

doi: 10.7690/bgzd.2023.01.010

基于 ADASYN-随机森林的智能家电内部电路故障诊断

舒一飞¹, 樊博¹, 康洁滢¹, 郭汶昇¹, 曾来²

(1. 国网宁夏电力有限公司营销服务中心, 银川 750000; 2. 四川大学电气工程学院, 成都 610065)

摘要: 针对智能家电内部电路故障诊断中存在的数不平衡和分类器诊断精度低的问题, 提出一种基于 ADASYN 算法过采样和随机森林 (random forest, RF) 的故障诊断方法。将电流信号进行小波包分解, 提取最后一层各节点能量作为特征向量; 使用 ADASYN 算法扩充训练数据集, 得到随机森林故障诊断模型并进行测试。实验结果表明: ADASYN-随机森林故障诊断模型对智能家电内部电路故障具有较高的诊断精度, 对故障诊断有一定的实用价值和指导意义。

关键词: 故障诊断; 特征提取; 小波包分解; ADASYN 算法; 随机森林

中图分类号: TP306 **文献标志码:** A

Fault Diagnosis of Internal Circuits of Smart Home Appliances Based on ADASYN-random Forest

Shu Yifei¹, Fan Bo¹, Kang Jieying¹, Guo Wensheng¹, Zeng Lai²

(1. Marketing Service Center, State Grid Ningxia Electric Power Co., Ltd., Yinchuan 750000, China;

2. College of Electrical Engineering, Sichuan University, Chengdu 610065, China)

Abstract: Aiming at the problems of data imbalance and low accuracy of classifier in the fault diagnosis of the internal circuit of intelligent appliances, a fault diagnosis method based on ADASYN algorithm over-sampling and random forest is proposed. The current signal is decomposed by wavelet packet, and the energy of each node in the last layer is extracted as the feature vector. The training data set is expanded by ADASYN algorithm, and the random forest fault diagnosis model is obtained and tested. The experimental results show that the ADASYN-random forest fault diagnosis model has high diagnosis accuracy for the internal circuit fault of intelligent household appliances, and has certain practical value and guiding significance for fault diagnosis.

Keywords: fault diagnosis; feature extraction; wavelet packet decomposition; ADASYN algorithm; random forest

0 引言

随着科学技术与信息技术的不断发展, 出现了各式各样的智能家电。不同于传统家电, 大部分智能家电内部结构复杂, 内部电路发生故障时难以用传统方法排查。设计一种便捷有效的算法诊断智能家电故障, 对于智能家电制造商而言, 在提升其产品竞争力的同时, 可以节约售后成本; 对于家庭用户而言, 可以提升使用体验, 及时排查风险, 减少安全隐患。

文献[1]使用傅里叶分析法诊断电路故障, 首先采用傅里叶变换将时域信息变换到频域中, 然后进行频域故障特征抽选, 实现故障诊断, 但该方法仅针对开路故障进行分析。文献[2]系统地介绍了采用支持向量机算法诊断短路故障的过程及能量特征, 但最终得到的模型诊断精度较低。文献[3]使用小波神经网络诊断电路故障, 但仅分析了单故障模式的

情况, 对诊断算法的具体分析较少。文献[4]提出了一种优选小波包和极端学习机相结合的电路故障诊断方法, 使用极端学习机对故障进行分类识别, 训练时间较短, 但仍存在诊断准确率较低的问题。文献[5]针对 SVM 中核参数的选择问题, 提出一种基于数据最大方差关联度准则的方法来衡量数据集的线性可分性, 并结合 PSO 算法建立了 RBF 核函数的自动优选方法, 利用其进行电路故障诊断, 诊断准确率较高, 但没有考虑各类样本分布不平衡的问题。

为了解决上述问题, 笔者提出采用 ADASYN 算法和随机森林算法相结合的诊断智能家电的内部电路故障方法, 首先在仿真软件中建立并运行智能家电内部电路模型, 采集智能家电在发生各种故障时的电流数据, 通过 5 层小波包分解, 提取最后 1 层各节点的小波包能量作为特征向量, 接着使用

收稿日期: 2022-09-12; 修回日期: 2022-10-20

基金项目: 国家电网公司总部科技项目 (5700-202155204A-0-0-00)

作者简介: 舒一飞 (1989—), 男, 贵州人, 工程师, 从事主站建设运维、线损管理、反窃电大数据应用研究。E-mail: 240374472@qq.com。

ADASYN 算法平衡各类故障样本数据, 最后得到随机森林故障诊断模型。

1 小波包分解

原始电流信号维数较高。考虑到训练时长以及资源消耗, 需要对原始电流信号进行特征提取。笔者提取小波包能量特征作为特征向量, 把原始高维的电流信号转化为低维的特征向量, 有利于减少计算量, 缩短训练时间。

在发生故障的情况下, 电流信号会产生高次谐波, 时域特征不能完整地表示原始信号。笔者通过小波包分解, 提取最后 1 层各节点的能量作为特征向量。小波包分解通过 1 组正交、迅速衰减的小波基拟合原始信号, 具有较高的时频分辨率。小波变换只对信号的低频分量进行分解, 而小波包分解同时对信号的高频分量和低频分量进行分解, 能展现更多原始信号的细节, 便于后续的特征提取^[6-7]。3 层小波包分解如图 1 所示。

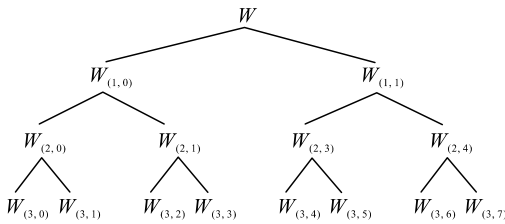


图 1 3 层小波包分解

图中: W 为原始时域信号; $W_{(i,j)}$ ($i=0, 1, 2, 3; j=0, 1, 2, \dots, 7$) 为小波包分解后第 i 层第 j 个节点的信号。小波包分解的具体递推公式如下:

$$\left. \begin{aligned} d_k^{j+1, 2n} &= \sum_{k \in Z} d_p^{j, n} h_{(2^p-k)} \\ d_k^{j+1, 2n+1} &= \sum_{k \in Z} d_p^{j, n} g_{(2^p-k)} \end{aligned} \right\} \quad (1)$$

式中: d 为小波包, 上标表示小波包的分解级数, 下标表示该小波包在此级的位置; $h_{(2^p-k)}$ 为低通滤波器; $g_{(2^p-k)}$ 为高通滤波器。

原始时域信号经过小波包分解后, 节点能量可通过下式计算:

$$E_{p,j} = \int |W_{(p,j)}^m(t)| dt = \sum_{k=1}^n |d_{j,k}|^2 \quad (2)$$

式中: $E_{p,j}$ 为第 p 层第 j 个节点的能量; $d_{(i,k)}$ ($j=0, 1, 2, \dots, 7; k=1, 2, \dots, n$) 为节点 $W_{(p,j)}^m$ 的小波包分解系数。小波包分解后第 p 层各个节点的能量特征向量为:

$$E = [E_{p,0}, E_{p,1}, \dots, E_{p,2^p-1}]$$

将能量特征向量进行归一化处理:

$$X = E / \sum_{j=0}^{2^p-1} E_{p,j} \quad (3)$$

在实际应用中, 小波包分解层数过多会增加运算的复杂度, 降低诊断效率, 容易导致信号失真, 故一般选择 3~6 层较为合适。

2 不平衡数据处理

理想情况下每个类标签对应的样本比例应该近似相等, 但实际中采集到的各样本数据量往往相差很大。训练集中各个类的样本数量不平衡时, 模型的预测结果会偏向训练集中的多数类, 容易造成误判, 故训练前需要解决数据集不平衡的问题。解决数据集不平衡最常用的方法是数据采样和数据加权, 这 2 种方法在数学上是等价的, 但对不同的机器学习算法, 数据采样和数据加权的训练效果有所不同, 随机森林算法会对数据集进行随机抽样, 使用数据采样的方法往往效果更好。数据采样分为过采样和欠采样, 在计算资源足够并且少数类样本较多的情况下, 使用过采样的方法扩充少数类样本更合适。笔者使用 ADASYN 算法^[8]对少数类样本进行过采样, 解决样本数据不平衡的问题, 该算法根据学习难度使用不同的权重生成少数类样本, 算法步骤如下:

1) 确定需要生成的样本数量:

$$G = (m_l - m_s) \times \beta \quad (4)$$

式中: m_l 为现有多数类样本数量; m_s 为现有少数类样本数量; $\beta \in [0, 1]$, 当 $\beta=1$ 时, 合成后的少数类样本和多数类样本数量相等。

2) 计算每个少数类样本的 K 个近邻中多数类样本所占比例:

$$r_i = \Delta_i / K \quad (5)$$

式中: Δ_i 为第 i 个少数类样本的 K 个近邻中多数类样本的数量, $i=1, 2, \dots, m_s$ 。

3) 标准化 r_i :

$$\hat{r}_i = r_i / \sum_{i=1}^{m_s} r_i \quad (6)$$

4) 对每一个少数类样本计算需要生成新样本的数量:

$$g_i = \hat{r}_i \times G \quad (7)$$

式中 G 为总共需要生成的样本数量。

5) 对每一个少数类样本 x_i , 根据下式生成 g_i 个新样本:

$$s_i = x_i + (x_{z_i} - x_i) \times \lambda \quad (8)$$

式中： s_i 为生成的样本； x_{z_i} 为 x_i 的 K 个近邻中随机选取的一个少数类样本； λ 为一个随机数， $\lambda \in [0,1]$ 。

3 随机森林

随机森林(RF)是由随机构造的森林模型组成，而每个森林模型由许多决策树组成。决策树(decision tree, DT)是一种树形结构的监督学习算法，主要应用于分类问题。决策树的每个内部节点表示一个属性的一次测试，每条边表示对该属性的一个测试输出，每个叶节点表示某个类或者类分布。决策树将训练集中的所有样本置于根节点，然后选择最优特征划分数据集，若子集基本被正确分类，则构建叶节点，否则重新选择新的最优特征继续划分。如图 2 所示，决策树是一种简单并且高效的机器学习算法，在分类问题中应用广泛，但决策树容易生成一个复杂的模型，造成过拟合，影响模型的泛化能力^[9]。

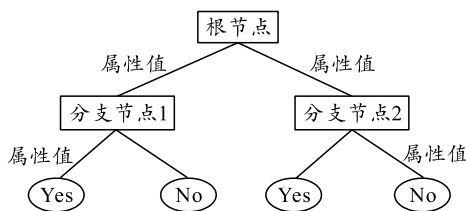


图 2 决策树

RF 是基于 bagging 框架^[10]的一种集成学习算法，通过对多个弱分类器的结果进行投票表决，形成一个强分类器^[11]。具体步骤可以描述为：

1) 对数据集 N 使用 Bootstrap 法重抽样，有放回地随机抽取 n 个样本，重复 K 次后，新的数据集记为：

$$N_k, k=1, 2, \dots, K;$$

2) 以 N_k 作为新的训练数据集，构建根节点，随机选取 s 个特征构建决策树 T_k ，并且使 T_k 最大可

能的生长，不进行剪枝；

3) 重复步骤 2)，生成 K 棵决策树，对所有决策树的结果投票表决，投票次数最多的类别作为最终结果，表决公式为：

$$R(x) = \arg \max_y \sum_{k=1}^K I[r_k(x) = y] \quad (9)$$

式中： x 为训练数据集； y 为样本的实际分类结果； $r_k(x)$ 为训练得到的决策树模型。

与决策树相比，随机森林算法抗干扰能力更强，可以避免无效数据点对训练结果造成的负面影响，防止模型过拟合。随机森林算法的泛化误差取决于森林中每颗数的分类能力以及任意 2 棵树之间的相关性，Bootstrap 法重抽样能够降低各子树之间的相关性，提高模型的泛化能力。此外，与深度学习相比，随机森林算法对样本数量的要求更少，有利于解决实际情况下故障数据难以采集的问题。

4 随机森林的故障诊断算例分析

为验证随机森林故障诊断模型对智能家电内部电路故障的诊断效果，在仿真平台中搭建智能家电内部电路模型，设置 4 种不同的故障状态，获取支路的电流数据用于训练及测试随机森林故障诊断模型。

4.1 故障模型搭建

智能家电内部电路仿真模型如图 3 所示。从左到右的支路编号分别记为 0、A、B、C、D、E、F、G、H、I、J、K，用电器的功率因素在 0.7~0.95 之间。在实际情况中，短路故障(F1)、断线故障(F2)、串联电弧故障(F3)和泄漏电流故障(F4)发生的可能性比较大，笔者主要针对这 4 种故障模式进行分析。发生短路电弧故障和泄漏电流故障时，输出的电流信号和发生故障的支路无关，故这 2 种故障只采集单个支路的电流数据。

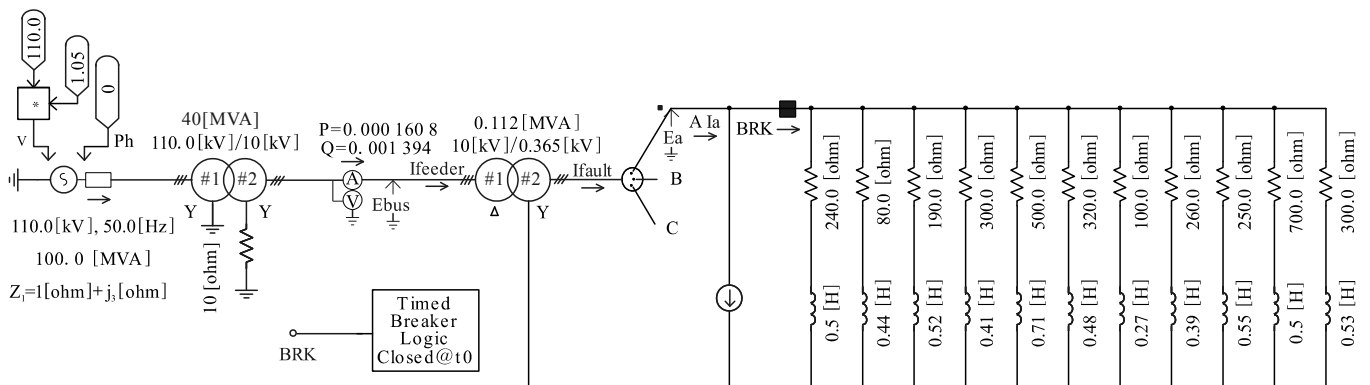


图 3 智能家电内部电路仿真模型

4.2 数据采集

在 PSCAD 仿真平台建立智能家电内部电路模型后, 电源电压有效值设置为 220 V, 总仿真时间为 5 s, 测量频率为 1 kHz, 模拟发生故障。在各种故障情况下, 采集发生故障支路的电流信号, 故障电流波如图 4 所示。

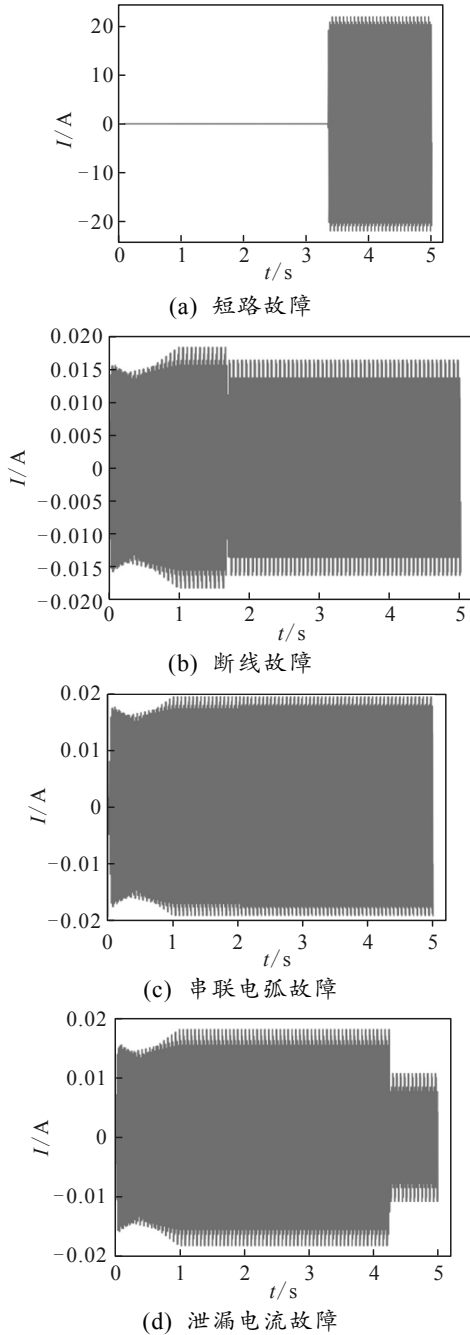


图 4 不同故障状态下的电流波形

4.3 评价指标

在评价模型的好坏时, 通常选择的指标有准确率(accuracy, A)、精确率(precision, P)、召回率(recall, R)和 F1 值(F1-score)等。在评价不平衡数

据集分类算法的性能时, 准确率会有一定缺陷。评价多分类问题时, 通常的做法是把 n 分类分解为 n 个 2 分类, 评价时每次以其中一个类为正类, 其余类统一为负类, 然后分别计算 2 分类指标, 最后通过平均得到多分类评价指标, 常用的平均方式有宏平均(macro)和微平均(micro), 但微平均受多数类影响较大, 故笔者采用宏平均。

假设有 $n(n > 2)$ 类数据, 对于类别 $i(i=1, 2, \dots, n)$ 的 TP、TN、FP 和 FN 值描述如下:

真正(true positive, TP), 将类别 i 的样本判断为类别 i 的样本;

真负(true negative, TN), 将其他类别 j 的样本判断为其他类别 $m(m \neq j)$ 的样本;

假正(false positive, FP), 将其他类别 j 的样本判断为类别 i 的样本;

假负(false negative, FN), 将类别 i 的样本判断为其他类别 j 的样本。

则多分类中类别的精确率、召回率和 F1 值定义如下:

$$P_i = TP_i / (TP_i + FP_i); \quad (10)$$

$$R_i = TP_i / (TP_i + FN_i); \quad (11)$$

$$F1\text{-score}_i = 2 \cdot P_i \cdot R_i / (P_i + R_i). \quad (12)$$

多分类中所有类别精确率、召回率和 F1 值的宏平均定义如下:

$$P_m = \sum_{i=1}^n P_i / n; \quad (13)$$

$$R_m = \sum_{i=1}^n R_i / n; \quad (14)$$

$$F1\text{-score}_m = 2 \cdot P_m \cdot R_m / (P_m + R_m). \quad (15)$$

4.4 随机森林故障诊断

随机森林故障诊断的具体步骤可以表述如下:

1) 采集数据。

采集得到的故障样本共 9 100 组, 其中: 短路故障 1 500 组, 断线故障 1 200 组, 串联电弧故障 1 400 组, 泄漏电流故障 5 000 组。在 9 100 组样本中, 随机选取 6 370 组作为训练样本, 其余 2 730 组作为测试样本。

2) 提取故障特征。

6 370 组训练样本经过 5 层小波包分解后, 每组样本得到 32 维的故障特征向量, 共得 6 370 组 32 维故障特征向量构成的训练集。

3) 处理不平衡样本。

上一步得到的原始训练集中, 短路故障 1 056

组, 断线故障 827 组, 串联电弧故障 976 组, 泄漏电流故障 3 511 组。采用 ADASYN 算法对数量较少的 3 种故障进行扩充后, 得到串联电弧故障数据 3 456 组, 短路故障数据 3 511 组, 断线故障数据 3 536 组, 泄漏电流故障特征数据 3 522 组, 作为最终的训练集。

4) 随机森林故障诊断模型。

将随机森林中决策树的数量设置为 300 棵, 训练后得到随机森林故障诊断模型。

5) 模型测试。

对其余的 2 730 组测试样本重复步骤 2), 得到 2 730 组 32 维特征向量构成的测试集, 使用随机森林故障诊断模型测试。

4.5 实验结果及对比

针对 4 种不同故障情况, 将电流数据进行 5 层小波包分解后, 不同故障情况下第 1 节点、第 32 节点小波包分解波形如图 5 所示。从图中可以看出, 无论在低频还是高频, 不同故障的分解波形差别很大, 故将 5 层小波包分解后的最后一层各节点的能量作为特征向量可以较好地地区分不同故障情况。

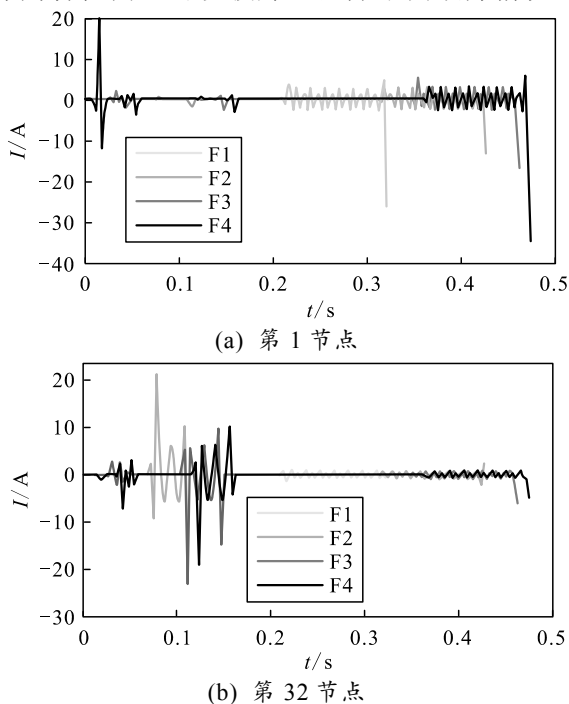


图 5 不同故障第 1 节点、第 32 节点分解波形

为了使实验结果更具说服力, 在 ADASYN 算法扩增数据集的基础上, 将随机森林故障诊断模型与 SVM^[12]、决策树对不同故障情况的测试结果进行比较, 使用精确率、召回率和 F1 值作为评价指标, 测试结果如图 6—8 所示。从图中可以看出, 针

对 4 种不同故障, ADASYN-随机森林算法对断线故障 (F2) 的诊断精确率略低于 ADASYN-SVM 算法, 但综合各项指标来看, ADASYN-随机森林算法明显优于 ADASYN-SVM 和 ADASYN-决策树算法。

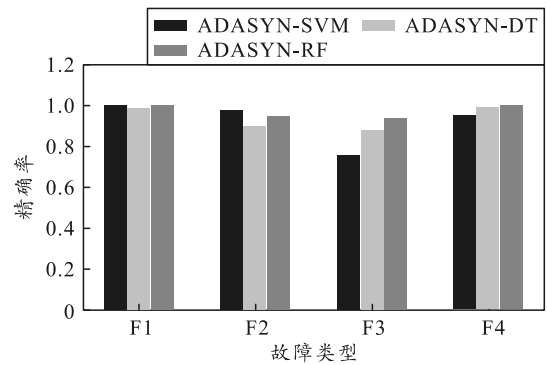


图 6 不同算法各种故障类型的精确率对比

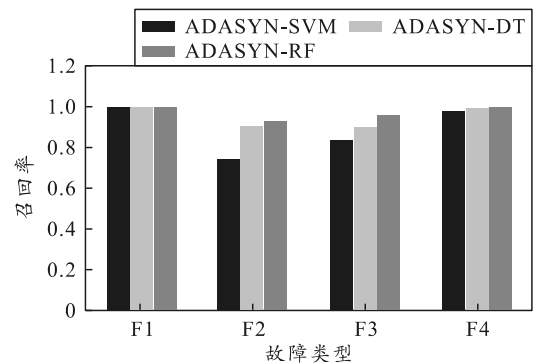


图 7 不同算法各种故障类型的召回率对比

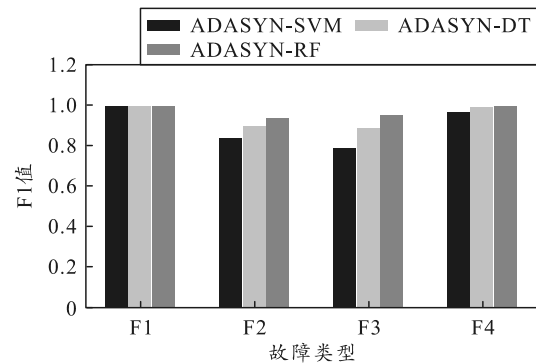


图 8 不同算法各种故障类型的 F1 值对比

为了进一步验证笔者提出方法的有效性, 分别在原始数据集和 ADASYN 算法扩增后的数据集的基础上, 比较随机森林、SVM 和决策树算法, 测试结果如表 1 所示。

从下表中可以看出, 无论是 SVM、决策树还是随机森林, 使用 ADASYN 算法扩充训练集后的模型测试效果均明显好于原始不平衡训练集。其中 ADASYN-随机森林的各项指标明显优于其他算法。

表 1 测试结果对比 %

方法	精确率	召回率	F1 值
SVM	91.31	82.16	85.56
决策树	93.56	93.68	93.62
随机森林	96.19	95.91	95.96
ADASYN-SVM	91.68	89.28	90.00
ADASYN-决策树	94.15	94.48	94.31
ADASYN-随机森林	97.12	97.18	97.15

5 结束语

笔者针对智能家电内部电路故障的诊断，搭建智能家电内部电路仿真模型，将随机森林算法用于诊断智能家电内部电路故障，通过小波包分解提取故障信号的能量特征，有利于解决时域下可能存在高次谐波的问题，提高故障特征提取的稳定性。采用 ADASYN 算法对训练集进行扩充，解决了实际情况下采集得到的各类样本分布可能不平衡的问题。实验结果表明：ADASYN-随机森林故障诊断模型有较高的诊断精度，对电路故障的诊断具有一定的实用意义和理论研究价值。

参考文献：

[1] 徐德洪, 程肇基, 崔学军. 用傅里叶分析法诊断电力电子电路的故障[J]. 浙江大学学报(自然科学版), 1994, 28(6): 667-673.

[2] 胡清, 王荣杰, 詹宜巨. 基于支持向量机的电力电子电路故障诊断技术[J]. 中国电机工程学报, 2008, 28(12): 107-111.

(上接第 41 页)

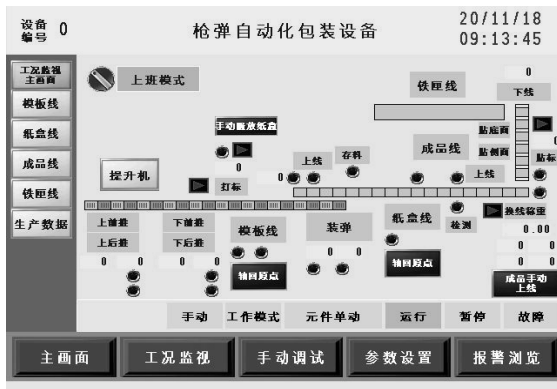


图 5 触摸屏工况监视页面

5 结束语

枪弹纸盒自动化包装设备已投入使用，系统运行可靠，生产效率得到有效提升，人工成本大大降低，生产节拍可达到 200 发/min。使用结果表明：该设备运行流畅，纸盒包装完整，满足实际生产需要，证明了该控制方案的可行性。

[3] 谭阳红, 何怡刚. 模拟电路故障诊断的小波方法[J]. 电工技术学报, 2005, 20(8): 90-93.

[4] 何星, 王宏力, 陆敬辉, 等. 基于优选小波包和 ELM 的模拟电路故障诊断[J]. 仪器仪表学报, 2013, 34(11): 2615-2618.

[5] 周绍磊, 廖剑, 史贤俊. RBF-SVM 的核参数选择方法及其在故障诊断中的应用[J]. 电子测量与仪器学报, 2014, 28(3): 240-245.

[6] 王阳, 乐健, 周谦, 等. 基于小波包分析与神经网络的变压器区内外故障判断方法[J]. 电测与仪表, 2020, 57(7): 1-7, 15.

[7] KUMAR M S, RAMANATHAN R, JAYAKUMAR M, et al. Physical layer secret key generation using discrete wavelet packet transform[J]. Ad Hoc Networks, 2021, 118(3): 102523.

[8] HE H, BAI Y, GARCIA E A, et al. ADASYN: Adaptive synthetic sampling approach for imbalanced learning[C]// Neural Networks, 2008. IJCNN 2008. (IEEE World Congress on Computational Intelligence). IEEE International Joint Conference on IEEE, 2008.

[9] 杨茂, 翟冠强. 基于决策树理论的风电功率实时预测方法[J]. 电测与仪表, 2018, 55(11): 120-124.

[10] BREIMAN L. Using Iterated Bagging to Debias Regressions[J]. Machine Learning, 2001, 45(3): 261-277.

[11] BREIMAN L. Random forest[J]. Machine Learning, 2001, 45: 5-32.

[12] SUYKENS J A K, VANDEWALLE J. Least squares support vector machine classifiers[J]. Neural Processing Letters, 1999, 9(3): 293-300.

参考文献：

[1] 李威, 胡翔, 张博. 枪弹包装线虚拟样机设计及仿真[J]. 兵工自动化, 2020, 39(7): 48-51.

[2] 冉松, 涂集林, 黎梅, 等. 智能制造在枪弹制造行业内的应用[J]. 兵工自动化, 2020, 39(11): 24-26.

[3] 史慧芳, 郭进勇, 胡翔, 等. 基于 Rocky/Ansys Workbench 的发射药自动称量包装过程仿真[J]. 兵工自动化, 2021, 40(8): 56-60.

[4] 戴秋洪, 代久双, 李建康, 等. 包装密封性在线检测锁紧装置设计与仿真计算[J]. 兵工自动化, 2021, 40(9): 89-91.

[5] 金翰林, 李锦, 谷严波. 某枪弹装配系统的控制系统设计[J]. 兵工自动化, 2020, 39(6): 27-29.

[6] 范庆辉, 钱俊松, 黄权, 等. 弹头涂色标机控制系统设计[J]. 兵工自动化, 2018, 37(1): 79-82.

[7] 徐亮, 李作武, 钟顺金. 基于 PLC 的冲压力控制系统[J]. 兵工自动化, 2013, 32(4): 64-66.

[8] 刘远博, 薛丽贤, 郑毅. 可编程控制器 PLC 控制系统程序设计探讨[J]. 黄金, 2011, 32(4): 34-36.