

基于机器学习和优化 AI 算法 预测壳芯强度

于 成¹, 刘建平², 陆仕平¹

(1. 浙江万丰轻合金研究院有限公司, 浙江新昌 312500; 2. 浙江万丰摩轮有限公司, 浙江新昌 312500)

摘要: 准确可靠的壳芯强度预测, 对提升铸造工艺性能、实现制造过程中材料的高效利用具有重要意义。本研究基于所构建的数据库, 结合皮尔逊相关系数及SHAP分析, 探讨了8种机器学习算法在预测壳芯抗拉强度方面的适用性。结果表明: XGBoost与GBR模型拟合效果优异, 其决定系数(R^2)分别达到93.5%和92.1%, 同时均方误差、平均绝对误差和均方根误差等指标也维持在较低水平; 对输入特征的重要性进行SHAP分析, 发现树脂含量与乌洛托品含量是影响抗拉强度的关键变量; XGBoost与GBR模型预测稳定性较优, 其预测值与实测值高度吻合, 变化趋势一致。综上所述, 本研究证实基于机器学习的建模方法可有效实现壳芯抗拉强度的预测, 相较于传统分析手段, 其在预测能力方面具有明显优势。

关键词: 铸造; 壳型成形; 拉伸性能; 机器学习; 强度预测; SHAP分析; AI算法

中图分类号: TG142.1; TP181 **文献标识码:** A **文章编号:** 1001-4977 (2026) 08-0885-08

DOI: 10.27014/j.cnki.zhuzao.2026.0114

Strength Prediction of Shell Core Using Machine Learning and Optimized AI Algorithms

YU Cheng¹, LIU Jian-ping², LU Shi-ping¹

(1. Zhejiang Wanfeng Light Alloys Research Institute Co., Ltd., Xinchang 312500, Zhejiang, China; 2. Zhejiang Wanfeng Motorcycle Wheel Co., Ltd., Xinchang 312500, Zhejiang, China)

Abstract: Reliable shell core strength prediction is essential for improving the casting process performance and ensuring efficient material utilization in manufacturing processes. Based on the constructed database, and combined with Pearson correlation coefficient and SHAP analysis, this study explores the applicability of eight machine learning techniques to predict the tensile strength of shell cores. The results show that the XGBoost and GBR models achieve excellent fitting performance, with determination coefficient (R^2) of 93.5% and 92.1%, respectively, along with comparatively lower mean squared error (MSE), mean absolute error (MAE), and root mean squared error (RMSE). Feature importance analysis using SHAP identifies that the resin and urotropine contents are dominant variables influencing the tensile strength. The XGBoost and GBR models show better predictive stability, as their predictions closely match observed values and exhibit consistent trends. Overall, the findings highlight the effectiveness of machine learning-based modelling for tensile strength estimation of shell cores, offering improved predictive capability relative to traditional analytical approaches.

Key words: foundry; shell mold forming; tensile property; machine learning; strength prediction; SHAP analysis; AI algorithm

壳芯是一种基于树脂覆膜砂精密铸造技术, 利用热固性树脂覆膜砂的热反应特性, 使砂粒在热金属型腔中固化成具有高几何稳定性与表面质量的壳体结构。由于成形精度优异、重复性良好且工艺过程高度可控, 壳型工艺在现代铸造工业中被视为制造高性能砂芯的重要方法, 并在高精度复杂铸件的生产中发挥

着关键作用^[1-2]。壳型工艺通常包括树脂覆膜砂制备与覆膜砂在金属模具内的成形两个阶段。第一阶段, 通过热混法制得树脂覆膜砂, 即在受控温度条件下, 将热固性树脂预先均匀包覆于每个砂粒表面, 这一过程不仅影响覆膜砂的流动性和热硬化行为, 还决定其在模具充填及固化阶段的结构完整性, 是保证成形质量

基金项目: 浙江省工程研究中心“汽车轻合金铸件数字孪生关键技术与应用”项目(ZJTX2305)。

收稿日期: 2026-03-31 收到初稿, 2026-05-07 收到修订稿。

作者简介: 于成(1989-), 男, 硕士, 工程师, 主要从事金属新材料、新技术、新工艺研究工作。E-mail: cheng.yu@wft.com

引用格式: 于成, 刘建平, 陆仕平. 基于机器学习和优化 AI 算法预测壳芯强度 [J]. 铸造, 2026, 75(8): 885-892.

Yu Cheng, Liu Jianping, Lu Shiping. Strength prediction of shell core using machine learning and optimized AI algorithms [J]. Foundry, 2026, 75(8): 885-892.

的基础^[3]。第二阶段,树脂覆膜砂在专用壳型成形设备中被充填至预热的金属模具内,覆膜砂迅速升温,树脂发生固化反应,促使砂粒之间形成强度较高的三维网状结构。随着固化反应进行,砂粒逐渐结合形成壳体,其几何形状严格复制模具内腔。固化过程的热动力学特性可使壳型在短时间内获得高强度和优异尺寸稳定性,从而显著提升壳芯的成形精度^[3-4]。壳型工艺的优势主要体现在以下几个方面:①尺寸精度高,典型公差可保持在约0.005 mm/mm量级,为高复杂度部件提供可靠的几何控制能力;②壳型表面粗糙度通常在0.3~4.0 μm 范围内,显著优于传统砂型铸造工艺;③可制造厚度5 mm以下的薄壁复杂曲面结构,具有卓越的几何形状再现能力;④壳型形成过程易于机械化与自动化,适用于中小尺寸铸件的大批量生产;⑤覆膜砂与多种有色合金体系具有良好适配性,使壳型工艺在不同材料体系中具有较高的工艺通用性;⑥工艺重复性强,壳芯性能在批次间具有良好一致性;⑦在实现高质量成形的同时可减少加工余量,显著降低整体制造成本。总体而言,壳型工艺凭借高精度、高可靠性与高效率,以及在表面质量控制、复杂结构复制以及成本效益方面的综合优势,成为高端精密铸件制造中的关键工艺技术。

准确评估壳芯强度对确保铸件的结构可靠性与运行安全至关重要。传统强度评估方法多依赖经验模型与试验测定,但此类方法通常工作量大、成本高,且结果易受操作者及试验条件影响而产生波动^[5-6]。此外,实验室环境下难以真实再现工业生产中的复杂工况,导致标准化试样测量结果与实际性能之间出现偏差。仅依赖传统手段难以满足现代铸造工业对高精度、高一致性的质量控制需求。因此,开发更为先进的数据驱动预测方法显得尤为重要。基于数据挖掘与机器学习的智能评估技术能够综合考虑多变量因素,克服传统模型的简化假设,并在无需大量破坏性试验的情况下实现对壳芯强度的准确预测,有助于提高生产过程的可控性与稳定性,保障整体制造效率与产品一致性。

近年来,机器学习(machine learning, ML)技术在材料科学领域的应用迅速增长,尤其在材料力学性能预测方面展现出巨大潜力^[7-9]。与传统的物理模型或经验模型相比,机器学习能够更有效地捕捉材料属性、工艺参数与性能输出之间的非线性关联,因此被视为促进材料设计与制造智能化的重要工具。已有诸多研究尝试评估了不同机器学习算法在预测材料性能方面的适用性,并通过参数调优、特征工程与模型集成等技术提升其预测精度^[10-11]。尽管这些研究取得了一

些成果,但关于不同算法在同一数据库和相同试验条件下的系统性比较仍相对有限。

对于壳芯材料而言,其力学性能受原材料性质(如原砂粒度和树脂含量等)与加工工艺参数(如温度、压力和固化时间等)共同影响,这些变量之间存在高度耦合和复杂的非线性关系。因此,构建能够反映这些输入变量的相互作用,并准确预测壳芯强度演变规律的机器学习模型,对提升铸造工艺质量与实现生产过程的智能优化具有重要意义。然而,目前利用机器学习对壳芯或壳型材料强度进行预测的研究仍处于起步阶段,且研究方法存在以下不足:其一,多数研究未对广泛的机器学习算法进行全面比较,难以得出具有普适性的结论;其二,关于模型解释性与预测偏差来源的分析较为薄弱,难以满足工程实际对模型透明度和可解释性的需求;其三,输入特征贡献的量化评估尚不充分,这限制了工程人员依据模型结果指导材料配比或工艺优化的能力。

针对以上研究空白,本课题系统性地比较了8类常见且代表性强的机器学习算法,并通过严格的验证程序评估其对壳芯强度预测的表现。所用模型包括基于树结构的梯度增强回归模型(GBR)、极端梯度增强模型(XGBoost)、随机森林回归模型(RFR)、支持向量回归模型(SVR)、人工神经网络模型(ANN)、多层感知器模型(MLP)、套索回归模型(Lasso)以及 k -近邻算法(KNN)。这些算法是当前材料信息学领域常用的主要建模策略,涵盖从线性到非线性、从浅层到深层、从参数化到非参数化的多种模型结构,能够实现对不同模型的全面对比分析。

在模型构建过程中,对每种算法均实施系统化的超参数优化,以确保其性能处于最优或次优状态。同时,采用交叉验证作为模型评估框架,旨在提升预测性能的稳健性,减少过拟合对结果分析的影响。为解决机器学习模型在工程应用中可解释性不足的问题,采用SHAP方法,通过计算各特征在不同预测任务中的贡献度,使输入变量对预测结果的影响机制能够得到定量揭示。SHAP方法不仅能够提供全局特征重要性排序,还能通过局部解释揭示模型在个别样本中的决策依据,从而提升模型的透明性和工程可用性。

通过上述系统性比较,本课题旨在揭示不同算法在预测壳芯强度时的性能差异,探讨它们在捕捉复杂非线性关系方面的适用性和局限性。研究结果可为选择合适的机器学习方法以满足不同工程需求提供依据,同时对输入变量的敏感性分析提供参考。

1 研究方法

1.1 试样制备

试验数据来自不同工艺条件下制备的标准化壳芯试样。首先制备树脂覆膜砂，过程如下：将15 kg硅砂预热至150 ℃，然后倒入机械搅拌机中充分加热；待砂料受热均匀后，加入硅砂质量2%~8%的商用酚醛树脂，持续搅拌40~80 s，使树脂在砂粒表面形成均匀的包覆层；依次加入占树脂质量5%~15%的乌洛托品（六甲基四胺）作为固化剂，以及占树脂质量2%~6%的硬脂酸钙作为润滑与脱模助剂；继续搅拌混合，直至覆膜砂冷却并呈现自由流动性，此时砂粒表面的树脂膜已成功固化成形。

采用标准拉伸试样进行力学性能测试，试样截面积为135 mm²，见图1。具体制备过程如下：首先将金属模具预热至200~350 ℃，随后通过喷吹装置将预制的树脂覆膜砂填充至金属模具中，加热保压时间控制在5~10 min，以确保壳体结构均匀完整并具备足够强度；固化结束后，将试样从模具中脱出并冷却至室温，即可用于后续测试。

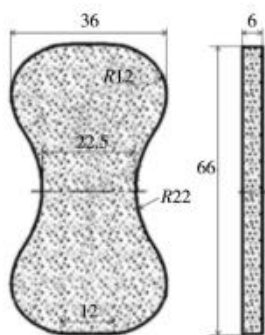


图1 覆膜砂标准拉伸试样示意图

Fig. 1 Schematic diagram of the standard tensile test specimen for precoated sand

1.2 数据库的构建

定量研究壳芯试样的抗拉强度，需要综合考虑原砂粒度、酚醛树脂含量、乌洛托品含量以及硬脂酸钙含量等多个参数。由于数据库的质量直接关系到预测模型的准确性与可靠性，因此在构建数据库时应格外注重数据来源与变量控制。数据库基于试验数据建立，其中砂粒骨料采用两种规格：AFS细度50/100和100/200；树脂含量设定为砂质量的2%、3%、4%、5%、6%、7%和8%；乌洛托品含量为树脂质量的5%、7%、9%、11%、13%和15%；硬脂酸钙含量为树脂质量的2%、3%、4%、5%和6%。每种工艺条件下均制备

5~8个平行拉伸试样进行测定，取其平均值作为该工艺条件下的最终结果。测试完成后，将数据库按80%与20%划分为训练集与测试集，用于模型建立与验证。除本研究获得的数据外，数据库还采用了已发表文献中的相关数据。为确保数据质量与可比性，仅纳入对材料成分、加工工艺及力学性能有完整、明确描述的文献数据，以提升模型训练的可靠性与严谨性。

1.3 人工智能算法

一般认为，在人工智能（AI）分析中，大规模数据能够显著提升模型的训练效率与预测性能，尤其在处理复杂系统或探索因果关系时。然而，在覆膜砂和壳型制造领域，受实际生产条件以及研究者主观选择等因素的限制，无论是通过受控试验构建大规模数据库，还是从已发表文献中提取高质量数据，都面临较大挑战，样本量难以充分扩展，因此选择适合的小样本建模方法尤为关键。

总体而言，极端梯度增强模型^[12-13]在处理结构化数据方面性能卓越，能够自动处理特征间的相互作用，并对缺失数据具有良好适应性，因此特别适用于样本量有限、变量关系复杂的场景。梯度增强回归模型^[14]作为一种稳健的集成学习方法，通过构建一系列逐步优化的决策树，有效最小化损失函数，能够处理连续与不连续数据。随机森林回归^[15]和支持向量回归^[16-18]则是材料研究中被广泛验证的成熟技术，它们在小样本条件下也能很好地模拟复杂的非线性关系，并保持稳定可靠的预测性能。人工神经网络^[19-20]是一类经典模型，常用于多响应指标优化，以及依据给定工艺参数预测材料的力学与物理性能。其中，多层感知器^[21-22]作为基础性的前馈神经网络，由输入层、隐藏层和输出层构成，神经元之间采用非线性激活函数并通过反向传播进行训练，可用于回归与分类任务，适合处理非线性数据，本质上是解决线性不可分问题。此外，基于最小绝对偏差收缩与变量选择方法的套索回归^[23-25]是一种具备正则化与自动特征选择能力的回归方法，通过在线性回归损失函数中引入L1惩罚项，将不重要的特征系数压缩甚至归零，从而提升模型的简洁性与可解释性，尤其适用于高维数据场景。 k -近邻算法^[26-27]作为一种非参数监督学习方法，通过计算样本间的距离并根据最近的 k 个相邻样本进行预测，操作简单、适应性强，常用于分类与回归问题。尽管上述模型大多面向较大规模数据库设计，但结合其特征提取能力与表示能力，本研究进一步探索它们在样本量受限条件下的潜在适用性，为覆膜砂壳芯强度预测提供可能的建模途径。

1.4 壳芯力学性能评估准确性和效率指标

利用测试数据库进行模型评估时, 需要对各模型的预测准确性和误差水平进行系统检验。为提高模型评估的可靠性, 降低过拟合风险, 采用交叉验证技术, 并对所有模型进行参数调优。AI模型的预测效果通过标准的回归误差指标进行评价, 包括决定系数 (R^2)、均方根误差 (RMSE) 以及平均绝对误差 (MAE)^[28]。其中, 决定系数的计算方式如下:

$$R^2 = 1 - \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (1)$$

式中: n 为观测值数; y_i 为测量值; $\hat{y}_i$ 为对应的模型预测值; $\bar{y}$ 为平均观测值。

R^2 用于量化预测模型的因变量方差比, 其数值范围为0到1 (或0到100%)。数值越高表示拟合度越好。

均方根误差用于评估预测误差的离散度, 其计算方式如下:

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

它反映了预测值与观测值的平均偏差程度。RMSE值越小, 预测准确率越高, 数值为0表示完全拟合。MSE (均方误差) 与RMSE相关, 为RMSE的平方。

平均绝对误差用于衡量预测值与实际值之间的平均绝对差, 表达式如下:

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

与RMSE不同, MAE对所有误差均给予相同权重, 提供了直观且易于理解的预测准确度衡量标准。MAE值越低, 表示模型预测性能越好。

通过皮尔逊相关系数 (r)^[29-30]对所考察变量进行系统的相关性分析, 表达式如下:

$$r = \frac{n \sum xy - \sum x \sum y}{\sqrt{[n \sum x^2 - (\sum x)^2] [n \sum y^2 - (\sum y)^2]}} \quad (4)$$

式中: x 和 y 为被比较的变量; $\sum xy$ 为成对观测的变量乘积之和; $\sum x$ 和 $\sum y$ 为各个观测值的和; $\sum x^2$ 和 $\sum y^2$ 为各个观测值的平方和。

SHAP是机器学习中用于解释所有模型输出的综合框架, 通过赋予每个输入特征重要性 (SHAP值) 量化该特征对特定预测的贡献。它将模型的预测分解为每个输入特征贡献的总和, 确保总和等于实际预测与平均预测的差值。Lundberg和Lee的工作解释了该方法的基本原理^[31]。SHAP框架已被用于评估合金元素和工艺参数对复杂相互作用下微观结构特征的影响^[32]。每个样本按照式 (5) 进行计算。

$$f(x) = \Phi_0 + \sum_{i=1}^M \Phi_i(x) \quad (5)$$

式中: $f(x)$ 为预测的微观结构特征值; Φ_0 为基线值的常数项; M 为输入变量的总数; $\Phi_i(x)$ 为SHAP值。

输入变量对目标特征的贡献会被加权并以不同输入值组合相加, 其定义如式 (6) 所示:

$$\Phi_i(x) = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|! (|F| - |S| - 1)!}{|F|!} [f_{S \cup \{i\}}(x) - f_S(x)] \quad (6)$$

式中: F 为所有输入特征的集合; $|F|$ 为输入特征的总数; S 为输入特征的一个子集; $|S|$ 为输入特征子集的数量; $f_{S \cup \{i\}}(x)$ 为用该特征训练模型的预测值; $f_S(x)$ 为模型的预测值。

2 结果与讨论

2.1 机器学习模型性能比较

8种AI算法的性能指标如表1所示, 包括关键值预测准确性指标及对应误差值。这些指标均来源于测试数据库的最终评估, 可反映模型在处理未见数据时的泛化能力。

表1 测试数据库下机器学习模型的准确度指标和误差值
Tab. 1 Accuracy metrics and error values for the machine learning models of test datasets

算法	数据个数	R^2	MSE	MAE	RMSE
XGBoost	136	0.935	11.303	2.938	3.362
GBR	136	0.921	15.461	3.722	3.932
RF	136	0.896	27.857	3.781	5.278
ANN	136	0.822	31.855	3.735	5.644
MLP	136	0.813	34.727	4.836	5.893
SVR	136	0.771	45.914	4.908	6.776
KNN	136	0.726	79.762	7.181	8.931
Lasso	136	0.561	113.145	8.748	10.637

图2展示了抗拉强度预测值与实测值之间的相关性, 以及AI辅助评估指标的数据分布。可以看出, 集合类学习算法在抗拉强度预测中表现优异。其中, XGBoost表现最佳, 决定系数 R^2 达到0.935, 说明其能解释93.5%的数据方差, 同时误差指标最低, 显示出高度可靠的预测能力。GBR预测精度同样较高 ($R^2=0.921$), 能够有效模拟材料的非线性关系。RF模型也保持较强预测能力 ($R^2=0.896$), 其性能优于所有非集合类算法。综上, 集合类算法在复杂回归任务中能够在保持高预测精度的同时有效控制误差。

在人工神经网络方面, ANN和MLP模型表现中等, 决定系数分别为0.822、0.813。如图2所示, 这两类模型虽然具备一定基线预测能力, 但较高的误差范围表明, 在缺乏充分超参数优化的情况下, 它们难以充分捕捉复杂参数之间的交互关系, 需要通过改进网

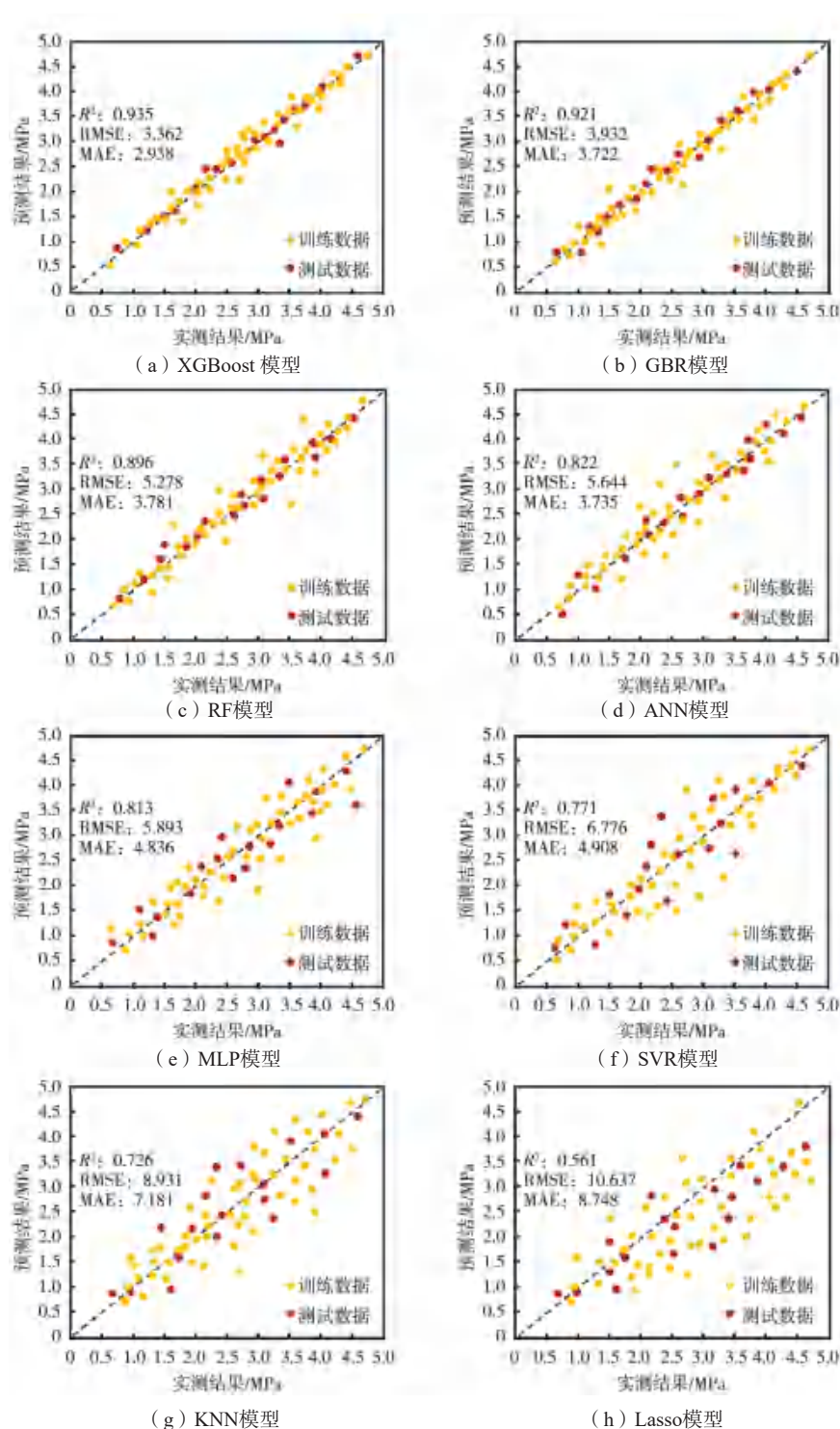


图2 壳芯试样抗拉强度预测值和实测值与评估指标数据之间的相关性

Fig. 2 Correlation between predicted and measured values of the tensile strengths of the shell core samples and the evaluation metrics data

络结构或优化特征选择来提升性能。SVR、KNN和Lasso的预测精度相对较低, 决定系数 R^2 分别为0.771、0.726和 0.561。这些模型在预测中的准确度差异明显, 表明其泛化能力有限, 对于壳芯抗拉强度的预测效果不佳。

总体来看, 选择稳健机器学习模型有利于准确预

测壳芯抗拉强度。集合类学习方法, 尤其是XGBoost和GBR, 具有高预测精度和低误差, 适合处理复杂的回归问题。人工神经网络如ANN和MLP表现中等, 通过进一步调优有提升预测性能的潜力。相比之下, SVR、KNN和Lasso等单模型在准确性和稳定性方面表现较弱, 在预测抗拉强度时可靠性较低。

2.2 皮尔逊变量的相关系数

成分变量与壳芯抗拉强度之间的皮尔逊相关性分析如图3所示。树脂含量与数据库中决定系数呈现强负相关($r=-0.65$),说明树脂比例与抗拉强度密切相关。在实际生产中,适量增加树脂含量通常能够提升强度,反映出树脂作为粘结剂在砂粒间形成有效连接从而增强结构完整性的核心作用。乌洛托品含量与抗拉强度之间存在中等程度的负相关($r=-0.26$)。这表明,虽然适量乌洛托品有助于促进树脂固化进而提升强度,但其过量添加可能会削弱混合体系的整体力学性能。硬脂酸钙含量与抗拉强度的相关性更弱($r=-0.18$),轻微的负向趋势可能源于其主要作用是改善流动性,而非直接增强承载能力。

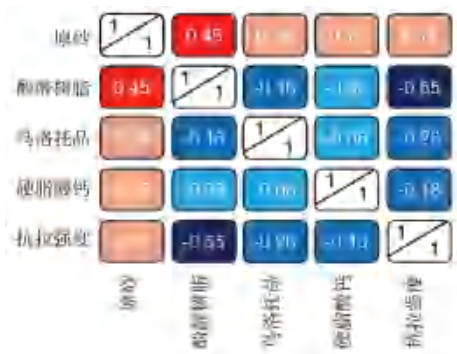


图3 壳芯的抗拉强度皮尔逊相关性分析

Fig. 3 Pearson's correlation analysis for the tensile strengths of shell cores

2.3 参数 / 变量的数值特征

图4展示了各输入变量的统计特征,有利于理解数据的分散程度和分布趋势。相对频率分布图直观体现了不同数值出现的频率,可反映数据的稳定性与可重复性。树脂含量作为壳芯与模具材料中的核心参数,其取值范围为原砂质量的2%~8% (平均5.2%),通过调节树脂比例可以满足不同强度需求,为工艺优化提供较大的灵活空间。乌洛托品占树脂质量的5%~15% (平均9.8%),生产过程中其用量调整十分明确,可用于优化树脂固化效率和壳芯力学性能。硬脂酸钙含量变化相对较小,为树脂质量的2%~6% (平均3.7%),其添加量需更精确地控制,以在覆膜砂流动性与壳芯结构完整性之间取得平衡。分析并未将原砂粒度作为连续变量纳入,而是仅考虑了两种工业常用砂型。主要原因在于实际生产中原砂粒度调整幅度有限,适用于壳型工艺的砂粒分布变化较小,难以作为具有连续性影响的变量使用。

如图5所示,作为输出变量的抗拉强度范围为1.3~5.3 MPa,平均值为3.4 MPa,充分反映了不同成分组合下材料力学性能的差异性。图4和图5中拟合的正态分布曲线强调了输入与输出变量的中心趋势及其相对均匀的分布特征。总体而言,数据库结构平衡、覆盖范围适度,为后续机器学习建模与性能分析提供了可靠且稳定的数据基础。

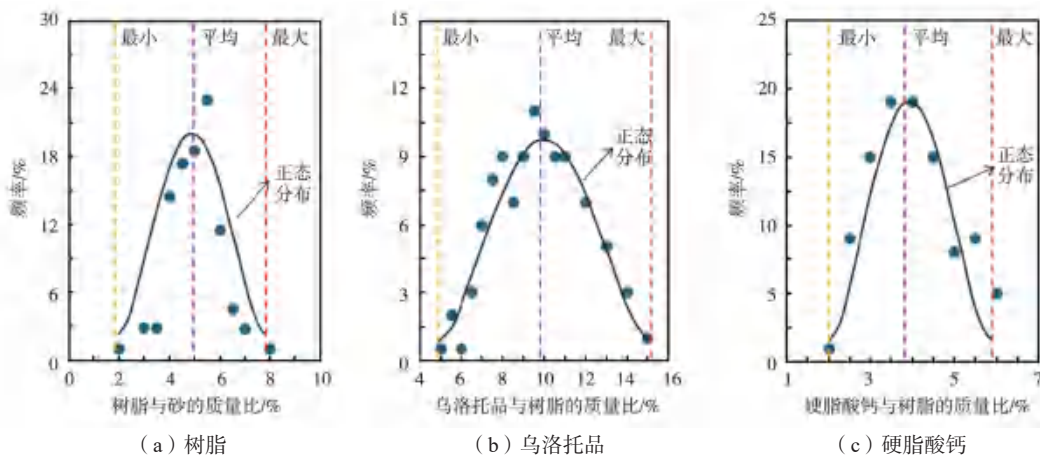


图4 壳芯中各成分的输入参数统计分析

Fig. 4 Statistical analysis of the input parameters in the shell cores

2.4 SHAP 分析: 输入参数对抗拉强度的影响

如图6所示,SHAP分析结果清晰揭示了各输入变量对抗拉强度预测的贡献程度。通过平均SHAP值,可以直观比较不同变量对模型输出的重要性。可以看出,酚醛树脂含量是影响最大且最关键的变量,平均SHAP值为+6.72,其在决定壳芯抗拉强度方面起核心

作用。正向SHAP值说明树脂含量越高,预测强度越大,且其贡献超出基线水平。因此,在实际生产中合理控制树脂用量尤为重要,尤其是在制造结构复杂、对强度要求较高的铸件时。乌洛托品的影响排在第二位,平均SHAP值为+4.15,其在树脂固化过程中具有关键作用。正向SHAP值表明适当增加乌洛托品含量有

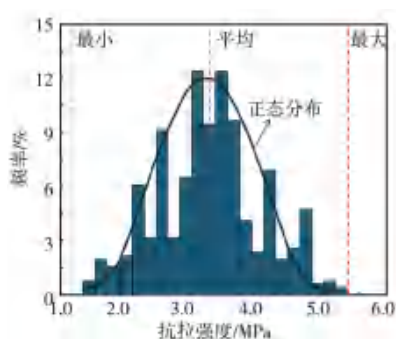


图5 壳芯抗拉强度输出参数统计分析

Fig. 5 Statistical analysis of tensile strength output parameters of the shell cores

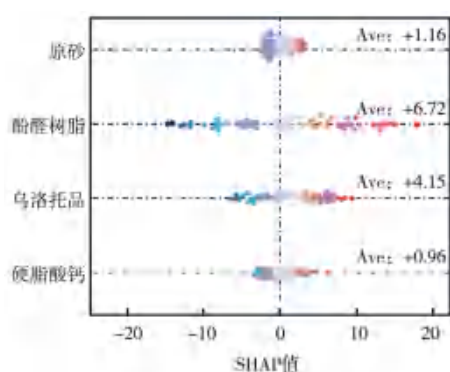


图6 特征输入变量和输出结果的SHAP分析及平均SHAP值

Fig. 6 SHAP analysis of feature input variables and output results and the average SHAP values

助于提高壳芯的抗拉强度，但过量添加可能削弱材料的整体性能，因此需保持合理比例，以兼顾固化效率与材料稳定性。原砂骨料作为壳芯的主体材料，其平均SHAP值为+1.16，贡献虽不及前两项显著，但仍具有正面影响。原砂颗粒在树脂粘接下形成结构框架，因此颗粒分布直接影响强度表现。细颗粒需要更高树脂用量才能形成充分覆盖的树脂膜，从而实现足够的粘接强度；粗颗粒紧实后孔隙率较大，接触点减少，强

度上限相对较低。由于工业生产中原砂粒度变化范围较小，其对抗拉强度的影响贡献适中。硬脂酸钙的影响相对有限，平均SHAP值为+0.96，其主要作用在于改善覆膜砂的流动性与加工性能，而非直接提升最终强度，因此对抗拉性能的贡献较小。总体而言，SHAP分析不仅明确了各变量的影响层级，还为材料配方优化与工艺控制提供了可靠依据，有助于提升壳芯制造过程的可控性与产品性能。

3 结论

本研究系统评估了8种机器学习算法在预测壳芯抗拉强度方面的效果，主要结论如下：

(1) XGBoost模型和GBR模型的预测性能最优，在测试集上，二者分别达到93.5%和92.1%的准确率，这主要归因于其在非线性关系和复杂特征交互建模方面的优势。RFR模型、ANN模型和MLP模型也展现出较为可靠的预测能力。相比之下，SVR、KNN和Lasso的表现较弱，其中Lasso最低，仅为56.1%。

(2) XGBoost和GBR表现出较优的预测稳定性，未出现严重过拟合。SVR、KNN和Lasso的误差较大，稳定性也相对不足。此现象表明基于集成类学习的方法在预测壳芯抗拉强度时具有更高的稳健性。

(3) 酚醛树脂含量是决定抗拉强度的核心变量，两者在皮尔逊相关分析中具有最显著的正相关性，且酚醛树脂的平均SHAP值高达+6.72，为所有特征中影响最大的因素，乌洛托品含量的相关性次之，硬脂酸钙的影响则相对有限，平均SHAP值为+0.96。

(4) 本研究为壳芯抗拉强度预测模型的选择提供了实用的参考框架，同时强调了超参数调优与高质量数据库在提升机器学习模型性能中的关键作用。该研究框架不仅能够提高预测准确性，还可为铸造及相关工程领域的工艺优化与材料设计提供有效支持。

参考文献：

- [1] ASM International Handbook Committee. ASM handbook, volume 15: casting [M]. OH, USA: ASM International, 2008.
- [2] Campbell J. Complete casting handbook: metal casting processes, metallurgy, techniques, and design [M]. 2nd ed. Oxford: Butterworth-Heinemann, 2015.
- [3] Groover M P. Fundamentals of modern manufacturing: materials, processes, and systems [M]. 6th ed. New Jersey: Wiley, 2016.
- [4] Brown J R. Foseco ferrous foundryman's handbook [M]. 11th ed. Oxford: Butterworth-Heinemann, 2018.
- [5] Stefanescu D M. Science and engineering of casting solidification [M]. 3rd ed. Berlin: Springer, 2015.
- [6] Sue T. Mold & core test handbook [M]. Illinois: American

Foundry Society, 2001.

- [7] Ehsan K, Mohamed A H, Inqiad W B, et al. Multi expression programming and interpretable machine learning for determining compressive strength of coral sand aggregate concrete [J]. Materials Today Communications, 2025, 45: 112370.
- [8] Tang H T, Zhang W, Li X X, et al. A discrete group teaching optimization algorithm for solving many-objective sand casting whole process production scheduling problem [J]. Computers & Operations Research, 2024, 164: 106563.
- [9] Jia H Q, Jiang J G, Li F, et al. Artificial neural network model based on improved VMD algorithm to monitor sand mass flow rate in natural gas pipeline [J]. Measurement, 2025, 254: 117887.

- [10] Tang H T, Fang B, Liu Y, et al. A hybrid teaching and learning-based optimization algorithm for distributed sand casting job-shop scheduling problem [J]. *Applied Soft Computing*, 2022, 120: 108694.
- [11] Chao Z M, Li Z K, Dong Y K, et al. Estimating compressive strength of coral sand aggregate concrete in marine environment by combining physical experiments and machine learning-based techniques [J]. *Ocean Engineering*, 2024, 308: 118320.
- [12] Chen T Q, Guestrin C. XGBoost: a scalable tree boosting system [C]// *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016: 785–794.
- [13] Otchere D A, Ganat T O A, Ojero J O, et al. Application of gradient boosting regression model for the evaluation of feature selection techniques in improving reservoir characterization predictions [J]. *Journal of Petroleum Science and Engineering*, 2021, 208: 109244.
- [14] Saraiva F M G, Guimaraes A C R, da Silva Júnior O S, et al. Performance analysis of gradient boosting regressor using K-fold and group K-fold for predicting permanent deformation in tropical soils [J]. *Journal of Materials Research and Technology*, 2026, 40: 3555–3567.
- [15] Breiman L. Random forests [J]. *Machine Learning*, 2001, 45: 5–32.
- [16] Cortes C, Vapnik V. Support-vector networks [J]. *Machine Learning*, 1995, 20: 273–297.
- [17] Polewczyk F, Leyssale J M, Aurel P, et al. Elasticity of dense anisotropic carbons: a machine learning model of the structure-property relationship informed by large scale molecular dynamics data [J]. *Acta Materialia*, 2024, 270: 119851.
- [18] Walczak S. Artificial neural networks [M]// *Advanced methodologies and technologies in artificial intelligence, computer simulation, and human-computer interaction*. Pennsylvania: IGI Global Scientific Publishing, 2019.
- [19] Dagli C H. Artificial neural networks for intelligent manufacturing [M]. Berlin: Springer, 2012.
- [20] Qamar R, Zardari B A. Artificial neural networks: an overview [J]. *Mesopotamian Journal of Computer Science*, 2023, 2023: 130–139.
- [21] Kotsiopoulos T, Sarigiannidis P, Ioannidis D, et al. Machine learning and deep learning in smart manufacturing: the smart grid paradigm [J]. *Computer Science Review*, 2021, 40: 100341.
- [22] Chan K T, Abu-Salih B, Qaddoura R, et al. Deep neural networks in the cloud: review, applications, challenges and research directions [J]. *Neurocomputing*, 2023, 545: 126327.
- [23] Cui C H, Wang D H. High dimensional data regression using Lasso model and neural networks with random weights [J]. *Information Sciences*, 2016, 372: 505–517.
- [24] Tibshirani R. Regression shrinkage and selection via the Lasso [J]. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 1996, 58 (1): 267–288.
- [25] Cui Z X, Gong G L. The effect of machine learning regression algorithms and sample size on individualized behavioral prediction with functional connectivity features [J]. *NeuroImage*, 2018, 178: 622–637.
- [26] Huang S, Huang M M, Lyu Y J. A novel approach for sand liquefaction prediction via local mean-based pseudo nearest neighbor algorithm and its engineering application [J]. *Advanced Engineering Informatics*, 2019, 41: 100918.
- [27] Richman J S. Chapter thirteen - multivariate neighborhood sample entropy: a method for data reduction and prediction of complex data [J]. *Methods in Enzymology*, 2011, 487: 397–408.
- [28] Wang W J, Lu Y M. Analysis of the mean absolute error (MAE) and the root mean square error (RMSE) in assessing rounding model [C]// *IOP Conference Series: Materials Science and Engineering*, 2018, 324: 012049.
- [29] Rodgers J L, Nicewander W A. Thirteen ways to look at the correlation coefficient [J]. *The American Statistician*, 1988, 42: 59–66.
- [30] Benesty J, Chen J D, Huang Y T, et al. Pearson correlation coefficient [M]// *Noise reduction in speech processing*. Berlin: Springer, 2009: 37–40.
- [31] Lundberg S M, Lee S I. A unified approach to interpreting model predictions [J]. *Neural Information Processing Systems*, 2017, 30: 4765–4774.
- [32] Mastelini S M, Cassar D R, Alcobaa E, et al. Machine learning unveils composition-property relationships in chalcogenide glasses [J]. *Acta Materialia*, 2022, 240: 118302.