

# 基于 SCADA 数据的风机叶片故障检测技术与展望

蒋志伟 李永坚

湖南工程学院电气与信息工程学院，湘潭

**摘要** | 以基于 SCADA 数据的风机叶片故障检测技术为研究对象，首先对风机叶片故障类型和故障成因进行分析，并对近十年以来 SCADA 系统监控数据的预处理技术、叶片故障诊断模型的建立进行了研究，经过综合对比，本文建议采用交叉验证的递归特征消除法（RFECV）和基于随机森林算法的故障诊断模型来处理更多的叶片故障。最后对风机叶片故障检测技术的未来发展方向进行了总结和建议。

**关键词** | 风机叶片；预处理技术；特征选择；模型；故障诊断

Copyright © 2021 by author (s) and SciScan Publishing Limited

This article is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/). <https://creativecommons.org/licenses/by-nc/4.0/>



中国作为世界第一的风电大国，风能的开发和利用得到了迅速的发展，然而随着风力发电机装机容量的逐渐增大，故障率也随之增加。根据凯斯内斯风电场信息论坛提供的事统计数据，截止 2020 年底全球风电场事故大多是由叶片故障所引起的，占事故总数的 16.5%。若不及时进行检查和维护，可能会引发电力系统的二次损坏，甚至对风电场操作人员构成安全隐患。因此发展叶片故障诊断技术对提高叶片的可靠性、安全性和使用寿命具有重要意义。

为了解决叶片故障给机组稳定运行带来的不利影响，国内外研究人员提出

基金项目：湖南省自然科学基金联合项目（2020JJ6017）。

通讯作者：李永坚（1971.05-），男，教授，硕士生导师，研究方向：特种电机和风电技术研究。

文章引用：蒋志伟，李永坚：基于 SCADA 数据的风机叶片故障检测技术与展望 [J]. 电气工程与技术, 2021, 2 (3) : 33-47.

<https://doi.org/10.35534/eet.0203011>

了多种无损检测技术<sup>[1-3]</sup>来为风力发电机提供更好的安全性和较长的检测周期,以实现可靠的电力供应。但上述方法需要借助外界设备才能够实现故障诊断,在叶片上大范围的安装传感器可能会影响叶片的力学性能并且传感器长期暴露在外界环境下容易发生损坏,增加额外的运维成本<sup>[4]</sup>。目前大多数风电场都采用数据采集与监测控制(SCADA)系统用于叶片的结构健康状态监测,并采集叶片工况数据用于后期分析和应用。为了能够更好的利用风场大数据,并在不需要借助外界设备的情况下对风机叶片进行有效检测,许多学者提出了基于SCADA数据的故障诊断技术<sup>[5-7]</sup>,结合数据分析,提取与叶片故障相关性较强的特征用作模型训练的输入,建立故障诊断模型后对其进行训练,针对训练结果进行参数优化,可以清晰的识别风机叶片故障的原因及过程,最终为风电场管理人员提出相应的风机叶片维护策略。

本文综述了近十年来研究风机叶片运行过程中出现的故障类型和风机叶片故障检测技术的最新进展。其中第一部分对风机叶片故障类型和故障成因以及对叶片的影响进行分析;第二部分介绍了针对SCADA系统监控数据预处理的最新研究成果;第三部分详述了关于建立风机叶片故障诊断模型的研究成果;第四部分是对风机叶片故障检测行业面临的挑战和未来发展方向进行了总结和建议。

## 1 风机叶片的故障分析

### 1.1 风机叶片故障的类型

风机叶片在运行过程中往往会由于内部应力和外部环境而导致不同程度的损坏,其中的典型损坏主要有裂纹<sup>[8]</sup>、粘合剂脱粘<sup>[9]</sup>、分层、叶片断裂、叶尖排水孔堵塞、边缘腐蚀<sup>[10]</sup>等。其中,裂纹、分层、脱粘是最主要的损坏。叶片中的损坏可能只是一种类型,也有可能是多种类型同时存在,因此需要定期对叶片进行状态监测,及时采取针对性预防措施,减少损坏的形成,保证结构的质量和机组的稳定运行。

### 1.2 风机叶片故障的原因

造成叶片故障的原因可以分为以下两个类别:

(1) 叶片制造安装过程中的人为失误引起的故障。目前市场上大多数风机叶片的制造仍以手工操作为主，而人为因素的不确定性是引起叶片损坏的一个极为重要的原因，如：叶片粘结处开裂和脱粘、粘结处存在气泡、铺面不平整等。据统计，每年由于叶片粘结处开裂而导致的叶片损坏居故障原因的首位，而叶片粘结处开裂就是属于生产过程中出现的制造工艺不严谨<sup>[11]</sup>。另一方面，叶片在运输安装过程中由于工人使用不当容易导致叶片表面磨损，该磨损将在风机今后的日常运行中埋下隐患。

(2) 外部条件引起的故障。一方面是外部环境引起的故障，如台风、暴雨、积冰、雷击、风沙颗粒物等，都会导致叶片不同程度的损坏。有数据表明，因极端环境因素导致风机叶片的故障占故障总数的 76%<sup>[12]</sup>。雷击会导致结构损坏，严重时会导致叶片涂层剥落；积冰会导致叶片旋转不平衡，增加叶片疲劳和安全风险；风沙颗粒物会导致叶片前缘磨损，前缘粘合处会因此开裂；台风可以使叶片载荷超过其额定载荷，即使叶片处于静止状态也会损坏叶片；在极端寒冷的环境下运行的风机可能会产生材料脆性断裂。另一方面是运行时的不当引起的故障。当风电机组运行过程中刹车系统失灵，转速不断增加，风轮越转越快，最终引发事故<sup>[13]</sup>。除风电机组自身故障，还有某些风电场为了提高经济效益，风电机组长期保持超负荷运行，最后使得叶片因疲劳失效引发整个机组故障。

风机叶片故障类型总结如表 1 所示。

表 1 风机叶片故障类型一览表

Table 1 List of fault types of wind turbine blades

故障类型	故障区域	故障成因	对叶片的影响
纤维断裂与屈曲	外部	过载或疲劳而导致应力屈服	导致材料的拉伸强度和压缩强度下降
纤维与基体界面开裂	外部	吸湿、热疲劳、应力疲劳、基体老化	增加了叶片表面粗糙度
粘合剂脱粘	外部	胶粘剂老化、应力屈服	丧失了层板间粘合的能力
拉力和屈曲载荷驱动分层	内部 / 外部	疲劳、基体老化	降低叶片的压缩强度和刚度，影响结构的完整性
叶尖排水孔堵塞	内部	空气中的颗粒物混入排水孔	影响排水、产生噪音
边缘腐蚀	外部	基体环境老化	降低机组的效率、增加叶片的振动和负荷

## 2 SCADA 数据预处理技术

在实际工程中, 由于 SCADA 系统采集到的叶片运行实时数据大多为正常数据, 从而使得原始数据集中的正常数据与故障数据产生严重的比例失衡, 这将会严重影响后期的检测精度。另一方面, 原始数据集中存在着多种与叶片故障具有相关性的特征, 若将所有特征都作为模型训练的输入, 势必会增加计算的复杂度, 甚至造成维数灾难。因此, 在模型训练之前需要对原始数据集进行预处理和特征提取。

### 2.1 数据预处理

针对原始数据集中存在的类别不平衡问题, 一般通过过采样、欠采样和混合采样三种方法进行处理, 使各类样本数量分布均衡, 提高数据集的分类效果。

过采样是通过复制样本的方式来增加类别数量较少的样本, 其常用方法有随机过采样、综合少数类过采样 (SMOTE)、分界线—过采样 (Borderline-SMOTE) 等。但是目前的过采样方法存在着合成重叠样本或噪声样本以及过拟合的问题, 对不平衡数据的分类效果不尽人意。重庆大学的李刘杰<sup>[14]</sup>针对此类问题提出了一种基于层次聚类的加权过采样法, 该方法首先对类别数量较少的样本进行聚类, 再结合每个类簇的采样倍率和少数类样本的采样权重在各合成区域中完成过采样。通过在多个数据集上的应用, 结果表明该方法可以有效提高整体的分类精度。石洪波<sup>[15]</sup>等介绍了关于 SMOTE 过采样的研究现状和工作原理, 以及多种扩展方法和应用场景, 并指出 SMOTE 算法在不同应用背景下可能遇到的问题和处理方法。

欠采样是通过随机挑选样本作为训练集来减少样本数量以达到平衡数据集的效果, 其常用方法有编辑最近邻居规则 (ENN)、邻居清除规则 (NCR)、Tomck 链接欠采样 (Tomck-links) 等。何云斌<sup>[16]</sup>等提出了一种基于聚类的加权边界点集成欠采样算法, 该算法可以实现对原始数据集的约减操作, 提高执行效率。相较于其他欠采样方法而言, 该方法可以有效的处理类别不平衡的问题。同时孟东霞<sup>[17]</sup>等通过特征边界的数据点来区分原始数据集中的分类信息, 对远

离分类决策面的多数类样本进行随机选择，分布在决策面附近的少数类样本进行保留，将两者综合可得到平衡数据集。与其他欠采样相比保留了多数类样本中的几何分布特征，避免了样本的信息流失率，有利于提高不平衡数据中少数类的分类精度。

在特殊情况下，单独使用过采样处理可能会引发模型的过拟合，而单独使用欠采样处理可能会导致样本重要信息的流失。为了提高对不平衡数据的分类精度通常采样混合采样的方法进行预处理，即欠采样与过采样相结合的方法。山西财经大学的陈鑫<sup>[18]</sup>针对欠采样易丢失样本重要信息的问题提出了基于数据局部分布的混合采样方法；针对过采样易造成过拟合、加重类间重叠的问题，提出了基于数据整体分布的混合采样法；将上述两种采样法应用于入侵检测领域，结果表明该方法可有效缓解数据类别不平衡的问题，并在不影响总体识别率的情况下提高少数类样本的正确识别率。

## 2.2 特征选择

在风机叶片故障诊断的过程中，要想实现较高的检测精度，首先要通过特征选择确定与叶片故障相关性的敏感特征，对特征数据进行一定的筛选，再结合算法模型强大的表征能力，才能实现有效的故障诊断。根据特征选择的形式，目前常见的特征选择方法主要分为三种：过滤法、包裹法和嵌入法。

包裹法是根据模型的效果对特征的优劣进行评价，通常将分类器的分类准确率作为特征子集的评价标准。而用来评价特征优劣的算法模型有很多，其中最为常用的是随机森林分类算法。但是由于在特征选择的过程中需要对分类器进行不断的训练，时间成本较高，并且特征选择的结果依赖于分类方法，容易产生过拟合。但相较于过滤法，包裹法找特征子集的方式更具有合理性。李大中<sup>[19]</sup>等利用随机森林算法来挖掘原始数据中与叶片结冰关联性较强的特征，该算法既可以用于样本特征分类，又可以用于分析各个分类特征的重要程度。重庆邮电大学的陈高升<sup>[20]</sup>提出了一种基于 OVO 的多评价准则的改进的 LVW 特征选择算法，以基分类器的精度作为评价标准，筛选出了最佳的特征子集，并将该算法与 BP 神经网络相结合建立网络入侵检测模型，有效的提高了检测精度并

降低了误警率。

嵌入法将特征选择与模型训练结合在一起，在训练模型的过程中就可以得到各个特征的重要性，其中最典型的嵌入法是决策树算法。决策树学习本质上属于递归选优，经过不断地筛选最终留下对训练数据有着足够分类能力的特征，这一过程本身就在进行特征选择。周晓敏<sup>[21]</sup>等利用梯度提升决策树算法进行特征选择，筛选出影响振动的关键性因素，降低了模型的复杂度。在基金交易场景中，董纪阳<sup>[22]</sup>使用决策树算法对客户交易行为进行特征选择，可以较为准确的筛选出对目标信息相关性较强的特征，有效避免了特征提取过程的主观性，大幅提升了召回率。但是这种方法只适用于模型的局部选择，并且容易产生过拟合现象。

过滤法是按照特征的发散性或相关性对每一个特征进行评价，选出评价最高的特征，并按其权重进行排序。该过程与后续的模型训练相互独立，因此它与其他分类方法结合使用时具有较高的灵活性，分类时间较短。最为常用的过滤法是 Relief 算法和主成分分析（PCA），曹渝昆<sup>[5]</sup>等采取 Relief 特征选择的方法，筛选出与叶片结冰相关性较强的特征，以降低数据的维度，该算法的关键是如何确定相关统计量。傅晓骏<sup>[23]</sup>等采用主成分分析技术进行特征选择，分析样本数据在主成分和非主成分方向的投影，可有效的构造出敏感特征。Relief 算法与主成分分析在降维方面有着较大的相似性，但 Relief 特征选取是通过特征权重的大小进行排序，不改变特征的形式，而主成分分析是通过奇异值的大小进行排序，改变了原有特征的形式。张林林<sup>[24]</sup>等引入了递归特征消除法对原始数据进行特征选择，并采用交叉验证法优化其网络结构，为验证有无递归特征消除法对后期模型训练的影响，利用原始数据集训练模型并将两者的测试结果进行对比，发现加入特征分析后的模型准确率提高到 97%，而未进行特征分析的模型准确率仅为 91%，结果表明特征分析可提高模型的准确率，改善模型的性能。王彬容<sup>[25]</sup>等采用了 3 种特征选择算法对原始数据进行排序选择，并对 3 种算法的分类准确率进行比较，得知交叉验证的递归特征消除法（RFECV）筛选出的特征子集分类准确率优于基于最大互信息系数法（MIC）和基于最大相关最小冗余特征筛选方法（mRmR），并且在选取 15 维特征的情况下，RFECV 算

法的准确率为 98.28%，既达到了降维的目的，又提高了分类准确率。

### 3 风机叶片的故障诊断模型

风机叶片的故障诊断实质上是一个故障的分类过程，其原理是通过对已知数据的学习得到相关分类准则，然后以这一准则来判断其他未知数据类别的方法。近年来常用的分类方法有支持向量机（SVM）、人工神经网络（ANN）、决策树和混合分类等。

支持向量机（SVM）是基于特征空间中的最大分类间隔求解现有样本下最优解的分类算法。由于在分类过程中不需要大量的样本数据，所以该分类方法在解决非线性、小样本问题上具有较大的优势，但当样本数据量过大或者维度过高时，需要涉及到高维度矩阵运算，增加了运算复杂程度的同时模型训练时间也会随之正常，所以支持向量机不适用于样本数据量较大的分类问题。郑楠<sup>[26]</sup>等采用支持向量机数据描述（SVDD）的方法建立故障诊断模型，通过对正常样本数据的训练得到超球体，超球体可以尽可能的将待测数据中的正常数据聚集在一起，远离其他故障数据，进而实现故障诊断。王宇鹏<sup>[27]</sup>等采用动态柯西蜂群算法对样本数据特征进行寻优，并将其作为支持向量机模型中的输入，实验证明改进后的支持向量机模型可有效提高故障诊断的准确率。Xu<sup>[28]</sup>等利用粒子群优化算法对支持向量机进行优化建立了风机叶片故障诊断模型，同时傅晓骏<sup>[23]</sup>也采用了该方法对风机叶片结冰进行检测，均得到了良好的检测精度。

人工神经网络（ANN）可以通过人工神经元的连接单元或节点构成不同的网络结构，建立模型时根据样本数据的模式选取对应的网络结构，进而实现特征表达。根据网络结构的不同将其分为前馈神经网络、反馈神经网络和深度神经网络。此外，人工神经网络在处理复杂问题时需要学习大量的特征参数，导致其训练时间明显高于其他分类方法，并且在处理数据量较少的样本数据时容易发生拟合，泛化能力较差，但神经网络算法的检测精度与其他分类方法相比误差较小，这便是其在工程实践中应用较为广泛的原因。李辉<sup>[29]</sup>等利用极限学习机（ELM）模型对风电机组齿轮箱进行故障诊断，有效的解决了诊断过程中的模态混叠现象，经验证其故障诊断率达到了 100%。李飞羽<sup>[30]</sup>等构建了

RBF 神经网络模型, 通过叶片固有频率的变化对叶片覆冰位置和质量进行预测, 结果表明误差随着覆冰质量的增加而降低。Chen 等<sup>[31]</sup>利用改进的 BP 自聚类+ELM 的混合预测模型对风机叶片的结冰故障进行了预测, 结果表明该算法相比于单一的 ELM 模型预测误差更小, 模型运行时间更短。在数据采集条件有限的情况下, 姜娜<sup>[32]</sup>等提出了一种时序上采样卷积神经网络模型, 利用重构原理将向量型数据转化为二维网格型数据, 结果表明卷积神经网络的误差很小且具有一定的有效性和稳定性。Yang 等<sup>[33]</sup>提出了一种基于 SCADA 数据的深度全连接神经网络模型用于风机叶片结冰状态的诊断, 与传统的 BP 神经网络模型相比, 该算法模型具有更高的检测精度。Wang<sup>[34]</sup>等提出了一种基于深度学习神经网络的故障诊断模型, 其中包含了特征学习和度量学习两个模块, 由于度量模块强大的学习能力, 引入了标签平滑技术缓解过拟合现象, 进一步提高分类性能。李大中<sup>[35]</sup>等提出了一种使用深度学习算法进行优化的深度全连接神经网络模型, 用于风机叶片结冰预测。最后与最近邻法 (KNN)、支持向量机 (SVM)、未使用深度学习优化算法的 BP 神经网络的预测结果进行对比, 结果表明所提出的预测算法模型预测误差和计算量均小于另外 3 种模型。

决策树是建立随机森林和 XGBoost 模型的基分类器, 对已知样本数据具有良好的分类能力, 但难以对未知数据进行准确分类, 并且决策树的生成只适用于模型局部的选择, 应用场景有限。随机森林算法使用 Bagging 方法将多个决策树组合在一起, 并在样本数据预处理阶段引入了随机性, 其中各决策树之间是相互独立的, 极大程度上改善了其泛化能力。极端梯度提升算法 (XGBoost) 是 Boosting 算法的改进算法, 它采用梯度下降的方式来生成决策树, 每一步的迭代都是朝着最小化给定目标函数的方向, 提高运算速度的同时也提高了精度, 避免过拟合现象。该算法是目前使用较多的一种提升算法。梁川<sup>[36]</sup>提出了一种基于随机森林算法的风机叶片故障诊断模型, 用于处理大样本、多维的 SCADA 数据, 结果表明该模型具有较快的检测速度和较高的检测精度。叶春霖<sup>[7]</sup>等采用了 K 近邻 (KNN)、朴素贝叶斯 (NB)、随机森林 (RF) 以及 AdaBoost4 种分类算法建立风机叶片结冰故障诊断模型, 经过对比发现基于随机森林算法的故障诊断模型具有更好的诊断性能和泛化能力。山东科技大学的孙英倩<sup>[37]</sup>选取了



样本数据中的 7 个特征向量作为极端梯度提升算法 (XGBoost) 模型的输入, 经过对模型的训练和评估, 证明该模型可以高效准确的预测风机叶片早期结冰状态。Guo<sup>[6]</sup> 等构造了一个极端梯度提升算法 (XGBoost) 模型, 根据转子转速、输出功率和环境温度三种变量确定叶片是否发生结冰现象, 实例分析证明, 该方法可以提前 5h 触发吸冰警报, 为保证风机的稳定运行提供了充足的时间。

当使用单一的算法模型进行故障诊断时, 很难保证检测的高精度, 此时将两种或两种以上的模型组合实行混合诊断可以较大程度上提高检测精度。近年来混合模型在风机叶片故障检测行业中广泛应用。刘萍兰<sup>[38]</sup> 等将 CART、RF 和 MLP 三种单一模型组合构建了一个 Stacking 模型对风机叶片结冰进行预测, 经综合对比, 该模型的预测性能要优于单一基分类器。熊冒全<sup>[39]</sup> 等提出了一种基于 Bi-LSTM 模型, 其预测输出值再输入到 SVM 模型中进行训练, 进而对风机叶片的覆冰状态实时预测。通过实验证明, 该预测方法的准确率为 95.0%。董健<sup>[40]</sup> 等在传统 BP 神经网络算法的基础上提出了 BP-adaboost 算法模型, 结果表明该模型构成的强分类器的平均得分为 75.3 分, 比单独 BP 神经网络组成的弱分类器平均得分要高。上海师范大学的张丹峰<sup>[41]</sup> 将 LightGBM、XGBoost、ERT 三个模型进行嵌套融合得到一个混合模型, 用于风机叶片结冰预测, 实证研究表明混合模型具有更好的检测性能, 其中 XGBoost 和 ERT 模型的效果较好, LightGBM 模型在运行速度上具有较大的优势。西南科技大学的雷云峰<sup>[42]</sup> 研究了一种基于 ARIMA 模型、随机森林、XGBoost 和 LSTM 神经网络的混合模型, 其中预测性能最好的模型占据最大的组合权重, 使其在组合中发挥最大作用, 结果证明该混合模型取得了比单一模型更好的结冰预测效果。

表 2 不同故障诊断模型汇总

Table 2 Summary of different fault diagnosis models

故障诊断模型	按不同网络结构分类	研究方法	研究效果
神经网络	前馈神经网络	BP 神经网络、径向基函数神经网络 (RBF)、极限学习机 (ELM) 等	样本数量较少时, 容易导致神经网络过拟合, 无法得到很好的泛化能力。样本数量较大时则需要大量的参数学习, 训练时间较长

续表

故障诊断模型	按不同网络结构分类	研究方法	研究效果
	反馈神经网络	Hopfield 网络、Elman 网络等	具有联想记忆功能，可以用来解决样本数据快速寻优问题
	深度学习神经网络	卷积神经网络 (CNN)、深度全连接神经网络 (FCNN)、堆栈自编码网络 (SDAE) 和长短期记忆神经网络 (LSTM) 等	所需训练数据较多，模型复杂程度较为复杂，但是故障诊断精度和分类准确率较高
决策树	/	随机森林、GBDT、XGBoost 等	各决策树在运行期间可以并行处理，提高了运行速度，改善了决策树的泛化能力
支持向量机	/	标准 SVM、概率 SVM、LSSVM 和多核 SVM 等	在解决非线性、小样本的问题上具有一定的优势，但不适用于大规模数据集的分类问题
混合模型	/	PSO+ANN、CART+RF+MLP、BP-adaboost、Bi-LSTM、LSTM+CNN 等	将多种诊断模型进行组合可以较大程度上提高检测精度和检测效率

## 4 总结与展望

风力发电行业将来的发展会越发迅速，更为先进的故障检测技术陆续提出可以有效减少因机组故障引起的电力损失。风力发电机叶片故障检测研究未来的发展方向主要包括以下几个方面：

(1) 故障检测技术向智能化一体化发展，例如开发故障检测知识图谱，自动匹配识别叶片故障原因以及故障类型，提高诊断效率。

(2) 针对 SCADA 数据类别不平衡的特点，许多专家学者采用 SMOTE 算法进行采样处理，可以极大程度的保留原始数据；综合不同采样方式的优点采用混合采样法，可达到数据平衡的效果。例如 SMOTE-ENN 混合采样法、SMOTE-Tomek links 混合采样法。

(3) 针对 SCADA 数据维度过高的特点，一方面可以结合各学科领域知识分析叶片故障机理，确定与叶片故障相关性较强的关键特征；另一方面可以采用多种特征选择方法相结合的方式降低特征维度和冗余性。

(4) 针对故障诊断模型的选择，一方面由于深度学习神经网络模型相较于

其它检测模型误差较小，可以加大对该模型的研究。此外混合模型可有效减小检测误差，这也是未来重要的研究领域。同时还应加强对新型诊断模型的开发；另一方面模型的好坏取决于模型的参数，因此对模型参数进行调优也可以达到准确的诊断效果。

随着风力发电机功率的不断提高，捕获风能的风叶片越做越大，对叶片的要求也越来越高，发展叶片的故障检测技术对提高叶片的可靠性、安全性和使用寿命具有重要意义。本文首先对风机叶片故障类型和故障成因进行分析，并对近年来 SCADA 数据的预处理技术和风机叶片故障诊断模型进行了研究，经过综合对比，提出了一种新型风机叶片故障检测技术：

(1) 使用 SMOTE-Tomek links 混合采样法解决样本数据类别不平衡问题，提高分类器的分类性能。同时采用交叉验证的递归特征消除法 (RFECV) 对高维数据进行特征提取，既可达到高维数据降维的目的，又能准确筛选出样本数据中的关键特征。

(2) 随机森林算法在处理大样本、多维数据方面具有较大优势，并且其中的各决策树都是相互独立的，在运行过程中可以并行处理，从而提升运行速度。因此，采用随机森林算法构建的叶片故障诊断模型具有更好的泛化能力。

## 参考文献

- [1] Kalkanis K, Kaminaris S D, Psomopoulos C S, et al. Structural Health Monitoring for the Advanced Maintenance of Wind Turbin: A review [J]. International Journal of Energy and Environment, 2018, 12: 69-79.
- [2] FPG Márquez, AMP Chacón. A review of non-destructive testing on wind turbines blades [J]. Renewable Energy, 2020, 161: 998-1010.
- [3] Du Y, Zhou S, X Jing, et al. Damage detection techniques for wind turbine blades: A review [J]. Mechanical systems and signal processing, 2020 (141).
- [4] 于筱然. 浅析各种无损检测技术的优缺点 [J]. 电子测试, 2019 (19): 86-88.

- [ 5 ] 曹渝昆, 朱萌, 王晓飞. 基于特征选择和 XGBoost 的风机叶片结冰预测 [ J ] . 电气自动化, 2019, 41 ( 3 ) : 31-33+118.
- [ 6 ] Guo Peng, Infield David. Wind turbine blade icing detection with multi-model collaborative monitoring method [ J ] . Renewable Energy, 2021, 179: 1098-1105.
- [ 7 ] 叶春霖, 邱颖宁, 冯延晖. 基于数据挖掘的风电机组叶片结冰故障诊断 [ J ] . 噪声与振动控制, 2018, 38 ( S2 ) : 643-647.
- [ 8 ] Wei Q, Lu D. A Survey on Wind Turbine Condition Monitoring and Fault Diagnosis—Part I: Components and Subsystems [ J ] . IEEE Transactions on Industrial Electronics, 2015, 62 ( 10 ) : 6546-6557.
- [ 9 ] Beganovic N, Sffker D. Structural health management utilization for lifetime prognosis and advanced control strategy deployment of wind turbines: An overview and outlook concerning actual methods, tools, and obtained results—ScienceDirect [ J ] . Renewable and Sustainable Energy Reviews, 2016, 64: 68-83.
- [ 10 ] Wei Q, Lu D. A Survey on Wind Turbine Condition Monitoring and Fault Diagnosis—Part II: Signals and Signal Processing Methods [ J ] . IEEE Transactions on Industrial Electronics, 2015, 62 ( 10 ) : 1-1.
- [ 11 ] 王峥. 大型风机叶片结构损伤诊断方法研究 [ D ] . 兰州交通大学, 2015.
- [ 12 ] Chou J S, Chiu C K, Huang I K, et al. Failure analysis of wind turbine blade under critical wind loads [ J ] . Engineering Failure Analysis, 2013, 27.
- [ 13 ] 徐宏, 胡广顺. 电动变桨距风机飞车的原因分析及预防 [ J ] . 自动化应用, 2012 ( 1 ) : 65-66.
- [ 14 ] 李刘杰. 基于过采样与集成学习的不平衡数据分类方法研究 [ D ] . 重庆: 重庆邮电大学, 2019.
- [ 15 ] 石洪波, 陈雨文, 陈鑫. SMOTE 过采样及其改进算法研究综述 [ J ] . 智能系统学报, 2019, 14 ( 6 ) : 1073-1083.
- [ 16 ] 何云斌, 冷欣, 万静. 不平衡数据加权边界点集成欠采样方法 [ J ] . 西

- 安电子科技大学学报: 1-8
- [17] 孟东霞, 李玉鑑. 基于特征边界欠采样的不平衡数据处理方法 [J]. 统计与决策, 2021, 37 (11): 30-33.
- [18] 陈鑫. 基于数据分布的不平衡数据混合采样研究 [D]. 山西财经大学, 2021.
- [19] 李大中, 王超, 李颖宇. 基于 XGBoost 算法的风机叶片结冰状态评测 [J]. 电力科学与工程, 2019, 35 (9): 43-48.
- [20] 陈高升. 基于机器学习的网络入侵检测方法研究 [D]. 重庆邮电大学, 2020.
- [21] 周晓敏, 郝勇凯, 丛文韬, 等. 基于梯度提升决策树模型的冷连轧机颤振研究 [J]. 振动与冲击, 2021, 40 (13): 154-158.
- [22] 董纪阳. 基于决策树自动化特征选择的基金客户流失预测研究——后疫情时代下的思考 [J]. 山东社会科学, 2020 (9): 74-80.
- [23] 傅晓骏, 孙海龙, 章珈宁. 基于 PSO-SVM 的风机叶片结冰检测方法 [J]. 科技风, 2021 (2): 189-190.
- [24] 张林林, 胡熊伟, 李鹏, 等. 基于极限学习机的电力系统暂态稳定评估方法 [J]. 上海交通大学学报, 2019, 53 (6): 749-756.
- [25] 王彬蓉, 王维博, 周超, 等. 基于 EMD 自适应重构的心音信号特征筛选及分类 [J]. 航天医学与医学工程, 2020, 33 (6): 533-541.
- [26] 郑楠, 于广亮, 刘娟楠, 等. 一种基于 SVDD 和 PCA 的风机故障诊断方法 [J]. 自动化技术与应用, 2019, 38 (11): 143-146.
- [27] 王宇鹏, 王致杰, 刘琦, 等. 基于动态柯西蜂群算法优化支持向量机的风机叶片故障诊断 [J]. 电气工程学报, 2018, 13 (1): 16-22.
- [28] Jiaohui Xu, Wen Tan, Tingshun Li. Predicting fan blade icing by using particle swarm optimization and support vector machine algorithm [J]. Computers and Electrical Engineering, 2020, 87: 106751.
- [29] 李辉, 邓奇. 基于改进 LMD 方法的风电机组齿轮箱故障诊断研究 [J]. 自动化仪表, 2021, 42 (3): 60-65.

- [ 30 ] 李飞宇, 崔红梅, 苏宏杰, 等. 基于神经网络的风力机叶片覆冰预测方法 [ J ] . 水电能源科学, 2021, 39 ( 6 ) : 171-174.
- [ 31 ] Cheng Peng, et al. Icing Prediction of Fan Blade based on a Hybrid Model [ J ] . International Journal of Performability Engineering, 2019, 15 ( 11 ) : 2882-2890.
- [ 32 ] 姜娜, 严蜜, 李柠. 基于时序上采样卷积神经网络的风机叶片结冰检测 [ J ] . 控制与决策, 1-9.
- [ 33 ] Xiyun Yang, et al. Diagnosis of Blade Icing Using Multiple Intelligent Algorithms [ J ] . Energies, 2020, 13 ( 11 ) .
- [ 34 ] Wang Cunjun, Xu Zili. An intelligent fault diagnosis model based on deep neural network for few-shot fault diagnosis [ J ] . Neurocomputing, 2021, 456: 550-562.
- [ 35 ] 李大中, 刘家瑞, 张华英. 基于深度全连接神经网络的风机叶片结冰预测方法 [ J ] . 电力科学与工程, 2019, 35 ( 4 ) : 39-44.
- [ 36 ] 梁川. 基于特征提取和随机森林的风机故障诊断 [ J ] . 科学技术创新, 2020 ( 26 ) : 55-58.
- [ 37 ] 孙英倩. 基于数据驱动的风机叶片结冰预测研究 [ D ] . 山东科技大学, 2020.
- [ 38 ] 刘萍兰, 汤占军, 蒋鹏程, 等. 基于 Stacking 模型的风力机叶片结冰预测研究 [ J ] . 数据通信, 2021 ( 3 ) : 48-53.
- [ 39 ] 熊昌全, 何泽其, 张宇宁, 等. 基于 Bi-LSTM 和支持向量机的风机叶片短期覆冰状态预测模型 [ J ] . 四川电力技术, 2021, 44 ( 3 ) : 88-94.
- [ 40 ] 董健, 柳亦兵, 滕伟, 等. 基于 BP \_ Adaboost 算法的风电机组叶片结冰检测 [ J ] . 可再生能源, 2021, 39 ( 5 ) : 632-636.
- [ 41 ] 张丹峰. 基于 LightGBM, XGBoost, ERT 混合模型的风机叶片结冰预测研究 [ D ] . 上海师范大学, 2018.
- [ 42 ] 雷云峰. 基于数据挖掘的监测设备故障诊断及预测研究 [ D ] . 西南科技大学, 2021.

## Research and Prospect of Fan Blade Fault Detection Technology based on SCADA Data

Jiang Zhiwei Li Yongjian

*Hunan, Institute of Engineering, Xiangtan*

**Abstract:** The fan blade fault detection technology based on SCADA data is taken as the research object. Firstly, the fault types and causes of fan blades are analyzed, and the pre-processing technology of SCADA system monitoring data and the establishment of blade fault diagnosis model in the past ten years are studied. After comprehensive comparison, In this paper, a cross-validated recursive feature elimination (RFECV) method and a fault diagnosis model based on random forest algorithm are proposed to deal with more blade faults. Finally, the future development direction of fan blade fault detection technology is summarized and suggested.

**Key words:** Wind Turbine Blade; Pretreatment technology; Feature selection; Model; Fault diagnosis