Supplementary Materials

1 Data Collection

1.1 High Entropy Alloy Hardness Dataset

The high entropy alloys (HEAs) hardness dataset contains 583 data points collected in our previous work [1]. This dataset includes composition, processing condition, Vickers hardness value. The processing conditions include five categories: as-cast, homogenized, work hardening, powder metallurgy and additive manufacturing. In addition, 12 empirical physical parameters are calculated as the material features, which involves the valence electron concentration, the difference of Pauling electronegativity, the difference of atomic radii, the mixing enthalpy, the configuration entropy, the number of itinerant electrons, the six square of work function, the melting point, the cohesive energy, the modulus mismatch in strengthening model, the difference of shear modulus and the shear modulus [1]. The probability density distribution of the target property values across this dataset is presented in Fig. S1a.

1.2 Steel Hardness Dataset

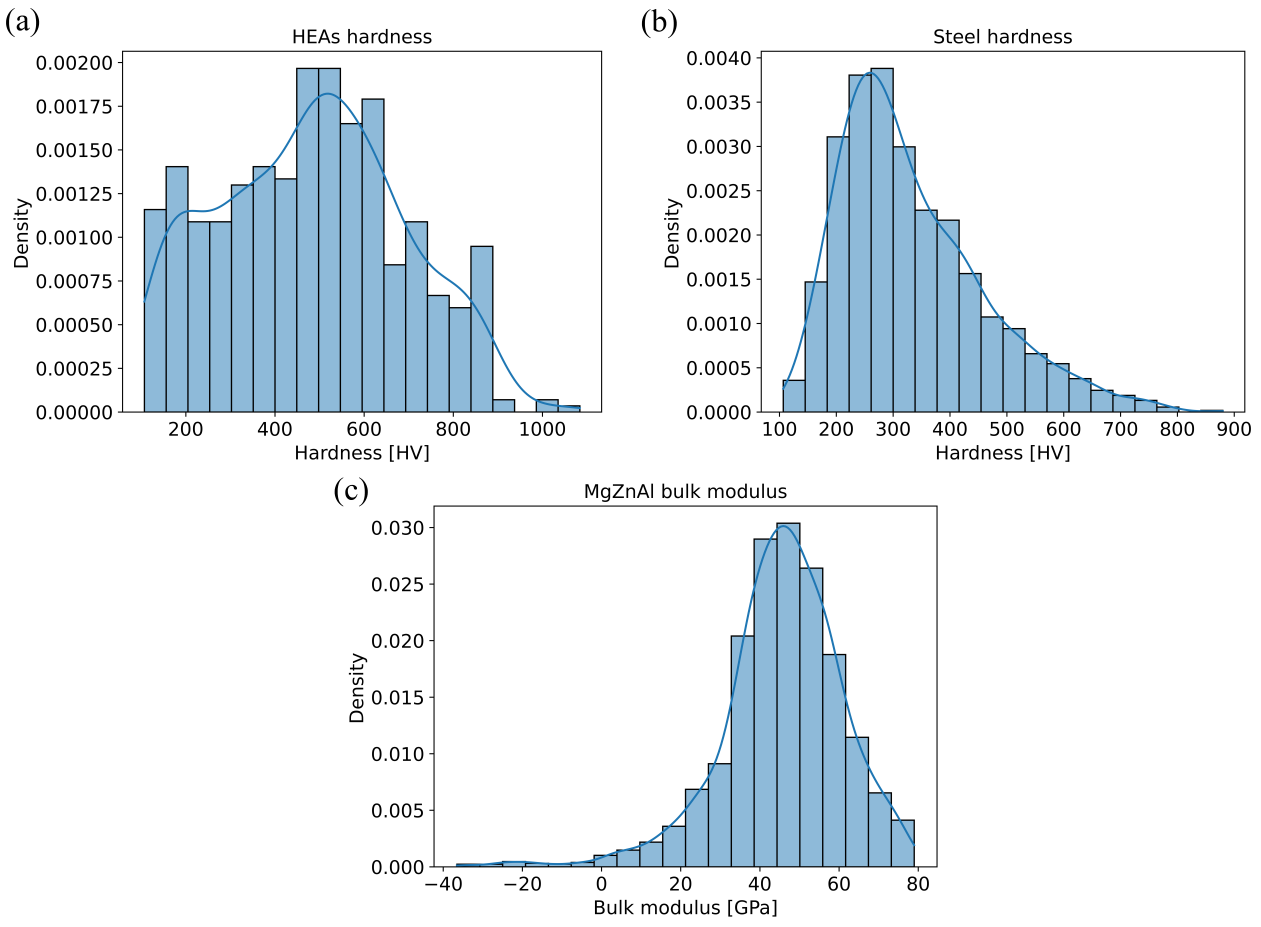
The steel hardness dataset specifically focuses on quenched and tempered martensitic low-alloy steels. The data were obtained from literature [2], which included hardness data for both high-alloy and low-alloy steels. These steel classifications are defined by their total alloy content, as established in prior research [3]. For this study, only low-alloy steel data were selected due to their superior data quality and broader practical engineering application value.

The dataset comprises 1373 low-alloy steel samples, with hardness (HV) as the target property. The features include: Chemical composition (wt. %): C, Mn, Si, Cr, V, Mo, Ni, Al, Cu, Ti, N, B and S. Processing conditions: Tempering temperature (*T*, K) and tempering time (*Time*, s), where time was log10-scaled (log10(*t*)) [4]. Additional metadata: Data source and a binary *DROP* indicator (True/False), this indicator is true if sample possesses mixed or potentially mixed microstructures. Categorical features were encoded as dummy variables to ensure machine learning (ML) compatibility. The probability density distribution of the target property values across this dataset is presented in Fig. S1b.

1.3 Photocatalyst Formation Energy Dataset

Two-dimensional (2D) metals stabilized in an octahedral coordination geometry with six ligands (2DO materials) exhibit significant potential as photocatalysts for hydrogen evolution [5,6]. The tunability of metal-ligand combinations enables a vast design space for such 2D materials. Ritesh Kumar et al. constructed a comprehensive database—termed the 2DO database—containing 3099 2DO photocatalysts from the 2D layered double hydroxide (LDH) family [7]. Each material in this dataset is characterized by its formation energy and a set of 162 descriptors, including elemental properties and chemical hardness parameters. The elemental features (151 in total) were derived using the Magpie framework via the Matminer package [8], while an additional 11 descriptors based on local chemical hardness were incorporated to quantify metal-ligand interactions. The probability density distribution of the target property values across this dataset is presented in Fig. S1c.

Notably, unlike other tasks that explicitly pursue maximum values, this task distinctly aims to minimize the target value. Consequently, in terms of data partitioning, data sorting, and model configuration of the present model, the corresponding settings were inversely adjusted compared to those used for other dataset.



**Figure S1:** The probability density distributions of the target property values across the three datasets.

2 The Hyperparameter Optimization of Extrapolation Random Forest

Based on extensive experimental results, several hyperparameter optimization details of Extrapolation Random Forest (ETR) model were identified to enhance the model’s extrapolation robustness. Firstly, the number of trees using linear regression should be bigger than those employing quadratic or cubic regression. For example, the number range of (100, 500) for the former and (1, 50) for the latter. Secondly, to ensure stable training of the regression models, the minimum number of samples per node should not be set too low: values above 20 are advised. When hyperparameter optimization or model training requires excessive computational resources, the manner of optimizing the split thresholds by minimizing the target variable’s variance during tree construction can be adopted to enhance efficiency.

3 The Details of ML Models and Improved Genetic Algorithm

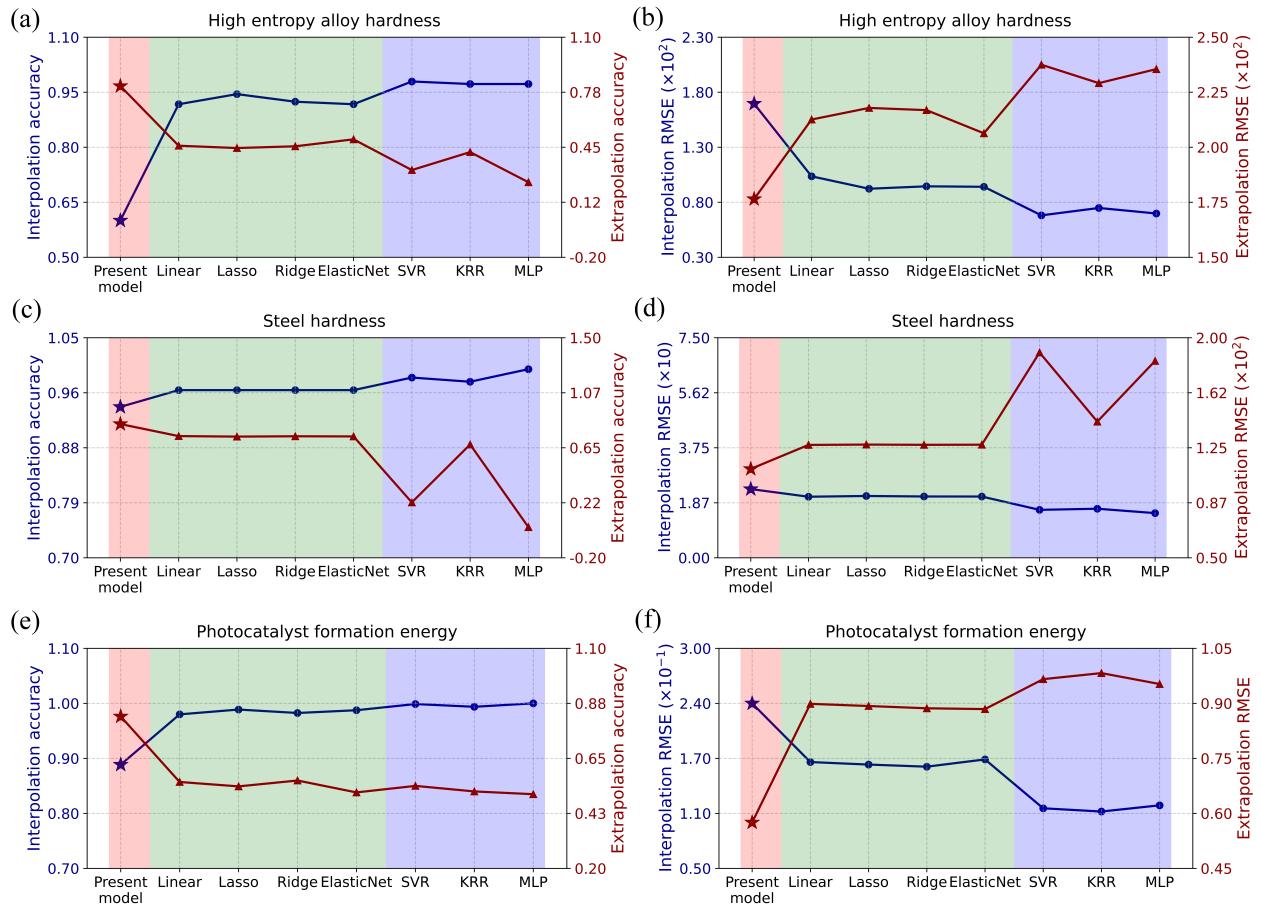
The study initially considered various tree-based ML models (regression trees, random forests, extremely randomized trees, LightGBM, and XGBoost) for model comparison. However, they were excluded due to their inherent limitations in extrapolation: conventional tree-based models (regression trees, random forests and extremely randomized trees) completely lack extrapolation capability, and boosted tree models (LightGBM and XGBoost) demonstrate poor performance in extrapolation tasks.

The feature selection method used for different models is the improved genetic algorithm (GA) proposed in our previous work [9]. The core objective of the improved GA is to find the global optimal feature subset accurately, quickly and steadily. The concepts of feature importance and gene manipulation are introduced into the improved GA to make it more comprehensible. Comparative analysis demonstrates that the improved GA is superior to the traditional GA and other typical feature selection methods in the aspects of accuracy, stability and efficiency obviously [9].

In this study, in order to obtain the global optimal feature subsets, the hyperparameters of improved GA was set as follows. Compared to traditional GA, there are two specific hyperparameters in improved GA. The certain number *d* of individuals were constructed before initial population creation, where this hyperparameter was set to 300. Among them, the certain number *e* of individuals with the lowest prediction errors were collected, where this hyperparameter was set to 150. Furthermore, as common hyperparameters in GA, the population consisted of 100 individuals, crossover probability was 0.8, a mutation probability was 0.01, and iterations was 50. The improved GA was run independently five times, and the optimal feature subset obtained during the iterations across these five runs was selected as the final result. Since hyperparameter tuning and training process of the ERF is cumbersome, for simplicity, the screening results from the RF can be used. Similarly, the Multilayer Perceptron (MLP) model could utilize all features. In practice, MLP hyperparameters were hardly optimized and typically defined empirically through repeated experimentation. To facilitate this process, two alternative strategies were adopted separately to: random partitioning of the original training data into 80% training and 20% validation sets; specifically allocating the most extreme 20% of samples as the validation set while using the remainder for training.

4 The Results of Different Models in Three Datasets

The classification accuracy and prediction errors for interpolation and extrapolation data across different models and three datasets are shown in Fig. S2.



**Figure S2:** **Classification accuracy and prediction errors for interpolation and extrapolation across different models and materials datasets.** The red background indicates the present model, the green background represents the simple model, and the purple background denotes the complex model.

|  |
| --- |
|  |
| true_vs_predicted |
| true_vs_predicted |

**Figure S3: Scatter plots of predicted versus actual values across the three benchmark tasks.** Blue and red points represent extrapolation and interpolation predictions, respectively. Dashed lines indicate the maximum/minimum labels in the training set, distinguishing interpolation from extrapolation regions.

The top-*n* accuracy (top-*n* fitness) of different models on three datasets are shown in Tables S1–S3.

**Table S1:** The top-*n* accuracy (top-*n* fitness) of different models for high entropy alloy hardness prediction.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***n*** | **10** | **20** | **30** | **40** |
| Present model | 2 (0.475) | 7 (0.460) | 8 (0.391) | 10 (0.370) |
| Linear | 1 (0.311) | 1 (0.065) | 6 (0.072) | 13 (0.115) |
| Lasso | 1 (0.240) | 2 (0.120) | 7 (0.177) | 15 (0.247) |
| Ridge | 1 (0.256) | 1 (0.155) | 6 (0.227) | 15 (0.284) |
| Elasticnet | 1 (0.173) | 1 (0.865) | 6 (0.094) | 11 (0.147) |
| SVR | 0 (0.416) | 0 (0.393) | 1 (0.399) | 3 (0.391) |
| KRR | 0 (0.568) | 1 (0.542) | 3 (0.527) | 6 (0.531) |
| MLP | 0 (0.042) | 1 (0.156) | 2 (0.221) | 6 (0.309) |

**Table S2:** The top-*n* accuracy (top-*n* fitness) of different models for steel hardness prediction.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***n*** | **10** | **20** | **30** | **40** |
| Present model | 5 (0.551) | 8 (0.405) | 11 (0.400) | 17 (0.396) |
| Linear | 2 (0.648) | 5 (0.580) | 10 (0.603) | 14 (0.594) |
| Lasso | 2 (0.648) | 6 (0.594) | 10 (0.612) | 15 (0.595) |
| Ridge | 2 (0.648) | 6 (0.594) | 10 (0.603) | 14 (0.594) |
| Elasticnet | 2 (0.649) | 6 (0.595) | 10 (0.604) | 14 (0.595) |
| SVR | 0 (0.568) | 0(0.565) | 0 (0.526) | 0 (0.495) |
| KRR | 4 (0.688) | 7 (0.645) | 10 (0.659) | 12 (0.654) |
| MLP | 0 (0.694) | 3 (0.679) | 6 (0.647) | 6 (0.634) |

**Table S3:** The top-*n* accuracy (top-*n* fitness) of different models for photocatalyst formation energy prediction.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***n*** | **10** | **20** | **30** | **40** |
| Present model | 4 (0.497) | 7 (0.366) | 11 (0.315) | 13 (0.264) |
| Linear | 0 (0.542) | 2 (0.557) | 6 (0.571) | 12 (0.597) |
| Lasso | 2 (0.670) | 7 (0.644) | 8 (0.632) | 10 (0.653) |
| Ridge | 0 (0.565) | 3 (0.596) | 5 (0.611) | 6 (0.604) |
| Elasticnet | 3 (0.681) | 6 (0.641) | 10 (0.636) | 13 (0.645) |
| SVR | 0 (0.595) | 0 (0.601) | 0 (0.576) | 0 (0.562) |
| KRR | 0 (0.609) | 0 (0.579) | 1 (0.564) | 1 (0.524) |
| MLP | 0 (0.693) | 6 (0.693) | 7 (0.690) | 12 (0.691) |

5 The Design of Lightweight Mg-Zn-Al Alloy with High Bulk Modulus and Debye Temperature

5.1 Mg-Zn-Al Alloy Sample Generation

The alloy samples within the Mg-Zn-Al ternary system were investigated using the Universal Structure Predictor: Evolutionary Xtallography (USPEX) software [10]. In the computational framework, crystal structures containing up to 20 atoms were systematically enumerated, with the evolutionary algorithm performing a maximum of 40 population iterations. A comprehensive alloy space of 2295 Mg-Zn-Al configurations with varying composition ratios was generated, and the binary systems were also considered.

5.2 Density Functional Theory Calculation

In the present study, density functional theory (DFT) calculations were performed using the Vienna *Ab initio* Simulation Package (VASP) [11]. The electron-ion interactions were described by the projector-augmented wave (PAW) method within the generalized gradient approximation (PBE). A plane-wave cutoff energy of 520 eV was used for all simulations. Brillouin zone integration was performed using Monkhorst-Pack *k*-point meshes, with grid spacings of 0.04 Å−1 for electronic structure calculations and 0.03 Å−1 for elastic constant determinations. Structural optimization was achieved through conjugate gradient relaxation of both lattice parameters and atomic positions, with convergence criteria set at 10−4 eV for the total energy.

5.3 Feature and Target Generation

The target property bulk modulus (*B*) and Debye temperature () were calculated based on elastic tensor *C*ij by formulas (S1)–(S3):

|  |  |
| --- | --- |
|  | (S1) |
|  | (S2) |
|  | (S3) |

where and are the Voigt approximation and the Reuss approximation of the bulk modulus, respectively; and are the Voigt approximation and the Reuss approximation of the shear modulus, respectively; *ρ* is the density of the crystal structure; is the volume of the crystal structure; ℏ is Planck’s constant (1.05457 × 10−34 ); is the Boltzmann constant (1.38065 × 10−23 ). The density is calculated by dividing the total mass of atoms in the structure by its volume.

5.4 The Results of Mg-Zn-Al Alloy Design

The USPEX-generated structure of these two alloys are shown below.

|  |  |  |
| --- | --- | --- |
| Alloy−1 | | |
| 1.0 | | |
| 4.106973 | 0.025430 | −0.009567 |
| −0.037852 | 6.203845 | −0.045439 |
| −0.003030 | 1.318734 | 3.838185 |
| Zn Al | | |
| 1 5 | | |
| Direct | | |
| 0.497658 | 0.079804 | 0.080994 |
| 0.999677 | 0.410588 | 0.913477 |
| 0.998536 | 0.077375 | 0.577778 |
| 0.499714 | 0.410085 | 0.415530 |
| 0.998656 | 0.744117 | 0.244144 |
| 0.498926 | 0.746730 | 0.746402 |
| Alloy−2 | | |
| 1.0 | | |
| 5.771499 | −0.002245 | 0.008882 |
| −2.878240 | 5.004281 | 0.036675 |
| −1.493981 | 0.821999 | 4.757158 |
| Mg Al | | |
| 1 7 | | |
| Direct | | |
| 0.128213 | 0.630670 | 0.240220 |
| 0.628567 | 0.130365 | 0.239652 |
| 0.628618 | 0.631156 | 0.241245 |
| 0.381206 | 0.886744 | 0.739499 |
| 0.873099 | 0.378612 | 0.739639 |
| 0.379519 | 0.380971 | 0.748375 |
| 0.128869 | 0.131338 | 0.240850 |
| 0.875894 | 0.884596 | 0.734759 |

References

1. Li S, Li S, Liu D, Yang J, Zhang M. Hardness prediction of high entropy alloys with periodic table representation of composition, processing, structure and physical parameters. J Alloys Compd. 2023;967:171735. doi:10.1016/j.jallcom.2023.171735.
2. Murdoch HA, Field DM, Szajewski BA, McClenny LD, Garza A, Rinderspacher BC, et al. Tempered hardness optimization of martensitic alloy steels. Integr Mater Manuf Innov. 2023;12(4):301–20. doi:10.1007/s40192-023-00311-9.
3. ASTM A255. Standard test methods for determining hardenability of steel. West Conshohocken, PA, USA: ASTM International; 2003.
4. Speich G. Tempering of low-carbon martensite. Trans Metall Soc AIME. 1969;245:2553–64.
5. Gunjakar JL, Kim IY, Lee JM, Lee NS, Hwang SJ. Self-assembly of layered double hydroxide 2D nanoplates with graphene nanosheets: an effective way to improve the photocatalytic activity of 2D nanostructured materials for visible light-induced O2 generation. Energy Environ Sci. 2013;6(3):1008–17.
6. Zhao Y, Zhang S, Shi R, Waterhouse GIN, Tang J, Zhang T. Two-dimensional photocatalyst design: a critical review of recent experimental and computational advances. Mater Today. 2020;34:78–91. doi:10.1016/j.mattod.2019.10.022.
7. Kumar R, Singh AK. Chemical hardness-driven interpretable machine learning approach for rapid search of photocatalysts. npj Comput Mater. 2021;7:197. doi:10.1038/s41524-021-00669-4.
8. Ward L, Agrawal A, Choudhary A, Wolverton C. A general-purpose machine learning framework for predicting properties of inorganic materials. npj Comput Mater. 2016;2:16028. doi:10.1038/npjcompumats.2016.28.
9. Li S, Li S, Liu D, Zou R, Yang Z. Hardness prediction of high entropy alloys with machine learning and material descriptors selection by improved genetic algorithm. Comput Mater Sci. 2022;205:111185. doi:10.1016/j.commatsci.2022.111185.
10. Glass CW, Oganov AR, Hansen N. USPEX—Evolutionary crystal structure prediction. Comput Phys Commun. 2006;175(11–12):713–20. doi:10.1016/j.cpc.2006.07.020.
11. Hafner J. *Ab-initio* simulations of materials using VASP: density-functional theory and beyond. J Comput Chem. 2008;29(13):2044–78. doi:10.1002/jcc.21057.