始化特征集 $S = \emptyset$, 通过贪心算法并行计算各候选特征的混合得分, 迭代选取得分前 20 的特征作为最优特征子集保存至特征集中, 混合得分计算如式(2)所示。

$$\text{Score} = \alpha \times MI + (1 - \alpha - \gamma) \times$$

$$\text{Accuracy} - (\gamma \times \text{Redundancy}) \quad (2)$$

式中: MI 为各候选特征数据归一化后的互信息量; Accuracy 为 XGBoost 模型基于当前候选特征的诊断性能指标; Redundancy 为该候选特征与已选特征的平均皮尔逊相关系数绝对值。

权重参数 α, γ 随迭代轮次动态调整, 具体如式(3)和(4)所示。

$$\alpha_{\text{adj}}^{(t)} = \alpha_0 \times (1 - (t/T)^2) \quad (3)$$

$$\gamma_{\text{adj}}^{(t)} = \gamma_0 \times (t/T)^2 \quad (4)$$

式中: α_0 是初始互信息权重系数, 设为 0.7; γ_0 是初始冗余惩罚权重系数, 设为 0.1; t 和 T 分别是当前迭代轮次和最大迭代轮次。

2.3 海上风电机组故障诊断模型构建

海上风电机组故障诊断大模型架构如图3所示, 主要分为3层, 包含时序数据输入模块、大语言模型基础架构、分类输出模块。

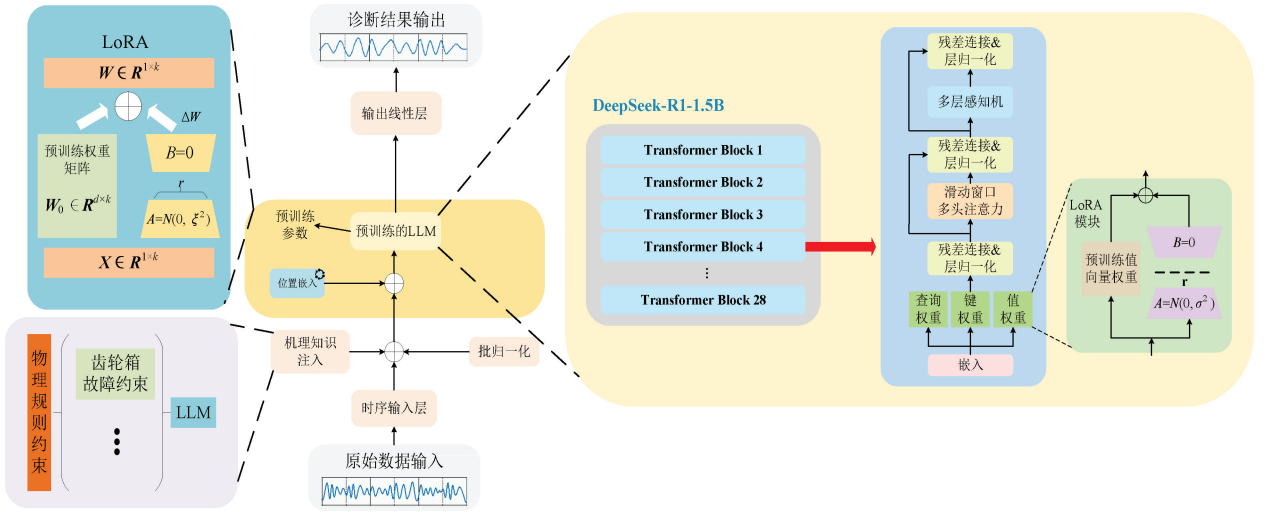


图3 海上风电机组故障诊断大模型架构

Fig. 3 Large model architecture for fault diagnosis of offshore wind turbines

1) 时序数据输入模块

时序数据输入模块将所选特征列的数据采用滑动窗口对原始数据进行切分, 通过两层全连接层、GELU 层和

层归一化映射到高维空间。

设输入最优特征为 $X \in \mathbf{R}^{B \times W \times d}$, B 代表批次大小; W 代表滑动窗口大小; d 代表特征维度。通过 2 层堆叠变

换,即:

$$\mathbf{H}^{(0)} = \mathbf{X} \quad (5)$$

$$\mathbf{H}^{(1)} = \text{LN}(\text{GELU}(\mathbf{W}_1 \mathbf{H}^{(0)} + \mathbf{b}_1)) \quad (6)$$

$$\mathbf{H}^{(2)} = \mathbf{W}_2 \mathbf{H}^{(1)} + \mathbf{b}_2 \quad (7)$$

$$\text{GELU}(x) = x \cdot \frac{1}{2} \left[1 + \frac{x}{\sqrt{\pi}} \int_0^{\frac{x}{\sqrt{\pi}}} e^{-t^2} dt \right] \quad (8)$$

2) 大模型基础架构

大模型基础架构由 DeepSeek-R1-1.5B 预训练模型作为基础架构,将多维时空特征表示注入模型的嵌入输入层;预训练模型由 28 层 Transformer 层堆叠,每层包含多头自注意力和前馈网络,即:

$$\mathbf{X}^L = \text{FFN}^L(\text{LN}(\mathbf{X}^{L-1} + \text{MHAttn}^L(\mathbf{X}^{L-1}))) \quad (9)$$

$$\text{MHAttn}(\mathbf{X}) = \text{Concat}(\text{Attn}_1, \dots, \text{Attn}_h) \mathbf{W}_o \quad (10)$$

$$\text{Attn}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Softmax}(\mathbf{Q} \mathbf{K}^T / d^k) \mathbf{V} \quad (11)$$

式中: $\mathbf{Q}, \mathbf{K}, \mathbf{V}, \mathbf{W}_o$ 分别为查询投影矩阵、键投影矩阵、值投影矩阵、输出投影矩阵。

3) 分类输出模块

分类输出模块由双层 MLP 组成,首层将输入特征压缩到原特征维度的一半并施加 GELU 激活,第 2 层映射到对应的故障类别空间,即:

$$\mathbf{y}_2 = \mathbf{W}_2 \cdot \text{LN}(\text{GELU}(\mathbf{y}_1)) + \mathbf{b}_2 \quad (12)$$

$$\mathbf{y}_1 = \mathbf{W}_1 \mathbf{h}_{\text{pool}} + \mathbf{b}_1 \quad (13)$$

$$\mathbf{h}_{\text{pool}} = \text{AdaptiveAvgPool1D}(\mathbf{H}^T) \quad (14)$$

式中: $\mathbf{H} \in \mathbf{R}^{B \times W \times D}$ 为模型输出的隐状态,其中 D 为隐藏维度。

2.4 基于先验知识损失函数的海上风电故障诊断模型训练

本文提出一种类别均衡与物理约束的先验知识损失函数,旨在通过动态平衡不同故障类别的损失贡献以强化稀有故障类别的学习权重,同时依托设备物理运行规律约束模型决策边界,提升模型对稀有故障的识别精度与整体的泛化性能。

将物理约束损失函数与类别均衡损失函数加权融合,构建总损失函数,如式(15)所示。

$$L_{\text{total}} = \lambda_1 L_{\text{efl}}(P_i) + \lambda_2 L_{\text{phy}} \quad (15)$$

式中: λ_1 与 λ_2 分别控制各损失函数的权重; $L_{\text{efl}}(P_i)$ 为类别均衡损失函数损失值; L_{phy} 为物理约束损失函数损失值。

1) 物理约束损失函数

物理约束损失函数依托风电机组关键部件的物理运行机理,引入设备运行状态先验信息,对模型输出施加显式约束,以避免预测结果偏离设备实际工作逻辑。该损失函数通过度量关键物理指标在实际运行状态下的统计均值与其安全运行阈值之间的偏差,构造损失违反项,从而引导模型决策始终处于设备可接受的物理运行区间内。本文选取齿轮箱油温与转子转速作为物理约束指标,两者在齿轮箱相关故障中具有较强的物理意义。

在损失函数构造中,首先计算物理指标的监测均值与对应阈值之间的偏差,当指标处于安全区间内时,偏差项对总损失的影响较小;当偏差逐渐逼近阈值上限或超过阈值时,通过引入非线性梯度放大机制,显著提升违反物理约束的惩罚力度,使模型在反向传播过程中更强烈地修正不合理预测。最终形成以齿轮箱油温与转子转速为核心的物理约束损失项,并与分类损失函数共同优化模型参数,从而确保模型在追求分类性能的同时,满足设备安全运行的物理约束。具体公式如式(16)所示。

$$L_{\text{phy}} = \frac{1}{B} \sum_{i=1}^B \left(\text{ReLU} \left(\frac{T_{\text{th}} - \mu_i}{\sigma_i + \varepsilon} - \bar{T}_i \right) + \text{ReLU}(n_{\text{th}} - \sigma_{s_i}) \right) \cdot P_{f,i} \quad (16)$$

式中: B 为批量大小; i 用于指代批次中的第 i 个样本; μ_i 和 σ_i 分别表示齿轮箱油温在标准化处理前的平均值和标准差;特别地, ε 是一个常数,这里设为 10^{-6} ,作为平滑项以避免在 σ_i 接近 0; $\bar{T}_i$ 指的是第 i 个样本在当前时间窗口内齿轮箱油温的平均值;而 σ_{s_i} 则表示第 i 个样本在当前时间窗口内转子转速的标准差; $P_{f,i}$ 代表模型诊断第 i 个样本属于齿轮箱故障类别的概率。

2) 类别均衡损失函数

针对海上风电机组故障诊断中故障样本稀缺、类别分布极度失衡的问题,原始 Focal Loss 通过引入固定的调制因子 $(1-P_i)^\gamma$ 与类别平衡权重 α_i ,在一定程度上缓解了类别不平衡问题,计算公式如式(17)所示。

$$L_{\text{efl}}(P_i) = -\alpha_i (1 - P_i)^\gamma \log(P_i) \quad (17)$$

式中: P_i 表示为模型预测的置信度; γ 为固定的聚焦参数,用于降低易分类样本的权重; α_i 为固定的类别平衡权重,用于平衡正负样本的重要性。

然而,原始 Focal Loss 损失函数存在聚焦强度固定与平衡能力有限的局限,其中,调制因子 γ 是全局统一的超参数,无法根据不同类别的样本分布和学习难度进行动态调整,导致其在处理多类别、多难度的复杂任务时难以实现精准聚焦,同时类别平衡权重 α_i 通常基于样本频率手动设置,缺乏对稀有类别学习过程的主动保护,容易被高频类别主导。

类别均衡损失函数通过引入类别自适应聚焦因子 γ ,动态调整每个类别的聚焦强度,缓解数据集中正负样本失衡问题;同时设计类别加权因子,提升稀有类别的贡献占比,避免其学习过程被高频类别主导。计算如式(18)所示。

$$L_{\text{efl}}(P_i) = -\sum_{j=1}^C \alpha_i \left(\frac{\gamma}{\gamma_b} \right) (1 - P_i)^\gamma \log(P_i), \quad \gamma = \gamma_b + \gamma_v^j \quad (18)$$

式中: P_i 表示为模型预测的置信度; α_i 是类别平衡权重,用来平衡正负样本重要性的参数; $(1-P_i)$ 是调制因子; γ_b 是基准聚焦因子,作为动态调整的基准值; γ_v^j 是类别

相关的可变聚焦因子。

在模型训练流程中,首先将筛选的正常数据的特征参数对应的数据通过分层抽样法按8:1:1比例进行样本划分,生成时序训练样本并将数据集划分为训练集、验证集和测试集,为后续评估奠定可靠数据基础;然后加载DeepSeek1.5B预训练模型,采用高效微调策略,对后3层参数进行更新并冻结其余层级参数,并通过物理先验损失函数引导模型开展微调训练。

在训练阶段,每轮迭代通过总损失函数计算预测误差后,仅对大模型后3层参数进行反向传播更新,同时实时监测训练集损失变化以判断模型对数据统计模式的拟合程度,并在验证集上持续计算各故障类别的精确率(precision,P)、召回率(recall,R)及F1分数(F1 Score,F1)以评估分类稳定性。待训练完成并达到收敛标准后,保存训练好的模型,最终用于独立测试集的性能评估,海上风电机组故障诊断模型训练流程如图4所示。

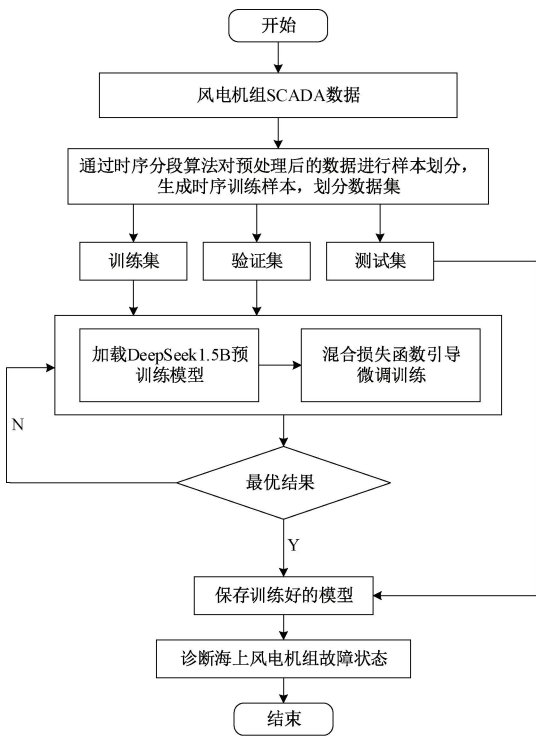


图4 海上风电故障诊断模型训练流程
Fig. 4 Flow chart of offshore wind power fault diagnosis model training

3 实验验证与分析

为验证本文所提方法的诊断性能,本文的实验数据包括两部分,一部分来自EDP开放数据平台的风电数据集,另一部分来自江苏省国信研究院公司提供的某海上风电场的风电数据。该实验在Python3.8环境下运行,深度学习框架为Pytorch。选用精确率、召回率、F1分数

评估模型诊断性能。

3.1 实验1

1) 实验1数据集

实验1数据集来自EDP开放数据平台的风电场数据,记录了葡萄牙风电场的5台风力发电机的运行数据。该数据集在多个运行过程中出现了变压器故障、液压系统故障、齿轮箱故障、发电机轴承故障。

本实验使用的原始数据参数包括时间、实际风电功率、风速、风向、环境温度、齿轮箱油温、发电机轴承温度等80个维度信息。首先根据风速与有功功率的散点图,采用DBSCAN密度聚类算法识别异常数据点,获得清洗后的风电机组数据。依据上述数据预处理方法,得到正常运行状态下SCADA数据清洗前和清洗后的对比结果如图5所示。

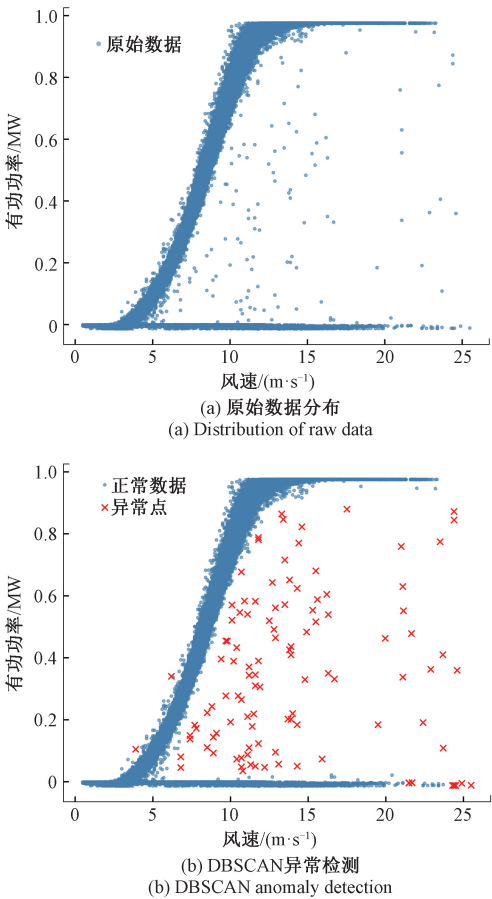


图5 DBSCAN聚类算法异常值清洗前后对比
Fig. 5 Comparison chart of DBSCAN clustering algorithm before and after outlier cleaning

特征筛选从数据层面入手,通过贪心算法并结合所设计的混合得分选择方法,结合极端梯度提升模型(extreme gradient boosting,XGBoost)计算各候选特征的混合得分,最终选取排名前20位的特征作为故障诊断模型的输入,特征评估结果如图6所示。

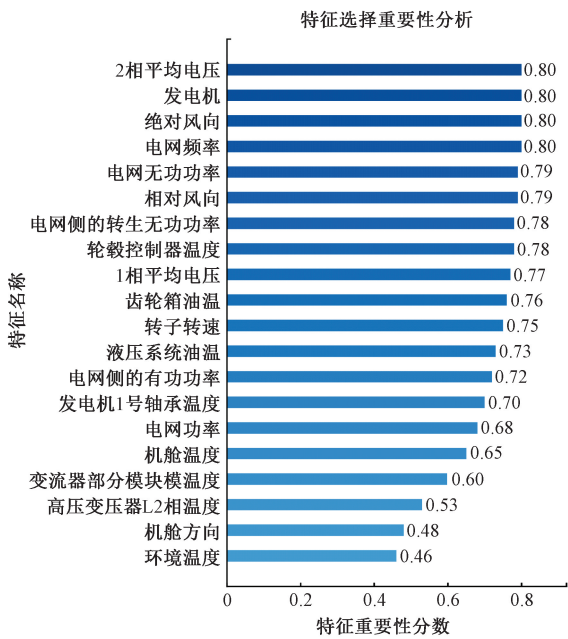


图 6 混合评估选择前 20 个特征

Fig. 6 Selecting top twenty features by hybrid evaluation

模型的训练参数包括:使用 AdamW 优化器,学习率设置为 0.000 02,权重衰减设置为 0.01,训练轮次为 10 次,批量大小为 32。这些参数的设置经过多次实验验证,使模型在诊断任务中具有良好性能。

2) 实验结果分析

为验证本文方法在风电故障诊断任务中的有效性,本研究选取大模型+Focal Loss、XGBoost、卷积神经网络(convolutional neural network, CNN)、卷积神经网络-Transformer 组合模型(convolutional neural network-transformer, CNN-Transformer)进行对比实验。风电故障诊断实验结果如表 1 和图 7 所示。

本文所提方法在精确率、召回率、F1 分数等评估指标上整体表现最优,精确率均值达 97%,明显优于大模型+Focal loss、CNN、CNN+Transformer 及 XGBoost,且平均 F1 分数在各方法中最高,表明其在整体分类任务上具有突出优势。在正常类别识别中,该方法精确率、召回率及 F1 分数均达 99%,较 CNN 的召回率 90%、XGBoost 的 F1 分数 97%体现出高稳定性与低误判率;在变压器故障识别中,3 项指标均达 98%,优于大模型+Focal loss 的 F1 分

表 1 各模型在公共风电数据集诊断结果

Table 1 Fault diagnosis results of each model in public wind power data set (%)

故障类型	本文方法			大模型+Focal loss			CNN			CNN+Transformer			XGBoost		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1	P	R	F1
正常	99	99	99	99	98	99	98	90	94	99	98	99	96	99	97
变压器故障	98	98	98	97	98	96	54	91	68	93	97	95	82	83	82
液压系统故障	94	93	96	89	98	93	72	80	76	84	89	88	85	70	77
齿轮箱故障	96	98	97	91	98	94	56	90	69	81	100	90	80	62	70
发电机轴承故障	96	99	98	97	98	98	79	97	85	96	98	97	95	91	93

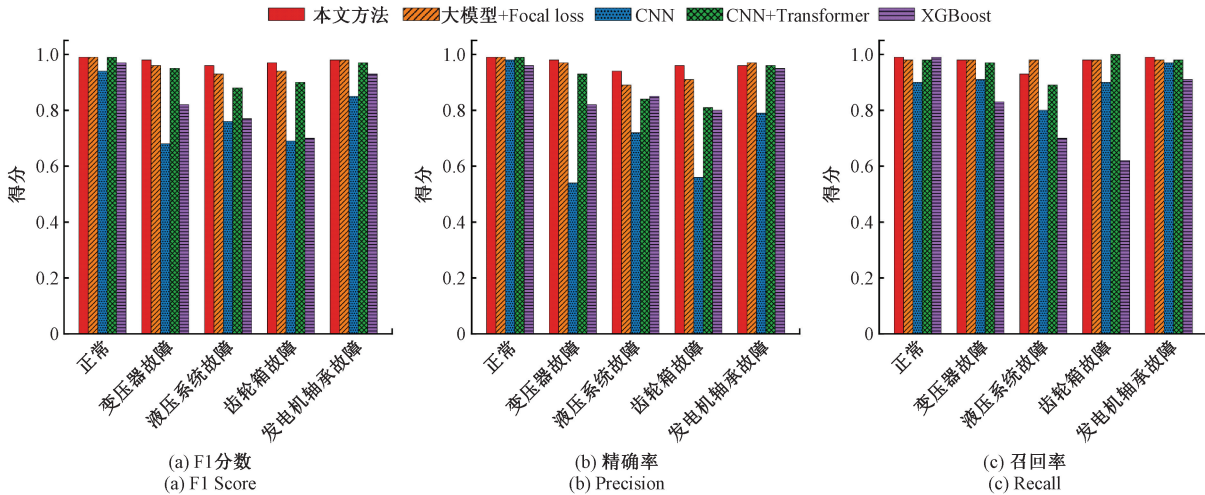


图 7 各模型在公共风电数据集诊断性能对比

Fig. 7 Comparison chart of diagnostic performance of each model on public wind power dataset

数 96%, 展现出对小样本类别的精准识别能力; 齿轮箱故障识别中, 精确率与 F1 分数均达到 96% 以上, 远超 CNN 的精确率 56%、XGBoost 的 F1 分数 70%, 高特征重叠场景下效果突出; 液压系统故障诊断中, 性能较 CNN 的 F1 分数 76%、XGBoost 的 F1 分数 77% 提升显著, 有效削弱类别间特征混淆; 发电机轴承故障识别中, F1 分数达 98%, 优于 CNN 的 F1 分数 85%、XGBoost 的 F1 分数 93%, 充分印证了在该故障识别任务中的高准确性与可靠性。

3.2 实验 2

1) 实验 2 数据集

实验 2 数据来自江苏省国信研究院公司提供的某海上风电场的风电数据, 该数据由 70 台风力发电机组成, 包含 2024 年 9 月至 2024 年 12 月的风电机组数据, 数据采样频率为 1 次/min。其中分别在多个运行过程中出现冷却系统故障、变桨系统故障、液压系统故障、变流器故障、齿轮箱故障。

依据实验 1 的数据预处理和特征筛选方法, 得到 SCADA 数据异常值清晰前后的对比结果如图 8 所示, 特征评估结果如图 9 所示。

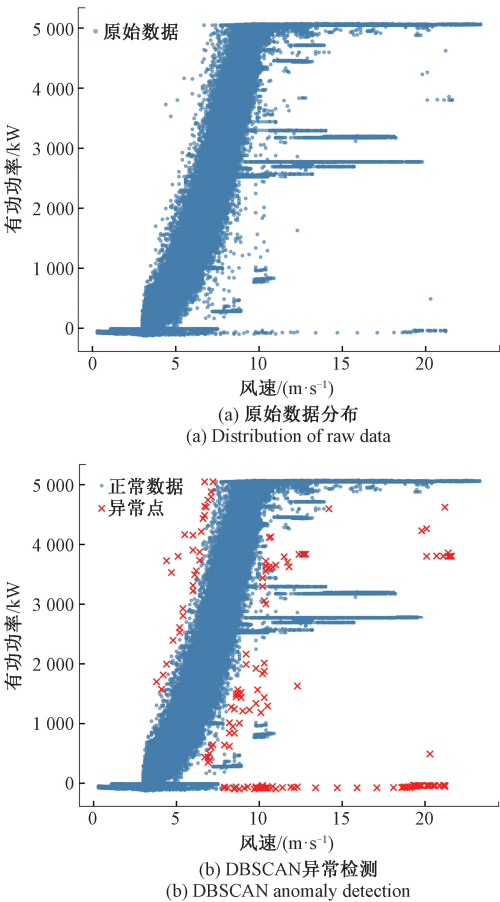


图 8 DBSCAN 聚类算法异常值清洗前后对比
Fig. 8 Comparison chart of DBSCAN clustering algorithm before and after outlier cleaning

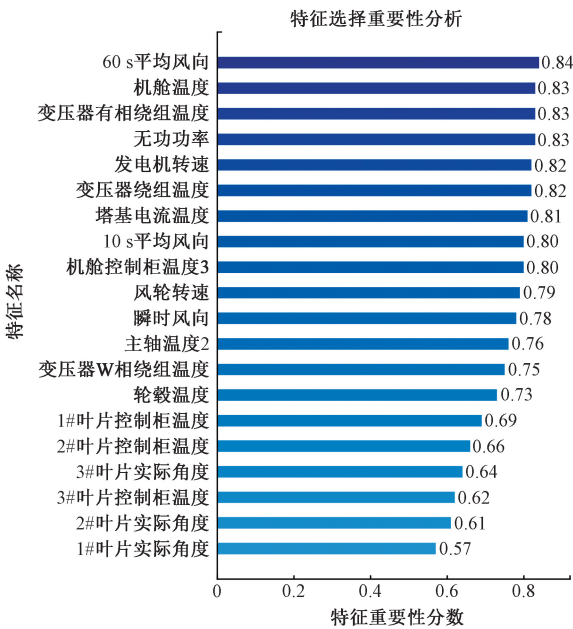


图 9 混合评估选择前 20 个特征
Fig. 9 Selecting top twenty features by hybrid evaluation

模型的训练参数包括: 使用 AdamW 优化器, 学习率设置为 0.000 02, 权重衰减设置为 0.01, 训练轮次为 15 次, 批量大小为 32。这些参数的设置经过多次实验验证, 使模型在诊断任务中具有良好的诊断性能。

2) 实验结果分析

为验证本文方法在风电故障诊断任务中的有效性, 本研究选取大模型+Focal Loss、XGBoost 模型、CNN 模型、CNN-Transformer 模型作为对比。风电故障诊断实验结果如表 2 和图 10 所示。

本文所提方法在实验任务中展现出显著综合性能优势, 精确率均值达 97%、平均 F1 分数为 96%, 处理不同类别样本分布时, 均衡性与稳定性优于大模型+Focal loss、CNN、CNN+Transformer 以及 XGBoost 等对比模型。在具体故障诊断中, 冷却系统故障 F1 分数达到 94%, 显著优于其他模型; 变桨系统故障精确率、召回率、F1 分数均达到 98% 以上, 诊断效果突出; 液压系统故障 F1 分数为 93%、精确率为 95%, 较 CNN87% 的 F1 分数、XGBoost76% 的 F1 分数, 性能提升显著, 有效削弱类别间的特征混淆; 变流器故障精确率为 99%、召回率为 95%、F1 分数为 97%, 优于大模型+Focal loss 的 93% 的精确率、CNN+Transformer 的 92% 的召回率、XGBoost 的 77% 的 F1 分数, 印证了本文方法的高准确性与可靠性; 齿轮箱故障精确率与 F1 分数均达 99%, 优于 CNN 的 93% 的精确率、XGBoost 的 89% 的 F1 分数, 在故障数量较少的场景下诊断效果突出, 同时也验证改进后损失函数对类别不平衡问题的优化效果。

表 2 各模型在某海上风电场风电数据集诊断结果

Table 2 Diagnostic results of each model in the wind power data set of an offshore wind farm (%)

故障类型	本文方法			大模型+Focal loss			CNN			CNN+Transformer			XGBoost		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1	P	R	F1
正常	96	99	98	98	98	98	94	96	95	97	76	85	98	99	98
冷却系统故障	98	90	94	95	93	94	87	79	83	55	86	68	98	38	55
变桨系统故障	99	98	98	97	99	98	95	90	93	78	97	87	97	79	88
液压系统故障	95	91	93	89	92	91	81	94	87	61	93	75	97	62	76
变流器故障	99	95	97	93	95	94	79	93	86	45	92	62	95	65	77
齿轮箱故障	99	98	99	98	98	98	93	90	91	74	91	85	97	82	89

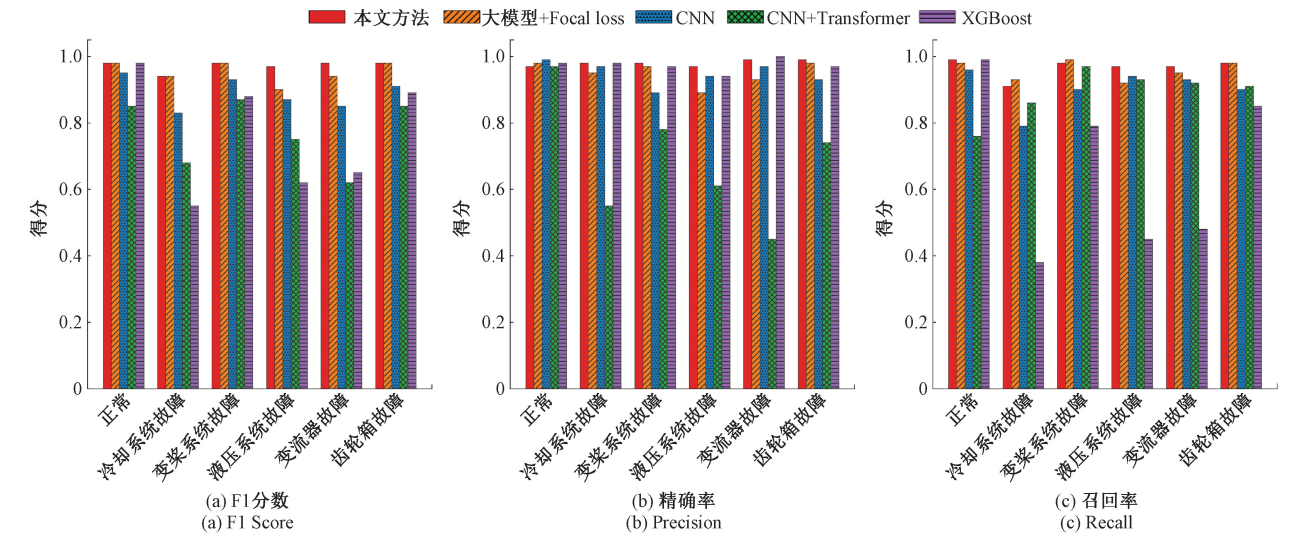


图 10 各模型在某海上风电场风电数据集故障诊断性能对比图

Fig. 10 Comparison chart of fault diagnosis performance of each model in wind power dataset of an offshore wind farm

4 结 论

针对海上风电机组故障诊断中存在的故障识别精度低和模型泛化性不足的问题,提出一种基于人工智能大模型的海上风电机组故障诊断方法。该方法以大模型为基座,结合时序转换、分类识别模块,深度挖掘时序故障特征,有效解决模型对时序故障信息捕捉不充分的问题,为后续精准故障诊断提供高质量特征支撑;同时提出一种融合类别均衡与物理约束的先验知识损失函数,解决数据集中类别失衡导致故障识别精度低的问题。经实验验证,该方法在海上风电机组故障诊断具备更优的诊断精度与泛化能力。

然而,该方法仍依赖一定数据量,对更多机组的适配性有待提升。未来将设计适配大模型的学习策略,融入故障机理先验以降低数据依赖,优化模型参数微调方式,

进一步提升诊断精度与泛化能力,更好适配实际工程需求。

参考文献

[1] 郑崇伟, 李崇银. 海上风能等级区划研究: 瓶颈与对策[J]. 中国科学院院刊, 2023, 38(4): 654-665.
ZHENG CH W, LI CH Y. Research on offshore wind energy classification and zoning: Bottlenecks and countermeasures[J]. Bulletin of the Chinese Academy of Sciences, 2023, 38(4): 654-665.

[2] MEHLAN F C, NEJAD A R. Rotor imbalance detection and diagnosis in floating wind turbines by means of drivetrain condition monitoring[J]. Renewable Energy, 2023, 212: 70-81.

[3] 金晓航, 孙毅, 单继宏, 等. 风力发电机组故障诊断与预测技术研究综述[J]. 仪器仪表学报, 2017,

- 38(5):1041-1053.
- JIN X H, SUN Y, SHAN J H, et al. Review of fault diagnosis and prognosis technologies for wind turbines[J]. Chinese Journal of Scientific Instrument, 2017, 38(5): 1041-1053.
- [4] 金晓航, 杨宇辰, 喻轩昂. 改进多层自组织映射网络驱动的风电机组状态监测[J]. 仪器仪表学报, 2025, 46(3):231-241.
- JIN X H, YANG Y CH, YU X ANG. Wind turbine condition monitoring driven by an improved multilayer self-organizing map network [J]. Chinese Journal of Scientific Instrument, 2025, 46(3): 231-241.
- [5] 黄玲玲, 符杨, 任浩瀚, 等. 基于状态信息的风电机组维护研究综述[J]. 中国电机工程学报, 2020, 40(21):7065-7078.
- HUANG L L, FU Y, REN H H, et al. Research review on wind turbine maintenance based on condition information[J]. Proceedings of the CSEE, 2020, 40(21): 7065-7078.
- [6] XIE T M, XU Q F, JIANG C X, et al. The fault frequency priors fusion deep learning framework with application to fault diagnosis of offshore wind turbines[J]. Renewable Energy, 2023, 202: 143-153.
- [7] LIU D D, CUI L L, CHENG W D. Fault diagnosis of wind turbines under nonstationary conditions based on a novel tacho-less generalized demodulation[J]. Renewable Energy, 2023, 206: 645-657.
- [8] 申艳杰, 李鹏飞, 武英杰, 等. 基于双馈式风电机组定子电流信号现象学模型的齿轮箱故障监测方法[J]. 仪器仪表学报, 2025, 46(9):348-359.
- SHEN Y J, LI P F, WU Y J, et al. Gearbox fault monitoring method based on a phenomenological model of stator current signals in doubly fed wind turbines[J]. Chinese Journal of Scientific Instrument, 2025, 46(9): 348-359.
- [9] 曹利宵, 贾晓冰, 李宁, 等. 基于自回归嵌入的风电机组传动系统双阈值状态监测[J]. 仪器仪表学报, 2025, 46(9):360-371.
- CAO L X, JIA X B, LI N, et al. Dual-threshold condition monitoring of wind turbine drivetrain based on autoregressive embedding [J]. Chinese Journal of Scientific Instrument, 2025, 46(9): 360-371.
- [10] ZHOU F N, YANG SH, FUJITA H, et al. Deep learning fault diagnosis method based on global optimization GAN for unbalanced data [J]. Knowledge-Based Systems, 2020, 187: 104837.
- [11] PLIATSIOS D, SARIGIANNIDIS P, LAGKAS T, et al. A survey on SCADA systems: Secure protocols, incidents, threats and tactics[J]. IEEE Communications Surveys & Tutorials, 2020, 22(3): 1942-1976.
- [12] LIANG J P, ZHANG K, AL-DURRA A, et al. A novel fault diagnostic method in power converters for wind power generation system [J]. Applied Energy, 2020, 266: 114851.
- [13] GAO ZH W, ODGAARD P. Real-time monitoring, fault prediction and health management for offshore wind turbine systems [J]. Renewable Energy, 2023, 218: 119258.
- [14] LIU X Z, YANG Y H, ZHANG J. Resultant vibration signal model based fault diagnosis of a single-stage planetary gear train with an incipient tooth crack on the sun gear[J]. Renewable Energy, 2018, 122: 65-79.
- [15] LI D D, ZHAO Y, ZHAO Y. A dynamic-model-based fault diagnosis method for a wind turbine planetary gearbox using a deep learning network [J]. Protection and Control of Modern Power Systems, 2022, 7(2): 1-14.
- [16] NIE Y Y, LI F Y, WANG L M, et al. A mathematical model of vibration signal for multistage wind turbine gearboxes with transmission path effect analysis [J]. Mechanism and Machine Theory, 2022, 167: 104428.
- [17] 张健. 基于 ACSEEMD 和模糊 C 均值聚类的风电机组齿轮箱故障预警[J]. 电子器件, 2019, 42(6):1410-1415.
- ZHANG J. Fault early warning of wind turbine gearbox based on ACSEEMD and fuzzy C-means clustering[J]. Chinese Journal of Electron Devices, 2019, 42(6): 1410-1415.
- [18] 施耀华, 冯延晖, 任铭, 等. 融合电流和振动信号的永磁同步风电系统变流器故障诊断方法研究[J]. 中国电机工程学报, 2020, 40(23):7750-7760.
- SHI Y H, FENG Y H, REN M, et al. Research on fault diagnosis method for converter of permanent magnet synchronous wind power system fusing current and vibration signals [J]. Proceedings of the CSEE, 2020, 40(23): 7750-7760.

- [19] HUANG B Y, ZHAO X M, CHEN H L, et al. Time-frequency weighted spectral amplitude modulation and its applications in fault diagnosis of rolling bearings [J]. *Mechanical Systems and Signal Processing*, 2025, 237: 113135.
- [20] ZHANG ZH Y, WANG SH K, WANG P G, et al. Research on fault early warning of wind turbine based on IPSO-DBN[J]. *Energies*, 2022, 15(23): 9072.
- [21] ZHAO D Z, SHAO D P, CUI L L. CTNet: A data-driven time-frequency technique for wind turbine fault diagnosis under time-varying speeds[J]. *ISA Transactions*, 2024, 154: 335-351.
- [22] SUN SH L, HU W Y, LIU Y K, et al. Matching contrastive learning: An effective and intelligent method for wind turbine fault diagnosis with imbalanced SCADA data[J]. *Expert Systems with Applications*, 2023, 223: 119891.
- [23] 任磊, 王海腾, 董家宝, 等. 工业大模型: 体系架构、关键技术与典型应用[J]. *中国科学: 信息科学*, 2024, 54(11): 2606-2622.
- REN L, WANG H T, DONG J B, et al. Industrial foundation model: Architecture, key technologies, and typical applications[J]. *Scientia Sinica (Informationis)*, 2024, 54(11): 2606-2622.
- [24] 彭宇, 季拓, 郭楚亮. 故障预测与健康管理领域大语言模型: 应用与展望[J]. *仪器仪表学报*, 2025, 46(10): 2-21.
- PENG Y, JI T, GUO CH L. Large language models in fault prognosis and health management: Applications and prospects[J]. *Chinese Journal of Scientific Instrument*, 2025, 46(10): 2-21.
- [25] ZHOU B, LI X Y, LIU T Y, et al. CausalKGPT: Industrial structure causal knowledge-enhanced large language model for cause analysis of quality problems in aerospace product manufacturing [J]. *Advanced Engineering Informatics*, 2024, 59: 102333.
- [26] LU J W, WANG X X, CHENG X, et al. Parallel factories for smart industrial operations: From big AI models to field foundational models and scenarios engineering [J]. *IEEE/CAA Journal of Automatica Sinica*, 2022, 9(12): 2079-2086.
- [27] REN L, WANG H T, LI J W, et al. AIGC for industrial time series: From deep generative models to large

generative models [J]. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2025, 55(11): 7774-7791.

- [28] 雷亚国, 李熹伟, 李响, 等. 面向机械设备通用健康管理智能运维大模型[J]. *机械工程学报*, 2025, 61(6): 1-13.

LEI Y G, LI X W, LI X, et al. Intelligent operation and maintenance foundation model for universal health management of mechanical equipment [J]. *Journal of Mechanical Engineering*, 2025, 61(6): 1-13.

- [29] WU J D, WANG Z Y, HONG M X, et al. Medical SAM adapter: Adapting segment anything model for medical image segmentation[J]. *Medical Image Analysis*, 2025, 102: 103547.

- [30] ZHENG H L, SHEN L, TANG AN K, et al. Learning from models beyond fine-tuning [J]. *Nature Machine Intelligence*, 2025, 7(1): 6-17.

作者简介



张泽辉, 2012 年于上海海事大学获得学士学位, 2019 年于武汉理工大学获得硕士学位, 2022 年于南开大学获得博士学位, 现为杭州电子科技大学特聘副研究员, 主要研究方向为海上风电机组故障诊断, 新能源设备健康运维及控制。

E-mail: zhangzehui@hdu.edu.cn

Zhang Zehui received his B.Sc. degree from Shanghai Maritime University in 2012, his M.Sc. degree from Wuhan University of Technology in 2019, and his Ph.D. degree from Nankai University in 2022. He is currently a distinguished associate researcher at Hangzhou Dianzi University. His main research interests include fault diagnosis of offshore wind turbines, health operation and maintenance as well as control of new energy equipment.



刘洲, 2017 年于南京航空航天大学获得硕士学位。现为江苏省国信研究院的研究员, 主要研究方向为图像识别和故障诊断在能源领域的应用。

E-mail: liuzhou@jsgx.net

Liu Zhou received his M.Sc. degree from Nanjing University of Aeronautics and Astronautics in 2017. He works as a researcher at Jiangsu Guoxin Research Institute. His main research interest includes applications of image recognition and fault diagnosis in the energy sector.



黄煜明,2017 年于南京邮电大学获得学士学位,现为金风科技海上风电大数据及智能运维技术高级工程师,主要研究方向为海上风电机组故障诊断,新能源设备健康运维及控制。

F-mail:huangyuming@goldwind.com

Huang Yuming received his B.Sc. degree from Nanjing University of Posts and Telecommunications in 2017. He is currently a senior engineer specializing in offshore wind power big data and intelligent O&M technology at Goldwind Science & Technology. His main research interests include fault diagnosis of offshore wind turbines, health operation and maintenance as well as control of new energy equipment.



徐晓滨(通信作者),2003 年于河南大学获得学士学位,2006 年于河南大学获得硕士学位,2009 年于上海海事大学获得博士学位,现为杭州电子科技大学自动化学院教授,主要研究方向为海上风电机组故障诊断,新能源设备健康运维及控制。

E-mail:xuxiaobin1980@hdu.edu.cn

Xu Xiaobin (Corresponding author) received his B.Sc. degree from Henan University in 2003, his M.Sc. degree from Henan University in 2006, and his Ph.D. degree from Shanghai Maritime University in 2009. He is currently a professor and Ph.D. supervisor with the School of Automation at Hangzhou Dianzi University. His main research interests include fault diagnosis of offshore wind turbines, health operation and maintenance as well as control of new energy equipment.