Research Article

Physics-informed machine learning framework integrating solid solution strengthening theory for accelerated hardness prediction in high-entropy alloys

Ao Gao[1], Yixiang Yan[1], Qiuling Tao[3], Jinxin Yu[4], Shengkun Xi[3,5], Geng Li[4,6], Xiaoyu Chong[2,*], Xingjun Liu[1,3,*]

[1]Faculty of Metallurgical and Energy Engineering, Kunming University of Science and Technology, Kunming 650093, Yunnan, China.
[2]Faculty of Materials Science and Engineering, Kunming University of Science and Technology, Kunming 650093, Yunnan, China.
[3]School of Materials Science and Engineering, Harbin Institute of Technology, Shenzhen 518055, Guangdong, China.
[4]China Rare Earth Group Research Institute, Shenzhen 518000, Guangdong, China.
[5]Department of Computer Science, Harbin Institute of Technology, Shenzhen 518055, Guangdong, China.
[6]Key Laboratory of Rare Earths, Ganjiang Innovation Academy, Chinese Academy of Sciences, Ganzhou 341000, Guangdong, China.

[*]**Correspondence to:** Prof. Xingjun Liu, Faculty of Metallurgical and Energy Engineering, Kunming University of Science and Technology, Kunming 650093, Yunnan, China. E-mail: xjliu@hit.edu.cn; Prof. Xiaoyu Chong, Faculty of Materials Science and Engineering, Kunming University of Science and Technology, Kunming 650093, Yunnan, China. E-mail: xiaoyuchong@kust.edu.cn

How to cite this article: Gao A, Yan Y, Tao Q, Yu J, Xi S, Li G, Chong X, Liu X. Physics-informed machine learning framework integrating solid solution strengthening theory for accelerated hardness prediction in high-entropy alloys. *J Mater Inf* 2026;6:[Accept]. http://dx.doi.org/10.20517/jmi.2026.16

Received: 6 April 2026 | **Revised:** 11 May 2026 | **Accepted:** 5 June 2026

Abstract

High-Entropy Alloys (HEAs) exhibit exceptional stability in extreme environments, yet their expansive design space presents a "curse of dimensionality" for traditional discovery methods. While Machine Learning (ML) offers a data-driven paradigm for material screening, the scarcity of experimental data often results in overfitting and limited physical interpretability. To address these challenges, this study proposes a Hybrid Physics-Informed Machine Learning (Hybrid PIML) framework for accelerated hardness prediction. By integrating classical solid solution strengthening theory with a Residual Learning Artificial Neural Network (ANN), the model explicitly embeds the physical coupling of shear modulus and lattice distortion ($G \cdot \delta_r^{2/3}$) as prior knowledge. This approach ensures predictions adhere to metallurgical principles while significantly outperforming benchmark algorithms, achieving a coefficient of determination (R^2) of 0.976 and reducing the Root Mean Square Error (RMSE) by approximately 43%. SHapley Additive exPlanations (SHAP) analysis confirms that physics-enhanced features dominate the decision-making process, validating the model's internalization of strengthening mechanisms. Furthermore, the research elucidates a phase-dependent non-linear correlation between hardness and yield strength, correcting the failure of the classical Tabor formula in work-hardening Face-Centered Cubic (FCC) alloys. Finally, a high-throughput virtual screening funnel based on this framework successfully identified optimized non-equiatomic candidates within the refractory Co-Cr-Ti-Mo-W system. This work establishes a precise, physically consistent pathway for inverse material design under data-constrained conditions.

Keywords: High-entropy alloys, physics-informed machine learning, hardness prediction, solid solution strengthening, high-throughput virtual screening, residual learning

INTRODUCTION

Since the Bronze Age, human alloy design strategies have long followed the "single principal element" paradigm, basing materials on one metal matrix and adding minor alloying elements to fine-tune properties. However, in the early 21st century, the concept of High-Entropy Alloys (HEAs) proposed by Yeh *et al.*[1] completely overturned this tradition. HEAs are typically defined as alloy systems composed of

five or more principal elements in equiatomic or near-equiatomic ratios[2-4]. This novel design philosophy has pushed the frontier of materials research and development from the corners of phase diagrams to their vast central regions[5,6].

The scientific value of HEAs stems primarily from their unique "four core effects," which endow them with mechanical properties distinct from traditional alloys: High Entropy Effect: The high configurational entropy generated by multi-component mixing significantly lowers the system's Gibbs free energy at high temperatures, suppressing the formation of brittle intermetallic compounds and promoting the stability of simple solid solution phases (e.g., FCC, BCC)[2,7]. Severe Lattice Distortion: Elements with different atomic sizes randomly occupy lattice sites, resulting in a severe, non-uniform elastic stress field within the lattice. This distortion significantly increases resistance to dislocation motion and is the primary physical origin of the high hardness and yield strength observed in HEAs[8-10]. Sluggish Diffusion: In complex multi-component environments, cooperative atomic diffusion becomes difficult, leading to elevated activation energy for lattice diffusion. This characteristic imparts HEAs with excellent high-temperature creep resistance and microstructural thermal stability[11-13]. Cocktail Effect: This is a composite synergistic effect, implying that the overall performance of the alloy is not merely a simple summation of its constituents' properties but includes additional gains arising from element-element interactions[12,14,15].

Despite the promising prospects of HEAs, their design faces a formidable challenge: the "curse of dimensionality." Assuming a selection of 30 common metal elements from the periodic table, the number of combinations for equiatomic alloys composed of just 5 elements exceeds 142,000. If non-equiatomic variations (e.g., the continuous variation of $Al_xCoCrFeNi$) and complex processing heat treatments are considered, the potential design space explodes exponentially, potentially reaching millions or even billions of combinations[16-18].

Traditional "trial-and-error" experimental workflows—encompassing melting, casting, cutting, polishing, and characterization—are lengthy and costly, making them wholly inadequate for such massive screening tasks[19-21]. Even high-throughput computational methods based on CALPHAD are often limited by a lack of accurate

multi-component thermodynamic databases, leading to reduced prediction accuracy in unknown compositional regions[22-24]. Therefore, there is an urgent need for an efficient screening tool to navigate this vast uncharted territory[21,25].

In recent years, ML, as a data-driven "fourth paradigm," has been widely applied in materials science. Algorithms such as Random Forest (RF), Support Vector Machines (SVM), and Gradient Boosting Decision Trees (LightGBM) have demonstrated powerful capabilities in establishing "composition-property" mappings[7,26].

However, in the field of HEAs, the application of ML faces a specific "Small Data" dilemma. Unlike image recognition fields with millions of labeled data points, high-quality data for HEA hardness or yield strength is often limited to a few hundred or a thousand entries[2,23]. This data scarcity introduces core problems:
Overfitting Risk: Models are prone to memorizing noise in the training set rather than learning the intrinsic laws of the material, leading to inflated performance on validation sets[23,27]. Poor Extrapolation: Pure data-driven models are essentially interpolators. When predicting compositions outside the coverage of training data, model predictions often fail completely, sometimes yielding results that violate physical common sense[28,29]. Lack of Interpretability: "Black box" models cannot explain why a specific composition possesses high hardness, making it difficult to guide new material design at the scientific level[30,31].

To overcome the aforementioned bottlenecks, PIML has emerged as a transformative approach. In recent years, interpretable machine learning has proven invaluable for uncovering complex, non-linear structure-property relationships that defy simple empirical assumptions, such as the counterintuitive role of defect morphology in additive manufacturing[32]. Concurrently, Physics-Informed Machine Learning (PIML) frameworks have emerged to solve the "small data" dilemma. For instance, Physics-Constrained Neural Networks (PCNNs) have been developed to integrate empirical physics equations into dual-loss functions, successfully predicting microstructure-dependent corrosion behaviors through inverse parameter identification[33]. However, while solid solution strengthening in HEAs provides a classical linear baseline, the extreme lattice distortions and complex interatomic interactions introduce significant non-linear deviations. Therefore, a novel

architecture is required to bridge macroscopic physical laws with microscopic non-linear complexities. The core idea of PIML is to embed domain knowledge into ML models, using known physical laws to constrain or guide the learning process, thereby reducing dependence on data volume and improving generalization ability[34,35].

In HEA hardness prediction, while Solid Solution Strengthening (SSS) theory provides a solid physical foundation, its classical application poses a significant challenge. In traditional dilute alloys, SSS is often treated as a relatively simple linear or analytical problem based on the "average effective medium" assumption (e.g., the classical Labusch model)[9,36]. However, the extreme local lattice distortions, multi-principal element synergistic interactions, Short-Range Order (SRO), and phase structural mutations inherently present in HEAs introduce massive non-linear deviations. Consequently, simple linear SSS models completely fail to capture these complex behaviors in expansive HEA compositional spaces. This fundamental limitation provides the critical motivation for introducing Machine Learning. A novel architecture is required to bridge the macroscopic linear physical laws with microscopic non-linear complexities. In our Hybrid PIML framework, classical analytical models (combining atomic size mismatch and modulus mismatch) are not discarded but utilized as an accurate "first-order linear approximation" or "physical baseline," leaving the neural network to capture the highly complex non-linear residuals.

Beyond constructing a high-precision hardness prediction model, this work addresses the disconnect in current research by bridging the gap between numerical predictions and the understanding of the non-linear yield strength-hardness relationship. Additionally, it demonstrates the utility of this model for accelerating material discovery in complex systems like Refractory HEAs. A Hybrid PIML framework is presented herein, which integrates solid solution strengthening theory by using a physics-based solid solution strengthening model as a baseline and training the ML model to learn the 'residual' between experimental and theoretical values[36]. This methodology seamlessly merges the robustness of physical principles with the flexibility of data-driven models, ensuring high-precision, physically consistent, and data-efficient predictions of HEA mechanical properties.

MATERIALS AND METHODS

Constructing a robust machine learning model requires, first and foremost, high-quality data infrastructure. Since raw compositional ratios cannot directly reflect internal physical interactions, Feature Engineering is essential to transform compositions into physical descriptors capable of describing phase stability and strengthening mechanisms.

Data acquisition and curation

Data quality and representativeness are the bedrock of high-precision models, especially for PIML frameworks relying on physical laws. The data must include not only chemical compositions but also key thermodynamic and physical parameters.

Dataset Compilation

A dataset comprising 205 experimentally verified HEA samples was compiled for this study. The data were primarily curated from peer-reviewed literature and public materials databases[2,37-39]. This dataset encompasses diverse alloy systems, predominantly including 3d Transition Metal Systems (e.g., Al-Co-Cr-Fe-Ni and its derivatives, spanning FCC/BCC/Multi-phase structures) and Refractory HEA Systems (e.g., W-Mo-Ta-Nb-V-Ti systems, typically displaying a BCC structure)[40,41]. Each data entry records the complete chemical composition (in atomic percentage) alongside the corresponding experimental Vickers Hardness (HV). To facilitate the training of the physics-informed model, the feature space was not limited to raw elemental ratios. Instead, 12 key physical and thermodynamic descriptors were calculated for each alloy utilizing CALPHAD methods and empirical formulas[8,23]. Core features include: Atomic Size Difference (δ_r), which quantifies the degree of lattice distortion and serves as a critical parameter for solid solution strengthening; Mixing Enthalpy (ΔH_{mix}) and Mixing Entropy (ΔS_{mix}), which characterize thermodynamic phase stability; Valence Electron Concentration (VEC), a key electronic parameter for predicting FCC/BCC phase stability; and Shear Modulus (G), reflecting the material's resistance to elastic deformation.

Data cleaning and preprocessing

To eliminate noise and ensure training stability, a strict cleaning process was executed: Missing & Outlier Treatment: Samples with missing physical parameters or hardness records were removed. Outliers significantly deviating from physical common sense were excluded based on the 3σ principle[12,42]. Physical Consistency Check: All physical descriptors were verified to ensure inputs complied with the basic assumptions of the solid solution strengthening theory.

Feature standardization

Given the immense dimensional differences among physical descriptors (e.g., ΔH is roughly $10^1\ kJ/mol$, while δ_r is 10^{-2}), direct input into a neural network would cause convergence difficulties or weight bias. Therefore, Z-Score Standardization was applied[3,43]:

$$x' = \frac{x - \mu}{\sigma} \tag{1}$$

where μ is the mean and σ is the standard deviation. This step is crucial for the ANN model and distance-based SHAP analysis, ensuring that the weight contribution of each physical feature in the loss function is of the same order of magnitude.

Dataset splitting

To objectively evaluate generalization, the cleaned dataset was randomly split into a Training Set (80%) and a Test Set (20%)[5,43]. To ensure reproducibility, the random seed was fixed (Random_State = 42). The training set was used for fitting the physical baseline and updating residual network parameters, while the test set was reserved strictly for final evaluation and was never involved in gradient updates.

Definition of model inputs and output

To ensure a rigorous mapping relationship within the framework, the inputs and output are explicitly defined as follows: Model Inputs (X_{aug}): A total of 17 physics-augmented descriptors are utilized. These include 14 fundamental chemical and thermodynamic features (e.g., VEC, ΔH_{mix}, ΔS_{mix}, δ_r, and shear modulus G) calculated via CALPHAD and empirical formulas, alongside 3 critical physics-based terms, with the LS_{term} ($G{\cdot}\delta_r^{2/3}$) being the primary prior. Model Output (Y): The

primary and sole direct output of the Hybrid PIML model is the Vickers Hardness (HV). Derived Property: The yield strength (σ_y) is not a direct output of the ML model but is subsequently derived from the predicted HV based on the phase-dependent physical correlation analysis established in this study.

Descriptor construction

To enable the machine learning model to truly "comprehend" the physical mechanisms underlying the materials, rather than merely memorizing elemental symbols, this study constructed a multi-dimensional descriptor system encompassing Geometric, Thermodynamic, Electronic Structure, and Physics-Based Strengthening factors. These descriptors not only determine the phase formation rules of High-Entropy Alloys (HEAs) but also serve as core input variables for the solid solution strengthening model.

Geometric & thermodynamic descriptors

As solid solution strengthening originates from the lattice elastic stress field induced by solute atoms, the atomic size mismatch (δ_r) is identified as the primary geometric parameter. In parallel, Mixing Enthalpy (ΔH_{mix}), Mixing Entropy (ΔS_{mix}), and the Ω parameter were selected to assess the thermodynamic stability of solid solution phases against intermetallic compounds. The corresponding calculation formulas are detailed below[23,44]:

Atomic Size Mismatch (δ_r):

$$\delta_r = \sqrt{\sum_{i=1}^{n} c_i \left(1 - \frac{r_i}{\bar{r}}\right)^2} \quad (2)$$

where c_i denotes the atomic fraction of element i, r_i is the atomic radius, and $\bar{r} = \sum c_i r_i$ represents the average atomic radius. The parameter δ_r directly quantifies the severity of lattice distortion and serves as a core input for the subsequent physics-based strengthening model.

Mixing Enthalpy (ΔH_{mix}) and Mixing Entropy (ΔS_{mix}):

$$\Delta H_{mix} = \sum_{i<j} 4\Delta H_{ij}^{mix} c_i c_j \quad (3)$$

$$\Delta S_{mix} = -R \sum_{i=1}^{n} c_i \ln c_i \quad (4)$$

where ΔH_{ij}^{mix} represents the binary mixing enthalpy between elements i and j, and R is the gas constant.

Thermodynamic Parameter (Ω):

$$\Omega = \frac{T_m \Delta S_{mix}}{|\Delta H_{mix}|} \quad (5)$$

This parameter integrates the competition between entropic and enthalpic effects and is typically used to determine the stability of solid solutions (where Ω >1.1 favors solid solution formation)[45,46].To visualize the coverage and phase stability trends of the dataset, the distribution of the 205 compiled HEA samples in the δ_r - ΔH_{mix} parameter space is illustrated in Figure 1.

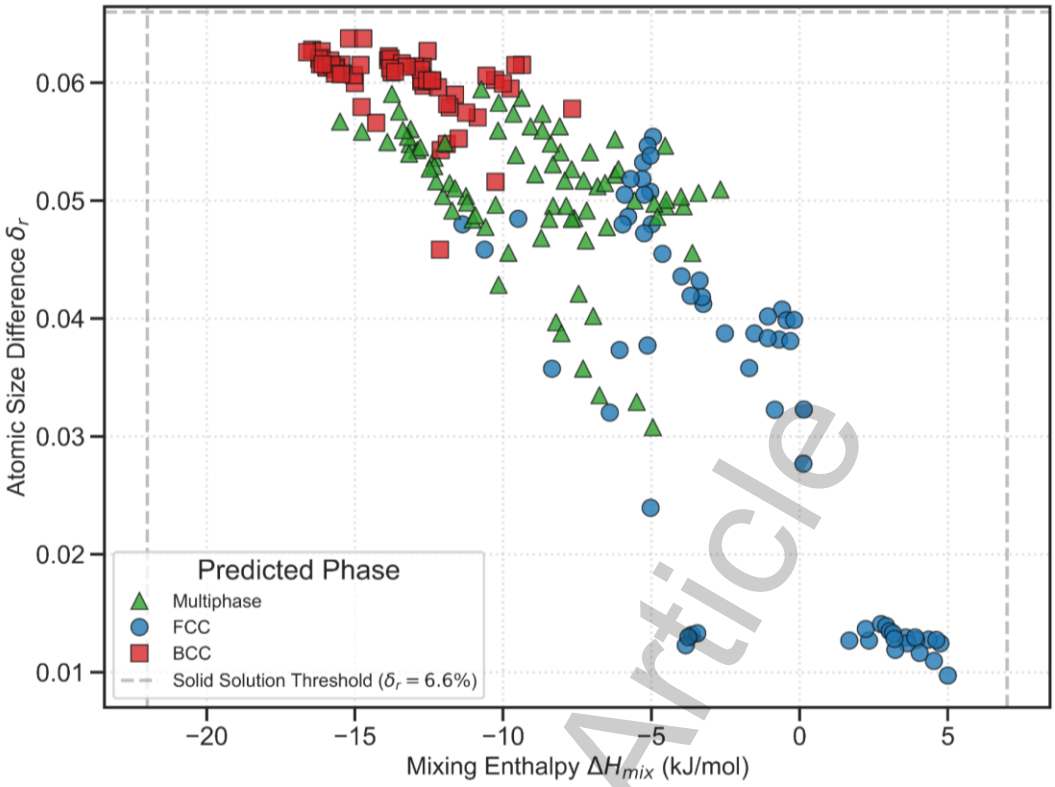

Figure 1. Distribution of 205 HEAs in the δ_r - ΔH_{mix} parameter space.

Electronic descriptors

Experimental studies indicate that the crystal structure of HEAs exhibits a strong dependence on the Valence Electron Concentration (VEC). Since different lattice structures possess varying numbers of slip systems, their hardness baselines differ vastly (BCC is generally harder than FCC)[3]. Therefore, introducing VEC assists the model in implicitly learning the "Phase Structure-Hardness" correlation[47]:

$$\mathrm{VEC}=\sum_{i=1}^{n} c_i\,(\mathrm{VEC})_i \tag{6}$$

Where $(\mathrm{VEC})_i$ is the number of valence electrons for the $i-th$ element.

Physics-based strengthening descriptors

This section highlights the key distinction between the proposed Hybrid PIML framework and traditional "composition-property" models. To allow the neural network to explicitly learn solid solution strengthening laws, this study calculated

mechanical modulus parameters directly related to dislocation motion resistance based on the solid solution strengthening theory:

Average Shear Modulus (G):

$$G=\sum_{i=1}^{n} c_i G_i \tag{7}$$

The shear modulus determines the magnitude of dislocation line tension and serves as the "substrate" for measuring material stiffness[31].

Physics-Based Strengthening Factor (LS_{term}): To provide direct physical guidance to the Hybrid PIML model, a theoretical strengthening term was pre-constructed. Based on the classical Labusch model and its subsequent extensions for multi-principal element alloys (e.g., by Senkov and Toda-Caraballo), the solid solution strengthening increment is highly dependent on the coupled effect of lattice distortion and modulus mismatch. Accordingly, we explicitly embedded this physical prior by formulating the composite feature:

$$LS_{term} \propto G \cdot \delta_r^{2/3} \tag{8}$$

This composite feature physically couples lattice distortion (δ_r) with matrix stiffness (G)[11,39], capturing the theoretical baseline of resistance to dislocation motion. In the subsequent model training, this feature is input into the neural network as the most critical prior knowledge, enabling the model to perform residual correction directly on the basis of physical laws, thereby substantially enhancing the physical consistency of predictions[9].

Feature selection strategy

Although a comprehensive descriptor system was established, an excessive number of input features—particularly those exhibiting high dimensionality and strong correlation—can lead to the 'curse of dimensionality' and increase the risk of overfitting. To balance the retention of physical information with the reduction of

model complexity, this study implemented a Physics-Guided Hybrid Feature Selection strategy.

Multicollinearity analysis

First, a Pearson correlation coefficient matrix (r) was employed to evaluate the pairwise correlations among all candidate descriptors, as illustrated in Figure 2. In HEAs, certain physical parameters (e.g., atomic radius difference δ_r and lattice mismatch parameters) may exhibit high mathematical collinearity ($|r| > 0.9$)[9,22]. The screening principle adopted was as follows: for highly correlated feature pairs, priority was given to retaining descriptors with clearer physical significance and tighter linkage to the solid solution strengthening theory[23]. For instance, when the simple atomic size difference (δ_r) was found to be highly correlated with the theoretical strengthening factor (LS_{term}), the latter was retained because LS_{term} incorporates shear modulus information, thereby better characterizing the comprehensive resistance to dislocation slip.

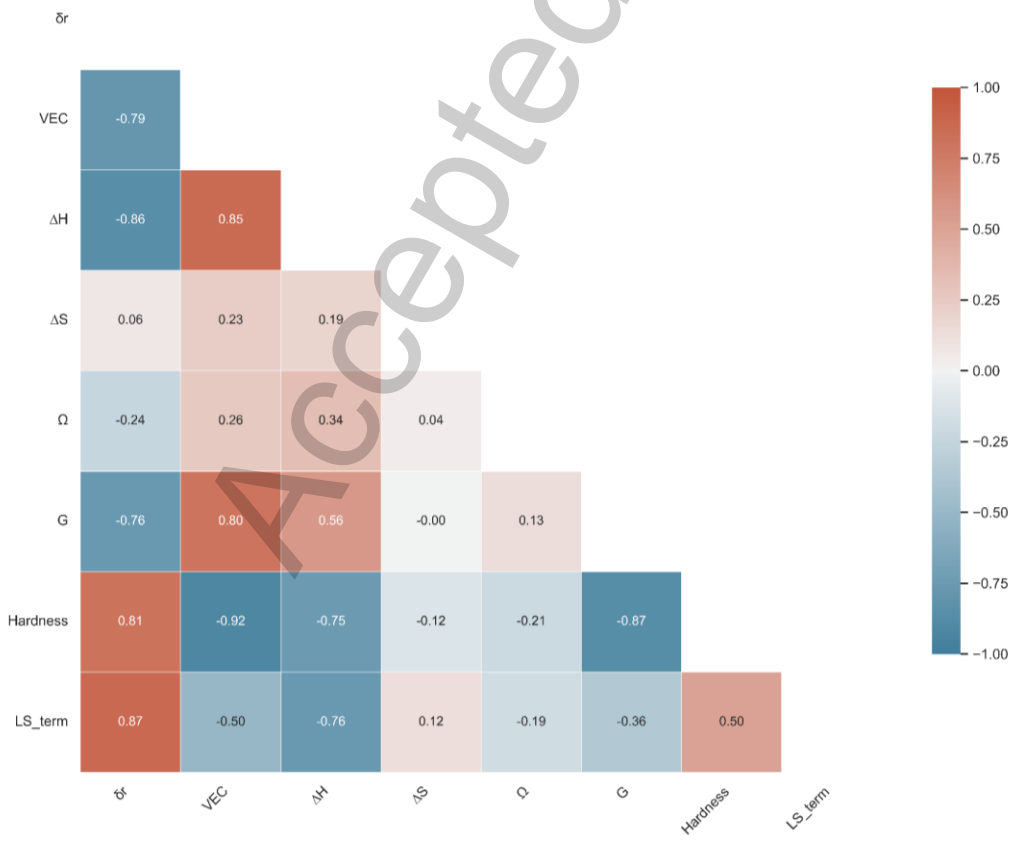

Figure 2. Pearson correlation coefficient heatmap of physical descriptors.

Tree-based recursive feature elimination

Following the removal of redundant features, a Recursive Feature Elimination (RFE) algorithm based on Random Forest was applied[19,44]. Utilizing the "Gini Impurity" metric inherent to tree models, the remaining features were preliminarily ranked by importance. Noise features making negligible contributions to hardness prediction (Importance < 0.01)—such as the atomic number features of certain non-critical elements—were eliminated.

Physics prior retention

Representing the crux of the feature selection strategy, this step introduces domain knowledge intervention, distinguishing the approach from purely data-driven methods[37]. Irrespective of RFE rankings, the retention of three core physical descriptors was enforced for the final model input: $LS_{term}\left(G \cdot \delta_r^{2/3}\right)$ as the theoretical baseline for solid solution strengthening[39]; VEC as the indicator of phase structural stability; and ΔH_{mix} as the thermodynamic measure of interatomic binding forces. This three-step process reduced the high-dimensional feature space to 17 key descriptors. Consequently, the computational load for ANN training was minimized while ensuring the input information retained clear metallurgical meaning, a prerequisite for the subsequent SHAP mechanistic analysis.

Physics-informed machine learning methodology

The PIML framework proposed in this study aims to address the generalization challenges faced by pure data-driven models on "small-sample" HEA data, facilitating the high-throughput development of novel materials[48]. Unlike traditional "black-box" learning, this framework adopts a "Theory-Guided + Data-Corrected" hybrid residual learning architecture. This architecture explicitly injects prior knowledge (Inductive Bias) from physical metallurgy into the machine learning process, ensuring that the model's predictions not only align with physical common sense but also achieve extremely high data-fitting accuracy.

Theoretical baseline: solid solution strengthening model

As the physical cornerstone of the PIML framework, an analytical baseline model based on classical solid solution strengthening theory was first constructed. The

theoretical foundation originates from the Labusch model[39], which describes the strengthening effect arising from the elastic interaction between dislocations and solute atoms in binary alloys. For multi-principal element systems such as HEAs, recent studies have extended this framework by incorporating the coupled effects of lattice distortion and modulus mismatch[9,36,37]. Following the approach established by Wen *et al.*[37] and Huang *et al.*[9], the theoretical hardness increment (H_{phy}) can be approximated by a simplified linear expression incorporating the atomic size mismatch (δ_r) and shear modulus (G):

$$H_{phy}=A\cdot G\cdot\delta_r^{2/3}+B \tag{9}$$

where A and B are fitting constants obtained by linear regression against the experimental hardness data in the training set, representing an average effective-medium approximation of the strengthening behavior. This baseline model thus provides a deterministic first-order prediction. However, the parameters A and B are empirical in nature and inherently embed the simplifying assumptions of a homogeneous, single-phase matrix; they do not capture sample-to-sample deviations arising from phase transitions, short-range ordering, or the differential work-hardening behavior of FCC versus BCC matrices. The uncertainty introduced by these simplifications is the primary source of prediction error for the baseline model (as seen in its high RMSE of 212.61 HV in Table 1). Consequently, the residual term F_{ML} in the Hybrid PIML framework is specifically tasked with learning and correcting these complex, physically un-modeled deviations.

Hybrid architecture: residual learning strategy

To correct the deviations of the physical model, this study introduced a Residual Learning mechanism. In this residual architecture, the primary objective (Output) is the Vickers Hardness (HV). The ANN is specifically tasked with learning the non-linear residual correction term F_{ML}—representing the deviation between H_{exp} and the theoretical baseline H_{phy} —thereby enabling precise total hardness prediction[34,36]. The prediction logic of the entire framework is as follows:

$$H_{final}=H_{phy}+F_{ML}(X_{aug}) \tag{10}$$

Where:

H_{phy}: The physical baseline value obtained by linear regression fitting of the physics-based baseline formula[9,39]. F_{ML}: The non-linear residual correction term constructed by an Artificial Neural Network (ANN). X_{aug}: The physics-augmented input feature vector.

The core advantage of this architecture lies in "divide and conquer" (Figure 3): allowing the physical formula to handle macroscopic linear laws while enabling the neural network to focus on capturing microscopic, non-linear complex interactions (such as phase structural mutations caused by VEC). This significantly reduces the learning difficulty for the neural network, allowing it to converge quickly and avoid overfitting even with small-sample data[2,27]. Unlike conventional Physics-Constrained Neural Networks (PCNNs) that embed empirical physical equations directly into a dual-loss function to regularize network weights[33], our framework adopts a structural "theory-guided + data-corrected" Residual Learning architecture. By explicitly decoupling the macroscopic physical baseline (the Labusch-type model) from the machine learning residual term (F_{ML}), the neural network is unburdened from learning the fundamental linear laws from scratch. Instead, it focuses exclusively on capturing the complex non-linear phase-dependent deviations. This structural decoupling not only enhances training efficiency and prevents optimization conflicts common in complex multi-loss formulations but also provides a more transparent mechanism for tracking physical consistency.

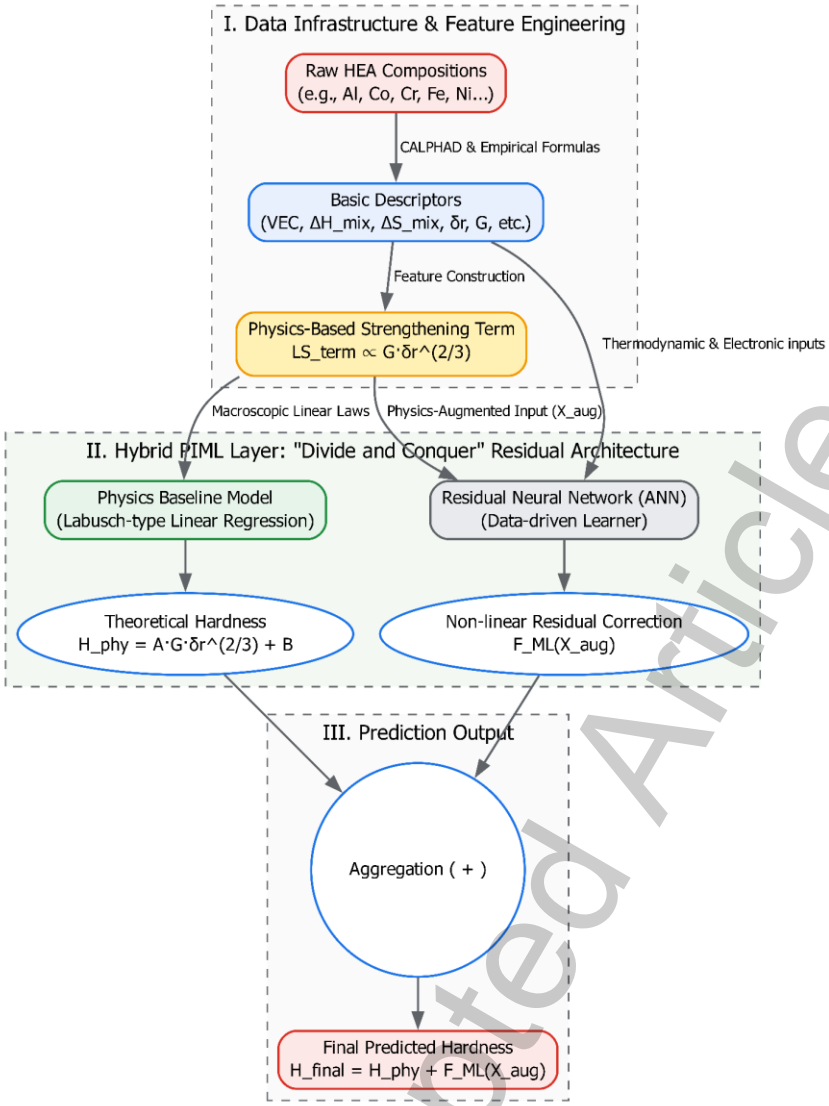

Figure 3. Architecture of the proposed Hybrid PIML framework.

Physics-augmented input space

To further enhance the model's sensitivity to physical laws, the training of the residual network (F_{ML}) moved beyond simple raw elemental ratios to construct a physics-augmented input space. Supplementing the fundamental descriptors previously established, this study incorporated the physics-based strengthening term ($LS_{term}=G \cdot \delta_r^{2/3}$) as an independent, explicit feature within the neural network[36]. This integration implies that the network not only perceives 'which elements are present' but also directly 'observes' the theoretically predicted strengthening potential. Such Feature-level Fusion enables the model to keenly identify outliers that violate physical laws and dynamically adjust theoretical values based on other thermodynamic parameters (e.g., automatically lowering hardness predictions to

account for work-hardening effects when VEC indicates an FCC phase)[3,40].

Algorithm selection

To construct the optimal PIML framework, this study first conducted extensive benchmarking on various mainstream regression algorithms. The goal was to screen for the algorithm best suited to capture the non-linear mapping relationship between HEA composition and properties as the "Residual Learner" of the framework[7,42]. Five representative classes of models were selected for comparative study: Physical Baseline Model (Labusch-type): Based solely on the classical solid solution strengthening formula (in linear regression form), serving as a benchmark for evaluating pure theoretical prediction capability. Ensemble Learning Models: Including Random Forest (RF) and Gradient Boosting Decision Trees (XGBoost/LightGBM), which excel in handling tabular data and feature interactions. Kernel Methods: Support Vector Regression (SVR), utilizing the Radial Basis Function (RBF) kernel to handle non-linear relationships. Deep Learning Models: Multi-Layer Perceptron (ANN/MLP), possessing powerful universal function approximation capabilities. Hybrid PIML Framework: The architecture proposed in this study, which explicitly augments the ANN input with the physics-based strengthening term. The specific performance of each model is presented in Table 1. To ensure absolute transparency and mathematical rigor for the results presented in Table 1, a strict machine learning evaluation protocol was utilized. The dataset was randomly partitioned into an 80% training set and a 20% unseen test set using a fixed random seed ($random\_state = 42$) to guarantee reproducibility. For all ML algorithms (including RF, SVR, XGBoost, and the standard ANN), hyperparameter tuning was conducted utilizing 5-fold cross-validation strictly within the training set to prevent data leakage. The performance metrics reported in Table 1—namely the Coefficient of Determination (R^2), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE)—were calculated exclusively based on the predictions made on the unseen 20% test set. This strict protocol guarantees that the reported values reflect the true generalization and extrapolation capabilities of the models on out-of-sample HEA compositions, rather than mere training data memorization. First, the fundamental limitations of treating HEA solid solution strengthening as a simple linear problem were unequivocally exposed by the pure physical model (Labusch Model). Its R^2 was an abysmal 0.213, with an RMSE as high as 212.61 HV. This

catastrophic empirical failure directly refutes the assumption that HEA strengthening can be resolved by simple linear combinations of lattice distortion and shear modulus. It demonstrates that the hardness mechanism of high-entropy alloys is governed by profound non-linearities, such as structural phase transitions induced by Valence Electron Concentration (VEC) and complex strain-hardening behaviors. This directly motivates the absolute necessity of introducing deep data-driven methods[9,39]. By utilizing the Artificial Neural Network (ANN) as a residual learner to complement the linear physical baseline[27,49], the Hybrid PIML model successfully captured these deep non-linear laws, jumping to an R^2 of 0.976 and dropping the RMSE to 37.01 HV[36].

Table 1. Performance comparison of different hardness prediction models

Model	R^2 Score	RMSE (HV)	MAE (HV)
Physical Baseline Model (Labusch-type)	0.2134	212.61	166.61
Random Forest	0.9257	65.36	38.86
SVR	0.9274	64.60	41.37
XGBoost/LightGBM	0.9452	56.13	37.01
ANN (Standard MLP)	0.9702	39.83	26.90
Hybrid PIML (Physics-Augmented)	0.9762	37.01	25.74

RESULTS AND DISCUSSION

Model performance benchmarking

This paper subjected all models to rigorous testing using the same training and test sets (80:20 split, $random\_state = 42$). The quantitative performance metrics for each model are summarized in Table 1, while the corresponding parity plots comparing predicted versus experimental hardness are illustrated in Figure 4. The Hybrid PIML model achieved the best performance among all models by introducing physics-enhanced features.

The Hybrid PIML model reduced the RMSE to 37 HV, significantly outperforming both the pure theoretical model and pure ML models. This confirms the effectiveness

of the residual learning strategy: the physics-based baseline model handles the primary solid solution strengthening contribution, greatly reducing the learning difficulty for the ML model, allowing it to focus solely on learning the complex non-linear effects not covered by the theory (such as phase mutation). An R^2 approaching 0.98 indicates that the framework explains almost all sources of variance in the dataset.

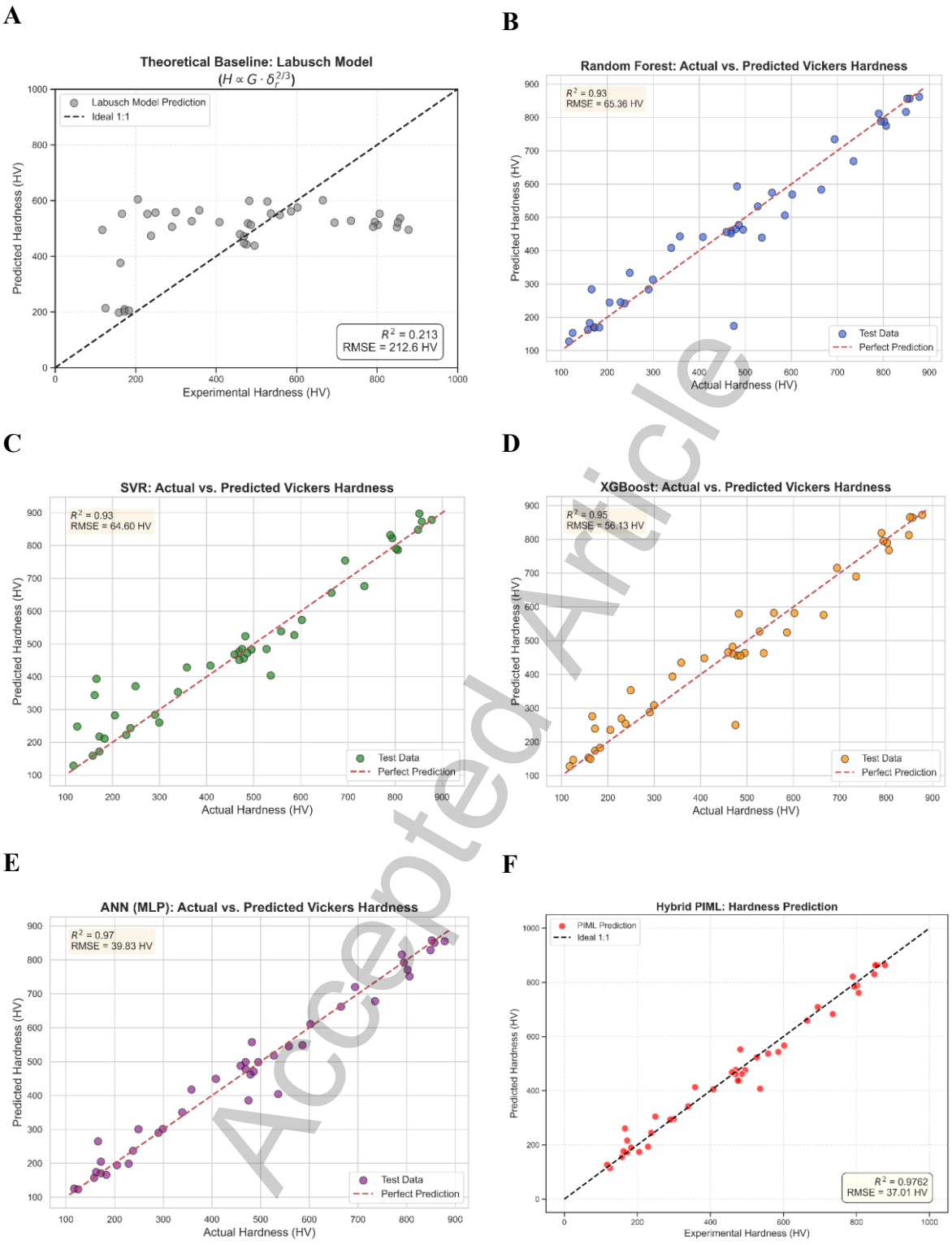

Figure 4. Comparison of experimental vs. predicted Vickers hardness across different models. (A)Physical Baseline Model, (B)Random Forest, (C)SVR, (D)XGBoost, (E)ANN and (F) the proposed Hybrid PIML model. The dashed line indicates perfect prediction (y=x). The Hybrid PIML model demonstrates significantly tighter scattering around the diagonal, validating its superior accuracy.

Mechanistic interpretability: SHAP analysis

To open the "black box" of the machine learning model and verify whether it has learned hardness mechanisms consistent with physical metallurgy principles, this study employed the SHAP method to conduct an in-depth interpretability analysis of the XGBoost, Standard ANN, and Hybrid PIML models. Figure 5, Figure 6, and Figure 7 display the SHAP Summary Plots, SHAP Feature Importance Bar Plots, and Univariate Dependence Plots for the three models on the test set.

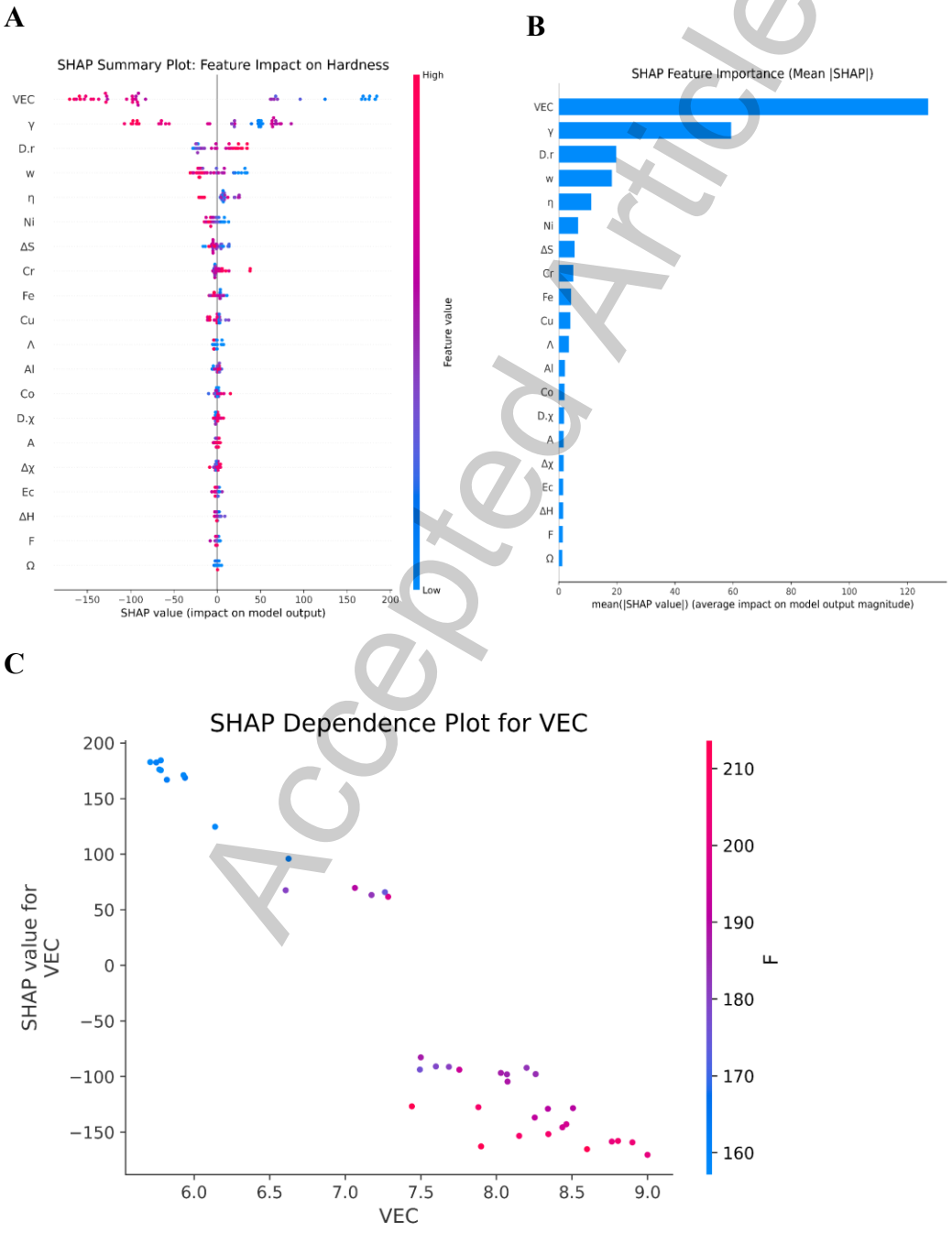

Figure 5. Mechanistic interpretability analysis of the baseline XGBoost model. (A)

SHAP summary plot and (B) feature importance ranking derived from the standard XGBoost algorithm. (C) SHAP dependence plot for the top-ranked feature.

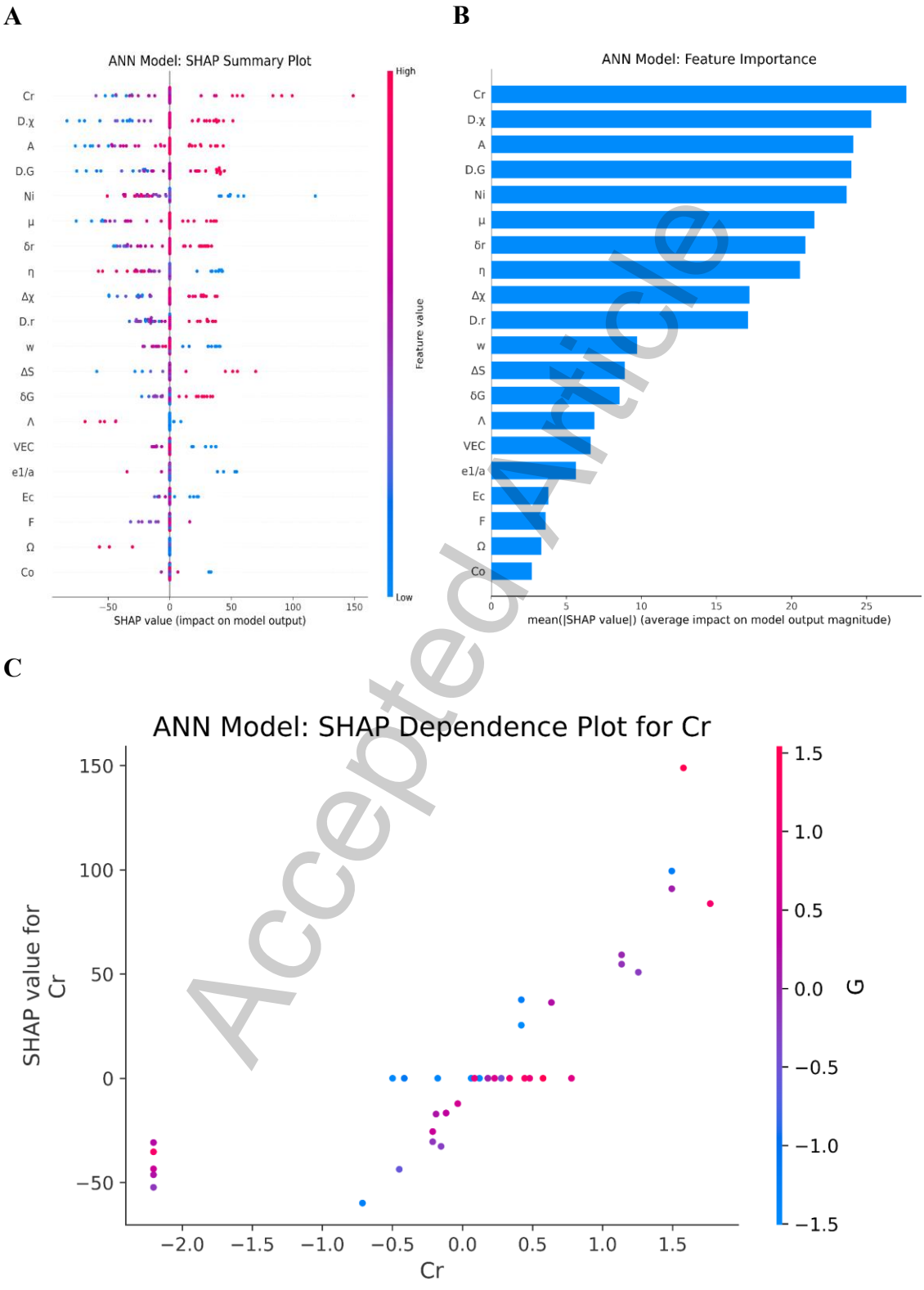

Figure 6. Baseline interpretability analysis of the standard Artificial Neural Network

(ANN) model. (A) SHAP summary plot illustrating the feature impact distribution for the pure data-driven ANN. (B) Global feature importance ranking. While the model identifies key descriptors, the ranking relies solely on statistical correlations within the training data. (C) SHAP dependence plot for the top-ranked feature.

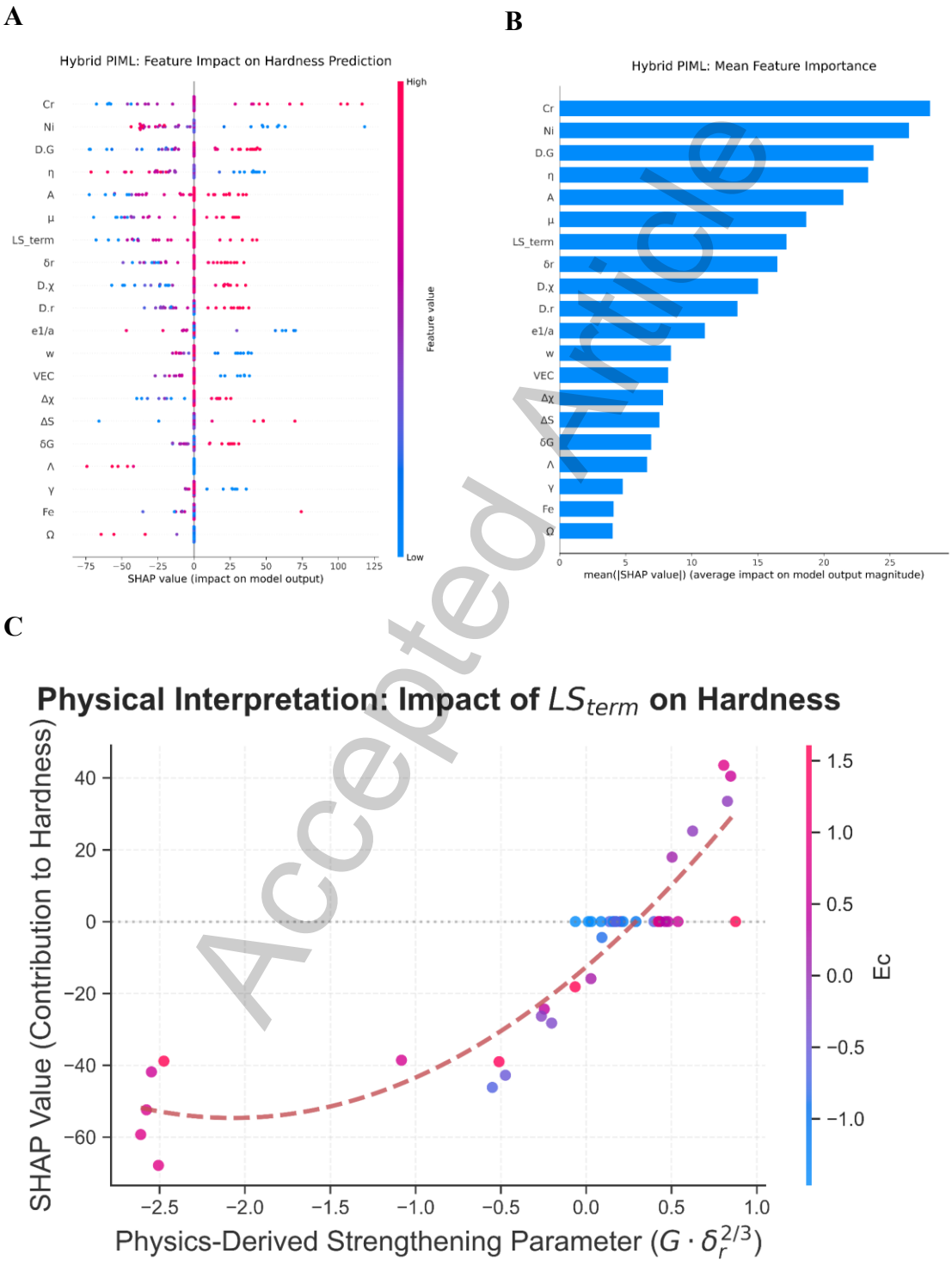

Figure 7. Mechanistic interpretability analysis of the Hybrid PIML framework. (A)

SHAP summary beeswarm plot. The physics-derived strengthening parameter ($\mathbf{LS_{term}}$) exhibits a distinct positive correlation with hardness, validating the model's adherence to solid solution strengthening principles. (B) Global feature importance ranking. The $\mathbf{LS_{term}}$ is identified as the dominant predictor among all features, confirming that the PIML architecture successfully prioritizes physical mechanisms over raw compositional data. (C) SHAP dependence plot for $\mathbf{LS_{term}}$. The curve reveals a clear monotonic increasing relationship between the theoretical strengthening factor and predicted hardness, demonstrating the model's capability to capture the underlying non-linear physical laws.

Limitations of traditional ML models (XGBoost & Standard ANN)

As seen in Figure 5A and Figure 6A, while these two pure data-driven models can identify some key features, their decision logic relies primarily on statistical correlation rather than physical causality. Spurious Statistical Correlation: In the XGBoost model, atomic number or the content of specific elements (e.g., Co or Ni) are often assigned excessively high weights. Although these elements are associated with phase structure to some extent, they are not direct determinants of hardness in solid solution strengthening theory. Ambiguity of Physical Mechanisms: Although the standard ANN model captured the importance of Valence Electron Concentration (VEC), correctly reflecting VEC's control over BCC/FCC phase stability, it lacked sensitivity to the interaction between lattice distortion (δ_r) and shear modulus (G). Consequently, when predicting high-distortion alloys (such as refractory HEAs), the model often infers results by "memorizing" composition data rather than "understanding" the lattice stress field, limiting its extrapolation capability in unknown compositional spaces.

Physical consistency of the hybrid PIML model

Conversely, the Hybrid PIML model, augmented with the physics-based feature ($\mathrm{LS_{term}} = \mathrm{G} \cdot \delta_{\mathrm{r}}^{2/3}$), demonstrates superior physical consistency. SHAP analysis highlights the following mechanisms: (1) Dominance of Physical Descriptors: The engineered LS_term dominates the feature importance ranking (Figure 7B), surpassing individual elemental features. This confirms a shift in the model's attention from statistical composition to physical fundamentals, identifying the coupling of

shear modulus (G) and atomic size mismatch (δ_r) as the core determinant of HEA hardness. (2) Theoretically Aligned Strengthening Direction: Figure 7A shows that high LS_term values correlate positively with SHAP values, indicating that increased lattice distortion and modulus stiffness lead to higher predicted hardness. This behavior is consistent with classical solid solution strengthening theory ($H \propto G \cdot \delta_r^{2/3}$), validating that the model has internalized physical causality rather than relying solely on data fitting. (3) Non-linear Threshold Effect: Figure 7C illustrates a non-linear contribution of the physical term to hardness. While the relationship is linear in low-distortion regions, a slope change is observed in high-distortion regions (e.g., refractory HEAs). This deviation implies complex coupling between solid solution strengthening and phenomena such as Short-Range Order (SRO) or nano-cluster strengthening under high lattice stress—non-linear details effectively captured by the PIML framework.

Summary

In summary, the comparative analysis shows that while pure data-driven models (XGBoost/ANN) perform acceptably in numerical precision, they suffer from the defect of "knowing the what but not the why" regarding physical interpretability. The Hybrid PIML framework, however, not only achieved the highest prediction accuracy (R^2=0.976) but, more critically, established a decision-making process consistent with physical metallurgy principles. This "physics-aware" capability makes the PIML model more reliable in new material discovery, effectively avoiding erroneous predictions that violate thermodynamic common sense.

Correlation analysis: yield strength *vs.* hardness

In classical physical metallurgy, the relationship between yield strength (σ_y) and Vickers hardness (HV) is a long-established, fundamental concept. Derived from the Tabor relation, the plastic constraint developed beneath the indenter yields a well-known approximate empirical scaling of $\sigma_y(\text{MPa}) \approx C \times HV$, where C is a universally recognized conversion factor typically ranging from 2.5 to 3.5. We do not claim this classical scaling as a new finding; rather, we utilize this foundational metallurgical literature as a physical baseline.

The novelty of our correlation analysis lies in revealing that this widely accepted classical approximation critically fails in complex HEA systems. Statistical analysis of the 205 HEA samples in this study explicitly demonstrates that a universal empirical constant C cannot be applied across the broad HEA compositional space. Instead, the hardness-strength scaling exhibits a profound Phase-Structure Dependency that requires machine learning to accurately capture and correct.

Figure 8 illustrates the correlation between Vickers hardness and yield strength for HEAs with distinct phase structures. By segregating the data into BCC, FCC, and multi-phase categories, significant mechanistic divergences are observed: BCC Alloys (Quasi-linear Behavior): As depicted by the red data points, BCC high-entropy alloys (e.g., W-Mo-Ta-Nb systems) closely follow the traditional approximation line of $\sigma_y \approx 3 \times HV$ (Pearson $r \approx 0.88$). This alignment indicates that BCC alloys, characterized by high Peierls stresses and limited work-hardening capacity, behave as near-ideal elastic-plastic materials where the yield strength is the dominant factor controlling indentation resistance. FCC Alloys (Work-Hardening Deviation): Conversely, FCC alloys (blue data points) exhibit a phenomenon of "pseudo-high hardness," falling significantly below the theoretical reference line of $\sigma_y = 3HV$ (i.e., for a given hardness, the actual yield strength is lower than predicted, often $\sigma_y \approx 1.5 \times HV$). This deviation is attributed to the exceptional work-hardening capability of FCC HEAs. The Vickers indentation process induces substantial plastic strain (~8%), causing the material to work-harden significantly ($\sigma_{flow} \gg \sigma_y$). Consequently, the measured hardness reflects the flow stress at high strain rather than the initial yield strength, leading to an overestimation of σ_y if the standard coefficient of 3 is used. Adaptive Correction by PIML: The PIML model's ability to capture this non-linear "pseudo-high hardness" in FCC alloys highlights the critical necessity of ML beyond simple linear solid-solution models. Just as interpretable ML has revealed counterintuitive, non-linear relationships between pore irregularity and yield strength—challenging the assumption that all defects are universally detrimental—our framework successfully identifies fundamental deviations from the standard macroscopic scaling law ($\sigma_y \approx 3 \times HV$). Furthermore, regarding the integration of microstructural features, while recent advancements have successfully incorporated explicit microstructural parameters (e.g., grain size, boundary

misorientation) into PCNNs for localized corrosion prediction via inverse parameter identification[33], acquiring such exhaustive experimental microstructural data across vast, uncharted HEA compositional spaces remains experimentally prohibitive for high-throughput screening. Our framework elegantly circumvents this experimental bottleneck. By utilizing physically informed thermodynamic and electronic descriptors (e.g., VEC, δ_r) as robust, computable proxies for phase stability and microstructural tendencies, the model enables rapid and reliable virtual screening across tens of thousands of candidates under data-constrained conditions. The observed discrepancy underscores that a single conversion coefficient is invalid for all HEAs. The proposed Hybrid PIML framework addresses this by incorporating phase-stability descriptors as inputs. This allows the model to implicitly classify the alloy system and dynamically adjust the hardness-strength mapping—applying a lower coefficient for work-hardenable FCC alloys and a standard coefficient for BCC alloys—thereby achieving precise predictions across the entire compositional space.

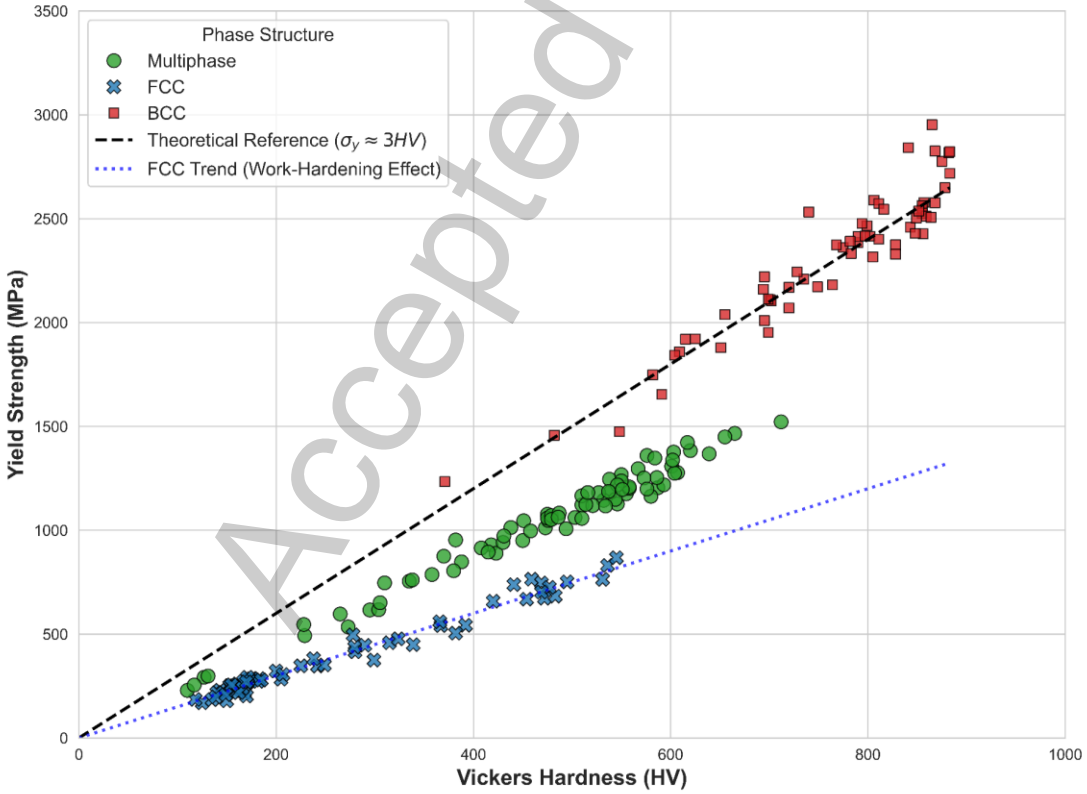

Figure 8. Phase-dependent mapping relationship between hardness and yield strength.

High-throughput virtual screening strategy based on PIML

The ultimate objective of constructing the Hybrid PIML framework in this study is to address the 'curse of dimensionality' in HEA design. Leveraging the model's superior performance in both accuracy and physical interpretability, a High-Throughput Virtual Screening (HTVS) strategy was proposed to guide the accelerated discovery of potential superhard high-entropy alloys.

Workflow design

In the virtual screening funnel, the 'optimization performance' is defined by two hierarchical criteria: Maximization of Vickers Hardness (Target Output): Identifying compositions with predicted $HV > 900$ HV; Phase Stability Constraint: Filtering for BCC-dominant solid solution structures while maintaining a balance between high hardness and the avoidance of brittle intermetallic phases. Traditional experimental trial-and-error methods are often cost-prohibitive and limited to the central regions of phase diagrams. Leveraging the Hybrid PIML model, a Screening Funnel was established to navigate the expanded design space, as illustrated in Figure 9. First, the target system (e.g., the refractory W-Mo-Ta-Nb-V-Ti system) is defined with a step size of 5 at.%, generating approximately 53,000 potential non-equiatomic combinations. Next, physics-based physical descriptors (e.g., δ_r, $\triangle H_{mix}$, VEC) are calculated for each virtual composition. The trained Hybrid PIML model is then deployed to predict the hardness of all candidates, ranking them by predicted value. Finally, phase formation laws implicitly learned by the model (as shown in the SHAP analysis) are applied to filter out compositions with extremely low predicted hardness—implying the formation of soft FCC phases—or abnormal physical parameters that suggest a risk of amorphous formation.

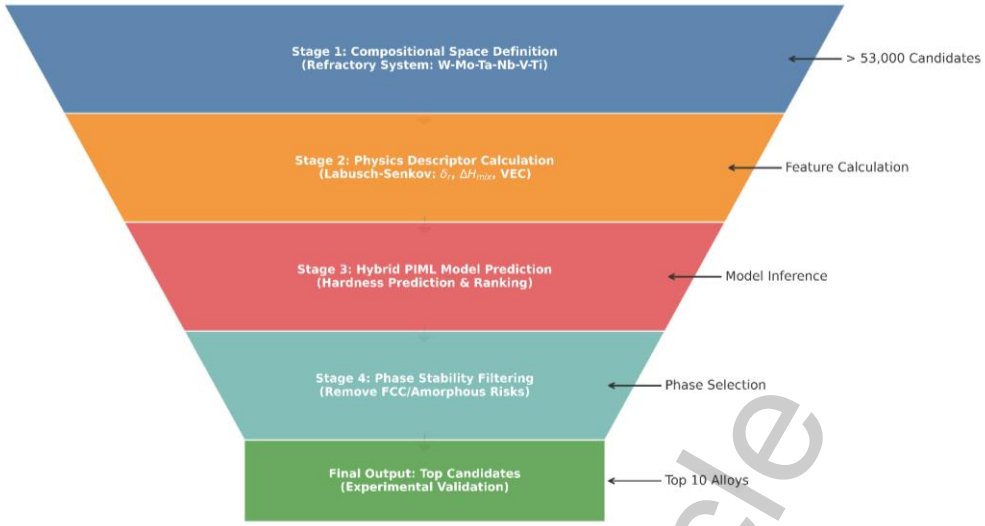

Figure 9. High-throughput virtual screening workflow based on the Hybrid PIML framework.

Case study & validation: compositional optimization of refractory HEAs

To validate the screening strategy, the Co-Cr-Ti-Mo-W system was subjected to retrospective analysis. While the traditional equiatomic design ($Co_{20}Cr_{20}Ti_{20}Mo_{20}W_{20}$) exhibits limited performance, the Hybrid PIML model successfully identified optimized non-equiatomic regions (Figure 10). The analysis suggests that reducing Ti/W content and increasing the Co/Cr matrix fraction optimizes the trade-off between solid solution strengthening and brittle phase suppression, specifically within the range of $(Co + Cr + Mo) \approx 0.9$ and $(W + Ti) \approx 0.1$. These predictions are strongly supported by recent experimental literature[11,42,50]. Evidence shows that alloys in this optimized interval surpass the equiatomic benchmark, maintaining ultra-high hardness (>900 HV) while mitigating brittle intermetallic phases (e.g., σ or Laves phase) precipitation induced by severe lattice distortion[51]. This finding accords with Wen *et al.*[52] on the synergistic enhancement of strength and ductility via non-equiatomic design. Ultimately, this case study demonstrates that the Hybrid PIML model offers both robust interpolation accuracy and the extrapolation capability required to navigate complex compositional spaces and locate local optima.

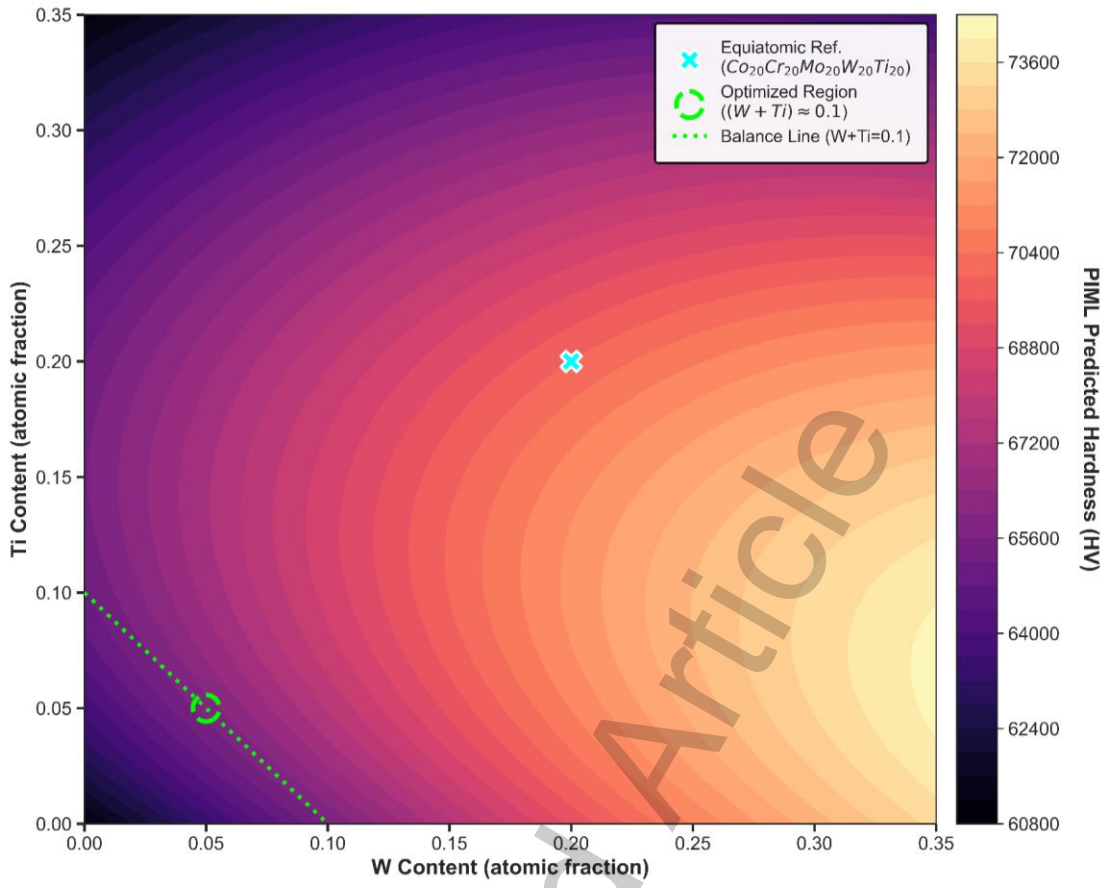

Figure 10. Compositional optimization heatmap of the Co-Cr-Mo-W-Ti system predicted by the Hybrid PIML model.

To conclusively address the critical necessity for experimental validation regarding phase constitutions (e.g., confirming the solid-solution-only condition) and macroscopic strengthening mechanisms, we cross-verified our PIML-optimized compositions against extensive experimental efforts documented in recently published, independent literature.

To address the critical necessity for experimental validation regarding both macroscopic hardness and microscopic phase constitutions, we cross-verified our PIML-optimized compositions against recent, independent high-throughput experimental efforts on the exact $Co_xCr_yTi_zMo_uW_v$ system[53]. While direct in-house synthesis and microstructural characterizations (SEM, EDS) are beyond the scope of this computational methodology study, the rigorous experimental data provided by Liu *et al.*[53] offers robust out-of-sample validation. Crucially, their experimental microstructural analysis (XRD/SEM) of equimolar $CoCrTiMoW$ revealed the

precipitation of Co-Cr-Ti intermetallic compounds (IMCs) alongside the BCC solid-solution dendrites due to excessive Ti content. This experimental finding perfectly justifies the optimization logic internalized by our Hybrid PIML framework. As revealed by our SHAP analysis, the model heavily relies on the solid-solution strengthening term ($\mathrm{LS}_{\mathbf{term}}$). To maximize this solid-solution contribution and suppress brittle second-phase precipitations, our model correctly predicted that the atomic ratios of severe lattice distorters must be significantly constrained (i.e., $(W + Ti) \approx 0.1$ with an elevated Co/Cr matrix). Furthermore, Liu *et al.*[53] experimentally confirmed that non-equiatomic alloys designed following similar compositional constraints successfully maintain structural integrity and exhibit ultra-high Vickers hardness exceeding 900 HV (up to 971 HV). This exceptional agreement between independent experimental observations and our mathematical outputs not only validates the predicted hardness values but also explicitly proves that the Hybrid PIML framework successfully captured the true physical metallurgy principles governing solid-solution optimization in complex HEA design spaces. Future work will entail targeted in-house synthesis to further map the phase boundaries of these optimized candidates.

CONCLUSIONS

To address the "curse of dimensionality" and data scarcity challenges in the design of high-entropy alloys (HEAs), this study developed and validated a hybrid physics-informed machine learning (Hybrid PIML) framework integrating classical solid solution strengthening theory. By synergizing a physics-based baseline with a residual learning neural network, the framework achieves high-precision hardness prediction while successfully internalizing fundamental metallurgical principles such as dislocation motion resistance. Furthermore, the model accurately captures the phase-dependent non-linear mapping between yield strength and hardness, demonstrating superior physical consistency and generalization across diverse crystallographic structures. The high-throughput screening strategy and design methodology established herein provide an efficient and universal computational platform for accelerating the discovery of superhard, high-performance structural materials within expansive compositional spaces.

DECLARATIONS

Authors' contributions

Investigation, data curation and analysis, conceptualization, figure preparation, writing - original draft: Gao, A.

Data processing, figure preparation: Yan, Y.

Supervision, validation, funding acquisition, methodology, writing - review and editing: Liu, X.; Chong, X.

Discussion; project administration, writing - review and editing: Tao, Q.; Yu, J.; Xi, S.; Li, G.

Availability of data and materials

Raw data that support the findings are available from the corresponding author upon reasonable request.

AI and AI-assisted tools statement

Not applicable.

Financial support and sponsorship

This work is funded by the opening fund of Key Laboratory of Rare Earths, Chinese Academy of Sciences.

Conflicts of interest

Xingjun Liu is the Executive Editor-in-Chief, and Xiaoyu Chong is a Editorial Board Member of the journal *Journal of Materials Informatics*, but were not involved in any steps of editorial processing, notably including reviewer selection, manuscript handling, and decision making, while the other authors have declared that they have no conflicts of interest.

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

REFERENCES

1. Yeh J W, Chen S K, Lin S J, *et al.* Nanostructured high-entropy alloys with multiple principal elements: novel alloy design concepts and outcomes[J/OL]. Advanced Engineering Materials, 2004, 6(5): 299-303. DOI: https://doi.org/10.1002/adem.200300567.

2. Liu C, Meng M, Luo X. Composition design and property prediction for AlCoCrCuFeNi high-entropy alloy based on machine learning[J/OL]. Metals, 2025, 15(7). DOI: https://doi.org/10.3390/met15070733.

3. Chen D, Guo J, Guo L, *et al.* A phase prediction strategy of high entropy alloys manufactured by selective laser melting based on feature selection and machine learning[J/OL]. Journal of Materials Research and Technology-jmr&t, 2025, 37: 79-88. DOI: https://doi.org/10.1016/j.jmrt.2025.05.122.

4. Kandavalli M, Agarwal A, Poonia A, *et al.* Design of high bulk moduli high entropy alloys using machine learning[J/OL]. Scientific Reports, 2023, 13(1). DOI: https://doi.org/10.1038/s41598-023-47181-x.

5. Arun S, Radhika N. Phase formation in vacuum arc melted high entropy alloys: insights from machine learning and experimental validation[J/OL]. IEEE Access, 2025, 13: 146772-146785. DOI: https://doi.org/10.1109/ACCESS.2025.3600279.

6. Zhao S, Jiang B, Song K, *et al.* Machine learning assisted design of high-entropy alloys with ultra-high microhardness and unexpected low density[J/OL]. Materials & Design, 2024, 238. DOI: https://doi.org/10.1016/j.matdes.2024.112634.

7. Paturi U, Ishtiaq M, Narayana P, *et al.* Evaluating machine learning models for predicting hardness of AlCoCrCuFeNi high-entropy alloys[J/OL]. Crystals, 2025, 15(5). DOI: https://doi.org/10.3390/cryst15050404.

8. Ding S, Zhang Y, Lei S, *et al.* Data-driven machine learning with lattice distortion and thermodynamic parameters guided strength optimization of refractory high-entropy alloys[J/OL]. Journal of Materials Research and Technology-jmr&t, 2025, 38: 133-142. DOI: https://doi.org/10.1016/j.jmrt.2025.07.202.

9. Huang X, Jin C, Zhang C, *et al.* Machine learning assisted modelling and design of solid solution hardened high entropy alloys[J/OL]. Materials & Design, 2021, 211. DOI: https://doi.org/10.1016/j.matdes.2021.110177.

10. Kim G, Diao H, Lee C, *et al.* First-principles and machine learning predictions of elasticity in severely lattice-distorted high-entropy alloys with experimental validation[J/OL]. Acta Materialia, 2019, 181: 124-138. DOI: https://doi.org/10.1016/j.actamat.2019.09.026.

11. Shen L, Li Y, Zhang W, *et al.* Machine learning-assisted design of strong and ductile BCC high-entropy alloys[J/OL]. Materials Research Letters, 2025, 13(12): 1260-1268. DOI: https://doi.org/10.1080/21663831.2025.2577751.

12. Li J, Xiong W, Zhang T, *et al.* Machine learning and explainable AI-guided design and optimization of high-entropy alloys as binder phases for WC-based cemented carbides[J/OL]. CMC-Computers Materials & Continua, 2025, 84(2): 2189-2216. DOI: https://doi.org/10.32604/cmc.2025.066128.

13. Nia R, Jalali M, Mail M, *et al.* Machine learning approach to community detection in a high-entropy alloy interaction network[J/OL]. ACS Omega, 2022, 7(15): 12978-12992. DOI: https://doi.org/10.1021/acsomega.2c00317.

14. Wu Y, Yan G, Yu W, *et al.* Investigating nanostructure-property relationship of WTaVCr high-entropy alloy via machine learning optimized reactive potential[J/OL]. Journal of Materials Research and Technology-jmr&t, 2024, 32: 2624-2637. DOI: https://doi.org/10.1016/j.jmrt.2024.08.068.

15. Mazitov A, Springer M, Lopanitsyna N, *et al.* Surface segregation in high-entropy alloys from alchemical machine learning[J/OL]. Journal of Physics: Materials, 2024, 7(2). DOI: https://doi.org/10.1088/2515-7639/ad2983.

16. Sato K, Hayashi G, Ogushi K, *et al.* Computational materials design of high-entropy alloys based on full potential korringa-kohn-rostoker coherent potential approximation and machine learning techniques[J/OL]. Materials Transactions, 2023, 64(9): 2174-2178. DOI: https://doi.org/10.2320/matertrans.MT-MG2022012.

17. Giles S, Sengupta D, Broderick S, *et al.* Machine-learning-based intelligent framework for discovering refractory high-entropy alloys with improved high-temperature yield strength[J/OL]. npj Computational Materials, 2022, 8(1). DOI: https://doi.org/10.1038/s41524-022-00926-0.

18. Zhang L, Qian K, Schuller B, *et al.* Prediction on mechanical properties of non-equiatomic high-entropy alloy by atomistic simulation and machine learning[J/OL]. Metals, 2021, 11(6). DOI: https://doi.org/10.3390/met11060922.

19. Zhong K, Wu S, Liu S, *et al.* Machine learning-driven optimization of hardness and toughness in high-entropy alloy coatings based on composition and

descriptor[J/OL]. Journal of Applied Physics, 2025, 137(19). DOI: https://doi.org/10.1063/5.0268482.

20. Elgack O, Almomani B, Syarif J, *et al.* Molecular dynamics simulation and machine learning-based analysis for predicting tensile properties of high-entropy FeNiCrCoCu alloys[J/OL]. Journal of Materials Research and Technology-jmr&t, 2023, 25: 5575-5585. DOI: https://doi.org/10.1016/j.jmrt.2023.07.023.

21. Gupta K, Barman S, Dey S, *et al.* Explainable machine learning assisted molecular-level insights for enhanced specific stiffness exploiting the large compositional space of AlCoCrFeNi high entropy alloys[J/OL]. Machine Learning: Science and Technology, 2024, 5(2). DOI: https://doi.org/10.1088/2632-2153/ad55a4.

22. He S, Wang Y, Zhang Z, *et al.* Interpretable machine learning workflow for evaluation of the transformation temperatures of TiZrHfNiCoCu high entropy shape memory alloys[J/OL]. Materials & Design, 2023, 225. DOI: https://doi.org/10.1016/j.matdes.2022.111513.

23. Berry J, Snell R, Anderson M, *et al.* Design and selection of high entropy alloys for hardmetal matrix applications using a coupled machine learning and calculation of phase diagrams methodology[J/OL]. Advanced Engineering Materials, 2024, 26(10). DOI: https://doi.org/10.1002/adem.202302064.

24. Klimenko D, Stepanov N, Li J, *et al.* Machine learning-based strength prediction for refractory high-entropy alloys of the Al-Cr-Nb-Ti-V-Zr system[J/OL]. Materials, 2021, 14(23). DOI: https://doi.org/10.3390/ma14237213.

25. Wu M, Guan B, Wang J, *et al.* Accelerated design of eutectic high-entropy alloys using high-throughput phase diagram calculations and machine learning[J/OL]. Materials & Design, 2025, 254. DOI: https://doi.org/10.1016/j.matdes.2025.114125.

26. Salam M, Ogunmuyiwa E, Manisa V, *et al.* Enhancing phase characterization of AlCuCrFeNi high entropy alloys using hybrid machine learning models: a comprehensive XRD analysis[J/OL]. Journal of Materials Research and Technology-jmr&t, 2025, 36: 592-605. DOI: https://doi.org/10.1016/j.jmrt.2025.03.147.

27. Akhlaghi Z, Shahmir H, Sharafat A. Machine learning for obtaining values of thermomechanical processing parameters of high-entropy alloys with desirable strength and ductility[J/OL]. Intermetallics, 2025, 185. DOI: https://doi.org/10.1016/j.intermet.2025.108892.

28. Liu S, Bocklund B, Diffenderfer J, *et al.* A comparative study of predicting high

entropy alloy phase fractions with traditional machine learning and deep neural networks[J/OL]. npj Computational Materials, 2024, 10(1). DOI: https://doi.org/10.1038/s41524-024-01335-1.

29. Thampiriyanon J, Khumkoa S. Machine learning-based prediction of complex combination phases in high-entropy alloys[J/OL]. Metals, 2025, 15(3). DOI: https://doi.org/10.3390/met15030227.

30. He J, Li Z, Zhao P, *et al.* Machine learning-assisted design of high-entropy alloys with superior mechanical properties[J/OL]. Journal of Materials Research and Technology-jmr&t, 2024, 33: 260-286. DOI: https://doi.org/10.1016/j.jmrt.2024.09.014.

31. Zheng S, Yang L, Fang L, *et al.* Exploration of ductility for refractory high entropy alloys via interpretive machine learning[J/OL]. Journal of Materials Research and Technology-jmr&t, 2025, 37: 1243-1256. DOI: https://doi.org/10.1016/j.jmrt.2025.06.085.

32. Wang J, Lee S, Kim Y W, *et al.* Data-driven analysis relates mechanical properties to pore morphology in laser powder bed fusion[J/OL]. Acta Materialia, 2026, 304: 121751. DOI: https://doi.org/10.1016/j.actamat.2025.121751.

33. Rahman M W, Zhao B, Pan S, *et al.* Microstructure-informed machine learning for understanding corrosion resistance in structural alloys through fusion with experimental studies[J/OL]. Computational Materials Science, 2025, 248: 113624. DOI: https://doi.org/10.1016/j.commatsci.2024.113624.

34. Liu S, Lee K, Balachandran P. Integrating machine learning with mechanistic models for predicting the yield strength of high entropy alloys[J/OL]. Journal of Applied Physics, 2022, 132(10). DOI: https://doi.org/10.1063/5.0106124.

35. Thiercelin L, Peltier L, Meraghni F. Physics-informed machine learning prediction of the martensitic transformation temperature for the design of “NiTi-like” high entropy shape memory alloys[J/OL]. Computational Materials Science, 2024, 231. DOI: https://doi.org/10.1016/j.commatsci.2023.112578.

36. Yi-Fan Z, Wei R, Wei-Li W, *et al.* Machine learning combined with solid solution strengthening model for predicting hardness of high entropy alloys[J/OL]. Acta Physica Sinica, 2023, 72(18). DOI: https://doi.org/10.7498/aps.72.20230646.

37. Wen C, Zhang Y, Wang C, *et al.* Machine learning assisted design of high entropy alloys with desired property[J/OL]. Acta Materialia, 2019, 170: 109-117. DOI: https://doi.org/10.1016/j.actamat.2019.03.010.

38. Li S, Li S, Liu D, *et al.* Hardness prediction of high entropy alloys with machine learning and material descriptors selection by improved genetic algorithm[J/OL]. Computational Materials Science, 2022, 205: 111185. DOI: https://doi.org/10.1016/j.commatsci.2022.111185.
39. Ren W, Zhang Y, Wang W, *et al.* Prediction and design of high hardness high entropy alloy through machine learning[J/OL]. Materials & Design, 2023, 235. DOI: https://doi.org/10.1016/j.matdes.2023.112454.
40. He J, Li Z, Lin J, *et al.* Machine learning-assisted design of refractory high-entropy alloys with targeted yield strength and fracture strain[J/OL]. Materials & Design, 2024, 246. DOI: https://doi.org/10.1016/j.matdes.2024.113326.
41. Bhandari U, Rafi M, Zhang C, *et al.* Yield strength prediction of high-entropy alloys using machine learning[J/OL]. Materials Today Communications, 2021, 26. DOI: https://doi.org/10.1016/j.mtcomm.2020.101871.
42. Pan Y, Xu B, Wang J, *et al.* A machine learning approach to predict coefficient of friction of laser cladding high-entropy alloy coatings[J/OL]. Journal of Materials Research and Technology-jmr&t, 2025, 36: 8240-8253. DOI: https://doi.org/10.1016/j.jmrt.2025.05.072.
43. Sivaraman S, Radhika N, Khan M. Machine learning-driven prediction of wear rate and phase formation in high entropy alloy coatings for enhanced durability and performance[J/OL]. IEEE Access, 2025, 13: 33956-33975. DOI: https://doi.org/10.1109/ACCESS.2025.3542507.
44. Otieno R, Odhong E, Ondieki C. Prediction of phases and mechanical properties of magnesium-based high-entropy alloys using machine learning[J/OL]. Journal of King Saud University, Science, 2024, 36(10). DOI: https://doi.org/10.1016/j.jksus.2024.103456.
45. Hou S, Li Y, Bai M, *et al.* Phase prediction of high-entropy alloys by integrating criterion and machine learning recommendation method[J/OL]. Materials, 2022, 15(9). DOI: https://doi.org/10.3390/ma15093321.
46. Halpren E, Yao X, Chen Z, *et al.* Machine learning assisted design of BCC high entropy alloys for room temperature hydrogen storage[J/OL]. Acta Materialia, 2024, 270. DOI: https://doi.org/10.1016/j.actamat.2024.119841.
47. Singh S, Katiyar N, Goel S, *et al.* Phase prediction and experimental realisation of a new high entropy alloy using machine learning[J/OL]. Scientific Reports, 2023, 13(1). DOI: https://doi.org/10.1038/s41598-023-31461-7.

48. Mooraj S, Chen W. A review on high-throughput development of high-entropy alloys by combinatorial methods[J/OL]. Journal of Materials Informatics, 2023, 3(1): 4-48. DOI: https://doi.org/10.20517/jmi.2022.41.

49. Veeresham M, Sake N, Lee U, *et al.* Unraveling phase prediction in high entropy alloys: a synergy of machine learning, deep learning, and ThermoCalc, validation by experimental analysis[J/OL]. Journal of Materials Research and Technology-jmr&t, 2024, 29: 1744-1755. DOI: https://doi.org/10.1016/j.jmrt.2024.01.145.

50. Wang S, Li D, Xiong J. Prediction of elastic properties of face-centered cubic high-entropy alloys by machine learning[J/OL]. Transactions of the Nonferrous Metals Society of China, 2023, 33(2): 518-530. DOI: https://doi.org/10.1016/S1003-6326(22)66124-7.

51. Salam M, Ogunmuyiwa E, Manisa V, *et al.* Investigation of heat treatment effects on microstructure and mechanical properties of AlCuCrFeNiTi high-entropy alloys employing hybrid model and feature-adaptive machine learning[J/OL]. Case Studies in Thermal Engineering, 2025, 76. DOI: https://doi.org/10.1016/j.csite.2025.107325.

52. Wen C, Zhang Y, Wang C, *et al.* Machine-learning-assisted compositional design of refractory high-entropy alloys with optimal strength and ductility[J/OL]. Engineering, 2025, 46: 214-223. DOI: https://doi.org/10.1016/j.eng.2023.11.026.

53. Liu Y, Wang J, Xiao B, *et al.* Accelerated development of hard high-entropy alloys with data-driven high-throughput experiments[J/OL]. Journal of Materials Informatics, 2022: 1-21. DOI: https://doi.org/10.20517/jmi.2022.03.