

DOI: 10.13382/j.jemi.B2407877

基于特征提取与改进 POA 的光伏阵列故障诊断*

韩茂林 杨琛 牛锋杰 周宁 周定璇

(上海海洋大学工程学院 上海 2013061)

摘要: 光伏阵列常处于复杂恶劣的环境中,易发生多种类型和不同程度的故障。为提高光伏阵列在恶劣环境下的故障诊断准确率,提出一种基于特征提取与改进鹈鹕算法(IPOA)优化支持向量机(SVM)的光伏阵列故障诊断模型算法。首先,在MATLAB/Simulink仿真平台对15种典型故障状态进行模拟,构建12维故障特征向量,并采用核主成分分析(KPCA)进行特征融合与提取,以增强特征表达能力;其次,针对传统鹈鹕算法在全局搜索与局部开发中的局限性,引入改进的Tent混沌映射、惯性权重、非线性收敛因子及自适应 t 分布变异策略,较大程度提升算法寻优性能;最后通过IPOA对SVM模型的惩罚因子 C 与核参数 γ 进行优化,建立IPOA-SVM光伏阵列故障诊断模型,并分别通过仿真模拟与实验测试对模型进行验证。结果表明,与传统6维特征量相比,采用所提12维特征量的诊断准确率更高;改进的算法模型基于仿真数据和实验数据的故障诊断分类准确率分别达到98.55%和97.93%,明显优于其他对比算法模型,在光伏阵列故障诊断中具有更高的准确率。

关键词: 光伏阵列;故障诊断;特征提取;支持向量机;改进鹈鹕优化算法

中图分类号: TM615;TN06

文献标识码: A

国家标准学科分类代码: 470.40

Photovoltaic array fault diagnosis based on feature extraction and improved pelican optimization algorithm

Han Maolin Yang Chen Niu Fengjie Zhou Ning Zhou Dingxuan

(School of Engineering Science and Technology, Shanghai Ocean University, Shanghai 2013061, China)

Abstract: Photovoltaic (PV) arrays often operate in complex and harsh environments, making them susceptible to various types and degrees of faults. To enhance the accuracy of fault diagnosis in such challenging conditions, this study proposes a novel fault diagnosis model based on feature extraction and an improved pelican optimization algorithm (IPOA) optimized support vector machine (SVM). Firstly, 15 typical fault states are simulated on the MATLAB/Simulink platform, from which a 12-dimensional fault feature vector is constructed. Kernel principal component analysis (KPCA) is then applied for feature fusion and extraction to improve feature representation capabilities. Secondly, to address the limitations of traditional pelican optimization algorithms in balancing global search and local exploitation, enhancements are introduced, including the Tent chaotic map, inertia weight, nonlinear convergence factors, and an adaptive t -distribution mutation strategy, all of which significantly improve the algorithm's optimization performance. Finally, the IPOA is used to optimize the penalty factor C and kernel parameter γ of the SVM model, establishing the IPOA-SVM PV array fault diagnosis model, which is then validated through both simulation and experimental tests. The results show that, compared to the traditional 6-dimensional feature set, the proposed 12-dimensional feature set achieves higher diagnostic accuracy. The improved model demonstrates fault diagnosis classification accuracies of 98.55% and 97.93% for simulation and experimental data, respectively, significantly outperforming other comparison models and demonstrating higher accuracy in PV array fault diagnosis.

Keywords: photovoltaic array; fault diagnosis; feature extraction; support vector machine; improved pelican optimization algorithm

0 引言

随着全球能源短缺问题的加剧,光伏发电作为可再生清洁能源的重要性日益凸显^[1-2]。然而,光伏阵列长期处于严苛的外部环境中,且在生产、运输、安装等环节中容易受损,导致光伏系统在运行过程中发生故障的概率较高,这不仅影响发电效率,还可能引发大范围设备损坏,进而导致更严重的事故^[3-4]。因此,深入研究光伏阵列故障的诊断技术具有重要的意义。

目前,国内外学者针对光伏阵列故障诊断提出了多种诊断方法^[5-7],主要包括红外图像法、电路结构法以及基于 I - V 曲线的智能检测法^[8-9]。红外图像法通过检测故障区域的温度异常来进行诊断,该方法依赖昂贵的红外成像设备,且只能处理有限的故障类型,同时容易受到环境因素的干扰^[10]。电路结构法通过在光伏阵列的各关键节点安装传感器,监测电压、电流等参数,实现故障检测。然而,此方法要求大量传感器,增加了安装和维护的成本及复杂度^[11]。与上述方法相比,基于 I - V 曲线的智能检测法则通过深入挖掘 I - V 曲线的特征信息,能够较为全面地反映光伏阵列的运行状态,具有较高的诊断准确性,且不需要大量硬件设备,具有更高的经济性和普适性^[12]。文献[13]通过评估多种机器学习算法对光伏阵列故障的诊断效果,为基于 I - V 曲线的故障诊断分类研究提供了经验。文献[14]通过选取开路电压 U_{oc} 、短路电流 I_{sc} 等作为故障特征量,有效地对正常、短路、老化、开路、阴影 5 类故障进行分类。文献[15]利用 U - I 曲线的二阶导数特征,实现了对局部阴影遮挡故障的精确诊断。文献[16-17]则引入填充因子 (fill factor, FF) 作为诊断特征,提出了包含 5 个特征的向量模型,用于区分不同模式的阴影和老化故障。同时,现有研究表明故障诊断与预测性维护的协同应用可显著提升光伏电站的可靠性与经济性。文献[18]强调预测性维护与故障诊断的协同可提升光伏电站安全性;文献[19]提出基于趋势的光伏系统健康管理框架,验证了故障诊断与预测联动的可行性。尽管现有研究在故障类型区分方面取得了一定的进展,但在故障程度的进一步区分,尤其是在不同程度的老化、遮挡故障及其模式区分方面,仍有待进一步的研究和改进。因此,如何在保证高效诊断的同时,增强对故障类型及故障程度的区分,为预测模型提供高精度输入,助力光伏电站从“事后维修”向“事前预防”的运维模式转型,成为光伏阵列故障诊断研究中的关键问题。

综上所述,提出一种基于特征提取与改进鹈鹕算法 (improved pelican optimization algorithm, IPOA) 优化支持向量机 (support vector machine, SVM) 的光伏阵列故障诊断方法,以期提高在多故障模式和不同故障程度下的

诊断准确率。首先,通过分析不同故障类型的输出特性,构建融合耦合因子与动态电学参数的 12 维故障特征量,并利用核主成分分析 (kernel principal components analysis, KPCA) 进行特征降维与噪声抑制,显著提高特征的表达能力。然后,设计融合 Tent 混沌初始化、惯性权重、非线性收敛因子及自适应 t 分布变异的多策略改进传统鹈鹕优化算法,再对 SVM 模型的超参数进行寻优,以进一步提升故障诊断的精确性。最后,通过仿真与实验验证所提方法与诊断模型在光伏阵列故障识别与诊断中的有效性。

1 光伏阵列故障分析与特征提取

1.1 光伏阵列仿真模型

光伏阵列作为光伏发电系统的重要组成部分,容易出现多种故障,影响系统整体性能。为便于对光伏阵列的不同运行状态进行研究分析,利用 Simulink 平台搭建光伏阵列的仿真模型,选用光伏组件在辐照度为 $1\,000\text{ W/m}^2$ 、温度为 $25\text{ }^{\circ}\text{C}$ 的标准测试条件 (standard test condition, STC) 下的参数如表 1 所示。

表 1 光伏组件参数
Table 1 Parameters of the PV module

参数	数值
开路电压 U_{oc}/V	48.5
短路电流 I_{sc}/A	11.31
最大功率点电压 U_m/V	40.7
最大功率点电流 I_m/A	10.57
最大功率 P_m/W	430

对实际工程中故障频率较高且故障影响较为严重的开路故障 (open circuit fault, OC)、线间短路故障 (line-to-line short circuit fault, LL)、异常老化故障 (abnormal degradation, AD) 和局部阴影遮挡故障 (partial-shading fault, PS) 进行研究,设计模型如图 1 所示。

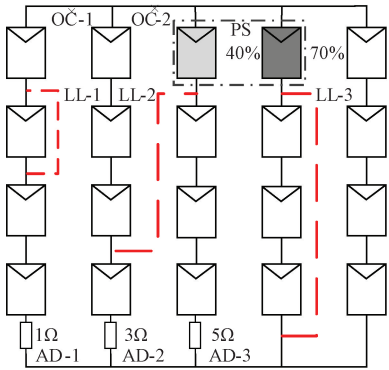


图 1 光伏阵列模型及故障设计

Fig. 1 Photovoltaic array model and fault design

在光伏阵列实际运行过程中,局部阴影遮挡是一种常见且复杂的故障类型,其发生情形多样,对光伏系统的性能影响显著。考虑到常出现光伏阵列的某一角或某一边被遮挡情况,且受遮挡的位置和程度出现差异,将遮挡模式细分为 6 种不同的情况,如图 2 所示。其中不同颜色表示遮挡的遮光率,反映不同程度的遮挡。

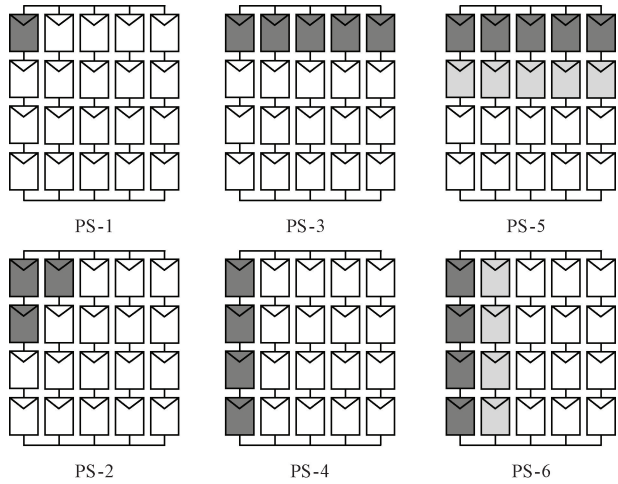


图 2 局部阴影遮挡故障模式

Fig. 2 Local shadow occlusion failure mode

1.2 故障特征选择

光伏阵列由大量光伏组件相互串并联组成,而光伏组件是使用光来产生电流的多参数模型,其输出受到内部参数和外界环境的影响。考虑环境条件变化对光伏阵列输出特性的影响,在不同温度 T_{pv} 和太阳辐照度 G 下,正常运行状态 (normal) 的光伏阵列输出特性曲线如图 3 所示。

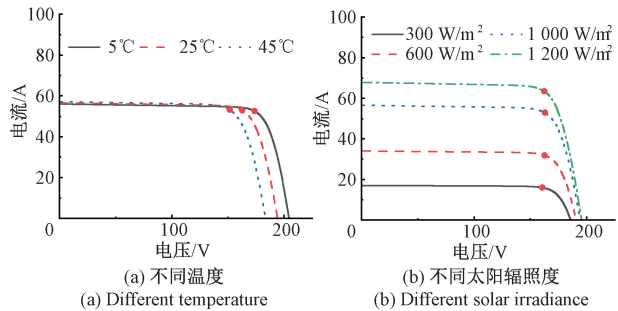


图 3 不同环境条件的 $I-V$ 特性曲线

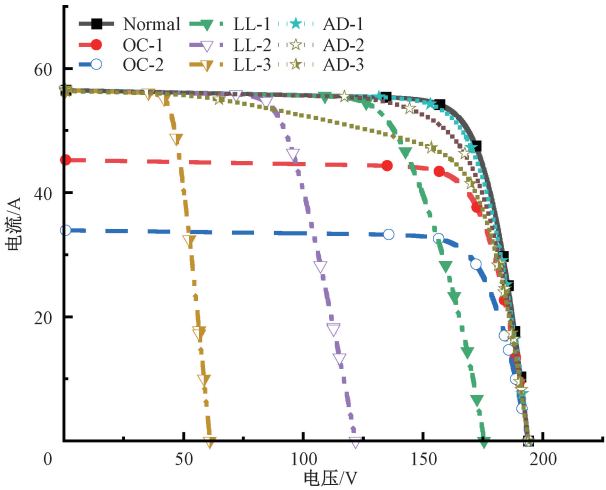
Fig. 3 $I-V$ characteristic curves under different environmental conditions

由图 3(a)可知,随着温度上升,开路电压 U_{oc} 降低,短路电流 I_{sc} 呈上升趋势,最大输出功率 P_m 降低;由图 3(b)可知,随着太阳辐照上升,开路电压 U_{oc} 呈上升趋势,短路电流 I_{sc} 升高,最大输出功率 P_m 升高。可见,光

伏阵列的输出容易受到环境变化的影响。

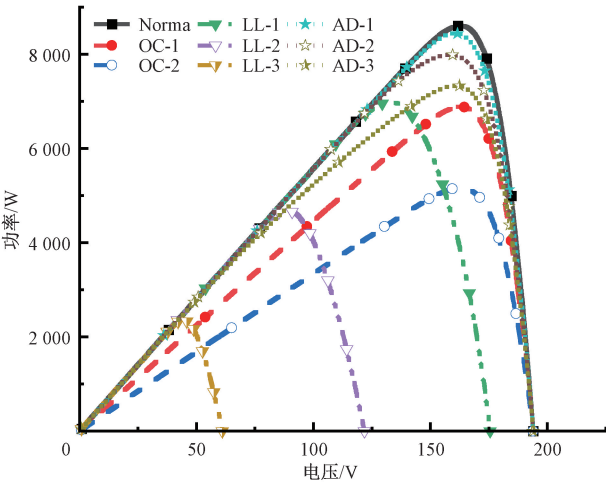
在光伏阵列故障诊断研究中,通常直接采用开路电压 U_{oc} 、短路电流 I_{sc} 、最大功率点电压 U_m 、最大功率点电流 I_m 、最大输出功率 P_m 、 $P-V$ 曲线峰值数 N_m 这 6 种特征量作为模型输入的特征量,记为 $F_1 \sim F_6$ 。正常及开路、线间短路、异常老化、局部阴影遮挡故障在 STC 下的输出特性曲线如图 4 所示。其他环境条件下,输出曲线变化趋势相似。

由图 4 可见,不同程度的老化故障之间仅最大功率存在偏差,AD-1 与 Normal 的输出特性曲线近乎重叠,仅在最大功率点处存在稍许偏差;PS-4、PS-6 不存在“多峰”现象,其输出特性曲线变化趋势与开路故障的相近。由图 3 可知,同一故障类型在不同环境下的输出特性也会发生变化,易导致不同故障状态在不同环境条件下出现相似输出。因此,仅选取 $F_1 \sim F_6$ 不能完全涵盖完整故障信息,无法充分反映不同环境下的故障类型,这可能导致诊断模型出现“误判”,从而影响诊断结果的准确性。



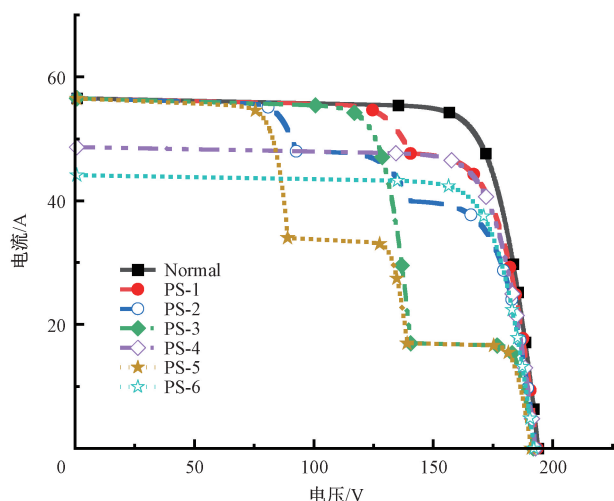
(a) 开路、短路、老化故障 $I-V$ 曲线

(a) $I-V$ curve of open circuit, short circuit and abnormal degradation

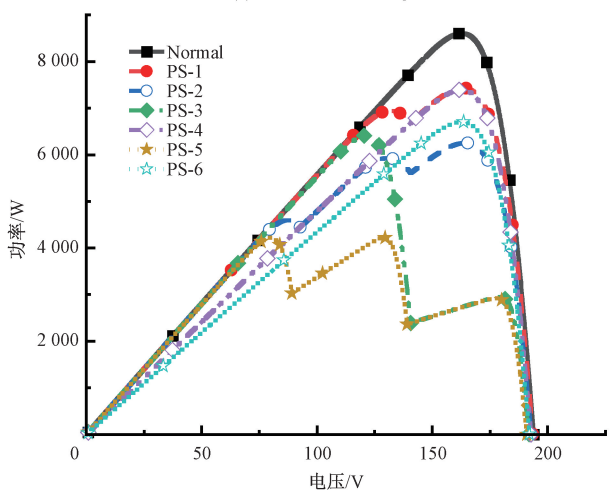


(b) 开路、短路、老化故障 $P-V$ 曲线

(b) $P-V$ curve of open circuit, short circuit and abnormal degradation



(c) 遮挡故障 I - V 曲线
(c) I - V curve of shading fault



(d) 遮挡故障 P - V 曲线
(d) P - V curve of shading fault

图4 不同故障的输出曲线

Fig. 4 Output curves for different faults

综上所述,又提出构建以下新的故障诊断特征量,记为 $F_7 \sim F_{10}$:

$$F_7 = \frac{\Delta U_m}{\Delta I_m} \quad (1)$$

$$F_8 = \frac{I_m \times U_m}{I_{sc} \times U_{oc}} \quad (2)$$

$$F_9 = \frac{I_m}{U_m - U_{oc}} \quad (3)$$

$$F_{10} = \frac{I_{sc} - I_m}{-U_m} \quad (4)$$

式中: F_7 为 I - V 曲线点 (U_m, I_m) 处的斜率; F_8 为填充因子,表征阵列的故障程度; F_9 为 I - V 曲线点 (U_m, I_m) 与 $(U_{oc}, 0)$ 之间连线的斜率; F_{10} 为 I - V 曲线点 (U_m, I_m) 与 $(0, I_{sc})$ 之间连线的斜率。

将上述 10 维电气参数特征 $F_1 \sim F_{10}$ 与环境条件 T_{pv} 、 G 组合成 12 维故障特征量,能够反映同一环境条件下更多故障特征信息,更准确地反映输出特性曲线上关键数据点的特征,以此提升对各类故障及其不同程度的诊断精度。

1.3 基于 KPCA 的故障特征提取

考虑到实际运行中不同光伏阵列选用光伏组件参数的差异以及阵列规模的差别,对 10 维电气参数故障特征进行标准化处理。通过保留原有数据特征并消除不同维度或数量级的干扰,确保特征数据具有可比性,从而增强诊断方法的适应性、稳定性及准确性。

$$F_i^* = \frac{F_i}{F_{STC}} \quad (5)$$

式中: F_i 表示第 i 个特征量的原始值; F_{STC} 表示 STC 正常状态下该特征量的值; F_i^* 表示 F_i 标准化处理后的值。

针对多维故障特征中存在的信息冗余和噪声干扰问题,通过 KPCA 对标准化处理后的特征进行融合和有效提取。在将高维数据映射到低维空间的过程中,尽可能保留原始数据的结构特征,最终得到的 Pareto 如图 5 所示。

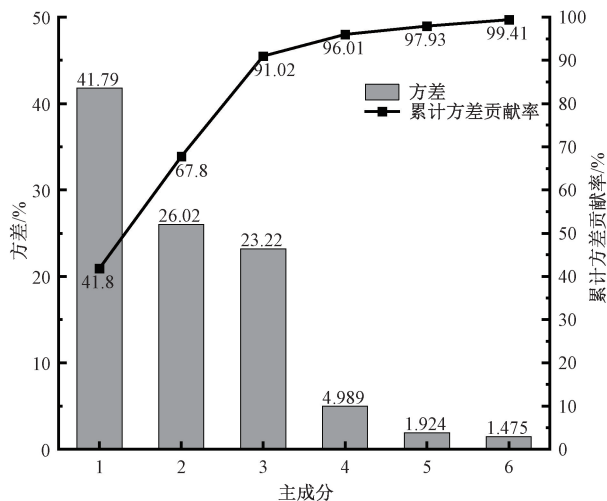


图5 KPCA 主成分分析

Fig. 5 Kernel principal component analysis

由图 5 可以看出,利用 KPCA 进行处理后,前 6 个核主成分特征的累计方差贡献率达到了 99.41%,充分说明这些主成分已涵盖了绝大部分的故障信息。因此,将高维数据降维至 6 个合理且有效,能够在保留关键信息的同时减少数据维度,提高后续诊断方法的效率和准确性。

2 融合多策略改进鹈鹕优化算法

2.1 鹈鹕优化算法 (POA)

POA 是由 Trojovský 等^[20]于 2022 年提出的一种启发

式智能优化算法,基本思想是模仿鹈鹕在攻击和捕猎时的行为和策略。POA 算法的寻优过程分为种群初始化阶段和狩猎阶段,鹈鹕种群的初始化公式如式(6)所示。

$$x = lb + rand \cdot (ub - lb) \quad (6)$$

式中: x 为鹈鹕的位置; $rand$ 是区间 $[0,1]$ 的随机数; ub 、 lb 分别表示鹈鹕上下位置边界。

POA 算法的狩猎(迭代寻优)过程分为移向猎物 and 掠过水面两个阶段。移向猎物(探索阶段),该阶段猎物在搜索空间内随机分布,鹈鹕识别猎物的位置并向其移动。每次迭代中,鹈鹕新位置的数学模型为:

$$x_{i,j}^{p_1} = \begin{cases} x_{i,j} + rand \cdot (p_j - I \cdot x_{i,j}), F_p < F_i \\ x_{i,j} + rand \cdot (x_{i,j} - p_j), F_p \geq F_i \end{cases} \quad (7)$$

式中: $x_{i,j}^{p_1}$ 为第 1 阶段更新后第 i 只鹈鹕在第 j 维的位置; P_j 为猎物在第 j 维的位置; I 为 $[1,2]$ 的随机整数; F_p 为猎物的目标函数值; F_i 为第 i 个候选解的目标函数值。

掠过水面(开发阶段),该阶段鹈鹕到达水面后展开翅膀,将猎物放在喉袋中向上移动。假设狩猎的领域半径为 R ,每次迭代中,鹈鹕新位置的数学模型为:

$$x_{i,j}^{p_2} = x_{i,j} + R \cdot (2 \cdot rand - 1) \cdot x_{i,j} \quad (8)$$

$$R = 0.2 \cdot (1 - \frac{t}{T}) \quad (9)$$

式中: $x_{i,j}^{p_2}$ 为第 2 阶段更新后第 i 只鹈鹕在第 j 维的位置; t 为当前迭代次数; T 为最大迭代次数。

在鹈鹕狩猎过程的两个阶段中,若更新后的新位置更优,则进行位置更新,否则保持不变,表达式如式(10)所示。

$$X_i = \begin{cases} X_i^{p_k}, F_i^{p_k} < F_i \\ X_i, \text{其他} \end{cases} \quad (10)$$

式中: k 为 1 或 2; $X_i^{p_k}$ 表示第 k 阶段更新后第 i 只鹈鹕的新位置。

2.2 多策略改进鹈鹕优化算法

1) 改进 Tent 混沌映射初始化

在 POA 算法中,鹈鹕种群在目标搜索空间中随机初始化导致种群个体分布杂乱,难以保证种群的多样性,影响后续的寻优性能。为提高鹈鹕种群初始的丰富性,引入一种基于 Tent 映射的种群初始化方法,使种群个体在解空间中均匀且随机分布。

混沌映射是一种非线性映射方法,具有随机性、遍历性、有序性等优点,常被用于改进优化算法领域^[21]。相较于其他混沌映射,Tent 映射遍历性更好,性能更佳,其表达式如式(11)所示。

$$x_{i+1} = \begin{cases} \frac{x_i}{\alpha}, 0 \leq x_i < \alpha \\ \frac{1 - x_i}{1 - \alpha}, \alpha \leq x_i < 1 \end{cases} \quad (11)$$

式中: α 是 $[0,1]$ 的随机数; x_i 为第 t 次迭代映射值; x_{i+1} 为第 $t+1$ 次迭代映射值。

Tent 映射产生混沌序列过程中,当 $\alpha = 0.5$ 时,系统呈现短周期状态;当初值与 α 相同时,系统将演化成周期状态。因此,为避免陷入以上两种特殊情况,在 Tent 映射的基础上加入干扰因子,修改后的 Tent 映射表达式为:

$$x_{i+1} = \begin{cases} \frac{x_i}{\alpha} + rand \cdot \frac{1}{\mu}, 0 \leq x_i < \alpha \\ \frac{1 - x_i}{1 - \alpha} + rand \cdot \frac{1}{\mu}, \alpha \leq x_i < 1 \end{cases} \quad (12)$$

式中: μ 为控制因子。

2) 惯性权重和混沌干扰因子

POA 在探索阶段,个体位置更新受当前所处位置影响较大,为了更好的平衡全局寻优和局部探索能力,选择引入一种非线性惯性权重因子 ω ,用于调整位置更新于当前位置的相关程度,表达式定义为:

$$\omega = \omega_l + (\omega_h - \omega_l) \cdot (\frac{e^{1/T} - 1}{e - 1})^\varphi \quad (13)$$

式中: ω_l 为最小惯性权重; ω_h 为最大惯性权重; φ 为系数。

算法迭代初期,惯性权重因子较小,位置更新与当前位置相关性小,有利于个体尽可能探索整个空间,避免过早陷入局部最优解;迭代后期,个体位置更新与当前位置相关性变大,从而围绕当前位置进行探索开发,增强算法的局部寻优能力,加快收敛速度。

同时,为更好地提高算法的搜索精度,进一步引入 Iterative 混沌映射作为干扰因子,代替原有的随机数 $rand$ ^[22]。通过改变不同的步长,增强 POA 的遍历性和随机性,提高其收敛速度和算法精度,如式(14)所示。

$$\rho_{i+1} = \frac{1 + \sin(0.7\pi/\rho_i)}{2} \quad (14)$$

式中: ρ_i 为第 t 次迭代的干扰因子值。

3) 非线性收敛因子

POA 在开发阶段,个体位置更新时,狩猎半径 R 随着迭代次数的增加呈线性减少,难以有效协调鹈鹕种群的全局搜索与局部开发能力。为此,通过引入非线性收敛因子对算法进行了改进,以增强前期的全局搜索效果,避免算法出现“早熟”现象;在后期则缩小狩猎范围,提升局部开发效率。改进后的狩猎范围半径公式为:

$$R = \frac{0.4}{\pi} \cdot \arccos(\frac{t}{T})^2 \quad (15)$$

4) 自适应 t 分布变异策略

为进一步提升算法的全局寻优能力,并防止在算法后期随着迭代次数的增加而出现最优解附近个体陷入局部最优的现象,引入一种自适应 t 分布变异策略^[23]。该

策略通过动态调整鸚鵡个体的位置,增强种群的多样性,从而提高全局搜索能力。具体表达式如式(16)所示。

$$X^* = X_{\text{best}} + X_{\text{best}} \cdot t(T) \quad (16)$$

式中: X^* 为变异扰动后最佳位置; X_{best} 为当前最佳位置; $t(T)$ 为以迭代次数为自由度 n 的 $t(n)$ 分布。

t 分布概率曲线如图 6 所示,随着迭代次数的增加,自由度 n 逐渐增大,分布曲线逐渐趋于标准正态分布。这意味着,随着迭代的进行,变异幅度逐渐减小,算法能够更好地平衡全局探索与局部开发。

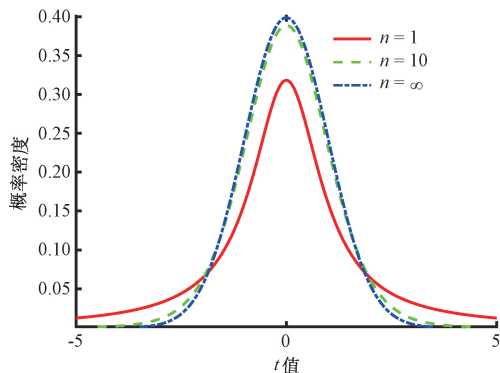


图 6 t 分布概率密度曲线

Fig. 6 t -distribution probability density curve

2.3 基于 IPOA-SVM 的故障诊断模型

1) SVM

SVM 是一种二分类模型,具有样本需求低、训练时间短、分类效果好、高维特征表现良好等优点,但对大规模数据集和多分类问题的处理相对困难,对噪声和异常值敏感^[24]。其核心在于构建一个最优超平面将不同数据分隔,使得分类距离最远,相应的分类决策函数为:

$$f(x) = \text{sgn} \left[\sum_{i=1}^n a_i y_i K(x_i \cdot x_j) + b \right] \quad (17)$$

式中: a 为拉格朗日因子,满足 $0 \leq a \leq C$; C 为 SVM 超参数惩罚因子; $K(x_i \cdot x_j)$ 为核函数; b 为相应的阈值实数。

由于光伏阵列的故障特征和分类类别较多,呈现出非线性多分类问题,因此需要对 SVM 进行多分类扩展。为此,需选择合适的核函数,将低维非线性数据映射到高维空间,使其能够线性可分。采用一对一(one-vs-one, OvO)分类器,将多分类问题转化为多个二分类任务。鉴于高斯核函数能够映射到无限维,具备多样化的决策边界和较快的收敛速度,选择使用高斯核函数,其表达式为:

$$K(x_i \cdot x_j) = \exp \left(-\frac{x_i - x_j^2}{2\gamma^2} \right) \quad (18)$$

式中: γ 为核函数参数。

2) 基于 IPOA-SVM 的故障诊断模型流程

所提的基于特征提取与 IPOA-SVM 的光伏阵列故障

诊断方法,主要包括模型输入特征选择、特征提取和模型优化。首先,分析输出特性曲线,选择合适的特征向量;其次,利用 KPCA 方法对选取的输入特征进行提取;最后,采用 IPOA 算法对 SVM 模型的惩罚因子 C 和核函数参数 γ 进行迭代寻优,构建最优诊断模型,并对测试数据集进行故障诊断。

构建基于 KPCA-IPOA-SVM 的光伏阵列故障诊断模型,诊断流程如图 7 所示。

步骤 1) 采集不同条件下光伏阵列的输出特性曲线;

步骤 2) 数据预处理,以输出曲线为基础提取 10 维电气参数特征,同环境条件组成 12 维故障样本数据集,标准化处理后作为模型输入;

步骤 3) 特征提取,通过 KPCA 方法对故障特征进行特征融合和有效特征提取;

步骤 4) 数据划分,将处理后的样本数据集按 8 : 1 : 1 的比例随机划分为训练集、验证集和测试集;

步骤 5) 适应度函数设定为最大化识别准确率,因此将诊断错误率定义为适应度函数,其表达式为:

$$\text{fitness} = 1 - \frac{1}{n} \sum_{z=1}^n |y'_z - y_z| \quad (19)$$

式中: y_z 和 y'_z 表示第 z 个样本对应的预测标签和实际标签; n 为样本总数;

步骤 6) 模型训练,融合多策略改进 POA 算法对 SVM 模型的超参数进行寻优,输出最佳超参数组合;

步骤 7) 故障诊断,将超参最佳值代入模型进行训练验证,将测试集数据输入故障诊断模型,得到对应的故障类型及故障诊断准确率。

3 仿真模拟与结果分析

3.1 仿真数据采集

基于所构建的光伏阵列仿真模型,分别对 15 种光伏阵列运行状态进行仿真模拟。为覆盖多种环境条件状况,设计温度 T_{PV} 范围为 $5 \sim 50 \text{ }^\circ\text{C}$,步长变化为 $5 \text{ }^\circ\text{C}$;太阳辐照度 G 范围为 $100 \sim 1\,200 \text{ W/m}^2$,步长变化为 50 W/m^2 。根据在不同环境条件和运行状态下的模拟仿真,共采集 $15 \times 10 \times 23 = 3\,450$ 条输出特性曲线,并从中提取相应的故障特征量,建立包含 3 450 组样本的光伏阵列仿真故障数据集。

3.2 特征有效性验证

为验证所选取电气参数特征作为待选故障特征向量的有效性,设置仿真数据标准化结果的箱型图,分别呈现每一个参数特征的数据分布,如图 8 所示。

图 8 结果表明,标准化后的特征量表现出良好的统一性和聚合度,不同故障间差异明显,不同故障程度间也呈现出一定的分布差异。这表明选取的故障特征与故障

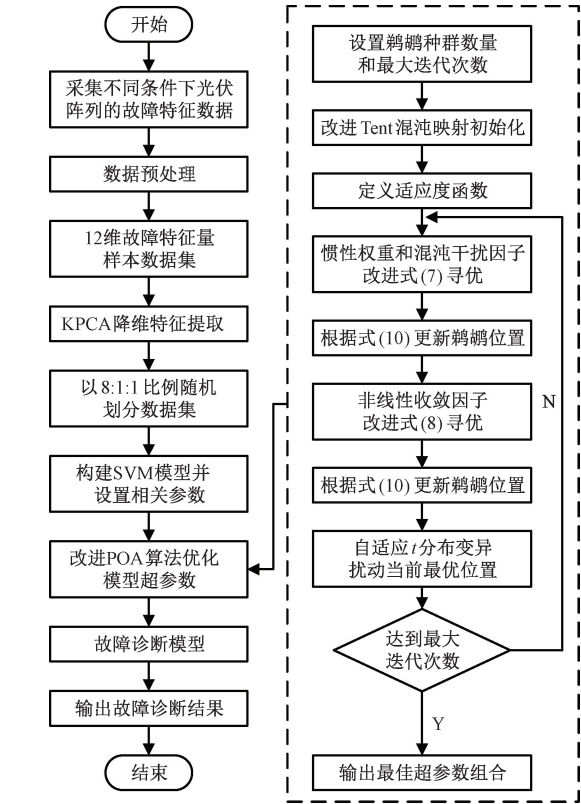


图 7 光伏阵列故障诊断流程

Fig. 7 Flow chart of fault diagnosis for PV array

类别具有强相关性,有助于避免“误诊”现象的发生,提高故障诊断模型的诊断效果。

3.3 仿真结果与分析

设置 IPOA-SVM 模型参数如下:惩罚因子 C 的搜索范围为 $(0.1, 100)$,核函数参数 γ 的搜索范围为 $(0.1, 5)$,设种群数量 $N=30$,最大迭代次数 $T=100$,训练集、验证集和测试集按 $8:1:1$ 的比例随机划分。为量化模型的性能,采用混淆矩阵和分类准确率作为诊断效果评价指标,定义如表 2 所示。

表 2 混淆矩阵

Table 2 Confusion matrix

真实类	预测类	
	真正类 (TP)	假负类 (FN)
	假正类 (FP)	真负类 (TN)

混淆矩阵能够显示分类结果中被正确和错误分类的样本数量,其中行表示实际类别,列表示预测类别。分类准确率 (accuracy, Acc) 表示样本正确预测的比例。

$$Acc = \frac{1}{N} \sum_{i=1}^N Acc_i = \frac{1}{N} \sum_{i=1}^N \frac{TP_i + TN_i}{TP_i + TN_i + FP_i + FN_i} \quad (20)$$

式中: N 为类别总数; Acc_i 表示第 i 类故障的分类准确率;

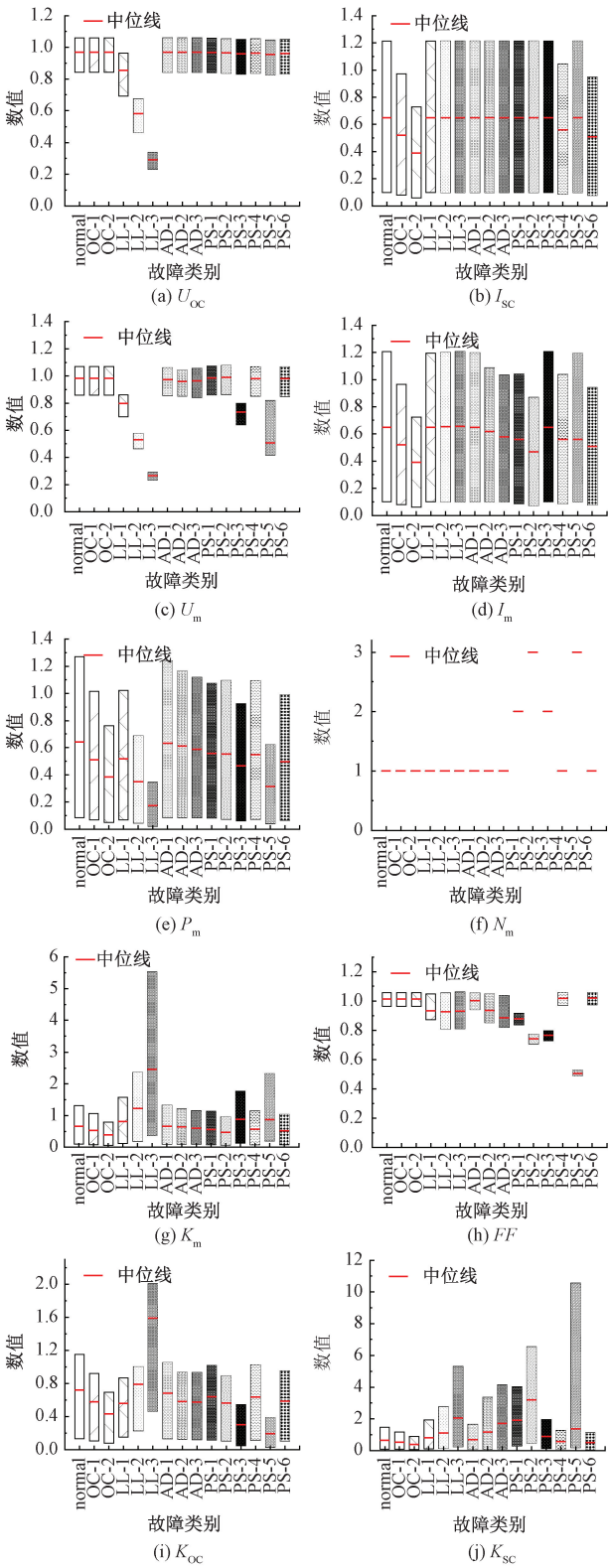


图 8 不同故障特征标准化箱型图

Fig. 8 Standardized box plots for different fault characteristic

TP_i 和 TN_i 分别表示模型正确地分类为第 i 类和非 i 类的样本数量; FP_i 和 FN_i 表示错误地将样本分类为第 i 类和

非 i 类的样本数量。

1) KPCA 特征提取对模型诊断的影响分析

为验证基于 KPCA 故障特征提取方法对改善模型性能的有效性以及较其他特征提取方法的优越性。采用不同特征提取方法对 12 维故障特征进行降维处理,并将降维后的特征数据作为 IPOA-SVM 模型的输入进行故障诊断,诊断结果如表 3 所示。

表 3 诊断结果对比

Table 3 Comparison of diagnostic results

特征提取方法	特征维度	平均分类准确率/%
Pearson	6	87. 83
灰色关联分析	6	88. 70
Spearman	6	87. 25
KPCA	6	98. 84

由表 3 可知,采用基于 KPCA 的特征提取方法,IPOA-SVM 模型的平均诊断准确率最高,达到 98. 84%,而使用 Pearson、灰色关联分析、Spearman 的诊断准确率分别为 87. 83%、88. 70%、87. 25%。诊断结果表明基于 KPCA 的故障特征提取方法能有效提高光伏阵列故障诊断准确率,模型改善效果更加优越。

2) IPOA-SVM 模型诊断结果分析

将 KPCA 特征提取后的仿真数据集输入 IPOA-SVM 模型进行诊断验证,其故障诊断结果如图 9 所示。可以看出,在 345 个测试集样本中有 5 个样本诊断错误,实验中 OC-2 的 1 个样本被诊断为 OC-1,PS-4 的 1 个样本被诊断为 PS-6,从而导致这两类故障准确率较低;Normal 和 AD-1 之间出现了较多样本误诊的情况,这是因为光伏阵列在发生 AD-1 故障时,其输出曲线与正常运行时的曲线差异较小,仅在最大功率点附近有细微变化。事实上,轻微的老化故障对光伏阵列整体运行影响较小,因此可以将其视为正常状态。

IPOA-SVM 模型能够准确识别单一故障类型,考虑到光伏实际运行中可能会出现不同组串支路同时发生不同故障的复合型故障 (mixed fault, Mix),设置 4 种复合型故障如表 4 所示。

表 4 复合型故障类型

Table 4 mixed fault types

标签	故障类型	具体描述
Mix-1	复合故障	一条组串开路+一个组件短路
Mix-2		一个组件短路+老化(电阻为 1 Ω)
Mix-3		一条组串开路+老化(电阻为 3 Ω)
Mix-4		一个组件短路+局部阴影遮挡(PS-1)

在 3. 1 节设置的仿真条件下获取每种复合故障数据

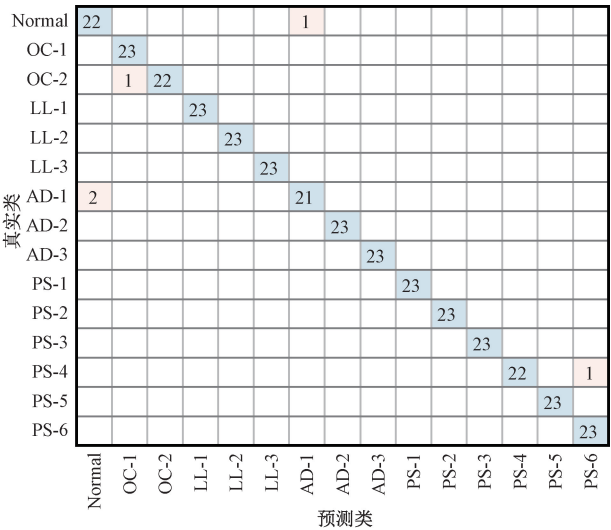


图 9 故障诊断结果的混淆矩阵

Fig. 9 Confusion matrix of fault detection results

230 组,与单一故障数据组成新的仿真数据集代入 IPOA-SVM 模型进行多次测试,结果如图 10 所示。

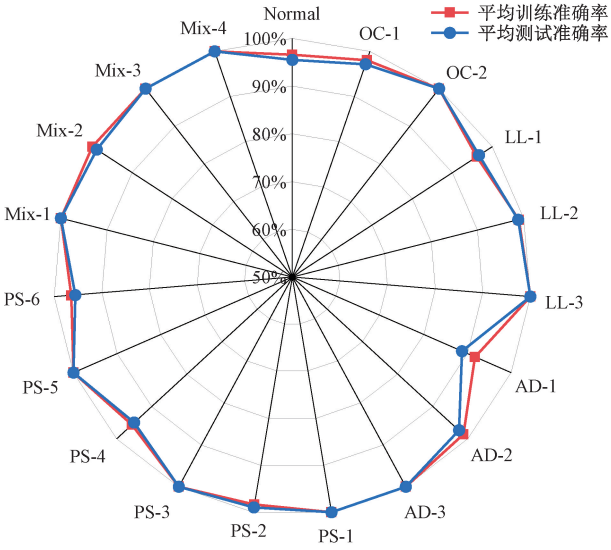


图 10 复合型故障诊断结果

Fig. 10 Mixed fault diagnosis results

同单一故障数据集诊断结果相比,复合故障的存在导致 LL-1 与 Mix-2 故障出现“误诊”,其他故障类型变化不大,4 种复合故障类型识别率为 100%。结果表明,IPOA-SVM 模型对复合型故障同样具有较高的适应性。

3) 不同故障特征量和算法模型对比分析

为验证所提方法的有效性和准确性,将 6 维特征数据集与 12 维特征量数据集经 KPCA 特征提取后分别输入到 PSO-SVM、GWO-SVM、POA-SVM 及 IPOA-SVM 模型中进行诊断。考虑到数据集划分对诊断结果的影响,进

行多次实验以获取诊断模型的平均准确率,诊断结果如表 5 所示,适应度曲线如图 11 所示。

表 5 不同故障特征量及算法模型的诊断准确率

Table 5 Diagnosis accuracy of different fault characteristic vectors and algorithmic models (%)			
故障特征量	算法模型	平均训练准确率	平均测试准确率
6 维特征量	PSO-SVM	82.75	81.16
	GWO-SVM	82.79	80.87
	POA-SVM	83.45	81.73
	IPOA-SVM	84.12	82.56
12 维特征量	PSO-SVM	98.73	95.36
	GWO-SVM	98.48	96.81
	POA-SVM	98.3	97.97
	IPOA-SVM	98.51	98.55

由表 5 可以看出,6 维特征量的测试准确率不足 90%,而 12 维特征量的测试准确率均达到 95% 以上,达到高准确率水平。这说明提高故障特征维度,向数据集加入更多有效故障特征信息,对于提高模型分类效果有较大帮助;从不同算法模型的对比来看,所提出的 IPOA-SVM 故障诊断模型平均诊断分类准确率为 98.55%,相较于 PSO-SVM、GWO-SVM、POA-SVM 模型分别提高了 3.19%、1.74%、0.58%,故障诊断分类准确率更高,更适用于所研究故障的诊断分类。

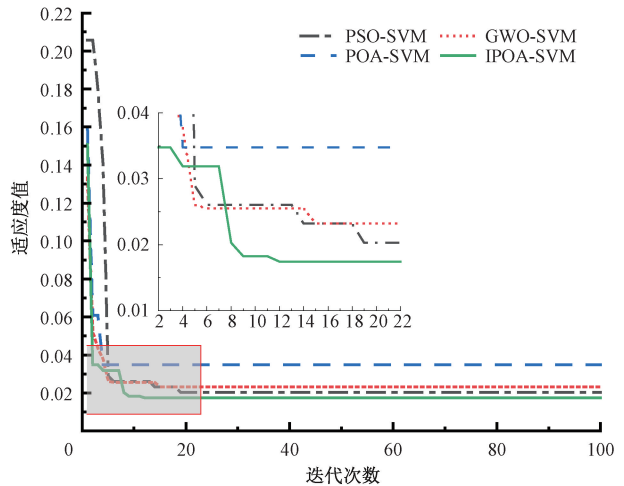


图 11 适应度曲线对比
Fig. 11 Comparison chart of adaptation curves

适应度曲线如图 11 所示,可以看出,POA-SVM 模型达到最佳适应度所需迭代次数最少,但相较其他模型最佳适应度值表现较差。PSO-SVM、GWO-SVM 模型分别需要 19 次、15 次得到最佳适应度值,而 IPOA-SVM 模型所需迭代次数最少,收敛速度更快,仅 12 次即可达到最佳,且最佳适应度值更优。

4 实验验证与分析

4.1 实验平台搭建

为进一步验证所提故障诊断方法的实际应用效果并评估模型的适用性,采用 12 块参数相同的光伏组件,构建了 4×3 的光伏阵列进行测试。选用光伏组件的标况参数为开路电压 21.6 V、短路电流 1.83 A、最大功率点电压 18 V、最大功率点电流 1.67 A、最大输出功率 30 W。实验时,通过环境监测仪分别测量环境中的温度和太阳辐照度,利用 I - V 测试仪和万用表测量光伏阵列的输出特性,故障实验平台如图 12 所示。

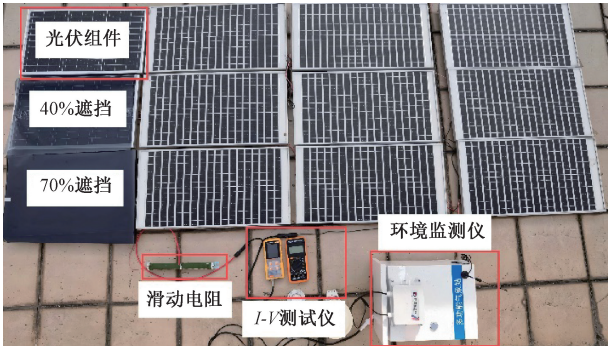


图 12 光伏阵列故障实验平台

Fig. 12 Experimental platform for PV array faults

故障实验设计如下:

- 1) 采用与仿真实验相同的设计,断开阵列内一条、两条支路,模拟 2 种开路故障。
- 2) 短接阵列内一块、两块光伏组件,模拟 2 种短路故障。
- 3) 调节光伏阵列中串联滑动变阻器的电阻大小,以 1、3、5 Ω 电阻值模拟 3 种异常老化故障。
- 4) 通过采用两种不同遮光率的薄膜遮挡光伏组件模拟不同的遮光程度。覆盖透光性较强的薄膜用于模拟轻微遮挡,其遮光率约为 40%;覆盖透光性较差的薄膜用于模拟严重遮挡故障,其遮光率约为 70%。以与仿真模拟相同的遮挡模式设计 6 种局部阴影遮挡故障。

实验在每天 8:00~16:00 时间段内采集数据,环境条件为 10~40 $^{\circ}\text{C}$ 、50~1 100 W/m^2 ,每种故障类型采集 100 组数据,共计 1 400 组。根据实验设计的不同运行状态,每条数据标注相应的标签,组成 1 400 组数据作为故障诊断的实验数据集。

4.2 实验结果与分析

分别将 6 维实验数据集与 12 维实验数据集经 KPCA 特征提取后作为 IPOA-SVM 模型的输入进行故障诊断,得到的诊断分类结果如表 6 所示。

表 6 不同运行状态的诊断分类准确率
Table 6 Diagnostic classification accuracy
for different operating states (%)

运行状态	6 维特征量	12 维特征量
Normal	68.45	85.64
OC-1	61.43	93.62
OC-2	96.71	100
LL-1	100	100
LL-2	100	100
AD-1	46.24	95.1
AD-2	72.1	100
AD-3	76.88	96.66
PS-1	100	100
PS-2	100	100
PS-3	100	100
PS-4	59.92	100
PS-5	100	100
PS-6	63.74	100
合计	81.82	97.93

由表 6 可见,受环境因素和数据采集过程中噪声、扰动等影响,实验数据的故障诊断分类准确率低于仿真数据。采用 6 维特征量进行故障诊断存在大量故障类型被误诊的情况,诊断效果远低于 12 维特征量。实验结果表明,采用 12 维特征量存在少许 Normal、AD-1 样本误诊与 OC-1、AD-3 样本误诊情况,但 IPOA-SVM 模型的平均整体分类准确率仍达到 97.93%,能够为光伏电站的早期故障预警提供高置信度判据。结合电站运维管理系统,实时诊断结果可触发预防性维护指令(如清洗遮挡组件或更换老化部件),从而降低故障扩大风险,提升电站运行安全性。

对比 PSO-SVM、GWO-SVM、POA-SVM 及 IPOA-SVM 模型的实验数据诊断结果,并进行多次重复实验,各模型的平均分类准确率如图 13 所示。

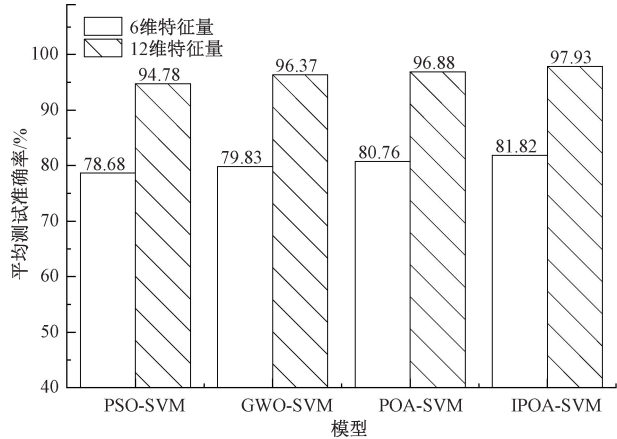


图 13 不同模型的诊断结果

Fig. 13 Diagnostic results of different models

由图 13 可知,采用 12 维特征量的模型诊断准确率明显优于 6 维特征量,且采用 IPOA-SVM 模型的平均诊断分类准确率优于 PSO-SVM、GWO-SVM 和 POA-SVM 模型,分别提高了 3.15%、1.56%和 1.05%。

5 结 论

本文提出了一种基于特征提取与改进鹌鹑算法优化支持向量机的光伏阵列故障诊断方法,能够有效识别光伏阵列多种类型及不同程度的故障。通过仿真模型分析光伏阵列在不同故障条件下的 I - V 输出特性,基于 6 维基本特征量构建了 10 维电气参数特征,并结合环境条件形成 12 维故障特征量,精确反映了 I - V 曲线的关键特性;采用 KPCA 对 12 维特征进行特征融合与有效提取,成功提取 6 个关键主成分,有效减少了高维特征量之间的冗余信息和噪声干扰,显著提升了诊断模型的准确性;融合改进的 Tent 混沌映射、惯性权重、非线性收敛因子和自适应 t 分布变异多策略改进鹌鹑优化算法,构建了基于 KPCA-IPOA-SVM 的故障诊断模型。通过仿真与实验数据的对比验证,12 维故障特征在提高诊断准确率方面展现出明显的优势,并进一步证明了 IPOA-SVM 模型在故障诊断中的高准确性与适用性,为光伏阵列故障诊断的研究提供了可靠的技术参考。

此外,本文提出的方法能够准确诊断光伏阵列的故障类型及其严重程度,为故障预测和预防性维护提供了可靠的数据基础。未来的研究可以结合历史诊断数据与时间序列分析(如 LSTM 网络),构建光伏组件健康状态评估模型,预测性能退化趋势并制定动态维护策略,以进一步提升光伏电站的整体安全性。同时,还可探讨将所提方法部署到光伏逆变器的嵌入式系统中,实现实时故障诊断与边缘计算,从而增强模型在实际应用中的鲁棒性与广泛适用性。

参考文献

[1] WAQAR A M, LI G, JIN Y, et al. Failures of photovoltaic modules and their detection: A review [J]. Applied Energy, 2022, 313: 118822.

[2] 徐进. 2023 年我国能源电力发展综述 [J]. 能源, 2024 (3): 29-35.

XU J. An overview of energy and power development in China in 2023 [J]. Energy Industry, 2024 (3): 29-35.

[3] QIU T, WANG L, LU Y, et al. Potential assessment of photovoltaic power generation in China [J]. Renewable and Sustainable Energy Reviews, 2022, 154: 111900.

[4] 董华展. 太阳能光伏发电发展趋势探析 [J]. 现代工业经济和信息化, 2023, 13 (4): 282-284.

DONG H ZH. Development trend of solar photovoltaic power generation [J]. Modern Industrial Economy and

- Informationization, 2023, 13(4): 282-284.
- [5] 刘行行, 帕孜来·马合木提, 程志江, 等. 基于 XGBoost 的光伏阵列故障诊断方法研究[J]. 电子测量技术, 2023, 46(12): 8-14.
- LIU H H, PAZILAI M, CHENG ZH J, et al. Research on fault diagnosis method of photovoltaic array based on extreme gradient boosting[J]. Electronic Measurement Technology, 2023, 46(12): 8-14.
- [6] 吕游, 郑茜, 齐欣宇, 等. 基于改进 EfficientNet 的红外图像光伏组件故障识别研究[J]. 仪器仪表学报, 2024, 45(4): 175-184.
- LYU Y, ZHENG Q, QI X Y, et al. A study on fault recognition of photovoltaic module with infrared images based on improved efficientnet[J]. Chinese Journal of Scientific Instrument, 2024, 45(4): 175-184.
- [7] RAHMAN M M, KHAN I, ALAMEH K. Potential measurement techniques for photovoltaic module failure diagnosis: A review[J]. Renewable and Sustainable Energy Reviews, 2021, 151: 111532.
- [8] WANG J, GAO D, ZHU S, et al. Fault diagnosis method of photovoltaic array based on support vector machine[J]. Energy Sources, Part A: Recovery, Utilization, and Environmental Effects, 2019, 45(2): 5380-5395.
- [9] HONG Y Y, PULA R A. Methods of photovoltaic fault detection and classification: A review[J]. Energy Reports, 2022, 8: 5898-5929.
- [10] 蒋琳, 苏建徽, 李欣, 等. 基于可见光和红外热图像融合的光伏阵列热斑检测方法[J]. 太阳能学报, 2022, 43(1): 393-397.
- JIANG L, SU J H, LI X, et al. Hot spot detection of photovoltaic array based on fusion of visible and infrared thermal images[J]. Acta Energaie Solaris Sinica, 2022, 43(1): 393-397.
- [11] 刘强, 郭珂, 毛明轩, 等. 一种基于串联等效电阻的光伏故障检测方法[J]. 太阳能学报, 2020, 41(10): 119-126.
- LIU Q, GUO K, MAO M X, et al. A photovoltaic fault detection method based on series equivalent resistance[J]. Acta Energaie Solaris Sinica, 2020, 41(10): 119-126.
- [12] 朱青云, 刘凡, 曾伟. 基于 I-V 曲线逆推法的光伏组件故障诊断策略[J]. 可再生能源, 2024, 42(8): 1029-1036.
- ZHU Q Y, LIU F, ZENG W. Fault diagnosis method of solar cell based on inverse inference of I-V curves[J]. Renewable Energy Resources, 2024, 42(8): 1029-1036.
- [13] LI B, DELPHA C, MIGAN-DUBOIS A, et al. Fault diagnosis of photovoltaic panels using full I-V characteristics and machine learning techniques[J]. Energy Conversion and Management, 2021, 248: 114785.
- [14] 孙培胜, 陈堂贤, 程陈, 等. 基于 SOA-SVM 模型的光伏阵列故障诊断研究[J]. 电源学报, 2025, 23(1): 143-150.
- SUN P SH, CHEN T X, CHENG CH, et al. Research on fault diagnosis of photovoltaic array based on SOA-SVM model[J]. Journal of Power Supply, 2025, 23(1): 143-150.
- [15] 赵清英, 吴晶晶, 张雪辉, 等. 基于萤火虫扰动麻雀搜索算法-极限学习机的光伏阵列故障诊断方法研究[J]. 电网技术, 2023, 47(4): 1612-1625.
- ZHAO J Y, WU J J, ZHANG X H, et al. Fault diagnosis of photovoltaic arrays based on sparrow search algorithm with firefly perturbation-extreme learning machine[J]. Power System Technology, 2023, 47(4): 1612-1625.
- [16] 李斌, 高鹏, 郭自强. 改进蛱蝶算法优化 LSTM 的光伏阵列故障诊断[J]. 电力系统及其自动化学报, 2024, 36(8): 70-78.
- LI B, GAO P, GUO Z Q. Improved dung beetle optimizer to optimize LSTM for photovoltaic array fault diagnosis[J]. Proceedings of the CSU-EPSCA, 2024, 36(8): 70-78.
- [17] 李斌, 郭自强, 高鹏. 改进北方苍鹰算法在光伏阵列中应用研究[J]. 电子测量与仪器学报, 2023, 37(7): 131-139.
- LI B, GUO Z Q, GAO P. Application of improved northern goshawk optimization algorithm in photovoltaic array[J]. Journal of Electronic Measurement and Instrumentation, 2023, 37(7): 131-139.
- [18] MELLIT A, TINA G M, KALOGIROU S A. Fault detection and diagnosis methods for photovoltaic systems: A review[J]. Renewable and Sustainable Energy Reviews, 2018, 91: 1-17.
- [19] MARANGIS D, LIVERA A, TZIOLIS G, et al. Trend-based predictive maintenance and fault detection analytics for photovoltaic power plants[J]. Solar RRL, 2024, 8(24): 2400473.
- [20] TROJOVSKY P, DEGHANI M. Pelican optimization algorithm: A novel nature-inspired algorithm for engineering applications[J]. Sensors, MDPI AG, 2022, 22(3): 855.
- [21] 李智杰, 赵铁柱, 李昌华, 等. 融合多策略的改进鹈鹕优化算法[J/OL]. 控制工程, 1-15[2024-09-14].
- LI ZH J, ZHAO T ZH, LI CH H, et al. Improved pelican optimization algorithm fused with multi-strategy[J/OL]. Control Engineering of China, 1-15[2024-09-14].
- [22] LI Y, LIU Y, LIN C, et al. An improved pelican optimization algorithm based on chaos mapping factor[J]. Engineering Letters, 2023, 31(4): 1605-1634.

[23] 张家玮,李琳,张奇志. 混合多策略改进的樽海鞘群算法及其应用[J]. 计算机工程与设计, 2024, 45(3): 822-829.

ZHANG J W, LI L, ZHANG Q ZH, et al. Hybrid multi-strategy improved slap swarm algorithm and its application [J]. Computer Engineering and Design, 2024, 45(3): 822-829.

[24] 周晓华,冯雨辰,陈磊,等. 改进秃鹰搜索算法优化 SVM 的变压器故障诊断研究[J]. 电力系统保护与控制, 2023, 51(8): 118-126.

ZHOU X H, FENG Y CH, CHEN L, et al. Transformer fault diagnosis based on SVM optimized by the improved bald eagle search algorithm [J]. Power System Protection and Control, 2023, 51(8): 118-126.

作者简介



韩茂林, 2021 年于哈尔滨理工大学获得学士学位, 现为上海海洋大学硕士研究生, 主要研究方向为光伏系统故障诊断、能源互联网。

E-mail: maolinH0812@ 163. com

Han Maolin received his B. Sc. degree from Harbin University of Science and Technology in 2021. Now he is a M. Sc. candidate at Shanghai Ocean University. His main research interests include Fault diagnosis of photovoltaic systems and energy internet.



杨琛(通信作者), 2001 年于中国石油大学获得学士学位, 2004 年于重庆大学获得硕士学位, 2016 年于同济大学获得博士学位, 现为上海海洋大学副教授, 主要研究方向为光伏故障诊断、能源互联网、物联网技术、渔光互补技术。

E-mail: cyang@shou. edu. cn

Yang Chen (Corresponding author) received her B. Sc. degree from China University of Petroleum in 2001, M. Sc. degree from Chongqing University in 2004 and Ph. D. degree from Tongji University in 2016, respectively. Now she is an associate professor at Shanghai Ocean University. Her main research interests include Photovoltaic fault diagnosis, energy internet, Internet of Things technology and fishery-photovoltaic complementary technology.