

基于实验、分子动力学模拟和 QSPR 分析的玻璃力学性能 机器学习模型

严静萍^{1,2}, 王飞梅^{1,2}, 李波源^{1,2}, 邓 路¹, 胡丽丽¹

(1. 中国科学院上海光学精密机械研究所, 上海 201800; 2. 中国科学院大学, 北京 100049)

摘 要: 硅酸盐玻璃被广泛应用于建筑窗户、电子设备、光学器件、核废料处理和日用品等领域。随着科技的发展, 当代社会对新型玻璃材料的性能要求越来越高, 其中, 力学性能直接影响其使用寿命、安全性和功能性, 因而得到了广泛关注。鉴于玻璃组成的复杂性, 传统的试错法已经难以满足新材料的快速开发需求; 另一方面, 随着计算机技术的不断进步, 运用材料计算方法设计满足应用需求的玻璃组成成为新型玻璃研发的又一选择, 并逐渐在新型玻璃材料的快速迭代中起到越来越关键的作用。本文围绕多组分硅酸盐玻璃, 以力学性能为主要目标, 结合实验数据、分子动力学模拟和定量结构性能关系分析法, 构建了 3 种机器学习预测模型: 1) 基于实验测得数据的玻璃密度预测模型; 2) 基于实验密度数据, 运用分子动力学模拟计算获得性能数据所构建的玻璃杨氏模量预测模型; 3) 以少量实验数据为基础, 运用 QSPR 方法进行数据增强采样所构建的玻璃硬度预测模型。通过实验制备和测试, 进一步验证模型预测所得玻璃组成的相关性, 证明了模型预测的有效性。这为基于机器学习的玻璃性能预测提供了新思路, 有助于加快新型高性能玻璃的研发进程。

关键词: 硅酸盐玻璃; 机器学习; 定量结构-性能关系分析; 力学性能

中图分类号: J527.3 **文献标志码:** A **文章编号:** 0454-5648(2025)10-2808-12

网络出版时间: 2025-07-02



硅酸盐玻璃因其优异的物理化学性能, 被广泛应用于炊具制造、实验室仪器、建筑幕墙、电子设备显示器、生物医学器械和核废料固化处理等领域^[1-2]。提高玻璃的机械性能对于应对能源、通信和基础设施领域的重大挑战至关重要^[3]。特别是, 玻璃的刚度(如杨氏模量 E) 在柔性基板、显示器、光纤、建筑玻璃、超硬复合材料、手术设备或轻质建筑材料的加工中起着关键作用^[3-6]。因此, 开发具有可调控机械性能的新型玻璃体系已成为重要研究方向^[7-8]。玻璃是亚稳态的, 若从液态冷却得足够快, 其几乎可以包含元素周期表中的所有元素^[9]。目前, 可能的玻璃组合物总数约为 10^{52} , 且只有一小部分的玻璃组成被探索过^[9]。玻璃的巨大组成范围为组分调整带来无限可能, 这为开发具有特定性能的新材料带来了极大挑战。同时, 采用传统的试错法开发新材料实验周期长, 易受环境因

素影响, 效率大大降低。新玻璃的设计需要优化组成, 降低研究成本, 确保可加工性^[10]。

随着计算机技术的不断发展, 通过材料计算模拟开发具有预测能力的精确性能模型已成为必要。玻璃材料的建模方法包括基于经验到基于理论的多种方法^[11], 偏经验的模型有基于统计建模的逆向蒙特卡罗(Reverse Monte Carlo)方法、拓扑约束理论、有限元分析、数据驱动的机器学习方法等, 偏理论的模型有分子动力学(MD)模拟、Metropolis 方法、密度泛函理论等。其中定量结构-性能关系(QSPR)分析法是一种极具应用前景的方法, 其可以关联玻璃的结构与性能, 并表现出对未知的玻璃性能良好的预测性, 从而加速对目标性能的玻璃设计。QSPR 分析中的结构信息可以来自实验表征或计算模拟, 但有很多研究表明结合分子动力学模拟的 QSPR 分析更具有优势^[12-14]。分子动力学模拟可

收稿日期: 2025-04-10。 修订日期: 2025-05-06。

基金项目: 国家自然科学基金(52472017); 中国科学院高层次人才计划。

第一作者: 严静萍(1997—), 女, 博士研究生。

通信作者: 邓 路(1987—), 男, 博士, 研究员。

Received date: 2025-04-10. Revised date: 2025-05-06.

First author: YAN Jingping (1997—), female, Doctoral candidate.

E-mail: jpyan@siom.ac.cn

Correspondent author: DENG Lu (1987—), male, Ph.D., Professor.

E-mail: denglu@siom.ac.cn

以模拟材料的微观结构并提取复杂的结构信息,以便关联玻璃性能,为玻璃性能的预测提供支撑。早在 20 世纪 80 年代,Thomas F. Soules 等人便已经开始借助分子动力学模拟对玻璃的弹性模量进行模拟计算,不过当时他们通过计算所得的弹性模量比实验值高了一个数量级^[15]。随着算法优化和计算能力的提升,Yuan 等利用 BKS 势函数研究了块体石英玻璃的弹性行为,得到的弹性模量约为 85 GPa,高于实验值 75 GPa^[16]。Lu 等的研究表明,ZrO₂ 掺杂的钠钙硼硅酸盐和钙铝硅酸盐玻璃的弹性模量可以通过基于 MD 的 QSPR 分析进行预测^[17]。目前,基于 MD 的 QSPR 分析已经成功应用于多种玻璃性能,包括密度、热膨胀系数、杨氏模量、硬度、玻璃的初始溶解速率等^[12,17-19]。然而,目前的 QSPR 研究多聚焦于小范围玻璃组成的精准预测,对于多系列的大数据拟合效果欠佳。

近年来,机器学习(ML)方法的引入为预测玻璃的性能提供了全新视角^[20-22]。与基于物理的模型相比,基于 ML 的模型依赖从现有数据库中“学习”,将玻璃数据与合成参数建立关系从而生成预测模型。目前这种方法已经应用于玻璃材料中^[4,11,23-24],用于预测机械性能^[24-25]、溶解动力学^[20,26-28]、光学性能^[29-31]、玻璃的形成能力^[32]等,显示了 ML 方法在设计高级玻璃材料中潜在的应用前景。Yang 等利用高通量分子动力学模拟和基于不同模型的机器学习方法来预测硅酸盐玻璃的杨氏模量,结果表明神经网络模型在预测杨氏模量的非线性变化精度上表现最佳^[24]。Bishnoi 等则发现基于稀疏数据集预测硅酸盐玻璃的杨氏模量中,高斯过程回归明显优于神经网络^[25]。此外,Hu 等通过对多种钙铝硅酸盐玻璃体系进行大规模的分子动力学模拟,证明了随机森林模型是弹性常数预测的有效工具,其预测结果与分子动力学模拟结果具有高度一致性^[33]。Du 等也通过结合机器学习和高通量分子动力学模拟,有效预测了钙铝硅玻璃的机械性能,并揭示成分和制备条件对性能的影响^[34]。Liu 等对玻璃领域中使用的机器学习算法进行了综述^[35]。尽管许多机器学习算法可以达到高精度,但模型的性能在很大程度上依赖于数据,其对数据的要求包括可用(易于获取)、完整(参数范围大)、一致性(用相同的测试方案获得)、准确(避免“垃圾进,垃圾出”模式),以及代表性(数据集需要提供足够的信息来训练模型)。虽然已经有了一些关于玻璃性能的数据库^[36],但是玻璃的生产条件或测试方式的不一致可能会

使其直接用作机器学习方法的训练集成为挑战,需要在数据清理和无偏离群值检测方面做出一些努力。

应用机器学习预测材料性能通常需要大量的高质量数据进行训练,这对数据集要求较高。本研究为加快设计具有目标力学性能要求的玻璃配方,针对不同的玻璃性能(密度、杨氏模量、硬度)构建了 3 个机器学习模型,其中数据集分别源于 SciGlass 数据库、MD 模拟、QSPR 模型。其中 SciGlass 数据库拥有 40 多万种玻璃及其相应的物理和化学性能,在玻璃界得到广泛使用^[37]。然而,该数据库的开发和维护在 2013 年左右停止,最终版本在 GitHub 上开源共享。Cassar 开发了一个 Python 模块 PyGlass 用于访问 SciGlass 的数据,并为许多玻璃性能构建预测 ML 模型^[38]。文章主要由以下几个部分组成:研究方法包括对分子动力学模拟方法、QSPR 分析方法、机器学习算法及实验方法的详细介绍;结果与讨论部分包括对玻璃样品数据集的分析、机器学习模型预测性能的分析以及机器学习模型的可解释性;最后得出结论。

1 研究方法

1.1 分子动力学模拟

采用 LAMMPS 软件对由 SciGlass 数据库筛选得到的六元硅酸盐玻璃(SiO₂-B₂O₃-Al₂O₃-MgO-CaO-Na₂O)进行分子动力学模拟^[38-39]。为准确描述硼硅酸盐玻璃体系的网络结构,本研究选用了 Teter 势函数^[40-41]。与其他势函数相比,Teter 势在描述硼在多组分玻璃网络结构特点方面,特别是硼硅酸盐玻璃中关键的硼配位数转变具有明显优势^[42],因而更适合研究本文中的玻璃体系。此外,其也被广泛应用于多个玻璃体系,包括硅酸钠玻璃^[43-44]、磷酸盐玻璃^[14,45]等。该势函数包含长程库伦项和短程 Buckingham 项:

$$V_{ij}(r) = \frac{Z_i Z_j e^2}{4\pi\epsilon_0 r_{ij}} + A_{ij} \exp\left(\frac{-r_{ij}}{\rho_{ij}}\right) - \frac{C_{ij}}{r_{ij}^6} \quad (1)$$

式中: ϵ_0 为真空介电常数; e 为单电子电荷; r_{ij} 为粒子对 i 和 j 的原子间距离; Z_i 和 Z_j 表示离子的偏电荷; A_{ij} 、 ρ_{ij} 和 C_{ij} 值是 Buckingham 项的拟合参数。后续修正了近距离内粒子对的势能表达:

$$V'(r_{ij}) = \frac{B_{ij}}{r_{ij}^n} + D_{ij} \cdot r_{ij}^2 \quad (2)$$

其中 B_{ij} 、 D_{ij} 、和 n 是拟合参数。相关势函数的参数

列于表 1。玻璃的熔融淬火过程与我们以往的研究相同^[14]。玻璃的初始结构基于实验密度随机生成，原子数约 6000 个。模拟过程包括：系统先在 0 K 下进行能量最小化，300 K 下弛豫后升温至 6000 K

完全熔融，之后以 2.5 K/ps 的速率冷却至 300 K，最终分别在 NPT、NVT 及 NVE 系综下进行弛豫。对于室温弛豫的最后 100 ps，每 50 步记录 1 次信息，用于获得模拟玻璃的结构信息^[46]。

表 1 Buckingham 势的参数^[40,42,47]
Table 1 Parameters for the Buckingham potential^[40,42,47]

Pairs	A_{ij}/eV	$\rho_{ij}/\text{\AA}$	$C_{ij}/(\text{eV} \cdot \text{\AA}^6)$	B_{ij}	n_{ij}	D_{ij}
Si ^{2.4} -O ^{-1.2}	13 702.905	0.193 817	54.681	28.9489	3.9329	-3.0565
B ^{1.8} -O ^{-1.2}	13 441.648	0.171 271	28.500	13.2985	4.3014	-2.5465
Al ^{1.8} -O ^{-1.2}	12 201.417	0.195 628	31.997	51.7248	3.1794	-10.1935
Mg ^{1.2} -O ^{-1.2}	7063.490	0.210 901	19.210	54.0765	2.8416	-11.6575
Ca ^{1.2} -O ^{-1.2}	7747.183	0.252 623	93.109	74.7448	3.1676	-3.9866
Na ^{0.6} -O ^{-1.2}	4383.756	0.243 838	30.700	48.2619	2.8938	-4.7252
O ^{-1.2} -O ^{-1.2}	2029.220	0.343 645	192.580	46.3392	3.4243	-0.3275

通过分子动力学模拟可以计算得到多种玻璃的力学性能，包括体积模量、剪切模量、杨氏模量以及泊松比，具体方法如下。作为一种描述材料刚度的重要方式，弹性常数矩阵通常可以通过分析变形过程应力张量和应变张量之间的关系来求解，该关系可表示为：

$$S_{ij} = C_{ijkl} * \epsilon_{kl} \quad (3)$$

其中： S_{ij} 为对称应力张量元素； C_{ijkl} 是 4 阶弹性常数； ϵ_{kl} 是对称应变张量元素。由于材料的对称性，独立的弹性常数可减少至 21 个。利用 LAMMPS 软件从 6 个方向(x, y, z, xy, yz, xz)对熔融-淬火过程得到的最终玻璃结构施加形变，并采用零能量最小化方法计算应力响应来求解弹性常数矩阵。使用 Hill 方法^[48]来计算弹性模量，即对 Voigt 和 Reuss 近似模型的结果取几何平均值。在 Voigt 和 Reuss 近似下的体积模量可表示为式(4)和式(5)：

$$K_R = \frac{1}{(S_{11} + S_{22} + S_{33}) + 2 \times (S_{12} + S_{23} + S_{31})} \quad (4)$$

$$K_v = \frac{1}{9} \times [(C_{11} + C_{22} + C_{33}) + 2 \times (C_{12} + C_{23} + C_{31})] \quad (5)$$

剪切模量可表示为式(6)和式(7)：

$$G_R = \frac{15}{4 \times (S_{11} + S_{22} + S_{33}) - 4 \times (S_{12} + S_{23} + S_{31}) + 3 \times (S_{44} + S_{55} + S_{66})} \quad (6)$$

$$G_v = \frac{1}{15} \times [(C_{11} + C_{22} + C_{33}) - (C_{12} + C_{23} + C_{31}) + 3 \times (C_{44} + C_{55} + C_{66})] \quad (7)$$

杨氏模量 E 和泊松比 ν 可通过式 (5) 和式 (6) 代入体积模量和剪切模量得到：

$$E = \left(\frac{1}{3 \times G} + \frac{1}{9 \times K} \right)^{-1} \quad (8)$$

$$\nu = \frac{1}{2} \times \left[1 - \frac{3 \times G}{3 \times K + G} \right] \quad (9)$$

1.2 QSPR 分析

QSPR 理论结构描述符最早由 Lusvardi 等^[18]提出，用于氟磷硅酸盐体系的 T_g 和化学稳定性预测。Lu 等^[12]采用单键键强等对原有结构描述符中能量描述项进行了优化，并成功应用于多种氧化物玻璃体系，结果的拟合度和模型预测准确度较好。但现有的 QSPR 方法受到实验能量数据不完备的限制，其应用范围有一定的局限性，例如，现有单键键强参数体系缺少五配位 Al—O 的单键键强数据。相比于基于传统键强参数的 QSPR 模型，基于电负性的 QSPR 模型的优势包括其可以覆盖大多数元素、实现最小化实验输入以及兼容同一体系不同成分含量的玻璃系列^[49]。因此，采用 Tantardini 等^[50]的电负性表示的能量项替代传统的能量项。描述符主要包含组成、结构以及电负性表示的能量项。鉴于玻璃的硬度与键强、网络连通性相关，额外的修饰体会导致玻璃网络解聚，从而影响硬度。因此 N_M 与硬度呈现负相关，且选取 Q^n 结构与硬度关联。所用的理论结构描述符如下所示：

$$F_{\text{net}} = \frac{1}{N_M} \sum_X^{\text{formers}} Q^n \cdot \Delta \chi_{XO}^2 \quad (10)$$

其中： N_M 是玻璃网络修饰体(Mg、Ca、Na)的总原子数， $\Delta \chi_{XO}$ 是元素 X 和 O 的电负性差值， $\Delta \chi_{XO}^2$ 在数值上等于离子稳定能^[51]。 Q^n 代表玻璃中的桥氧数量。

为构建硬度相关的 QSPR 模型, 从 SciGlass 数据库中筛选得到了 43 个硼硅酸盐玻璃组成及相应的密度和硬度数据, 实现玻璃组成-模拟结构-实验性能的关联。由该 QSPR 模型进行增强数据采样, 即进一步预测了 50 个玻璃组成的硬度值, 最后结合实验数据和预测数据(共 93 个数据)构建了机器学习的训练数据库。利用 QSPR 模型进行数据增强, 可实现同类数据的快速生成, 弥补传统数据库中实验数据量不足、质量参差以及采集周期长等问题, 有效提高后续机器学习建模的准确性与可靠性。

1.3 机器学习算法

图 1 为机器学习驱动的玻璃性能模型构建流程图。构建了关于硅酸盐玻璃的密度、杨氏模量

以及硬度的机器学习模型, 数据来源于 SciGlass 数据库筛选得到的 425 个硅酸盐玻璃密度和 43 个硬度值、MD 模拟得到的 425 个杨氏模量值以及 QSPR 模型预测得到的 50 个硬度值。SciGlass 数据库提供了本研究中使用的密度和部分硬度数据, 该数据库在 GitHub 存储库中对外开放^[52]。主要针对六元硼硅酸盐玻璃($\text{SiO}_2\text{-B}_2\text{O}_3\text{-Al}_2\text{O}_3\text{-MgO-CaO-Na}_2\text{O}$), 组成筛选条件为形成体含量(摩尔分数)总和大于 50%, 各修饰体含量小于 50%。该筛选条件有助于得到具有稳定网络结构的玻璃组成, 同时可以避免极端组成对模型的干扰。获得初始玻璃数据集后进一步开展数据清洗, 删除极值与重复的数据, 并进一步对氧化物组成进行归一化处理。

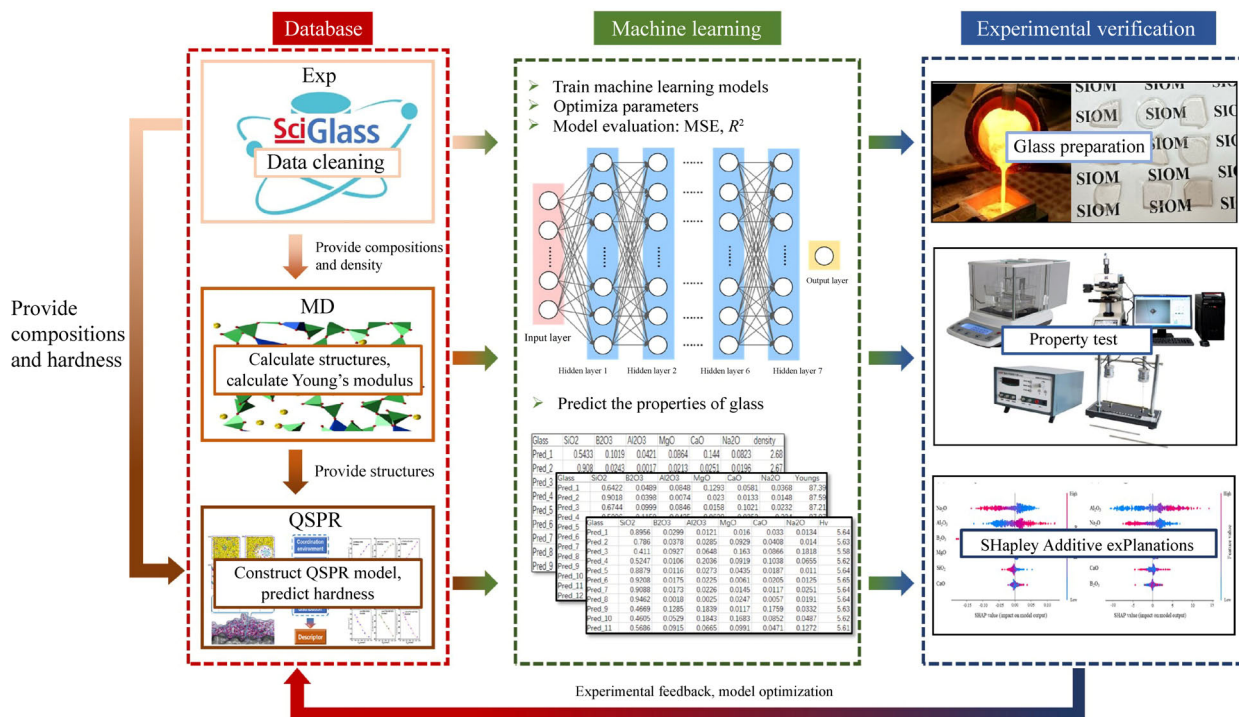


图 1 机器学习驱动的玻璃性能模型构建流程

Fig. 1 Process of constructing the machine learning model of glass properties

采用多层神经网络 (Multilayer Perceptron, MLP) 来建立硅酸盐玻璃组成与其物理性能(密度、杨氏模量、硬度)之间的映射关系。神经网络由一个输入层、若干隐藏层和一个输出层组成, 且各层神经元通过加权连接^[53]。每个模型以氧化物的摩尔百分比作为输入变量, 输出性能数据。隐藏层中的每个神经元通过 ReLU 激活函数引入非线性, 使模型能够拟合复杂的非线性关系。输出层使用线性激活函数, 预测玻璃的性能。Keras 框架具有简洁易用、高度模块化、支持快速构建和调试神经网络模型,

且可与 TensorFlow 无缝集成的优势^[54], 因此采用 Keras 框架构建和训练模型。在模型训练过程中, 鉴于数据量有限, 按照 9:1 的比例将自建的数据集划分为训练集和测试集^[55,56]。该划分方式在保证训练样本充分性的同时, 也保留了足够的测试数据用于评估模型的泛化能力。此外, 以均方误差 (Mean Squared Error, MSE) 和决定系数 (R^2) 作为所得模型的评估指标, 其定义分别为:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (11)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (12)$$

其中： y_i 为实验值； $\bar{y}_i$ 为实验平均值； $\hat{y}_i$ 为模型预测值。训练过程基于反向传播算法(Back Propagation)和 Adam 优化器，dropout 率为 0.2。采用随机搜索进行超参数优化，针对神经网络模型设定的超参数搜索空间如下：隐藏层数(2~10 层)、每层神经元数(10, 16, 32)。在设定的搜索空间内，随机采样 20 组超参数组合，基于均方误差(MSE)进行评估，以获得最优模型。最终确定模型的输入层后接 7 个隐藏层，每层包含 10 个神经元，训练 5000 次。之后利用训练完成的模型预测玻璃的密度、杨氏模量和硬度，并对预测得到的 9 个配方进行实验验证。玻璃的详细组成、性能及预测的玻璃配方参看补充信息。此外，结合 SHAP 解释结果，反馈修正预测模型，构成一个数据-模型-实验三位一体的闭环。

1.4 实验

采用传统的熔融冷却法对机器学习预测的玻璃样品进行制备。所有实验制备的玻璃样品均由其组成中相应元素的氧化物、碳酸盐、氢氧化物等作为原料。称取原料共计 40 g，将原料充分混合后，在空气气氛下置于 1450 °C 刚玉坩埚中熔融 3 h，中途偶尔加以搅拌以帮助样品完全熔融，之后将玻璃液倒在铁板模具上冷凝成型，快速放入退火炉中于 590 °C 进行退火。

对退火后的玻璃进行进一步的性能测试，包括密度、杨氏模量值以及硬度。玻璃的密度基于阿基米德排水法原理，采用日本 ALFA MIRAGE SD-200L 型电子固体密度计进行测试。玻璃的杨氏模量采用瑞泰克公司的 TECLAB UMS-200 型杨氏测量仪，通过超声波测声速法进行测量，其中纵波和横波探头的频率分别为 5 MHz 和 5 MHz，为避免超声波在空气中传播，需在样品和探头间填充耦合剂^[57]。相较于传统的静态或动态测试方法，超声波法在精度与重复性方面具有更大优势。玻璃的泊松比 ν 、剪切模量 G 和杨氏模量 E 分别可以由以下公式得到：

$$\nu = \frac{\left(\frac{1}{2}\right)c_l^2 - c_t^2}{c_l^2 - c_t^2} \quad (13)$$

$$G = \rho c_t^2 \quad (14)$$

$$E = 2G(1 + \nu) \quad (15)$$

其中： c_l 和 c_t 分别代表纵波和横波声速； ρ 表示玻璃

的密度，杨氏模量的测试误差在 $\pm 1\%$ 左右。玻璃的硬度采用日本 Future Tech 公司的 FM-800 型数字显微硬度计进行测试，测试载荷 100 g，持续 10 s。为减小误差，每个样品测量 5 个点，取平均值得到维氏硬度 H_v 值。

2 结果与讨论

2.1 玻璃样品数据集的数据组成

图 2 为玻璃的密度、杨氏模量及硬度数据集中组成和性能的分布情况，为后续分析提供了直观依据。其中，玻璃的密度数据集源于 SciGlass 数据库，经过筛选后得到 425 组六元硼硅酸盐玻璃组成以及其对应的密度数据。同时，以这些数据作为 MD 模拟的输入，通过计算得到相应玻璃组成的杨氏模量值，并建立了杨氏模量数据集。玻璃的密度和杨氏模量数据集的组成范围相同，涵盖了二氧化硅含量 40~90% (摩尔分数，下同)，硼铝氧化物含量 0~40%，修饰体含量 0~40%。玻璃的性能主要集中于三元图的右下角。在密度数据集中，玻璃密度分布在 2.3~2.8 g/cm³，集中于 2.4~2.6 g/cm³，约占数据集的 78%。由图 2a 可以看出，低氧化硅、高修饰体氧化物含量的玻璃密度较高，反之玻璃密度较低。而杨氏模量数据集中，玻璃的杨氏模量主要集中在 82~115 GPa，且高氧化硅含量玻璃组成的杨氏模量相对较高。

玻璃的硬度数据集共包含 93 组数据，主要由两部分组成：第一部分是由 SciGlass 数据库筛选得到的 43 组六元硼硅酸盐玻璃组成及硬度数据，并结合 MD 模拟所得结构构建 QSPR 模型；第二部分是运用所建立的 QSPR 模型预测得到的 50 组数据，为后续机器学习模型的建立提供了数据，以便提高模型的泛化能力和预测准确性。以往的很多工作已经证实了由 MD 模拟的玻璃结构与实验结果吻合，验证了 MD 模拟在描述玻璃微观结构方面的可靠性^[14,58,59]。在这项工作中，选择 Q^n 结构作为 QSPR 建模的结构输入，将模拟结构和实验性能相关联，建立硬度相关的 QSPR 模型，如图 3 所示。模型显示硬度与构建的 F_{net} 描述符之间具有良好的线性关系， R^2 值大于 0.80，表明该模型具有较好的拟合能力和预测能力。硬度数据集的玻璃组成中，二氧化硅含量 45%~80%，硼铝氧化物含量 3%~28%，修饰体氧化物含量 9%~30%，硬度值分布在 5~7 GPa。相比于实验获得的数据，应用基于 MD 模拟的 QSPR 方法得到的预测数据不仅可以减少实验工作

量,降低研究成本,同时能保证数据的同源性和一致性,避免因实验条件变化引起的误差,从而提高后续机器学习模型的稳定性。

2.2 机器学习模型预测准确性分析

构建了 3 个机器学习模型,分别关联硅酸盐玻璃的密度、杨氏模量及硬度性能。这 3 个机器学习模型

的数据集来源各不相同,密度数据源于 SciGlass 数据库,杨氏模量数据源于分子动力学模拟,硬度数据使用了 QSPR 模型进行数据增强。如图 4 所示是机器学习模型的预测结果,3 个模型的 R^2 值表现出明显的差距,其中密度相关的机器学习模型 R^2 值最低,测试集 R^2 值仅为 0.300, MSE 值为 0.01。

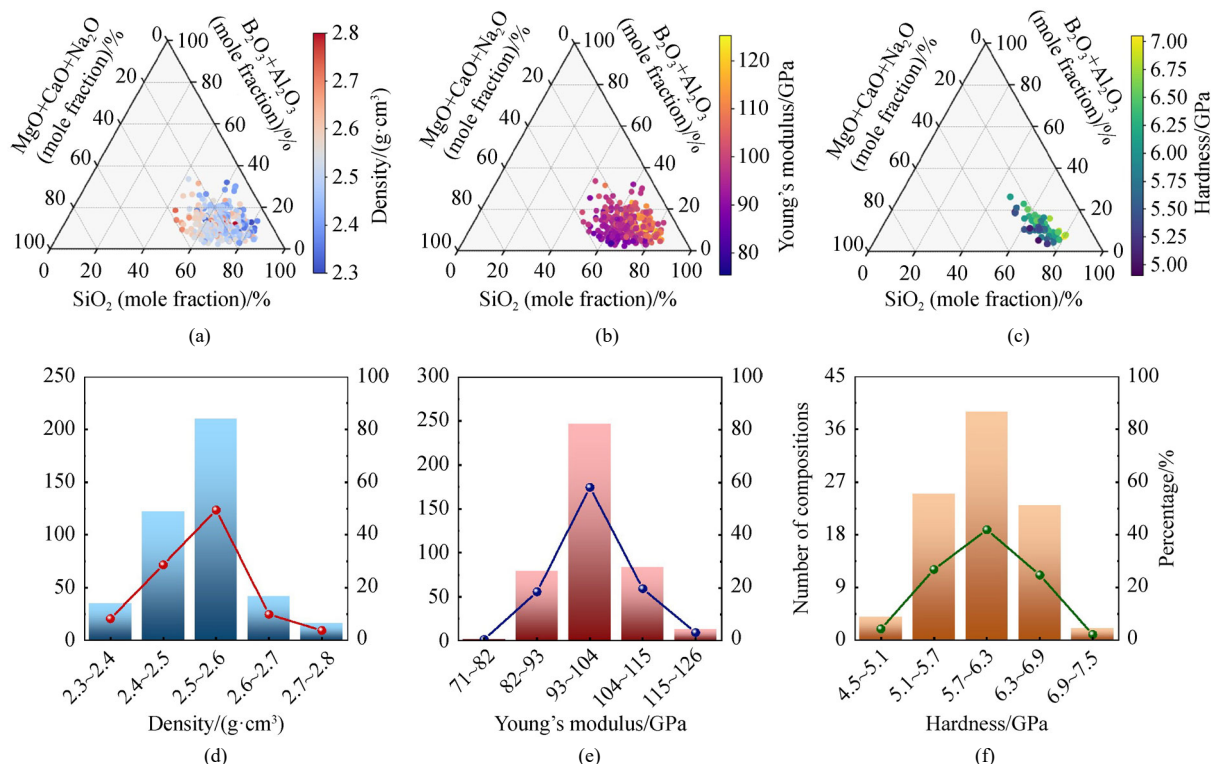


图 2 玻璃数据集中组成和性能分布情况: (a, d) 密度数据集、(b, e) 杨氏模量数据集以及 (c, f) 硬度数据集。(d~f) 中的折线表示数据占比

Fig. 2 Distribution of composition and properties in the glass dataset: (a, d) density, (b, e) Young's modulus and (c, f) hardness. The line-symbol curves in (d~f) represent the percentages of the species in each dataset

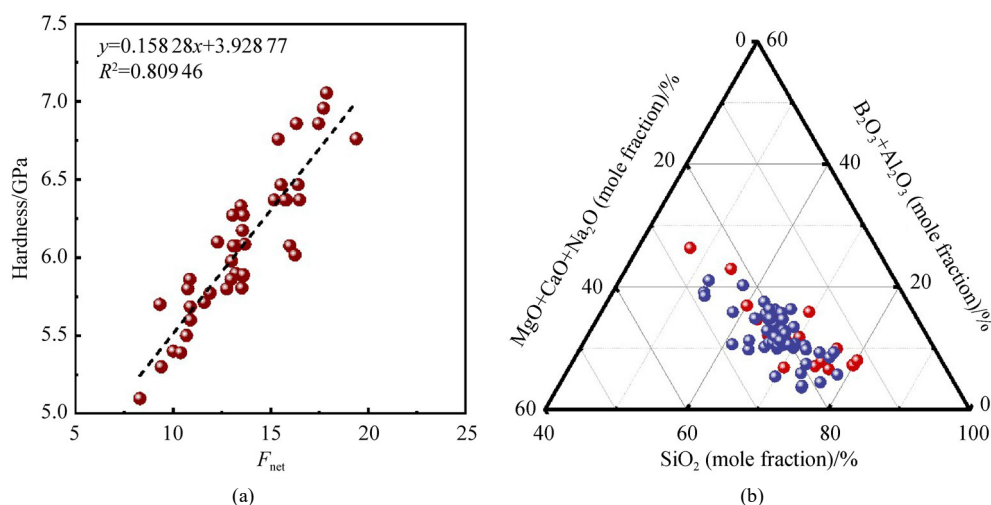


图 3 关于硅酸盐玻璃硬度的 (a) QSPR 模型及 (b) 组成分布, 其中红球表示实验值, 蓝球表示 QSPR 模型预测值

Fig. 3 (a) QSPR model and (b) composition distribution for the hardness of silicate glasses, where red spheres represent experimental values and blue spheres represent predicted values

利用 SciGlass 数据库, 可以实现玻璃组成以及目标性能的快速获取, 但不同制备条件、冷却方法、玻璃组成挥发性、气泡含量、测试人员及方法等因素的影响, 导致数据质量不一, 玻璃制备方法和性能测试的多样性可能是密度预测机器学习模型 R^2 不高的原因之一, 因而降低了模型的稳定性和预测准确性。分子动力学模拟可以很好地保证数据的同源性, 且相对实验来说, 可以快速地为机器学习模型

积累数据, 同时达到一定的 R^2 值。杨氏模量相关的机器学习模型 R^2 可以达到 0.7, 而 MSE 值为 15.35, 误差值较高。尽管 MD 能够模拟含有多组分氧化物玻璃, 然而用相同的模拟过程生成数千或上万个玻璃的 MD 模拟结构仍需要大量的计算资源, 特别是模拟的玻璃尺寸和冷却速率对最终模拟结构会产生影响^[60]。此外, 通过 MD 模拟很难准确获得玻璃所有的性能, 而 QSPR 方法可以实现基于少量数据的大量同源数据的快速预测, 从而实现数据增强采样的作用, 为机器学习积累数据。如图 4c 所示, 硬度相关的机器学习模型 R^2 大于 0.95, 且 MSE 值为 0.01。本工作中的硬度数据 46.24% 源于 SciGlass, 剩余 53.76% 源于 QSPR 模型预测。QSPR 模型中结构特征的引入可以提高模型的可解释性, 克服传统机器学习模型仅使用化学组成作为模型特征的局限性。

由上述 3 个机器学习模型进行性能预测, 选取 9 个玻璃配方进行实验验证, 结果如表 2 所示。预测配方详见补充材料 S3, 配方中的 SiO_2 控制在 40~60 mol% 左右以便于实验制备玻璃。尽管密度相关的 ML 模型的 R^2 并不高, 但其仍然表现出较好的预测性, 最高的相对误差(Exp3)仅 2.67%, 最低的相对误差 0.28%。杨氏模量相关的 ML 模型的预测值几乎高于实验值 10~15 GPa, 相对误差值大多在 10%~20% 之间。其中, 实验测量误差为 $\pm 1\%$ (约 1 Ga), 而模拟误差约为 10 GPa。由文献可知, 模拟所得的玻璃杨氏模量值通常受到经验势的影响, BKS 势和 Teter 势函数模拟的结果与实验相比, 通常高约 10~15 GPa^[16,46]。因此, ML 模型预测值偏高主要源于经验势函数导致的模拟偏差, 但总体而言, 模拟产生的数据一致性较好, 可用于 ML 模型的构建。然而也有个别相对误差较小, 例如 Exp6 样品的相对误差仅为 1.92%。硬度相关的模型的预测值与实验值相差 0.05~0.4 GPa, 相对误差为 1%~6%。其中一部分误差来源于仪器测试误差, 约 0.049 GPa, 另一部分误差可能来源于 MD 模拟计算的玻璃结构, 如模拟得到的 Al 几乎为四配位, 与实际情况存在差异, 这主要是由于经验势函数的限制。综上所述, 本工作构建的 3 个机器学习模型均表现出了一定的性能预测性且偏差在可接受范围内。如果没有充足的高质量实验数据作为训练集, 那么利用 MD 模拟和 QSPR 模型无疑是一种较好的选择。未来研究可以通过多种措施来进一步提高模型预测准确性, 包括: 优化势函数参数, 使其

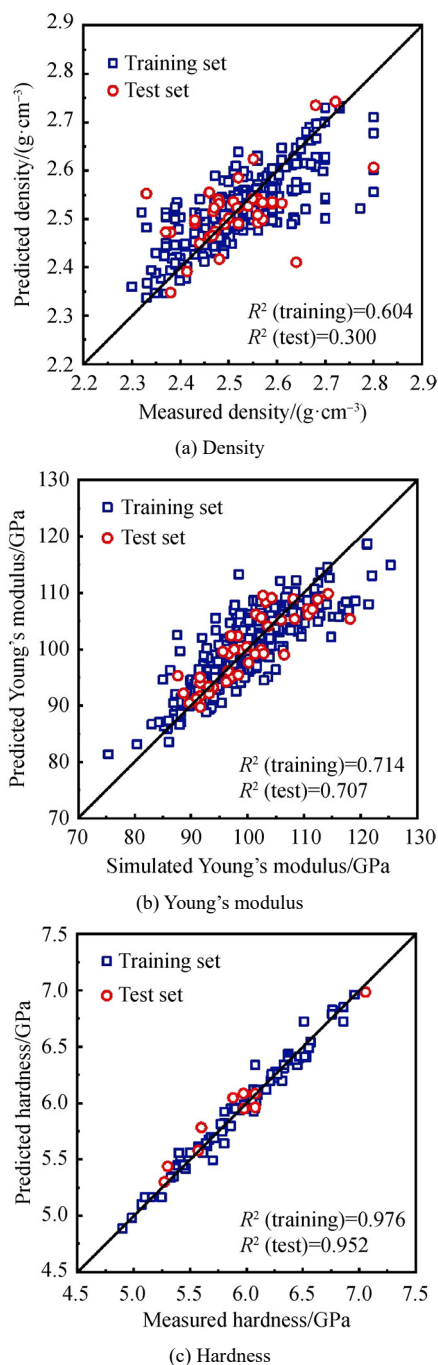


图4 不同性能的机器学习模型的预测结果

Fig. 4 Predicted results of machine learning models for different properties

更适配六元硼硅酸盐体系；采用更大体系尺寸，提升模拟的准确性；增加误差传播分析，更准确量化各误差源的贡献。

2.3 机器学习模型的可解释性

机器学习模型，特别是基于神经网络的模型，因其大量的层和参数通常被视为“黑盒”，难以直接理解其内部决策机制。此外，机器学习模型在训练过程中依赖有限的实验数据，并且涉及非线性函数的拟合，这进一步增加了其结果的解释难度。这些挑战凸显了对透明和可解释的机器学习模型的需求，以促进机器学习方法在玻璃材料设计中的解

释和应用。SHAP(SHapley Additive exPlanations)分析是一种基于合作博弈论的方法，能够为机器学习模型提供解释^[61]。SHAP 值衡量每个特征在预测中的边际贡献，考虑了特征的个体效应以及其与其他特征的交互作用。相比于传统的特征重要性评估方法，如排列特征重要性(FIP)^[62]，SHAP 分析提供了更有细粒度的解释，使得研究者能够更直观地理解特征如何影响预测结果。

在本研究的玻璃材料性能预测中，SHAP 分析被用于探究不同氧化物对密度、杨氏模量及硬度的影响。图 5 是相关性能预测模型的 SHAP 分析蜂群

表 2 3 个 ML 模型的预测配方验证结果
Table 2 Predictive recipe validation results for three ML models

Glass	Density			Young's modulus			Hardness		
	Pred./($\text{g}\cdot\text{cm}^{-3}$)	Exp./($\text{g}\cdot\text{cm}^{-3}$)	Rel. Err./%	Pred./GPa	Exp./GPa	Rel. Err./%	Pred./GPa	Exp./GPa	Rel. Err./%
Exp1	2.57	2.58	0.28	95.85	83.67	14.55	6.38	6.50	1.92
Exp2	2.57	2.59	0.75	96.97	86.33	12.32	6.42	6.63	3.12
Exp3	2.57	2.53	1.61	92.42	79.63	16.06	6.68	6.33	5.52
Exp4	2.41	2.44	1.18	94.76	79.47	19.23	5.87	5.73	2.50
Exp5	2.60	2.62	0.90	94.57	85.93	10.18	6.34	6.47	1.97
Exp6	2.61	2.68	2.67	94.71	96.56	1.92	6.81	7.07	3.62
Exp7	2.55	2.53	0.78	95.6	83.68	14.25	5.99	6.38	6.11
Exp8	2.53	2.52	0.53	90.84	79.18	14.73	6.01	6.34	5.20
Exp9	2.45	2.49	1.74	87.33	76.04	14.85	5.99	6.05	1.03

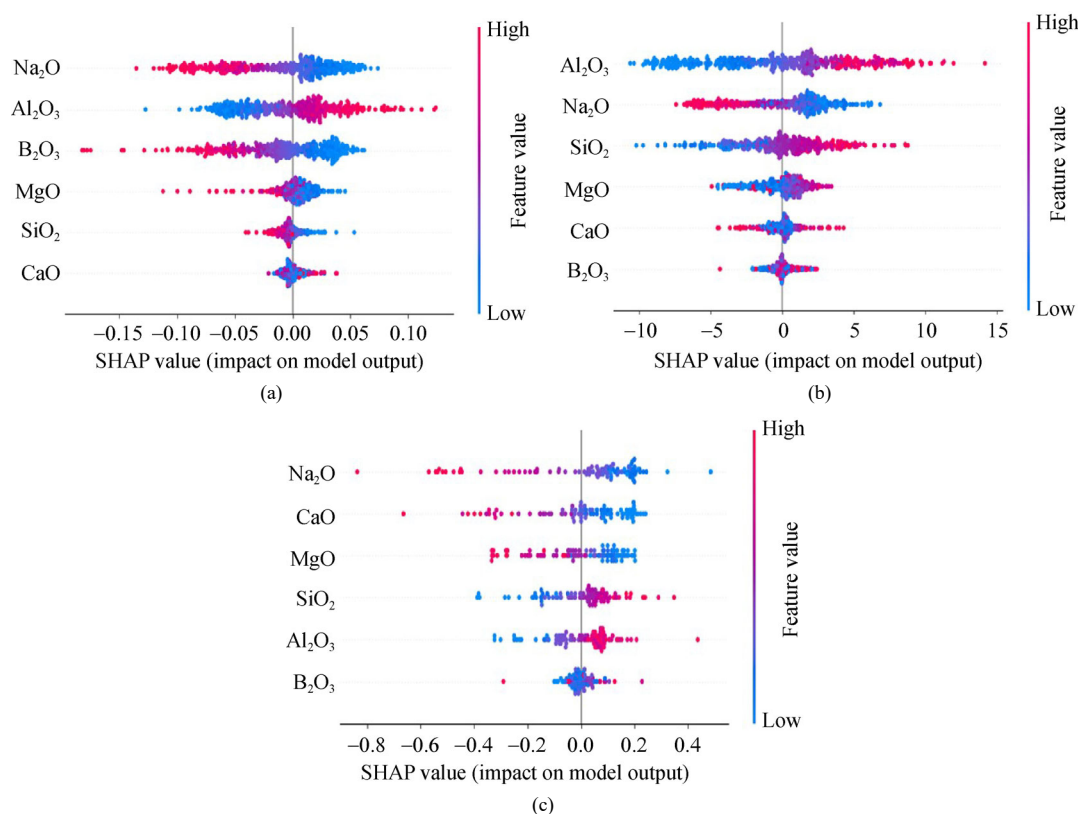


图 5 ML 模型的 SHAP 分析蜂群图，其中关联的玻璃性能分别为(a)密度、(b)杨氏模量和(c)硬度
Fig. 5 Beeswarm plots of SHAP values for (a) density, (b) Young's modulus, and (c) hardness

图, 不仅能够提供整体上的特征重要性排名, 还能具体到每个样本, 揭示其对模型输出的正向或负向贡献。在蜂群图的左侧, 氧化物根据 SHAP 值的平均绝对值进行了重要性排序。图中水平轴上的有色点代表包含该氧化物的一个玻璃样本, 而点的颜色反映该氧化物在样本中的含量, 随着右侧色标中的冷色到暖色, 对应氧化物含量由低到高。点在 x 轴的位置则表示该氧化物在样本中的 SHAP 值, 具有相同 SHAP 值的点在垂直方向堆叠。如图 5a 所示, 硅酸盐玻璃的密度受 Na_2O 、 Al_2O_3 、 B_2O_3 的影响较大。作为修饰体的 Na_2O 和 MgO 含量的增加会导致玻璃网络解聚从而引起密度降低, 而 CaO 则表现出一定的两面性。 Al_2O_3 的加入会使得玻璃结构发生交联, 致使网络更加致密。 SiO_2 和 B_2O_3 的增加会导致硅酸盐玻璃的密度降低, 这是因为大量的形成体在玻璃中形成网络聚合, 玻璃的空隙增多, 自由体积增加, 从而导致玻璃密度降低^[63]。对于杨氏模量而言, SiO_2 和 Al_2O_3 的增加通常会增强玻璃的杨氏模量。这是因为 $[\text{AlO}_4]$ 结构能够连接 $[\text{SiO}_4]$ 单元及网络外体, 提高其刚性。然而, B_2O_3 作为网络形成体并未表现出显著的正向关联性, 这与先前研究^[64]的结果相符。据报道, 在硅酸钙玻璃中, SiO_2 、 B_2O_3 和 Al_2O_3 的共同作用可能导致杨氏模量的非单调变化。此外, Na_2O 的增加会降低玻璃的杨氏模量, 而具有高阳离子场强的 MgO 则对杨氏模量表现出正向影响。值得注意的是, CaO 对杨氏模量的影响与其对密度的影响类似, 呈现出双重作用。图 5c 表明, Na_2O 、 CaO 和 MgO 含量的增加均会降低玻璃的硬度, 而 SiO_2 、 Al_2O_3 和 B_2O_3 含量的增加则会增强玻璃的硬度。这一现象可由刚性理论解释: 玻璃的硬度与其网络的拓扑约束数量成正比, 严格来说, 是与单位体积内的角约束数量成正比, 而非与每个原子的约束总数成正比^[65]。

3 结论

建立了包含不同来源的玻璃性能数据集, 数据包括 SciGlass 数据库提供的 425 个玻璃实验密度值和 43 个实验硬度值、MD 模拟计算得到的 425 个杨氏模量值以及 QSPR 模型预测的 50 个硬度值。基于上述数据集, 分别构建了基于实验测试结果、基于计算模拟结果、基于 QSPR 数据增强的 3 种机器学习预测模型。基于以上 3 个模型预测出 9 个配方并进行了 3 个性能(密度、杨氏模量和硬度)的实验验证, 相对误差较小(其中, 杨氏模量模型考虑了

模拟结果带来的误差影响), 预测结果较为理想。随着机器学习方法的兴起, 数据质量对于模型预测准确性的影响非常重要。本工作在获取实验数据的基础上, 通过 MD 模拟方法获取同源性的玻璃性能数据, 基于 QSPR 方法实现快速的数据增强采样, 打破了传统数据库构建过程中实验数据参差不齐、数据量获取慢等局限, 为实现数据驱动的玻璃材料研究新范式提供了新的思路。

(第二十七届中国科协年会学术论文)

参考文献:

- [1] SHELBY J E. Introduction to Glass Science and Technology[M]. The Royal Society of Chemistry, 2005: 249–261.
- [2] GIN S, ABDELOUAS A, CRISCENTI L J, et al. An international initiative on long-term behavior of high-level nuclear waste glass[J]. Mater Today, 2013, 16(6): 243–248.
- [3] WONDRACZEK L, MAURO J C, ECKERT J, et al. Towards ultrastrong glasses[J]. Adv Mater, 2011, 23(39): 4578–4586.
- [4] ONBAŞLI M C, TANDIA A, MAURO J C. Mechanical and Compositional Design of High-Strength Corning Gorilla® Glass[C]// Handbook of Materials Modeling. Cham: Springer International Publishing: 1–23.
- [5] ROUXEL T. Elastic properties of glasses: A multiscale approach[J]. Comptes Rendus Mécanique, 2006, 334(12): 743–753.
- [6] ROUXEL T. Elastic Properties and Short-to Medium-Range Order in Glasses[J]. J. Am. Ceram. Soc., 2007, 90(10): 3019–3039.
- [7] MAURO J C, PHILIP C S, VAUGHN D J, et al. Glass science in the United States: Current status and future directions[J]. Int J Appl Glass Sci, 2014, 5(1): 2–15.
- [8] MAURO J C, ZANOTTO E D. Two centuries of glass research: Historical trends, current status, and grand challenges for the future[J]. Int J Appl Glass Sci, 2014, 5(3): 313–327.
- [9] ZANOTTO E D, COUTINHO F A B. How many non-crystalline solids can be made from all the elements of the periodic table?[J]. J Non Cryst Solids, 2004, 347(1–3): 285–288.
- [10] MAURO J C. Decoding the glass genome[J]. Curr Opin Solid State Mater Sci, 2018, 22(2): 58–64.
- [11] MAURO J C, TANDIA A, VARGHEESE K D, et al. Accelerating the design of functional glasses through modeling[J]. Chem Mater, 2016, 28(12): 4267–4277.
- [12] LU X N, DU J C. Quantitative structure-property relationship (QSPR) analysis of calcium aluminosilicate glasses based on molecular dynamics simulations[J]. J Non Cryst Solids, 2020, 530: 119772.
- [13] DU J C, LU X N, GIN S, et al. Predicting the dissolution rate of borosilicate glasses using QSPR analysis based on molecular dynamics simulations[J]. J Am Ceram Soc, 2021, 104(9): 4445–4458.
- [14] YAN J, ZHANG Y, WANG F, et al. Effect of Al_2O_3 on the structure–property relationship of sodium aluminophosphate glasses: A combined study of experiments, MD simulations, and QSPR analysis[J]. J Am Ceram Soc, 2023, 106(11): 6510–6526.
- [15] SOULES T F, BUSBEY R F. The rheological properties and fracture of a molecular dynamic simulation of sodium silicate glass[J]. J Chem Phys, 1983, 78(10): 6307–6316.

- [16] YUAN F L, HUANG L P. Molecular dynamics simulation of amorphous silica under uniaxial tension: From bulk to nanowire[J]. *J Non Cryst Solids*, 2012, 358(24): 3481–3487.
- [17] LU X N, DENG L, GIN S, et al. Quantitative structure-property relationship (QSPR) analysis of ZrO_2 -containing soda-lime borosilicate glasses[J]. *J Phys Chem B*, 2019, 123(6): 1412–1422.
- [18] LUSVARDI G, MALAVASI G, TARSITANO F, et al. Quantitative structure-property relationships of potentially bioactive fluoro phospho-silicate glasses[J]. *J Phys Chem B*, 2009, 113(30): 10331–10338.
- [19] KALAHE J, LU X N, RILEY B J, et al. Structure-property relations of sodium iron phosphate nuclear waste glasses: Effects of iron redox ratio and glass composition[J]. *J Am Ceram Soc*, 2025, 108(2): e20174.
- [20] ANOOP KRISHNAN N M, MANGALATHU S, SMEDSKJAER M M, et al. Predicting the dissolution kinetics of silicate glasses using machine learning[J]. *J Non Cryst Solids*, 2018, 487: 37–45.
- [21] KRAUS M A, DRASS M. Artificial intelligence for structural glass engineering applications: Overview, case studies and future potentials[J]. *Glass Struct Eng*, 2020, 5(3): 247–285.
- [22] ABOUD A W, GUILLEN D P, CHRISTENSEN B A. Prediction of melter cold-cap topology from plenum temperatures with computational fluid dynamics and machine learning[J]. *Int J Ceramic Engine & Sci*, 2022, 4(4): 257–269.
- [23] CASSAR D R, DE CARVALHO A C P L F, ZANOTTO E D. Predicting glass transition temperatures using neural networks[J]. *Acta Mater*, 2018, 159: 249–256.
- [24] YANG K, XU X Y, YANG B, et al. Predicting the Young's modulus of silicate glasses using high-throughput molecular dynamics simulations and machine learning[J]. *Sci Rep*, 2019, 9(1): 8739.
- [25] BISHNOI S, SINGH S, RAVINDER R, et al. Predicting Young's modulus of glasses with sparse datasets using machine learning[J]. *J Non Cryst Solids*, 2019, 524: 119643.
- [26] LIU H, ZHANG T, ANOOP KRISHNAN N M, et al. Predicting the dissolution kinetics of silicate glasses by topology-informed machine learning[J]. *npj Mater Degrad*, 2019, 3: 32.
- [27] LILLINGTON J N P, GOÛT T L, HARRISON M T, et al. Predicting radioactive waste glass dissolution with machine learning[J]. *J Non Cryst Solids*, 2020, 533: 119852.
- [28] HAN T H, STONE-WEISS N, HUANG J, et al. Machine learning as a tool to design glasses with controlled dissolution for healthcare applications[J]. *Acta Biomater*, 2020, 107: 286–298.
- [29] CASSAR D R, SANTOS G G, ZANOTTO E D. Designing optical glasses by machine learning coupled with a genetic algorithm[J]. *Ceram Int*, 2021, 47(8): 10555–10564.
- [30] CASSAR D R, MASTELINI S M, BOTARI T, et al. Predicting and interpreting oxide glass properties by machine learning using large datasets[J]. *Ceram Int*, 2021, 47(17): 23958–23972.
- [31] ZAKI M, VENUGOPAL V, BHATTOO R, et al. Interpreting the optical properties of oxide glasses with machine learning and Shapely additive explanations[J]. *J Am Ceram Soc*, 2022, 105(6): 4046–4057.
- [32] WILKINSON C J, TRIVELPIECE C, HUST R, et al. Hybrid machine learning/physics-based approach for predicting oxide glass-forming ability[J]. *Acta Mater*, 2022, 222: 117432.
- [33] HU Y J, ZHAO G, ZHANG M F, et al. Predicting densities and elastic moduli of SiO_2 -based glasses by machine learning[J]. *npj Comput Mater*, 2020, 6: 25.
- [34] DU T, CHEN Z M, JOHANSEN S M, et al. Predicting stiffness and toughness of aluminosilicate glasses using an interpretable machine learning model[J]. *Eng Fract Mech*, 2025, 318: 110961.
- [35] LIU H, FU Z P, YANG K, et al. Machine learning for glass science and engineering: A review[J]. *J Non Cryst Solids*, 2021, 557: 119419.
- [36] PRIVEN A I, MAZURIN O V. Glass property databases: Their history, present state, and prospects for further development[J]. *Adv Mater Res*, 2008, 39–40: 147–152.
- [37] REKHSON S M. On the 90th birthday of oleg V. mazurin, scholar and teacher[J]. *J Non Cryst Solids*, 2017, 471: 483–489.
- [38] CASSAR D R. GlassNet: A multitask deep neural network for predicting many glass properties[J]. *Ceram Int*, 2023, 49(22): 36013–36024.
- [39] PLIMPTON S. Fast parallel algorithms for short-range molecular dynamics[J]. *J Comput Phys*, 1995, 117(1): 1–19.
- [40] DU J. Challenges in Molecular Dynamics Simulations of Multicomponent Oxide Glasses[C]//MASSOBRIO C, DU J, BERNASCONI Meds, *Molecular Dynamics Simulations of Disordered Materials*. Cham: Springer, 215: 157–180.
- [41] DENG L, DU J. Borosilicate and Boroaluminosilicate Glasses[C]//DU J, CORMACK AN eds. *Atomistic Simulations of Glasses: Fundamentals and Applications*. Cham: Wiley-VCH Verlag GmbH & Co. KGaA, 2022: 224–260.
- [42] DENG L, DU J C. Development of boron oxide potentials for computer simulations of multicomponent oxide glasses[J]. *J Am Ceram Soc*, 2019, 102(5): 2482–2505.
- [43] DU J C, CORMACK A N. The structure of erbium doped sodium silicate glasses[J]. *J Non Cryst Solids*, 2005, 351(27–29): 2263–2276.
- [44] DU J, CORMACK A N. The medium range structure of sodium silicate glasses: A molecular dynamics simulation[J]. *J Non Cryst Solids*, 2004, 349: 66–79.
- [45] DU J C, KOKOU L, RYSEL J L, et al. Structure of cerium phosphate glasses: Molecular dynamics simulation[J]. *J Am Ceram Soc*, 2011, 94(8): 2393–2401.
- [46] DENG L, DU J C. Development of effective empirical potentials for molecular dynamics simulations of the structures and properties of boroaluminosilicate glasses[J]. *J Non Cryst Solids*, 2016, 453: 177–194.
- [47] TIAN J, LI Y, ZHENG J J, et al. Insight into the structure and crystallization of $\text{SiO}_2\text{-Al}_2\text{O}_3\text{-P}_2\text{O}_5\text{-Na}_2\text{O-MgO-CaO}$ glass system with Mg-Ca substitution: A molecular dynamics study[J]. *J Non Cryst Solids*, 2024, 629: 122869.
- [48] HILL R. The elastic behaviour of a crystalline aggregate[J]. *Proc Phys Soc A*, 1952, 65(5): 349–354.
- [49] YAN J P, ZHANG Y J, WANG F M, et al. Electronegativity-based QSPR analysis for understanding structure-property relationships of glass materials[J]. 2025, 129(20): 5033–5046.
- [50] TANTARDINI C, OGANOV A R. Thermochemical electronegativities of the elements[J]. *Nat Commun*, 2021, 12(1): 2087.
- [51] PAULING L. The nature of the chemical bond[M]. Ithaca: Cornell University Press, 1960.
- [52] Epam/SciGlass, EPAM systems, 2025, (Accessed: Mar. 02, 2025). [EB/OL]. Available: <https://github.com/epam/SciGlass>.
- [53] GOODFELLOW I, BENGIO Y, COURVILLE A. Deep Learning[M]. Cambridge: The MIT Press, 2016.
- [54] SUBASI A. Practical Machine Learning for Data Analysis Using Python[M]. Amsterdam: Elsevier, 2020: 91–202.
- [55] NGUYEN Q H, LY H B, HO L S, et al. Influence of data splitting on performance of machine learning models in prediction of shear

- strength of soil[J]. Math Probl Eng, 2021, 2021: 4832864.
- [56] SIVAKUMAR M, PARTHASARATHY S, PADMAPRIYA T. Trade-off between training and testing ratio in machine learning for medical image processing[J]. PeerJ Comput Sci, 2024, 10: e2245.
- [57] TIEGEL M, HOSSEINABADI R, KUHN S, et al. Young's modulus, Vickers hardness and indentation fracture toughness of alumino silicate glasses[J]. Ceram Int, 2015, 41(6): 7267–7275.
- [58] FURUTANI K, OHKUBO T, DU J C, et al. Elucidating the atomic structures of the gel layer formed during aluminoborosilicate glass dissolution: An integrated experimental and simulation study[J]. J Phys Chem C, 2022, 126(18): 7999–8015.
- [59] YUE Y, TUHEEN M I, DU J. Borosilicate Glasses[M]//POMEROY M ed. Encyclopedia of Materials: Technical Ceramics and Glasses. Amsterdam: Elsevier, 2021: 519–539.
- [60] DENG L, DU J C. Effects of system size and cooling rate on the structure and properties of sodium borosilicate glasses from molecular dynamics simulations[J]. J Chem Phys, 2018, 148(2): 024504.
- [61] LUNDBERG S M, ERION G, CHEN H, et al. From local explanations to global understanding with explainable AI for trees[J]. Nat Mach Intell, 2020, 2(1): 56–67.
- [62] ALTMANN A, TOLOŞI L, SANDER O, et al. Permutation importance: A corrected feature importance measure[J]. Bioinformatics, 2010, 26(10): 1340–1347.
- [63] VARSHNEYA A K, MAURO J C. Composition-structure-property relationship principles[M]//VARSHNEYA A K, MAURO J C eds. Fundamentals of Inorganic Glasses. Amsterdam: Elsevier, 2013: 165–172.
- [64] KAZEMBAYKI M, YANG K, MAURO J C, et al. Decoupling of indentation modulus and hardness in silicate glasses: Evidence of a shear- to densification-dominated transition[J]. J Non Cryst Solids, 2021, 553: 120518.
- [65] SMEDSKJAER M M, MAURO J C, YOUNGMAN R E, et al. Topological principles of borosilicate glass chemistry[J]. J Non Cryst Solids, 2011, 115(44): 12930–12946.

Machine Learning Models for Predicting Mechanical Properties of Glass Based on Experiments, Molecular Dynamics Simulations and QSPR Analysis

YAN Jingping^{1,2}, WANG Feimei^{1,2}, LI Boyuan^{1,2}, DENG Lu¹, HU Lili¹

(1. Shanghai Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, Shanghai 201800, China; 2. University of Chinese Academy of Sciences, Beijing 100049, China)

Extended Abstract

Introduction Silicate glass is a versatile material widely used in architectural windows, electronics, optics, and nuclear waste storage. As science and technology advance, demand grows for high-performance glass materials. The mechanical properties, which directly impact service life, safety, and functionality, have attracted much attention. Conventional glass development relies on empirical experience and trial-and-error methods, which are slow and limited in scope. In addition, the amorphous nature and broad compositional range of glass materials make designing tailored materials challenging, requiring composition optimization, cost reduction, and process assurance. Recent advancements on computational modeling, machine learning (ML), and materials informatics offer promising solutions. ML models can predict glass properties and accelerate the discovery of high-performance compositions *via* leveraging high-dimensional compositional data and integrating computational and experimental insights. However, ML effectiveness depends on dataset quality—availability, completeness, consistency, accuracy, and representativeness. In this work, three ML-based models were constructed from three distinct sources, and datasets were sourced from SciGlass (density), molecular dynamics simulations (Young's modulus), and a QSPR model (hardness). This approach could provide an innovative dataset construction strategy for advancing glass composition design.

Methods In this study, ML models were developed to predict the density, Young's modulus, and hardness of silicate glasses. The dataset comprised 425 density values and 43 hardness values selected from the SciGlass database, 425 Young's modulus values obtained through MD simulations, and 50 hardness values predicted using a QSPR model. In the MD simulations, the Teter potential was employed, while the QSPR modeling incorporated descriptors was related to electronegativity. This work focused on six-component borosilicate glasses composed of SiO₂-B₂O₃-Al₂O₃-MgO-CaO-Na₂O. For initial data collection, a data preprocessing step was conducted, which included outlier removal, elimination of duplicate entries, and normalization of the compositional variables. A multilayer perceptron (MLP) neural network was used to model the relationship between glass composition and properties. The neural network architecture consisted of an input layer, several hidden layers, and an output layer, with weighted connections between neurons. The input features were the molar percentages of the oxide components, while the outputs were the corresponding physical properties. Non-linearity was introduced through the use of the rectified linear unit (ReLU) activation function in the hidden layers, enabling the model to capture complex nonlinear relationships. The output layer used a linear activation function to provide direct predictions of glass properties. To train and validate the models, the dataset was split into training and test sets in a ratio of 9:1. This partitioning ensured sufficient training data, while reserving a subset for independent evaluation of model generalization. Model performance was assessed using mean squared error (MSE) and the coefficient of determination (R^2). The trained models were subsequently used to predict the density, Young's modulus, and hardness of silicate glasses. Finally, nine

formulations and corresponding properties were predicted from the models and experimentally verified.

Results and discussion The compositions of the dataset are widely distributed. The density of the glass varies between 2.3 g/cm³ and 2.8 g/cm³, Young's modulus is primarily distributed between 82 GPa and 115 GPa, and hardness values range from 5.0 GPa to 7.0 GPa. The QSPR model related to hardness demonstrates a good linear correlation between hardness and the F_{net} descriptor, with the R^2 value of greater than 0.80, indicating good model fitting and predictive performance. In this study, three ML models are developed to predict the density, Young's modulus, and hardness of silicate glasses. The R^2 values of the models show clear differences. The density-related model shows a relatively low predictive accuracy, with an R^2 of only 0.30 on the test set, the model for Young's modulus achieves an R^2 of approximately 0.70 and the hardness-related model exhibits the maximum performance, with R^2 of > 0.95. However, the predictive accuracy of all three models was satisfactory. The maximum relative error for density prediction is only 2.67%. The predictions of Young's modulus are overestimated by approximately 10 GPa, mainly due to the limitations associated with empirical interatomic potentials. The prediction error for hardness ranges from approximately 0.05 GPa to 0.40 GPa. The SHAP analysis is further used to investigate the contributions of individual oxide components to the properties. The results reveal that SiO₂ and Al₂O₃ contents positively affect both Young's modulus and hardness, while Na₂O negatively affects the glass density. As a result, the use of MD simulations and QSPR method represents a promising alternative in the absence of abundant high-quality experimental data for model training.

Conclusions This study presented a comprehensive, multi-source modeling framework for predicting the mechanical properties of silicate glasses by integrating experimental data, MD simulations, and QSPR-based data augmentation. A total of 425 experimental density values, 425 MD-derived Young's modulus values, and 93 hardness values (43 experimental, 50 QSPR-predicted) were used to train three distinct ML models. Nine formulations were predicted and experimentally verified for three properties (*i.e.*, density, Young's modulus and hardness) based on the three models above. The experimental results exhibited small relative errors, indicating a reasonable agreement with model predictions. The results of Young's modulus took into account the effect of errors introduced by simulation. The data quality could become a critical factor affecting the accuracy of predictive models as using ML method. This work incorporated homologous glass property data obtained from MD simulations and employed QSPR method for rapid data augmentation based on experimentally acquired data. This strategy could address the limitations commonly encountered in conventional database construction, including inconsistent data quality and slow acquisition of experimental measurements. It could provide a novel approach toward establishing a data-driven paradigm for glass materials research.

Keywords silicate glass; machine learning; quantitative structure-property relationship analysis; mechanical properties