

· 研究论文 ·

基于应力-应变曲线的机器学习预测聚乙烯和聚丙烯力学性能

张乐康¹ 孙晓宇¹ 李景庆¹ 刘伦洋² 廖涛² 卢影² 李宏飞^{2*} 门永锋^{2*} 蒋世春^{1*}

(¹天津大学材料科学与工程学院 天津 300072)

(²中国科学院长春应用化学研究所 高分子科学与技术全国重点实验室 长春 130022)

摘要 根据聚乙烯(PE)和聚丙烯(PP)的工程应力-应变曲线以及杨氏模量(YM)、断裂伸长率(EB)和拉伸强度(TS)构建了多特征机器学习预测模型,探究了材料属性、制样工艺、测试参数对力学性能的影响.通过最小绝对值收敛和选择算子(Lasso)、决策树回归(DTR)、随机森林回归(RF)等7种算法,结合5折交叉验证和超参数优化,实现了力学性能的定量预测;基于极端梯度提升树回归(XGB)算法完成了应力-应变曲线的高保真度拟合.结果表明:PE力学性能的预测受数据量限制虽然存在精确度局限,但预测趋势表明预测方法有效;YM在PP数据集上预测性能最优(测试集 $R^2 \geq 0.80$,部分模型超过0.90),无明显过拟合;EB和TS预测受数据量所限,仅RF、XGB、KNN模型在PP数据集中表现出有效预测能力;控制变量条件下,PP的应力-应变曲线预测 R^2 大于0.90,可精准再现温度依赖的力学性能.证明了机器学习对聚烯烃力学性能预测的可行性,可为高分子材料力学性能快速预测提供技术支撑.

关键词 应力-应变曲线;机器学习;力学性能

引用: 张乐康,孙晓宇,李景庆,刘伦洋,廖涛,卢影,李宏飞,门永锋,蒋世春.基于应力-应变曲线的机器学习预测聚乙烯和聚丙烯力学性能.高分子学报, doi: 10.11777/j.issn1000-3304.2026.26005.

Citation: Zhang, L. K.; Sun, X. Y.; Li, J. Q.; Liu, L. Y.; Liao, T.; Lu, Y.; Li, H. F.; Men, Y. F.; Jiang, S. C. Machine learning predictive models for the tensile properties of polyethylene and polypropylene. *Acta Polymerica Sinica* (in Chinese), doi: 10.11777/j.issn1000-3304.2026.26005.

聚烯烃是各工业领域中一类至关重要的高分子材料,凭借其轻质、化学惰性、价廉以及性能可调等特点,已被广泛应用于从包装材料到工程塑料等多个领域^[1].聚烯烃以聚乙烯(PE)和聚丙烯(PP)为主,它们的年产量超过2.2亿吨,在全球每年合成超过4亿吨的高分子材料中占有重要地位^[2].中国每年PE、PP产量已超7000万吨^[3],是全球聚烯烃生产、出口与进口第一大国,PE、PP的高性能化与功能化升级已成为行业高质量发展的核心需求.

聚烯烃材料高性能化的关键是结构的精准调控,而实现结构调控的挑战在于了解和认识其多

尺度结构与宏观性能之间的关系^[4-6].例如,PP的分子量分布和支链结构等不仅影响加工性能,而且影响使用性能.高分子材料的多尺度结构对加工条件(温度、压力、冷却速率)和使用环境非常敏感,从而导致材料的设计变量增多、调控难度提升,给结构与性能的精准设计带来了巨大挑战^[7].这种高维、非线性的多因素耦合关系,使得基于经典高分子物理的解析模型难以实现对力学性能的定量、精准预测.

机器学习作为数据驱动的建模范式,无需明确物理建模即可挖掘复杂变量间的非线性关系,是对数据分布进行学习的过程,已在高分子材料

2026-01-05收稿,2026-02-09录用, 网络出版.

基金项目:高分子科学与技术全国重点实验室开放基金.

*通信联系人, E-mail: hfli@ciac.ac.cn; men@ciac.ac.cn; scjiang@tju.edu.cn

doi: 10.11777/j.issn1000-3304.2026.26005; CSTR: 32057.14.GFZXB.2026.7563

性能预测领域展现出巨大潜力^[8-10]. 机器学习涉及概率论、统计学、逼近论、凸分析、算法复杂度理论等多学科, 是数学、机器学习理论与工程实践相结合的交叉学科. 随着计算机硬件的发展, 机器学习的应用范围越来越广. 近年来, 相关研究已实现从分子设计到性能预测的突破: 变分自编码器与图神经网络相结合, 实现了高分子单体结构的自动化生成与指纹识别效率的量级提升^[11-13]; Batra 等^[12]构建了高斯过程回归(Gaussian process regression, GPR)模型, 从变分自编码器得到的隐空间中进行表示学习并预测了高分子的玻璃化转变温度和带隙能; 随机森林(random forest, RF)在预测高分子材料热性能方面表现出优势; 而用于聚氨酯弹性体的分层机器学习模型则优于基于直接成分的方法^[14]; 神经网络可以将 PP 材料的应力-应变曲线与温度和应变速率关联起来^[13]; 结合分子动力学模拟和机器学习构建的人工神经网络模型成功地提高了环氧树脂的综合性能^[15]. 然而, 现有研究主要针对单一性能或特定材料体系, 关于 PE 和 PP 多力学参数协同预测及全应力-应变曲线拟合的研究仍鲜有报道, 数据量不足与特征冗余对预测精度的影响尚未得到系统探究, 缺乏对 YM、断裂伸长率(EB)、拉伸强度(TS)等多参数的协同预测^[13-19]; 应力-应变全曲线的高保真预测研究较少, 且多数预测模型受限于小样本数据集, 难以覆盖材料、工艺、环境的多维度变量组合^[20-23]; 现有研究较少系统探究特征相关性与数据分布对性能预测的影响, 导致模型的实际应用受限^[24-27].

我们根据 2 类(5 种)聚烯烃材料、3 项制样工艺与 4 套测试设备的实验数据, 构建了涵盖材料属性、样品尺寸参数、测试条件的多维度特征体系数据集, 通过 7 种机器学习算法(含 Lasso、RF、XGB 等), 实现了对 PE 和 PP 关键力学参数(YM、EB、TS)的定量预测, 并基于 XGB 算法建立了工程应力-应变曲线的高保真预测模型, 揭示了多因素耦合下聚烯烃力学性能的演变规律, 可为

材料性能快速评估与精准设计提供技术支撑, 推动聚烯烃研发模式向“数据驱动”转型.

1 实验部分

1.1 实验材料与样品制备

5 种测试材料: M1 (等规聚丙烯(*i*PP) T1701) 和 M2 (高密度聚乙烯(HDPE) 5000S) 购自中国石油化工股份有限公司北京燕山分公司; M3 (无规共聚聚丙烯(PPR) 200P) 购自晓星集团; M4 (HDPE) 由中国石化北京化工研究院提供; M5 (*i*PP 427888 与 *i*PP 427853) 购自 Sigma-Aldrich.

3 种制样方法: S1 为模压后裁成哑铃型标准样条, S2 为模压后切成哑铃形非标样品(尺寸为 4 mm × 1 mm × 0.2 mm)^[28], S3 是模压后裁成哑铃形非标样条(尺寸为 10 mm × 5 mm × 0.5 mm)^[29].

1.2 数据采集与测试方法

4 套单轴拉伸的测试设备: E1 (英国 Testometric M250 电子力学万能试验机)、E2 (岛津 Autograph AGS-X 1kN 电子力学万能试验机)、E3 (凯尔测控技术(天津)有限公司, 电子力学万能试验机)^[28] 和 E4 (英国 Linkam TST350, 电子力学万能试验机)^[29].

测试条件: T1 组测试温度包括室温、85 °C、95 °C, 应变速率为 10、20、50、100、200 mm/min; T2 组在室温下以 9 mm/min 应变速率进行单轴拉伸测试, 同一条件下重复实验超过 100 组; T3 组在 30~140 °C 范围内以 10 °C 为梯度, 在 0.02 mm/min 应变速率下进行单轴拉伸测试.

通过上述测试获取工程应力-应变曲线, 并计算杨氏模量(YM)、断裂伸长率(EB)和拉伸强度(TS), 重复实验结果取平均值. 测试对应情况见表 1.

1.3 数据集构建与特征工程

根据表 1 数据构建 PE 数据集和 PP 数据集, 并对两个数据集分别进行数据清洗. 经清洗后 PE 数据集有数据样本 15 个, 包含 5 个特征: 长度(length, mm)、宽度(width, mm)、厚度

Table 1 Composition of the dataset and basic data acquisition information.

Data composition	Data 1	Data 2 ^[28]	Data 3 ^[29]
Materials	M1, M2, M3	M4	M5
Sample preparation method	S1	S2	S3
Testing equipment	E1, E2	E3	E4
Test conditions	T1	T2	T3

(thickness, mm)、应变速率(strain rate, mm/min)和测试温度(test temperature, °C). PP数据集有数据样本75个,除了包含PE的5个特征外还包含:材料种类(*i*PP、PPR)、数均分子量(M_n , 10^6 g/mol)和重均分子量(M_w , 10^6 g/mol).对材料种类的分类特征集通过独热编码转化成二元特征,由于形

变速率跨越多个数量级,采用对数(lg)处理.预测目标为YM (MPa)、EB (%)和TS (MPa)的lg结果.将2个数据集分别按照训练:测试集=4:1进行随机划分,见图1(a).在训练集中,对所有特征进行线性归一化处理.使用训练的归一化模型对测试集特征进行归一化,以防止信息泄露.

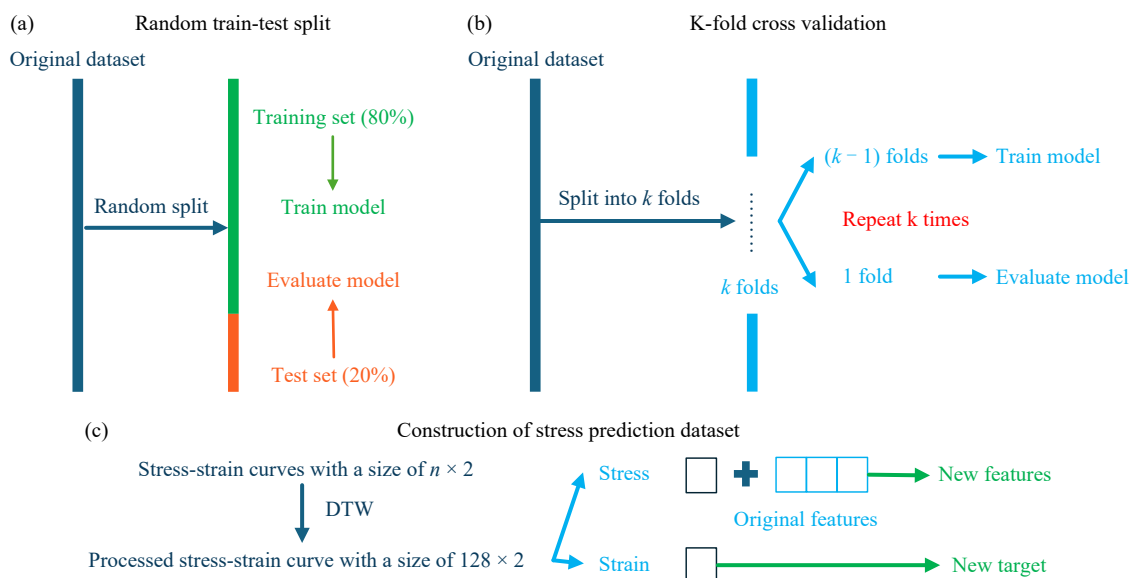


Fig. 1 Schematic diagrams of dataset processing and splitting strategies for mechanical property prediction, comprising (a) the random train-test split workflow (partitioning the original dataset into 80% training set and 20% test set for model training and performance evaluation), (b) the k -fold cross-validation strategy (repeatedly utilizing $(k - 1)$ folds for model training and 1 fold for evaluation to improve model generalization ability), and (c) the stress prediction dataset construction process (standardizing $n \times 2$ raw stress-strain curves to 128×2 processed curves *via* dynamic time warping (DTW), then fusing original features with strain features to generate new input features and designating stress as the new prediction target).

1.4 机器学习与模型训练

选取线性模型、树模型、集成学习、核方法等不同类型的7种机器学习算法并对比预测性能.最小绝对值收敛和选择算子(least absolute shrinkage and selection operator, Lasso)核心是在线性回归损失函数中引入L1正则化项,不仅能压缩系数幅值降低模型复杂度,更可通过将冗余/弱关联特征系数置零实现稀疏化特征选择,解决传统线性回归在高维特征场景下的多重共线性问题^[24];决策树回归(decision tree regression, DTR)作为非参数模型,基于均方误差(MSE)或平均绝对误差(MAE)最小化准则,通过二叉树递归划分特征空间,每个叶节点对应独立特征子区域,以区域内样本目标值均值作为预测输出,具有可解释性的优势^[29];随机森林回归(random forest regression, RF)是Bagging集成框架的核心代表,以无剪枝完全生长决策树为基学习器,通

过Bootstrap自助采样实现样本随机、每棵树随机选取部分特征分裂实现特征随机(双重随机性),依托袋外样本(OOB)完成内部泛化能力评估,无需额外验证集^[30];极端梯度提升树回归(eXtreme gradient boosting regression, XGB)是Boosting框架的优化实现,以决策树为基学习器,核心差异于RF的并行集成,采用梯度下降迭代策略—每棵新树均针对前序模型的预测残差拟合,同时融入L1/L2正则化、树结构剪枝及缺失值自适应处理机制^[31];核岭回归(Kernel ridge regression, KRR)是将Ridge与核技巧(Kernel trick)结合的产物,通过核函数(如RBF核)隐式将低维非线性特征映射至高维线性空间,在高维空间执行岭回归以规避过拟合,无需显式计算映射过程,有效解决“维度灾难”,适合小样本、特征非线性弱关联场景^[32,33];高斯过程回归(Gaussian process regression, GPR)基于贝叶斯概

率框架, 将预测函数视为服从高斯过程的随机变量, 通过定义核函数刻画样本间相关性构建先验分布, 结合观测数据更新为后验分布, 核心优势是精准的不确定性量化能力, 可同步输出预测值与置信区间^[34]; K近邻回归(*k*-nearest neighbor, KNN)属于惰性学习算法, 无需预先训练模型参数, 预测时通过距离度量(如欧氏距离、曼哈顿距离)筛选 *k* 个近邻样本, 基于近邻样本目标值加权平均输出结果, 对局部数据分布高度敏感, 与其它进行全局建模的差异明显。

模型在训练集中采用5折交叉验证的网格法进行超参数调优, 随后基于最优参数完成训练. 交叉验证的示意图见图1(b). 其中网格法调参以最大化决定系数 R^2 为目标, R^2 的计算如下:

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

其中, n 为样本数, y_i 和 \hat{y}_i 分别为第 i 个数据的实测值与预测值, \bar{y} 为实测值的平均值。

为了实现聚烯烃力学行为的全周期精准描述与预测, 在核心力学参数预测基础上, 构建了应力-应变曲线预测模型. 基于PP数据集集中的60个控制变量法获取的样本, 在原有特征体系中新增加全区间应变特征, 以应力为预测目标; 通过动态时间规整(DTW)对原始曲线进行对齐处理并下采样至128个特征点, 确保曲线长度统一; 为避免信息泄露, 按4:1比例以完整曲线为单位划分训练集与测试集; 综合预测准确性与训练效率, 选取XGB算法建模, 采用以最大化决定系数 R^2 为目标的贝叶斯优化策略(100次迭代, 5折交叉验证)完成超参数优化与模型训练. 应力-应变曲线重构数据集的构建示意图见图1(c).

2 结果与讨论

2.1 数据集特征

PE和PP的应力-应变曲线形状复杂, 难以通过单一简单函数(如线性函数、幂函数)对其完整曲线形态进行精准表征. 图2为PE和PP在室温下测得的应力-应变曲线, 二者均呈现“弹性—屈服—成颈—应变硬化”的力学行为, 表现出明显的屈服与韧性断裂特征, 且在弹性应变(YM关联)、屈服强度、断裂应变等特征基本一致, 这一点为使用具有特征的力学性能值作为预测目

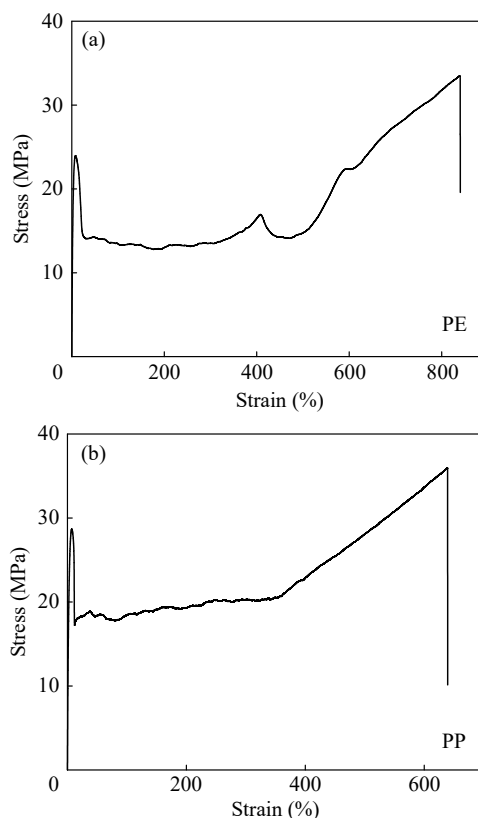


Fig. 2 Engineering stress-strain curves of polyethylene (a) and polypropylene (b) during uniaxial stretching at room temperature.

标以达到对PE和PP应力-应变曲线进行描述提供了可能。

为了便于理解训练模型的目标, 我们对PE和PP实验数据进行了归一化处理, 归一化结果见图3. 图3中的结果表明: PE在应变硬化阶段曲线形态相似度较高, 但成颈阶段的应力变化存在明显差异, 部分样品屈服后应力持续下降(可能与成型过程中的结晶行为有关); *i*PP与PPR的曲线形态基本一致, 部分*i*PP在成颈与应变硬化阶段的应力增加较大。

相似的应力-应变曲线可以用YM、EB和TS表示, 图4为YM、EB、TS经非线性降维方法*t*-SNE降维处理后的分布结果. 图4中YM、EB和TS的数值分布表明: PE数据集内部无明显聚类现象, 证明不同样品的力学性能特征相似度较高; PP数据集中*i*PP与PPR的分布区域部分重叠, 仅在边缘区域存在分离, 证明2种材料力学性能的同源性。

综上所述, 通过YM、EB、TS这3个特征可以表达曲线的差异化信息. 但图3中的结果表

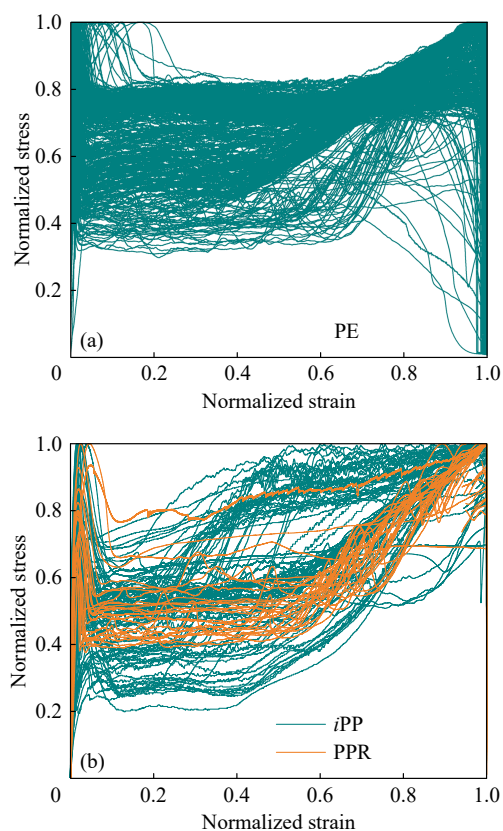


Fig. 3 Normalized stress-strain curves of (a) PE and (b) PP. The curve normalization method is defined as follows: within the same curve, the normalized stress and strain are calculated by $\text{Stress}_{\text{norm}} = \text{Stress}/\text{Stress}_{\text{max}}$, $\text{Strain}_{\text{norm}} = \text{Strain}/\text{Strain}_{\text{max}}$. This is used to eliminate the differences in absolute stress and strain values, focusing on the characteristics of curve shape differentiation of different samples in the necking stage and strain hardening stage.

明这3个特征仍缺乏对曲线之间差异的描述，只有应力-应变全曲线预测能提供曲线原始信息。

为了进一步研究数据特征和训练模型对预测结果的影响，我们考察了形变速率的影响，结果

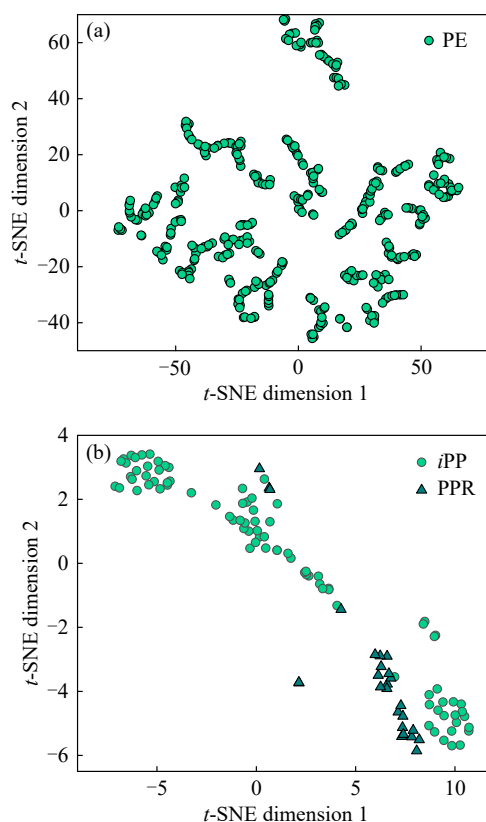
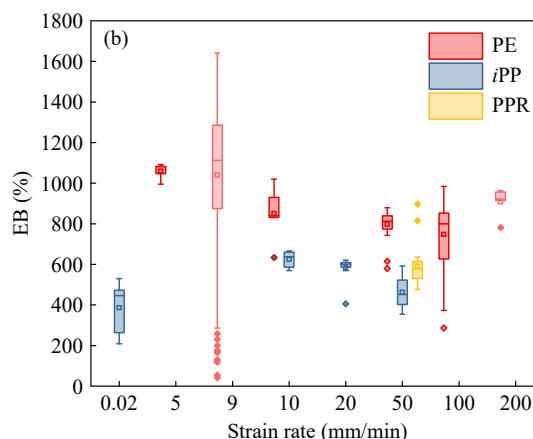
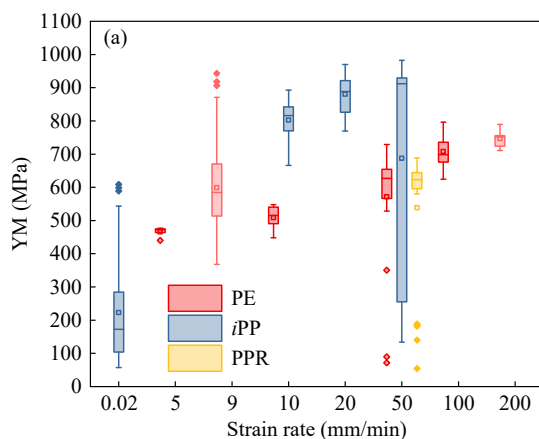


Fig. 4 Distribution maps of YM, EB, and TS of (a) PE data and (b) PP data after t-SNE dimensionality reduction.

见图5。图5中的结果表明：由于多特征影响，形变速率对PE和PP的测试力学性能影响不显著，但仍能呈现出可能的变化趋势。其中，对于PE的TS随形变速率的增加呈现出上升趋势；在YM中，也能观察到类似的变化趋势。

我们通过计算特征与预测目标间的线性关联考察了特征与预测目标之间的关系，并采用皮尔逊相关矩阵表示预测效果，结果见图6。图中的皮尔逊相关系数(r)为：



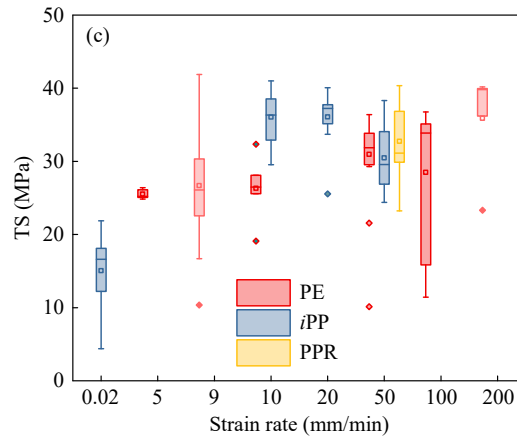


Fig. 5 Distribution of mechanical properties of PE and PP with different strain rates. The strain rates range from 0.02 mm/min to 200 mm/min, covering the actual test rates of samples from different sources. The effects of strain rate on the three mechanical properties are shown in (a) YM, (b) EB, and (c) TS, respectively.

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (2)$$

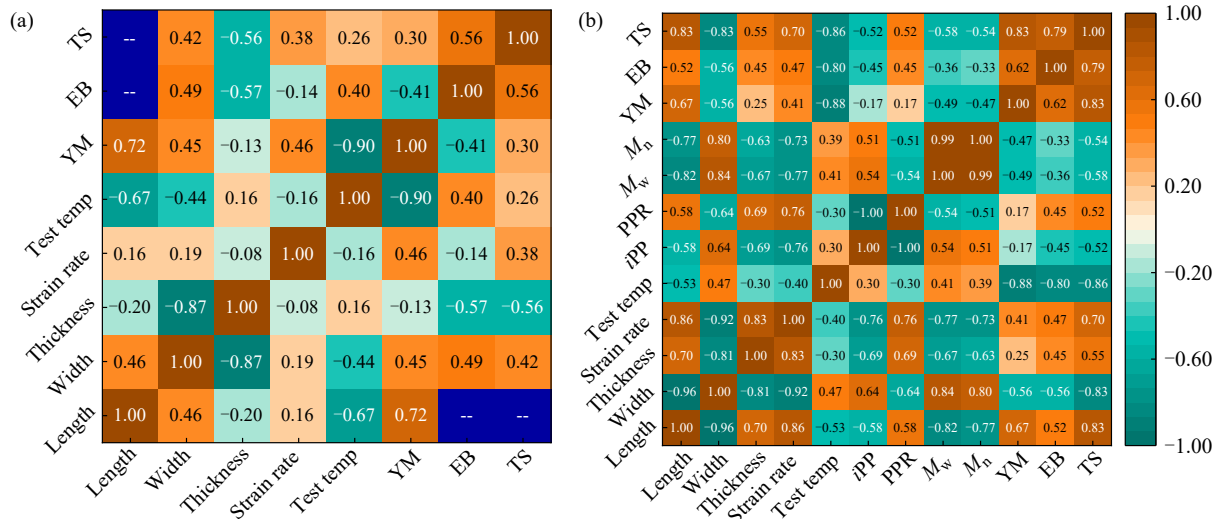
其中, n 为样本数, x_i 和 y_i 分别为 2 个变量第 i 个值, \bar{x} 和 \bar{y} 分别为 2 个变量的平均值.

图 6 中的皮尔逊相关矩阵揭示了特征与力学性能的线性关联规律. PE 数据集: 测试温度与 YM 呈强负相关 ($r = -0.90$), 表明温度升高显著降低材料刚性; 长、宽、厚与力学性能相关性较弱 ($|r| < 0.3$), 说明制样尺寸对性能影响有限, 可作为辅助校正特征. PP 数据集: M_n 与 M_w 呈强正相关 ($r = 0.99$), 存在严重共线性; 测试温度与 YM ($r = -0.82$)、TS ($r = -0.79$) 均呈强负相关, 与 EB 呈弱负相关 ($r = -0.36$), 表明温度对刚性和强度的影响更为显著; 材料种类 (iPP/PPR) 与力学性能相关性较弱 ($|r| < 0.2$), 验证了二者结构与性能的相

似性. 图 7 为训练集与测试集的力学性能分布. 由于 PE 数据集测试集仅含 3~4 个样本, 且分布与训练集偏差明显, 代表性不足; PP 数据集样本量相对充足, 训练集与测试集分布一致性显著提升, 但仍存在局部偏差, 反映出小样本数据集划分的局限性.

2.2 模型训练结果

7 种模型在 PE 和 PP 数据集上的测试集 R^2 对比见图 8. 图中结果表明: YM 的预测最优, PP 数据集所有模型 R^2 均 ≥ 0.80 , XGB、RF 模型超过 0.90, 且训练集与测试集 R^2 差异 ≤ 0.10 , 无明显过拟合; PE 数据集因样本量不足 (测试集仅 3 个样本), 模型稳定性差, 但 Lasso 模型凭借测试温度与 YM 的强线性关联, 仍实现了相对可靠的预测 ($R^2 = 0.72$); EB 与 TS 预测性能较差, PE 数据集所有模型 R^2 均 < 0 , 表明模型无法捕捉有效关联;



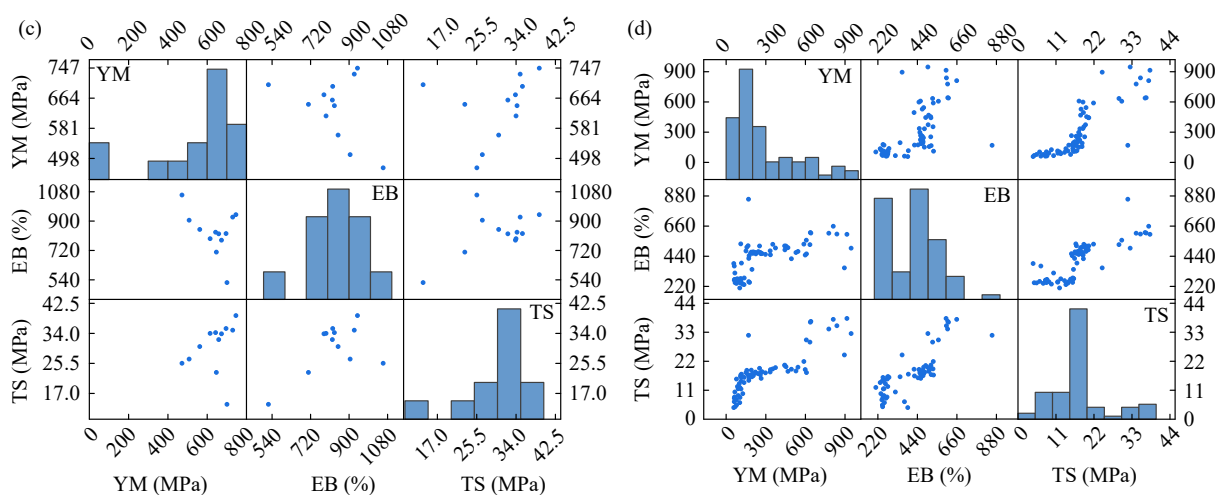
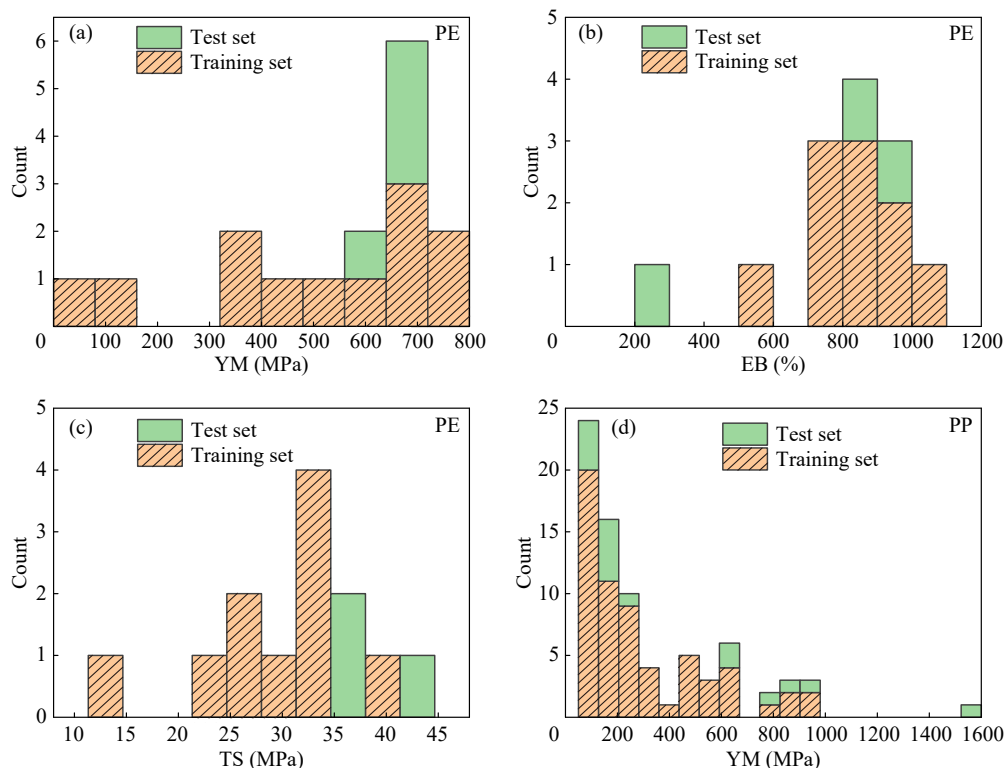


Fig. 6 Pearson correlation matrix of feature-mechanical properties for (a) PE dataset and (b) PP dataset. Fracture-related values are missing in some data samples. Specifically, all samples with the same initial inter-fixture length in tensile tests have complete fracture-related data, whereas data gaps occur in samples with different initial lengths, which results in blank corresponding r -values. Pearson correlation visualization *via* scatter plots for toughness-strength trade-off: (c) PE dataset, (d) PP dataset.

PP数据集中, RF、XGB、KNN模型表现突出($R^2=0.65\sim 0.78$), 其余模型 $R^2<0.5$, 可能与EB、TS受多因素耦合影响(如结晶形态、链取向)有关, 非线性关联更复杂; 集成学习算法(RF、XGB)整体优于线性模型(Lasso)、单一树模型(DTR)及核方法(KRR、GPR), 表明集成算法对多特征耦合关系的挖掘能力更强, 更适用于聚烯烃力学性能预测.

图9为RF模型的预测散点图, 图中结果进一步证明: PE数据集EB、TS预测点偏离对角线, 无法实现有效预测; PP数据集YM预测点集中于对角线附近, EB、TS预测虽存在偏差, 但趋势基本一致, 与 R^2 分析结果吻合. 可见, 模型的预测性能受到了数据量的限制, 特别是PE数据集上难以将训练集与测试集上 R^2 的差异解释为过拟合现象, 此时预测性能的描述本身



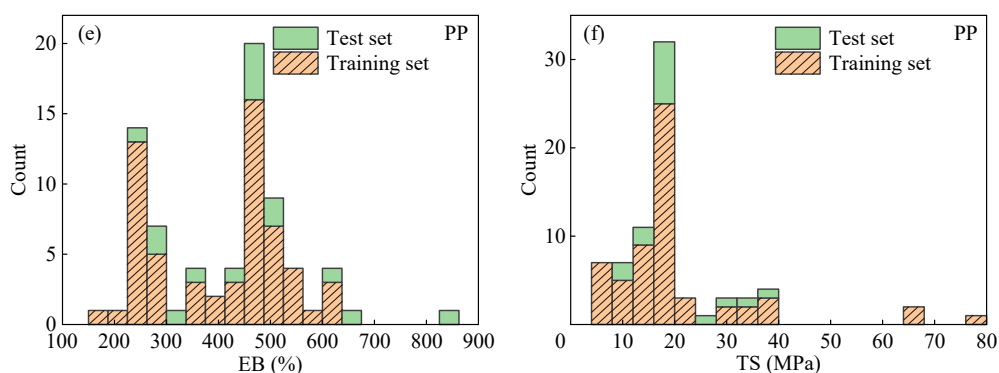


Fig. 7 Comparison of the distribution of mechanical properties of PE and PP datasets in the training set and test set: (a–c) show the distributions of YM, EB, and TS of the PE dataset in sequence; (d–f) show the distributions of the corresponding mechanical parameters of the PP dataset in sequence.

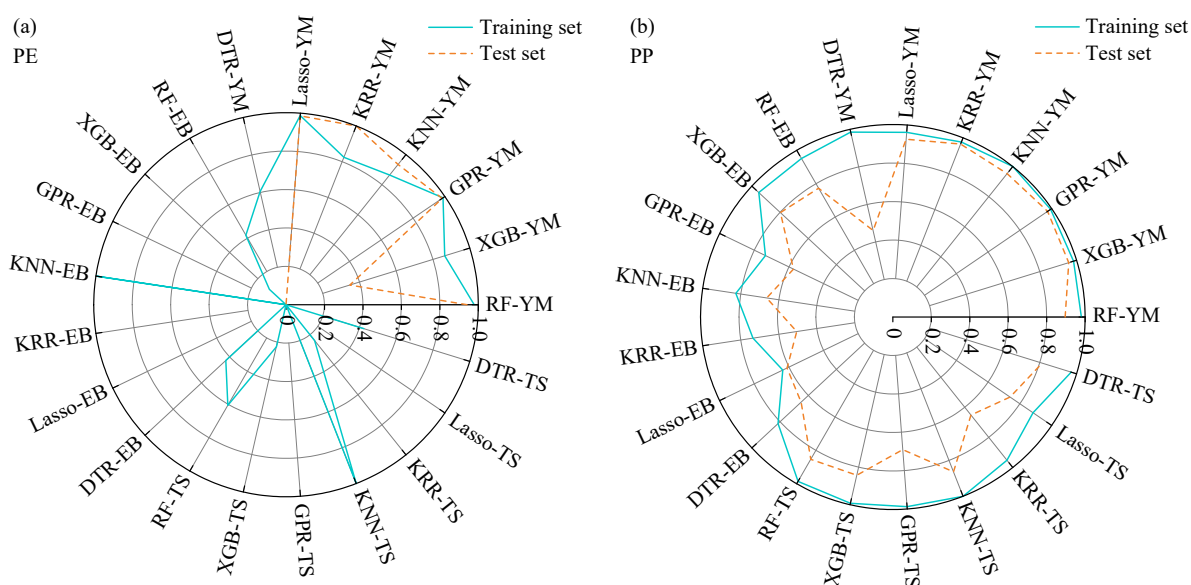
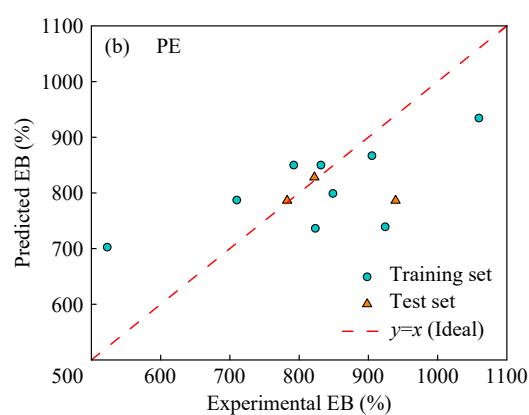
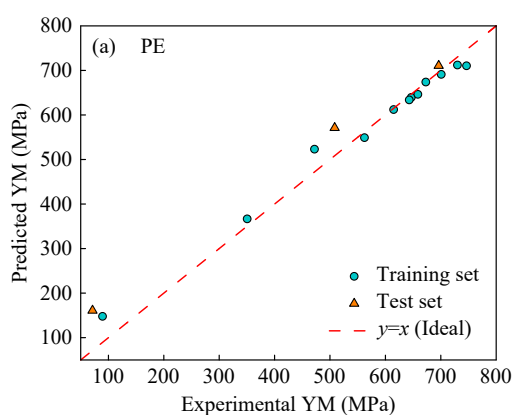


Fig. 8 Seven machine learning models are compared for predicting mechanical properties on (a) PE and (b) PP datasets. The models include Lasso, Decision Tree Regression (DTR), Random Forest (RF), Extreme Gradient Boosting Tree (XGB), Kernel Ridge Regression (KRR), Gaussian Process Regression (GPR), and K-Nearest Neighbor Regression (KNN).

就缺失代表性，同时稀缺的数据限制了模型对数据分布的学习。在 PP 数据集中数据量的影响有了明显的改善，但仍然存在数据不足的情况，

少数点的预测偏差对模型性能的评估存在明显的影响。补充数据集的样本数量是后续研究的重点。



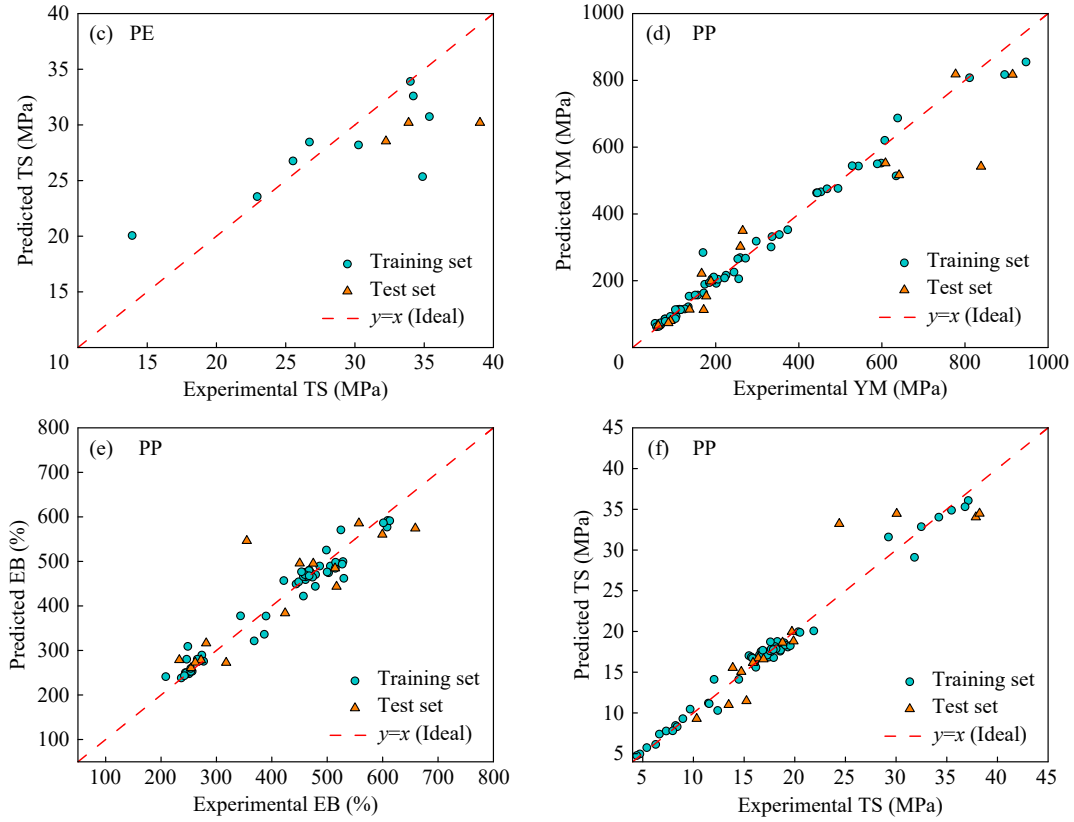


Fig. 9 Scatter plots of mechanical property predictions by the RF model on (a–c) PE and (d–f) PP datasets.

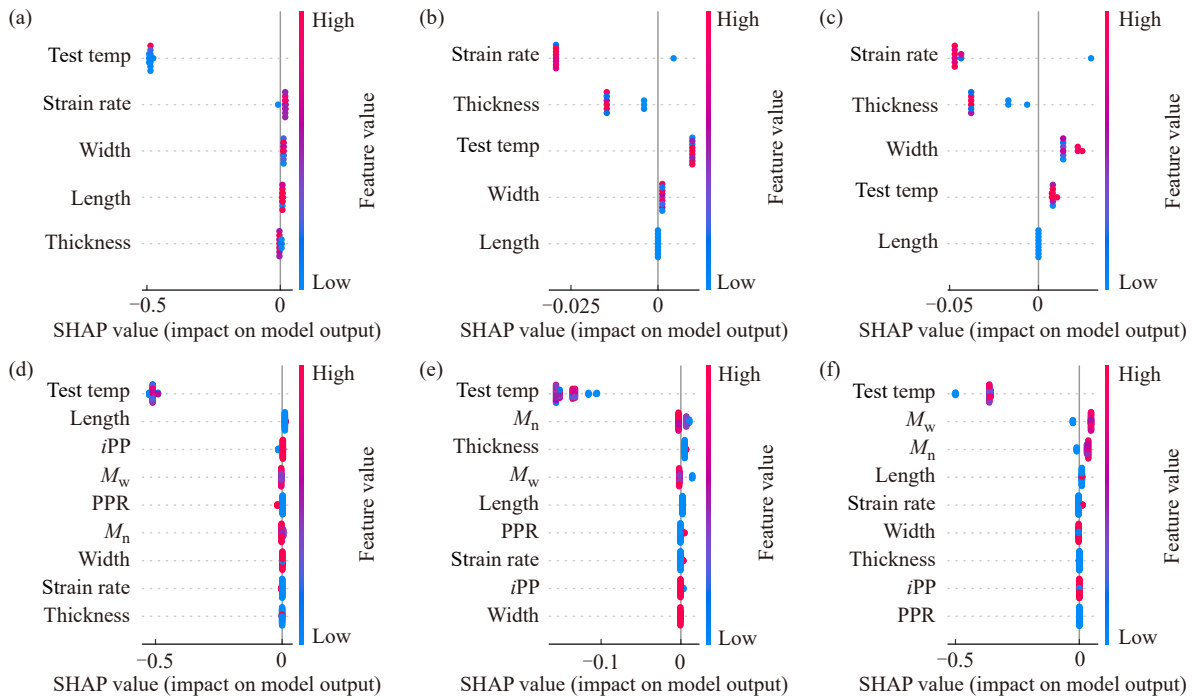


Fig. 10 SHAP feature importance distribution of polyolefin (PE/PP) mechanical property prediction models. (a) PE dataset prediction target YM; (b) PE dataset prediction target EB; (c) PE dataset prediction target TS; (d) PP dataset prediction target YM; (e) PP dataset prediction target EB; (f) PP dataset prediction target TS.

通过计算 SHAP 值可了解各输入特征对 PE 和 PP 数据集力学性能预测模型输出的影响趋势，

训练集上 RF 模型的 SHAP 特征重要性蜂群图见图 10。图中每一点对应单个样本的特征取值，红

色代表特征高值、蓝色代表特征低值；横坐标 SHAP 值表示特征对模型输出的影响程度，正值会让预测结果变大，负值会让预测结果变小。图中结果表明：在 PE 数据集内，模型仅对 YM 的预测效果较优；而 EB、TS 的预测性能整体较差，从而使得对应任务下 SHAP 值的代表性受限。测试温度对力学性能的影响显著，对于 PE 来说，温度升高导致 YM 显著降低，而 EB 则上升，与温度升高降低材料刚性(降低 YM)一致。而制样尺寸特征的 SHAP 值分布于 0 附近，表明其对 PE

数据集中力学性能预测结果的影响甚微，这与特征相关性分析的结果吻合；形变速率则因与数据来源存在关联，且当前数据量有限，其真实作用尚未明确。对于 PP 数据集(图 10(d)~10(f))，除测试温度(含 60 个梯度样本)、形变速率外， M_n 和 M_w 也是力学性能预测的关键特征；不过从图中 M_n 和 M_w 的分布来看，受限于测试样本的取值范围，二者的数值呈现集中化特征，可结合预测结果进一步分析(见图 11)。综上，SHAP 分析明确了 PE 与 PP 数据集力学性能预测的核心驱动特征，

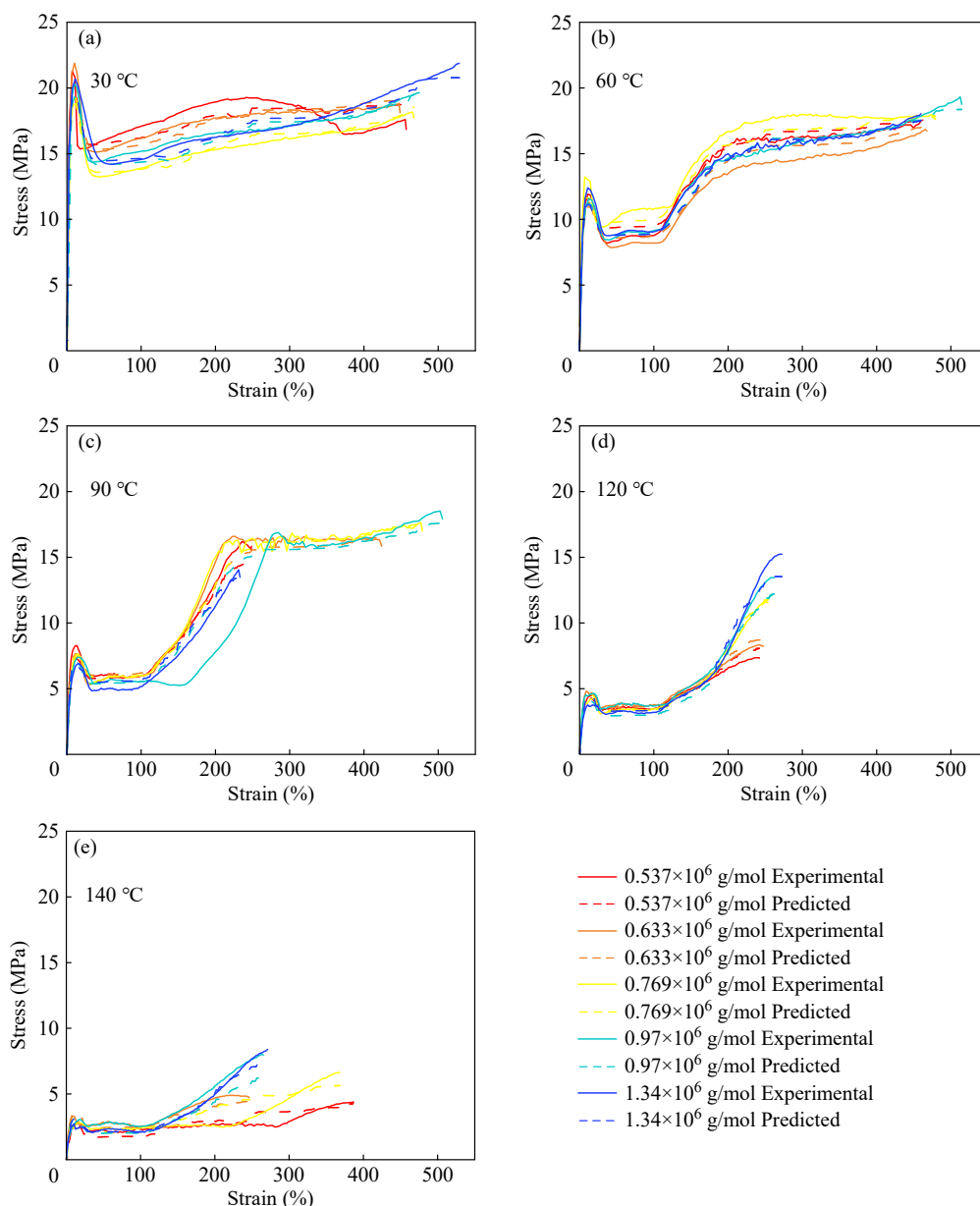


Fig. 11 The prediction effect of the XGB model on the stress-strain curves of PP control variable dataset contains 60 samples (including systematic temperature and molecular weight changes). The solid lines represent the experimental stress-strain curves, and the dashed lines represent the stress-strain curves predicted by the XGB model. The curve comparisons at (a) 30 °C, (b) 60 °C, (c) 90 °C, (d) 120 °C, and (e) 140 °C are displayed.

为将来数据采集的特征重点、数据集的特征筛选以及模型的简化优化提供了依据。

利用XGB模型预测的PP应力-应变曲线见图11。图中结果表明：预测曲线与实验曲线吻合， R^2 普遍大于0.90，可精准再现温度升高导致材料软化、屈服强度降低及应变硬化减弱等关键行为；即使在140 °C下，模型仍能捕捉应力-应变曲线的核心特征，表明该模型具有良好的泛化能力。此外，模型对成颈阶段的应力波动也可以有效拟合，弥补了单一力学参数预测的信息缺失。全样本曲线预测与大规模应力-应变曲线预测在文献[35]中同样表现优异，体现出充足的应用潜力。预处理阶段采用DTW算法对曲线进行降采样，通过压缩曲线预测数据集的规模，在保留曲线核心形态特征的同时，显著提升了模型训练效率。数据量扩大时，DTW算法的预处理耗时明显增加，曲线预测数据集的规模膨胀也加重了模型训练的负担；采用生成式模型的高效建模方式，或将成为学习应力-应变曲线的优选路线。

3 结论

通过构建多材料、多工艺、多参数的聚烯烃数据集，结合7种机器学习算法，实现了对力学性

能及工程应力-应变曲线的精准预测；证明机器学习在聚烯烃力学性能预测中的可行性，YM预测效果最优，PP数据集上XGB、RF模型测试集 R^2 超0.90，无明显过拟合；EB、TS预测受数据量与非线性关联复杂性限制，仅PP数据集中集成学习模型表现有效；特征相关性分析表明，测试温度是影响YM的关键因素($r=-0.82\sim-0.90$)，制样尺寸对性能影响较小，可作为辅助校正特征； M_n 与 M_w 的强共线性提示建模时需进行特征筛选；XGB模型可实现应力-应变曲线的高保真预测($R^2\geq 0.90$)，能精准再现温度依赖的力学行为，为聚烯烃材料全周期力学性能评估提供了新方法。

研究局限性：PE数据集样本量不足导致模型稳定性较差，预测性能缺乏验证；特征集中未包含分子链支化、结晶等微观结构参数。未来研究可通过进一步补充数据、数据增强(如生成式模型合成虚拟样本)、迁移学习(利用同类高分子数据迁移训练)弥补数据量不足的缺陷；同时引入微观结构特征，进一步提升模型预测精度与普适性；此外，可扩展至长期服役性能、耐环境老化性能等更广泛的性能指标，构建全维度聚烯烃材料性能预测平台，为高性能聚烯烃的快速开发提供技术支撑。

REFERENCES

- 1 Wang, X. Y.; Gao, Y. S.; Tang, Y. Sustainable developments in polyolefin chemistry: progress, challenges, and outlook. *Prog. Polym. Sci.*, **2023**, 143, 101713.
- 2 Plastics—the fast Facts **2024**. 2024. <https://plasticseurope.org/knowledge-hub/plastics-the-fast-facts-2024/>.
- 3 吴长江. 我国聚烯烃产业技术的现状与发展建议. *科学通报*, **2022**, 67(17), 1853–1862.
- 4 Zhu, T. G.; Li, X. L.; Zhao, X. Y.; Zhang, X.; Lu, Y. L.; Zhang, L. Q. Stress-strain behavior and corresponding crystalline structures of four types of polyethylene under a wide range of strain rates. *Polym. Test.*, **2022**, 106, 107460.
- 5 Zhu, X. Y.; Zhou, Y. F.; Yan, D. Y. Influence of branching architecture on polymer properties. *J. Polym. Sci. Part B Polym. Phys.*, **2011**, 49(18), 1277–1286.
- 6 Seyfi, S. K.; Ali, A. F. Properties of linear low density polyethylene. *IJITEE*, **2020**, 9(9), 348–352.
- 7 Suljovrujic, E.; Milicevic, D.; Djordjevic, K.; Rogic Miladinovic, Z.; Stamboliev, G.; Galovic, S. Structure-property relationship in isotactic polypropylene under contrasting processing conditions. *Polymers*, **2025**, 17(14), 1889.
- 8 刘伦洋, 丁芳, 李云琦. 高分子材料大数据研究: 共性基础、进展及挑战. *高分子学报*, **2022**, 53(6), 564–579.
- 9 Xie, C. H.; Qiu, H. K.; Liu, L.; You, Y.; Li, H. F.; Li, Y. Q.; Sun, Z. Y.; Lin, J. P.; An, L. J. Machine learning approaches in polymer science: progress and fundamental for a new paradigm. *SmartMat*, **2025**, 6, e1320.
- 10 Butler, K. T.; Davies, D. W.; Cartwright, H.; Isayev, O.; Walsh, A. Machine learning for molecular and materials science. *Nature*, **2018**, 559(7715), 547–555.
- 11 Kingma, D. P., Welling, M. Auto-Encoding Variational Bayes. *arXiv*, **2022**, arXiv: 1312.6114. <https://arxiv.org/abs/1312.6114>.
- 12 Batra, R.; Dai, H. J.; Huan, T. D.; Chen, L. H.; Kim, C.; Gutekunst, W. R.; Song, L.; Ramprasad, R. Polymers for extreme conditions designed using syntax-directed variational autoencoders. *Chem. Mater.*, **2020**, 32(24), 10489–10500.

- 13 Jordan, B.; Gorji, M. B.; Mohr, D. Neural network model describing the temperature- and rate-dependent stress-strain response of polypropylene. *Int. J. Plast.*, **2020**, 135, 102811.
- 14 Menon, A.; Thompson-Colón, J. A.; Washburn, N. R. Hierarchical machine learning model for mechanical property predictions of polyurethane elastomers from small datasets. *Front. Mater.*, **2019**, 6, 87.
- 15 Jin, K.; Luo, H.; Wang, Z. Y.; Wang, H.; Tao, J. Composition optimization of a high-performance epoxy resin based on molecular dynamics and machine learning. *Mater. Des.*, **2020**, 194, 108932.
- 16 Kopal, I.; Labaj, I.; Harničárová, M.; Valíček, J.; Hrubý, D. Prediction of the tensile response of carbon black filled rubber blends by artificial neural network. *Polymers*, **2018**, 10(6), 644.
- 17 Duong, H. T.; Phan, H. C.; Tran, T. M.; Dhar, A. S. Assessment of critical buckling load of functionally graded plates using artificial neural network modeling. *Neural Comput. Appl.*, **2021**, 33(23), 16425–16437.
- 18 Zhou, Q.; Zhao, A. G.; Wang, H.; Liu, C. Machine learning guided design of mechanically efficient metamaterials with auxeticity. *Mater. Today Commun.*, **2024**, 39, 108944.
- 19 Kopal, I.; Harničárová, M.; Valíček, J.; Kušnerová, M. Modeling the temperature dependence of dynamic mechanical properties and visco-elastic behavior of thermoplastic polyurethane using artificial neural network. *Polymers*, **2017**, 9(10), 519.
- 20 Zhang, F.; Miyao, T.; Izumiya, Y.; Chen, C. H.; Funatsu, K. Designing heat-resistant and moldable polyester resin by the integration of machine learning models with expert knowledge. *ACS Appl. Polym. Mater.*, **2024**, 6(8), 4579–4586.
- 21 Ding, F.; Liu, L. Y.; Liu, T. L.; Li, Y. Q.; Li, J. P.; Sun, Z. Y. Predicting the mechanical properties of polyurethane elastomers using machine learning. *Chinese J. Polym. Sci.*, **2023**, 41(3), 422–431.
- 22 Ding, F.; Liu, T. L.; Zhang, H.; Liu, L. Y.; Li, Y. Q. Stress-strain curves for polyurethane elastomers: a statistical assessment of constitutive models. *J. Appl. Polym. Sci.*, **2021**, 138(39), 51269.
- 23 Zhang, H.; Ding, F.; Liu, T. L.; Liu, L. Y.; Li, Y. Q. Additivity of the mechanical properties for acrylonitrile-butadiene-styrene resins. *J. Appl. Polym. Sci.*, **2022**, 139(15), 51923.
- 24 Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **1996**, 58(1), 267–288.
- 25 Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly accurate protein structure prediction with AlphaFold. *Nature*, **2021**, 596(7873), 583–589.
- 26 Kazemi-Khasragh, E.; Fernández Blázquez, J. P.; Garoz Gómez, D.; González, C.; Haranczyk, M. Facilitating polymer property prediction with machine learning and group interaction modelling methods. *Int. J. Solids Struct.*, **2024**, 286–287, 112547.
- 27 Blum, L. C.; Reymond, J. L. 970 million druglike small molecules for virtual screening in the chemical universe database GDB-13. *J. Am. Chem. Soc.*, **2009**, 131(25), 8732–8733.
- 28 Liao, T.; Yang, X.; Zhao, X. T.; Tang, Y. J.; Jiang, Z. Y.; Men, Y. F. Gaussian and non-Gaussian distributions of fracture properties in tensile stretching of high-density polyethylene. *Macromolecules*, **2021**, 54(18), 8860–8874.
- 29 Lu, Y.; Lyu, D.; Han, C. H.; Men, Y. F. Molecular weight and distribution dependent critical stress of fragmentation of highly oriented amorphous network in stretched isotactic polypropylene. *Macromolecules*, **2023**, 56(18), 7467–7478.
- 30 Breiman, L. Random forests. *Mach. Learn.*, **2001**, 45(1), 5–32.
- 31 Chen, T. Q.; Guestrin, C. XGBoost: a scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. San Francisco California USA, **2016**, 785–794.
- 32 Kim, C.; Chandrasekaran, A.; Jha, A.; Ramprasad, R. Active-learning and materials design: the example of high glass transition temperature polymers. *MRS Commun.*, **2019**, 9(3), 860–866.
- 33 Yang, Y. P.; Patil, N.; Askar, S.; Kumar, A. Machine learning-guided study of residual stress, distortion, and peak temperature in stainless steel laser welding. *Appl. Phys. A*, **2024**, 131(1), 44.
- 34 Rasmussen, C. E.; Nickisch, H. Gaussian processes for machine learning (GPML) toolbox. *Journal of Machine Learning Research*, **2010**(11), 3011–3015.
- 35 Zhang, L. K.; Sun, X. Y.; Li, J. Q.; Liu, L. Y.; Liao, T.; Lu, Y.; Yuan, W. B.; Chen, H. Y.; Li, H. F.; Men, Y. F.; Jiang, S. C. Machine learning prediction of mechanical properties of polyolefins from stress-strain curves. *Polymer*, **2026**, 345, 129574.

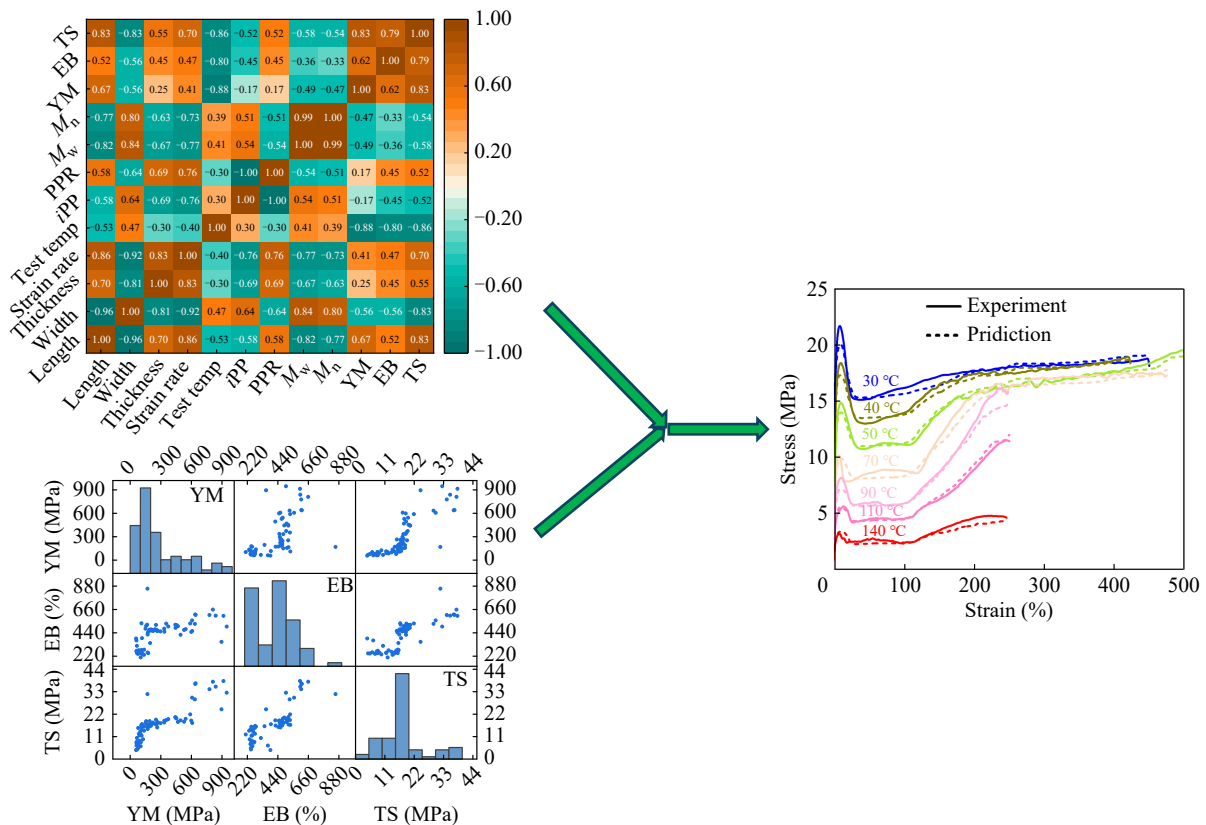
Machine Learning Predictive Models for the Tensile Properties of Polyethylene and Polypropylene

Le-kang Zhang¹, Xiao-yu Sun¹, Jing-qing Li¹, Lun-yang Liu², Tao Liao², Ying Lu²,
 Hong-fei Li^{2*}, Yong-feng Men^{2*}, Shi-chun Jiang^{1*}

(¹School of Materials Science and Engineering, Tianjin University, Tianjin 300072)

(²State Key Laboratory of Polymer Science and Technology, Changchun Institute of Applied Chemistry,
 Chinese Academy of Sciences, Changchun 130022)

Abstract Polyethylene (PE) and polypropylene (PP) are among the most productive and extensively applied polymer materials worldwide, and precise regulation of their properties is a core prerequisite for attaining high-performance characteristics. In this study, a multi-feature machine learning prediction model was established based on the engineering stress-strain curves and three key mechanical indicators (Young's modulus (YM), elongation at break (EB), and tensile strength (TS)) of PE and PP. The model was employed to investigate the influence mechanisms of material attributes, sample preparation processes, and testing parameters on the mechanical properties of target polymers. Seven distinct algorithms, including Lasso, Decision Tree Regression (DTR), and Random Forest Regression (RF), were integrated with five-fold cross-validation and hyperparameter optimization strategies to realize the quantitative prediction of the mechanical properties. Furthermore, Extreme Gradient Boosting Regression (XGBoost, XGB) was adopted to achieve a high-fidelity fitting of the stress-strain



* Corresponding authors: Hong-fei Li, E-mail: hfli@ciac.ac.cn

Yong-feng Men, E-mail: men@ciac.ac.cn

Shi-chun Jiang, E-mail: scjiang@tju.edu.cn

curves. Due to the limitation of the amount of data, the prediction accuracy of the mechanical properties of PE has certain limitations. However, from the perspective of the overall trend of the prediction, this prediction method still has a good effect. The results demonstrate that the prediction performance for YM was optimal on the PP dataset, with the coefficient of determination (R^2) on the test set reaching no less than 0.80 and even exceeding 0.90 for certain models, without any obvious overfitting. In contrast, the predictive efficacy for EB and TS was constrained by the limited dataset size, where only the RF, XGBoost, and K-Nearest Neighbor (KNN) models exhibited reliable predictive capabilities on the PP dataset. Under variable control, the R^2 value for the prediction of the PP stress-strain curves surpassed 0.90, enabling the accurate reproduction of the temperature-dependent mechanical properties. This study verifies the feasibility of applying machine learning techniques to predict the mechanical properties of polyolefins, thereby providing promising technical support for the rapid evaluation of the mechanical performance of polymer materials.

Keywords Engineering stress-strain curve; Machine learning; Mechanical properties