

# 基于随机森林算法和数据增强策略的拘束 相关断裂韧性预测研究\*

单康中<sup>1</sup> 王骁晓<sup>2</sup> 刘芳<sup>3</sup> 崔元元<sup>3</sup> 杨杰<sup>1\*\*</sup>

(<sup>1</sup> 上海理工大学能源与动力工程学院,上海市动力工程多相流动与传热重点实验室,上海,200093)

(<sup>2</sup> 华东理工大学承压系统与安全教育部重点实验室,上海,200237)(<sup>3</sup> 上海理工大学机械工程学院,上海,200093)

**摘要** 选用核电用钢 A508 为研究对象,对 K 近邻回归(KNN)、核回归(KR)、线性回归(LR)和随机森林回归(RF)四种算法拘束相关断裂韧性的预测能力进行了考察,发现四种算法对拘束相关断裂韧性的预测效果为 RF > LR > KNN > KR. 进一步,基于 RF 算法,在数据中加入了平面应变状态下的数据进行数据增强,对单边缺口弯曲 (SENB) 试样的拘束相关断裂韧性进行了预测与验证,并基于验证过的模型,向单边裂纹拉伸 (SENT)、紧凑拉伸 (CT) 和中心裂纹拉伸 (CCT) 试样进行了移植. 结果表明:通过添加数据增强策略,试样在边界点处的断裂韧性预测能力得到有效提高,预测结果更加准确,基于 RF 算法并添加数据增强策略的模型具有较好的泛化能力. 最后,在基于 RF 算法和数据增强的基础上,加入试样类别这一特征,构建了准确度较高的统一预测模型.

**关键词** 拘束,断裂韧性,机器学习,随机森林算法,数据增强

**DOI:** 10.19636/j.cnki.cjasm42-1250/o3.2024.044

## 0 引言

拘束可以看作结构对裂纹尖端区域材料塑性变形的阻碍,包括面内拘束和面外拘束. 面内拘束受裂纹扩展方向上结构尺寸的影响,面外拘束则受与裂尖平面相平行方向上结构尺寸的影响. 拘束显著影响着材料的断裂行为,拘束的丧失会导致材料断裂韧性的增加<sup>[1]</sup>. 为进行准确的结构完整性评定,需要考虑拘束对材料断裂韧性的影响.

在过去的几十年里,针对拘束对断裂韧性的影响已进行了很多研究. 近几年,张经纬等<sup>[2]</sup>对紧凑拉伸 (CT) 试样在不同厚度及裂纹长度下的断裂韧性进行了试验研究,并通过有限元分析了上述试样的拘束水平,建立了表观断裂韧度与拘束参数  $Q$  的关联方程. 张宏等<sup>[3]</sup>基于细观损伤力学理论,采用数值模拟方法建立了环焊缝裂纹扩展阻力数值计算模型,针对低强、等强以及高强 3 种焊缝强度匹配条件,

分别计算了面内、面外拘束耦合下以裂纹尖端张开位移 (CTOD) 为表征的环焊缝断裂韧性,并比较了使用不同拘束参数来描述断裂韧性与拘束关联的合理性. 孙彤等<sup>[4]</sup>采用理论分析、宏观试验、数值模拟相结合的方法,探究了面内、面外拘束对 G20Mn5QT 铸钢断裂韧性的影响,并建立了拘束与该材料断裂韧性之间的关联. Huang 等<sup>[5]</sup>建立了一个利用弹塑性相场模型和改进的边界层公式预测拘束效应对断裂韧性影响的新框架.

随着科学的发展,科学研究在历经经验科学、理论科学、计算科学三个范式后,如今已迈入数据驱动科学范式<sup>[6,7]</sup>. 以机器学习为主的数据驱动方法正在被广泛应用于各个领域<sup>[8,9]</sup>. 在断裂韧性预测领域, Ma 等<sup>[10]</sup>基于采集的 Nb-Si 基合金数据,利用机器学习模型分析了 Nb-Si 基合金室温断裂韧性与合金元素的相关性. Ibrahim 等<sup>[11]</sup>基于中心直切口巴西圆盘 (CSNBD) 测试生成的 420 个数据集,提出了 15 种有监督和无监督学习算法,对混凝土的断裂韧

\* 国家自然科学基金项目 (Nos. 52375154 和 52311530067) 资助.

2024-09-15 收到修改稿, 2024-09-20 网络首发.

\*\* 通讯作者. E-mail: yangjie@usst.edu.cn

性进行了估算. 并通过与试验结果的对比, 评估了所应用模型的性能和准确性. Hu 等<sup>[12]</sup>使用了六种机器学习模型(支持向量机(SVM)、随机森林(RF)、反向传播神经网络(BPNN)、反向传播-粒子群优化(BP-PSO)、卷积神经网络(CNN)和径向基函数神经网络(RBF))预测了岩石在不同温度下的断裂韧性. 可以发现, 机器学习在断裂韧性预测中的应用研究还不多, 且在研究中并没有考虑拘束的影响.

此外, 虽然数据驱动方法正在被广泛应用, 但囿于决策的透明度和结果的可解释性, 数据驱动机器学习模型尚难以部署到一些性能要求高的关键领域<sup>[13]</sup>. 在最近的研究中, 一些学者将数据驱动与机理驱动相结合, 通过增加边界条件对模型参数进行约束, 以达到增强数据驱动可解释性、提高泛化能力、得到符合物理规律预测结果的效果<sup>[14-16]</sup>. 然而, 目前结合机理驱动的方法仍处于探究阶段, 且对边界条件的要求较为苛刻, 数据机理双驱动下的融合建模和合理解释仍然是学者们面临的一个难题.

鉴于此, 本文选用核电用钢 A508 为研究对象, 首先考察了 K 邻近回归(KNN)、核回归(KR)、线性回归(LR)和 RF 回归四种算法对拘束相关断裂韧性的预测能力, 在筛选出预测结果最好的算法后, 基于该算法, 对不同试样的拘束相关断裂韧性进行了预测. 进一步, 通过添加数据增强策略, 对单边缺口弯曲(SENb)试样的拘束相关断裂韧性进行了预测与验证, 并将验证过的模型向单边裂纹拉伸(SENT)、CT 和中心裂纹拉伸(CCT)试样进行了移植. 最后, 在基于 RF 算法和数据增强的基础上, 加入试样类别特征, 构建了准确度较高的统一模型, 实现了不同试样准确的拘束相关断裂韧性预测.

## 1 数据来源及机器学习算法

### 1.1 不同试样的拘束相关断裂韧性

本文从文献<sup>[17-19]</sup>中获取了 A508 钢在不同拘束下的断裂韧性. 这些断裂韧性分别在不同试样中获得, 包括 CT、SENb、SENT 及 CCT 试样. 针对每一种试样, 通过改变裂纹尺寸(裂纹长度  $a$ 、裂纹长度与试样宽度的比值  $a/W$ ) 达到改变面内拘束的目的, 通过改变试样厚度(试样厚度  $B$ 、试样厚度与

试样宽度的比值  $B/W$ 、平面应变状态) 达到改变面外拘束的目的. 四种试样的示意图如图 1 所示, 不同试样数据集的数量如表 1 所示.

表 1 不同试样数据集的数量

Table 1 Number of datasets for different specimens

Specimens	Numbers
SENb	18
SENT	20
CT	10
CCT	20

### 1.2 四种机器学习算法

#### 1.2.1 KNN 算法

KNN 具有简便性和易实现性, 因受数据点噪声的影响较小, 因此被广泛应用. 该算法假设输入的新数据对应的输出值在各个样本对应的输出值空间中相似, 最相似的 K 个数据点及其各个特征具有相同权重. 故输入新数据后, KNN 提取样本中与新数据最相似的 K 个数据, 通过取 K 个数据对应输出值的均值作为新数据的输出值.

KNN 通常采用欧式距离对数据点和新数据点进行距离度量, 新数据点  $x_0 = [x_{0,1}, x_{0,2}, \dots, x_{0,j}, \dots, x_{0,n}]$  和样本点  $x_i = [x_{i,1}, x_{i,2}, \dots, x_{i,j}, \dots, x_{i,n}]$  之间的距离为:

$$d(x_0, x_i) = \left[ \sum_{j=1}^n (x_{i,j} - x_{0,j})^2 \right]^{1/2} \quad (1)$$

式中:  $x_{i,j}$  和  $x_{0,j}$  分别为第  $i$  个数据点与新数据点的第  $j$  个特征属性值.

#### 1.2.2 KR 算法

KR 算法是一种为建立随机变量  $X$  和  $Y$  之间的非线性关系而存在的利用非参数的随机变量条件期望进行预测的回归方法. 对于所有的非参数回归, 在变量  $X$  的情况下, 变量  $Y$  的条件期望数学表达式可以写成:

$$E(X|Y) = r(X) \quad (2)$$

其中,  $r$  是回归函数. 常用的核函数包括高斯核函数(Gaussian Kernel)、多项式核函数(Polynomial Kernel)和线性核函数(Linear Kernel)等.

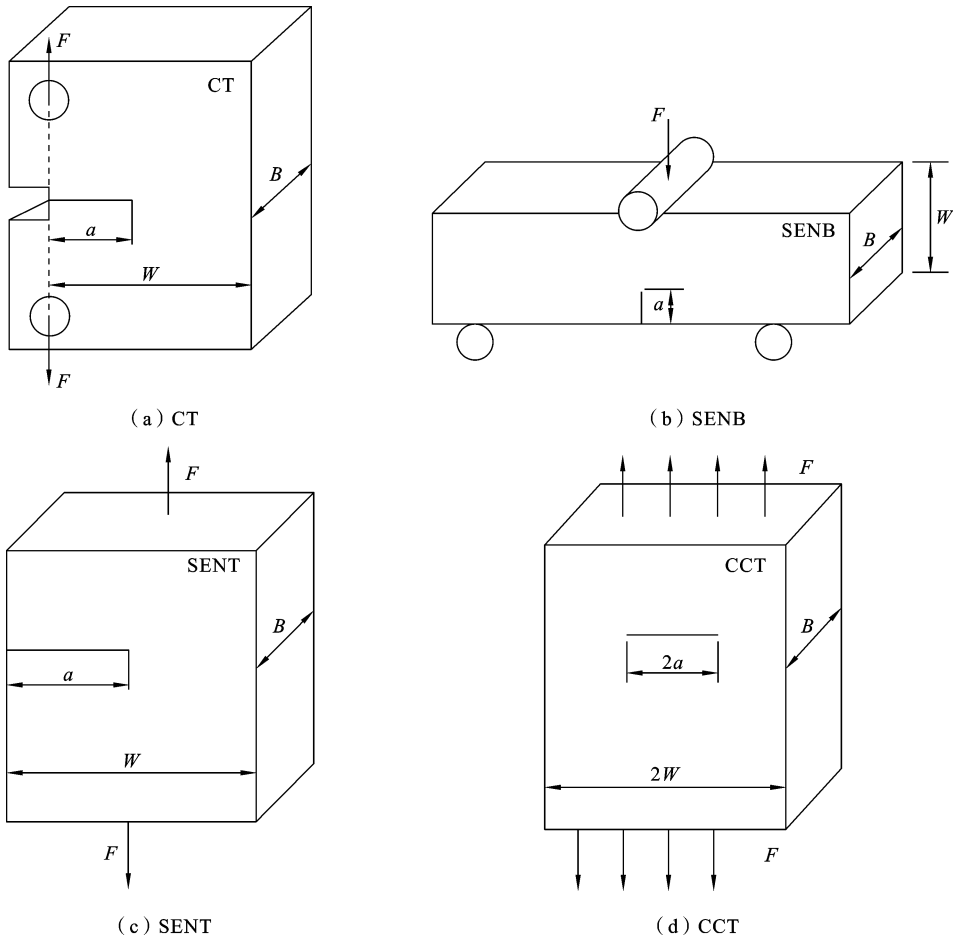


图 1 四种试样的示意图<sup>[17-19]</sup>

Fig. 1 Schematic diagram of four specimens<sup>[17-19]</sup>

### 1.2.3 LR 算法

LR 算法是在给定的训练集上学习得到一个线性函数,在损失函数的约束下,求解相关系数,最终在测试集上测试模型的回归效果.若回归分析中仅包括一个自变量  $x$  和一个因变量  $y$ ,且二者的关系可用一条直线近似表示,这种回归分析称为一元线性回归分析:

$$y = a + bx \tag{3}$$

式中  $a$  和  $b$  为回归系数;若回归分析中包括 2 个或 2 个以上的自变量,且因变量和自变量之间是线性关系,这种回归分析称为多元线性回归分析:

$$y = a + b_1x_1 + b_2x_2 + e \tag{4}$$

式中:  $x_1, x_2$  为自变量;  $y$  为因变量;  $a$  为截距;  $b$  为回归系数;  $e$  为误差项.

### 1.2.4 RF 算法

RF 算法由多个个体决策树构成,并利用投票

机制进行决策分类及回归预测.其中每棵决策树的分割节点数目均根据样本特征数随机确定,这种随机性也使得集合决策树具备更优的预测性能.一个包含  $K$  棵决策树的 RF 算法模型如图 2 所示.

在图 2 中,首先对原始样本集进行抽样以获得  $K$  个样本,并将其建模成决策树,即弱分类器;接着每个弱分类器均会产生相应的决策结果,再选择相应的变量建模成决策树,以此获得具有  $K$  棵决策树的随机森林;最后利用投票机制寻得最高分的树,且将其作为最终结果输出. RF 算法模型 RF 可表示为:

$$R_F = \max_{i,K} \prod_{k=1}^K \eta(C_k, S_i) \tag{5}$$

式中,  $S_i$  是第  $i$  个样本集;  $C_k$  是第  $k$  棵决策树;  $\eta(C_k, S_i)$  是  $C_k$  对  $S_i$  的决策值;  $\max$  是投票机制的函数式.

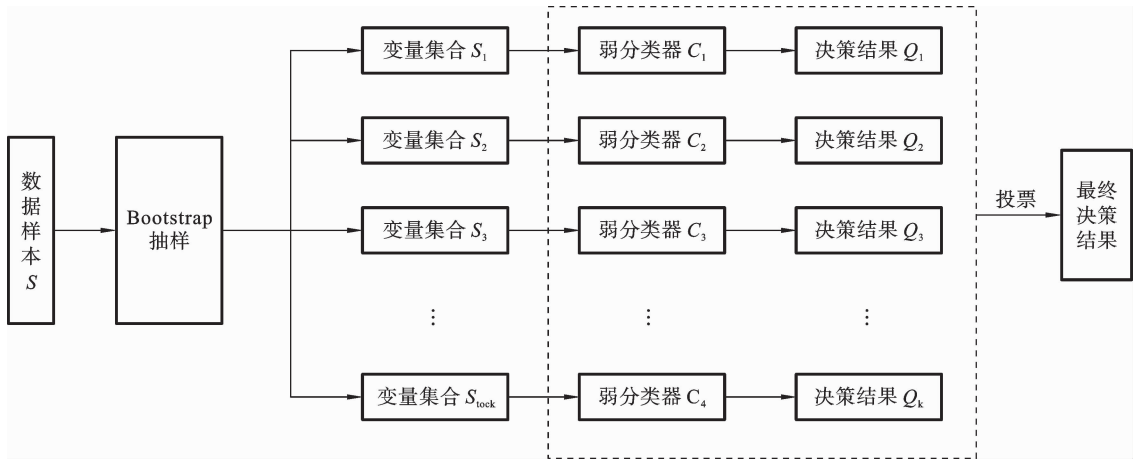


图 2 随机森林(RF)算法决策树的构建过程

Fig. 2 The constructing process of a decision tree based on random forest (RF) algorithm

### 1.3 数据增强策略

数据增强是一种通过让有限的的数据产生更多的等价数据来人工扩展数据集的技术. 在对比学习中, 数据增强不仅可以生成正样本, 也能改善对比学习在源领域与目标领域间的迁移能力. 在传统对比学习方法中, 数据增强除拥有数据增强的基本特性(如增加训练样本数量)外, 主要用于生成锚点样本的正样本.

本文中, 采用平面应变状态下的数据进行数据增强.

## 2 结果与讨论

### 2.1 不同机器学习算法的对比

首先对所有试样进行特征工程, 所有试样输入特征( $a/W, B/W, B, W$  和  $a$ )之间的相关性矩阵如表 2 所示. 从表中可以发现,  $a/W$  与  $a, B/W$  与  $B$  间相关性较强, 其余各特征之间相关性较弱, 不存在明显强相关性, 而  $a/W$  与  $a, B/W$  与  $B$  是后期需要进行对比的对象.

表 2 特征相关性

Table 2 The correlation of feature

	$a/W$	$B/W$	$B$	$W$	$a$
$a/W$	1.000000	-0.048381	-0.047423	-0.004881	0.931996
$B/W$	-0.048381	1.000000	0.929033	-0.106963	-0.140845
$B$	-0.047423	0.929033	1.000000	0.111798	-0.039838
$W$	-0.004881	-0.106963	0.111798	1.000000	0.219458
$a$	0.931996	-0.140845	-0.039838	0.219458	1.000000

为了探索更适合拘束相关断裂韧性预测的算法, 针对不同面内拘束下的 SENB 试样, 对节 1.2 中所介绍四种机器学习算法的预测结果进行了对比, 如图 3 所示.

由图可见, 随着面内拘束( $a/W$ )的增加, SENB 试样的拘束相关断裂韧性逐渐降低, 符合其变化规

律. 四种算法中, RF 算法的预测结果与实际结果最接近, 预测效果最好, KR 算法的预测效果最差. 四种算法的拟合系数  $r^2$  如表 3 所示, 从表中可以更加清晰的发现四种算法对 SENB 试样拘束相关断裂韧性的预测效果为  $RF > LR > KNN > KR$ .

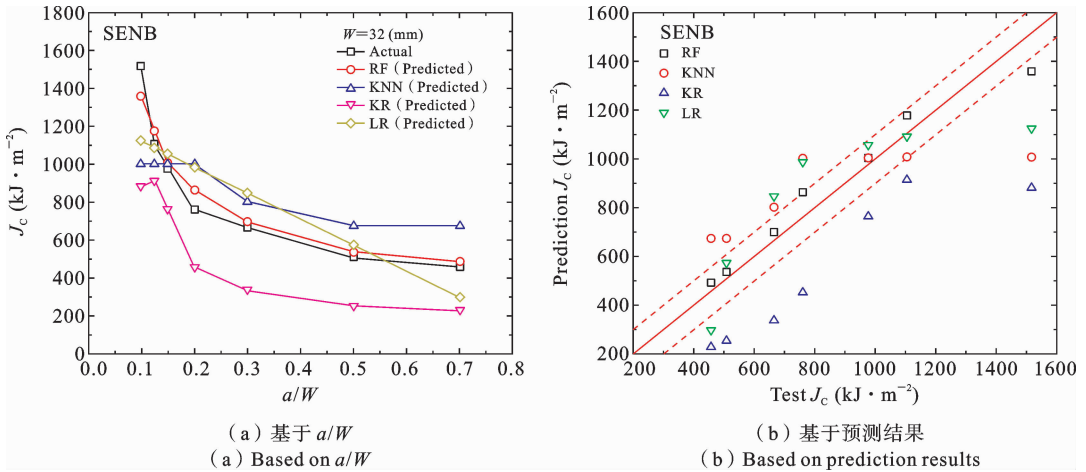


图 3 四种机器学习算法预测结果的对比

Fig. 3 Comparison of prediction results of four machine learning algorithms

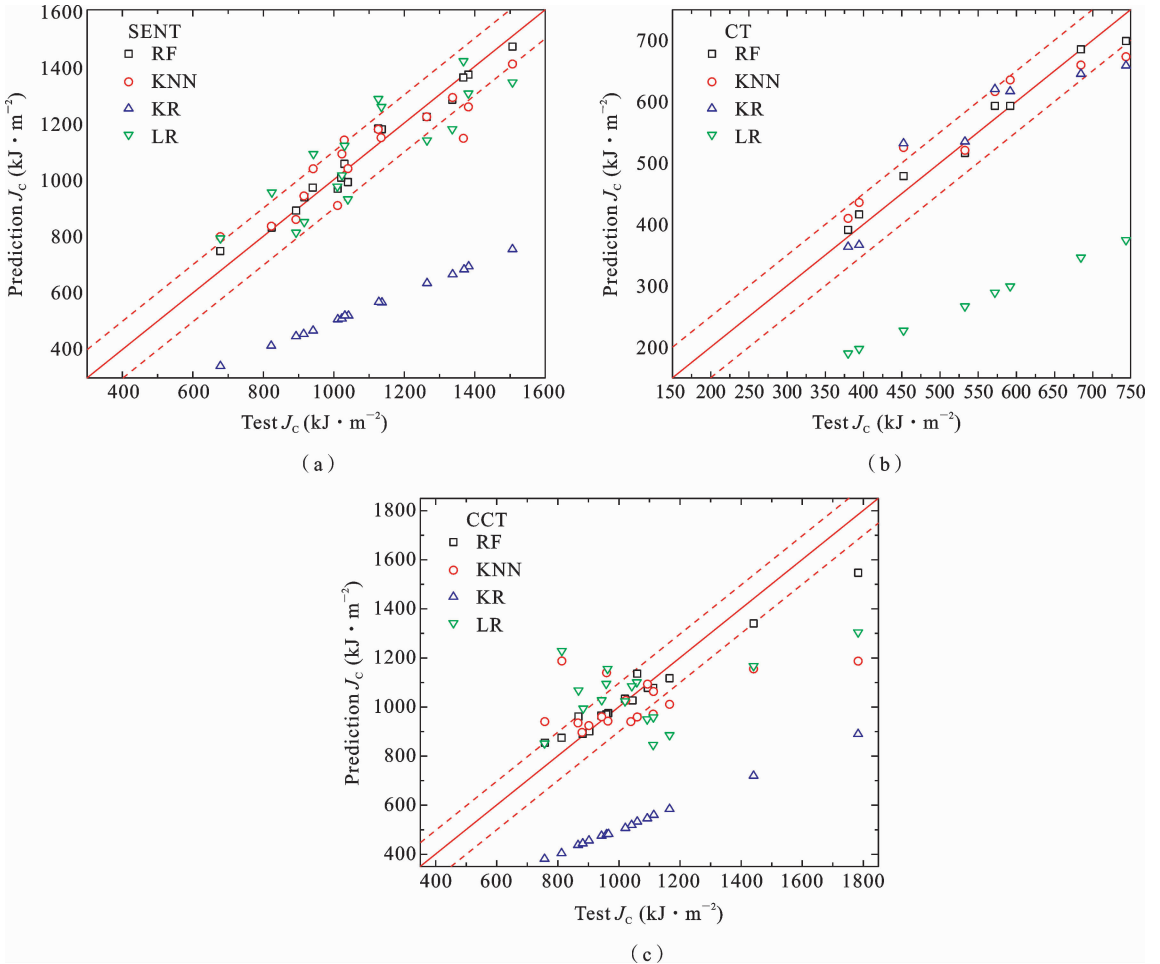


图 4 SENT(a)、CT(b)和 CCT 试样(c) 四种机器学习算法预测结果的对比

Fig. 4 Comparison of the prediction results of four machine learning algorithms for SENT (a), CT (b) and CCT specimens (c)

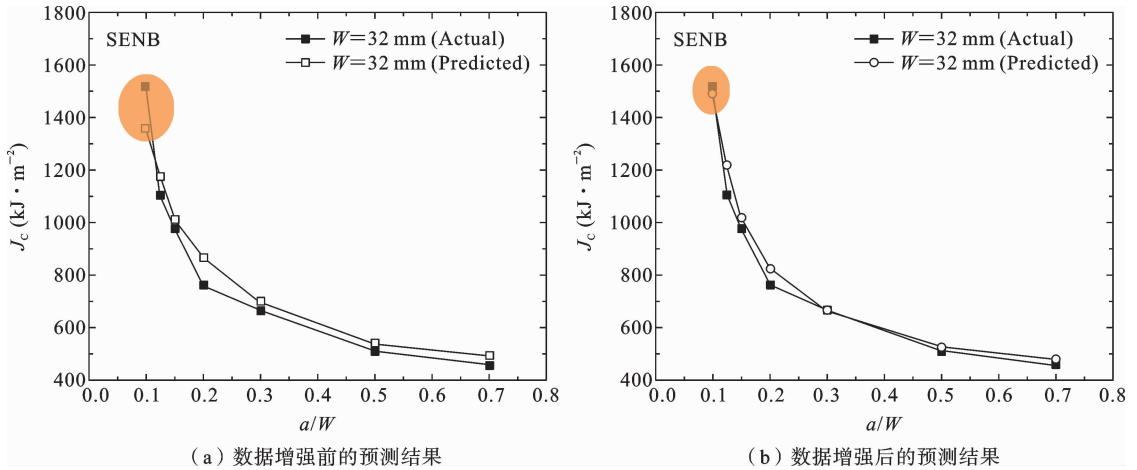


图 5 不同面内拘束 SENB 试样数据增强前后预测结果对比

Fig. 5 Comparison of prediction results for SENB specimens with different in-plane constraints before and after data enhancement

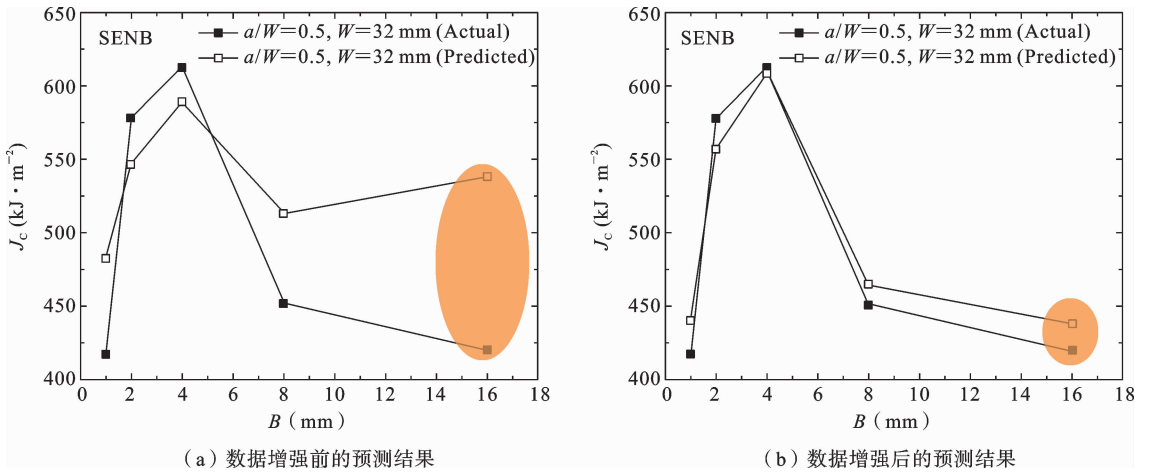


图 6 不同面外拘束 SENB 试样数据增强前后预测结果对比

Fig. 6 Comparison of prediction results for SENB specimens with different out-of-plane constraints before and after data enhancement

表 3 四种算法拟合系数  $r^2$  对比

Table 3 Comparison of the fitting coefficients  $r^2$  of the four algorithms

Model	$r^2$
RF	0.96495
KNN	0.48469
KR	-0.003487
LR	0.71528

了四种算法预测效果的对比,如图 4 所示.由图可见,四种算法的预测效果中均为 RF 最好,故后文中将选取 RF 算法进行进一步的研究.

### 2.2 数据增强对预测结果的影响

在预测过程中,发现对于试样边界点的预测效果较差(如图 5(a)中  $a/W = 0.1$  时),为提高预测的准确性,针对不同面内拘束下的 SENB 试样,在使用 RF 算法进行预测的同时在数据中加入平面应变状态下的数据以进行数据增强,数据增强前后其预测结果对比如图 5 所示.由图可见,进行数据增强后,预测结果更加准确,预测效果更好.

进一步,针对 SENT、CT 和 CCT 试样,均进行

尤其是在  $a/W=0.1$  边界点处, 预测结果得到了改善. 图 5(a) 的拟合系数  $r^2$  为 0.96495, 图 5(b) 的拟合系数  $r^2$  为 0.98375, 数据增强后预测准确度进一步提高.

针对不同面外拘束下的 SENB 试样, 同样使用 RF 算法进行预测, 并加入平面应变时的数据进行数据增强, 数据增强前后其预测结果对比如图 6 所示. 由图可见, 进行数据增强后, 预测效果得到明显改善, 尤其是在  $B=16$  mm 边界点处. 图 6(a) 的拟合系数  $r^2$  为 0.89373, 图 6(b) 的拟合系数  $r^2$  为 0.93841, 数据增强后预测准确度显著提高.

使用训练好的模型, 对面内、面外拘束同时变化时的 4 个试样进行预测结果验证, 如图 7 所示. 由图可见, 预测结果与实际结果吻合良好,  $r^2$  为 0.98036, 说明使用 RF 算法并添加数据增强策略后的模型具有较好的泛化能力.

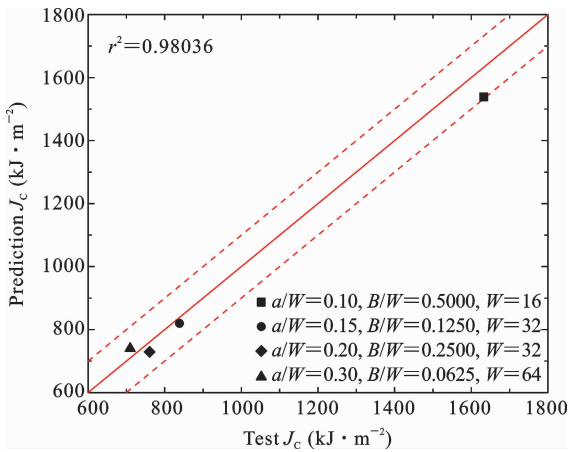


图 7 面内、面外拘束同时变化时的预测结果验证

Fig.7 Verification of prediction results when in-plane and out-of-plane constraints change simultaneously

在进行试样设计时, 笔者发现一些学者通过改变  $a$  和  $B$  来改变拘束, 还有一些学者通过改变  $a/W$  和  $B/W$  来改变拘束, 故基于使用 RF 算法并添加数据增强策略后的模型, 对这 4 个拘束变化关键特征量进行提取并进行重要性排序, 如图 8 所示. 由图可见, 对于 SENB 试样, 4 个关键特征量的重要性排序为  $a/W$ 、 $a$ 、 $B/W$ 、 $B$ , 当设计试样来研究拘束变化的影响时, 通过改变  $a/W$  和  $B/W$  来改变拘束所达到的效果会更加明显.

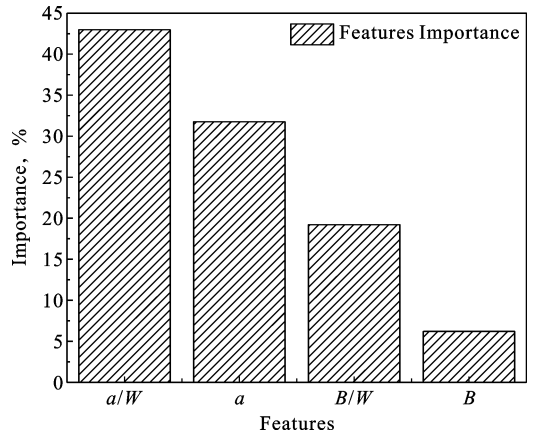


图 8 SENB 试样关键特征量重要性排序

Fig.8 The importance ranking of key characteristic quantities of SENB specimens

## 2.3 基于 RF 算法和数据增强的拘束相关断裂韧性预测

### 2.3.1 SENT 试样

为进一步考察模型的泛化能力, 针对面内、面外拘束同时变化的 SENT 试样, 使用 RF 算法和数据增强后的模型进行预测, 并与数据增强前的结果进行对比, 如图 9 所示. 由图可见, 进行数据增强后, 在之前预测效果略差的  $B=4$  mm 边界点处, 预测结果得到了进一步改善. 数据增强前的拟合系数  $r^2$  为 0.97045, 数据增强后的拟合系数  $r^2$  为 0.97596, 略有增加, 这主要是因为数据增强前对 SENT 试样的预测结果较好.

基于使用 RF 算法并添加数据增强策略后的模型, 对 SENT 试样进行拘束变化关键特征量重要性排序, 如图 10 所示. 由图可见, 对于 SENT 试样, 4 个关键特征量的重要性排序为  $a$ 、 $a/W$ 、 $B$ 、 $B/W$ , 与 SENB 试样不同. 这与其加载方式有关, 将在后续统一进行讨论.

### 2.3.2 CT 试样

同样, 针对面内、面外拘束同时变化的 CT 试样, 使用 RF 算法并添加数据增强策略后的模型进行预测, 数据增强前后预测结果对比如图 11 所示. 由图可见, 进行数据增强后,  $B=4$  和 32 mm 边界点处的预测效果得到明显改善. 数据增强前的拟合系数  $r^2$  为 0.97370, 数据增强后的拟合系数  $r^2$  为 0.98679.

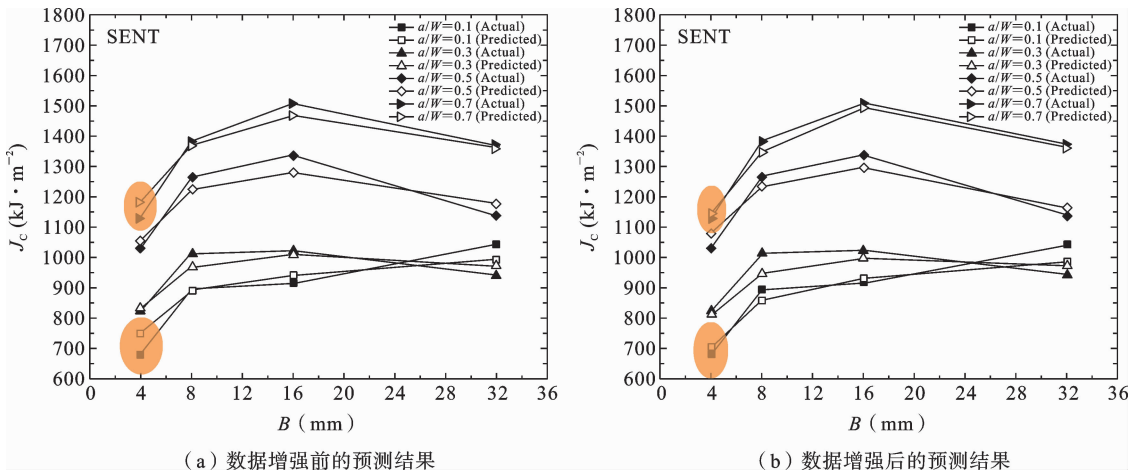


图 9 SENT 试样数据增强前后预测结果对比

Fig. 9 Comparison of prediction results for SENT specimens before and after data enhancement

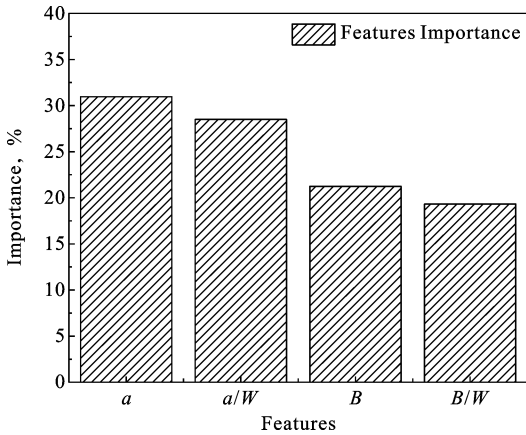


图 10 SENT 试样关键特征量重要性排序

Fig. 10 The importance ranking of key characteristic quantities of SENT specimens

基于使用 RF 算法并添加数据增强策略后的模型, CT 试样拘束变化关键特征量重要性排序如图 12 所示. 由图可见, 对于 CT 试样, 4 个关键特征量的重要性排序为  $B/W$ 、 $B$ 、 $a/W$ 、 $a$ , 且  $B/W$ 、 $B$  与  $a/W$ 、 $a$  相差较大. 这主要是因为所使用的 CT 试样数据中仅包含了两组  $a/W$  的变化 ( $a/W=0.5$  和  $a/W=0.7$ ), 且这两组试样拘束接近, 拘束相关断裂韧性相差不大.

### 2.3.3 CCT 试样

针对面内、面外拘束同时变化的 CCT 试样, 数据增强前后预测结果对比如图 13 所示. 由图可见, 进行数据增强后, 预测效果同样得到明显改善. 数据增强前的拟合系数  $r^2$  为 0.90673, 数据增强后的拟

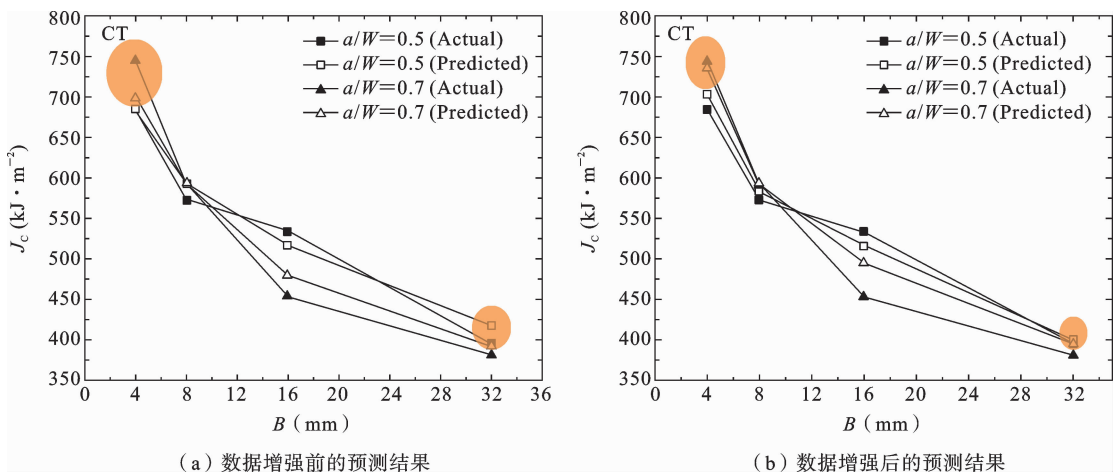


图 11 CT 试样数据增强前后预测结果对比

Fig. 11 Comparison of prediction results for CT specimens before and after data enhancement

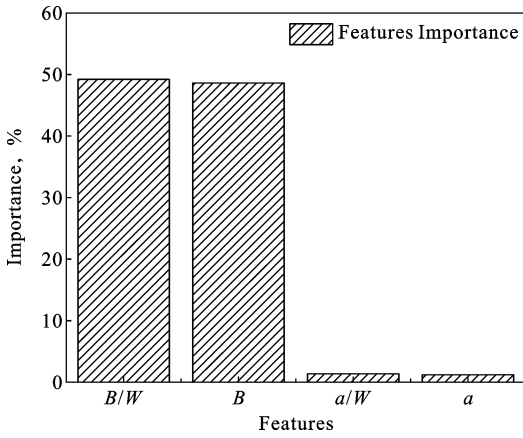
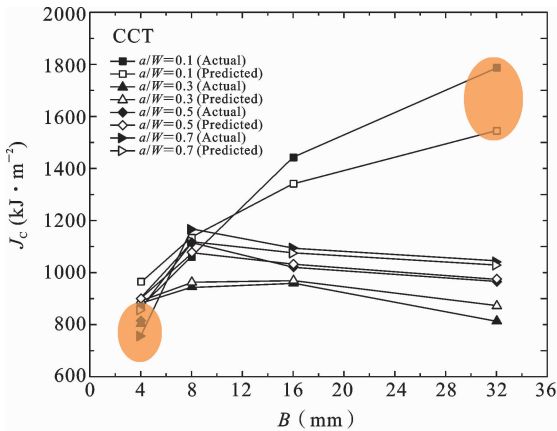


图 12 CT 试样关键特征量重要性排序

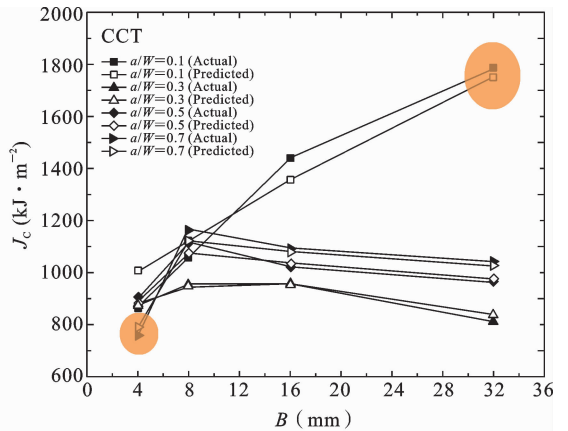
Fig. 12 The importance ranking of key characteristic quantities of CT specimens

合系数  $r^2$  为 0.93388. 综合 SENT、CT 和 CCT 试样的预测对比结果可以进一步说明, 基于 RF 算法和数据增强策略的模型具有较好的泛化能力, 且数据增强策略的加入有效提高了试样在边界点处的预测能力.

同样可以得到 CCT 试样拘束变化关键特征量重要性排序如图 14 所示, CCT 试样 4 个关键特征量的重要性排序为  $B$ 、 $B/W$ 、 $a$ 、 $a/W$ . 综合对比可以发现, 对于 SENB 和 CT 试样, 裂尖主要承受弯曲载荷, 4 个关键特征量中  $a/W$  和  $B/W$  重要性更大, 当设计试样来研究拘束变化的影响时, 通过改变  $a/W$  和  $B/W$  来改变拘束所达到的效果更好; 对于 SENT 和 CCT 试样, 裂尖主要承受剪切载荷, 4 个关键特



(a) 数据增强前的预测结果



(b) 数据增强后的预测结果

图 13 CCT 试样数据增强前后预测结果对比

Fig. 13 Comparison of prediction results for CCT specimens before and after data enhancement

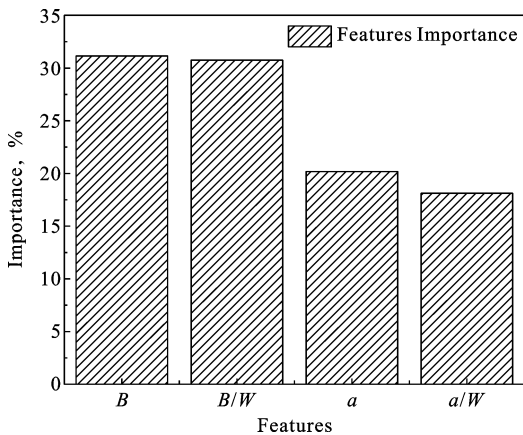


图 14 CCT 试样关键特征量重要性排序

Fig. 14 The importance ranking of key characteristic quantities of CCT specimens

征量中  $a$  和  $B$  重要性更大, 通过改变  $a$  和  $B$  来改变拘束所达到的效果更好.

### 2.4 考虑试样类别的统一预测模型

进一步, 同时考虑四个试样, 在基于 RF 算法和添加数据增强策略的基础上, 加入试样类别这一特征, 构建统一模型, 进一步对所有试样的拘束相关断裂韧性进行预测, 如图 15 所示. 由图可见, 该模型具有很高的预测精度, 其  $r^2$  达到了 0.98221, 后续可使用该模型对不同试样的拘束相关断裂韧性进行预测.

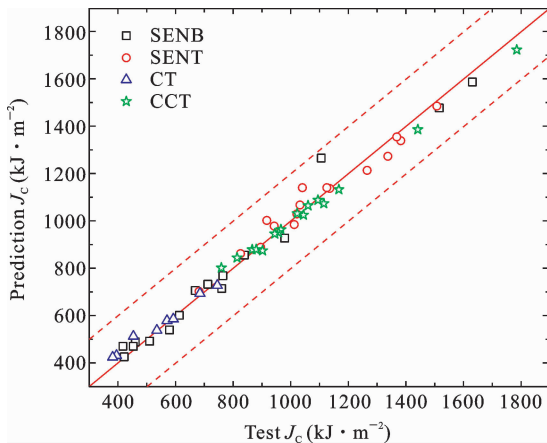


图 15 统一模型中不同试样的预测结果

Fig. 15 Prediction results of different specimens in the unified model

### 3 结论

选用核电用钢 A508 为研究对象,考察了 KNN、KR、LR 和 RF 四种算法对拘束相关断裂韧性的预测能力,在筛选出预测结果最好的算法后,基于该算法,对不同试样的拘束相关断裂韧性进行了预测.进一步,通过添加数据增强策略,实现了更加精准的拘束相关断裂韧性预测,并最终建立了一个统一预测模型.所得主要结论如下:

(1) 四种算法相比,RF 算法的预测效果最好, KR 算法的预测效果最差.四种算法对拘束相关断裂韧性的预测效果为 RF>LR>KNN>KR.

(2) 使用 RF 算法进行预测的同时在数据中进行数据增强,将使预测结果变得更加准确、预测效果更好.数据增强策略的加入有效提高了试样在边界点处的预测能力,且该模型具有较好的泛化能力,可以从 SENB 试样成功移植到 SENT、CT 和 CCT 试样.进一步,在基于 RF 算法和添加数据增强策略的基础上,加入试样类别特征,构建了统一预测模型.

(3) 对于 SENB 和 CT 试样,裂尖主要承受弯曲载荷,当设计试样来研究拘束变化的影响时,通过改变  $a/W$  和  $B/W$  来改变拘束所达到的效果更好;对于 SENT 和 CCT 试样,裂尖主要承受剪切载荷,通过改变  $a$  和  $B$  来改变拘束所达到的效果更好.

### 参考文献

- [1] Brocks W, Schmitt W. Second parameter in J-R curves; constraint or triaxiality in; Proceedings of the 1993 Conference on Constraint Effects in Fracture: Theory and Applications[C], ASTM, Dallas, TX, USA, 1993.
- [2] 张经伟,林嘉铭,谈建平,郭子键,刘康林.试样几何尺寸对 347L 材料断裂性能的影响[J].压力容器,2023,40(01):30-37. (Zhang J W, Lin J M, Tan J P, Guo Z J, Lin K L. Effect of specimen geometries on fracture toughness of 347L materials[J]. Pressure Vessel Technology, 2023, 40(01): 30-37. (in Chinese))
- [3] 张宏,吴锴,冯庆善,隋永莉,刘啸奔,杨悦,杨叠,戴联双,王东营,王冲.高钢级管道环焊缝断裂韧性与裂尖拘束关系[J].石油学报,2023,44(02):385-393. (Zhang H, Wu K, Feng Q S, Sui Y L, Liu X B, Yang Y, Yang D, Dai L S, Wang D Y, Wang C. Relationship between fracture toughness and crack tip constraint of high-strength pipe girth welds[J]. Acta Petrolei Sinica, 2023, 44(02): 385-393. (in Chinese))
- [4] 孙彤.考虑面内/面外拘束的 G20Mn5QT 铸钢材料断裂韧性研究[D].天津大学,2021. (Sun T. Research on Fracture Toughness of G20Mn5QT Cast Steel Considering In-plane and Out-of-plane Constraint [D]. Tianjin University, 2021. (in Chinese))
- [5] Huang X, Xie Q, Li S L, Qi H Y, Yang X G, Shi D Q. Novel framework for predicting constraint effects on fracture toughness using an elastic-plastic phase field model and modified boundary layer formulations [J]. Theoretical and Applied Fracture Mechanics, 2024, 130: 104279.
- [6] Agrawal A, Choudhary A. Perspective: Materials informatics and big data; Realization of the “fourth paradigm” of science in materials science[J]. APL Materials, 2016, 4(5): 053208.
- [7] Schleder G R, Padilha A C M, Acosta C M, Costa M, Fazzio A. From DFT to machine learning: recent approaches to materials science - a review[J]. Journal of Physics: Materials, 2019, 2: 032001.
- [8] 赵申坤,姜潮,龙湘云.一种基于数据驱动和贝叶斯

- 理论的机械系统剩余寿命预测方法[J]. 机械工程学  
报, 2018, 54(12): 115-124. (Zhao S K, Jiang C,  
Long X Y. Remaining useful life estimation of me-  
chanical systems based on the data-driven method and  
bayesian theory[J]. Journal of Mechanical Engineer-  
ing, 2018, 54(12): 115-124. (in Chinese))
- [9] Liu Y L, Niu C, Wang Z, Wang Z, Gan Y, Zhu Y,  
Sun S H, Shen T. Machine learning in materials ge-  
nome initiative: A review[J]. Journal of Materials  
Science & Technology, 2020, 57: 113-122.
- [10] Ma E H, Shin S H, Choi W J, Byun J M, Hwang B  
C. Machine learning approach for predicting the frac-  
ture toughness of Nb-Si based alloys[J]. International  
Journal of Refractory Metals and Hard Materials, 2023,  
117: 106420.
- [11] Albaijan I, Samadi H, Mahmood Z M F, Mah-  
moodzadeh A, Fakhri D, Ibrahim H H, Ouni E H  
M. Evaluation of concrete's fracture toughness under  
an acidic environment condition using advanced ma-  
chine learning algorithms[J]. Engineering Fracture  
Mechanics, 2024, 298: 109948.
- [12] Hu X J, Liao D, Ma D D, Xie S L, Xie N, Hu H B,  
Gong X N. Machine learning models for predicting rock  
fracture toughness at different temperature conditions  
[J]. Case Studies in Construction Materials, 2023,  
19: e02622.
- [13] 肖立志. 机器学习数据驱动与机理模型融合及可解  
释性问题[J]. 石油物探, 2022, 61(2): 205-212.  
(Xiao L Z. The fusion of data-driven machine learn-  
ing with mechanism models and interpretability issues  
[J]. Geophysical Prospecting for Petroleum, 2022,  
61(2): 205-212. (in Chinese))
- [14] Karniadakis G E, Kevrekidis I G, Lu L, Perdikaris P,  
Wang S, Yang L. Physics-informed machine learning  
[J]. Nature Reviews Physics, 2021, 3: 422-440.
- [15] Karpatne A, Atluri G, Faghmous J H, Steinbach M,  
Banerjee A, Ganguly A, Shekhar S, Samatova N,  
Kumar V. Theory-guided data science: A new para-  
digm for scientific discovery from data[J]. IEEE  
Transactions on Knowledge and Data Engineering,  
2017, 29(10): 2318-2331.
- [16] Yang J Y, Kang G Z, Kan Q H. Rate-dependent  
multiaxial life prediction for polyamide-6 considering  
ratchetting: Semi-empirical and physics-informed ma-  
chine learning models[J]. International Journal of Fa-  
tigue, 2022, 163: 107086.
- [17] 杨杰. 面内/面外统一拘束参数及其与材料及焊接接  
头断裂韧性的关联[D]. 华东理工大学, 2014.  
(Yang J. Unified Characterization Parameter of In-  
plane and Out-of Plane Constraint and Its Correlation  
with Fracture Toughness of Materials and Welded  
Joints[D]. East China University of Science and  
Technology, 2014. (in Chinese))
- [18] Yang J, Wang G Z, Xuan F Z, Tu S T. Unified  
characterisation of in-plane and out-of-plane con-  
straint based on crack-tip equivalent plastic strain[J].  
Fatigue & Fracture of Engineering Materials &  
Structures, 2013, 36(6): 504-514.
- [19] Yang J, Wang G Z, Xuan F Z, Tu S T. Unified correla-  
tion of in-plane and out-of-plane constraints with fracture  
toughness[J]. Fatigue & Fracture of Engineering  
Materials & Structures, 2014, 37(2): 132-145.

# A Study on Constraint-Related Fracture Toughness Prediction Based on Random Forest Algorithm and Data Enhancement Strategies

Kangzhong Shan<sup>1</sup>   Xiaoxiao Wang<sup>2</sup>   Fang Liu<sup>3</sup>   Yuanyuan Cui<sup>3</sup>   Jie Yang<sup>1</sup>

*(<sup>1</sup> Shanghai Key Laboratory of Multiphase Flow and Heat Transfer in Power Engineering, School of Energy and Power Engineering, University of Shanghai for Science and Technology, Shanghai, 200093)*

*(<sup>2</sup> Key Laboratory of Pressure Systems and Safety, Ministry of Education, East China University of Science and Technology, Shanghai, 200237)*

*(<sup>3</sup> School of Mechanical Engineering, University of Shanghai for Science and Technology, Shanghai, 200093)*

**Abstract** The study of data-driven predictions for constraint-related fracture toughness is an interdisciplinary scientific problem relevant to mechanics, mechanical engineering, as well as computer science and technology, and is of great significance for accurate structural integrity assessment. This research focused on nuclear power steel A508. The predictive capabilities of four algorithms, namely the K-nearest neighbors (KNN) regression, kernel regression (KR), linear regression (LR), and random forest (RF) regression, for constraint-related fracture toughness predictions were investigated. The RF algorithm outperformed the others, while the KR algorithm had the least effective predictions. The prediction accuracy ranked as follows:  $RF > LR > KNN > KR$ . Furthermore, based on the RF algorithm, data under plane strain conditions were added for data enhancement, enabling the prediction and verification of constraint-related fracture toughness for single-edge notch bending (SENB) specimens. The validated model was successfully transplanted to single-edge notch tension (SENT), compact tension (CT), and central crack tension (CCT) specimens. Results indicated that the RF algorithm with data augmentation improved prediction accuracy and capability, particularly at boundary points. The RF-based model, enhanced with additional data strategies, demonstrated strong generalization across different specimen types. For SENB and CT specimens, bending loads dominate at the crack tip; thus, altering  $a/W$  and  $B/W$  enhances restraint. For SENT and CCT specimens, where shear loads predominate at the crack tip, adjusting  $a$  and  $B$  proves more effective. Finally, a unified, high-accuracy prediction model was developed by incorporating sample category features using the RF algorithm and data enhancement strategies.

**Key words** constraint, fracture toughness, machine learning, random forest algorithm, data enhancement