

基于集成学习-堆叠模型的高钛渣中 TiO_2 品位预测与优化

周开敏¹, 蔡金秋², 王凯旋², 侯彦青², 侯 郊^{1*}, 王建国^{2,3}

(1. 云南驰宏锌锗股份有限公司, 云南 曲靖 655011; 2. 昆明理工大学冶金与能源工程学院, 云南 昆明 650093; 3. 云南迪庆有色金属有限责任公司, 云南 香格里拉 674408)

摘 要: TiO_2 是高钛渣的主要成分, 广泛应用于涂料、塑料、造纸、化妆品等多个工业领域。提出一种基于堆叠模型的 TiO_2 品位预测方法, 结合随机森林(RF)、梯度提升机(GBM)和支持向量回归(SVR), 采用集成学习优化高钛渣中 TiO_2 品位的预测精度。数据来自冶金厂的原始生产数据, 结合数据处理、特征衍生等手段, 通过降维技术将特征变量从 33 个减少到 15 个, 筛选出了更有价值的特征变量。试验结果表明, 堆叠回归模型在验证集和测试集上表现优异, R^2 值为 0.924 9, MAPE 值为 0.29% 和 0.30%, MSE 值为 0.177 和 0.182, 统计表现均优于单一模型。SHAP 值分析显示, TiO_2 与 FeO 质量比和 C 百分比含量等特征变量在一定范围内能够提高 TiO_2 品位。此外, 结合堆叠模型, 基于蒙特卡洛试验确定了关键特征变量的最佳范围, 如 TiO_2/FeO 比率(1.70 至 2.12)和 TiO_2/C 比率(0.50 至 0.58)。该方法避免了传统经验型方法中的配料尝试与能源浪费, 有助于节能减排, 提升钛渣的生产效率和产品质量。

关键词: TiO_2 品位预测; 集成学习; 堆叠模型; 冶金过程优化; 特征重要性分析

中图分类号: TF823, TP181

文献标志码: A

文章编号: 1004-7638(2025)05-0111-12

DOI: 10.7513/j.issn.1004-7638.2025.05.012

开放科学(资源服务)标识码(OSID):



听
语
音
与
作
者
互
动
聊
科
研

Stacked ensemble learning model-based prediction and optimization of the grade of titanium dioxide in high titanium slag

ZHOU Kaimin¹, CAI Jinqiu², WANG Kaixuan², HOU Yanqing², HOU Jiao^{1*}, WANG Jianguo^{2,3}

(1. Yunnan Chihong Zinc and Germanium Co., Ltd., Qujing 655011, Yunnan, China; 2. School of Metallurgy and Energy Engineering, Kunming University of Science and Technology, Kunming 650093, Yunnan, China; 3. Yunnan Diqing Nonferrous Metals Co., Ltd., Shangri-La 674408, Yunnan, China)

Abstract: Titanium dioxide (TiO_2), the primary component of high-titanium slag, is widely used in various industrial fields such as coatings, plastics, papermaking, and cosmetics. This paper proposed a TiO_2 grade prediction method based on a stacked model that combines random forest (RF), gradient boosting machine (GBM), and support vector regression (SVR) to optimize the prediction accuracy of TiO_2 grade in high-titanium slag through ensemble learning. The dataset was obtained from raw production data in a metallurgical plant. After data processing and feature derivation, dimensionality reduction techniques were applied to reduce the number of feature variables from 33 to 15, thereby identifying the most informative variables. Experimental results demonstrate that the stacked regression model achieves excellent performance on the validation and test sets, with an R^2 value of 0.924 9, MAPE values of 0.29% and

收稿日期: 2025-07-25; 修回日期: 2025-09-01; 接受日期: 2025-09-01

基金项目: 国家自然科学基金项目(22168019, 52474380), 云南省重大科技项目(202202AB080014, 202102AD080005, 202202AD080006)。

作者简介: 周开敏, 1969 出生, 男, 云南曲靖人, 硕士, 正高级工程师, 长期从事冶金环保研究工作, E-mail: 1683976251@qq.com; * 通讯作者: 侯郊, 1984 年出生, 男, 湖南岳阳人, 硕士, 高级工程师, 长期从事冶金环保研究工作, E-mail: 86334663@qq.com。

0.30%, and MSE values of 0.177 and 0.182, respectively—outperforming individual models. SHAP value analysis further revealed that feature variables such as the TiO_2/FeO mass ratio and C content positively influence the TiO_2 grade within certain ranges. Moreover, by integrating the stacked model with Monte Carlo simulations, the optimal ranges of key feature variables were determined, such as the TiO_2/FeO ratio (1.70–2.12) and the TiO_2/C ratio (0.50–0.58). This approach avoids the trial-and-error process and energy waste associated with traditional empirical methods, thereby contributing to energy conservation and emission reduction while enhancing the production efficiency and product quality of titanium slag.

Key words: TiO_2 grade prediction, ensemble learning, stacked model, metallurgical process optimization, feature importance analysis

0 引言

二氧化钛 (TiO_2) 是高钛渣的主要成分, 由于其优异的物理和化学性质, 二氧化钛被广泛应用于许多工业领域^[1-5], 其在太阳能电池和光催化剂领域也展示出了巨大的潜力^[6-7]。由于高钛渣中 TiO_2 的品位将影响上游产品的质量和应用效果, 因此提高渣中 TiO_2 的提取和品位已成为工业生产中的一个重要挑战。

传统上, 改善高钛渣中 TiO_2 品位的方法主要依赖于经验配料和改进工艺。LÜ 等人^[8] 提出了一种通过低温还原和分离钛铁矿制备高钛渣的方法, 该方法能够在较低的温度下完成还原过程, 从而提高高钛渣的品位和金属回收率。DONG 等人^[9] 则通过添加 Na_2CO_3 对矿物进行焙烧, 实现了矿物相的重建。尽管上述成功地从高钙镁钛渣中获得了高品位的 TiO_2 , 但该过程涉及复杂的化学过程同时难以工业化。KONDORI 等人^[10] 通过一系列探索性试验, 采用化学分析、X 射线衍射、扫描电子显微镜和能量色散光谱对样品进行了表征, 但其大部分 TiO_2 品位仍然低于 90%。

尽管上述在一定程度上有效, 但它们通常面临高成本、低效率和环境污染等问题。因此, 寻找更环保且具成本效益的方法是一个重要方向。同时, 高钛渣中 TiO_2 品位的确定依赖于精确的化学和物理检测技术。例如, QU 等^[11] 人通过 X 射线荧光光谱 (XRF) 精确测定了渣中的 TiO_2 含量。WU 等人^[12] 也通过 XRF 测定了高钛渣的主要和次要元素。ZHENG^[13] 等人则使用电感耦合等离子体质谱法测定了钛铁矿和钛磁铁矿中的钛含量。这些方法能够提供相对准确的测量结果, 但在时间、成本和操作难度方面都有一定的不足。值得注意的是, 以上方法都需要对已经生产出的样品进行取样和测试, 而在当今工业智能化的总体趋势下, 这些方法在快速和实时决策中的应用存在一定限制。

近年来, 机器学习等人工智能技术在工业预测中的应用越来越广泛。例如, CAO 等人^[14] 通过混合

策略改进了麻雀搜索算法, 优化了支持向量机模型 (ISSA-SVM), 并将该模型应用于宁夏工业用水预测, 展示了机器学习在大规模数据处理中的潜力。LI 等人^[15] 基于因子分析方法定义了高炉状态, 并使用 AdaBoost 算法进行预测, 建立了大数据平台的高炉综合状态评估和预测系统。WANG 等人^[16] 提出的 PixelRNN 模型在复合材料的成型应用中预测准确度达到了 97.53%, 有助于解决复合材料制造过程中不良的随机问题。CAO 等人^[17] 提出了 CRIM 模型促进了激光粉末床熔融 Ti6Al4V 合金的参数优化和微观结构定制, 并增强了强度-延展性协同作用。GUNASEGARAM 等人^[18] 通过机器学习辅助的 CLC 在增材制造中进行缺陷和异常控制。HE 等人^[19] 使用机器学习方法预测高熵合金相结构, 试验结果与预测结果一致。BAI 等人^[20] 对热模拟数据和微观结构特征进行分析, 采用 BP-ANN 方法建立了流动应力和位错密度预测模型, 预测精度在 90% 以上。HOSSAIN 等人^[21] 探索了一种预测高温奥氏体不锈钢应力开裂的机器学习方法。NAUTIYAL 等人^[22] 将机器学习方法应用于石油和天然气行业, 为石油上游工艺数据的智能预测提供了新的方向。在材料强度预测领域, 施等人开发了一种基于机器学习的模型, 预测 V-N 钢热轧板的屈服强度, 展示了机器学习在处理复杂工业数据方面的有效性^[23]。此外, ÁLVARO 等人^[24] 利用人工智能技术优化了钢铸造行业的操作系统, 提高了吹氧过程的效率。

尽管机器学习模型在多个领域的应用已经取得了广泛进展, 但在钛渣中 TiO_2 品位预测以及涉及无烟煤和钛铁矿混合物的工业试验中, 其潜力仍未得到充分挖掘。因此, 引入了一种创新的集成学习框架^[25], 该框架结合了支持向量回归 (SVR)、随机森林和梯度提升等基础模型, 并采用线性堆叠作为元模型。通过利用这些不同模型的优势, 该框架旨在提高预测精度并优化整体模型性能。文章强调了集成学习方法在预测高钛渣中 TiO_2 品位方面的表现, 通过对工厂提供的 100 个生产周期数据集进行细致地统计分析和数据清洗, 结合 SHAP 解释的堆叠模

型的应用,并通过进行 5 000 次蒙特卡洛试验,生成了一个大型的数据集,确定了关键成分的最优范围,蒙特卡洛仿真与堆叠模型的协同作用使我们能够充分考虑工业过程中的测量误差、数据噪声和各种不确定性,为 TiO₂ 品位预测和物料配比的优化提供了新的思路。这一方法不仅提高了 TiO₂ 品位预测的准确性,还提供了更具解释性的可视化指标,从而减少了能源浪费,推动了更高效、环保的工业过程。

1 方法论与模型

1.1 数据清洗与预处理

实际冶金生产场景当中,工厂数据的多样性、不一致性以及脏数据都会导致后续决策的精准度下降与误操作。因此,采用了插值、删除和相关特征估计的方法,并使用归一化技术将数据在同一尺度上进行处理比较,此外使用 3σ 原则将超出 3 倍标准差范围的数据视为异常值,进行可能的更正和删除。图 1 展示了原始数据与数据处理后的箱线图,清晰地反映了数据处理前后的分布差异。

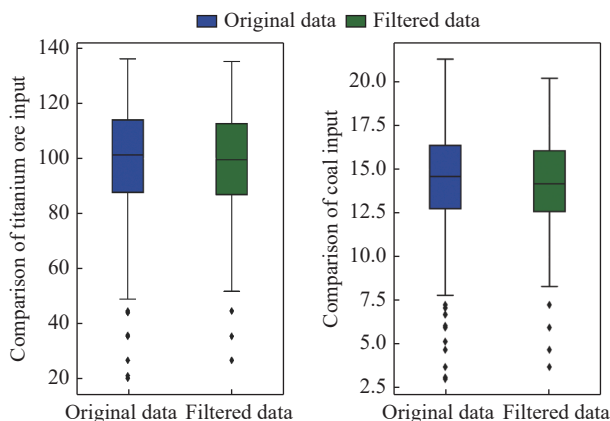


图 1 原料输入的箱线图比较
Fig. 1 Boxplot comparison of raw material input

在对原始数据进行清洗和预处理的基础上,确保了数据的准确性和分布一致性,最终成功优化了数据集的质量,为后续分析提供了可靠的数据支持。

1.2 特征变量选择

对数据集进行了特征衍生操作,使用特征比例衍生、二项式衍生等方法,从单一的各个原材料含量输入,到组合原始特征与数学变换得到了新的特征变量。数据集从原始特征扩展到了 33 个特征变量,丰富了特征空间,补充了潜在的非线性特征。

随后,使用主成分分析(PCA)对这 33 个特征变量进行降维处理,对数据进行标准化处理,计算协方差矩阵分解得到特征值和特征向量,根据累计解释方差比例确定保留主成分数量,通过分析原始特征在主要主成分上的载荷大小来识别最具信息量的特

征,将数据压缩至有效的较低维度。PCA 通过提取主要成分(图 2),减少了特征空间的复杂度,同时保留了数据中的关键信息。在 PCA 降维后,最终选择了 15 个具有较高解释性的特征变量用于后续建模分析。

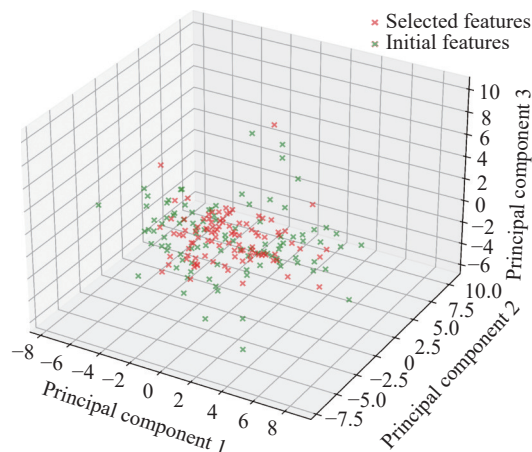


图 2 3D PCA 散点图中特征
Fig. 2 Features in a 3D PCA scatter plot

1.3 机器学习模型和堆叠模型

单一模型在其他领域的有效性如表 1 所示,这些模型在工业生产性能和指标预测的工业建模研究中得到了广泛的应用。因此,研究采用了堆叠模型(stacked model),这是一种集成学习技术^[26-27],通过将多个机器学习模型结合来提高目标值的预测能力。其堆叠模型使用两层架构,其中第一层由多个基础学习器组成,独立学习数据并提供预测结果;随后将这些预测作为输入传递到第二层。第二层是一个线性元模型,它进一步学习和训练第一层的输出,形成最终的综合预测模型(图 3)。

文中选择了三种基础学习器:随机森林(RF)、梯度提升机(GBM)和支持向量机(SVR)^[28-30]。每个基础学习器在其特定能力范围内提供不同数值的预测结果,通过集成框架能够实现更准确的综合预测。不同的单一机器学习模型在特定应用场景中的优缺点有所不同。如表 2 所示,线性回归具有良好的可解释性和简单的模型结构,但在处理非线性数据时效果较差。随机森林在预测准确性上表现优异,但其模型复杂且训练时间较长。支持向量机在处理高维数据和复杂分类问题时具有出色表现,但内核函数的选择与长时间训练的问题也较为棘手。梯度提升机具有高预测准确性和对异常值的鲁棒性,但在高维稀疏数据上可能表现不佳,并且容易过拟合。电热法生产高钛渣提高 TiO₂ 品位的关键在于严格控制低价钛与高价钛之比小于 0.1,主要通过四种方法实现:一是采用沥青粘合剂压制的团料入炉,形成

"乌龟壳"状结构,使还原过程早于成渣过程,提高铁氧化物金属化率,减少低价钛生成;二是精确控制还原剂过碳量,避免还原不充分或过还原导致炉底沉积问题;三是实施两段熔炼法,先通过隧道窑等预处理使铁氧化物还原成金属铁,提高炉料电导率,加快升温速率,缩短冶炼周期,降低电耗;四是采用吹氧出炉放渣辅助工艺,使部分低价钛重新氧化成高价

钛,降低炉渣熔点并提高温度,显著改善高品位钛渣($\text{TiO}_2 > 75\%$)的流动性,解决出炉困难问题,从而有效提高 TiO_2 品位、降低能耗、提高产量^[31]。机器学习方法通过数据驱动的智能决策,能更全面、精准地控制电热法生产高钛渣过程中的复杂物理化学变化,突破传统方法的经验局限,实现更高 TiO_2 品位、更低能耗和更稳定生产的综合优化目标。

表 1 工业生产性能和指标预测的机器学习建模研究总结

Table 1 Summary of machine learning modeling studies for prediction of industrial production performance and metrics

ML models	Target	Number of samples	Input variables	Reference
ANN, SVM, DT	Estimation oil production performance of LSWI core flooding	117	Petrophysical properties, oil viscosity, oil density, residual oil saturation, temperature, brine properties	[32]
ANN, DT, ERT, GB, RF, EXBoost	Estimation the CO_2 foam strength	157	Shear rate, temperature, pressure salinity, surfactant concentration foam quality	[33]
SVM, XGBoost, MLP, RF	Strip crown	1 809	Strip width, slab thickness, exit thickness, entrance temperature, exit temperature, rolling force, rolling speed, strip yield strength, bending force, rolling shifting, roll diameter roll thermal expansion, roll wear	[34]
Linear SVR, ANN Gaussian SVR	The hotspot temperature	16	Mould temperature, melt temper-ature, holding time, holding pressure, shot size, switch-over position, injection speed, and cooling time	[35]
Stacked, BPNN, LSTM	Underground reservoir pressure	4 000	Spatio-temporal data, Spatial coordinates and time information	[36]
RFR, KNN, ANN, SVR	Tempered martensite hardness	1 900	Tempering temperature, C, tempering time, Cr, Si, Mn, Mo, Ni on HV	[37]
ANN	Optimization of water alternating CO_2 gas	166	Injection rate, production rate limit, start of depletion, end of depletion, average pressure	[38]
PSO-SVM, KNN, RF	Crop yield	1 200	Weather, soil fertility, water availability, water quality, crop pricing	[39]

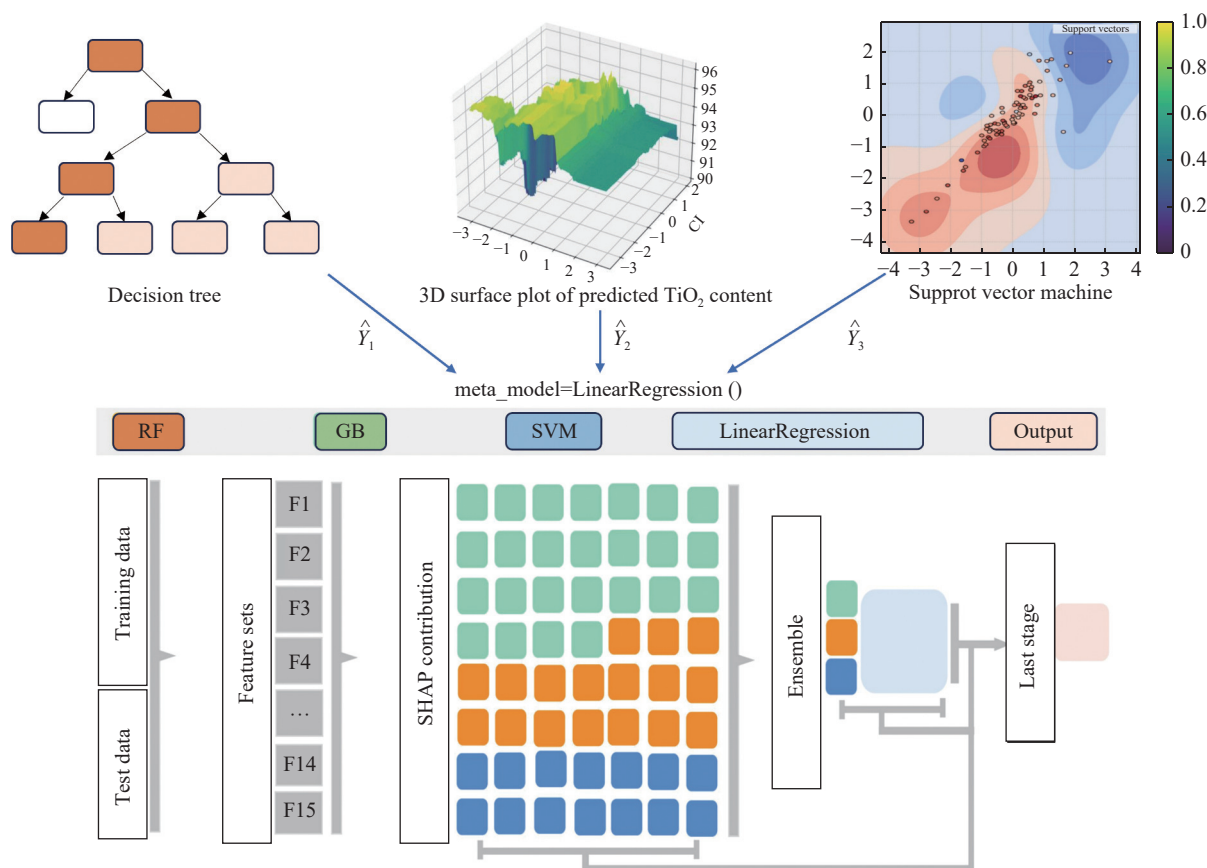


图 3 堆叠模型架构

Fig. 3 Stacked model architecture flowchart

堆叠模型则是通过结合不同的模型, 训练多个基础模型和元模型, 能够减少单个模型的局限性影

响。通过大量的训练优化与参数调整确保了最终模型能够提供最佳的预测效果。

表 2 机器学习模型优缺点比较
Table 2 Comparison of pros and cons of machine learning models

Methods	Pros	Cons
Liner Regression	Simple, easy to implement, highly interpretable	Assumes linear relationship between variables, not suitable for nonlinear data, sensitive to outliers
Random Forest	Generally high accuracy, can handle large amounts of data, reduces the risk of overfitting	Larger model size, long training time, lower interpretability
Support Vector Machine	Effective in handling high-dimensional data, suitable for complex classification problems, insensitive to feature scaling	Long training time, difficulty in choosing the appropriate kernel function, sensitive to parameter selection
Gradient Boosting	High accuracy, robust to outliers, can handle data with complex relationships	High computational cost, prone to overfitting, complex parameter tuning, not suitable for high-dimensional sparse data
Logistic Regression	Suitable for binary classification problems, provides probabilistic output, easy to understand and implement	Sets linear boundaries, limited ability to fit complex or nonlinear relationships
Stacked	Combines the strengths of different models, reduces the bias of a single model, flexible model structure	Complex model structure, multiple models such as base models and meta-models need to be trained

1.4 流程架构

采用关键输入变量, 如 TiO_2/FeO 比率($\text{TiO}_2_F\text{eO_Ratio}$)、 $\text{TiO}_2/\text{Fe}_2\text{O}_3$ 比率($\text{TiO}_2_F\text{e}_2\text{O}_3_Ratio$)和 TiO_2/C 比率($\text{TiO}_2_C_Ratio$)等, 同时对数据集进行了划分。通过统计分析和模型效果评估, 优化参数并调整模型。成功利用堆叠模型并结合蒙特卡洛试验, 确定了关键特征变量的最优取值范围, 从而确立了整个流程架构。图 4 展示了这些关键过程在文中的参与情况。

2 结果与讨论

2.1 数据集的统计分析

通过分析钛渣生产中原料含量和 TiO_2 品味数据, 整理了钛铁矿和无烟煤的原料统计数据, 统计结

果如表 3 所示, 其中 m_i 表示组分 i 的输入量(单位: t), i 为 TiO_2 , C ; w_j 表示钛矿石中对应元素 j 的百分比含量(单位: %), j 为 TiO_2 , Fe , FeO , Fe_2O_3 , SiO_2 , MgO , CaO ; w_c 表示煤炭中碳含量(单位: %), $w_{(\text{TiO}_2)}$ 表示炉渣中的二氧化钛含量(单位: %)。为了进一步探讨原料成分对目标值的影响, 通过热图分析了原料化学成分与 TiO_2 含量之间的关系(图 5), 热图分析结果表明, 在一定比例的原材料投入下, 炉渣中 TiO_2 的百分比与原料的总投入呈现出很强的相关性, 相关系数高达 0.97。与此同时, 热图还显示, SiO_2 和 CaO 与目标值(即炉渣中的 TiO_2 百分比)之间没有显著的相关性。因此, 在特征衍生阶段, 可以适当降低这两种成分的权重, 优先选择与目标值具有较高相关性的特征变量, 以优化模型的预测性能。

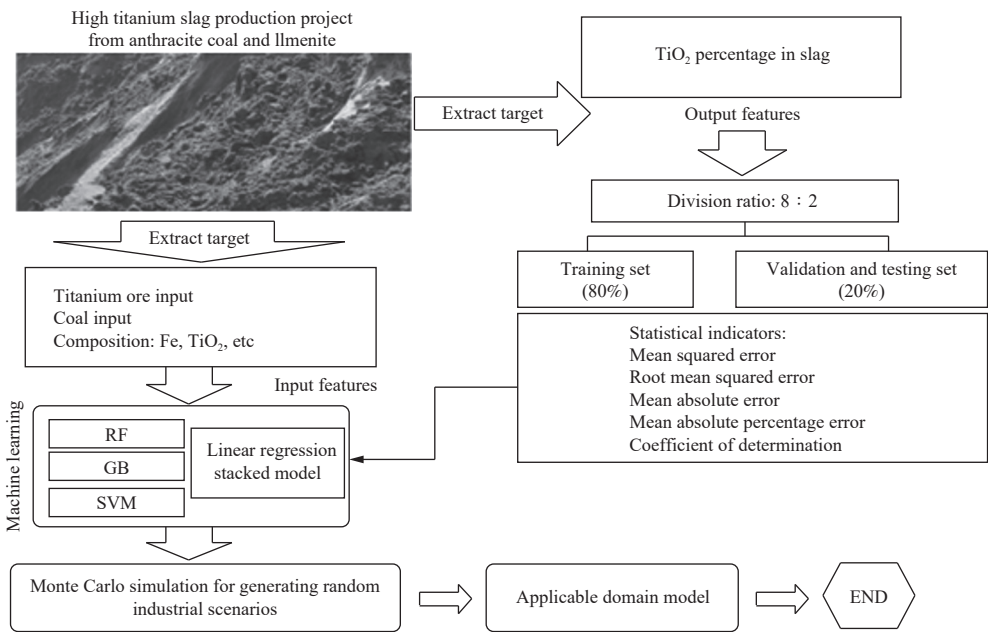


图 4 流程架构

Fig. 4 Workflow architecture

表 3 工厂数据统计
Table 3 Statistical parameters of factory raw data

Statistical parameter	$m_{\text{TiO}_2}/\text{t}$	m_{C}/t	$w_{\text{TiO}_2}/\%$	$w_{\text{Fe}}/\%$	$w_{\text{FeO}}/\%$	$w_{\text{Fe}_2\text{O}_3}/\%$	$w_{\text{SiO}_2}/\%$	$w_{\text{MgO}}/\%$	$w_{\text{CaO}}/\%$	$w_{\text{C}}/\%$	$w_{(\text{TiO}_2)}/\%$
Mean	99.89	14.39	48.51	34.94	25.92	21.13	1.03	0.65	0.04	90.50	94.48
Standard	19.74	2.92	0.73	1.00	1.31	1.82	0.37	0.09	0.04	1.92	2.99
Minimum	20.04	3.02	46.50	31.97	19.68	16.58	0.07	0.03	0.01	83.71	83.09
Maximum	136.08	21.31	49.93	47.33	30.04	34.33	2.54	1.42	0.52	93.40	99.61
Median	101.19	14.6	48.65	34.98	25.86	20.93	0.96	0.66	0.03	91.07	94.79

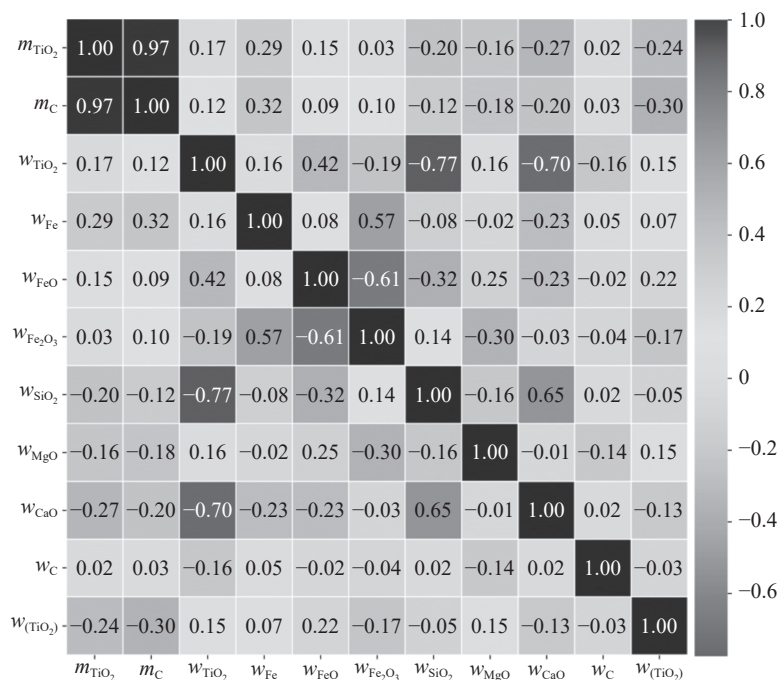


图 5 工厂数据热力图(相关系数)
Fig. 5 Heatmap of correlations in factory raw data(correlation coefficient)

2.2 原料的化学反应及特征变量分析

分析了钛铁矿和无烟煤之间的化学反应,其中钛铁矿作为主要原料,无烟煤作为还原剂。特征变量分析表明,钛铁矿的化学成分和无烟煤的纯度对于提高高钛渣的生产效率和最终产品质量至关重要。在高钛渣生产过程中,主要任务是最大限度地去除杂质元素,如铁、钙和镁,同时专注于钛的提取。核心的化学反应可概括为: $\text{FeTiO}_3 + \text{C} \rightarrow \text{TiO}_2 + \text{Fe} + \text{CO}$ 。在此过程中,钛铁矿的主要化学成分包括 TiO_2 、 FeO 、 Fe_2O_3 、 SiO_2 、 MgO 和 CaO ,这些组分的总量会影响冶炼过程,且适当的成分比例对最终产品的质量有直接影响。

结果强调了各种原材料投入的比例对提高 TiO_2 提取率的重要作用。热图分析还揭示了 TiO_2 的质量百分比与钛铁矿中 SiO_2 的百分比之间存在显著的负相关关系(相关系数为-0.77)。基于上述发现,原始数据基础上生成了 33 个特征变量,这些特征变量包括原始化学成分及其衍生的多重比率特征。

通过绘制特征的散点图矩阵(图 6)进一步分析了新特征与目标值之间的关系,揭示了数据的分布特征。特别地, TiO_2 与 Fe_2O_3 的比率($\text{TiO}_2\text{Fe}_2\text{O}_3\text{Ratio}$)与目标值之间呈现出正相关关系,进一步验证了这些特征在预测 TiO_2 提取效率中的重要性。然而,钛铁矿中 SiO_2 与 MgO 的比率($\text{SiO}_2\text{MgORatio}$)与目标值的关系较弱,表明该特征对炉渣中 TiO_2 含量影响不明显或可能受其他潜在因素的影响。分析选定的 15 个特征(图 7)表明,“Fe”的重要性最高,重要性系数为 0.45,其次是“ $\text{TiO}_2\text{FeORatio}$ ”(0.30)和“ $\text{Fe}_2\text{O}_3\text{Weight}$ ”(0.25)。这些特征在预测模型中发挥了重要作用。此外,“C_Weight”、“C1”和“ Fe_2O_3 ”等特征也显示出有效的权重,表明 TiO_2 与二价铁(FeO)和三价铁(Fe_2O_3)的比率,以及还原剂和杂质的含量,均能显著影响钛白粉的产量和品质。

另一方面,某些特征的重要性较低,如“ $\text{TiO}_2\text{MgORatio}$ ”(0.04)、“MgO”(0.05)和“Mg_Weight”(0.05)。这表明这些特征对目标值的解释作用较小,

可能在实际应用中对预测模型贡献有限。基于这些发现, 优化关键原材料比例成为提高钛渣生产品质的关键因素。

通过上述分析, 最终确定的 15 个特征变量显示出更为集中且一致的数据分布, 反映了它们在解释目标变量方面的高度相关性和有效性。

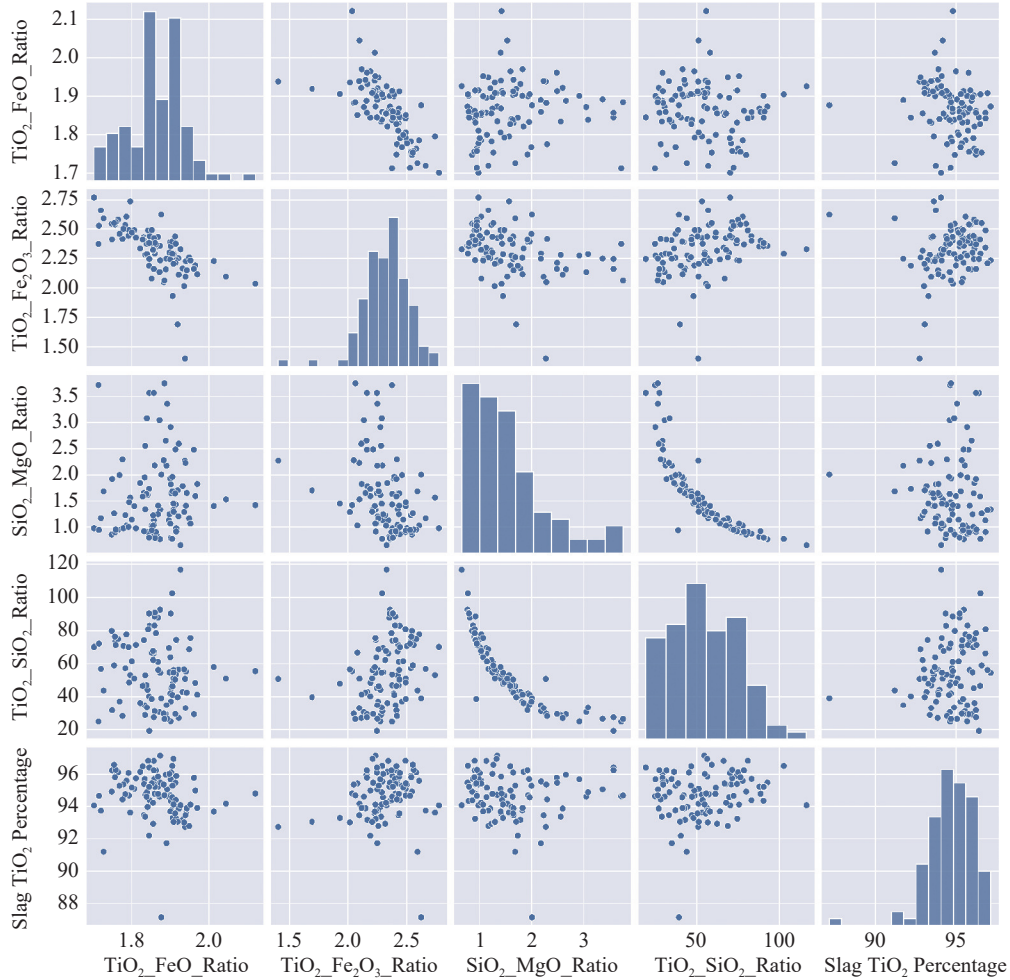


图 6 特征与目标值散点图矩阵

Fig. 6 Partial feature and slag TiO_2 scatter plot matrix

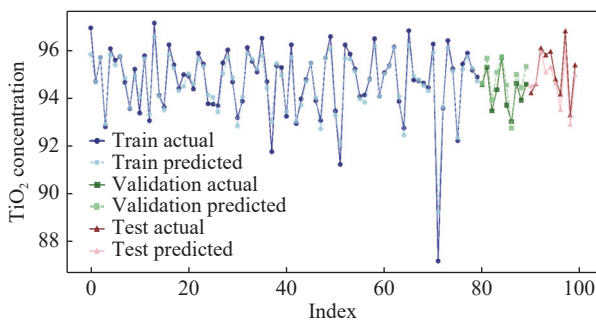


图 7 堆叠模型拟合曲线

Fig. 7 Fitting curve of the stacked model on the dataset

2.3 机器学习模型优化与验证

模型使用 80% 的数据作为训练集, 剩余 20% 的数据用于验证集和测试集, 并在模型训练期间, 监控模型的性能以评估其拟合度和泛化能力。

采用传统的性能指标来评估模型的表现, 包括

均方误差(MSE)、根均方误差(RMSE)、平均绝对误差(MAE)、平均绝对百分比误差(MAPE)和决定系数(R^2)。这些指标的数学定义如式(1)~(5)所示。

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (1)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2} \quad (2)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |Y_i - \hat{Y}_i| \quad (3)$$

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right| \quad (4)$$

$$R^2 = 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}} \quad (5)$$

其中 Y_i 是第 i 个观测值的实际响应值, $\hat{Y}_i$ 是模型预测的响应值, n 是样本点总数, SS_{res} 是残差平方和, ss_{tot} 是总平方和。

在模型构建过程中,使用了堆叠模型(Stacked Model),集成了三个基础模型:随机森林(RF)、梯度提升机(GBM)和支持向量机(SVM)。元模型则采用了线性回归。模型的超参数通过 Grid Search 方法进行了优化,具体的参数设置见表4。

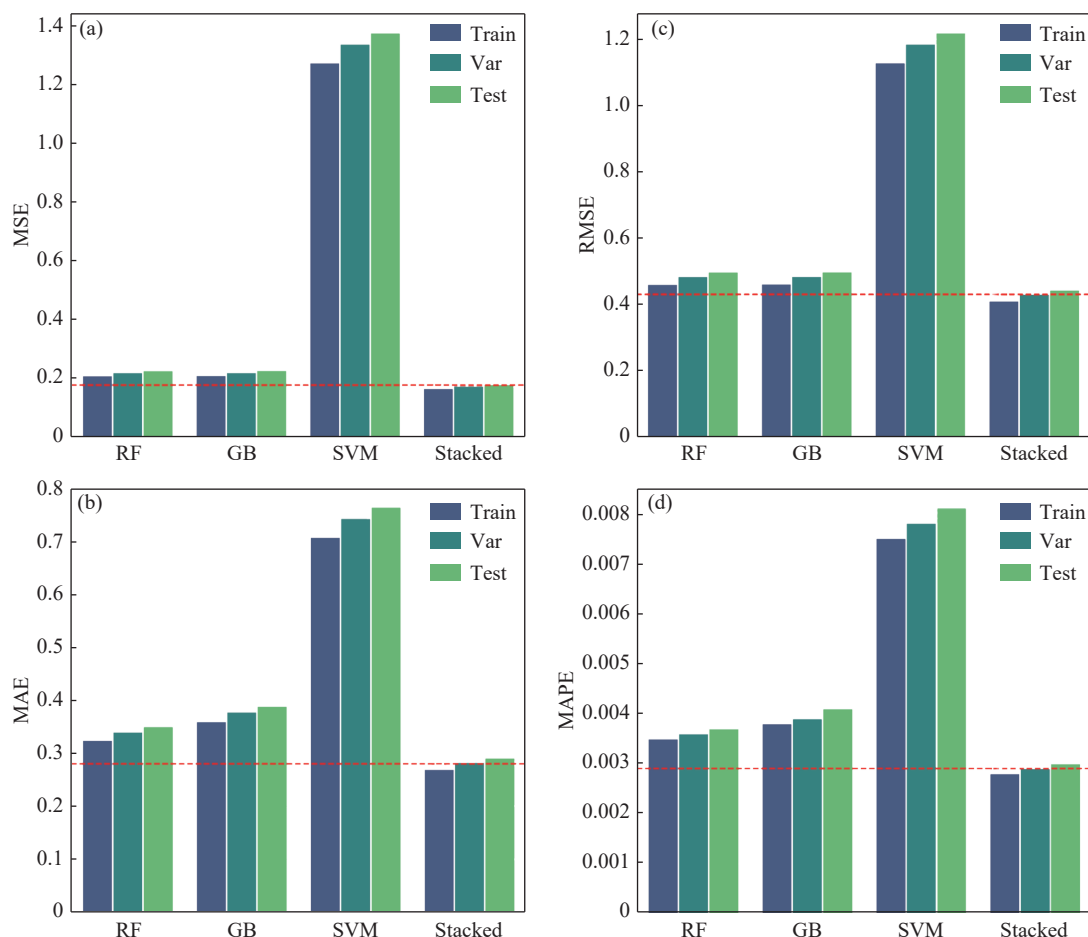
通过这一综合方法,成功地协同多个模型,并进行了统计分析和目视检查。图7显示了堆叠模型的预测曲线与实际数据曲线紧密吻合,说明该模型成功学习了关键特征并有效预测了目标值。观察统计指标(图8)后发现,训练集的 R^2 值达到0.9249,验证集和测试集的表现也优于单一模型,尤其在MAPE和MSE指标上,堆叠模型的表现非常出色。在验证集和测试集上的MAPE值分别为0.29%和

0.30%,MSE分别为0.177和0.182,表明堆叠模型的预测误差维持在非常小的范围内,符合实际工程需求。

表4 超参数调优表

Table 4 Hyperparameter tuning for this study

Model	Parameters	value
RF	n_estimators	400
	random_state	42
	max_depth	None
	min_samples_split	2
GBM	learning_rate	0.1
	max_depth	6
	n_estimators	200
	random_state	42
SVM	C	1.0
	kernel	Rbf
	gamma	Scale
	epsilon	0.1
Stacked	Base Estimators	RF GB SVM
	Meta Model	Linear Regression



(a) 均方误差 (MSE); (b) 平均绝对误差 (MAE); (c) 均方根误差 (RMSE); (d) 平均绝对百分比误差 (MAPE)

图8 四种机器学习模型的统计指标比较

Fig. 8 Comparison of statistical indicators for the four ML models

此外,通过泰勒图(图9),可以直观地观察到每个模型的预测性能差异。堆叠模型在多个指标上均

优于其他模型,展示了最佳的整体准确和实用性。总体而言,堆叠模型是此项工作的首选预测模型,但

随机森林、梯度提升机和支持向量回归也表现出良好的性能。与传统的冶金理论模型相比, 其存在黑箱问题, 但机器学习方法通过数据驱动能捕捉复杂的非线性关系, 显著提高预测精度。这些结果为集成学习模型在相关领域和工业应用中提供了广泛的应用潜力。

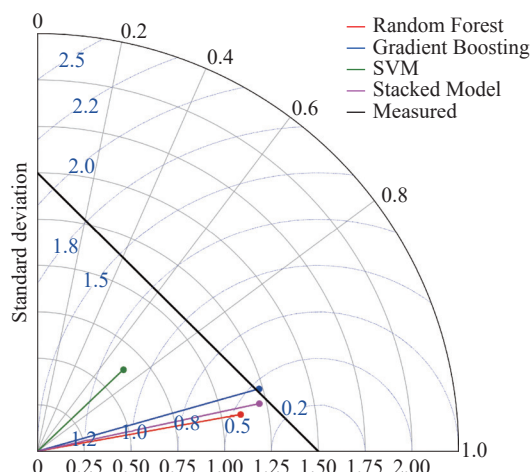


图 9 模型预测性能的泰勒图

Fig. 9 Taylor diagrams of model predictive performances

2.4 SHAP 分析

此研究不仅成功利用堆叠模型预测了钛渣中 TiO_2 的品位, 而且计算了各个特征的加权平均贡献度, 并通过 SHAP 图(图 10)进行了直观的解释, 从而在一定程度上解决了传统机器学习模型的不可解释性问题。

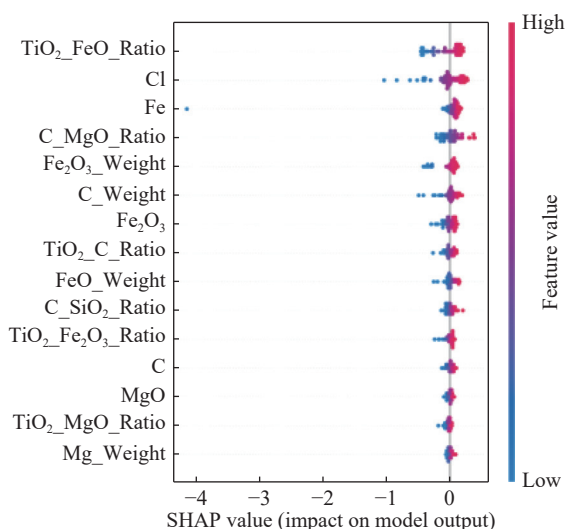


图 10 特征变量的平均 SHAP 贡献示意

Fig. 10 Average SHAP contribution plots for each feature

在确定的 15 个关键特征变量中, SHAP 分析表明, 钛铁比率(TiO_2 _FeO_Ratio)和碳输入量(CI)在所有特征中拥有最高的平均绝对 SHAP 值, 这表明它

们在预测 TiO_2 品味方面发挥着至关重要的作用。尽管钛铁质量比是后来加入的特征, 但其对最终预测结果的显著贡献证明了新特征在优化模型预测性能中的重要性。同时, TiO_2 _FeO_Ratio 和 TiO_2 _C_Ratio 等关键特征的贡献也非常明显, 为后续工艺优化提供了有力的指导。

尽管一些特征的 SHAP 值较低, 它们在模型中仍然发挥着作用。这不仅扩展了工艺参数的范围, 也为未来的工艺优化提供了额外的信息和潜在的改进空间。

2.5 特征范围的优化确定

模型通过 5 000 次蒙特卡洛模拟, 模拟了各种工业场景下不同输入组合的影响^[40-44]。利用堆叠模型对这些数据进行分析, 旨在筛选出能够使钛渣中 TiO_2 百分比超过 95% 的关键因素。通过动态调整输入特征的范围并观察模型的输出, 我们迭代优化了关键变量的范围。模拟结果揭示了输入变量间的复杂相互作用对 TiO_2 品位的影响, 本次模拟仅耗时 CPU 52 s, 与传统工业试验相比, 后者可能需要数月甚至数年的时间配料试验。图 11 展示了蒙特卡洛模拟预测的 TiO_2 百分比分布, 仿真结果(以 P90、P50 和 P10 分位数表示)显示, 预测值主要集中在 95% 的目标附近, 其中 P90 约为 95.4%, P50(中位数)为 95.1%, P10 为 94.8%, 显示出较小的偏差, 反映了模型预测的高精确度。大多数预测的 TiO_2 百分比值集中在所需范围内, 表明模型能够有效预测最佳目标值, 减少变异性, 增强预测准确性。

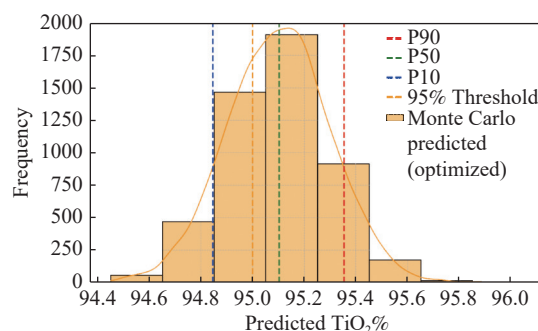


图 11 蒙特卡罗模拟的 TiO_2 % 分布

Fig. 11 Monte Carlo predicted TiO_2 % distribution with stacked model

相比之下, 图 12 显示实际生产数据的 TiO_2 百分比分布较为广泛且不一致, 尽管存在部分样本达到了高 TiO_2 百分比, 但整体变异性较大, P90 约为 97.5%, P50 为 94.8%, P10 约为 91%, 这反映了在实际生产中依靠经验配料维持目标范围内 TiO_2 百分

比仍存在优化空间。与传统冶金相比,蒙特卡洛模拟结合堆叠模型分析关键变量的范围,其目标值变异性更小。通过对比两图,优化特征范围后,该模型成功缩小了 TiO_2 百分比范围,使预测结果更集中,更接近目标 95%。这表明堆叠模型在生产过程控制中的潜力,能够提高预测的可靠性,减少生产过程中的波动,并有助于更高效地优化冶金过程。

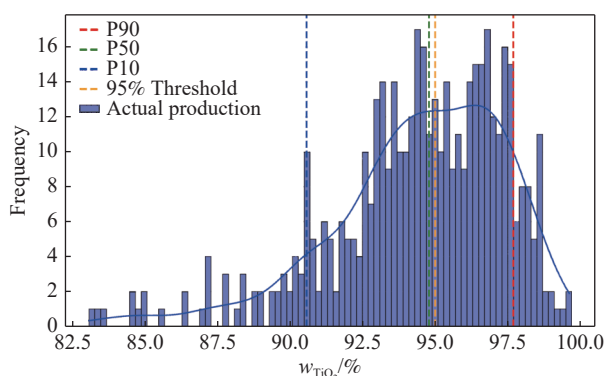


图 12 数据集中实际 w_{TiO_2} 的分布

Fig. 12 Actual $\text{TiO}_2\%$ distribution in dataset

综上所述,堆叠模型在预测 TiO_2 百分比方面的巨大潜力,不仅能够通过蒙特卡洛模拟快速生成可靠的预测结果,节省时间和资源,还能更好地实现冶金过程的智能化优化。

3 结论

采用数据驱动的建模技术,旨在提高涉及无烟煤和钛铁矿混合的工业过程中的高钛渣 TiO_2 品位。为实现更高的预测精度和优化模型整体性能,结合

了支持向量回归(SVR)、随机森林(Random Forest)和梯度提升机(Gradient Boosting)作为基础学习模型,并通过线性堆叠(Linear Stacking)方法作为元模型,构建了集成学习框架。该框架基于工厂的 100 个生产周期数据集进行广泛的统计分析和数据清洗,以提取更有深度的特征变量,并通过 Shapley 方法评估了输入特征对输出响应的影响。

集成模型的表现出色, R^2 达到了 0.924 9,且在验证集上实现了 0.29% 和 0.30% 的绝对均值百分比误差(MAPE),显示出相较于单一模型的明显优势。此外,还进行了 5 000 次蒙特卡洛试验,基于物理化学机理有效扩展了数据集,多模型结合的方法有效降低了方差和减轻了过拟合的风险。探索了关键特征,如 TiO_2/FeO 比率(1.70 ~ 2.12)、 $\text{TiO}_2/\text{Fe}_2\text{O}_3$ 比率(1.40 ~ 2.77)、 TiO_2/C 比率(0.50 ~ 0.58)、碳含量(84.41% ~ 92.91%)和 MgO 含量(0.41% ~ 1.30%)对 TiO_2 品位的影响,发现这些特征的组合有助于生成最佳的 TiO_2 品位。

通过结合堆叠模型与 SHAP 解释方法,不仅提高了 TiO_2 品位预测的精度,还为优化原材料比例提供了可操作的见解,进一步优化了冶金过程的原料配比。这种方法能有效提升最终产品的质量,并为相关工业决策提供有力支持。通过将这一方法集成到用户友好的软件包中,不仅为高钛渣中钛白粉品位的预测提供了实用工具,也能进行原材料比例的分析与优化。该工具消除了一定程度上因经验配比造成的能源浪费,并使整个生产过程更加智能化、高效、环保,进而有效实现节能目标。

参考文献

- [1] WANI T U, RATHER A H, KHAN R S, *et al.* Titanium dioxide functionalized multi-walled carbon nanotubes and silver nanoparticles reinforced polyurethane nanofibers as a novel scaffold for tissue engineering applications[J]. *Journal of Industrial and Engineering Chemistry*, 2023, 121 : 200-214.
- [2] AKINNAWO S. Synthesis, modification, applications and challenges of titanium dioxide nanoparticles[J]. *Research Journal of Nanoscience and Engineering*, 2019, 3(4).
- [3] LI Y, ZHONG M, HE X, *et al.* The combined effect of titanium dioxide nanoparticles and cypermethrin on male reproductive toxicity in rats[J]. *Environmental Science and Pollution Research*, 2023, 30(9) : 22176-22187.
- [4] HARUNA A, CHONG F K, HO Y C, *et al.* Preparation and modification methods of defective titanium dioxide-based nanoparticles for photocatalytic wastewater treatment—a comprehensive review[J]. *Environmental Science and Pollution Research*, 2022, 29: 70706-70745.
- [5] TIAN C, HUANG S, YANG Y. Influences of hydrolysis temperature and hydrolysis time on titanium white pigment via short sulfate process[J]. *Advanced Materials Research*, 2012, 602-604: 1255.
- [6] MIU B A, DINISCHOTU A. Green synthesized titanium dioxide nanoparticles and their future applications in biomedicine, agriculture and industry[J]. *Rev Biol Biomed Sci*, 2021, 4(1): 1-21.

- [7] TIWARI P, VERMA M, AMBIKA, *et al.* Titanium dioxide-based nanoparticles and their applications in water remediation[J]. *Journal of Environmental Engineering Science*, 2023: 281-304.
- [8] LÜ X W, LÜ W, ZHANG Y Y, *et al.* Method for preparing high titanium slag through low temperature reduction and separation of titanium-containing iron ore: Chinese Patent CN106521139A[P]. 2017.
(吕学伟, 吕伟, 张颖异, 等. 一种低温还原分离含钛铁矿物制备高钛渣的方法 [P]. 中国专利: CN106521139A, 2017.)
- [9] DONG H G, GUO Y F, JIANG T, *et al.* Production of synthetic rutile from high Ca and Mg type titanium slag by mineral phase reconstruction process[J]. *The Chinese Journal of Nonferrous Metals*, 2012, 22(9): 2642-2647.
(董海刚, 郭宇峰, 姜涛, 等. 高钙镁型钛渣物相重构法制取人造金红石[J]. 中国有色金属学报, 2012, 22 (9): 2642-2647.)
- [10] KONDORI J, MIAH M I, ZENDEHBOUDI S, *et al.* Hybrid connectionist models to assess recovery performance of low salinity water injection[J]. *Journal of Petroleum Science and Engineering*, 2021, 197: 107833.
- [11] QU Y H. Rapid analysis of TiO₂ in blast furnace slag by X-ray fluorescence spectrometry[J]. *Metallurgical Analysis*, 2003(5): 40-42.
(曲月华. X 射线荧光光谱法快速分析高炉渣中 TiO₂[J]. 冶金分析, 2003(5): 40-42.)
- [12] WU L P, ZENG Y, LIU W, *et al.* Determination of major and minor elements in high-titanium slag by fusion sample preparation-X-ray fluorescence spectrometry[J]. *Multipurpose Utilization of Mineral Resources*, 2017, (1): 81-84.
(仵利萍, 曾英, 刘卫, 等. 熔融制样 - X 射线荧光光谱法测定高钛渣主次量元素 [J]. 矿产综合利用, 2017, (1): 81-84.)
- [13] ZHENG M Q, LI B M, CHENG X H. Rapid analysis of mineral phases in titanium ore[J]. *Rock and Mineral Analysis*, 2010, 29(1): 61-63.
(郑民奇, 李邦民, 程秀花. 钛矿石物相的快速分析[J]. 岩矿测试, 2010, 29 (1): 61-63.)
- [14] CAO J C, LU M. Industrial water consumption prediction based on improved SSA-SVM with hybrid strategy[J]. *Water Resources and Power*, 2023, 41(9): 28-31.
(曹敬椿, 卢敏. 基于混合策略改进 SSA-SVM 的工业用水量预测[J]. 水电能源科学, 2023, 41 (9): 28-31.)
- [15] LI H, BU X, LIU X, *et al.* Evaluation and prediction of blast furnace status based on big data platform of ironmaking and data mining[J]. *ISIJ International*, 2021, 61(1): 108-118.
- [16] WANG Y, XU S, BWAR K, *et al.* Application of machine learning for composite moulding process modelling[J]. *Composites Communications*, 2024, 48: 101960.
- [17] CAO Y H, CHEN C Y, XU S Z, *et al.* Machine learning assisted prediction and optimization of mechanical properties for laser powder bed fusion of Ti6Al4V Alloy[J]. *Additive Manufacturing*, 2024, 91: 104341.
- [18] GUNASEGARAM D R, BARNARD A S, MATTHEWS M J, *et al.* Machine learning-assisted in-situ adaptive strategies for the control of defects and anomalies in metal additive manufacturing[J]. *Additive Manufacturing*, 2024, 81: 104013.
- [19] HE Z, ZHANG H, CHENG H, *et al.* Machine learning guided BCC or FCC phase prediction in high entropy alloys[J]. *Journal of Materials Research and Technology*, 2024, 29: 3477-3486.
- [20] BAI M, WU X, TANG S, *et al.* Study on hot deformation behavior and recrystallization mechanism of an Al-6.3Zn-2.5Mg-2.6Cu-0.11Zr alloy based on machine learning[J]. *Journal of Alloys and Compounds*, 2024, 1000: 175086.
- [21] HOSSAIN M A, MIRELES A J, STEWART C M. A machine learning approach for stress-rupture prediction of high temperature austenitic stainless steels[C]. *Proceedings of the ASME Turbo Expo 2022: Turbomachinery Technical Conference and Exposition*, Rotterdam, Netherlands, June 13-17, 2022: V007T17A029. ASME.
- [22] NAUTIYAL A, MISHRA A K. Machine learning approach for intelligent prediction of petroleum upstream stuck pipe challenge in oil and gas industry[J]. *Environment Development and Sustainability*, 2022.
- [23] SHI Z, DU L, HE X, *et al.* Prediction model of yield strength of V-N steel hot-rolled plate based on machine learning algorithm[J]. *JOM*, 2023, 75(5): 1750-1762.
- [24] ÁLVARO R O, GERT G, MARTIN S, *et al.* Optimisation of operator support systems through artificial intelligence for the cast steel industry: a case for optimisation of the oxygen blowing process based on machine learning algorithms[J]. *Journal of Manufacturing and Materials Processing*, 2022, 6(2): 34-34.
- [25] MIENYE I D, SUN Y. A survey of ensemble learning: concepts, algorithms, applications, and prospects[J]. *IEEE Access*, 2022, 10: 99129-99149.
- [26] LEE C Y, LIU Y Y, TAN B S, *et al.* Design of hybrid machine learning system for efficient prediction of parameters in metal forming processes[J]. *International Journal of Advanced Manufacturing Technology*, 2018, 97: 4389-4405.
- [27] RAMESH G R, RAJU S N C B, NAIDU M S S R. Mathematical modeling and prediction of quality characteristics in CNC

- turning using machine learning algorithms[J]. Journal of Manufacturing Science and Engineering, 2023, 145(2): 029902.
- [28] YU M, ZHANG X, HAN Q, *et al.* Data-Driven method for prediction of residual stress of welding process[J]. Journal of Materials Processing Technology, 2023, 313: 117618.
- [29] HANY A B, HE W C, ZHANG T J, *et al.* Process control and optimization for high-quality stainless steel strip production[J]. Computers & Industrial Engineering, 2022, 165: 107896.
- [30] HE A S, LIU Z L, WANG F Q, *et al.* Application of artificial intelligence algorithms in predictive maintenance of manufacturing equipment[J]. Journal of Advanced Manufacturing Technology, 2022, 118: 1177-1189.
- [31] ZHU S M, YOU M L. Discussion on several problems in improving TiO₂ grade in high titanium slag production by electric heating method[J]. Iron Steel Vanadium Titanium, 1984, 1: 41-45.
- [32] RASHIDA S, PEYMAN P, LEI W. Data-Driven analyses of low salinity waterflooding in carbonates[J]. Applied Sciences, 2021, 11(14): 6651-6651.
- [33] ISKANDAROV J, FANOURGAKIS G S, AHMED S, *et al.* Data-Driven prediction of in situ CO₂ foam strength for enhanced oil recovery and carbon sequestration[J]. RSC Advances, 2022, 12: 35703-35711.
- [34] WANG Z, HUANG Y, LIU Y, *et al.* Prediction model of strip crown in hot rolling process based on machine learning and industrial data[J]. Metals, 2023, 13(5): 900.
- [35] KARIMINEJAD M, TORMEY D, O'HARA C, *et al.* Prediction of hotspots in injection moulding by using simulation, in-mould sensors, and machine learning[C]. 2023 9th International Conference on Control, Decision and Information Technologies (CoDIT), Rome, Italy, 2023: 309-314.
- [36] CHENG H, HE Y, ZENG P, *et al.* Deep Learning-based prediction of subsurface oil reservoir pressure using spatio-temporal well-logging data[J]. Advances in Water Resources, 2023: 1265-1275.
- [37] FIGUEIREDO S G, COSTA A M, GONÇALVES M J M, *et al.* Artificial intelligence algorithms in process systems engineering[M]. 1st ed. Cham: Springer, 2022.
- [38] LIU Y, ZHAN L, FANG Z, *et al.* Study on intelligent diagnosis and prediction of wear status in deep mining equipment under extreme environment[J]. Science Progress, 2023, 106: 009317132312036.
- [39] SINGH M, KUMAR S, NAIR P S K. Artificial intelligence-based approaches for predicting industrial asset reliability[J]. Materials, 2021, 14(23).
- [40] CHENG S F, CHENG W C. Application of machine learning algorithms in the predictive maintenance of industrial robots[J]. Procedia CIRP, 2022, 113: 595-600.
- [41] XIE Y Z, LIU B, ZHANG X L, *et al.* Predicting the wear life of railway track using machine learning: a case study[J]. Journal of Transportation Engineering, 2023, 149(5): 04023022.
- [42] GOHAR M A K, JAMIL A S, RAZAK A H. Machine learning approach for predicting energy consumption in mManufacturing industry[J]. Advanced Science Letters, 2023, 29(12): 10760-10766.
- [43] PRADHAN M M, DUTTA B S, NARAYAN H S. Real-time prediction of industrial equipment health using artificial intelligence and internet of things[J]. Computers & Industrial Engineering, 2022, 164: 107976.
- [44] ZHUANG S S R, GUO L Y, ZHUANG W D, *et al.* Application of machine learning for predicting optimal parameters in manufacturing[J]. International Journal of Advanced Manufacturing Technology, 2022, 118(7): 1411-1424.