https://doi.org/10.12442/j.issn.1002-185X.20250385

机器学习辅助设计穿甲用TiZr系高熵合金的可解释模型

罗 浩，刘天宇，魏华男，高兴勇，范飞高，翟安琪，刘 卓

（陆军工程大学 弹药保障与安全性评估国家级实验教学示范中心，河北 石家庄 050000）

摘 要：兼具高硬度、抗压强度及高温相稳定性的TiZr系高熵合金，因其在穿甲战斗部中的潜在应用价值受到广泛关注。本研究引入机器学习方法辅助开展高熵合金成分设计，探索成分与性能之间的复杂关系，提升合金设计效率。考虑到穿甲作用对合金材料硬度的高要求，搭建了包含组元摩尔分数和5项关键描述符在内的15维特征数据集，集成157组合金硬度数据，采用8种典型机器学习模型（随机森林、K近邻、支持向量机等）进行训练，经网格搜索超参数优化及交叉验证，筛选出精度最优的预测模型，并利用可解释性框架（Shapley additive explanations，SHAP）揭示特征贡献机制。结果表明，XGBoost（extreme gradient boosting）模型预测质量最高，决定系数R^2达到0.73，平均绝对百分比误差为14.0%，混合焓（ΔH_{mix}）、Nb元素含量、原子尺寸失配（δ）是影响合金硬度的最主要因素，硬度调控需协同热力学稳定性、电子结构与几何尺寸3要素，特征间补偿效应是性能优化的关键，该方法为穿甲用TiZr系高熵合金提供了成分设计新范式，验证了机器学习在高效毁伤领域的巨大工程价值。

关键词：机器学习；高熵合金；硬度预测；可解释模型

中图法分类号：TG139 **文献标识码：**A **文章编号：**1002-185X（2026）10-2629-10

1 引 言

现代战争形态下，穿甲战斗部材料的性能直接影响武器装备的作战效能。传统穿甲材料如贫铀合金和钨合金，在面临新战法、新目标时所具备的高效毁伤性能不足。贫铀合金虽具有高密度和良好的穿甲性能，但其放射性污染会引发不可逆环境危害；钨基重合金则存在高温性能不佳、韧性与强度难以均衡等问题[1]，因此开发新型穿甲战斗部材料具有重要战略意义。在此背景下，高熵合金因其独特的多主元特性，展现出高熔点、高硬度以及良好的高温强度等一系列优异性能[2–4]。在广泛的元素组合空间中，通过合理设置高熵合金中各组元的比例，能够实现合金的微观结构调控，进而优化材料硬度等影响穿甲表现的关键性能。然而，目前高熵合金成分设计一定程度上仍然依赖大量的重复实验，这种传统的试错法研发周期长、成本高，难以满足快速发展的高效毁伤能力需求。

随着材料数据的不断积累和计算技术的飞速进步，机器学习在材料设计领域崭露头角[5]。它能够处理海量数据，挖掘数据中隐藏的复杂关系，快速预测材料性能，从而显著缩短新材料研发周期、降低研发成本。Lee等[6]提出的归因深度可解释方案（DISMAA）表明，AlCoCrFeMnNi体系中Al和Cr的浓度与硬度呈正相关，其余元素浓度与硬度变化呈负相关。在AlCoCrFeMnNi体系中，与硬度相关的量级依次为|Ni|>|Fe|>|Cr|>|Al|>|Co|>|Mn|。Wen等[7]结合机器学习代理模型和实验设计算法的设计策略，预测合成的42种Al-Co-Cr-Cu-Fe-Ni系合金有35种合金的硬度值高于训练数据集中的最佳值，17种合金的硬度比训练数据集中的最大硬度提高了10%以上。Klimenko等[8]基于支持向量分类器，预测并制备得到的$Al_1Cr_9Nb_{35}Ti_5V_{40}Zr_{10}$合金在室温和高温下具有较高的强度，同时在室温下具有良好的延展性。

TiZr系高熵合金尤其是难熔高熵合金具备高强度、较好的延展性、高温性能稳定、良好的释能特性等性能优势。Huang等[9]利用随机森林分类器搜索分析得到对Nb-Ti-Zr随机固溶体弹性性能影响较大的3个关键特征为价电子浓度、铌含量和平均体积模量。Si等[10]以Zr-Ti-Nb-O合金作为靶材进行了模型选择、特征筛选和特征重要性排序，其结论表明合金的强度和塑性受Λ参数、原子半径差异、剪切模量差异等因素的影响，模型对强度和塑性的预测误差可分别达到10%和16%以下。Li等[11]使

收稿日期：2025-10-23

作者简介：罗 浩，男，1999年生，博士，陆军工程大学弹药保障与安全性评估国家级实验教学示范中心，河北 石家庄 050000，E-mail：sdgxy@126.com

用XGBoost算法为合金构建了一个高度准确的强度-延展性预测模型，最终得到的$Ti_{35}V_{35}Nb_{10}Mo_{20}$合金表现出优异的整体性能，其特定屈服强度可达到149.55 kPa·m^3/kg，比传统等原子比合金高出10.97%，压缩应变超过50%。赵凤媛等[12]则选用人工神经网络等5种机器学习方法以5个相预测参数为特征、139组数据为训练集，分类预测高熵合金的固溶体、混合固溶体及金属间化合物，其中人工神经网络算法的预测准确率最高，为90.72%，结论中还指出混合焓对相形成的贡献最显著，较大的混合焓、较小的原子尺寸差和价电子浓度利于固溶体相形成。

尽管机器学习在高熵合金成分设计领域已展现出显著技术优势，但在军事工程应用场景中仍存在明显研究缺口。穿甲战斗部材料作为弹药装备中的关键要素，需要满足良好的硬度指标，但现有的算法模型对TiZr系难熔高熵合金的硬度预测还没有充分展开，且超70%的学习模型呈现强黑盒特性，无法解析各关键参数的耦合作用机制，更影响了预测模型的进一步优化。因此，开展机器学习辅助设计穿甲用难熔高熵合金的设计与解释具有重要的现实意义。

本研究旨在利用机器学习方法，建立穿甲战斗部合金材料硬度的预测模型，揭示合金硬度的主要影响因素以及因素间的影响关系，提出适用于设计穿甲用TiZr系高熵合金性能的协同调控策略，期望能为军事装备的材料性能提升提供新的解决方案。

2　研究策略与方法

2.1　数据集建立

高熵合金材料设计的长久发展已经为机器学习提供了良好的数据基础，然而由于研究目标与客观条件不同，这些高熵合金的性能受制备方法和热处理工艺等影响因素影响往往导致出现相同成分不同性能的情况。为了降低工艺因素对特征标准化的影响，作者的数据集参考了Detor[13]、Hosseini[14]的数据统计结果，删减其中非难熔高熵合金和经过附加工艺处理的高熵合金，其样本统计数据覆盖了大部分现有的TiZr系高熵合金成分。结合文献调研与前期研究，挑选了多元合金中可能影响材料硬度性能的5组关键特征参量，如表1所示，这些参量中包含了影响材料力学性能的关键参数。

其中，VEC_i和c_i分别指i_{th}元素的价电子数和原子浓度。n是高熵合金中的元素种类总数。x_i和r_i分别表示i_{th}元素的电负性和原子半径。平均的电负性和平均原子半径分别计算为$\bar{x}=\sum_{i=1}^{n}c_i x_i$和$\bar{r}=\sum_{i=1}^{n}c_i r_i$。$R$表示气体常数。$H_{ij}$是Takeuchi和Inoue[15]研究中的数据，代表i_{th}元素和j_{th}元素之间的混合焓。表2给出了计算经验参数时需要使用的元素参量，根据这些基础参数最终计算得到一个包含157类合金的硬度数据集，并对数据集进行了Z-score标准化，以减少特征尺度对预测结果的影响，提升模型收敛速度和预测稳定性。依据标准化后的数据集绘制出特征参数的分布情况可视化示意图如图1所示。

表1 关键特征描述符

Table 1 Key feature descriptors

Feature	Equation
Valence electron concentration, VEC	$\mathrm{VEC}=\sum_{i=1}^{n}c_i\,\mathrm{VEC}_i$
Difference of the Pauling negativities, Δx	$\Delta x=\sqrt{\sum_{i=1}^{n}c_i\,(x_i-\bar{x})^2}$
Atomic size difference, δ	$\delta=100\sqrt{\sum_{i=1}^{n}c_i\,(1-r_i/\bar{r})^2}$
Mixing entropy, ΔS_{mix}	$\Delta S_{mix}=-R\sum_{i=1}^{n}c_i\ln c_i$
Mixing enthalpy, ΔH_{mix}	$\Delta H_{mix}=\sum_{i=1,i<j}^{n}4H_{ij}c_i c_j$

同时应当说明的是，为实现利用尽可能少的特征达成较好的合金硬度预测，本实验所使用的经验参数是具备充分意义的。一方面这些特征在前人的工作中已经得到运用并取得一定的成效，其在包含足够信息的同时避免冗余特征带来的过拟合风险，使得模型能够被更容易理解和解释[16–18]。另一方面，参数本身与材料的宏观性能之间存在着相互作用的密切关系，其中：(1)价电子浓度处于较高水平(VEC≥8)时，一定程度上能促进面心立方结构形成，此时材料中位错滑移容易，延展性提升，如CoCrFeMnNi(VEC≈8.4)，材料通常具有高延展性和韧性[19]，而当合金处于中低价电子浓度(VEC≤6.87)时，材料倾向于形成体心立方结构[20]，如AlCoCrFeNi(VEC≈6.6)，此时材料具有高强度和硬度[21]；(2)Δx较小(<0.2)时，元素间化学键合均匀，可促进单相固溶体形成，Δx较大(>0.5)时，元素间键合选择性增强，可能导致金属间化

表 2 元素的基本参量

Table 2 Basic parameters of elements

Parameter	Hf	Mo	Nb	Re	Ru	Ta	Ti	W	Zr	V
VEC	4	6	5	7	8	5	4	6	4	5
Atomic radius/pm	144	130	134	128	125	134	132	130	145	122
Electronegativity/Pauling	1.33	2.16	1.60	1.90	2.20	1.50	1.54	2.36	1.33	1.63

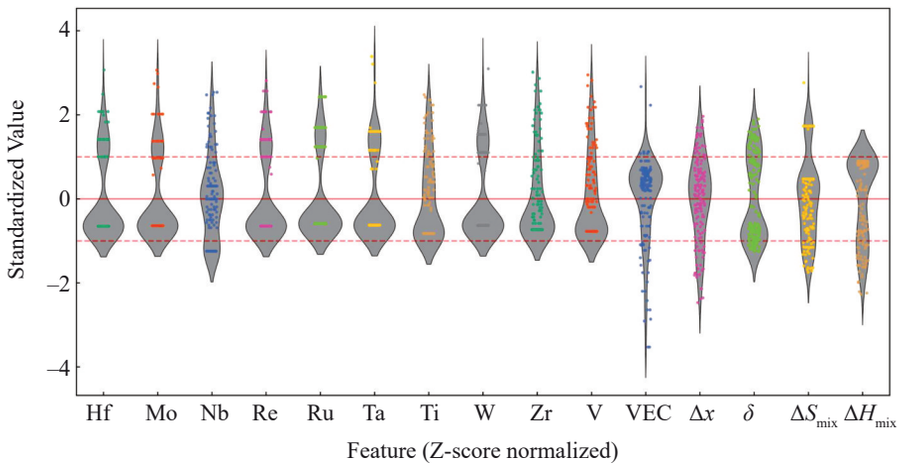


图 1　数据集的特征参数分布可视化图

Fig.1　Visualization diagram of the distribution of characteristic parameters of the dataset

合物或非晶相析出[22]；(3)δ最直接影响的是合金的晶格畸变效应，当δ位于适当的区间如5%~10%时，可最大化固溶强化效应，同时使合金保持一定的延展性[23]；(4)负的ΔH_{mix}有助于单相形成，提高均匀性；正的ΔH_{mix}可能促进析出相强化，但会降低合金的延展性。高的ΔS_{mix}带来高熵效应，能抑制扩散与相分离，稳定固溶体，提高合金的高温强度，因而最终在设计时往往要求ΔH_{mix}接近零或微负，同时$\Delta S_{mix}\geqslant 1.5R$以稳定单相固溶体[24]。

2.2　特征筛选

特征工程是提高机器学习模型性能的重要环节。在高熵合金性能预测中，从原始特征集合中提取和选择具有代表性的特征至关重要。通过主成分分析、相关性分析等方法，挖掘数据中的潜在特征，并去除冗余特征，降低特征维度，能有效提高模型的训练效率和泛化能力。如图 2 所示，为 TiZr 系高熵合金特征参数的皮尔逊相关系数（Pearson correlation coefficient，PCC）热图，用于分析判断特征参数之间的相关性强弱程度。一般认为|PCC|≥0.9 时，两个特征之间具有强相关性（线性相关），此时产生了机器学习预测中的冗余参量，需要对变量进行删除和保留；0.3<|PCC|<0.9 时，两个特征之间具有弱相关性；|PCC|≤0.3 时，两个特征之间没有相关性。由图可知，相关性最大的 3 组特征为元素 Hf 的摩尔浓度与Δx、元素 Ti 的摩尔浓度和ΔH_{mix}以及元素 V 的摩尔浓度和ΔH_{mix}，其 PCC 值分别为 0.69、0.69 和 0.67。根据上文确定的相关性判定阈值，这 3 组特征仅存在弱正相关关系，因此选择保留硬度数据集中的全部 13 个特征用于机器学习模型的训练。

2.3　机器学习算法

常规的机器学习预测问题通常可以分为回归事件和分类事件两类，两类场景对应了不同的算法模型、计算原理、评估效果和评价指标。表 3 中对比展示了 8 类模型的工作原理，针对数据集的特征，相应的适用场景存在一定的区分。利用这 8 类算法对 TiZr 系难熔高熵合金的硬度预测模型进行训练和对比，并选择平均绝对百分比误差（mean absolute percentage error，MAPE）和决定系数（R^2）作为模型预测质量的评价指标，其中 MAPE 越小，R^2越接近 1，代表机器学习模型效果越好，其计算方法如式（1）、（2）所示：

$$\mathrm{MAPE}=\frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_{\exp}-y_{\mathrm{pre}}}{y_{\exp}}\right| \tag{1}$$

$$r^2=\frac{1}{n-1}\frac{\sum_{i=1}^{n}(y_{\mathrm{pre}}-\overline{y_{\mathrm{pre}}})(y_{\exp}-\overline{y_{\exp}})}{s_{y_{\mathrm{pre}}}s_{y_{\exp}}} \tag{2}$$

其中，n为样本数量，$\overline{y_{\mathrm{pre}}}$和$\overline{y_{\exp}}$分别为预测值和真实值的均值，$s_{y_{\mathrm{pre}}}$和$s_{y_{\exp}}$分别为预测值和真实值的标准差。

3　结果与讨论

3.1　模型筛选

将标准化后的 15 个材料特征描述符作为输入参数，将材料硬度值作为目标输出变量，基于 Python 语言中的 Scikit-Learn 库调用了 7 种机器学习算法并使用 XGBoost 库导入了 XGBRegressor 模型，依次对以上 8 种算法进行训练。为了避免因单次测试集划分导致的误判，采用 10 折交叉验证法（10-fold）提升模型评估可信度。如图 3~4 所示，分别得到 8 种模型的 MAPE 和R^2得分。从计算结果可以看出，初始训练阶段，RandomForest、XGBoost 算法相比其余 6 类算法模型表现更好，其 MAPE 分别为 0.129、0.140，R^2分别为 0.68、0.58，即可初步解释目标变量 68% 和 58% 的方差，预测性能优于基线模型。以上结果表明，RandomForest 模型和 XGBoost 模型能相对更好地捕捉数

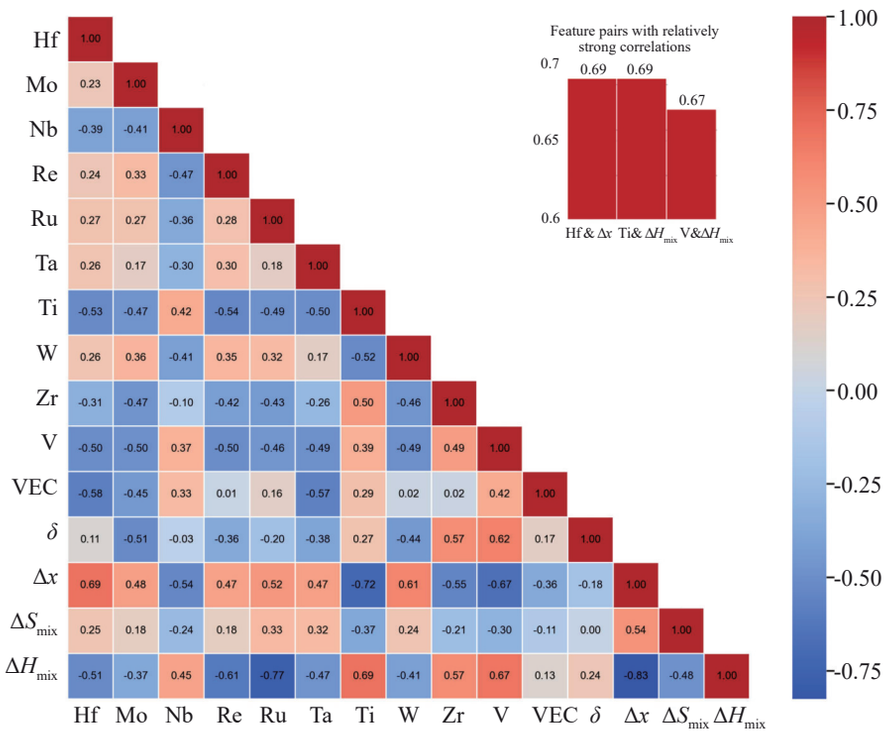


图2 特征参数皮尔逊相关系数热图

Fig.2 Heat map of Pearson correlation coefficient of characteristic parameters

表3 机器学习模型原理及适用场景

Table 3 Principles and applicable scenarios of machine learning models

Algorithm	Principle of operation	Applicable scenarios
Ridge	An L2 regularization term is added to the linear regression loss function to prevent overfitting	Data with many features and strong feature correlation
Lasso	Adding the L1 regularization term to the linear regression loss function yields sparse solutions	Feature selection or high-dimensional data
ElasticNet	The L1 and L2 regularization terms are combined	Scenarios where there is collinearity between features and feature selection is required
SVR	The data is mapped to a higher dimensional space by a kernel function to find the optimal hyperplane that makes the error within ε	Nonlinear relationships, small samples, or high-dimensional data
RandomForestRegressor	By voting with multiple decision trees, the variance is reduced and overfitting is reduced	Nonlinear relationships, high-dimensional data
DecisionTreeRegressor	The feature space is recursively partitioned to select the optimal splitting point	Data with high interpretability and nonlinear relationships
BayesianRidge	Based on a Bayesian framework, the marginal likelihood is maximized to estimate the parameters	Small sample data
KNeighborsRegressor	The target value is predicted based on the average of the nearest K samples	Local patterns are obvious, low-dimensional data
XGBRegressor	The loss function is optimized by gradient boosting tree, and regularization and pruning are added to control the complexity	Large scale data, high dimensional features

据中的非线性关系，且具有较高的预测稳定性，因而选用上述两个模型作为下一步训练和优化的目标算法模型。

3.2 超参数优化

作为一类集成学习模型，RandomForest 模型和

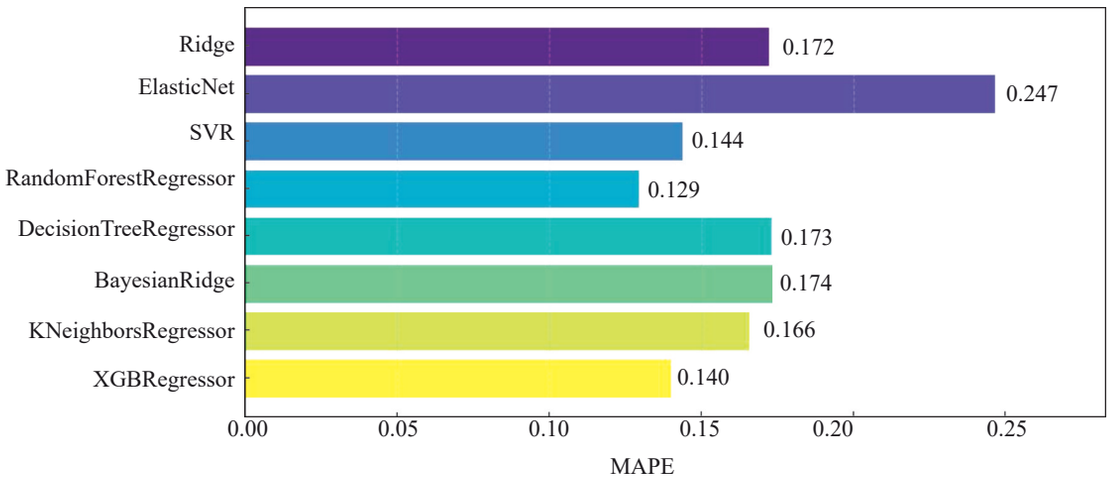


图 3　8 种算法的 MAPE 得分

Fig.3　MAPE scores for the eight algorithms

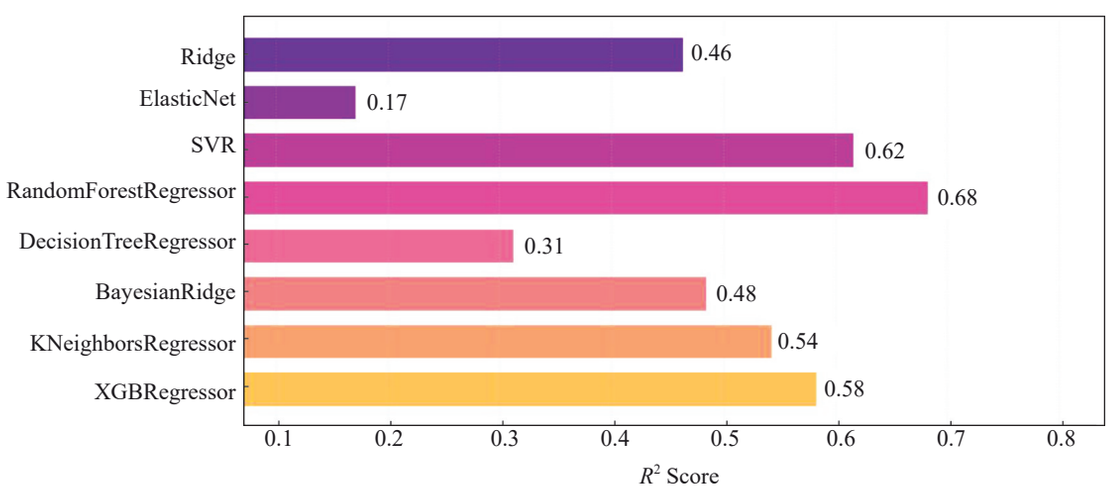


图 4　8 种算法的 R^2 得分

Fig.4　R^2 scores of the eight algorithms

XGBoost 模型在回归任务中能有效捕捉特征非线性关系、抑制过拟合，其性能高度依赖超参数选择。因而为提升模型的预测效果，进一步采用网格搜索（grid search）优化模型的主要超参数，并同样进行 10 折交叉验证。通过表 4 所示的搜索空间设置，计算得到了 RandomForest 算法的最佳参数组合为 n estimators=50，max depth=15，min samples split=5，max features=0.7，min samples leaf=2，对应验证集 R^2 得分为 0.69。XGBoost 算法的最佳参数组合为 colsample bytree=0.7，learning rate=0.2，max depth=3，n estimators=200，subsample=0.7，对应验证集 R^2 得分为 0.73，因此综合比较 XGBoost 模型能够较好满足计算需求。

3.3　性能预测

在预测前，首先通过绘制样本的散点图矩阵，将训练集与验证集的样本分布情况进行可视化对比，系统评估模型对特征关系的捕捉能力及泛化性能。如图 5 所示，训练集与验证集样本可视化展示了特征与真实目标变量的原始关联，两者散点分布特征趋于一致，一定程度上表明模型能够有效学习训练数据中的特征关系，并能准确迁移至验证数据，具备良好的泛化能力。最终在经超参数优化后的 XGBoost 模型预测下，得到图 6 所示的预测结果，预测硬度和合金实际硬度较为接近，基本都在对角线附近。结合模型拟合的置信区间（图中浅红色阴影），低硬度值区间阴影更宽，表明预测不确定性更高，在高实际值区间阴影较窄，预测更稳定，高硬度样本预测不确定性源于区间样本占比低（数据集中 21 个样本硬度值超过 10 GPa，占总样本数量 13%），该结果在同类预测研究中处于可接受范围，有待通过增加训练样本量来提升模型的预测稳定性。

表 4 超参数优化规范

Table 4 Hyperparameter optimization specification

Model	Parameter	Meaning
Random Forest Regressor	n estimators	The number of decision trees, with a search space of [50, 100, 150, 200]
	max depth	The maximum depth of a single decision tree has a search space of [10, 15, 20]
	min samples split	The minimum number of samples required for node splitting, with a search space of [5, 10, 15]
	min samples leaf	The minimum number of samples required for leaf nodes, with a search space of [2, 4, 6]
	max features	The maximum number of features considered when a single tree splits, with a search space of ['sqrt', 0.7]
XGBRegressor	n estimators	The number of trees, with a search space of [50, 100, 150, 200]
	learning rate	The learning rate and search space are [0.01, 0.05, 0.1, 0.2]
	max depth	The maximum depth of the tree has a search space of [3, 4, 5, 6]
	subsample	The sample sampling ratio and search space are [0.6, 0.7, 0.8, 0.9]
	colsample bytree	The feature sampling ratio and search space are [0.6, 0.7, 0.8, 0.9]

4 SHAP可解释性分析

在完成XGBoost模型的参数优化并确认其作为最佳模型后，需要深入理解模型的决策机制，以确保其预测结果的可解释性和可靠性。SHAP分析为此提供了强大的工具，其能够量化每个特征对模型输出的贡献，并揭示特征间的交互作用。通过SHAP分析，可以识别影响预测的关键因素，同时验证模型的逻辑是否符合领域知识，从而增强模型的可信度。图7为全局特征重要性图示，该图通过SHAP值量化了各特征对模型预测硬度的贡献程度，其中混合焓（ΔH_{mix}）、Nb元素含量、原子尺寸失配（δ）的SHAP值位列前三。其中，ΔH_{mix}反映了元素间结合能的强弱，负ΔH_{mix}有利于形成均匀固溶体，提升合金致密度与硬度，而正ΔH_{mix}可能促进析出相强化但需要考虑平衡延展性[11]；Nb元素的高权重则与难熔金属的晶格畸变效应相关，Nb原子半径（134 pm）与V（122 pm）、Zr（145 pm）的差异会引发晶格畸变，增大位错滑移阻力[14]，这种几何效应符合Koehler强化理论中畸变能与硬度的正相关关系[25]；原子尺寸失配（δ）则直接影响晶格畸变能，多数研究结果表明5%~10%的失配范围可最大化固溶强化效应且避免过度畸变导致脆性[26]，这些特征的重要性排序与高熵合金强化理论吻合。此外，V、Ru等元素的浓度差异反映原子尺寸与键合特性的协同作用[27]：V的小半径（122 pm）加剧畸变，但浓度过高时因价电子浓度（VEC）降低产生负面效应；Ru的高价电子（8e⁻）与高熔点（2607 ℃）则通过提升VEC与维持高温结构稳定性，间接强化室温硬度。

为直观展示各特征对合金硬度预测的全局影响，图8作出了预测模型的SHAP特征影响摘要图，图中以散点形式呈现各特征对硬度的影响方向与强度，其中横轴SHAP值的正负代表对硬度的促进或抑制作用，样本点颜色深浅则反映特征值大小。从图中混合焓ΔH_{mix}的散点分布可以看出，负ΔH_{mix}（深色点）多伴随正SHAP值，这是因为负混合焓促进元素间均匀键合，形成低自由能的单相固溶体，减少相界面缺陷，而原子尺寸差异δ的正向贡献则直接关联晶格畸变强化机制。在TiZr系难熔高熵合金中，较高的VEC意味着更多价电子参与形成离域电子云，增强原子间的静电相互作用，使晶格更难发生滑移或畸变，直接表现为硬度上升。从强化机制看，VEC的提升还会影响合金的电子云分布，促进s-d电子杂化，增强键合强度，同时高VEC体系更易形成无序固溶体结构，减少有序相析出，避免因相界面弱化导致的硬度下降[9]。电负性差异Δx的正贡献可能源于元素间电负性差促进化学短程有序[27]，增强电子云重叠程度，当Δx<0.2时，元素间电负性差异小，形成非极性共价键，原子排列无序度高，易形成单相固溶体，此时SHAP值多为负值，表明均匀固溶体对硬度的提升作用较弱；而Δx>0.5时，极性键占比增加，元素间键合选择性增强，易析出金属间化合物，通过第二相强化机制提升硬度，对应SHAP值向正值偏移，但可能伴随延展性下降。混合熵（ΔS_{mix}）的正向作用源于高熵效应：$\Delta S_{mix} \geqslant 1.5R$时，构型熵增抑制扩散与相分离，通过“熵稳定化”机制保持固溶体单相，避免脆性相析出[17]，这些规律与高熵合金相结构调控理论一致。特征交互机制方面，以混合焓（ΔH_{mix}）与原子尺寸差（δ）为例，当δ介于5%~10%时，ΔH_{mix}为正的样本SHAP贡献比ΔH_{mix}为负时低85%，符合“负ΔH_{mix}促进固溶体稳定，增强固溶强化”的热力学规律[24]。此外，通过计算交互特征的SHAP值协同贡献度，得到VEC与δ的协同贡献为两者SHAP值之和的96%，即是说两个特征共同作用时实际对硬度的贡献比各自独立贡献之和更小，这与样本中多数合金ΔH_{mix}>0（易析出相）有关，微观结构过度畸变可能引发析出相，析出相将在一定程度上抵消强化[20]。而后选取对硬度影响显著的一组核心特征Δx与ΔS_{mix}，当$\Delta S_{mix} \geqslant 1.5R$呈负贡献（降低硬度），因高熵效应引发的原子无序化会削弱位错滑移阻力，

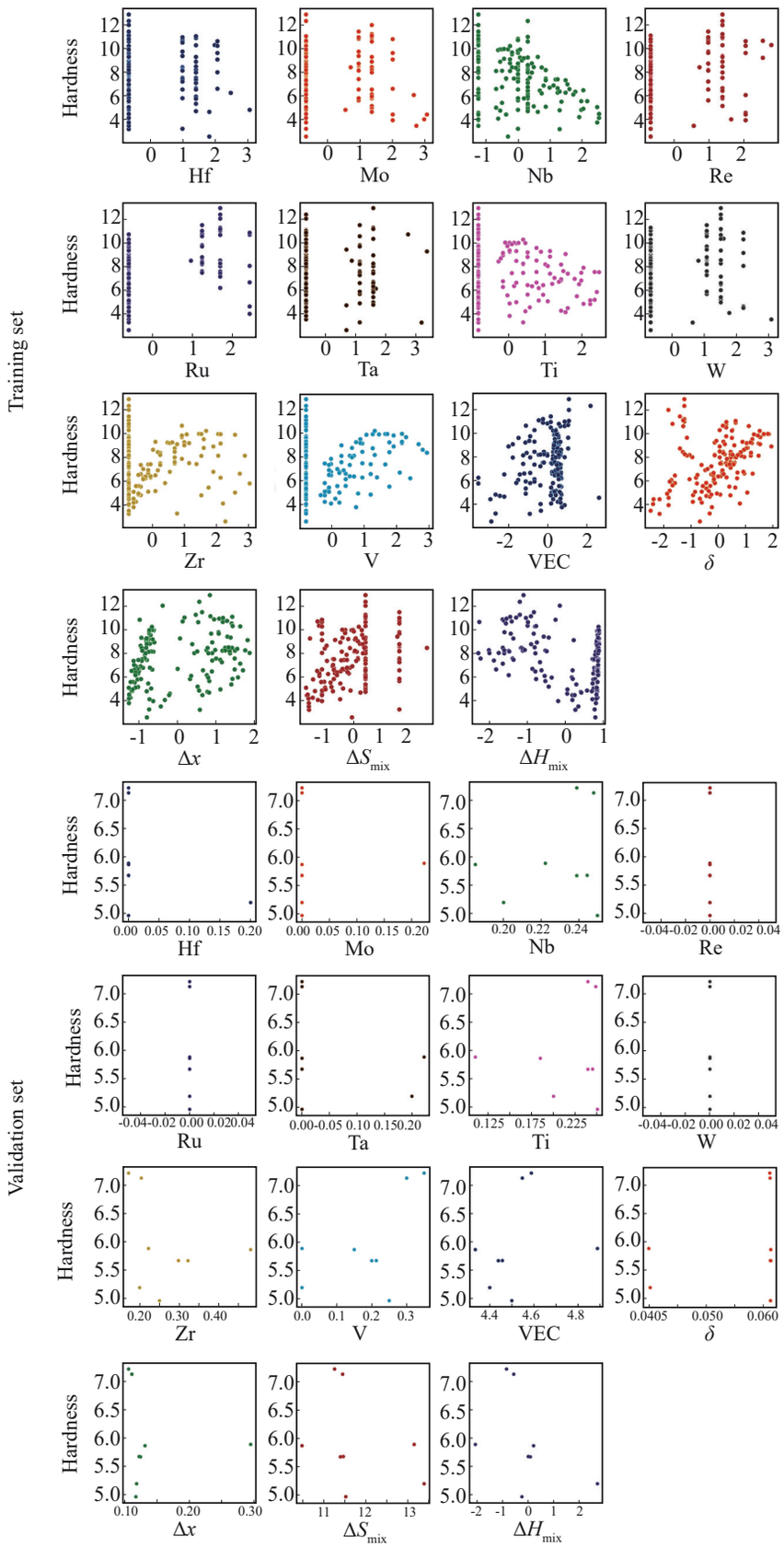


图 5　训练集和验证集的样本分布情况可视化对比

Fig.5　Visual comparison of the sample distribution between the training set and the validation set

可能导致整体流变应力的降低[28]，表现为材料软化，硬度下降。Δx<0.2（低电负性差）时，体系多为单相固溶体，高熵无序化最大化，且晶格畸变被分散，对硬度抑制作用最强，SHAP 值为–0.08；Δx>0.4（高电负性差）时，推动元素发生有

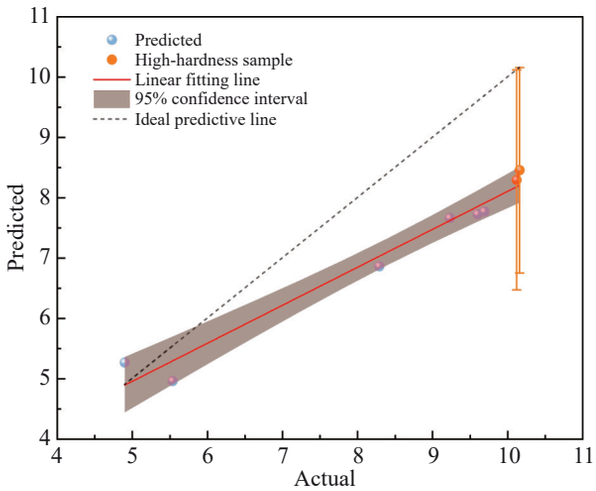


图 6 模型拟合结果

Fig.6 Model fitting result

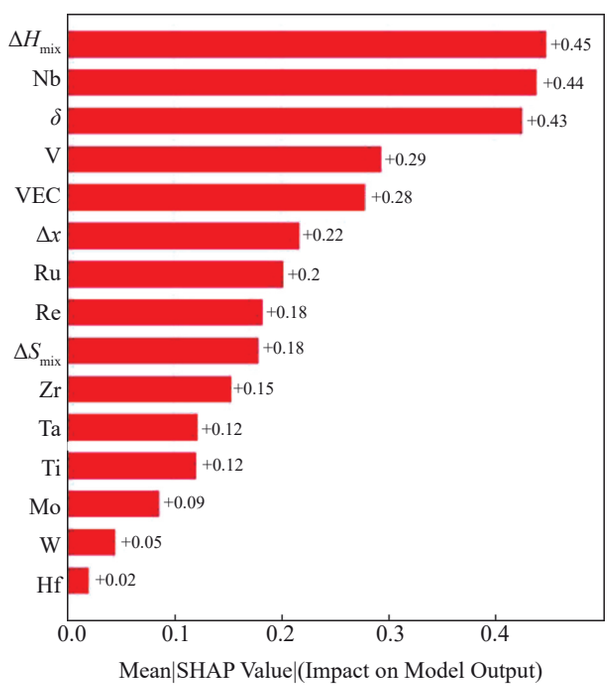


图 7 基于 SHAP 值的特征重要性排序

Fig.7 Feature importance ranking based on SHAP values

序化，形成化学短程有序或析出高硬度第二相（如ZrW_2）[29]，通过 Orowan 机制强化，抵消部分高熵抑制，SHAP 值升至–0.06。两个特征交互贡献值–0.13，量化了 Δx 对“相结构-高熵效应”的调控：单相中负协同更显著，表明控制 Δx 在 0.2~0.4，可平衡单相稳定与第二相强化，规避高熵负效应[30]。

Re 的价电子数为 7，原子半径 128 pm，除贡献高价电子提升 VEC 外，其与 Ti、Zr 的电负性差异（Re 为 1.90，Ti 为 1.50，Zr 为 1.33）可促进化学短程有序，增强电子云重叠程度，同时 Re 的加入能提高合金的混合焓绝对值，抑制无序固溶体向有序相转变，维持高畸变晶格结构，从而提高硬度[26]。三者的高浓度协同作用通过晶格畸变强

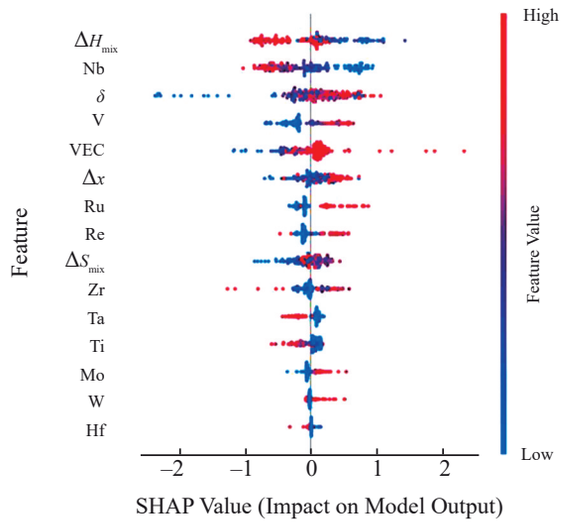


图 8 硬度预测模型 SHAP 特征影响摘要图

Fig.8 Summary diagram of the influence of SHAP features on hardness prediction models

化、固溶强化及电子结构调控，共同推动硬度提升。这与 XGBoost 模型的单个样本预测结果中每个特征对模型预测值的贡献度是一致的，如图 9 所示。

在该典型高硬度样本中，VEC（值为 4.333）以+1.64 的 SHAP 值成为核心正向贡献因素，其低 VEC 特性促进体心立方结构形成[20]，通过增强金属键合强度与晶格稳定性提升硬度；Ta（摩尔浓度为 0.17）贡献+0.21，其原子半径为 134 pm，通过影响原子半径差异 δ 引发适度晶格

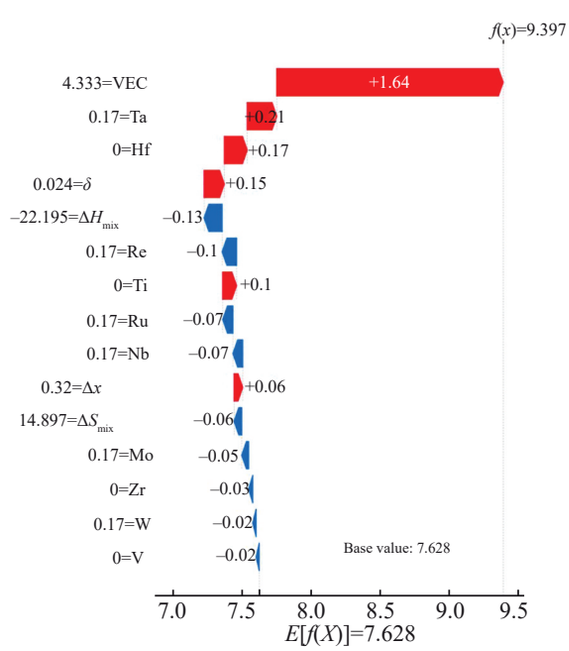


图 9 单一样本硬度预测的特征贡献度分布图

Fig.9 Distribution map of feature contribution degree for hardness prediction of a single sample

畸变，强化位错钉扎效应。而 ΔH_{mix}（值为–22.195）以–0.13 的 SHAP 值成为主要负向贡献因素，其过度负值可导致有序金属间化合物析出（如 Laves 相），削弱固溶强化效果[24]；Re/Ru（摩尔浓度均为 0.17）分别贡献–0.1 和–0.07，其中 Re 的高电负性引发元素偏析，Ru 在低 VEC 体系中未能充分发挥电子强化作用，且其小原子半径（125 pm）与基体差异可能导致过度畸变[31]。该样本整体呈现“强正向主导-弱负向抑制”的补偿平衡，即是说 VEC、Ta、δ 的协同增益（累计 SHAP+2.0）有效覆盖了 ΔH_{mix}、Re 等特征的抑制作用（累计 SHAP-0.3），验证了成分设计中需维持 VEC 和 δ 合理适配区间，并规避极负 ΔH_{mix} 与高电负性元素（如 Re）的叠加，凸显多特征协同优化在高熵合金设计中的必要性。

5　结 论

1）通过机器学习技术构建了 TiZr 系高熵合金硬度的可解释预测模型，证实 XGBoost 算法在捕捉成分-性能非线性关系中表现最优（验证集 R^2 达 0.73，平均绝对百分比误差为 14.0%），并借助 SHAP 分析揭示价电子浓度、混合焓等特征的协同作用机制，明确了电子结构调控与热力学参数平衡在合金设计中的核心地位。

2）构建了“特征重要性层级—全局影响规律—局部案例验证”的完整解释框架。研究发现不仅为合金成分设计提供了“优先调控核心特征、平衡元素协同效应”的量化依据，还通过训练集与验证集的对比确认了模型对实际合金体系的预测能力。

参考文献　　References

[1] Zhang Lang(章 浪), Zhao Fengpeng(赵丰鹏), Zhang Yuzhong(张钰忠) *et al*. *Explosion and Shock Waves*(爆炸与冲击)[J], 2025, 45(3): 107

[2] Tian Y S, Zhou W Z, Tan Q B *et al*. *Transactions of Nonferrous Metals Society of China*[J], 2022, 32(11): 3487

[3] Miracle D B, Senkov O N. *Acta Materialia*[J], 2017, 122: 448

[4] Tang Y, Wang R X, Xiao B *et al*. *Progress in Materials Science*[J], 2023, 135: 101090

[5] Zhang Tao(张 涛), Jiao Yongjun(焦拥军), Liu Zhenhai(刘振海) *et al*. *Rare Metal Materials and Engineering*(稀有金属材料与工程)[J], 2025, 54(7): 1882

[6] Lee C Y, Jui C Y, Yeh A C *et al*. *Journal of Alloys and Compounds*[J], 2024, 976: 173144

[7] Wen C, Zhang Y, Wang C X *et al*. *Acta Materialia*[J], 2019, 170: 109

[8] Klimenko D, Stepanov N, Ryltsev R *et al*. *Intermetallics*[J], 2024, 175: 108469

[9] Huang Z H, Chen H X G, Ye S B *et al*. *Materials Chemistry and Physics*[J], 2023, 299: 127537

[10] Si S P, Fan B J, Liu X W *et al*. *Materials & Design*[J], 2021, 206: 109777

[11] Li Y, Gong J J, Liang S L *et al*. *Journal of Materials Research and Technology*[J], 2025, 34: 1732

[12] Zhao Fengyuan(赵凤媛), Ye Yicong(叶益聪), Zhang Zhouran(张周然) *et al*. *Rare Metal Materials and Engineering*(稀有金属材料与工程)[J], 2023, 52(4): 1192

[13] Detor A, Oppenheimer S, Casey R *et al*. *Data in Brief*[J], 2022, 45: 108582

[14] Hosseini S, Vaghefi E, Lee C *et al*. *Materials Today Communications*[J], 2024, 40: 109607

[15] Takeuchi A, Inoue A. *Materials Transactions*[J], 2005, 46(12): 2817

[16] Ahmed I, Nan Q, Akhtar W *et al*. *Materials Today Communications*[J], 2025, 46: 112367

[17] Qiu M J, Huang P, Gu C *et al*. *Intermetallics*[J], 2025, 183: 108794

[18] Naqvi R B, Khan M I, Hakeem A S *et al*. *Journal of Materials Research and Technology*[J], 2025, 36: 2690

[19] Sun S J, Tian Y Z, Lin H R *et al*. *Materials Science and Engineering A*[J], 2019, 740–741: 336

[20] Yousefian C B, Magnussen J P, Lloyd M J *et al*. *Materialia*[J], 2025, 265: 116733

[21] Xiong T, Zheng S J, Pang J Y *et al*. *Scripta Materialia*[J], 2020, 186: 336

[22] Ji X. *International Journal of Cast Metals Research*[J], 2015, 28(4): 229

[23] Zhang X, Yan J, Chen Y H *et al*. *International Journal of Plasticity*[J], 2023, 166: 103643

[24] Dai D B, Xu T, Wei X *et al*. *Computational Materials Science*[J], 2020, 175: 109618

[25] Lee S H, Jung J G, Baik S I *et al*. *Materials Science and Engineering A*[J], 2021, 803: 140719

[26] Klimenko D, Stepanov N, Ryltsev R *et al*. *Intermetallics*[J], 2024, 175: 108469

[27] Ahmed I, Nan Q, Akhtar W *et al*. *Materials Today Communications*[J], 2025, 46: 112367

[28] Wang L, Ding J, Chen S S *et al*. *Nature Materials*[J], 2023, 22: 950

[29] Wu X L. *Journal of Materials Science & Technology*[J], 2023, 147: 189

[30] Cheng P, Li B Q, Sun Y Y *et al*. *Journal of Materials Research and Technology*[J], 2025, 38: 2168

[31] Yousefian C B, Magnussen J P, Lloyd M J *et al*. *Scripta Materialia*[J], 2025, 265: 116733

Machine Learning-Assisted Design of Interpretable Models for TiZr-Based High-Entropy Alloys Used in Armor-Piercing Applications

Luo Hao, Liu Tianyu, Wei Huanan, Gao Xingyong, Fan Feigao, Zhai Anqi, Liu Zhuo

(National Demonstration Center of Experimental Teaching for Ammunition Support and Safety Evaluation Education, Army Engineering University of PLA, Shijiazhuang 050000, China)

Abstract: TiZr-based refractory high-entropy alloys (RHEAs), known for their high hardness, compressive strength, and thermal phase stability, have garnered attention due to their potential application in armor-piercing warheads. This study introduced a machine learning (ML)-assisted approach to alloy design, aiming to uncover the complex relationships between composition and performance and to improve design efficiency. To address the critical requirement for hardness in armor-piercing applications, a 15-dimensional feature dataset was constructed from 157 experimental hardness data points, incorporating component molar fractions and five key descriptors. Eight ML models, including random forest, K-nearest neighbors, and support vector machines were trained, and XGBoost was identified as the most accurate through hyperparameter tuning via grid search and cross-validation. The SHAP (Shapley additive explanations) framework was applied to interpret feature contributions. Results indicate that the XGBoost model achieves the highest predictive performance (R^2=0.73, and the average absolute percentage error is 14.0%). The most influential factors affecting alloy hardness are mixing enthalpy (ΔH_{mix}), Nb content, and atomic size mismatch (δ). Effective hardness control relies on the synergistic regulation of thermodynamic stability, electronic structure, and geometric dimensions, where inter-feature compensation plays a critical role in optimizing overall performance.

Key words: machine learning; high-entropy alloy; hardness prediction; interpretable models

---

Corresponding author: Gao Xingyong, Ph. D., Professor, National Demonstration Center of Experimental Teaching for Ammunition Support and Safety Evaluation Education, Army Engineering University of PLA, Shijiazhuang 050000, P. R. China, E-mail: sdgxy@126.com