

海上大兆瓦风电机组故障预测与识别

张智伟¹, 王 靖¹, 黄煜明², 郑俊杰³

(1. 上海海湾新能风力发电有限公司, 上海 200433; 2. 江苏金风科技有限公司, 江苏 盐城 224003;
3. 北京金风科创风电设备有限公司, 北京 100176)

摘要:随着双碳战略的持续推进,可再生能源发电的重要性越来越被重视。海上风电机组位于偏僻地区,故障后维修难度较大,因此风电机组故障预测与识别技术研究至关重要。基于 Python sklearn 的机器学习框架和基于 TensorFlow 的故障预警方法,采用多种特征提取方式,并使用卷积神经网络进行特征融合和分类,可实现海上风电系统的故障诊断和预警。详细介绍该方法的设计思路、实验步骤和实验结果,并对该方法进行评估和分析。

关键词:机器学习;TensorFlow;卷积神经网络;风电机组故障

中图分类号:TP277 **文献标志码:**A **文章编号:**1671-1807(2023)17-0273-06

随着“双碳”目标战略的加速推进,可再生能源对于优化我国能源结构、保障能源安全、构建新型能源体系、助力实现“双碳”战略目标具有重要作用。风力发电作为重要的可再生能源之一,已经成为世界各国推广的重点。而海上风电机组具有风能资源丰富、单机装机容量大、利用率高等优势,已经成为风力发电的重要发展方向。然而,由于海上风电机组地理位置偏远、环境恶劣,因此故障率较高,维护难度也较大^[1-2]。因此,对于海上风电机组的故障预测和识别技术的研究具有重要的现实意义^[3-4]。

变流器是风力发电机组中最关键的部件,其主要作用是将风力发电机产生的交流电转换为高质量的交流电,以便通过电网进行输送和利用。变流器可以控制风力发电机的转速、电压、频率等参数,从而实现最佳的发电效率和稳定性^[5]。而变流器水冷系统则是为了保证变流器的稳定性和可靠性而设计的。由于变流器在工作时会产生大量的热量,如果不能及时散热,就会导致变流器温度过高而损坏。因此,变流器水冷系统可以通过水循环来吸收热量,并将其散发到外部环境中,从而保持变流器的正常工作温度,提高其工作效率和寿命^[6]。每年春夏季节,柳树和杨树的絮屑随气流飘荡后阻塞外部散热器的进风口^[7],给海上风电运维工作带来了严重困难。

夏博等^[2]在风电采集与监视控制系统数据问题上与深度学习方法相结合,并对今后的研究重点进行了展望。程江洲等^[8]提出了一种基于改进贝叶斯网络的风电机组故障诊断与风险预测模型,为风电机组故障诊断和风险预测提供有效依据。沈剑飞和翟相彬^[9]通过风电场选址、风机选择、融资方式选择 3 个手段来降低风电场在全寿命周期下的总成本。朱国栋^[10]为了解决风电故障寿命问题,对风电项目可持续发展能力进行了研究,从风资源、风电消纳、运行状况、财务状况 4 个考虑建立评价指标体系,对故障问题方面进行了阐述分析。苗青等^[11]对变流器结构性能进行对比分析,做出对变流器拓扑结构的合理选择。对风电机电变流器的故障研究方式众多,但对变流器的故障模式缺少更准确的研究。

针对变流器的主要故障模式,本文提出一种基于 Python sklearn 的机器学习框架,用于预测和识别海上大型兆瓦机组的故障。特别是使用支持向量机(support vector machine, SVM)、随机森林(random forest, RF)和深度学习(deep learning, DL)等算法来建立模型,分析和测试关于实际故障发生的数据。实验和结果表明,使用来自 sklearn 的机器学习技术在识别故障方面具有较高的准确率,可以有效地提高海上风电机组故障的预测和识别。

收稿日期:2023-04-06

作者简介:张智伟(1978—),男,上海人,上海海湾新能风力发电有限公司,工程技术部主任,高级工程师,研究方向为新能源发电;王靖(1975—),男,浙江温州人,上海海湾新能风力发电有限公司,生产运行部副主任,工程师,研究方向为风电运维管理;黄煜明(1994—),男,广东中山人,江苏金风科技有限公司,工程师,研究方向为新能源发电;郑俊杰(1986—),男,陕西西安人,北京金风科创风电设备有限公司,海工电气负责人,工程师,研究方向为风电运维技术。

风电大数据故障预警模型上线流程如图 1 所示。本文中着重对特征标准化、特征提取、数据转换、模型训练和模型评估 5 个方面进行详细阐述^[12]。

本文中提出的风力发电机组大数据故障预警模型是针对华东地区海上某风场的 6.45 MW 机组变流器主要故障模式进行设计的,监控风机变流器水冷系统的运行情况。为预测变流器水冷系统的温度^[13],筛选出 6 个直接决定和影响变流器温度的相关参数,包括水冷系统出口温度、水冷系统出口水压、水冷系统进口水压、水冷系统进口温度、环境温度和风机功率。

1 特征标准化

特征标准化是数据预处理过程中常用的一种方法^[14]。它可以将数据转化为均值为 0、标准差为 1 的标准正态分布,也可以将数据标准化到 0~1。特征标准化可以使得各个特征的大小范围相同,避免了某些数据的特征对结果的影响较大,从而提高了模型的准确性和预测能力,同时能够减少样本误差,稳定模型的性能。

在 Python 中,可以使用 MinMaxScaler 函数对数据进行标准化。MinMaxScaler 函数是 sklearn.preprocessing 模块中的一个类,可以将数据转化为 0~1 的范围之内。其中的关键步骤是对样本进行计算平均值和标准差的统计,再利用测试样本进行

min-max 缩放。
MinMaxScaler 实现特征标准化采用如下公式:
$$X_std = (X - \min(X)) / (\max(X) - \min(X))$$
$$X_scaled = X_std * (\max - \min) + \min$$
式中:axis=0 为计算每一列的最大值和最小值。函数将 X 中的每一列缩放到 0~1,然后进行平移和缩放,使得每一列都符合标准正态分布。

2 特征提取

特征提取是一个旨在识别和提取重要特征的任务,从而降低数据的复杂度,以便更容易进行模式识别和分类^[15]。在时间序列数据中,特征提取是一个特别重要的任务,因为时间序列数据具有高度相关性和复杂性。因此,通过提取出数据中最相关的、最有用的特征,可以帮助更好地理解 and 处理数据。

使用时间序列数据进行特征提取,并通过小波包分析和卷积神经网络(convolutional neural network,CNN)进行特征融合和分类。在特征提取过程中,将时间序列数据切分成固定长度的时间序列样本,可以减少数据噪声和冗余,更好地保留数据的有用信息;通过小波包分析,则可以更好地处理信号的这些高频成分;CNN 则可以从中提取出更高级、更抽象的特征。在将小波包分析和 CNN 结合起来时,小波包分析通常用于分解各个频率下的信

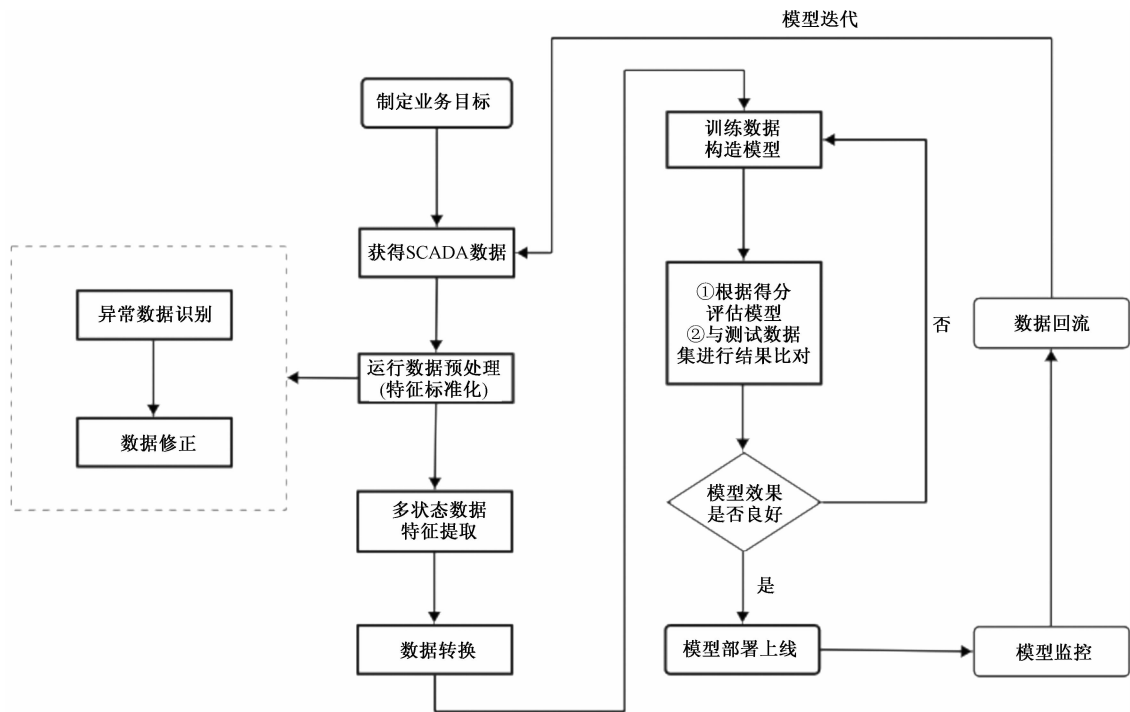


图 1 风电大数据预警模型上线流程

号,CNN 则用于处理来自各个分量的图像特征,从而实现特征融合。这样可以将时间序列问题转化为监督学习问题,从而可以使用监督学习算法来训练和测试模型。

综上所述,特征提取的原理是基于对原始数据进行转换和处理,提取出数据潜在的有用特征,从而更好地反映数据的本质特征。在时间序列数据的处理中,小波包分析和卷积神经网络联用是一种有效的方式,可以运用不同的技术取长补短,提取更为有效的数据特征,并加以融合,从而提高数据分类和预测的准确性,提高数据处理和监督学习的效率和精度。

3 数据转换

数据转换的作用是将原始数据转化为适用于机器学习和深度学习算法处理的有监督学习问题,使得机器可以从数据中学习规律和模式,并用于未来的预测和决策。特别地,对于时间序列数据,数据转换可以将观测值按照一定的时间范围进行滚动,以创建时间窗口,并将该时间窗口中的数据作为输入特征,窗口最后一个时间点的数据作为输出标签^[16]。通过这种方式,可以获得单位时间内的数据变化规律,使得机器可以对未来单位时间内的数据进行预测,而且可以帮助消除原始数据的噪声、消除数据中的周期性波动以及减少渐进性漂移等问题。特别是应用在风电机组数据的这种复杂时间序列数据中,数据转换还可以改进预测的准确性。

数据转换的原理是基于滚动时间窗口将时间序列数据转化为监督学习问题,具体步骤如下:

1)通过观察时序数据的特点来确定合适的窗口长度,在本例中,选择 30 min 和 60 min 作为窗口长度,以便学习时间尺度内的数据模式。

2)按照窗口长度滚动数据,并将窗口内的数据作为输入特征,窗口末尾时刻的数据作为输出标签。这样,就可以创建有标签的数据点,用于监督学习算法的训练。

3)基于这些有标签的数据点,训练预测模型,通过对未来时间点的输入特征进行预测,以预测未来的标签。

4 模型训练

LightGBM 是用于处理具有复杂特征和大量数据的问题,如图像识别、自然语言处理、预测和推荐等^[17-18]。LightGBM 使用了一种称为 GOSS(gradient-based one-side sampling)技术来加快训练速度,该技术可以通过去除梯度小的样本来减少数据数

量^[19-20],从而提高模型的效率。此外,LightGBM 通过使用直方图来对特征进行离散化,并且可以快速找到最佳划分点,从而提高特征筛选的效率。在训练时,LightGBM 使用了基于 leaf 的决策树来进行优化,即只考虑每个树叶节点的信息,而不是考虑整个树。

在风电数据领域,LightGBM 相比其他机器学习算法(如决策树、随机森林、神经网络等)有以下几个优势:

1)高效性。风电数据通常属于高维稀疏数据,包含大量的定量和定性特征,对于这种数据而言,LightGBM 在处理大规模数据时比其他算法更高效,由于使用了 GOSS,可以忽略那些梯度比较小的样本点,减少训练集的样本数。因此,它对于计算资源的要求更小,可以更快地训练模型。

2)准确性。由于使用切分点的直方图算法,LightGBM 能更准确地寻找最优划分点,同时使用 leaf 优化提高了模型的精度。

3)鲁棒性。LightGBM 能够在不平衡数据集、高维度数据集和缺失数据集上表现出更好的鲁棒性,具有更高的容错性和泛化性能。

4)可解释性。LightGBM 拥有容易理解的模型结构,输出的树形结构是可解释的,有助于深入理解变量和特征之间的关系,从而更好地进行数据分析和建模。

将时间序列数据转换为适合机器学习算法处理的数据格式,创建适合 LightGBM 模型处理的数据集,并使用 LightGBM 模型对数据进行训练和预测。

5 模型测试与评估

模型评估是使用 sklearn.metrics 库中的函数,用于评估机器学习模型的性能,其中包括以下几个指标^[21-22]:

1)均方误差(mean squared error, MSE)。MSE 表示预测值与真实值之间差值的平方和除以样本个数,是回归问题中最常用的评估指标。它的计算公式为

$$MSE = 1/n \sum (y_{pred} - y_{true})^2 \quad (3)$$

式中: n 为样本个数; y_{pred} 和 y_{true} 分别为模型的预测值和真实值。

2)平均绝对误差(mean absolute error, MAE)。MAE 表示预测值与真实值之间差值的绝对值之和除以样本个数,也是回归问题中常用的评估指标。它的计算公式为

$$MAE=1/n\sum |y_{pred}-y_{true}| \tag{4}$$

3)平均绝对百分误差(mean absolute percentage error, MAPE)。MAPE 表示预测值与真实值之间差值的绝对值除以真实值再乘以 100%的平均数。它对于处理数据单位不同的数据集时非常有用,也是回归问题中常用的指标之一。它的计算公式为

$$MAPE=1/n\sum (|y_{pred}-y_{true}|/y_{true})\times 100\% \tag{5}$$

4) R^2 得分(R squared score)。 R^2 得分表示模型的解释能力,即模型对因变量的变异能够解释多少。它的值为 0~1,值越接近 1 表示模型的解释能力越好。它的计算公式为

$$R^2=1-\sum (y_{pred}-y_{true})^2/\sum (y_{true}-y_{avg})^2 \tag{6}$$

式中: y_{avg} 表示真实值的平均数。

这些指标的作用在于对机器学习模型的性能进行评估,以确定模型预测准确性的好坏。在选择模型和调整模型参数时,可以根据不同的指标进行评估和比较,从而选择最合适的模型和参数组合。

在进行模型测试过程中,首先需要对模型进行训练,然后利用测试集对模型的性能进行评估,根据评估结果进行优化调整,得到最终的表现最佳的模型。

在本文中,采用了 LightGBM 模型进行训练和测试,并按照如下步骤进行调整优化^[23]:

1)采用默认参数进行训练,得到基准性能。在测试集上,得到 R^2 得分为 0.955,MAE 为 0.291。

2)将 learning rate 从默认值 0.1 调整为 0.01,并将 max_depth 从默认值 -1 调整为 5。在测试集上,得到 R^2 得分为 0.972,MAE 为 0.222,即已经比前一步有一定的提高。

3)将 min_child_samples 从默认值 20 调整为 3,并进一步调整 learning rate 和 max_depth。在测试集上,得到 R^2 得分为 0.978,MAE 为 0.184,此时模型的表现已经达到了最佳状态,可以作为最终模型进行应用。

6 模型结果展示

机组变流器主要故障模式进行设计测试集采用的数据是华东地区海上某风场的 6S 机组,单台容量 6.45 MW。图 2 为海上风力发电机组外观全貌。变流器为 2 套 3 MW 机组功率并联的结构,协助变流器 IGBT 模块散热的方法为水冷方式,由冷却液流经变流器带走热量,冷却液由循环泵的



图 2 海上风力发电机组外观图

进口经室外空气散热器与冷空气进行热交换,散热后再进入变流器。其中冷却液的正常温度范围一般处于 35~42 ℃。

对机组变流器温度进行采集,并采用本文的预测方法对变流器温度进行预测,预测结果如图 3 所示。

图 3 中,蓝色数据为机组实际表现值;桔红色数据为模型预测值;纵坐标为水冷系统出口温度,℃;横坐标为数据标签点,相邻标签点之间的时间间隔为 6 s。

为了评估模型的性能,抽取了 8 段时间的实际温度数据进行对比分析。

实际温度均值为 37.836、38.154、38.378、38.812、38.761、38.392、39.443、37.496 ℃。

预测温度均值为 38.103、38.019、38.701、38.932、38.848、38.635、39.254、37.592 ℃。

计算得出使用测试集测试模型的 MAE 为 0.183,MSE=0.039 97。

然后计算模型的精确率、召回率,得出模型在测试集上的精确率为 0.986,表示在所有被预测为高温(即温度达到故障门限)的样本中,有 98.6%的样本预测正确。召回率为 0.984 表示在所有实际高温样本中,有 98.4%的样本被正确地预测为高温。

评估结果显示,变流器水冷系统温度预测模型在预测高温事件方面具有高度准确性和可靠性。经过在测试集上评估该 LightGBM 模型的性能,发现该模型在预测目标值时表现出高准确性和稳定性,适用于生成高质量的预测结果,并已成功将预警模型软件挂载到大数据预警平台上,目前已应用

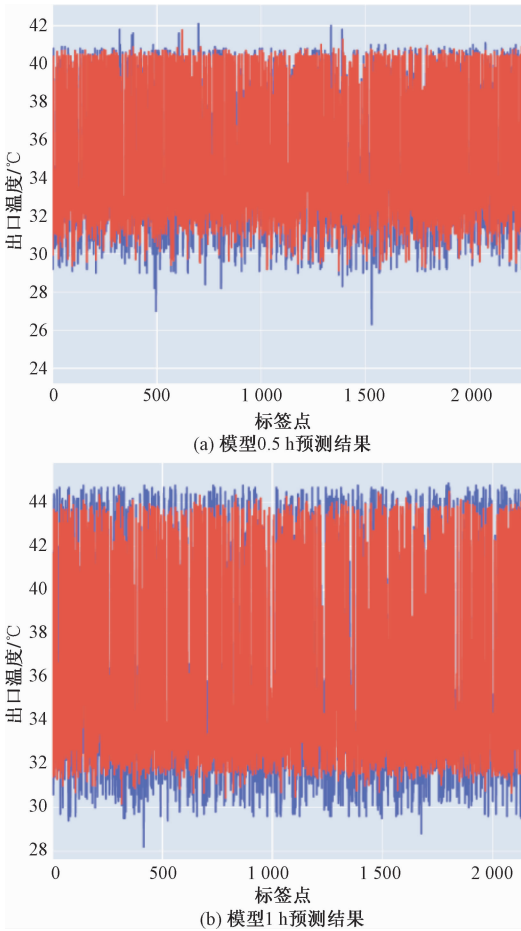


图3 预测结果

到数据预警和监控工作中。

此外,基于时间序列数据的特点,随着预测时间的增加,预测准确性会降低。因此,提出一种方法来实现预测模型的本地化效果。具体做法是,在使用 Python 运行模型后,利用 joblib 库中的 dump 函数将 LightGBM 模型对象保存到本地。随后,通过风机主控使用 joblib 库中的 load 函数重新加载该模型对象,从而实现预测模型的本地化。目前,预警模型已经挂载到主控程序算法的项目中,并正在控制系统部门进行稳定性测试。

7 结语

本文的基于有监督学习的风电机组故障预测模型,为风电行业提供了一种新的思路和工具。未来,该模型还可以进一步发展和优化,应用于风电机组的实时监测、故障诊断与预警等多个方面,促进新能源产业的发展和应用。随着信息技术的不断进步,相信不久将会看到风电机组故障预测模型呈现轻量化和实时化的发展趋势。这种趋势会使风力发电机组提供更多优质、高效的绿色能源。

参考文献

[1] 陈雪峰,郭艳婕,许才彬,等. 风电装备故障诊断与健康监测研究综述[J]. 中国机械工程,2020,31(2):175-189.

[2] 夏博,李春杨,万露露,等. 基于深度学习的风力发电机组故障预警方法研究综述[J]. 科学技术与工程,2023,23(9):3577-3587.

[3] 金晓航,许壮伟,孙毅,等. 基于 SCADA 数据分析和稀自编码神经网络的风电机组在线运行状态监测[J]. 太阳能学报,2021,42(6):321-328.

[4] 张帆,刘德顺,戴巨川,等. 一种基于 SCADA 参数关系的风电机组运行状态识别方法[J]. 机械工程学报,2019,55(4):1-9.

[5] 黄科元,刘静佳,黄守道,等. 永磁直驱系统变流器开路故障诊断方法[J]. 电工技术学报,2015,30(16):129-136.

[6] 董礼,许伟. 风电变流器预防性维护检测技术应用[J]. 电气技术,2020,21(9):94-98.

[7] 杨培文,熊国专,陈铁,等. 风电机组变流器故障缺陷析研[J]. 能源与环境,2022(5):47-49.

[8] 程江洲,冯馨以,冯梦婷,等. 计及气象因素的风电机组故障诊断与风险预测[J]. 科学技术与工程,2022,22(22):9645-9651.

[9] 沈剑飞,翟彬彬. 基于全寿命周期下风电场的成本控制[J]. 科技和产业,2013,13(10):92-96.

[10] 朱国栋. 风电项目可持续发展能力评价研究[J]. 科技和产业,2017,17(10):109-112.

[11] 苗青,杨剑锋,谢延凯,等. 分布式网中的变流器拓扑结构性能的对比分析[J]. 科学技术与工程,2014,14(13):184-190.

[12] 曹渝昆,朱萌. 基于主成分分析和 LightGBM 的风电场发电功率超短期预测[J]. 上海电力学院学报,2019,35(6):562-566.

[13] 陈朝春,赵瑞玲,金风 1.5 MW 风电机组水冷却系统探讨[J]. 电工技术,2019(15):95-96.

[14] 程诗雨,刘航,曾天生,等. 基于 SCADA 参数关系的风电机组部件重要度分析[J]. 可再生能源,2021,39(10):1335-1341.

[15] 靳志杰,霍志红,许昌,等. 基于特征选择和 XGBoost 的风电机组故障诊断[J]. 可再生能源,2021,39(3):353-358.

[16] 俞国燕,李少伟,董晔弘. 基于 XGBoost-LightGBM-LSTM 的风机齿轮箱轴承故障预警[J]. 轴承,2022(6):140-145.

[17] 周挺,杨军,周强明,等. 基于改进 LightGBM 的电力系统暂态稳定评估方法[J]. 电网技术,2019,43(6):1931-1940.

[18] 王愈轩,刘尔佳,黄永章. 基于 CNN-LSTM-lightGBM 组合的超短期风电功率预测方法[J]. 科学技术与工程,2022,22(36):16067-16074.

[19] 张博,向旭,贾俊龙,等. 基于 LightGBM 的天然气管道周围滑坡灾害预测方法[J]. 吉林大学学报(理学版),2023,61(2):338-346.

[20] 张航,史兆培,束垠,等. 基于 OOB-BO-LightGBM 的风

电机组故障诊断方法[J/OL]. 中国测试: 1-7[2023-06-30]. <http://kns.cnki.net/kcms/detail/51.1714.TB.20220623.1608.040.html>.

[21] 邢长征,徐佳玉. LightGBM 混合模型在乳腺癌诊断中的应用[J/OL]. 计算机工程与应用: 1-10[2023-06-30]. <http://kns.cnki.net/kcms/detail/11.2127.TP.20230228.1116.020.html>.

[22] 肖鹏飞. 基于 LightGBM 与 Catboost 融合的二手车估价模型研究[D]. 武汉: 中南财经政法大学, 2021.

[23] 刘付,秦华妮,赖惠慧. 基于 SMOTE 和贝叶斯优化的 Adj-LightGBM 人岗匹配算法[J]. 计算机与现代化, 2023(3): 90-95.

Fault Prediction and Identification of Wind Turbine Based on Offshore Megawatt Units

ZHANG Zhiwei¹, WANG Jing¹, HUANG Yuming², ZHENG Junjie³

(1. Shanghai Haiguan New Energy Wind Power Generation Co.,Ltd.,Shanghai 200433,China;
2. Jiangsu Jinfeng Technology Co.,Ltd.,Yancheng 224003,Jiangsu,China; 3. Beijing Goldwind
Science & Creation Wind Power Equipment Co.,Ltd.,Beijing 100176,China)

Abstract: With the continuous promotion of the dual-carbon strategy, the importance of renewable energy generation is becoming more and more important. Offshore wind turbines are located in remote areas, and maintenance after failure is more difficult. Therefore, the research on wind turbine fault prediction and identification technology is crucial. A machine learning framework based on Python sklearn and a fault warning method based on TensorFlow, multiple feature extraction methods is adopted and convolutional neural networks is used for feature fusion and classification, which can achieve fault diagnosis and warning for offshore wind power systems. The design ideas, experimental steps, and experimental results of the method was introduced in detail, and this method was evaluated and analyzed.

Keywords: machine learning; TensorFlow; convolutional neural network; wind turbine fault