

引用格式:李栋,廖毅桢,桑源,等.基于深度学习的海上风机多源缺失数据的填补方法[J].工业建筑,2026,56(5):176-186. LI Dong, LIAO Yizhen, SANG Yuan, et al. Deep Learning-Based Imputation of Multi-Source Missing Data for Offshore Wind Turbines[J]. Industrial Construction, 2026, 56(5):176-186 (in Chinese). DOI:10.3724/j. gyjzG26031308

基于深度学习的海上风机多源缺失数据的填补方法^{*}

李 栋¹ 廖毅桢¹ 桑 源¹ 李泽宇² 姚 博³ 陈洪兵²

(1. 福州大学土木工程学院,福州 350108;2. 北京科技大学未来城市学院/城镇化与城市安全研究院,北京 100083;
3. 中国五冶集团有限公司,成都 476300)

摘 要 针对海上风电机组在恶劣工况下因传感器失效或通信中断引发的多源监测数据缺失问题,构建了一种基于多头门控残差网络的填补模型。该方法通过特征级联将数据采集与监视控制系统(SCADA)数据和结构振动监测数据进行协同融合,并引入门控残差网络提取深度非线性耦合特征;同时设计出多头并行输出结构,实现对两类异构数据的独立重建。训练阶段采用动态掩码机制与混合损失函数,以增强模型对复杂气动工况的适应能力。基于某10 MW海上风电机组实测数据的验证结果显示,该模型在训练工况下具有较高的决定系数,能够实现多源缺失数据的精准重构。在面向非训练时段的泛化测试中,模型预测的决定系数虽略有波动,但仍能有效捕捉信号整体变化趋势;尤其对振动类数据,其泛化性能衰减幅度小于SCADA数据,展现出更强的稳定性。该方法可显著提升风机多源监测数据的完整性与可靠性,具备良好的工程应用前景。

关键词 海上风电机组;数据填补;门控残差网络;多源数据融合

Deep Learning-Based Imputation of Multi-Source Missing Data for Offshore Wind Turbines

LI Dong¹ LIAO Yizhen¹ SANG Yuan¹ LI Zeyu² YAO Bo³ CHEN Hongbing²

(1. College of Civil Engineering, Fuzhou University, Fuzhou 350108, China; 2. Future City College / Institute of Urbanization and Urban Safety, University of Science and Technology Beijing, Beijing 100083, China; 3. China MCC5 Group Co., Ltd., Chengdu 476300, China)

Abstract: To address the problem of missing multi-source monitoring data of offshore wind turbines caused by sensor failures or communication interruptions under harsh operating conditions, this paper proposes a novel imputation model based on a multi-head gated residual network. This method achieves collaborative fusion of supervisory control and data acquisition (SCADA) data and structural vibration monitoring data through feature concatenation, employs a gated residual network to extract deep nonlinear coupling features, and uses a multi-head parallel output architecture for the independent reconstruction of these two heterogeneous data types. During the training stage, a dynamic masking mechanism combined with a hybrid loss function is adopted to enhance the model's adaptability to complex aerodynamic operating conditions. Validated with field data from a 10 MW offshore wind turbine, the proposed model achieved a high coefficient of determination under training conditions, enabling accurate reconstruction of missing multi-source data. In generalization tests for non-training periods, although the coefficient of determination of the model's predictions fluctuated slightly, the model still effectively captured

^{*}福建省自然科学基金面上项目(2024J01266)。

第一作者:李栋,副研究员,主要从事大跨度空间薄膜结构、组合结构、海上风电、结构风工程等方面研究,dongli@fzu.edu.cn。

通信作者:陈洪兵,教授,主要从事工程结构损伤检测、服役安全智能诊断,钢-混凝土组合结构缺陷无损检测技术和多尺度多物理场仿真等领域研究,hongbingchen@ustb.edu.cn。

收稿日期:2026-03-13

the overall trends of monitoring signals. Notably, the degradation in generalization performance for vibration data was less pronounced than that for SCADA data, demonstrating its greater stability. The proposed method can significantly improve the completeness and reliability of multi-source monitoring data for wind turbines and holds considerable potential for engineering applications.

Keywords: offshore wind turbines; data imputation; gated residual network; multi-source data fusion

0 引言

风能已成为全球能源结构转型的关键支撑,其清洁、可再生特性驱动风电装机规模持续扩张^[1-2]。然而,海上风机长期处于高湿、盐雾及波浪耦合荷载等严酷环境,主轴承、齿轮箱与发电机等关键传动部件故障风险显著升高^[3-5]。统计表明,发电机与齿轮箱故障合计占风电机组非计划停运原因的比例超过 2/3^[6]。加之海上风电场多位于远海区域,受限于可达性差、作业窗口期短等因素,常规巡检与人工诊断手段难以及时响应,导致运维成本急剧攀升^[7-8]。鉴于此,充分挖掘既有监控与数据采集(SCADA)系统监测数据,构建高效的状态评估与早期预警机制,已成为提升海上风电机组运行可靠性与全寿命经济性的核心路径^[6,9]。

目前,风力发电机故障预警研究多基于 SCADA 系统数据,或依赖振动与应变测量数据^[10]。在状态监测领域,基于图结构的学习与预测方法展现出重要潜力。Zhao 等^[11]结合图卷积网络与多头注意力的条件扩散模型,能够捕捉风速数据的时空相关性,并通过概率化填补区间提升数据可靠性。Qu 等^[12]将数据排列重构为旅行商问题,优化了生成对抗网络框架,显著提高了大规模风机数据集的填补精度。Liao 等^[13]采用深度卷积网络的上下文编码器,有效建模非线性特征与多属性关联,在波动性风电数据中表现优异。Fan 等^[14]设计多尺度注意力掩码自编码器,通过局部化 Transformer 层提取时序与机组间依赖关系,在降低计算成本的同时实现高精度填补。Radev 等^[15]将双向长短期记忆网络(LSTM)的风参数预测与分数阶神经网络相结合,在多类运行场景下提升了功率预测准确性。Wang 等^[16]通过结合生成对抗填补网络与 XGBoost 算法,有效解决了 SCADA 数据缺失问题,并构建了高精度的失控预警模型;其创新性引入迁移学习技术,使模型能快速适配新机型,显著降低了计算与部署成本。

此外,振动响应是评估海上风机结构动力特

性、保障其安全稳定运行的关键参数,准确预测振动响应对风电机组长期服役性能至关重要^[17-18]。目前,海上风机振动响应预测主要依赖数值模拟与有限元分析。尽管此类方法能较精确地模拟结构在复杂荷载下的动力行为,但其计算资源消耗大、耗时长特点,限制了在风机全寿命周期分析与实时监测中的应用。因此,基于机器学习与深度学习的模型逐步应用于海上风电机组振动响应预测。Bai 等^[19]采用多输入递归 BiLSTM 网络,开发了一种混合深度学习和模拟框架,实现比传统模拟快 745 倍的计算速度,同时保持 91.94% 的预测准确率,为耗时数值方法提供了高效替代方案。

在 SCADA 数据驱动的状态监测与故障诊断方面,相关研究亦取得重要进展。赵洪山等^[20]基于堆叠自编码网络构建风电机组发电机状态监测模型,通过重构误差变化趋势实现异常预警与故障诊断。张舒翔等^[21]改进 Inception v1 结构,提出 1D_Inception v1 模型,有效提取 SCADA 数据空间多尺度特征,实现高精度状态监测。王凤等^[22]融合变分模态分解、卷积神经网络、多尺度图注意力网络及灰狼优化算法,构建面向大型风机集群的故障关联状态预测与识别方法,显著提升多状态故障预警的自动化水平与识别精度。上述方法为利用 SCADA 数据开展精细化状态监测提供了可行技术路径。

然而,当前研究仍存在一定局限性。当海上风机传感器损坏或数据传输中断时,由于远海维修困难,难以及时修正缺失数据,可能对风机安全运行构成威胁。尽管已有多种数据填补方法被提出,如基于生成对抗网络、扩散模型、Transformer 结构等,但现有方法尚未实现 SCADA 系统数据与振动监测数据的深度融合与相互补充。由于振动传感器成本远高于 SCADA 传感器,多数风机振动传感器布设数量有限,若能利用已有 SCADA 数据对缺失振动数据进行有效填补,不仅可实现利用 SCADA 数据等效替代振动数据进行动力响应分析,还能显著降低运营成本,提升风机系统整体可靠性与智能化水平。

因此,发展面向多源数据融合的缺失数据填补与状态预测方法,是未来研究的重要方向。

针对风电机组监测数据中因传感器失效或通信链路中断导致的缺失问题,提出一种基于多头门控残差网络(Multi-Head Gated Residual Network)的通用填补框架。该框架对 SCADA 及振动监测数据进行统一特征构造与归一化处理,依托 6 组 GRN 模块串联构建深层特征提取主干,并设置多头输出分支并行完成振动与 SCADA 特征的联合重构。训练环节引入自适应掩码、噪声扰动、混合损失函数及 L2 约束等机制,以增强模型在非理想工况下的稳健性与泛化迁移能力。进一步采用 K 折交叉验证对集成模型进行训练,结合 $N-1$ 诊断策略与多场景鲁棒性测试,系统评估模型在多种缺失模式下的数据还原效能。本方法面向风力发电机组状态监测系统中常见的采集缺失场景,可为设备健康评估与故障早期预警提供高质量、高连续性的数据基础。

1 模型理论基础

为应对海上风电机组多源传感器数据缺失问题,本研究构建了一个基于深度学习的通用数据填补框架。该框架旨在通过捕获 SCADA 系统与振动监测系统数据间的复杂非线性关系,实现对缺失值的高精度重建。本章系统阐述模型的理论基础,包括核心模块门控残差网络(Gated Residual Network, GRN)的结构原理、多头输出机制的设计思路、动态训练策略与损失函数优化方法,以及模型性能的评估指标体系。

1.1 模型整体架构

构建的缺失数据填充模型基于多头门控残差网络架构,包含公共特征提取主体及针对振动与 SCADA 数据的独立预测分支,其整体结构如图 1 所示。网络主体部分由 8 个门控残差块(GRN Block)串联而成,每个模块内部通过双路映射实现特征筛选与残差学习。其主路径依次包含全连接层、LeakyReLU 激活函数(负斜率为 0.01)及丢失层(Dropout),而门控路径利用 Sigmoid 函数调节信息流向,最终经残差加和与层归一化处理输出。主体网络隐藏层维度统一设置为 768,双分支预测头各包含一个 256 维的全连接子层。输入层接收由原始特征与掩码向量拼接而成的 $2 \times N$ 维向量。为了提升模型的泛化能力与收敛稳定性,设置丢失率为 0.15,

最大迭代轮次为 1500,并采用初始学习率为 0.001 的余弦退火衰减策略,最小学习率设为 1.0×10^{-6} 。模型训练的批次大小为 1024,权重衰减系数为 1.0×10^{-4} ,Adam 优化器参数 $\beta_1=0.9$ 、 $\beta_2=0.999$ 。网络层的具体参数设置见表 1。

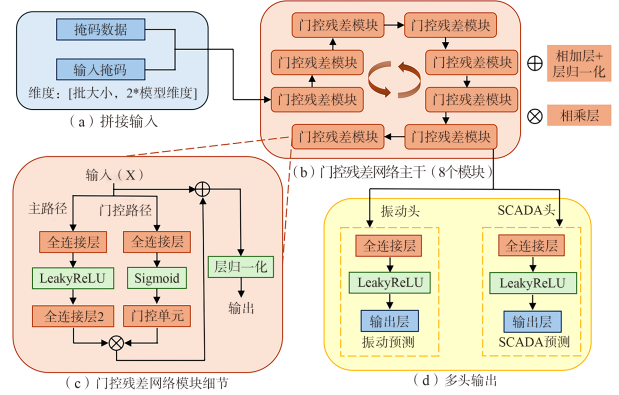


图 1 模型框架

Fig. 1 Model architecture

表 1 缺失数据填充网络层设置

Table 1 Network layer settings for missing data imputation

网络层	参数	网络层	参数
输入层	$2 \times N$ (特征+掩码)	丢弃层	0.15
全连接层	768	层归一化	Standard
激活函数	0.01	初始学习率	1.0×10^{-3}
门控激活	Sigmoid	最小学习率	1.0×10^{-6}
GRN 模块数	8	最大迭代轮次	1500
输出头全连接层	256	批次大小	1024
Vib 输出层	50	权重衰减	1.0×10^{-4}
SCADA 输出层	41	Adam 优化器	$\beta_1=0.9, \beta_2=0.999$

1.2 门控残差网络模块(GRN)

门控残差网络是所提出模型的基本构成单元,旨在从高维、不完整的数据中有效提取特征,同时缓解深层网络训练中常见的梯度消失问题。GRN 模块巧妙地结合了残差学习与门控机制,允许信息选择性地通过网络层,从而增强模型对复杂模式的表达能力。其内部结构如图 1(b)所示。

对于输入特征向量 x ,GRN 模块包含两条并行的计算路径:主路径负责对输入进行非线性变换以提取深层特征;门控路径则负责生成一个动态的调节信号,用于控制主路径信息的流出量。主路径的计算过程可表示为:

$$y = \text{Dropout}[\mathbf{W}_2 \delta(\mathbf{W}_1 x + \mathbf{b}_1) + \mathbf{b}_2] \quad (1)$$

式中: y 为主路径的输出向量; δ 为 LeakyReLU 激活

函数,用于引入非线性; W_1 和 W_2 分别为第一个和第二个全连接层的权重矩阵; b_1 和 b_2 为对应的偏置向量;Dropout层用于在训练过程中随机抑制一部分神经元,以防止模型过拟合。

门控路径通过一个独立的全连接层,并应用Sigmoid激活函数生成一个门控向量 g :

$$g = \sigma(W_3 x + b_3) \quad (2)$$

式中: g 为门控向量,其每个元素值介于0~1,起到信息流量“阀门”的作用; σ 为Sigmoid函数; W_3 与 b_3 分别为该全连接层的权重矩阵和偏置向量。

最终,主路径的输出 y 与门控向量 g 进行逐元素相乘,实现信息的选择性通过。相乘后的结果与原始输入 x 通过残差连接相加,以促进梯度流动和特征复用,最后再经过层归一化(Layer Normalization)处理,得到GRN模块的最终输出 z :

$$z = \text{LayerNorm}(x + g \odot y) \quad (3)$$

式中: $\odot$ 为Hadamard积,即逐元素相乘;LayerNorm为层归一化操作,用于稳定训练过程。该结构使模型能够根据输入 x 自适应地调节信息传递强度,从而更灵活、更有效地学习特征间的非线性依赖关系。

1.3 多头并行输出结构

考虑到振动监测数据与SCADA数据在物理含义、量纲及分布特性上存在显著差异,若采用单一的输出层同时预测两类特征,可能导致模型难以兼顾各自的精度需求。为此,本文在GRN主干网络之后设计了多头并行输出结构,分别针对振动特征与SCADA特征设置独立的预测头,如图1(d)所示。

主干网络的输出(即经过多层GRN模块提取的高阶特征)被同时馈入两个预测分支:振动预测头(Vibration Head)与SCADA预测头(SCADA Head)。每个预测头均由一个全连接层、一个LeakyReLU激活层和一个最终的输出层构成。这种解耦设计允许每个分支独立地学习专属于该类数据的映射函数,从而提升预测的专业性与准确性。模型最终输出与输入维度相对应的预测结果,其中振动头输出重构的振动特征,SCADA头输出重构的SCADA特征。

1.4 动态训练与损失函数优化

模型采用自监督学习范式进行训练,其核心思想是通过重构人为构造的缺失数据,使模型掌握特征间的内在关联。训练过程中的关键环节包括动

态掩码生成、数据增强及多目标损失函数的设计,流程如图2所示。

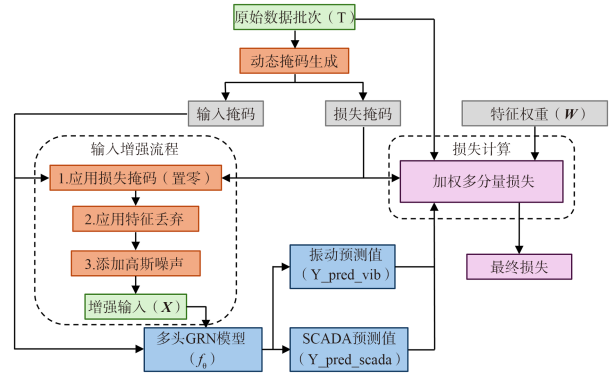


图2 模型训练策略流程

Fig. 2 Flowchart of model training strategy

在每个训练批次中,首先基于原始数据批次 T 为每个样本随机生成一个损失掩码 M 。该掩码用于标记当前批次中需要被模型预测的特征位置,即模拟的缺失特征。随后,对数据进行序列化增强以生成增强后的输入 X ,首先,根据损失掩码 M 将待预测的特征值置零,以模拟数据缺失的真实场景;其次,在已知特征中随机丢弃部分信息,迫使模型不过度依赖单一特征,提升其泛化能力;最后,向数据中添加微小的随机高斯噪声,以模拟传感器测量误差和信号扰动,从而增强模型对输入扰动的鲁棒性。增强后的数据 X 被输入多头GRN模型,从而得到预测值。

损失函数的设计是训练的核心,它兼顾了预测精度与模型泛化能力。总损失由振动头损失与SCADA头损失加权求和构成:

$$L_{\text{total}} = \lambda_{\text{vib}} L_{\text{vib}} + \lambda_{\text{scada}} L_{\text{scada}} \quad (4)$$

式中: L_{vib} 和 L_{scada} 分别为振动预测头和SCADA预测头的损失; λ_{vib} 和 λ_{scada} 为用于平衡两类任务重要性的权重系数。每个预测头的损失函数 L_{head} 进一步细化为多个目标的组合:

$$L_{\text{head}} = \sum_{i \in \text{Masked}} w_i \left[L_{\delta}(y_i, \hat{y}_i) + \alpha |y_i - \hat{y}_i| + \beta \|\hat{y}_i\|_1 \right] \quad (5)$$

式中:Masked表示由损失掩码 M 指定的需要计算损失的特征索引集合; w_i 为对应第 i 个特征的权重,用于调节不同特征对损失的贡献; L_{δ} 为Huber损失; α 为控制平均绝对误差(MAE)项权重的超参数; β 为 L_1 正则化项的权重,此项仅应用于振动预测头,用于促进预测值的稀疏性,鼓励模型捕捉更本质的振动模式。

Huber 损失是一种结合了均方差(MSE)和平均绝对误差(MAE)优点的鲁棒损失函数,其定义为:

$$L_{\sigma}(y, \hat{y}) = \begin{cases} \frac{1}{2}(y - \hat{y})^2 & |y - \hat{y}| \leq \delta \\ \delta \left(|y - \hat{y}| - \frac{1}{2}\delta \right) & \text{其他} \end{cases} \quad (6)$$

式中: y 为真实值; $\hat{y}$ 为预测值; δ 为一个超参数阈值。Huber 损失综合了均方误差对小误差敏感与平均绝对误差对异常值鲁棒的优点,当预测误差小于 δ 时采用MSE,误差较大时则转为MAE,从而在保证梯度稳定性的同时,提升模型对噪声数据的鲁棒性。

模型采用Adam优化器进行参数更新。为加速收敛并避免陷入局部最优,引入了带重启的余弦退火(cosine annealing with restarts)学习率调度策略。训练初期设置学习率预热(warm-up)阶段,有助于稳定训练过程。同时应用梯度裁剪(gradient clipping)技术,将梯度的 L_2 范数限制在预设阈值内,以防止梯度爆炸。

1.5 性能评估指标

为全面评估模型预测性能,选用决定系数(R^2)、均方根误差(RMSE)和平均绝对百分比误差(MAPE)作为核心评价指标,同时引入标准差(σ)辅助分析数据离散程度。

决定系数衡量模型对真实值变异性的解释能力,其值越接近1,表示拟合效果越好:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (7)$$

式中: n 为样本总数; y_i 为第 i 个样本的真实值; $\hat{y}_i$ 为第 i 个样本的预测值; $\bar{y}$ 为所有样本真实值的均值。

均方根误差反映预测误差的总体水平,它对较大误差敏感,计算式为:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (8)$$

平均绝对百分比误差以百分比形式呈现相对误差,便于直观理解,计算公式为:

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (9)$$

RMSE与MAPE的值越接近于0,表明预测精度越高。需注意,MAPE在真实值接近零时可能失真,

因此适用于目标值始终为正且远离零的场景。标准差 σ 用于描述数据波动范围,可辅助判断预测误差相对于数据自身变异的程度。

2 实验数据处理与特征筛选

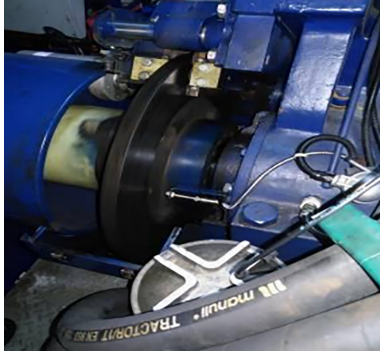
为构建高精度数据填补模型,需对原始SCADA系统与振动监测系统的多源数据进行系统化处理,包括数据对齐、特征提取、物理衍生特征构建及特征筛选降维。本章详细阐述数据来源、融合方法、衍生特征设计及特征选择过程,为后续模型训练奠定基础。

2.1 数据来源与特征衍生

2.1.1 数据来源

本研究基于中国某海上风电场项目,项目所处海域风能资源丰富,且大规模采用了便于施工的吸力筒导管架基础和大容量风机。本文采用的数据取自该项目的一台10 MW海上风力发电机。利用其监控与数据采集系统(SCADA)和独立的振动监测系统,采集并存储了该机组在2023年运行过程中的多维度数据,涵盖运行工况、环境参数以及机组机械与电气设备的功能状态。2023年下半年的运行数据,涵盖齿轮箱、发电机、变压器等关键部件的温度、压力、功率、风速、状态等参数,共874项原始特征,采样频率为10 Hz。振动监测系统同步采集了24项特征,包括塔筒顶部倾角、振动、位移及加速度等参数。传感器沿塔筒高度布置于5个截面(A~E),数据采集时段为2023年3月。采集风机转速以及风轮状态的设备如图3(a)、(b)所示,收集所得的功率曲线风速过程线与主轴转速过程线如图3(c)、(d)所示,这些曲线直观地展示了数据的连续性与可靠性。

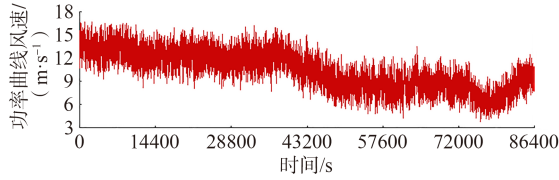
整体数据的采集时长为2023年3—12月。后续模型训练选取其中一天时间靠近的63840个样本进行训练,模型采用两阶段随机比例进行训练集、验证集以及测试集的划分。第一阶段为测试集划分,采用函数方法在全量数据中随机抽取15%的样本作为独立测试集,剩余部分作为训练集与验证集。第二阶段则使用五折交叉验证对剩余数据进行划分,每一折训练中,模型再次将剩余数据随机分为80%的训练集和20%的验证集。上述随机划分并非按照时间顺序进行,同时严格保证同一样本不会同时出现在测试集和训练集中,不存在样本重叠。



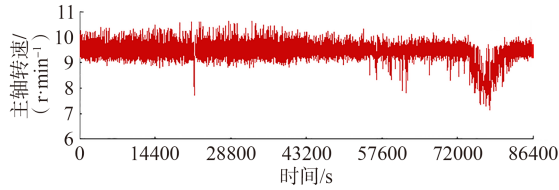
(a) 风机转速检测



(b) 风轮状态检测



(c) 功率曲线风速过程线



(d) 主轴转速过程线

图3 采集设备与所得曲线

Fig. 3 Data acquisition equipment and measured curves

2.1.2 数据对齐与特征提取

由于 SCADA 与振动系统原始采样频率不一致,需进行数据对齐与特征融合。首先,对所有振动通道信号采用零相位四阶巴特沃斯带通滤波器进行滤波,以分离主要动态响应并抑制噪声与基线漂移。滤波后信号按 1 s 间隔(对应 SCADA 系统 10 个采样点)分割为不重叠片段。对每个片段提取时域与频域特征。时域特征包括均方根值(RMS)、峰度(Kurtosis)和平均绝对值(MAV),计算公式如下:

$$Kurtosis = \frac{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^4}{\left[\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2 \right]^2} \quad (10a)$$

$$MAV = \frac{1}{N} \sum_{i=1}^N |x_i| \quad (10b)$$

式中: N 为采样点数; x_i 为滤波后的离散信号点; $\bar{x}$ 为该段信号的均值。

频域特征通过对每个片段进行快速傅里叶变换(FFT)获得频谱序列 $X(k)$,并提取主幅度(MainAmp)及其对应主频率(MainFreq):

$$X(k) = \sum_{n=0}^{N-1} x(n) \cdot e^{-j \frac{2\pi}{N} kn} \quad k \in (0, N-1) \quad (11a)$$

$$MainAmp = \max_k |X(k)| \quad (11b)$$

$$MainFreq = \frac{f_s}{N} \cdot \arg \max_k |X(k)| \quad (11c)$$

式中: n 为时间序列索引; k 为频率索引; j 为虚数单位; f_s 为振动信号的采样频率。

该流程每秒为每个振动通道生成 5 个特征(RMS, Kurtosis, MAV, MainAmp, MainFreq),振动传感器布置于塔筒的 A~E 五个截面,每个截面安装 2 个传感器,共计 10 个通道,由此构成 50 维振动特征集。

2.1.3 物理衍生特征构建

原始 SCADA 数据中的角度变量(如风向、偏航角)存在 $0^\circ \sim 360^\circ$ 的周期性数值突变,且这些变量与风机的气动载荷呈复杂的非线性关系。为增强模型对复杂气动工况的辨识能力,使其能够捕捉推力与转矩对风速的非线性响应,构造风速与风向正、余弦的耦合交互项,形成 2 个向量特征,以增强模型对风能矢量空间分布的表征能力。最后,保留归一化参考风速作为基准量。这 10 项衍生特征构成了一个富含物理先验信息的特征空间,有助于模型捕捉因气动载荷不对称而引发的异常振动模式。

2.2 特征筛选与降维

原始 SCADA 数据含 874 项特征,其中包含大量冗余或弱相关变量,直接输入模型可能导致过拟合与计算负担。因此需进行系统化特征筛选。首先剔除变化不明显的特征(如恒定值或近恒定值)及传感器存在但无有效数据输出的通道,初步保留 120 项特征。这些特征覆盖机组关键运行状态。

为量化剩余特征间的线性相关程度,本研究采用皮尔逊相关系数进行分析。皮尔逊相关系数 ρ 衡量

了两个变量 X 和 Y 之间线性相关的强度和方向,其定义为:

$$\rho_{X,Y} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad (12)$$

式中: X_i 和 Y_i 分别是变量 X 和 Y 的第 i 个观测值; $\bar{X}$ 和 $\bar{Y}$ 分别是 X 和 Y 的样本均值。系数的取值范围为 $[-1, 1]$, 绝对值越接近 1, 表示线性相关性越强。

基于初步筛选的 120 项特征, 绘制了相关性热

图, 如图 4 所示。为便于观察, 从中选取了高中低相关系数共 20 项特征进行细节展示, 如图 5 所示。

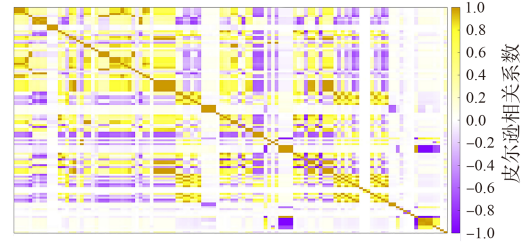


图 4 120 项特征参数相关性图

Fig. 4 Correlation matrix of 120 feature parameters

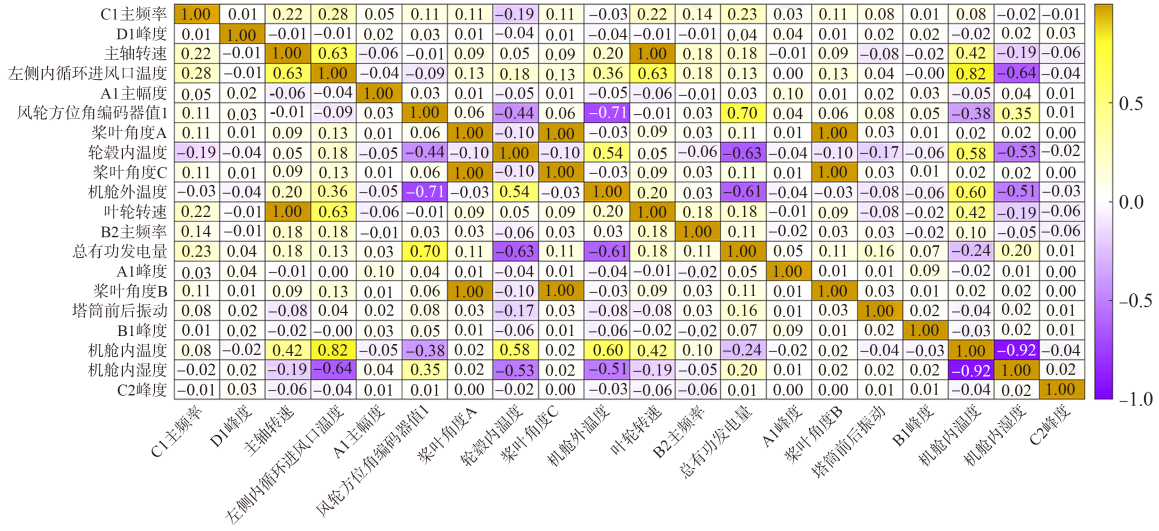


图 5 20 项特征参数相关性

Fig. 5 Correlation matrix of 20 feature parameters

在进行特征选择时, 首先以发电机转速, 即评估失控风险的关键指标为目标, 筛选出与其显著相关的特征。然后, 对于任意两个特征之间皮尔逊相关系数绝对值超过 0.95 的冗余对, 仅保留其中一个与目标变量更相关或更具物理意义的特征。通过上述步骤, 最终从 120 项中精选出 31 项关键 SCADA 特征, 如表 2 所示。这些特征全面涵盖了风速、转速、温度、电压、电流、振动、桨距角、风向等关键维度。

表 2 物理衍生特征变量

Table 2 Physically derived feature variables

变量类别	衍生变量名称	数量
气动能量因子	风速平方 (v^2)	1
角度连续化编码	风向正余弦、偏航角正余弦、偏航误差正余弦 ($\sin\theta_{wd}, \cos\theta_{wd}$)	6
向量交互特征	风速-风向耦合分量 ($v \cdot \sin\theta_{wd}, v \cdot \cos\theta_{wd}$)	2
基准参考量	归一化参考风速 (v_{ref})	1
总计		10

2.3 数据归一化

将筛选出的 31 项 SCADA 特征、经处理得到的 50 维振动特征以及 10 项物理衍生特征进行合并, 构成一个包含 91 维特征的融合特征向量(表 3)。该向量集成了机组运行参数、结构振动响应以及物理机理信息, 为模型提供了丰富而全面的输入。

由于各特征的量纲和数值范围差异巨大, 为消除其对模型训练的不利影响并加速梯度下降收敛, 采用最小-最大归一化方法将所有特征值线性映射到 $[0, 1]$ 区间:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (13)$$

式中: x 为原始特征值; x' 为归一化后的特征值; $\min(x)$ 和 $\max(x)$ 分别为该特征在数据集中的最小值和最大值。经过上述处理, 最终获得标准化后的 91 维特征数据集, 用于后续的模式训练与验证。

表 3 筛选后的特征参数

Table 3 Selected feature parameters

参数名称	单位	参数名称	单位	参数名称	单位
功率曲线风速	m/s	机舱内部温度	℃	有功功率	kW
瞬时风速	m/s	机舱外部温度	℃	无功功率	kW
30 s 平均风速	m/s	轮毂内部温度	℃	电网 A 相	V
				电压	
10 min 平均风速	m/s	发电机定子绕组	℃	电网 B 相	V
		温度		电压	
转子转速	m/s	变流器冷却介质	℃	电网 C 相	V
		温度		电压	
发电机转速	m/s	左内循环入口	℃	电网 A 相	kA
		温度		电流	
主轴转速	m/s	机舱内部	%	电网 B 相	kA
		湿度		电流	
塔筒前后振动	m	桨叶角度 A	(°)	电网 C 相	kA
位移				电流	
塔筒侧向振动	m	桨叶角度 B	(°)	总有功	kW·h
位移				发电量	
绝对风向		桨叶角度 C	(°)		
机舱绝对位置		转子方位角编码	(°)		
		器值			

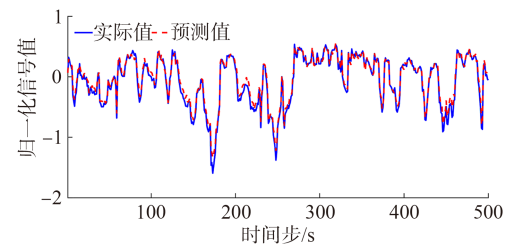
3 分析与验证

3.1 模型性能评估

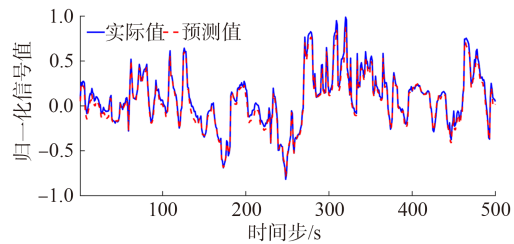
采用经预处理后的 SCADA 与振动监测数据对多头门控残差网络模型进行训练。为评估模型预测能力,分别从两个系统中选取关键特征参数进行对比分析。SCADA 系统选取发电机转速、电网 A 相电流、有功功率及电网 B 相电压;振动系统选取均方根值、平均绝对值及主幅值。将各参数的模型预测值与实际观测值进行对比,结果如图 6 与图 7 所示。

图 6 展示了 SCADA 系统四项关键参数的预测曲线与实际观测曲线的对比。可见:发电机转速、电网 A 相电流、有功功率的预测曲线与实际曲线高度吻合,模型表现出优异的拟合能力;电网 B 相电压虽波动性较大,但预测曲线仍能准确捕捉其整体变化趋势。图 7 展示了振动系统三项参数的预测结果。相较于 SCADA 数据,振动信号包含大量高频尖峰,波动更为剧烈。模型预测在尖峰处略显平滑,但对信号的整体趋势及关键波峰波谷的捕捉依然精准,表明模型对非线性、非平稳信号具有较强的泛化能力。

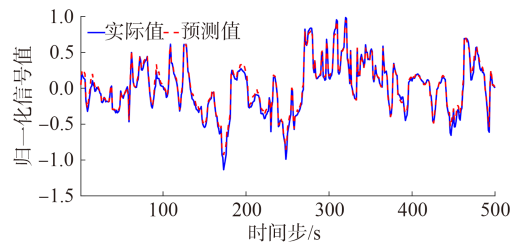
为进一步量化模型性能,采用决定系数 R^2 、均方根误差 RMSE 和平均绝对百分比误差 MAPE 作为评价指标,结果如表 4 所示,表中为进行 10 次独立随机



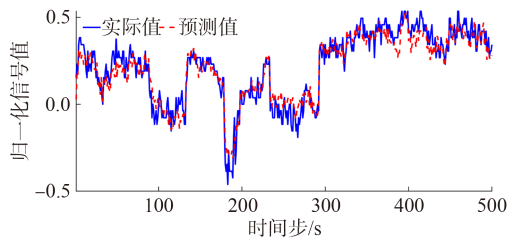
(a) 发电机转速



(b) 电网 A 相电流



(c) 有功功率



(d) 电网 B 相电压

图 6 SCADA 系统的预测值和观测值的比较

Fig. 6 Comparison of predicted and observed values of the SCADA system

抽取形成的平均值以及标准差值。SCADA 系统 4 项参数的 R^2 均大于 0.98, RMSE 低于 0.05, MAPE 小于 3%, 表明模型对 SCADA 数据具有极高的预测精度。振动系统三项参数的 R^2 均大于 0.92, RMSE 低于 0.03, MAPE 除主幅值略高于 10% 外, 其余均小于 10%, 同样验证了模型的有效性。

3.2 与基线填补方法的对比分析

为深入验证多头门控残差网络在多源数据重构任务中的性能优势,本研究选取线性插值法(LI)与随机森林回归算法(RFR)作为对比基线。其中,线性插值法基于时间邻域连续性假设,系工程实践中处理传感器数据缺失的常用基础方法;随机森林

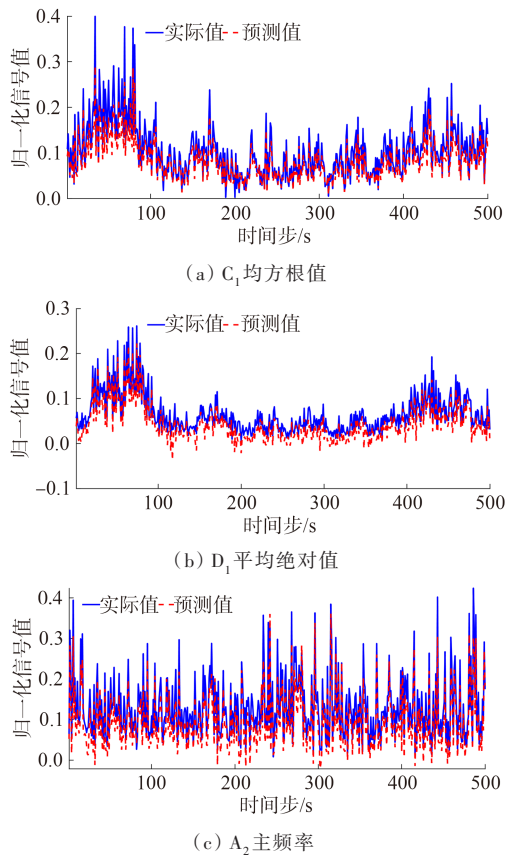


图7 振动系统的预测值和观测值比较

Fig. 7 Comparison of predicted and observed values of the vibration system

表4 模型预测性能指标

Table 4 Prediction Performance metrics of the proposed model

参数	R^2	RMSE	MAPE/%
发电机转速	0.9949 ± 0.0126	0.0468 ± 0.0005	0.5504 ± 0.2074
电网 A 相电流	0.9921 ± 0.0008	0.0229 ± 0.0004	2.2317 ± 0.4207
有功功率	0.9917 ± 0.0018	0.0289 ± 0.0007	2.8296 ± 0.3301
电网 B 相电压	0.9874 ± 0.0008	0.0310 ± 0.0002	0.1164 ± 0.0425
C_1 均方根值	0.9757 ± 0.0054	0.0114 ± 0.0003	8.0789 ± 0.2105
D_1 平均绝对值	0.9725 ± 0.0122	0.0097 ± 0.0001	8.9481 ± 0.3117
A_2 主频率	0.9237 ± 0.0054	0.0214 ± 0.0002	13.701 ± 0.4384

则作为经典机器学习模型,具备对多变量非线性映射关系的拟合能力。实验部分严格遵循前文所述的数据划分策略,以发电机转速、电网 A 相电流、 C_1 均方根值及 D_1 平均绝对值 4 项关键参数为对象进行填补效果评估,并统计 10 次独立重复实验所得评价指标的均值与标准差,以保证对比结果的稳健性与统计意义,结果如表 5 所示。

由表 4、5 所示量化评估结果可知,所提填补模型在 4 项关键参数的重构任务中均取得最优性能。

表5 模型对比性能指标

Table 5 Comparative performance metrics of models

参数	R^2	RMSE	MAPE/%
发电机转速(LI)	0.9521 ± 0.0234	0.0812 ± 0.0012	1.8420 ± 0.4125
电网 A 相电流(LI)	0.9315 ± 0.0187	0.0654 ± 0.0021	5.8741 ± 0.6542
C_1 均方根值(LI)	0.8423 ± 0.0315	0.0354 ± 0.0015	15.421 ± 1.0245
D_1 平均绝对值(LI)	0.8214 ± 0.0421	0.0287 ± 0.0018	16.874 ± 1.2451
发电机转速(RFR)	0.9103 ± 0.0000	0.1245 ± 0.0000	3.1254 ± 0.0000
电网 A 相电流(RFR)	0.8842 ± 0.0000	0.0912 ± 0.0000	8.9412 ± 0.0000
C_1 均方根值(RFR)	0.7154 ± 0.0000	0.0512 ± 0.0000	22.147 ± 0.0000
D_1 平均绝对值(RFR)	0.6841 ± 0.0000	0.0421 ± 0.0000	25.412 ± 0.0000

在发电机转速的填补中,该模型 R^2 达 0.9949,相较于线性插值法($R^2=0.9103$)提升约 9.29%;均方根误差由 0.1245 降至 0.0468,表明该架构对气动工况所致动态波动具有更强的捕捉能力。针对随机性更为显著的振动数据,随机森林回归算法在 D_1 平均绝对值项上的 R^2 仅为 0.6841,反映出传统插值策略在处理高频结构响应时的局限。相较之下,所提方法在同一参量上取得 $R^2=0.9725$ 的重构精度,较随机森林回归提升约 18.39%,验证了门控残差块对非线性耦合特征的有效提取。依托多头并行输出机制实现的协同重构框架,所提模型不仅在 SCADA 数据填补中展现出较高的复原精度,亦在复杂振动特征的还原中表现出较强的鲁棒性,可为海上风机多源监测数据的完整性提供技术支撑。

3.3 模型泛化能力验证

为了评估模型在实际应用中面对未知数据的泛化能力,选取了非训练日期的 SCADA 与振动监测系统非重叠时段数据进行预测分析。结合图 8 的预测曲线与表 6(表中为进行 10 次独立随机抽取形成的平均值以及标准差)的性能指标可以看出,模型在非训练数据集上依然能够较好地跟踪原始信号的整体变化趋势,具备优秀的泛化性能。从总体评价指标对比来看,相较于 3.1 节中模型在训练数据集上的表现,模型对非训练数据的预测结果呈现出决定系数 R^2 整体略微下降、平均绝对百分比误差 MAPE 有所上升的趋势。这符合深度学习模型在面对分布存在差异的未知工况数据时的常规表现。

进一步以在表 4 和表 6 中均被选作评估参数的电网 B 相电压和 C_1 均方根值为例进行量化分析。对于 SCADA 系统中的电网 B 相电压,其 R^2 由训练时段的 0.9874 下降至非训练时段的 0.9548,MAPE 则由 0.1164% 上升至 7.2971%;对于振动系统中的 C_1 均

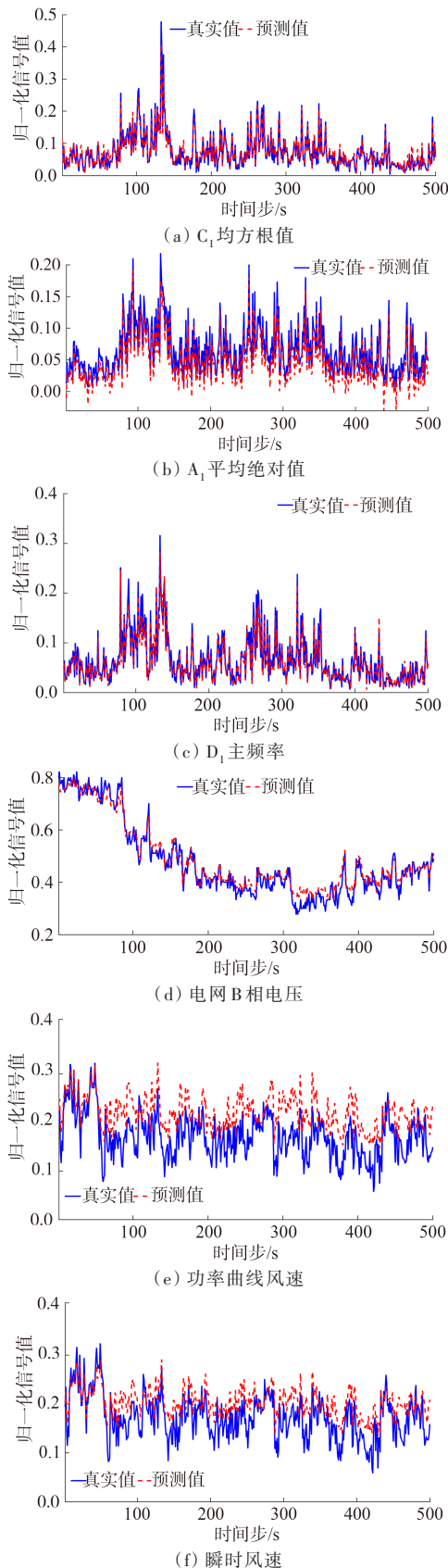


图 8 非训练日期的预测值和观测值比较

Fig. 8 Comparison of predicted and observed values for non-training periods

表 6 模型泛化性能指标

Table 6 Generalization performance metrics of the proposed model

参数	R^2	RMSE	MAPE/%
C_1 均方根值	0.9681 ± 0.0007	0.0184 ± 0.0003	11.0604 ± 0.1416
A_1 平均绝对值	0.9347 ± 0.0014	0.0256 ± 0.0002	22.6520 ± 0.3095
D_1 主频率	0.9332 ± 0.0017	0.0238 ± 0.0003	17.4078 ± 0.1439
电网 B 相电压	0.9548 ± 0.0006	0.0403 ± 0.0003	7.2971 ± 0.0795
功率曲线风速	0.8923 ± 0.0017	0.0671 ± 0.0005	21.7703 ± 0.3011
瞬时风速	0.8703 ± 0.0021	0.0744 ± 0.0007	18.5450 ± 0.2591

方根值,其 R^2 由 0.9757 微降至 0.9681, MAPE 由 8.0789% 上升至 11.0604%。数据对比及图像走势表明,尽管面对非训练日期的离散数据,模型在局部突变或极值处的拟合精度略有降低,但各项核心评价指标仍保持在较高水平,说明模型针对非训练日期的数据依然表现出良好的重构与预测功能,但在复杂非平稳工况的细节特征捕捉方面仍具有一定的提升空间。

此外,综合对比非训练数据中两类系统的预测指标可以观察到,模型对振动类型数据的泛化性能整体优于 SCADA 系统数据。在非训练时段下,代表振动系统特征的性能指标下降幅度相对较小。另外,从图 8 的对比曲线中可见其预测值与真实值的重合度更高。这表明本文所提出的多头门控残差网络框架,在融合多源特征并进行独立分支预测后,能够更加深入地学习到风电机组结构振动响应的深层物理规律。因此,模型在面对非训练日期的振动系统数据时,功能衰减更小,预测结果相对更为准确稳定。

4 结 论

针对海上风电机组多源传感器数据在复杂工况下易发生缺失的问题,本文提出了一种基于多头门控残差网络的通用数据填补框架。通过对真实 10 MW 海上风机 SCADA 数据与振动监测数据的特征融合与模型验证,主要得出以下结论:

1) 构建的多头门控残差网络模型能够有效提取 SCADA 系统与振动监测系统数据间的高维非线性物理关联。引入的动态掩码与混合损失函数策略,在无需完整标签数据的前提下,实现了对异构监测数据的高精度自监督联合重构。

2) 在训练集数据表现上,模型对 SCADA 关键特征和振动特征均展现出优异的拟合能力。SCADA

参数决定系数总体大于0.98,振动参数决定系数大于0.92,能够精准捕捉低频运行状态与高频结构动态响应的整体规律与关键波峰波谷。

3)对比实验显示,所提模型在发电机转速与振动参量重构中, R^2 较线性插值、随机森林分别提升9.29%与18.39%,验证了门控残差结构对非线性耦合特征提取的有效性,为后续泛化性评估提供参照。

4)在非训练日期的泛化能力验证中,模型对SCADA数据的预测整体决定系数略有下降,平均绝对百分比误差有所上升,局部极值拟合存在提升空间,但依然能够准确追踪信号变化趋势,保持了良好的预测功能。

5)模型对非训练日期的振动系统数据表现出比SCADA数据更小的功能衰减,表明该框架在挖掘深层结构动力学特征方面具有更强的鲁棒性,可为海上风电机组的全寿命周期状态评估提供高连续性的数据基础。

参 考 文 献

- [1] WANG F X, ZHANG S, HOU Y Z, et al. Extraction of offshore wind turbines in china by combining multispectral and SAR image data [J]. *Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2024, 17:3392695.
- [2] 蔡彦枫,李晓宇,任宗栋,等. 面向空中风力发电系统的风能资源分析[J]. *太阳能学报*, 2025, 46(2):624-632.
- [3] 张珺辉,董绍华,叶迎春. 基于LabVIEW的海上风机齿轮箱故障诊断系统设计[J]. *新技术新工艺*, 2025(1):29-37.
- [4] LEI Y G, YANG B, JIANG X W, et al. Applications of machine learning to machine fault diagnosis: a review and roadmap [J]. *Mechanical Systems and Signal Processing*, 2020, 138: 106587.
- [5] TAO S Y, QIAN Z, PEI Y, et al. Wind turbine failure detection based on SCADA data and data mining method [C]// *Renewable Power Generation Conference*. 2019.
- [6] ABORGEL T, SHEHATA A S, KOTB M A, et al. Heavy lift semi-submersible ships utilization in offshore wind turbines industry [J]. *Energy Reports*, 2022, 8: 834-847.
- [7] ZHANG W W, VATN J, RASHEED A. Gearbox pump failure prognostics in offshore wind turbine by an integrated data-driven model [J]. *Applied Energy*, 2025, 380: 124829.
- [8] HUSSAIN M, MIRJAT N H, SHAIKH F, et al. Condition monitoring and fault diagnosis of wind turbine: a systematic literature review [J]. *Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2024, 35:14747.
- [9] GOROSTIDI N, PARDO D, NAVA V. Diagnosis of the health status of mooring systems for floating offshore wind turbines [J]. *Renewable Energy and Power Quality Journal*, 2023, 287: 115862.
- [10] ENCALADA-DAVILA Á, TUTIVÉN C, MOYÓN L, et al. Main bearing fault prognosis in wind turbines based on gated recurrent unit neural networks [J]. *Renewable Energy and Power Quality Journal*, 2022, 20: 2172-2183.
- [11] ZHAO J H, SHEN X D, LIU T J, et al. Probabilistic imputation of missing offshore wind speed based on conditional diffusion models [J]. *Sustainable Energy, Grids and Networks*, 2025, 44: 101949.
- [12] QU F M, LIU J H, MA Y J, et al. A novel wind turbine data imputation method with multiple optimizations based on GANs [J]. *Mechanical Systems and Signal Processing*, 2020, 139: 106610.
- [13] LIAO W L, BAK-JENSEN B, PILLAI J R, et al. Data-driven missing data imputation for wind farms using context encoder [J]. *Journal of Modern Power Systems and Clean Energy*, 2022 (2) : 000894.
- [14] FAN Y W, FENG C L, WU R, et al. Multiscale-attention masked autoencoder for missing data imputation of wind turbines [J]. *Knowledge-Based Systems*, 2024, 299: 112114.
- [15] RADEV B, KASI V R, BINGI K. Enhancement of Texas wind turbine power predictions using fractional order neural network by incorporating machine learning models to impute missing data [J]. *Knowledge-Based Systems*, 2024, 300: 112176.
- [16] WANG B, ZHOU B, ZOU M H, et al. Research on condition monitoring of offshore wind turbine gearbox based on K-means clustering and extreme learning machine [C]// *International Conference on Artificial Intelligence and Power Engineering*. 2023.
- [17] JU S H, HUANG Y C. Analyses of offshore wind turbine structures with soil-structure interaction under earthquakes [J]. *Ocean Engineering*, 2019, 187: 106190.
- [18] GAO B, LI C, ZHOU F L, et al. Shaking table tests of offshore wind turbine systems with a suction bucket foundation in sandy seabed subject to earthquake and wind loads [J]. *Marine Structures*, 2025, 99: 103706.
- [19] BAI H, ZHANG J H, SUN K, et al. Nonlinear long-term vibration response prediction of offshore wind turbines under full operating conditions based on deep learning [J]. *Applied Ocean Research*, 2025, 159: 104625.
- [20] 赵洪山,刘辉海,刘宏杨,等. 基于堆叠自编码网络的风电机组发电机状态监测与故障诊断[J]. *电力系统自动化*, 2018, 42(11):102-108.
- [21] 张舒翔,刘晓彤,郭旭峰,等. 基于SCADA数据驱动的风电机组状态监测[J]. *电工技术*, 2022(22):47-50.
- [22] 王凤,刘伊彤,邹著,等. 考虑突发关联状态的风力机组故障预警方法[J/OL]. *系统工程理论与实践*, 2026[2026-03-09]. <https://link.cnki.net/urlid/11.2267.n.20240929.0955.009>.