谢可, 叶春明. 基于 XGBoost-INGO-KELM 的生产质量评估模型[J]. 智能计算机与应用,2026,16(7):161-171. DOI:10.20169/j.issn.2095-2163.24110403

基于 XGBoost-INGO-KELM 的生产质量评估模型

谢 可, 叶春明
(上海理工大学 管理学院, 上海 200093)

摘 要: 针对制造过程中多工序工艺质量评估的问题,提出了一种基于 XGBoost-INGO-KELM 模型的生产质量评估方法。首先利用 XGBoost 算法计算多工序生产过程中每个特征的重要性,删除没有重要性的特征,将剩余特征作为特征子集用于分类训练。再利用 INGO(Improved Northern Goshawk Optimization)算法找出 KELM(Kernel based Extreme Learning Machine)中的最佳参数,然后将最佳参数代入 KELM 中,用于生产质量的分类与预测。通过实验结果表明,相比较 SSA、NGO、INGO-KELM 等模型,本文所提出的 XGBoost-INGO-KELM 模型预测精度更高,为高精度生产质量评估提供了更好的实现方法和途径。

关键词: XGBoost; INGO; KELM; 特征选择; 生产质量评估
中图分类号: TP18 **文献标志码**: A **文章编号**: 2095-2163(2026)07-0161-11

A production quality evaluation model based on XGBoost-INGO-KELM

XIE Ke, YE Chunming
(Business School, University of Shanghai for Science and Technology, Shanghai 200093, China)

Abstract: Aimed at the problem of multi-process production quality evaluation in the manufacturing process, a production quality evaluation model based on XGBoost-INGO-KELM is proposed. Firstly, XGBoost algorithm is used to calculate the importance of each feature in the production process, delete features without importance, and use the remaining features as feature subsets for classification training. Then, the INGO algorithm is used to find the optimal parameters in KELM, and then the optimal parameters are brought into KELM for classification and prediction of production quality. The experimental results show that compared to models such as SSA, NGO, and INGO-KELM, the XGBoost-INGO-KELM model proposed in this paper has higher prediction accuracy, providing a better way to achieve high-precision production quality assessment. .

Key words: XGBoost; INGO; KELM; feature selection; evaluation of production quality

0 引 言

随着现代制造技术的不断发展,生产过程日趋复杂,涉及多道工序、各类材料及多种工艺参数的相互作用。在此背景下,如何在生产过程中有效保证产品质量合格性,同时持续提升制造效率,已成为企业面临的核心挑战之一[1]。长期以来,多工序工艺质量的评估主要依赖于生产人员的经验判断。然而,这样的评估方法效率低下,且容易受到评估者主观判断的影响,导致评估结果出现偏差,这对工艺质量评估工作产生了负面效果[2]。鉴于此,本研究旨在采集工艺生产过程中相关数据,并应用经过优化的 KLEM 算法,对其生产质量进行分类预测分析。

极限学习机(Extreme Learning Machine, ELM)是由 Huang 等学者[3]提出的学习算法。是一种用于训练单隐藏层前馈神经网络的机器学习算法。其关键思想是在输入层和隐藏层之间的权重是随机分配的,而输出权重是通过解析确定的。核极限学习机(Kernel based Extreme Learning Machine, KELM)通过在 ELM 框架中引入核函数,将数据隐式地映射到一个高维特征空间。在这个空间中,数据更有可能是线性可分的。KELM 保留了 ELM 的许多优点、

基金项目: 上海市哲学社会科学一般项目(2022BGL010)。
作者简介: 谢 可(2000—),女,硕士研究生。主要研究方向:智能算法。
通信作者: 叶春明(1964—),男,博士,教授。主要研究方向:生产调度,智能算法,工业工程。Email:yechunming6464@163.com 。
收稿日期: 2024-11-04

如快速训练能力和良好的泛化能力,并且通过使用核函数,可以有效地处理更复杂的数据分布和非线性问题[4]。还能将原本在低维空间中无法分割的数据集转换到高维空间中,并实现线性分割,从而显著提高了模型的预测精度[5]。因此,在分类预测领域,KELM 获得了广泛的应用[6]。汤庆峰等学者[7]将 KELM 用于微电网短期负荷预测。龙霞飞等学者[8]将 KELM 用于风电齿轮箱故障诊断。胡爱军等学者[9]将 KELM 用于滚动轴承多工况故障识别。

针对特征属性过多的问题,主成分分析(PCA)方法得到了广泛的应用。马杰等学者[10]将 PCA-PSO-GRU 模型用于重力拱坝诱导缝开合度预测。白雲杰等学者[11]将 PCA-EDT-CNN 模型用于柴油发动机的故障诊断。但 PCA 降维会导致信息损失,有时可能会失去关键信息,而且在某些情况下,PCA 无法找到最重要的特征。而随机森林(RF)能够对原始数据的信息进行保留。彭豪杰等学者[12]将随机森林用于 PM2.5 浓度预测。Wang 等学者[13]将随机森林用于冠心病风险评估。但 RF 使用的是随机特征选择和决策树的平均投票策略,可能会导致过拟合和欠拟合。相比之下,XGBoost 采用的是一种类似梯度提升的策略,能够在每一轮迭代中学习到数据的残差,从而提高整个模型的预测精度。Yang 等学者[14]将 XGBoost 用于预测高速公路交通事故严重程度。Ma 等学者[15]将 XGBoost 算法用于山洪风险评估。

另外,KELM 参数的选取会影响模型的性能和结果。针对这一问题,刘傲等学者[16]利用 PSO 找出 KELM 中的最佳参数对,并将 PSO-KELM 模型用于相控阵雷达识别。但 PSO 可能会陷入局部最优解,因其搜索范围受到已有的粒子位置和速度的限制,这使得 PSO 对于非凸优化问题的求解相对困难。周勇等学者[17]利用 SSA 找出 KELM 的最佳参数对,并将 SSA-KELM 模型应用于煤与瓦斯突出预测。但是麻雀搜索算法没有维护全局信息,导致搜索过程受限于当前解的领域范围。朱昶胜等学者[18]利用 DIGWO 算法找出 KELM 中的最佳参数,然后将 VMD-DIGWO-KELM 模型用于短期风电功率预测。DIGWO 算法中涉及的参数较多,而算法性能受参数设置的影响较为显著。为了进一步提升 KELM 参数选取的质量,近年来,出现了许多有着出众优化能力的智能算法。其中,具有代表性的北方苍鹰优化(Northern Goshawk Optimization, NGO)算法通过模仿北方苍鹰的行为模式,展现出良好的搜索能力和收敛性能。李思尧等学者[19]利用 NGO 找出 VMD 中的最佳分解参数,并将 NGO-VMD-HHT 模型用于电缆局部放电信号特征量提取。但是 NGO 也存在算法收敛精度低、容易陷入局部最优等缺陷。

为解决上述优化算法的不足,本文引入一种改进的北方苍鹰优化(Improved Northern Goshawk Optimization, INGO)算法,在传统 NGO 算法的基础上引入 Cubic 混沌映射解决种群初始化不均匀问题,提升算法的收敛速度[20-21]。对于算法中容易陷入局部最优解的情况,本文采取小孔成像反向学习策略,通过缩放因子动态调节反向解的数量、优化种群数量,避免种群陷入局部最优。

当数据集体量庞大时,INGO-KELM 模型的性能往往会受到影响。鉴于此,本研究提出了一种新的模型,该模型结合了 XGBoost 算法在数据降维上的优势和 INGO 算法在优化参数选择上的优势,进一步改进了 KELM 模型。为了验证改进后的 XGBoost-INGO-KELM 模型的有效性,本文以某制丝公司制造过程中的生产数据集作为输入,包含多个特征及质量类别信息。而输出则是生产质量的分类结果,即针对每个样本评估其生产质量类别。本文旨在通过相关操作,基于模型优化后的参数,最终得出高精度的预测结果。实验结果显示,本文提出的 XGBoost-INGO-KELM 模型在评估工艺生产质量的应用中展现出了良好的性能。

1 方法描述

1.1 核极限学习机

ELM(Extreme Learning Machine)的核心思想是快速训练神经网络。通过在输入层和隐藏层之间随机初始化权重,而无需进行迭代优化过程。这使得 ELM 能够在训练速度和准确性之间实现良好的平衡。

对于给定的样本 $S=\{(X_i, Y_i), i=1,2,3,\cdots, n\}$,这里 X_i 为网络输入集,Y_i 为网络映射输出集,i 为隐含层节点数, ELM 的数学表达式为:

$$\boldsymbol{F}(\boldsymbol{x})=\boldsymbol{h}(\boldsymbol{x})\cdot\boldsymbol{\beta}=\boldsymbol{H}\cdot\boldsymbol{\beta} \tag{1}$$

其中,$\boldsymbol{F}(\boldsymbol{x})$ 表示学习期望目标函数;$\boldsymbol{x}$ 表示网络输入的向量;$\boldsymbol{H}=\boldsymbol{h}(\boldsymbol{x})$ 表示隐含层节点输出矩阵;$\boldsymbol{\beta}$ 表示输出权重,通过最小二乘法可得出:

$$\boldsymbol{\beta}=\boldsymbol{H}^{\mathrm{T}}\left(\frac{\boldsymbol{I}}{\boldsymbol{C}}+\boldsymbol{H}\boldsymbol{H}^{\mathrm{T}}\right)^{-1}\boldsymbol{y} \tag{2}$$

其中,$\boldsymbol{I}$ 表示对角单位矩阵;$\boldsymbol{C}$ 表示正则化系

数；$\boldsymbol{y}=(y_1,y_2,\cdots,y_n)$ 表示期望输出。

然而，ELM 可能出现过拟合问题，也可能遇到较差的初始化导致模型性能不佳。为此，Zhu 等学者[22]在 ELM 中引入核函数思想，构造出 KELM，KELM 利用核函数将输入数据映射到高维特征空间，能够更好地处理非线性问题。

对于给定的样本 $(x_1,x_2,\cdots,x_n)$，KELM 的数学表达式为：

$$\boldsymbol{F}(\boldsymbol{x})=\boldsymbol{h}(\boldsymbol{x})\boldsymbol{\beta}=\boldsymbol{h}(\boldsymbol{x})\boldsymbol{H}^{\mathrm{T}}\left(\frac{\boldsymbol{I}}{\boldsymbol{C}}+\boldsymbol{H}\boldsymbol{H}^{\mathrm{T}}\right)^{-1}\boldsymbol{y}=\begin{bmatrix}K(\boldsymbol{x},\boldsymbol{x}_1)\\ \cdots\\ K(\boldsymbol{x},\boldsymbol{x}_n)\end{bmatrix}^{\mathrm{T}}\left(\frac{\boldsymbol{I}}{\boldsymbol{C}}+\boldsymbol{\Omega}_{\mathrm{ELM}}\right)^{-1}\boldsymbol{y} \tag{3}$$

其中，$K(\boldsymbol{x},\boldsymbol{x}_n)$ 表示核函数，$\boldsymbol{\Omega}_{\mathrm{ELM}}$ 表示核矩阵，$\boldsymbol{\Omega}_{\mathrm{ELM}}=\boldsymbol{H}\boldsymbol{H}^{\mathrm{T}}$。

本文选用多项式（Polynomial，Poly）核函数和 Sigmoid 核函数加权组合成混合核函数，可以更好地捕捉数据的非线性特征，并提高分类的准确性，其数学表达式为：

$$K(\boldsymbol{x},\boldsymbol{x}_n)=\tanh(a\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}_n+b) \tag{4}$$

其中，a 和 b 表示 Sigmoid 核函数的核参数，tanh 表示双曲正切函数。

1.2　XGBoost 算法

XGBoost 的目标函数由损失函数和正则化项两部分构成，其中损失函数衡量了模型预测结果与实际值之间的差异，正则化项则有助于控制模型的复杂度。其表达式为[22]：

$$\mathrm{Obj}(\Theta)=L(\Theta)+\Omega(\Theta) \tag{5}$$

XGBoost 的目标函数也可以写成：

$$\mathrm{Obj}^{(t)}=\sum_{i=1}^{n}l(y_i,\ y_i^{(t-1)}+f_t(x_i))+\Omega(f_t)+C \tag{6}$$

其中，C 表示常数项，将目标函数进行二次泰特公式展开，令 $g_i=\partial_{y^{(t-1)}}l(y_i,y^{(t-1)})$，$h_i=\partial^2_{y^{(t-1)}}l(y_i,y^{(t-1)})$，展开后的目标函数为：

$$\mathrm{Obj}^{(t)}\simeq\sum_{i=1}^{n}\left[l(y_i,\ y^{(t-1)})+g_if_t(x_i)+\frac{1}{2}h_i\ f_t^2(x_i)\right]+\Omega(f_t)+C \tag{7}$$

其中，$l(y_i,y^{(t-1)})$ 表示常数项，所以此时的损失函数为：

$$L\simeq\sum_{i=1}^{n}\left[g_if_t(x_i)+\frac{1}{2}h_i\ f_t^2(x_i)\right] \tag{8}$$

正则化项在 XGBoost 中起到控制模型复杂度的作用，且与叶子节点的个数 T 以及叶子节点权重 w 相关。这个关系可以通过如下公式来表示：

$$\Omega(f_t)=\gamma T+\frac{1}{2}\lambda\sum_{j=1}^{T}w_j^2 \tag{9}$$

综合上述分析，目标函数也可以写为：

$$\mathrm{Obj}^{(t)}\simeq\sum_{i=1}^{n}\left[g_if_t(x_i)+\frac{1}{2}h_i\ f_t^2(x_i)\right]+\gamma T+\frac{1}{2}\lambda\sum_{j=1}^{T}w_j^2=\sum_{j=1}^{T}\left[\left(\sum_{i\in v_j}g_i\right)w_j+\frac{1}{2}\left(\sum_{i\in v_j}h_i+\lambda\right)w_j^2\right]+\gamma T \tag{10}$$

其中，λ 表示正则化系数；γ 表示节点系数；v_j 表示第 j 个节点上的样本；w_j 表示节点的权重。令 $G_j=\sum_{i\in v_j}g_i$，$H_j=\sum_{i\in v_i}h_i$，对 w_j 求偏导等于 0，得到最优解为：

$$w_j^*=-\frac{G_j}{H_j+\lambda} \tag{11}$$

最后将最优解代入，得到最优目标函数如下：

$$\mathrm{Obj}^{(t)}=-\frac{1}{2}\cdot\sum_{j=1}^{T}\frac{G_j}{H_j+\lambda}+\gamma T \tag{12}$$

1.3　NGO 算法

北方苍鹰优化算法是由 Mohammad 等学者于 2021 年首次提出的，目标旨在解决多种复杂的优化问题[23-24]。在具体操作过程中，该算法遵循以下步骤：

（1）初始化阶段。在算法启动之初，会随机生成一群苍鹰个体，并分别为这些个体分配随机的初始位置和速度参数。定义公式为：

$$\boldsymbol{X}=\begin{bmatrix}\boldsymbol{X}_1\\ \vdots\\ \boldsymbol{X}_i\\ \vdots\\ \boldsymbol{X}_N\end{bmatrix}_{N\times m}=\begin{bmatrix}x_{1,1} & \cdots & x_{1,j} & \cdots & x_{1,m}\\ \vdots & & \vdots & & \vdots\\ x_{i,1} & \cdots & x_{i,j} & \cdots & x_{i,m}\\ \vdots & & \vdots & & \vdots\\ x_{N,1} & \cdots & x_{N,j} & \cdots & x_{N,m}\end{bmatrix}_{N\times m} \tag{13}$$

其中，$\boldsymbol{X}$ 表示北方苍鹰的种群矩阵；$\boldsymbol{X}_i$ 表示第 i 只北方苍鹰的位置；$x_{i,j}$ 表示第 i 只北方苍鹰在第 j 维的位置；N 表示北方苍鹰的种群数量；m 表示问题的维度。

种群的目标函数向量可表示为：

$$\boldsymbol{F}=\begin{bmatrix}F_1\\ \vdots\\ F_i\\ \vdots\\ F_N\end{bmatrix}_{N\times1}=\begin{bmatrix}F(X_1)\\ \vdots\\ F(X_i)\\ \vdots\\ F(X_N)\end{bmatrix}_{N\times1} \tag{14}$$

其中，F_i 表示第 i 只苍鹰的目标函数值。

(2)猎物识别阶段(勘探阶段)：每个个体根据其初始位置、速度，通过特定的适应度函数来评估自身的适应度。在此过程中，适应度较高的个体有更高的可能性被视作猎物。

$$P_i = X_k,$$
$$i = 1,2,\cdots,N,\ k = 1,2,\cdots,i-1,\ i+1,\cdots,N \quad (15)$$

$$X_{i,j}^{\text{new},P1} = \begin{cases} x_{i,j} + r(p_{i,j} - Ix_{i,j}), & F_{P_i} < F_i \\ x_{i,j} + r(x_{i,j} - p_{i,j}), & F_{P_i} \geqslant F_i \end{cases} \quad (16)$$

$$X_i = \begin{cases} X_i^{\text{new},P1}, & F_i^{\text{new},P1} < F_i \\ X_i, & F_i^{\text{new},P1} \geqslant F_i \end{cases} \quad (17)$$

其中，P_i 表示第 i 只北方苍鹰的猎物位置；F_{Pi} 表示对应位置的目标函数值；$X_i^{\text{new},P1}$ 表示北方苍鹰的新位置；$X_{i,j}^{\text{new},P1}$ 表示其 j 维的新位置；随机整数 k 取值范围为[1,N]；基于第 1 阶段更新后，$F_i^{\text{new},P1}$ 表示第 i 只北方苍鹰的目标函数值；随机数 r 取值范围为[0,1]；I 表示 1 或 2 的随机整数。

(3)追逐及逃生阶段(开发阶段)：根据适应度值的高低，个体将选择追逐适应度更高的个体或者逃离适应度较低的个体。通过不断更新位置和速度，个体之间实现了追逐和逃生的行为模式。研究推得的公式具体如下：

$$x_{i,j}^{\text{new},P2} = x_{i,j} + R(2r-1)x_{i,j} \quad (18)$$

$$R = 0.02\left(1 - \frac{t}{T}\right) \quad (19)$$

$$X_i = \begin{cases} X_i^{\text{new},P2}, & F_i^{\text{new},P2} < F_i \\ X_i, & F_i^{\text{new},P2} \geqslant F_i \end{cases} \quad (20)$$

其中，$X_i^{\text{new},P2}$ 表示第 i 只北方苍鹰的新位置；$X_{i,j}^{\text{new},P2}$ 表示基于第2阶段更新后第 i 只北方苍鹰的第 j 维的位置；$F_i^{\text{new},P2}$ 表示基于第2阶段更新后第 i 只北方苍鹰的目标数值；t 表示当前迭代次数；T 表示最大迭代次数。

(4)迭代与终止：以上 2 个阶段将不断循环执行，直到满足停止条件，例如达到预设的最大迭代次数或找到令人满意的优化解。

1.4 改进北方苍鹰优化算法

1.4.1 Cubic 混沌映射

北方苍鹰优化算法仍存在如低收敛精度和易受局部最优解限制等问题。为解决上述问题，本文在传统 NGO 算法的基础上引入 Cubic 混沌映射解决种群初始化不均匀，提升算法的收敛速度，Cubic 混沌映射的数学模型如下：

$$x_{n+1} = \rho x_n(1 - x_n^2) \quad (21)$$

其中，$x_n \in (0,1)$；ρ 表示控制参数，用于控制混沌性质以及非线性程度，这里 ρ 取 2.52。

1.4.2 小孔成像反向学习

对于算法中容易陷入局部最优解的情况，本文采取小孔成像反向学习策略，通过缩放因子动态调节反向解的数量优化种群数量，避免种群陷入局部最优。小孔成像的数学模型如下：

$$\frac{h}{H} = \frac{(a_i + b_i)/2 - x_{i,j}}{x_{i,j}^* - (a_i + b_i)/2} \quad (22)$$

$$x_{i,j}^* = \frac{a_i + b_i}{2} + \frac{a_i + b_i}{2m} - \frac{x_{i,j}}{m} \quad (23)$$

其中，$[a_i, b_i]$ 表示每只苍鹰的变化区间，$x_{i,j}$ 表示高度为 h 的火焰横坐标。通过小孔成像原理会在横坐标为 $x_{i,j}^*$ 的位置产生高度为 H 的倒影，其中缩放因子的公式为：

$$m = \frac{h}{H} \quad (24)$$

由此可知，正解值与反解值之差随着缩放因子 m 的减小而增大。当 $m = 1$ 时，就是一般的反向学习策略。

2 XGboost-INGO-KELM 预测模型的构建

首先，利用 XGBoost 算法计算出每个特征的重要性，将重要性较高的特征进行保留，将重要性较低和重要性为 0 的特征进行删除。然后，通过 INGO 算法优化 KELM 模型进行分类预测。XGBoost-INGO-KELM 模型的构建过程如图 1 所示，具体步骤如下：

步骤 1 给定初始数据集 W。

步骤 2 利用 XGBoost 算法降低输入特征维度、减少冗余和噪音特征，生成新的特征集 W^*。并对特征集 W^* 进行归一化处理。

步骤 3 初始化 NGO 参数。

步骤 4 Cubic 映射寻优初始化群体。

步骤 5 使用小孔成像原理反向学习策略改进种群多样性。

步骤 6 猎物识别阶段，根据式(15)~式(17)更新追捕者位置。

步骤 7 追击及逃生阶段，根据式(18)~式(20)更新追捕者位置。

步骤 8 判断迭代次数是否满足终止条件，如不满足，则重复步骤 6，反之则停止，输出最优参数。将最优正则化系数 C 和核函数参数 a、b 赋给 KELM，并进行分类预测。

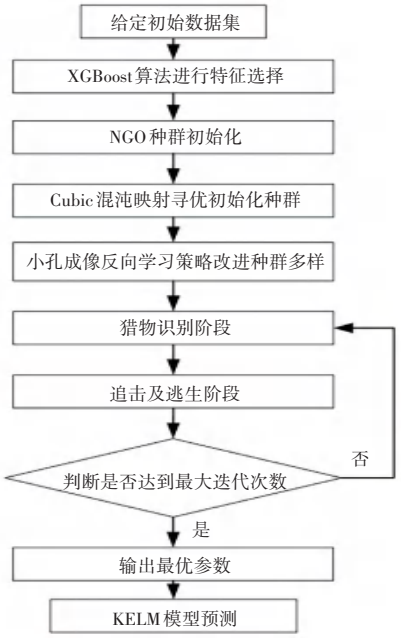


图 1　XGBoost-INGO-KELM 模型预测流程

Fig. 1　Prediction process of XGBoost-INGO-KELM model

3　实验与分析

3.1　生产数据集特征分析

为了验证提出的 XGBoost-INGO-KELM 模型的分类性能，本文利用中国浙江制丝 A 公司的制造数据进行实验分析。A 公司的生产数据集共有 192 个数据样本，13 个特征属性，依次是加料出口水分、加料出口温度、加料瞬时精度、加料料液温度、松散加水比例、累计加料量、烘前水分百分比、烘前水分 CPK、烘后水分百分比、烘后温度、风选后水分百分比、风选后水分 CPK、膨丝掺配瞬时精度。下文用 B_1 ~ B_{13} 表示。生产数据集共分为 3 类(分别代表生产质量一般、良好、很好)。由于生产数据集包含较多数据，所以本文只展示前 6 行的数据，具体见表 1。利用 XGBoost 算法计算出生产数据集中每个特征属性的重要性，删除重要性为 0 的数据后，将处理过后的数据集代入到 INGO-KELM 模型中进行预测。

表 1　生产数据集

Table 1　Production data set

Number	B_1	B_2	B_3	…	B_{11}	B_{12}	B_{13}	Category
1	19.06	44.15	18.89	…	11.67	2.82	3.20	0
2	18.57	44.15	18.88	…	11.99	3.51	3.33	0
3	18.64	44.57	18.33	…	12.07	2.71	2.82	0
4	18.56	44.95	18.34	…	12.07	3.09	3.11	0
5	18.50	44.70	18.63	…	11.87	2.71	2.72	0
6	18.46	45.07	18.76	…	11.99	3.03	3.09	0

为了能够清楚地看出生产数据集每个特征的分布情况，通过数据集特征散点图、数据集箱型图和数据集概率曲线图进行展示，如图 2~图 4 所示。

图 5 特征关系热力图显示出生产数据集中每个特征之间的相关性。

3.2　XGBoost 算法特征选择

基于对各特征情况的分析，利用 XGBoost 算法计算生产数据集各特征的重要性，删除特征重要性为 0 的特征，从而实现对数据的降维。每个特征的重要性如图 6 所示。

通过删除重要性为 0 的特征，剩余特征的程度见表 2。剩余特征还有 7 个分别为：B_2，B_4，B_6，B_{10}，B_{11}，B_{12} 和 B_{13}。删除的特征属性有 B_1，B_3，B_5，B_7，B_8 和 B_9。

3.3　多分类 XGBoost-INGO-KELM 模型的构建及求解

首先，使用 XGBoost 算法对生产数据集进行特征筛选，并将经筛选后的数据集作为训练数据。接着，利用 INGO 算法在 50 次迭代中寻找最佳正则化系数 C 和核函数参数 a、b，并将其应用于 KELM 模型进行分类预测。每次迭代都会得出一组最佳正则化系数 C 和核函数参数 a、b，当达到最大迭代次数时，将输出这组最佳参数。

构建了 XGBoost-INGO-KELM 模型后，以预测准确率作为评价指标来寻找最佳正则化系数 C 和核函数参数 a、b。当预测准确率最高时，对应的正则化系数 C 和核参数 a、b 为最优值，并与 SSA、NGO、INGO-KELM 模型进行对比，训练测试结果如图 7 所示，测试样本的混淆矩阵如图 8 所示。XGBoost-INGO-KELM 模型首先利用 XGBoost 算法对数据集进行特征筛选，然后利用 INGO 算法找出模型中的最佳参数。SSA、NGO、INGO-SKEM 利用原始生产数据集进行训练，再找出最佳参数，4 种模型求得的最优正则化系数 C 和核函数参数 a、b 数值见表 3。从图 8 中各模型的混淆矩阵可以看出，XGBoost-INGO-KELM 模型的预测准确率最高 95.8%，优于 SSA、NGO、INGO-KELM 模型，证明了本文所提模型的性能更好。

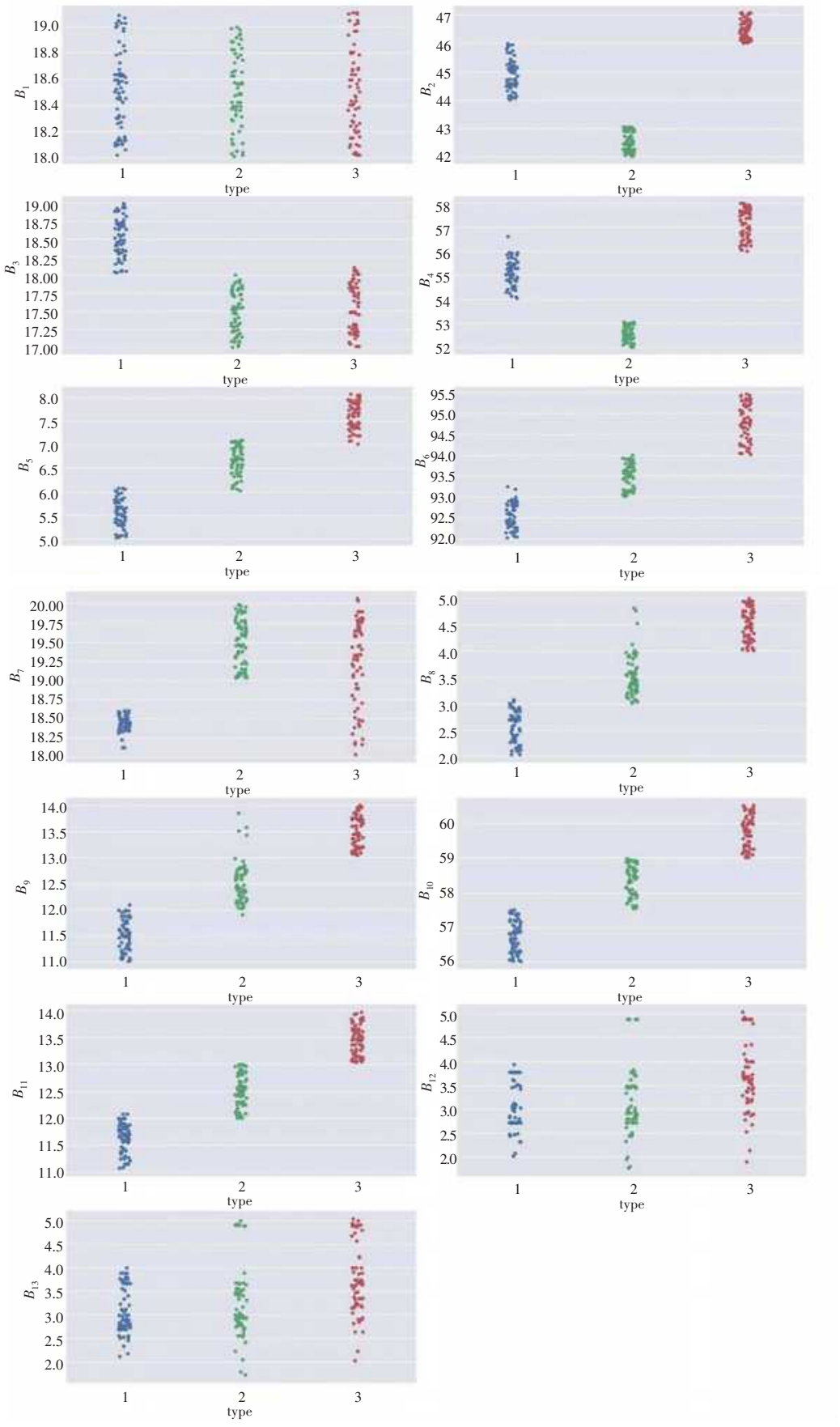


图 2 生产数据集特征散点图

Fig. 2 Scatter plot of features in the production dataset

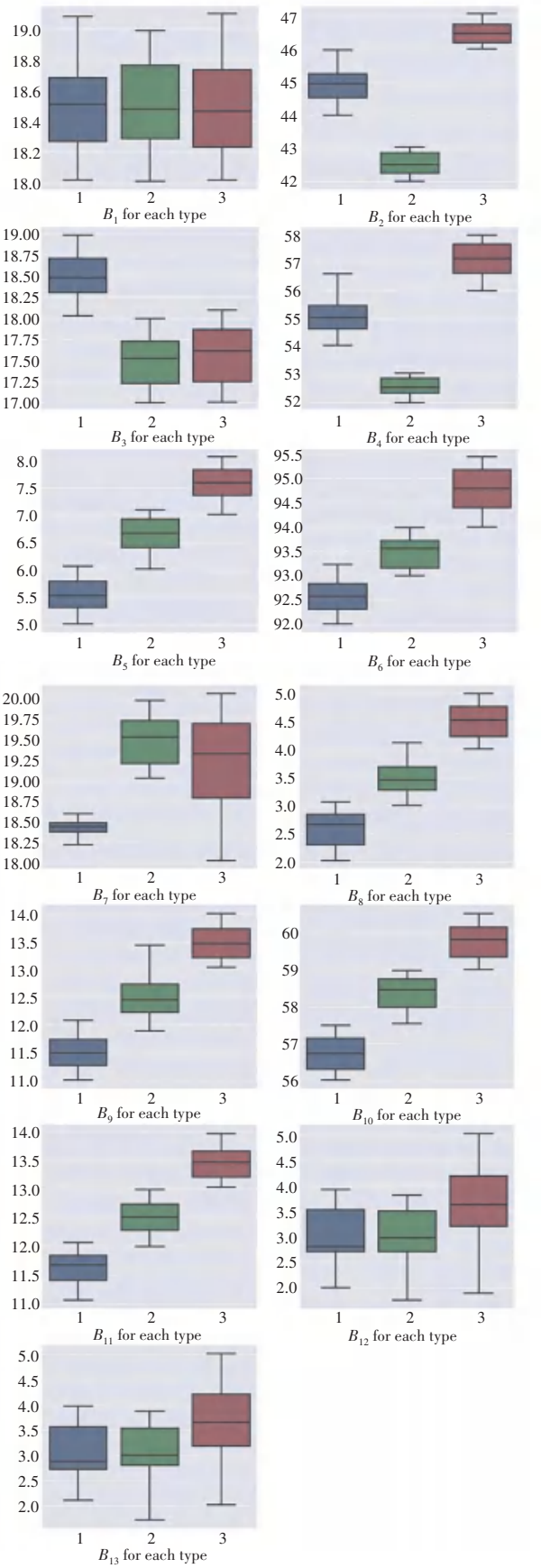


图 3　生产数据集特征箱型图

Fig. 3　Box plot of features in the production dataset

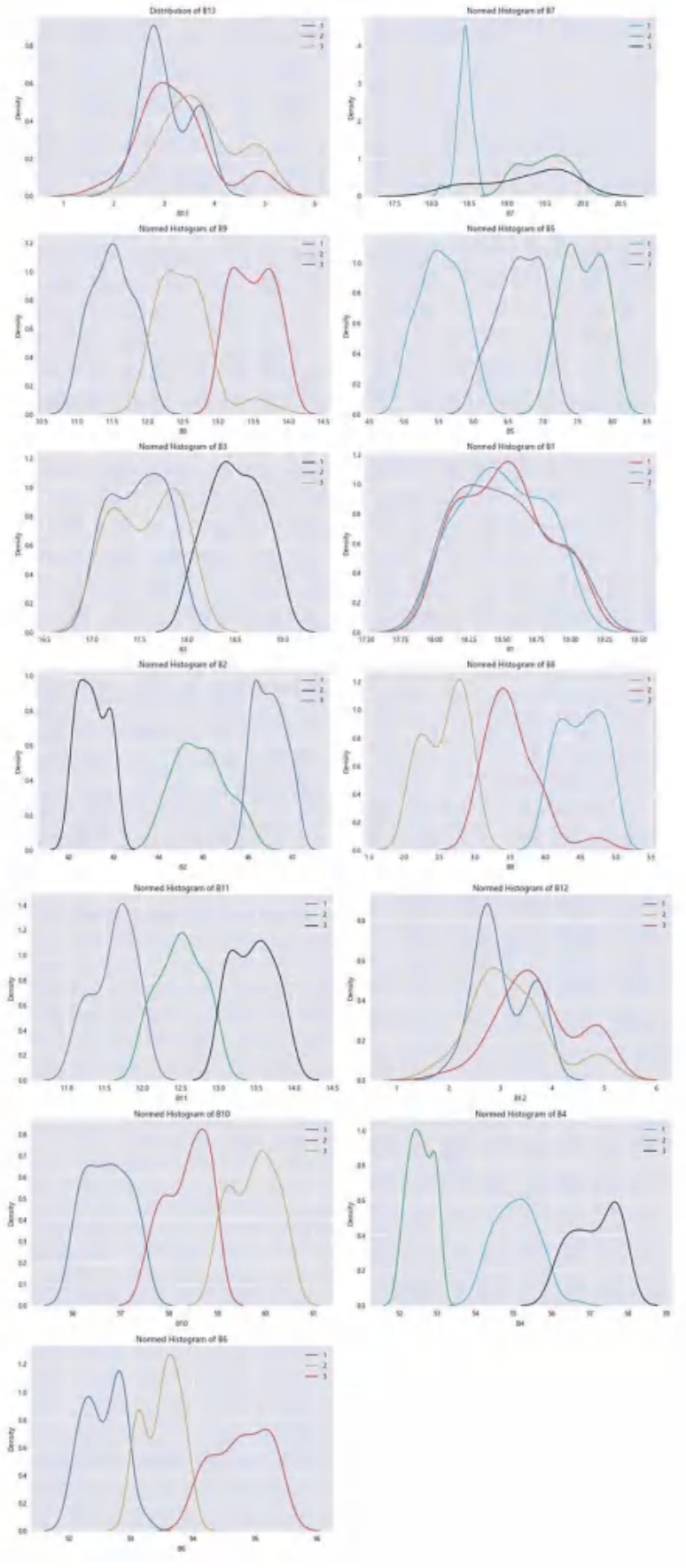


图 4　生产数据集特征概率曲线图

Fig. 4　Probability distribution curve of features in the production dataset

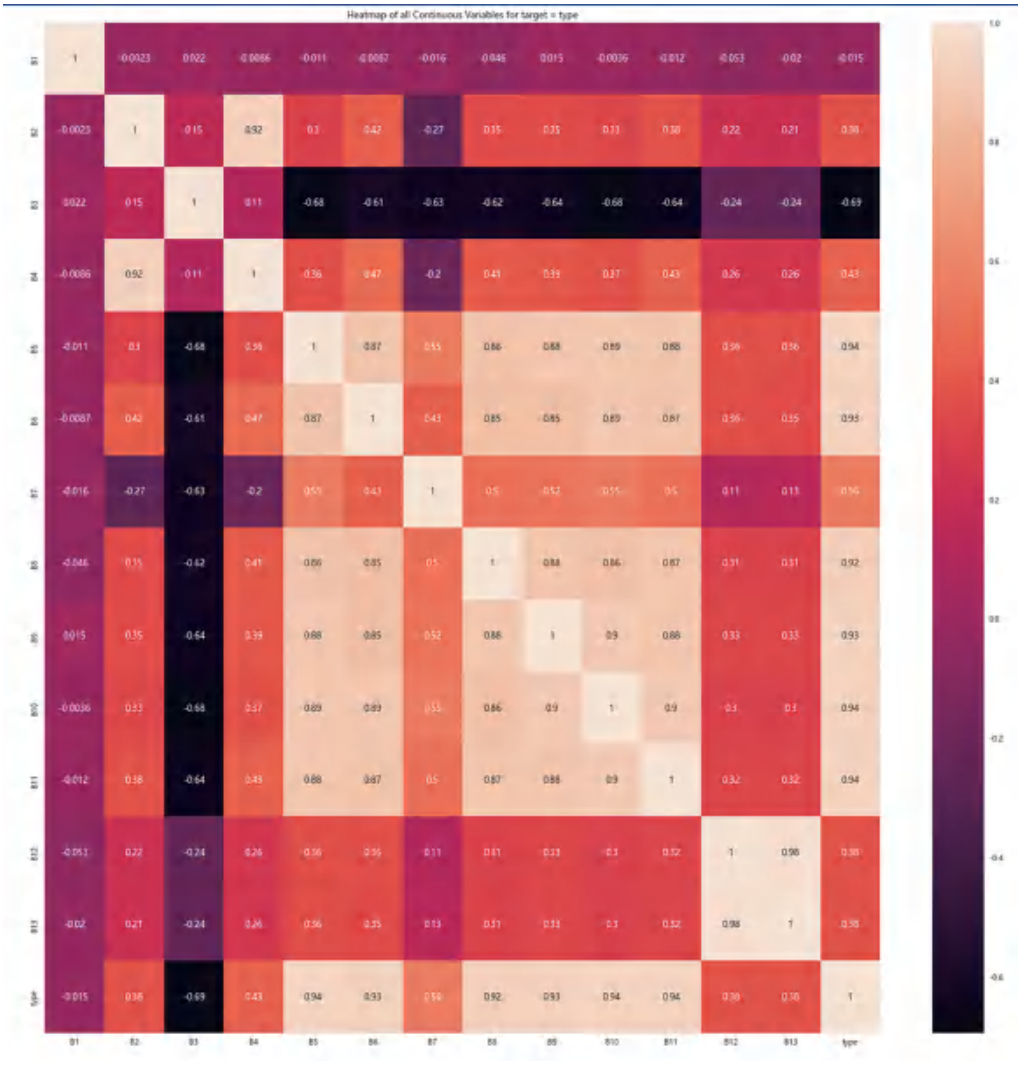


图 5　生产数据集特征关系热力图

Fig. 5　Heatmap of feature relationships in the production dataset

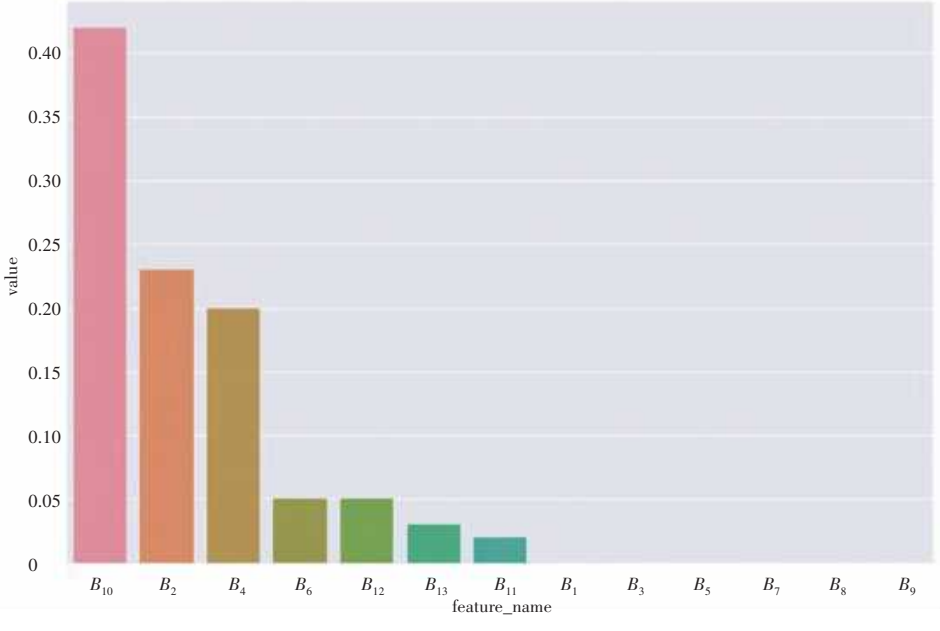


图 6　生产数据集每个特征重要度

Fig. 6　Importance of each feature in the production dataset

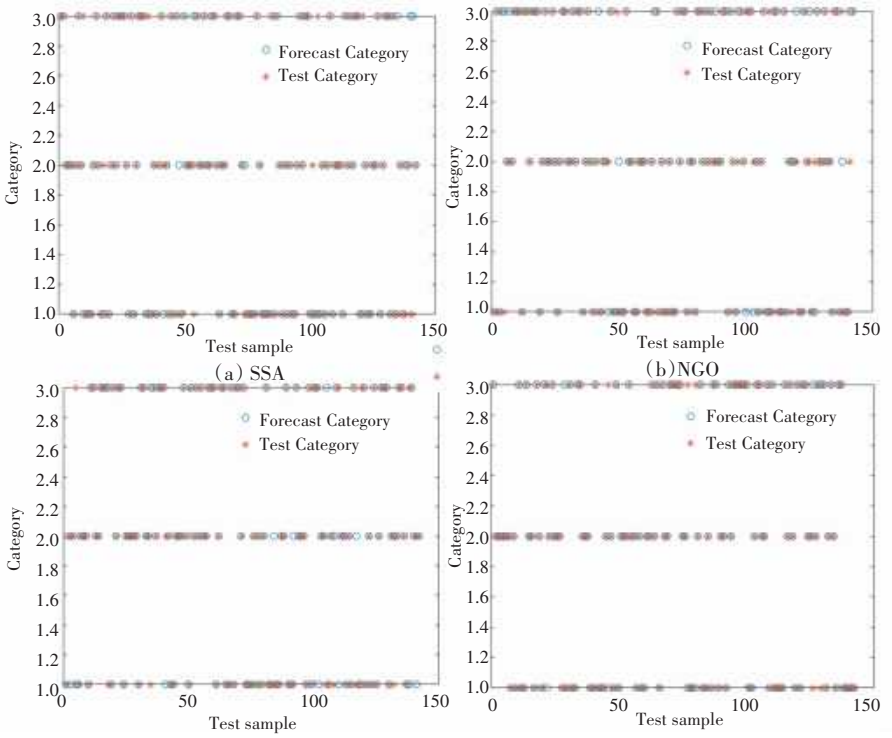


图 7 各模型的训练测试图

Fig. 7 Training test plot for each model



图 8 各模型的混淆矩阵

Fig. 8 Confusion matrix for each model

表 2　生产数据集属性的重要性

Table 2　Importance of production dataset attributes

特征	B_2	B_4	B_6	B_{10}	B_{11}	B_{12}	B_{13}
重要程度	0.23	0.20	0.05	0.42	0.02	0.05	0.03

表 3　不同模型参数及预测准确率对比

Table 3　Comparison of different model parameters and prediction accuracy

模型	正则化系数 C	核函数 a	核函数 b	预测准确率/%
SSA-KELM	1.699	1.332	0.878	84.5
NGO-KELM	4.756	0.455	0.523	87.3
INGO-KELM	7.396	0.718	0.781	90.8
XGBoost-INGO-KELM	3.122	2.136	3.121	95.8

通过仿真结果可以看出 XGBoost-INGO-KELM 模型的最优预测准确率为 95.8%，优于其他算法。说明 XGBoost-INGO-KELM 模型预测建模效果良好。

4　结束语

针对多工序工艺生产质量评估的问题，本文提出了一种基于 XGBoost-INGO-KELM 的生产质量评估模型。首先，利用 XGBoost 算法对生产数据集的特征进行降维。然后，利用 INGO 算法找出 KELM 模型中的最佳参数对。最后，将所提模型用于生产数据集进行分类预测。实验结果表明，XGBoost-INGO-KELM 模型在对生产数据集进行预测时，表现出优于 SSA-KELM、NGO-KELM 和 INGO-KELM 模型的预测效果，模型的最高预测准确率达到 95.8%，说明 XGBoost-INGO-KELM 模型性能良好，泛化能力更好。尽管如此，生产数据集的预测精度仍有提升空间，未来的研究将致力于探索更高效的模型以进一步提高预测准确率。

参考文献

[1] 阴艳超，洪志敏，顾文娟，等. 融合多通道 CNN-BiGRU 与时间模式注意力机制的多工序工艺质量预测方法[J]. 计算机集成制造系统，2025，31(8)：2905-2919.

[2] 杨龙飞. 基于机器学习的不同工艺参数对烟丝质量影响研究[J]. 电子测试，2022(16)：85-87.

[3] Huang G B，Zhu Q Y，Siew C K. Extreme learning machine：a new learning scheme of freed-forward neural networks[C]// Proceedings of 2004 IEEE International Joint Conference on Neural Networks. Piscataway，NJ：IEEE，2004：985-990.

[4] 薄青红，刘光定，张明，等. 基于 PCA 和 GA-SVM 联合模型的数控转台升降系统故障诊断[J]. 机械设计与研究，2022，38(5)：221-224.

[5] Liu L，Wei Y，Song X，et al. Fault diagnosis of wind turbine bearings based on CEEMDAN-GWO-KELM[J]. Energies，2022，16(1)：6-7.

[6] Fei S W，Liu Y Z. Fault diagnosis method of bearing utilizing GLCM and MBASA-based KELM[J]. Scientific Reports，2022，12(1)：3-4.

[7] 汤庆峰，刘念，张建华，等. 基于 EMD-KELM-EKF 与参数优选的用户侧微电网短期负荷预测方法[J]. 电网技术，2014，38(10)：2691-2699.

[8] 龙霞飞，杨苹，郭红霞，等. 基于 KELM 和多传感器信息融合的风电齿轮箱故障诊断[J]. 电力系统自动化，2019，43(17)：132-139.

[9] 胡爱军，张军华，刘随贤，等. 滚动轴承多工况故障的特征自动选择核极限学习机智能识别方法[J]. 振动与冲击，2020，39(23)：182-189.

[10] 马杰，刘晓青，黄永涛. 基于 PCA-PSO-GRU 模型的重力拱坝诱导缝开合度预测[J]. 水电能源科学，2023，41(2)：95-99.

[11] 白雲杰，贾希胜，梁庆海，等. 基于 PCA-EDT-CNN 的柴油发动机故障诊断方法[J]. 机械强度，2022，44(6)：1271-1278.

[12] 彭豪杰，周杨，胡校飞，等. 基于深度学习与随机森林的 PM2.5 浓度预测模型[J]. 遥感学报，2023，27(2)：430-440.

[13] Wang J，Rao C，Goh M，et al. Risk assessment of coronary heart disease based on cloud-random forest[J]. Artificial Intelligence Review，2023，56(1)：203-232.

[14] Yang Y，Wang K，Yuan Z，et al. Predicting freeway traffic crash severity using XGBoost-Bayesian network model with consideration of features interaction[J]. Journal of Advanced Transportation，2022，2022：4257865.

[15] Ma M，Zhao G，He B，et al. XGBoost-based method for flash flood risk assessment[J]. Journal of Hydrology，2021，598：126382.

[16] 刘傲，周正. 基于粒子群优化极限学习机的雷达识别方法[J]. 兵工自动化，2021，40(9)：1-5.

[17] 周勇，赵常辛，张德森，等. 基于 SSA—KELM 的煤与瓦斯突出预测[J]. 工矿自动化，2023，49(S2)：81-86.

[18] 朱昶胜，赵奎鹏. 改进灰狼算法的核极限学习机的风功率预测[J]. 计算机应用与软件，2022，39(5)：291-298.

[19] 李思尧，董章，陈雅旎，等. 基于 NGO-VMD-HHT 的电缆局部放电信号特征量提取方法[J]. 电气技术，2023，24(11)：35-41.

[20] 金爱兵，张静辉，孙浩，等. 基于 SSA-SVM 的边坡失稳智能预测及预警模型[J]. 华中科技大学学报(自然科学版)，2022，50(11)：142-148.

[21] 刘阳，张璐，吴德强，等. 基于 INGO-SVM 的输电铁塔地脚螺栓螺母缺失无损检测方法[J]. 高压电器，2025，61(2)：130-140.

[22] Zhu Q Y，Siew C K. Extreme learning machine：A new learning scheme of feedforward neural networks[J]. Neurocomputing，2006，70：489-501.

[23] 沈雅倩，黄志甲，周涛. 基于 XGBoost 分类算法的热舒适预测模型[J]. 安徽工业大学学报(自然科学版)，2022，39(1)：74-78.

[24] Dehghani M，Hubalovsky S，Trojovsky P. Northern goshawk optimization：A new swarm-based algorithm for solving optimization problems[J]. IEEE Access，2021，9：162059-162080.