小样本注浆材料强度预测模型对比研究*

翟东豪[1],王向鹏[1,2],姚磊钧[3],樊立博[3],吴旭[1],孔秋平[4],卢焱保[4]

(1.成都理工大学 计算机与网络安全学院, 四川 成都 610059;
2.成都理工大学 地球勘探与信息技术教育部重点实验室, 四川 成都 610059;
3.中国安能集团第三工程局有限公司, 四川 成都 611136;
4.福建永强岩土股份有限公司, 福建 龙岩市 364000)

摘要:注浆材料强度预测对提升注浆材料设计可靠性具有重要意义。针对传统经验模型难以准确捕捉材料成分与强度之间的关系,以及小样本导致的模型泛化能力不足的问题,提出了一种融合物理机理与数据增强的机器学习预测框架,并对随机森林回归(Random Forest Regression, RFR)、人工神经网络(Artificial Neural Network, ANN)、支持向量机回归(Support Vector Regression, SVR)和线性回归(Linear Regression, LR)4种机器学习方法的材料性能预测结果进行对比分析。结果表明:RFR模型的预测性能最优,其决定系数R^2提升至0.934,较LR、SVR、ANN模型分别提高了21.8%、10.4%、1.6%,均方误差*MAE*降至2.59,较LR、SVR、ANN模型分别降低了14.5%、13.4%、9.1%;特征重要性分析中,石墨比例对抗压强度贡献度达37.6%;物理约束的数据增强技术可显著提升小样本建模的鲁棒性,RFR和ANN模型的*MAE*较未纳入数据增强时分别提升39.8%和18.3%;RFR模型在各水固比条件下的预测误差均稳定低于10%。研究结果可为矿山注浆加固材料配比提供参考。

关键词:注浆材料;强度预测;机器学习预测框架;数据增强

中图分类号:TD265.4;TP183 **文献标识码**:A **文章编号**:1005-2763(2026)04-0170-08

Comparative Study on Strength Prediction Models of Small Sample Grouting Materials

ZHAI Donghao[1], *WANG Xiangpeng*[1,2], *YAO Leijun*[3], *FAN Libo*[3], *WU Xu*[1], *KONG Qiuping*[4], *LU Yanbao*[4]

(1.College of Computer and Network Security, Chengdu University of Technology, Chengdu, Sichuan 610059, China;
2.Key Laboratory of Earth Exploration and Information Technology of Ministry of Education, Chengdu University of Technology, Chengdu, Sichuan 610059, China;
3.China Anneng Group Third Engineering Bureau Co., Ltd., Chengdu, Sichuan 611136, China;
4.Fujian Yongqiang Rock and Soil Co., Ltd., Longyan, Fujian 364000, China)

Abstract: The strength prediction of grouting materials is of great significance to improve the reliability of material design for grouting. Aiming at the problems that the traditional empirical models are difficult to accurately capture the relationship between material composition and strength, and the insufficient generalization ability of models caused by small samples, a machine learning prediction framework combining physical mechanism and data enhancement was proposed. The material performance prediction results of four machine learning methods, Random Forest Regression (RFR), Artificial Neural Network (ANN), Support Vector Regression (SVR) and Linear Regression (LR), were compared and analyzed. The results show that the RFR model has the best predictive performance, and its determination coefficient (R^2) is increased to 0.934, which is 21.8%, 10.4% and 1.6% higher than that of LR, SVR and ANN models, respectively. The mean square error *MAE* is reduced to 2.59, which is 14.5%, 13.4% and 9.1% lower than that of LR, SVR and ANN models, respectively. In the analysis of feature importance, the contribution of graphite ratio to compressive strength is 37.6%. The data enhancement technique based on physical constraints can significantly improve the robustness of small sample modeling. The *MAE* of RFR

* **收稿日期**:2025-04-29
基金项目:龙岩市重点研发计划项目(2023LYF9003);四川自然科学基金项目(2025ZNSFSC0312)
作者简介:翟东豪(2002—),男,河南洛阳人,硕士研究生,主要研究方向为人工智能与地学大数据。E-mail:1102428556@qq.com
通信作者:王向鹏(1987—),男,陕西蓝田人,博士,副教授,主要研究方向为资源与灾害地球物理学。E-mail:xiangpengwang@cdut.edu.cn

and ANN models increases by 39.8% and 18.3% respectively compared to those without data enhancement. The prediction error of the RFR model remains stable below 10% under various water-solid ratio conditions. The research results can provide reference for the proportion of grouting reinforcement materials in mines.

Key words: Grouting material, Strength prediction, Machine learning prediction framework, Data enhancement

0 引言

注浆材料是由胶凝材料、导电相及添加剂等组成的多相复合材料。在矿业工程领域,注浆材料被广泛应用于采空区充填治理、巷道围岩加固和破碎带封堵等,其核心功能是通过浆液渗透-凝固过程填充岩土体裂隙,形成兼具力学承载与导电特性的硬化体,其力学性能(尤其是抗压强度)将直接影响矿山工程结构安全性与耐久性[1]。注浆材料可以作为一个电化学体系,其电学特性与内部孔隙结构、水化产物分布等关键参数存在显著关联性[2],在实际工程领域中,通常会在其中增添石墨等掺合料来增强其电学特征,使其具有更强的流变性能。然而,此类材料在服役过程中往往表现出显著的预测复杂性:一方面是由于注浆材料配比复杂、养护环境参数(温度、湿度、围压)及界面结合效应等多种非线性因素共同作用;另一方面是受限于试验的成本与周期,水泥的抗压强度测量一般将试件养护至 28 d,具有较大的滞后性[3]。

传统强度预测方法多依赖经验公式或简化物理模型,如回弹法[4]、钻芯法[5]、离散单元法[6],这些预测方法不仅存在滞后性,且难以表征材料组分之间的相互作用。近年来,机器学习(Machine Learning, ML)因能学习数据间的映射关系及具有事件预测能力,而被引入注浆材料性能预测领域[7-8]。随机森林回归(Random Forest Regression, RFR)基于集成学习策略,其特征子集随机抽样与多树投票机制可有效捕捉多组分材料的交互效应,同时提供特征重要性分析以辅助材料配比优化[9-11]。人工神经网络(Artificial Neural Network, ANN)通过仿生技术自动分析数据中的复杂关系,基于注浆材料抗压强度的影响因素构建多维特征,ANN 则通过这些参数建立材料组分、电学性能与力学性能的非线性映射关系[12]。支持向量机回归(Support Vector Regression, SVR)基于统计学习理论,通过核函数映射将低维非线性问题转化为高维空间线性回归问题,利用特征向量在超平面上的投影与目标值差距最小原则,在高维空间中选取最优超平面,为多因素交互作用提供定量解析框架[9]。线性回归(Linear Regression, LR)通过构建工艺参数与强度之间的显示数学关系,不仅能够快速量化单一变量的独立贡献,还可通过逐步回归筛选关键影响因素,适用于工程初期的多变量交互效应的敏感性分析[13]。然而在实际工程中,高成本与长周期性导致数据集规模有限,传统的机器学习方法易因数据稀疏陷入过拟合,对此,数据增强技术成为关键解决路径。在材料科学中,线性插值法因符合组分连续调控的物理规律,被应用于生成可信的虚拟数据[14]。

针对以上问题,本文提出了一种融合物理机理与数据增强的注浆材料强度预测框架。

(1) 构建多源特征空间。数据集涵盖材料类型(铝酸盐水泥、复合水泥等 6 类)、物质配比特征(水泥/石墨/水质量比)、电学特性(电阻率、极化率),通过水泥石墨比、总质量占比等衍生特征增加特征交互效应。

(2) 设计线性插值的特征空间数据增强方法:通过扰动控制将数据扩展至原始样本的 2 倍,改善数据分布的稀疏性。

(3) 建立多模型对比体系。采用随机森林回归(RFR)、人工神经网络(ANN)、支持向量机回归(SVR)和线性回归(LR)进行预测建模,RFR 的特征重要性分析为材料配比关系提供可解释的依据。

1 数据来源与预处理

1.1 数据来源

本次研究共浇筑 60 组试件,包括 6 种水泥,分别为铝酸盐水泥、复合水泥、粉煤灰水泥、火山灰水泥、矿渣水泥、普通硅酸盐水泥,水泥与石墨的质量比分别为 1∶0、20∶1、10∶1、5∶1、3∶1。电阻率、极化率采用 SCIP 样品岩芯 IP 测试仪进行测试。测量信号为电流 I 或初级电压 V_p,直接计算电阻率和极化。偏振测量分辨率为 1 μV/V,精度为 0.38%。采用微机控制电液伺服压力试验机(DYE-2000D)对阴干 28 d 的试样进行强度测试。

1.2 数据预处理

1.2.1 特征工程

在模型训练过程中,合理的数据预处理对提升

预测性能至关重要。原始数据集包括 6 类水泥(铝酸盐水泥、复合水泥、粉煤灰水泥、火山灰水泥、矿渣水泥和普通硅酸盐水泥)、配比特征(水泥、石墨、水的质量比)、电学特性(电阻率、极化率),通过计算水泥石墨比、总质量占比等衍生特征,增强变量间的交互作用。由于水泥质量、石墨质量等已被比例特征所代替,在数据处理中予以删除。

特征工程处理后的数据集包含材料类别(水泥材料)、配比参数(水泥质量 $m_{水泥}$、石墨质量 $m_{石墨}$、水质量 $m_{水}$)及电学特性(电阻率 ρ、极化率 η)。构造的交互特征包括水泥比例 $\varphi_{水泥}$、石墨比例 $\varphi_{石墨}$、水比例 $\varphi_{水}$、水泥石墨比 $R_{水泥/石墨}$,单一组分质量比例特征见公式(1),其中 $\varphi_{水泥}+\varphi_{石墨}+\varphi_{水}=1$,满足质量守恒约束;多组分交互特征即水泥石墨比计算见公式(2)。

$$\begin{cases}\varphi_{水泥}=\dfrac{m_{水泥}}{m_{水泥}+m_{石墨}+m_{水}}\\ \varphi_{石墨}=\dfrac{m_{石墨}}{m_{水泥}+m_{石墨}+m_{水}}\\ \varphi_{水}=\dfrac{m_{水}}{m_{水泥}+m_{石墨}+m_{水}}\end{cases}\tag{1}$$

$$R_{水泥/石墨}=\frac{\varphi_{水泥}}{\varphi_{石墨}}\tag{2}$$

1.2.2 编码与标准化处理

数据的数值特征和类别特征的预处理需采用不同方式。对于类别变量即水泥类型,采用独热编码方式,使其转换为可用于建模的数值特征。对于连续变量,如比例特征、电阻率、极化率,则采用标准分数法(Z-score 标准化)进行归一化处理,保证不同特征的量纲一致,避免特征值范围过大或过小影响模型训练。类别变量通过独热编码函数(OneHotEncoder)映射为 6 维稀疏向量 $\boldsymbol{M}$,见公式(3)。比例特征、电阻率和极化率通过公式(4)进行标准化处理。

$$\boldsymbol{M}=\mathrm{OneHotEncoder}(m)\tag{3}$$

$$\bar{x}=\frac{x-\mu_{\mathrm{train}}}{\sigma_{\mathrm{train}}}\tag{4}$$

式中:m 为输入的类别特征变量,即水泥类型,包括 6 种不同的类别;$\boldsymbol{M}\in\{0,1\}^{n\times 6}$;$\mu_{\mathrm{train}}$ 为训练集中特征的均值;σ_{train} 为训练集中特征的标准差;x 为原始特征值;$\bar{x}$ 为标准化处理后的特征值,服从均值为 0、标准差为 1 的正态分布。

1.2.3 物理约束下的数据增强

在样本量较小的情况下,数据增强是一种有效的优化策略。本文采用线性插值法扩展原始数据,即在同一材料类别内,随机选择两组样本进行线性组合,以一定权重生成新样本,从而改善数据的分布情况。增强倍数设定为 1,即生成的新数据量与原始数据量(60 组)相同,使得数据规模扩大一倍(120 组)。将这 120 组数据采用随机分层抽样方法,将特征工程处理后的原始数据集划分为训练集(80%,96 组)和测试集(20%,24 组),并确保各类材料的比例在两个子集中保持一致。为了保证增强数据的合理性,生成的新样本仍需满足物理约束,特别是比例特征需保持总和为 1。

基于材料配比的物理合理性,定义有效样本特征向量 $\boldsymbol{X}$,见公式(5)。其中,$\varphi_{水泥}$,$\varphi_{石墨}$,$\varphi_{水}$ 均满足 $[1-\delta,1+\delta]$,$\delta=0.02$ 为工程允许的配比误差阈值。若样本组分比例之和不属于[0.98,1.02]或任一比例超出(0,1)范围,则判定为异常数据并剔除。

$$\boldsymbol{X}=(\boldsymbol{M},\varphi_{水泥},\varphi_{石墨},\varphi_{水},R_{水泥/石墨},\rho,\eta)\tag{5}$$

从同种类材料类别中随机选取两组样本 $\boldsymbol{X}_1$ 和 $\boldsymbol{X}_2$,按随机权重系数 $\alpha\sim U(0.2,0.8)$ 进行线性插值,生成新的特征 $\boldsymbol{X}_{\mathrm{new}}$,见公式(6)。

$$\boldsymbol{X}_{\mathrm{new}}=\alpha\boldsymbol{X}_1+(1-\alpha)\boldsymbol{X}_2\tag{6}$$

同时引入物理约束,若插值后的组分比例(φ_i,$i\in$(水泥,石墨,水))之和不等于 1,则进行归一化处理,见公式(7)。

$$\varphi_i\leftarrow\frac{\varphi_i}{\sum\varphi_i}\tag{7}$$

2 模型搭建

本章采用 4 种方法(RFR, ANN, SVR, LR)构建注浆材料强度预测模型,涵盖集成学习、神经网络与经典回归算法,并介绍各模型核心架构和训练策略,以全面评估不同算法的性能差异和数据适用性。

2.1 RFR

RFR 属于集成学习中的装袋法(Bagging)模型,其核心思想是利用多个弱学习器构成强学习器,从而提升回归性能。自助法(Bootstrap)是一种非参数统计方法,其核心思想是通过有放回的重采样生成大量新样本,从而模拟统计量的抽样分布。模型引入双重随机性策略,对样本进行 Bootstrap 采样,并对输入特征进行随机子集选择,剩余样本则用于计算泛化误差。

本研究中搭建的 RFR 模型集成了 200 棵决策树,每棵树在限制最大深度为 5 的条件下独立训练。RFR 架构如图 1 所示。

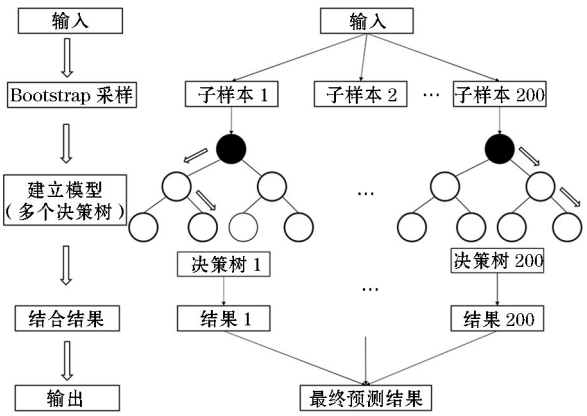


图 1 RFR 架构

Fig.1 Architecture of RFR

每棵树以如下递归方式构建:

$$\hat{y}_{\mathrm{RFR}}=\frac{1}{t}\sum_{\tau=1}^{t}f_{\tau}(\boldsymbol{X}) \tag{8}$$

式中:t 表示决策树数量,设定值为 200;$f_{\tau}(\boldsymbol{X})$ 表示第 τ 棵树对输入样本 $\boldsymbol{X}$ 的预测结果;$\hat{y}_{\mathrm{RFR}}$ 为 RFR 模型对输入样本的最终预测输出值。

RFR 主要超参数设置见表 1。

表 1 RFR 主要超参数

Table 1 Main hyperparameters of RFR

决策树数量	树深度	种子
200	5	随机种子

2.2 ANN

ANN 具有良好的非线性拟合能力,适用于高维、复杂特征的结构映射建模。本研究构建了一种三层全连接(Dense)结构,结合 ReLU 激活函数、丢弃(Dropout)机制及 Adam 优化器,以增强 ANN 的非线性表达能力并提高训练稳定性。

ANN 架构如图 2 所示,包含输入层、两层隐藏层与一层输出层。

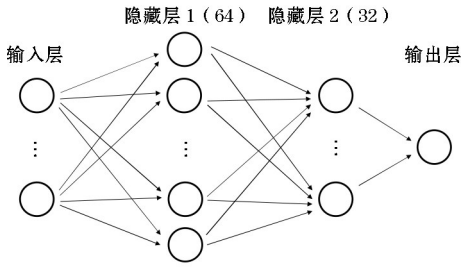


图 2 ANN 架构

Fig.2 Architecture of ANN

各层具体配置如下:

(1) 输入层,输入特征维度为 12;

(2) 隐藏层 1,64 个神经元,ReLU 激活函数,Dropout=0.2;

(3) 隐藏层 2,32 个神经元,ReLU 激活函数,Dropout=0.2;

(4) 输出层,1 个神经元,线性激活,输出预测值。

训练过程采用 MSE 作为损失函数,ANN 主要超参数配置见表 2。

表 2 ANN 主要超参数

Table 2 Main hyperparameters of ANN

优化器	损失函数	早停策略	训练轮数	学习率	批量大小	随机失活比率
自适应矩估计优化器(Adam)	均方误差(*MSE*)	耐心值(Patience)=50	500	0.001	16	0.2

2.3 SVR

SVR 源于支持向量机,通过最大间隔优化的方式在高维特征空间中寻找最佳回归超平面,其非线性回归形式主要依赖核函数映射。在本研究中采用径向基函数(Radial Basis Function, RBF)核函数[15]:

$$K(\boldsymbol{X}_1,\boldsymbol{X}_2)=\exp\left(-\gamma\left\|\boldsymbol{X}_1-\boldsymbol{X}_2\right\|^2\right) \tag{9}$$

式中,γ 为核函数参数。

通过网格搜索在验证集上优化超参数组合,得到 SVR 超参数搜索结果,见表 3。

表 3 SVR 超参数搜索结果

Table 3 Search results of SVR hyperparameters

参数	搜索范围	最优值
正则化参数	[0.1,1,10,100]	10
核函数参数	['scale',0.01,0.1,1]	'scale'
容差	[0.05,0.1,0.2]	0.1

2.4 LR

LR 是最基础的监督学习方法之一。本研究采用标准的最小二乘线性回归,模型训练过程通过最小化均方误差实现。线性回归主要作为基线模型,用于与更复杂的非线性模型(RFR、ANN、SVR)进行性能对比。

该方法通过构建输入特征与输出变量之间的线性映射关系,拟合目标变量:

$$\hat{y}_{LR}=\boldsymbol{\omega}^{T}\boldsymbol{X}+b \tag{10}$$

式中:$\boldsymbol{\omega}^{T}$ 为权重向量的转置;b 为偏置;$\hat{y}_{LR}$ 为 LR 模型对输入样本的最终预测输出值。

3 试验与结果分析

3.1 多模型对比分析

试验首先对数据集进行增强处理,分别训练 4 种模型,并以均方误差(MAE)、决定系数(R^2)、参数量和训练时间作为评价指标。试验结果见表 4,各模型预测结果的可视化如图 3 至图 6 所示。

表 4 模型对比分析结果

Table 4 Comparison analysis results of models

模型	MAE	R^2	参数量	训练时间/s
RFR	2.59	0.934	126 812	0.22
ANN	2.85	0.919	2 945	26.49
SVR	2.99	0.846	1 104	0.07
LR	3.03	0.767	13	0.01

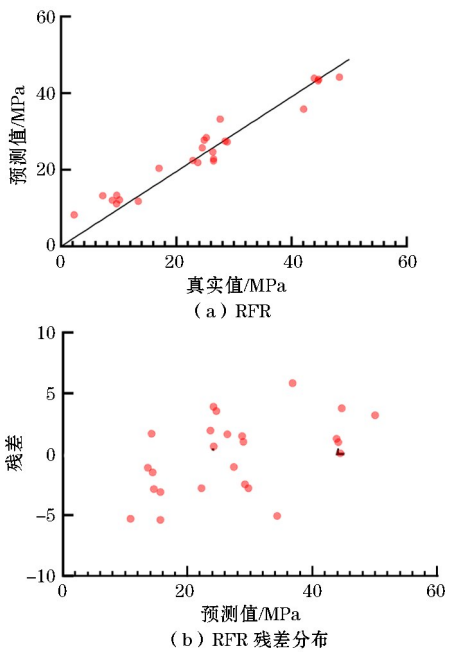
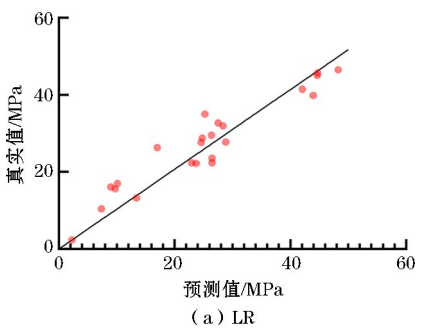


图 3 RFR 结果可视化

Fig.3 Visualization of RFR results



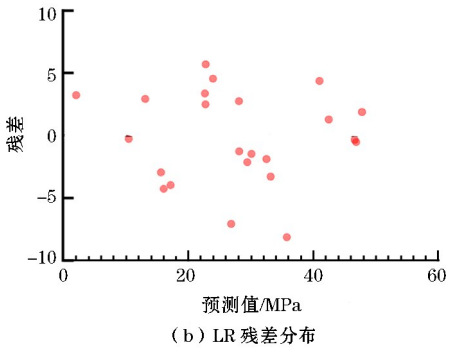


图 4 LR 结果可视化

Fig.4 Visualization of LR results

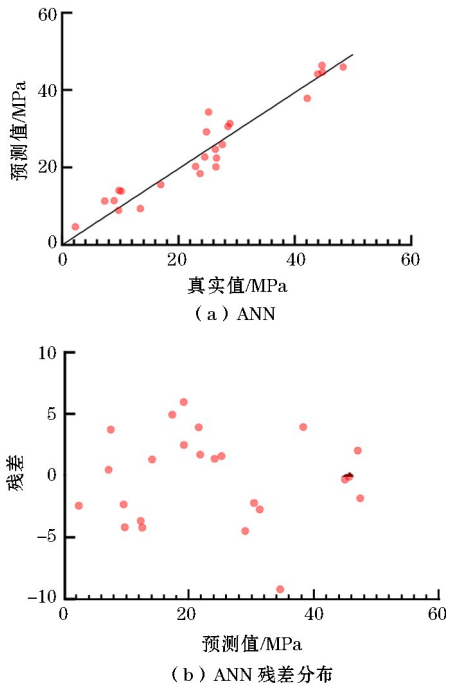


图 5 ANN 结果可视化

Fig.5 Visualization of ANN results

由表 4、图 3 至图 6 可知,RFR 模型表现最优,其在融合物理先验与数据增强机制后,准确率显著提升,测试集 R^2 高达 0.934,MAE 降至 2.59,训练时间(0.22 s)也位居前列,充分展现出了其在小样本场景下的高效性与稳定性。相比之下,ANN 也表现出良好的预测能力,R^2 达 0.919,表明其具备对非线性特征建模的潜力。但受限于迭代机制,ANN 模型训练时间为 26.49 s,存在一定的计算开销,限制了其在实时监测场景中的应用。SVR 和 LR 模型的性能表现一般,SVR 主要受限于参数选择敏感及交互建模能力,LR 则受限于单一性线性映射,导致两种模型的准确度下降。

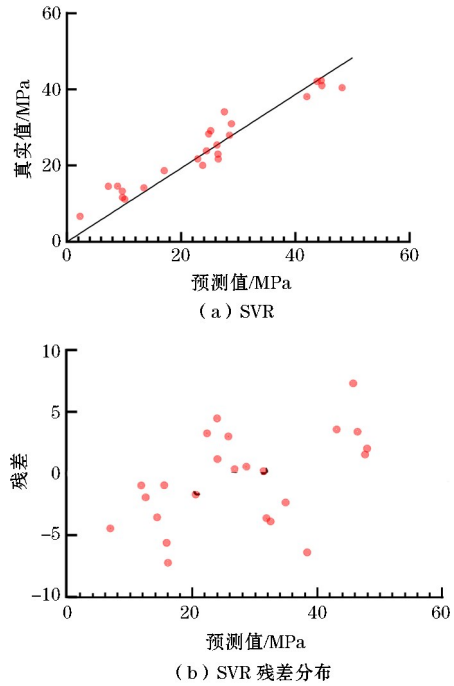


图 6　SVR 结果可视化

Fig.6　Visualization of SVR results

3.2　RFR 特征重要性分析

RFR 特征重要性分析是通过基尼指数反映各特征对预测误差的贡献度，特征重要性数值越高，表明该特征对材料强度的调控作用越显著。特征重要性分析结果如图 7 所示。由图 7 可知，石墨比例（特征重要性贡献 37.6%）与水泥比例（特征重要性贡献 23.5%）为主要决定因子，在调控材料强度中占据主导作用。铝酸盐水泥（特征重要性贡献 9.3%）的特征重要性高于其他水泥类型，符合其作为特种水泥的特点。上述量化分析结果可为工程实践中配比材料的优选提供参考依据。

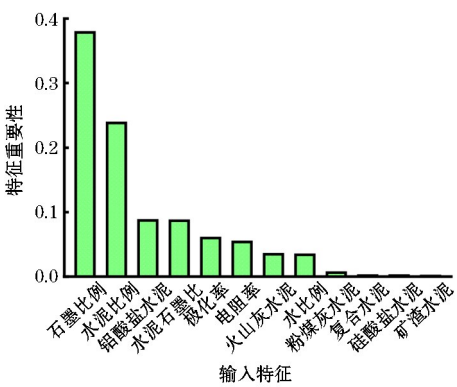


图 7　输入特征的 RFR 特征重要性分析结果

Fig.7　Analysis results of RFR feature importance for input features

综上可知，RFR 特征影响力排序依次为：石墨＞水泥＞极化率＞电阻率＞水。

水泥类型影响力排序依次为：铝酸盐水泥＞火山灰水泥＞粉煤灰水泥＞复合水泥＞普通硅酸盐水泥＞矿渣水泥。

3.3　数据增强对比试验

数据增强对比试验结果见表 5。由表 5 可知，数据增强显著提升了 RFR 与 ANN 两种模型的性能，RFR 的 *MAE* 由 4.30 降低至 2.59（下降 39.8%），R^2 从 0.741 升至 0.934，模型对强度变化的拟合能力显著增强。ANN 模型同样表现出稳定增益，*MAE* 下降 18.3%，R^2 由 0.818 提升至 0.920。SVR 和 LR 模型数据增强后的性能虽有所提升，但拟合能力有限，未纳入数据增强对比试验。综上可知，数据增强不仅能扩充样本空间，还可以提升模型的泛化能力与鲁棒性，特别是对 RFR 与 ANN 模型有显著作用。

表 5　数据增强对比试验结果

Table 5　Comparative test results of data augmentation

模型	数据增强	*MAE*	R^2
RFR	是	2.59	0.934
RFR	否	4.30	0.741
ANN	是	2.85	0.920
ANN	否	3.49	0.818

3.4　模型稳定性评估

为评估模型在小样本多类别材料数据中的鲁棒性，本研究采用分层 K 折交叉验证（$K=5$）对 120 组水泥性能数据进行验证，结果见表 6。通过分层抽样确保训练集（96 组样本）与验证集（24 组样本）的类别分布均衡，对比 RFR、ANN、SVR 和 LR 模型的预测性能。由表 6 可知，RFR 模型在稳定性与精度上均显著优于其他模型，其 *MAE* 均值最低，为 2.73，*MAE* 标准差仅为 0.31，较 SVR 模型降低65.2%，同时 R^2 均值达到了 0.912，R^2 标准差为0.021，验证了 RFR 模型的 Bagging 集成机制对数据波动的强适应能力。交叉验证结果显示，模型在未知数据上的误差增幅控制在5.4%～8.7%之间，表明该方法通过多轮数据划分有效规避了单次验证的随机偏差，为工程材料小样本建模提供了可靠性验证方案。

表 6　交叉验证性能对比结果

Table 6　Results of cross-validate performance comparisons

模型	*MAE* 均值	*MAE* 标准差	R^2 均值	R^2 标准差
RFR	2.73	0.31	0.912	0.021
ANN	3.02	0.45	0.874	0.035
SVR	3.25	0.89	0.803	0.068
LR	3.41	0.52	0.721	0.042

3.5 模型泛化性评估

为系统评估模型的泛化性能,本研究结合文献[16]中粉煤灰掺量对浆液性能的影响数据,筛选原始数据后对 RFR 模型进行参数微调(决策树数量=100,最大深度=3)。通过 RFR 模型预测不同水固比(1∶1、1∶1.1、1∶1.2、1∶1.3)及水泥-粉煤灰配比(1∶9、3∶7、5∶5)的 12 组样本的 28 d 强度,结果见表 7。由表 7 可知,RFR 模型对粉煤灰掺量样本展现出较好的泛化能力,各水固比条件下预测误差均稳定低于 10%,验证了 RFR 模型对材料主成分特征的有效捕捉;但在极端配比条件下,模型预测结果出现显著偏差。研究结果表明,RFR 模型在常规工程配比范围内具有较好的预测效果,但在极端低掺量条件下的预测存在一定瓶颈,后续需通过引入物理先验知识或改进特征工程等方式进一步提升模型的鲁棒性。

表 7 粉煤灰水泥 28 d 强度预测结果

Table 7 Prediction results of 28 d strength of fly ash cement

水固比	水泥∶粉煤灰	真实值/MPa	预测值/MPa	绝对误差/MPa
1∶1	1∶9	0.198	0.52	0.322
	3∶7	1.450	1.78	0.330
	5∶5	3.940	4.32	0.380
1∶1.1	1∶9	0.199	0.28	0.081
	3∶7	1.800	1.23	0.570
	5∶5	4.200	4.85	0.650
1∶1.2	1∶9	0.300	0.81	0.510
	3∶7	2.200	2.85	0.650
	5∶5	4.500	4.12	0.380
1∶1.3	1∶9	0.340	0.45	0.110
	3∶7	2.760	3.67	0.910
	5∶5	4.700	4.45	0.250

4 结论

本文针对注浆材料强度预测中的小样本建模难题,提出了融合物理机理与数据增强的机器学习框架,对随机森林回归(Random Forest Regression, RFR)、人工神经网络(Artificial Neural Network, ANN)、支持向量机回归(Support Vector Regression, SVR)和线性回归(Linear Regression, LR) 4 种机器学习方法的材料性能预测结果进行对比分析,得到如下主要结论。

(1) 在融合物理先验与数据增强机制后,RFR 模型的预测结果最优,其 R^2 高达 0.934,较 LR、SVR、ANN 模型分别提高了 21.8%、10.4%、1.6%;*MAE* 降至 2.59,较 LR、SVR、ANN 模型分别降低了 14.5%、13.4%、9.1%,训练时间为 0.22 s,验证了 RFR 模型在小样本场景下的高效性和稳定性。

(2) 材料配比特征对强度预测具有主导性影响。特征重要性分析中,石墨比例(特征重要性贡献为 37.6%)、水泥比例(特征重要性贡献为 23.5%)占主导作用,揭示了胶凝体系中固相占比对材料强度的关键调控作用,这一结果与材料力学中的孔隙率理论相吻合,增强了模型的可解释性。

(3) 物理引导的数据增强方法可有效提升小样本建模性能,RFR 和 ANN 模型的 *MAE* 较未纳入数据增强时分别提升 39.8%和 18.3%;RFR 模型在稳定性与精度上均显著优于其他模型,*MAE* 均值最低(为 2.73),*MAE* 标准差仅为 0.31(较 SVR 模型降低 65.2%),同时 R^2 均值达到 0.912,同时 R^2 标准差为 0.021,验证了 RFR 模型的 Bagging 集成机制对数据波动的强适应能力;RFR 模型在各水固比条件下的预测误差均稳定低于 10%。

(4) 本研究通过物理约束的数据增强提升了注浆材料强度预测的小样本建模能力,可为矿山注浆加固材料配比提供参考。未来研究需深度融合多尺度材料表征数据,结合微观数据与跨领域物理约束,构建动态配比优化框架的轻量化模型,并通过迁移学习共享不同材料体系的特征规律,以平衡模型预测精度及工程部署可行性的多重需求,为复杂岩层条件下的安全高效开采提供数字化解决方法。

参考文献(References):

[1] 吴元,王凯,杨晓婧,等.水泥基灌浆料基本力学性能试验研究[J].建筑结构,2014,44(19):95-98.
WU Yuan, WANG Kai, YANG Xiaojing, et al. Experimental study on basic mechanical properties of cementitious grout[J]. Building Structure,2014,44(19):95-98.

[2] 李化建,万广培,谢永江,等.矿物掺和料-水泥浆体电学特性与其微结构的关系[J].土木建筑与环境工程,2013,35(4):121-127.
LI Huajian, WAN Guangpei, XIE Yongjiang, et al. Relationship between electrical properties and microstructure of mineral admixture-cement paste[J]. Journal of Civil and Environmental Engineering,2013,35(4):121-127.

[3] 刘明利,申康,刘钟涛.多因素影响下的充填体强度预测与分析[J].矿业研究与开发,2023,43(7):5-11.
LIU Mingli, SHEN Kang, LIU Zhongtao. Prediction and analysis of backfill strength under the influence of multiple factors[J]. Mining Research and Development,2023,43(7):5-11.

[4] 董军锋,王耀南,雷波.《回弹法检测水泥基灌浆材料抗压强度技术规程》编制的试验研究[J].建筑结构,2022,52(2):85-88.

DONG Junfeng, WANG Yaonan, LEI Bo. Experimental study of *technical specification for inspecting compressive strength of cementitious grout by rebound method* [J]. Building Structure,2022,52(2):85-88.

[5] 陈海彬,卢建勇,李凯.钻芯法检测混凝土强度芯样尺寸效应的研究[J].混凝土,2014(3):141-143.
CHEN Haibin, LU Jianyong, LI Kai. Research on size effect of core samples for testing concrete strength with drilled core method[J]. Concrete,2014(3):141-143.

[6] 曹国栋,王辉,邓代强,等.考虑水化度的混凝土抗压强度数值预测研究[J].湘潭大学学报(自然科学版),2024,46(5):79-91.
CAO Guodong, WANG Hui, DENG Daiqiang, et al. Numerical prediction of concrete compressive strength considering hydration degree [J]. Journal of Xiangtan University (Natural Science Edition),2024,46(5):79-91.

[7] 李自强,崔素萍,马忠诚,等.机器学习算法用于水泥强度预测的研究进展[J].材料导报,2025,39(5):195-206.
LI Ziqiang, CUI Suping, MA Zhongcheng, et al. Research progress of machine learning algorithm for cement strength prediction[J]. Materials Reports, 2025, 39(5): 195-206.

[8] 刘德胜.基于机器学习算法的混凝土28 d抗压强度预测[J].混凝土与水泥制品,2022(9):20-24.
LIU Desheng. The prediction of 28 d compressive strength of concrete based on machine learning algorithms[J]. China Concrete and Cement Products,2022(9):20-24.

[9] 李其廉,陈佳尧,敦彦茹,等.基于机器学习的水泥基灌浆料强度预测方法[J].河北科技大学学报,2024,45(3):308-317.
LI Qilian, CHEN Jiayao, DUN Yanru, et al. Machine learning based strength prediction method for cement-based grouting material[J]. Journal of Hebei University of Science and Technology,2024,45(3):308-317.

[10] 崔晓宁,王起才,张戎令,等.基于随机森林的高性能混凝土抗压强度预测[J].兰州交通大学学报,2021,40(6):1-6+14.
CUI Xiaoning, WANG Qicai, ZHANG Rongling, et al. Prediction of compressive strength of concrete based on random forests[J]. Journal of Lanzhou Jiaotong University, 2021,40(6):1-6+14.

[11] 吴贤国,刘鹏程,陈虹宇,等.基于随机森林的高性能混凝土抗压强度预测[J].混凝土,2022(1):17-20+24.
WU Xianguo, LIU Pengcheng, CHEN Hongyu, et al. Characteristic screening and prediction of high-performance concrete compressive strength based on random forest method[J]. Concrete,2022(1):17-20+24.

[12] 张懂,张龙,李文波.基于人工神经网络的固化土抗压强度预测研究[J].安徽建筑,2025,32(2):89-92.
ZHANG Dong, ZHANG Long, LI Wenbo. Research on prediction of compressive strength of solidified soil based on artificial neural network[J]. Anhui Architecture,2025,32(2):89-92.

[13] 林海飞,杨二豪,赵鹏翔,等.类岩石材料力学特性参数多元线性回归模型[J].西安科技大学学报,2018,38(3):351-359.
LIN Haifei, YANG Erhao, ZHAO Pengxiang, et al. Multivariate linear regression model of the mechanical properties of rock-like materials [J]. Journal of Xi'an University of Science and Technology,2018,38(3):351-359.

[14] 杨波,靳壮壮,刘洋.基于有限元线性插值法的泡沫混凝土配合比优化[J].建筑结构,2023,53(增刊2):1376-1380.
YANG Bo, JIN Zhuangzhuang, LIU Yang. Mix proportion optimization of foam concrete based on finite element linear interpolation method [J]. Building Structure, 2023, 53(Suppl.2):1376-1380.

[15] LIU Xuewei, WANG Sai, LIU Bin, et al. Development of cement-based grouting materials and prediction of material properties using PSO-RBF machine learning [J]. Construction and Building Materials,2024,417:135328.

[16] 曹晓毅,刘小平,武博强.水泥-粉煤灰浆液试验及其在采空区注浆中的应用[J].西部探矿工程,2013,25(6):4-7.
CAO Xiaoyi, LIU Xiaoping, WU Boqiang. The test of cement fly ash grout and its application in goaf grouting[J]. West-China Exploration Engineering,2013,25(6):4-7.