

# 基于多源数据融合的光伏组件故障预警模型研究

赵禹茗

国电电力内蒙古新能源开发有限公司 内蒙古 呼和浩特 010010

**摘要:** 光伏产业规模化发展下,组件故障会降低发电效率、增加安全风险,传统检测方法难精准预警。本文提出基于多源数据融合的故障预警方案,梳理数据融合理论与故障机理,构建含数据预处理、特征融合、预警判断的完整模型。采集电气、环境、物理状态三类数据,采用特征层与决策层融合策略,搭配优化机器学习算法实现预警。该研究为光伏组件故障预警提供技术支撑,可降低运维成本,保障电站稳定运行,推动产业高质量发展。

**关键词:** 光伏组件;故障预警;多源数据融合;特征提取;机器学习

**引言:** 全球能源转型中,光伏作为清洁能源核心,装机容量持续攀升。光伏组件长期处于户外复杂环境,受光照、温度等因素影响易出现故障,降低发电效益甚至引发安全事故。当前主流检测方法如人工巡检、红外热成像等,存在效率低、成本高、滞后等问题,难以满足大规模电站运维需求。多源数据融合技术能整合不同维度数据互补信息,提升决策准确性。本文基于此,研究其在光伏组件故障预警中的应用,明确关键环节,通过实验验证模型,为电站运维提供精准预警方案。

## 1 多源数据融合理论基础

### 1.1 数据融合层次

数据融合层次是多源数据融合技术的核心框架,按处理阶段不同可分为数据层、特征层与决策层三类,各层次具有独特适用场景与技术特点。数据层融合又称像素级融合,直接对原始数据进行整合处理,保留原始信息最完整,适用于同源数据或格式一致的数据融合,如多个传感器采集的同一组件表面温度数据融合,但该层次对数据噪声敏感,计算量较大。特征层融合先对各数据源进行特征提取,再对特征向量进行融合,兼顾信息完整性与处理效率,是当前光伏领域应用较广的融合层次,可整合电气数据与环境数据的关键特征。决策层融合则基于各数据源独立决策结果进行最终决策,对数据预处理要求低,容错性强,适合处理异构性强的多源数据,但依赖各独立决策模块的性能。实际应用中需根据数据类型、融合目标及硬件条件,选择单一层次或多层次结合的融合方案,以实现融合效果与处理效率的平衡。

### 1.2 多源数据类型分析

光伏组件故障预警相关的多源数据可分为电气类、环境类与物理状态类三大类,各类数据从不同维度反映组件运行状态与故障特征。电气类数据是核心监测数据,包括组件输出电压、电流、功率、填充因子及逆变器运行

参数等,直接关联组件发电性能,故障发生时会出现显著异常波动,如热斑故障会导致局部功率骤降。环境类数据涵盖光照强度、环境温度、风速、降水量、湿度等,虽不直接反映故障,但影响组件运行状态与故障演化速度,如高温环境易加剧热斑故障发展<sup>[1]</sup>。物理状态类数据包括组件表面温度、红外图像特征、表面污染程度等,可直观反映组件物理结构与外观状态,如隐裂故障可通过红外图像灰度值变化识别。三类数据存在较强互补性,单一数据难以全面覆盖故障特征,需通过融合实现信息互补,为精准预警提供数据支撑。

### 1.3 数据融合方法

适用于光伏组件故障预警的多源数据融合方法多样,按技术原理可分为传统融合方法与智能融合方法两大类。传统融合方法包括加权平均法、贝叶斯估计法、D-S证据理论等,加权平均法通过赋予不同数据源权重实现数据整合,计算简单、实时性强,适用于数据层融合;贝叶斯估计法基于概率统计理论,可处理不确定性数据,适合小样本数据融合;D-S证据理论通过信任函数融合多源决策信息,容错性强,广泛应用于决策层融合。智能融合方法以机器学习与深度学习为核心,包括神经网络融合法、支持向量机(SVM)融合法、深度学习融合法等,神经网络融合法可自适应学习数据间映射关系,适合处理非线性数据;深度学习融合法如卷积神经网络(CNN)、长短期记忆网络(LSTM),可自动提取高维特征,适配复杂多源异构数据。实际应用中需结合数据类型、融合层次及预警需求,选择合适的融合方法,必要时进行方法改进与组合。

## 2 光伏组件故障机理与特征分析

### 2.1 常见故障类型

光伏组件常见故障按成因与表现形式可分为电气故障、物理故障与环境诱导故障三大类,各类故障具有不

同发生机理与危害程度。电气故障主要包括热斑故障、接线盒故障、旁路二极管故障等，热斑故障由组件局部遮挡导致，遮挡区域成为负载消耗其他区域电能并产生大量热量，长期可烧毁组件；接线盒故障多因密封不良、散热不畅导致，引发接触不良或短路，影响组件电流输出。物理故障涵盖隐裂、裂片、表面划伤、封装老化等，隐裂多由运输、安装过程中的机械冲击或温度应力导致，会降低组件功率输出并加剧老化；封装老化由长期紫外线照射、高温高湿环境引发，导致EVA胶膜黄变、背板开裂，影响组件绝缘性能。环境诱导故障包括表面污染、积雪覆盖、鸟粪堆积等，表面污染会降低透光率，减少光照吸收，积雪与鸟粪则造成局部遮挡，诱发热斑故障<sup>[2]</sup>。

## 2.2 故障特征提取

故障特征提取是实现精准预警的关键，需结合多源数据特性，提取各类故障的特异性特征。基于电气数据的特征提取方面，热斑故障表现为组件输出功率下降、开路电压轻微降低、短路电流显著减小，可提取功率衰减率、电压波动系数等特征；接线盒故障会导致组件串联电阻增大，可通过电阻变化率、电流稳定性指标表征。基于物理状态数据的特征提取中，隐裂故障在红外图像中呈现特定灰度值分布，可提取灰度均值、熵值、边缘特征等；表面污染可通过图像对比度、透光率衰减系数等特征量化。基于环境数据的辅助特征提取，可构建环境因子与故障发生概率的关联特征，如高温高湿环境下封装老化的加速特征。提取过程中需采用归一化、降维等处理，剔除冗余特征，保留对故障敏感的关键特征，为后续融合与预警提供高质量特征输入。

## 3 基于多源数据融合的光伏组件故障预警模型构建

### 3.1 模型总体框架

本文构建的光伏组件故障预警模型总体框架分为数据采集层、预处理层、特征融合层、预警模型层与结果输出层五个部分，形成完整的“数据-处理-融合-预警”闭环。数据采集层通过传感器、数据采集终端等设备，同步采集电气、环境、物理状态三类多源数据，传输至数据存储模块。预处理层对原始数据进行清洗、标准化、降维等处理，剔除异常数据与冗余信息，提升数据质量。特征融合层采用“特征层+决策层”的混合融合策略，先对电气数据与物理状态数据进行特征层融合，提取联合特征向量，再与环境数据的独立特征进行决策层融合。预警模型层采用优化后的LightGBM算法构建预警模型，输入融合后的特征数据进行模型训练与推理。结果输出层根据模型输出结果，结合预设阈值输出故障预警信息，包括故障类型、严重程度与发生位置，同时反馈至运维管

理系统。

### 3.2 数据预处理

数据预处理是保障模型性能的基础，针对多源数据的异构性与复杂性，设计多步骤预处理流程。首先进行数据清洗，采用3 $\sigma$ 原则识别电气数据中的异常值，通过邻近插值法填充环境数据中的缺失值，剔除物理状态数据中的无效图像数据，确保数据完整性与准确性。其次进行数据标准化，对电气数据与环境数据采用Z-Score标准化方法，消除量纲差异；对图像类物理状态数据采用归一化处理，将像素值映射至[0,1]区间。然后进行数据降维，采用主成分分析（PCA）方法对高维电气特征与环境特征进行降维，保留累计贡献率达95%以上的主成分；对图像特征采用局部二值模式（LBP）提取关键纹理特征，减少计算量。最后进行数据时序对齐，以秒级时间戳为基准，统一各类数据的时间维度，确保多源数据在时间上同步，为后续融合处理奠定基础。

### 3.3 特征融合策略

结合光伏多源数据特性，设计“特征层融合+决策层融合”的混合融合策略。特征层融合针对电气数据与物理状态数据开展，采用加权特征融合方法，基于互信息法计算各特征对故障的贡献度，赋予高贡献度特征更高权重，将电气特征（如功率衰减率、电压波动系数）与物理特征（如红外图像灰度熵、表面温度梯度）加权融合，形成联合特征向量，实现两类核心数据的信息互补。决策层融合针对特征层融合结果与环境数据开展，采用改进D-S证据理论，将特征层融合后的特征向量输入基础分类器得到初步决策结果，同时将环境特征输入另一基础分类器得到辅助决策结果，通过优化证据组合规则处理两类决策结果的冲突信息，得到最终融合决策结果<sup>[3]</sup>。

### 3.4 故障预警模型设计

选用LightGBM算法构建故障预警模型，并针对光伏数据特性进行优化。模型输入为融合后的特征向量，输出为故障类型与预警等级。首先对LightGBM算法的核心参数进行优化，采用贝叶斯优化方法确定学习率、决策树深度、叶子节点数等关键参数，提升模型收敛速度与泛化能力。其次引入注意力机制，对融合特征进行权重分配，强化对故障敏感特征的关注，提升模型对轻微故障的识别能力。模型训练过程中，采用交叉验证方法避免过拟合，将预处理后的数据集按7:2:1比例划分为训练集、验证集与测试集，基于训练集进行模型训练，通过验证集调整参数，利用测试集验证模型性能。模型预警逻辑为：将实时融合特征输入训练好的模型，输出故障概率，若概率超过预设阈值则触发预警，同时根据概率

大小划分“轻微、一般、严重”三级预警，为运维提供精准指导。

#### 4 实验验证与结果分析

##### 4.1 实验环境与数据集

实验硬件环境为Intel Core i7-12700H处理器、32GB内存、NVIDIA RTX 3060显卡，软件环境基于Python 3.8搭建，结合TensorFlow、Scikit-learn等库实现模型开发与验证。实验数据集来源于某100MW光伏电站，采集时间跨度为6个月，涵盖不同季节与天气条件。数据集包含电气数据（电压、电流、功率等，采样频率1次/分钟）、环境数据（光照、温度、风速等，采样频率1次/5分钟）、物理状态数据（红外图像、表面温度等，采样频率1次/小时），共包含正常样本10万条，热斑、隐裂、接线盒故障等故障样本各2万条。为提升模型泛化能力，采用数据增强技术扩充故障样本，通过添加微小噪声、调整数据幅值等方式生成新增样本，最终数据集规模达25万条，按比例划分为训练集、验证集与测试集。

##### 4.2 评价指标

选用多维度评价指标全面评估模型性能，核心指标包括预警精度、召回率、F1值、误报率与实时性。预警精度反映模型识别正确故障样本占总识别故障样本的比例，衡量预警准确性；召回率反映模型正确识别故障样本占实际故障样本的比例，衡量故障漏检程度；F1值为精度与召回率的调和平均数，综合评价模型识别性能；误报率反映模型将正常样本误判为故障样本的比例，衡量运维成本控制能力；实时性通过模型处理单条数据的平均时间衡量，评估模型适配实时预警需求的能力。同时设置对比实验，将所提模型与单一数据源模型（仅电气数据）、传统融合模型（加权平均+SVM）进行性能对比，突出所提模型的优势<sup>[4]</sup>。

##### 4.3 实验结果

实验结果显示，所提模型整体预警精度达95.8%，召

回率达94.2%，F1值达95.0%，误报率仅为2.3%，处理单条数据平均时间为0.08秒，完全满足实时预警需求。针对不同故障类型，模型对热斑故障的预警精度最高，达97.5%，对隐裂故障的预警精度为94.1%，主要因隐裂初期特征较微弱导致。对比实验结果表明，所提模型较单一数据源模型（精度82.3%、误报率11.5%）与传统融合模型（精度90.1%、误报率5.7%），在各项指标上均显著提升，其中精度分别提升13.5和5.7个百分点，误报率分别降低9.2和3.4个百分点。特征融合有效性验证显示，混合融合策略较单一特征层融合或决策层融合，F1值提升4.2~6.8个百分点，证明该策略可有效整合多源数据信息，提升模型性能。

#### 结束语

本文聚焦光伏组件故障预警，开展基于多源数据融合的模型研究，涵盖理论梳理、机理分析、模型构建与实验验证。研究揭示电气、环境、物理状态三类数据互补性，提出“特征层+决策层”混合融合策略，构建优化LightGBM预警模型，提升预警精度与实时性。实验表明，该模型性能优于传统方法，能精准识别常见故障。不过，极端环境下预警性能待验证，后续将拓展数据采集，结合数字孪生优化体系，并推进模型轻量化，推动其在中小电站应用。

#### 参考文献

- [1]白宇航,曹亮.基于多模型融合的雷电预警产品生成技术[J].电子技术与软件工程,2021(05):209-211
- [2]林翀,周皖奎,丁智华.基于HMM的分布式光伏电站故障预警方法研究[J].电力设备管理,2021(1):130-131.
- [3]徐峰,赵冠军,张志峥,等.基于最大准入功率计算的分布式光伏电源故障预警[J].科学技术与工程,2024,24(24):10394-10400.
- [4]吴春华,俞薛颖,李智华,等.基于FCM与高斯隶属度的光伏组件健康状态诊断[J].电网技术,2022,46(5):1887-1896.