Research Article

Accelerated development of hard high-entropy alloys with data-driven high-throughput experiments

Yi Liu[1,2,*], Jiong Wang[1], Bin Xiao[1], Jintao Shu[1]

[1]Materials Genome Institute, Shanghai University, Shanghai 200444, China.
[2]Zhejiang Laboratory, Hangzhou 311100, Zhejiang, China.

*****Correspondence to:** Prof. Yi Liu, Materials Genome Institute, Shanghai University, 333 Nanchen Road, Shanghai 200444, China. E-mail: YiLiu@shu.edu.cn

Abstract

The development of multicomponent alloys with target properties poses a significant challenge, owing to the enormous number of potential component combinations, high costs and the inefficiency of conventional empirical trial-and-error experimental approaches. To tackle this challenge, we develop a machine learning (ML)-guided high-throughput experimental (HTE) approach to accelerate the development of non-equimolar hard $Co_xCr_yTi_zMo_uW_v$ high-entropy alloys (HEAs). We first develop a set of all-process HTE facilities ranging from multi-tube ingredient assignment to multi-station electrical arc smelting and specimen preparation for bulk alloy samples with discrete compositions. Instead of random or combinatorial composition searching, HEAs with only ~1/28 of all the potential compositions are synthesized in two stages guided by the ML prediction. The final ML models, trained using 138 experimental data, predict the alloy hardness with mean relative errors of 5.3%, 6.3% and 15.4% at high (HV > 800), medium (HV = 600-800) and low (HV < 600) hardness ranges, respectively. In total, 14 superhard HEAs with HV > 900 are discovered by our ML-guided HTE approach. Moreover, the multiple ML models predict the hardness of 3876 hypothetical alloys covering the whole composition range, thereby revealing the systematic component effects based on the complete composition-hardness and descriptor-hardness correlations. The hardening mechanisms are elaborated by analyzing the microstructures of CoCrTiMoW. Furthermore, physical insights can be gained by transitioning from "*machine learning*" to "*learning from machine*". This work demonstrates that our ML-guided HTE approach provides an effective strategy for multicomponent alloy development with a potential hundred-fold overall increase in efficiency at a fraction of the cost compared to conventional methods.

INTRODUCTION

The development of materials with target properties is normally conducted using sequential single-sample experiments based on trial-and-error empirical experience, which makes the optimization process slow and costly. High-throughput experiments (HTE), originating from combinatorial experiments[1-5], can be used to prepare multiple samples with continuous composition gradients or to conduct experiments in batch mode with discrete compositions, featuring multi-station, automation and parallelization, for significantly enhanced efficiency compared to conventional preparation methods. HTEs can accelerate the development of materials, e.g., composition optimization for multicomponent alloys that often requires a large composition parameter space to be explored. Materials databases have now become increasingly important as a basis for materials design. Bligaard *et al.*[6] used large databases consisting of the lattice parameters, bulk moduli and heats of formation of ordered metallic alloys to find alloys with a compromise between low compressibility, high stability and cost. For most multicomponent materials, however, there are often insufficient property data available. HTEs can generate experimental data more efficiently, which is pivotal to building large materials property databases.

Although HTEs can improve the efficiency of composition searching, it remains challenging, if not impossible, to exhaust the enormous potential number of compositions of multicomponent alloys at a low cost. Reed and Warnken[7] utilized existing theoretical and empirical formulas to reduce the composition or process space to accelerate materials development. However, the reduction capability was limited by the theoretical assumption, and the extent of acceleration is still insufficient to meet real demands. Benefitting from the surge in artificial intelligence applications, machine learning (ML) methods based on statistical data analysis have been recently applied to develop materials[8-10] by correlating the input descriptors containing physical and chemical information with the target properties of the materials. Specifically, deep learning with artificial neural networks (ANNs) has been used by materials scientists in the development of new materials[11-18]. Conduit *et al.*[19-21] applied ANNs successfully to study nickel-based superalloys and molybdenum-based alloys. Islam *et al.*[22] employed a neural network method to recognize the underlying data pattern with an average accuracy of > 80% in cross-validation for phase selection and found that the valence electron concentration plays the most dominant role in the phase formation of multi-principal element alloys. Arisoy and Özel[23] used three-dimensional finite element methods to simulate the machining process and applied ML models to predict machining-induced microhardness and grain size. Khalaj and Pouraliakbar[24] used gene expressions with 17 different parameters, including the chemical compositions of steel and processing parameters, to build an ML model to predict the layer thickness of duplex-treated ceramic coatings on tool steels. ML methods have also been applied to predict other physical and mechanical properties, including the band gaps of inorganic crystals[25] and elastic constants[26].

As core principles of materials science, composition/process/structure (CPS)-property/performance (PP) relations are often very complex due to the high dimensional parameter space of CPS and the intrinsic nonlinear nature of CPS-PP correlations. Finding CPS-PP correlations is exactly the type of problem that ML methods can be used to solve. ML models can predict PP efficiently with input CPS descriptors, thereby guiding subsequent experiments until the completion of the optimization process after a few iteration cycles. In addition to the acceleration by HTE methods, ML prediction also can accelerate materials development by reducing the number of potential candidates to be studied experimentally. On the other hand, the use of HTEs represents an efficient method to generate experimental data for ML model training,

testing and validation. Combining HTE and ML methods can enhance the acceleration of materials development in a synergetic manner and is a major motivation of this work. In a related perspective, Correa-Baena *et al.*[27] proposed to accelerate materials development via the automation of experiment, ML and high-performance computing.

High-entropy alloys (HEAs) are typical multicomponent alloys that feature five or more principal elements, as originally proposed by Cantor *et al.*[28], Yeh *et al.*[29] and Ranganathan[30] in 2003 and 2004. The early development of HEAs emphasized the use of equimolar systems to maximize the configuration entropy. It is well known that the formation of solid solutions or intermetallic compounds often deviates from the equimolar stoichiometry, and the composition dependency of phase formation becomes more complex and critical among multiple principal elements. The compositions of HEAs need to be optimized beyond equimolar systems. Research into the effects of composition changes in HEAs has been widely carried out but is normally limited to one to two components and less than ten compositions[31-34]. Composition optimization requires a deep understanding of the full composition-property relations but the huge composition parameter space is very difficult to study with conventional approaches.

Refractory high-entropy alloys (RHEAs) have attracted significant attention due to their combined mechanical properties and thermal stability with potential applications in high-temperature environments. In 2010, Senkov *et al.*[35] developed two near-equimolar RHEAs, W-Nb-Mo-Ta and W-Nb-Mo-Ta-V, with HV values of 4.455 and 5.25 GPa, respectively. Later, Senkov *et al.*[36] prepared body-centered cubic (BCC) NbMoTaW and VNbMoTaW, with HV microhardness values of 4.46 and 5.42 GPa, respectively. By ion co-sputtering, Zou *et al.*[37] prepared WNbMoTa thin films and small-sized pillars consisting of strongly textured columnar nanometer-sized grains, reaching 6-8 GPa after 3 d of annealing at 1100 °C. Wu *et al.*[38] studied the effects of Mo and V on phase formation and solid solution strengthening in TiZrNbMoV HEAs. Melnick and Soolshenko[39] applied a thermodynamic approach to design $W_xTa_yMo_zNb_uTi_v$ RHEAs. Han *et al.*[40] studied the influence of Ti content on the phase structures and mechanical behavior of $Ti_xNbMoTaW$. Beyond the early development of equimolar single-phase HEAs, a large number of studies of near-equimolar single-phase alloys have been published, and multiphase non-equimolar alloys have drawn notable interest from researchers[41]. Both the complex compositions and processing conditions can affect the mechanical and physical properties of HEAs, meaning that HTEs and ML have become necessary approaches to explore the large parameter space[42-45].

Based on the combinatorial experiment concept, early HTEs used mainly multi-target magnetic or ion beam sputtering systems to prepare multilayer thin films with a continuous variation in composition. Such composition gradient materials provide significant details of the composition effects suitable for the construction of phase diagrams. However, the film properties often deviate from those of bulk materials. Furthermore, it remains challenging to explore the entire composition space for multicomponent alloys beyond ternary systems. Even using a pseudoternary diagram, the higher dimension information is greatly reduced after projection onto the three-dimensional representation. In addition, current HTE facilities focus mostly on a few major synthesis equipment. We suggest that the HTE concept should be applied to every step during the whole materials synthesis process, thereby maximizing the overall processing efficiency with few bottlenecks.

To tackle the challenges discussed above, here, we develop all-process HTE facilities to prepare bulk multicomponent alloy samples at the designed discrete compositions. The all-process HTE facilities are further combined with the prediction capability of ML to explore the important compositions for model building, validation, analysis and prediction within the complete high-dimension composition space. The

ML-guided HTE development is demonstrated to be able to accelerate the composition optimization of non-equimolar hard $Co_xCr_yTi_zMo_uW_v$ HEAs by carrying out fractional designed critical experiments. According to the ML predictions, we discuss the component effects according to the full composition-hardness and descriptor-hardness relations of the HEAs. Finally, the hardening mechanism and the roles of components are elaborated based on the microstructures of the CoCrTiMoW HEA.

The remainder of this study is organized as follows. First, the all-process HTE facilities and materials preparation process are introduced. We then describe the two-stage alloy development process guided by ML from the data generation, model training and testing to finalization and prediction of the ML models. The relations between composition/descriptor and hardness are also discussed, followed by a microstructural analysis of the CoCrTiMoW HEA. Finally, the conclusions of this study are drawn in the final section.

METHODS AND MATERIALS
High-throughput alloy synthesis facilities
In this work, we applied the HTE approach to synthesize $Co_xCr_yTi_zMo_uW_v$ HEAs using our in-house designed HTE facilities. We collaborated with the MTI Corporation[46] to develop a set of all-process HTE facilities for preparing bulk alloy samples with discrete compositions. Figure 1 shows the HTE facilities covering each step of the whole alloy synthesis process. The set of all-process HTE alloy synthesis facilities include (1) an automatic multi-tube (36) powder dispenser (MPD) for ingredient assignment; (2) a multi-station (16) ball mill (MBM) for powder mixing; (3) an automatic multi-station (16) pressing machine (MPM) for sample molding; (4) an automatic multi-station (32) electric arc melting furnace (MEAMF) for alloy smelting; (5) a multi-station (32) cold mosaic device (MCMD); (6) a multi-station (8) wire-cutting machine (MWCM); and (7) an automatic multi-station (16) polish-grinding machine (MPGM) for metallographic specimen preparation. Integrating the whole set of HTE equipment made the overall efficiency of the HTE alloy synthesis process at least ten times higher than that of the conventional single-sample mode preparation processes.

Materials preparation and characterization
In this work, $Co_xCr_yTi_zMo_uW_v$ HEAs with 138 designed compositions in total were synthesized using the all-process HTE facilities. The composition design process is described later in more detail. The alloy samples were prepared as follows. The alloy ingredients were assigned for each sample using the automatic MPD (36) according to the designed nominal compositions. Up to six types of components and 36 samples at different compositions can be made by the MPD. Pure metal powders (Co, Cr, Ti, Mo and W) at purity > 99.5 wt.% were used as raw materials to prepare the mixtures of alloy powders, with ~10 g of each sample measured at a precision of 0.01 g. The alloy powders were well mixed for 12 h at a speed of 150 rpm by the MBM (16). Up to 16 samples can be mixed simultaneously using the MBM. The mixed alloy powders were compressed to a dense cylinder sample (with a diameter of 2.05 cm and a height of 0.5 cm) at a pressure of 223 MPa for 1 min using the automatic MPM (16). Up to 16 samples can be compressed automatically by rotating samples sequentially under constant pressure.

The compressed samples were cast into alloy ingots by the automatic MEAMF (32). Each alloy sample (~10 g) was melted using a high-temperature electric arc and solidified quickly into a button ingot in a copper crucible with an underneath water cooling system. There were 32 copper crucibles in total placed in a closed glove box with water and oxygen controlled under 0.1 ppm. The whole alloy melting process was completed in an Ar atmosphere to prevent alloy oxidation. Up to 32 alloy samples can be prepared sequentially, each run by a programmed automatic zigzag scan of a tungsten electrode. The cast scan was
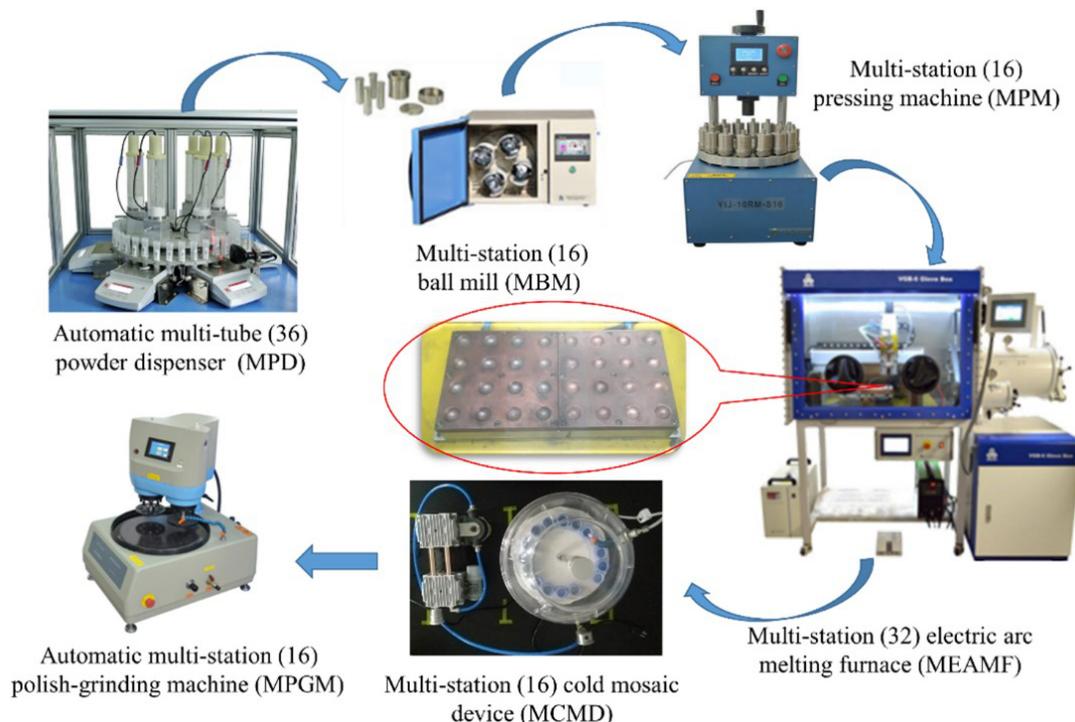
Figure 1. All-process high-throughput alloy synthesis facilities including an automatic MPD (36), a MBM (16), an automatic MPM (16), an automatic MEAMF (32), a MCMD (32) and an automatic MPGM (16).

conducted at least five times to reduce element segregation and increase composition homogeneity. The metallographic specimens were prepared by combining the MCMD (32), MWCM (8) and automatic MPGM (16). Multiple specimens can be prepared simultaneously.

The hardness of the alloy samples was measured by a Vickers hardness tester. Each sample was measured under a 294 N loading for 10 s at least five times to reduce statistic fluctuation. The root mean squared errors (RMSEs) of the measured values for a given sample were calculated, leading to a measurement uncertainty of 3.32%. The quality of materials property data is crucial for accurate ML models. In this work, the accuracy and consistency of the hardness data were achieved using the same preparation equipment and process, as well as the statistical average to minimize the experimental uncertainty due to the inhomogeneity of compositions or microstructures.

The crystal structure of the typical equimolar CoCrTiMoW HEA was identified by X-ray diffraction (XRD). The microstructures were characterized by scanning electron microscopy (SEM) combined with energy dispersive spectroscopy (EDS) and selected area electron diffraction (SAED). The SEM images provided the morphology of the microstructures, and their local chemical compositions and phases were determined by EDS and SAED.

RESULTS AND DISCUSSION
Alloy development guided by ML
In this work, the development process of hard HEAs was divided into two stages: stage I (exp-1) for data generation and model building; stage II (exp-2) for model validation and updates. Figure 2 describes the workflow of the ML-guided HTE development process.
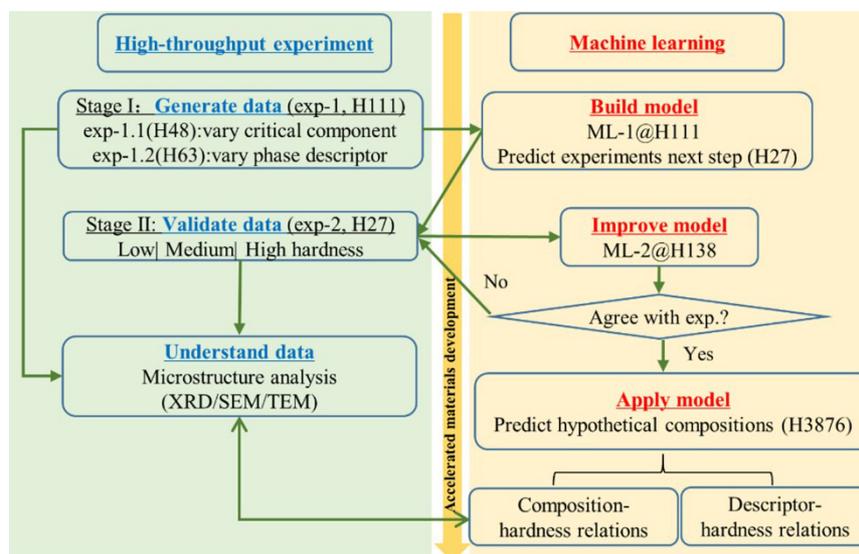
Figure 2. Workflow of ML-guided high-throughput alloy development process consisting of two stages (exp-1/ML-1 and exp-2/ML-2). H27, H111 and H3876 denote the experimental hardness datasets with the corresponding numbers of data, respectively. ML: Machine learning.

Data generation and model building (stage I)
Data generation
In stage I (exp-1), we designed the initial alloy compositions mainly for generating effective data for building the ML models. We designed the initial HEA compositions using two methods: exp-1.1 and exp-1.2. The first experimental design (exp-1.1) focused on the alloy components critical to hardness, similar to a conventional strategy. It is well known that Mo and W are refractory elements commonly used in hard alloys. Therefore, we designed 48 compositions in exp-1.1 (labeled as H48) by varying the Mo and W compositions of $Co_xCr_yTi_zMo_uW_v$ systematically, e.g., u and v vary between 0 and 1.5 at an interval of 0.25, while x, y and z are fixed at 1.

Furthermore, we also required the diversity of training data to cover a large variety of compositions and properties to include the correlations among the multiple components. Instead of the randomly scattered pick in the large five-dimension (5D) component space, we selected the compositions based on a two-dimensional (2D) descriptor map with reduced dimensions in exp-1.2. Kube *et al.*[47] showed that the valence electron concentration (VEC) and atomic radius difference (δ) can be used to classify HEA phases. We collected the data of 133 HEA alloys reported in the literature[48] and plotted their phase structures as a function of VEC and δ [Figure 3]. The various HEA alloys can be approximately divided into FCC, BCC, BCC + FCC, and amorphous phase regions in the VEC-δ map. The Mo and W compositions in exp-1.1 formed a line (yellow solid circles) in the BCC region in Figure 3. We calculated the VEC and δ of all possible compositions (labeled as H3876) of the $Co_xCr_yTi_zMo_uW_v$ systems and plotted them as black open circles in the VEC-δ map, with most of them falling into the BCC region. In exp-1.2, we then designed 63 compositions (labeled as H63), whose VEC and δ were evenly distributed (green solid circles) and covered the whole $Co_xCr_yTi_zMo_uW_v$ region in the VEC-δ map.

The HEAs with 111 compositions designed in exp-1.1 and exp-1.2 (labeled as H111) were synthesized by the HTE facilities, and their hardness was measured, as shown in Figure 4. Most of the HEAs designed by varying the Mo and W compositions (exp-1.1) have HV > 600, while those designed in the VEC-δ descriptor space (exp-1.2) have a wider hardness range of HV = 250-900, consistent with the initial design
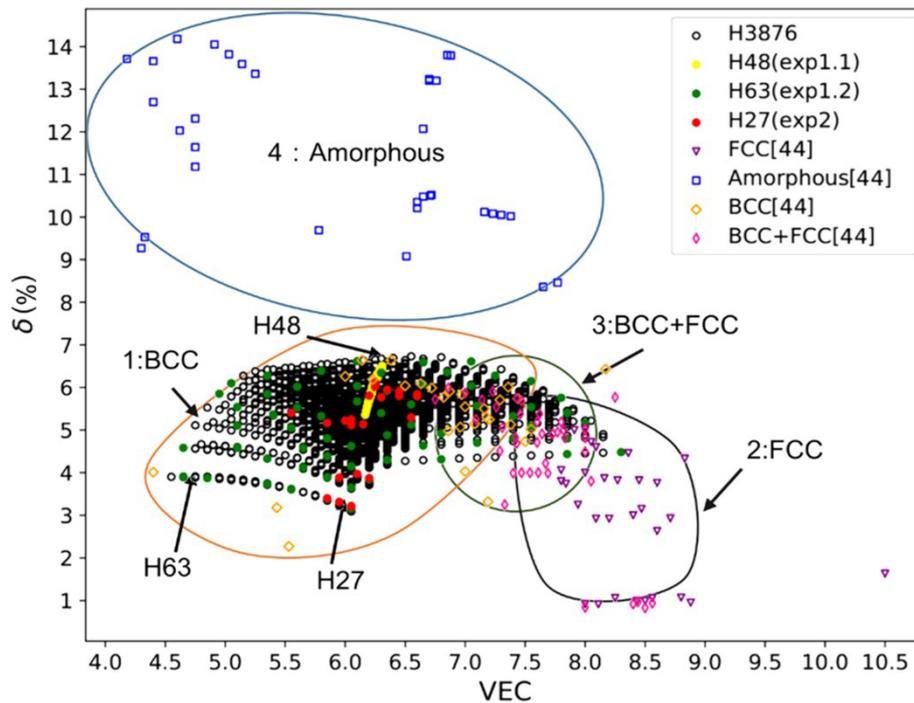
Figure 3. Various HEA phases from the literature (133 alloys)[44] and this work (138 alloys) classified by VEC and δ: BCC (region 1); FCC (region 2); BCC + FCC (region 3); amorphous (region 4). The HEAs designed in exp-1.1 (H48) are plotted as yellow solid circles, while those in exp-1.2 (H63) are green solid circles and those in exp-2 (H27) are red solid circles. The black open circles correspond to all hypothetical compositions of the $Co_xCr_yTi_zMo_uW_v$ system (H3876). HEAs: High-entropy alloys; VEC: valence electron concentration; BCC: body-centered cubic.
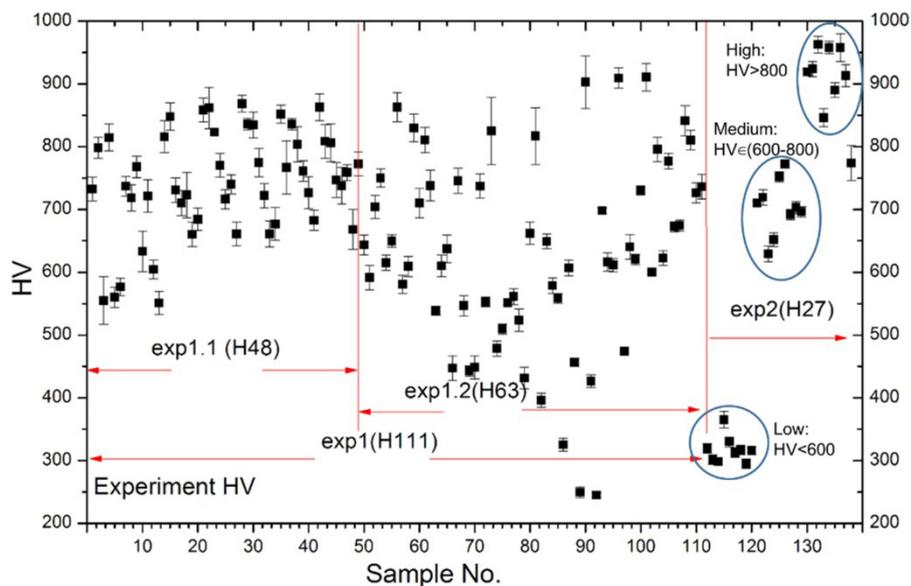
Figure 4. Vicker's hardness data (H138) of HEA $Co_xCr_yTi_zMo_uW_v$ measured at the stages of exp-1.1 (H48), exp-1.2 (H63) and exp-2 (H27). The error bars show the standard deviations among at least five measured values for each sample. HEA: High-entropy alloys.

idea. These 111 experimental data (H111) are used to build preliminary ML models in the next step.

Model building

To obtain a prediction model of hardness, we first employed a conventional statistical analysis approach, the multivariable linear regression method, to correlate the compositions of each component with the hardness (Supplementary Text 1, Supplementary Figures 1 and 2 of Supplementary Materials). However, the linear regression models were not effective for multiple component variables, as described in Supplementary Text 1 of the SM. Therefore, more advanced ML methods are necessary to reveal the complex nonlinear composition-hardness relationships of multicomponent alloys.

Next, we developed the ML prediction models using support vector machine (SVM) regression[49,50] and random forest (RF)[51] algorithms. In the SVM regression, two kernel functions were employed, including a radial basis function (RBF) and linear functions. In this work, all three ML algorithms, labeled as SVM_rbf, SVM_linear and RF, were applied using the Scikit-learn package implemented in Python[52].

The input descriptors to ML models are crucial to prediction accuracy. We considered 29 descriptors in total that were classified into three categories: (1) EMF (5): the molar ratio of the five component elements; (2) VDSHOT (6): the six important descriptors relevant to HEA phase stability as reported in the literature: the enthalpy of mixing ΔH_{mix}; the entropy of mixing ΔS_{mix}; the atom size difference δ[53]; the average melting temperature T_m; the compound parameter Ω[54] consisting of ΔS_{mix}, T_m and ΔH_{mix}; the VEC[55]; and (3) EP (18): the composition weighted sum of elementary properties[56]. To evaluate the effects of descriptor combination, we created four groups of descriptors: (1) AD (29): all descriptors; (2) EMF (5): molar ratio of components; (3) VDSHOT (6): phase structure related descriptors; and (4) SD: important descriptors selected by ML models (RF). The feature selection was determined based on the importance scores using the RF algorithm. More detailed descriptions of all descriptors and groups can be found in Supplementary Table 1 of the SM.

By combining the three ML algorithms and the four groups of descriptors, we created 12 ML methods, as shown in Table 1. Each ML method was employed ten times to generate ten ML models using different datasets split randomly for training and testing (8:2), leading to 120 ML models in total. Specifically, the experimental data were divided randomly into two parts: 80% for training and 20% for testing using different random seeds at each run. Each ML model was evaluated statistically by five-fold cross-validation on the 80% training dataset (labeled as "train"). The 20% dataset was chosen randomly and used for completely independent testing without being involved in the training process (labeled as "test").

The various ML methods and models often predicted different results, and their performance could not be readily compared using a single criterion. We needed to evaluate statistically the performance of the ML methods using multiple criteria. Moreover, reliable prediction results need to be confirmed statistically based on the scores evaluated by the multiple ML models. The multiple performance criteria, R^2, mean average error (MAE) and RMSE, were calculated and averaged over ten runs for each ML method, as shown in Figure 5A-C, respectively. Overall, the SVM_rbf method has larger R^2, smaller errors and less overfitting, though the other methods may also perform well in some cases.

Due to the intrinsic uncertainty of raw hardness data, the quantitative measures of prediction accuracy, e.g., MAE and RMSE, are less reliable compared with the qualitative correlation measure of R^2. Therefore, we selected the four best ML models from the four descriptor groups, respectively, according to the R^2 in the test datasets, and their results are shown in Figure 6A. These four best ML models had average $R^2 = 0.68$, MAE = 43 and RMSE= 77, representing the prediction accuracy of the preliminary ML-1 models based on the H111 experimental data in the first stage.

Table 1. Twelve machine learning methods combining three algorithms and four sets of descriptors used in this work

Descriptor algorithm	AD	SD	EMF	VDSHOT
SVM_rbf	SVM_rbf/AD	SVM_rbf/SD	SVM_rbf/EMF	SVM_rbf/VDSHOT
SVM_linear	SVM_linear/AD	SVM_linear/SD	SVM_linear/EMF	SVM_linear/VDSHOT
RF	RF/AD	RF/SD	RF/EMF	RF/VDSHOT

AD: All description; SD: selected descriptors; EMF: element mole fraction; VDSHOT: VEC, δ, S, H, Ω, T_m.

Model validation and finalization (stage II)
In experiment stage II (exp-2), we designed the compositions and synthesized the alloys to further validate the ML-1 models developed in stage I. By including the experimental hardness data obtained during stage II, we finalized the ML models for analysis and prediction.

Design compositions for validation (exp-2)
Given that the performance of the ML models may vary from each other, we need to apply multiple ML models statistically to ensure better prediction accuracy. We selected the top ten best ML-1 models from each of the four descriptor groups, respectively, based on the R^2 in the test datasets, leading to 40 good models in total. We designed 27 validation compositions (labeled as H27) at the high, medium and low hardness ranges (nine compositions for each range) according to the relative ranking scores evaluated by the 40 good models. Specifically, we used the 40 good ML-1 models to predict the hardness of 3876 hypothetical compositions (labeled as H3876) evenly distributed at an interval of 0.05 within the whole composition range of the $Co_xCr_yTi_zMo_uW_v$ systems. We then sorted the predicted hardness and selected nine compositions each at the low (#3867-3875), medium (#1900-1908) and large (#1-9) hardness ranges, respectively, using each model. The 40 good ML-1 models led to 360 compositions in total. Finally, the 27 compositions (H27) were chosen when they fell into the low, medium and high hardness ranges predicted by the most ML models. The HEAs at the designed 27 compositions (H27) were then synthesized using the HTE facilities, and their hardness values were measured, as shown in Figure 3 as exp-2. Indeed, the alloys synthesized in the second stage (exp-2) can be grouped into the high, medium, and low experimental hardness ranges, consistent with the predictions of the ML-1 models.

Figure 6B shows the hardness of the HEAs at the 27 validation compositions (H27) predicted by the three best individual ML-1 models selected by the R^2 in the test datasets, as well as the predicted results averaged over the 40 good ML-1 models with average R^2 = 0.97, MAE = 59 and RMSE = 72. These results validated the preliminary ML-1 models based on the H111 data.

Finalization and evaluation of ML models
We combined the experimental data from exp-1 (H111) and exp-2 (H27) to create the complete experimental dataset with the 138 hardness data (labeled as H138). We then rebuilt the 12 ML methods and 120 ML-2 models based on the complete H138 dataset. The 12 ML methods with different combinations of algorithms and descriptors were evaluated statistically, and the performance of the ML-2 models (R^2, MAE and RMSE) is plotted in Figure 5A-C (labeled with exp-2). The performance of the ML-2 models was generally better than the preliminary ML-1 models developed in the first stage (exp-1). We selected the best model according to the R^2 in the test datasets, SVM_rbf with all descriptors (AD), with R^2 = 0.88, MAE = 40 and RMSE = 56, as shown in Figure 6C. The predicted results based on the exp-2 H138 data are much better than those for exp-1 (H111) shown in Figure 6A. Furthermore, Figure 6D shows the prediction results for the 27 compositions (H27) averaged over the 40 good ML-2 models and the best individual model results based on the H138 data. The averaged performance of the 40 good ML-2 models was R^2 = 0.98, MAE = 40
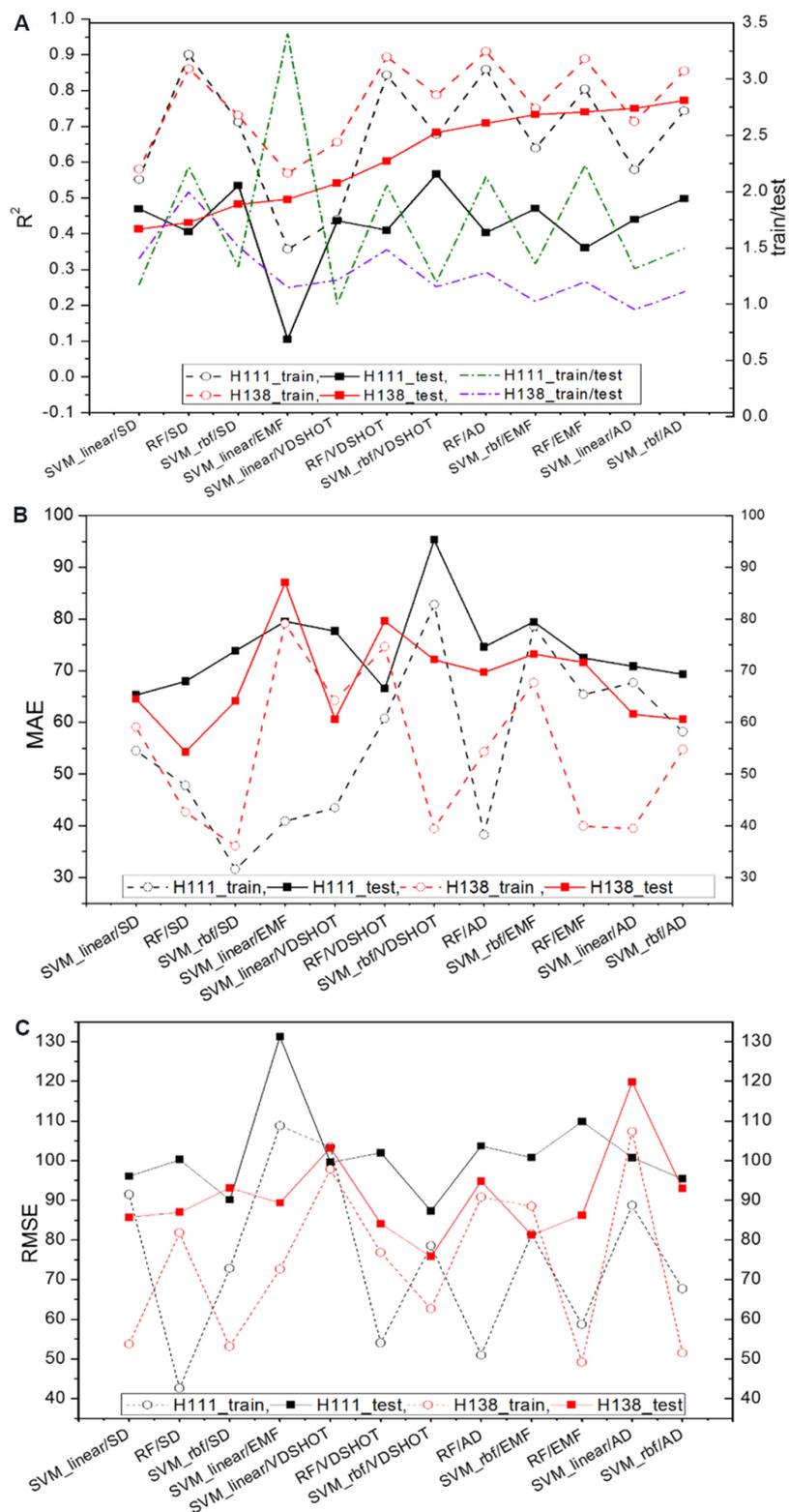
Figure 5. (A) R^2, (B) MAE and (C) RMSE of hardness predictions using 12 ML methods combining the three algorithms and the four groups of descriptors as described in the text. MAE: Mean average error; RMSE: root mean squared error; ML: machine learning.
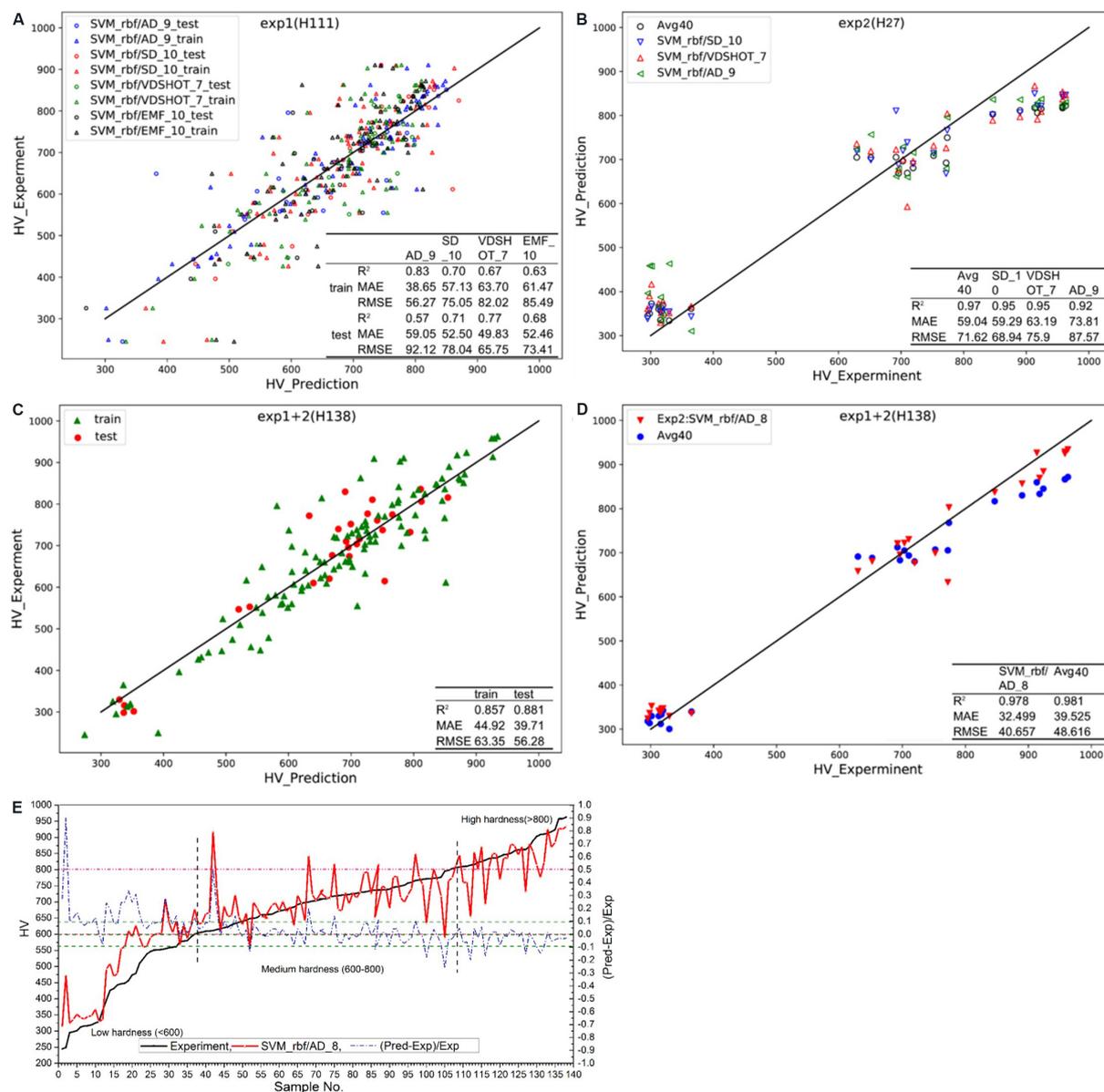
Figure 6. (A) Predicted hardness of H111 dataset by four best ML-1(H111) models from each descriptor group *vs.* experimental results (H111). (B) Predicted hardness of H27 dataset by three best ML-1(H111) models from each descriptor group and averaged 40 good ML-1(H111) models *vs.* experimental results (H27). (C) Predicted hardness of H138 dataset by best ML-2(H138) model *vs.* experimental results (H138). (D) Predicted hardness of H27 dataset by best ML-2(H138) model SVM_rbf/AD_8 (8th loop) and averaged 40 good ML-2(H138) models *vs.* experimental results (H27). (E) Predicted hardness of H138 dataset by best ML-2(H138) model SVM_rbf/AD_8 (8th loop) compared with experimental values (H138) and their relative difference. The inset tables show the statistical error analyses of the ML models. ML: Machine learning.

and RMSE = 49, again better than for the ML-1 models at exp-1 (H111) shown in Figure 6B. Therefore, we selected the SVM_rbf/AD (H138) model as the final single best model for further analysis.

Next, we evaluated the performance of the single best model SVM_rbf/AD (H138). Figure 6E shows the hardness predicted by the SVM_rbf/AD (H138) model compared with the sorted experimental results, as well as their relative difference. The hardness predicted by the ML-2 models followed the general trends of the experimental results. Specifically, the MAE/mean relative error of the ML prediction were 46.1%/5.3% at

the high hardness range (HV > 800), 43.5%/6.3% at the medium hardness range (HV = 600-800) and 63.2%/15.4% at the low hardness range (HV < 600). The deviations of ML predictions at the high and medium hardness ranges (HV > 600) were slightly larger than the experimental errors. At the low hardness range (HV < 600), the ML model always overestimated the hardness by 15.4% on average, probably due to the lack of low hardness data for training. Such systematic errors can be fixed in practice by a constant rigid shift, e.g., reducing by 15% as a correction.

Model prediction and analysis

The ultimate goal of ML is to acquire new knowledge using validated prediction models. We therefore need to learn further from the ML predictions, that is, "*learning from machine*" after "*machine learning*". After the model evaluation and validation discussed above, we applied the ML models to predict the hardness of HEAs at all hypothetical 3876 compositions (H3876) covering the full composition range of $Co_xCr_yTi_zMo_uW_v$. We can then discuss the component effects on hardness based on the composition-hardness databases generated by the ML prediction.

Composition-hardness correlation map

The prediction results often vary quantitatively among the ML models. We believe that the qualitative trends are more reliable than the quantitative absolute values. If similar qualitative trends are predicted by the several ML models, they are more likely to represent the true intrinsic tendency. On this basis, we selected the top ten best ML-2 (H138) models according to the R^2 in the test datasets using all descriptors (AD) with various algorithms and predicted the hardness corresponding to all the hypothetical 3876 compositions (H3876) to obtain the full composition-hardness relations. Supplementary Figure 3A-D show the hardness predicted by the top ten best ML-2 (H138) models from each of the four descriptor groups as functions of composition, whose component molar ratio varied systematically each at a time with an interval of 0.05. Such composition-hardness (CH) maps are the one-dimensional composition representation of the 5D component space of the HEAs. We found that the hardness in the CH maps presented the regular patterns of repeatable changes following the systematic change of each composition at a time. Seven out of the ten ML-2 (H138) models using AD descriptors predicted the qualitatively similar change patterns though they may differ quantitatively, as shown in Supplementary Figure 4, suggesting that the common change patterns represent the intrinsic tendency of component effects of the HEAs.

The hardness of all the 3876 hypothetical HEAs predicted by the single best model SVM_rbf/AD_8 presented a similar common change pattern as many others and gave the best R^2 in the test dataset compared with the experiments. Therefore, the SVM_rbf/AD_8 model predictions were used for the analysis of the component effect. Figure 7 presents the overall hardness change patterns within the full composition range and the enlarged portion of the large hardness range. Supplementary Figure 5A-G show the more detailed change patterns at the partial composition range to depict more clearly the component effects. The arrows labeled with the elements indicate the trends caused by the corresponding components. The contributions of Mo, Cr, Ti and Co to the hardness all had different optimal values and were correlated to each other, forming four curve envelopes with the maximized hardness at the optimal compositions.

Specifically, the hardness generally increased as the Mo concentration increased and the W concentration decreased. This is also consistent with Supplementary Equation (1) of the SM, where the Mo term has a positive contribution and the W term has a negative coefficient. As shown in Supplementary Figure 6A, Mo/W increased linearly with the sum of Mo and W and the hardness maximized at the ratio of Mo/W = 3-7 and Mo + W = 0.2-0.4. Furthermore, the maximum hardness values were found when Cr/Co = 1-3 and Cr + Co = 0.5-0.8, as shown in Supplementary Figure 6B. This indicates that Cr is an effective component
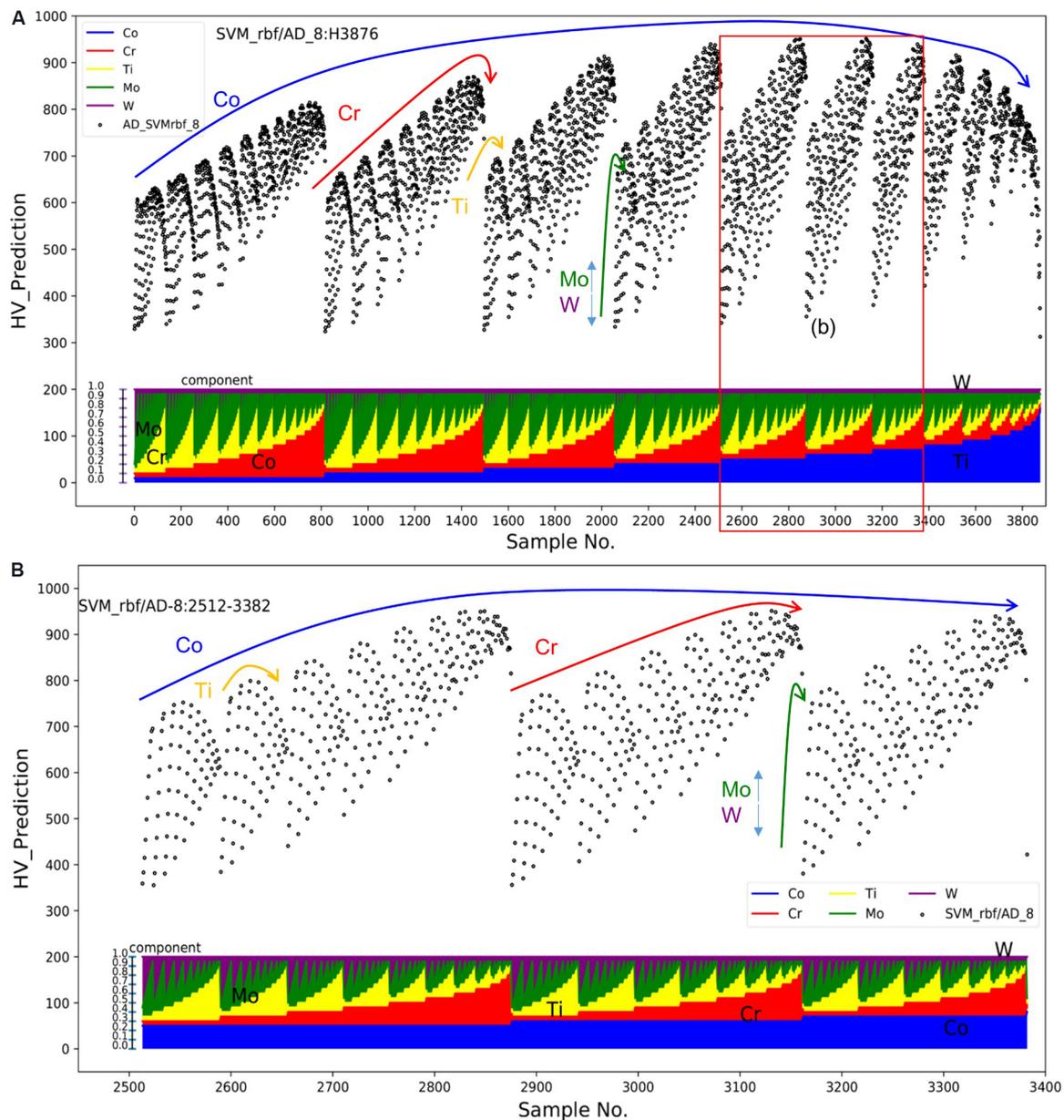
Figure 7. Hardness predicted by best ML-2 (H138) model, SVM_rbf/AD_8, for (A) all 3876 hypothetical HEAs (H3876) and (B) hardest HEAs with sample No. between 2512 and 3382. The arrows with elements indicate the trend of component effects. The insets at the bottom show the compositions of the HEAs. ML: Machine learning; HEAs: high-entropy alloys.

for increasing the hardness, consistent with the fact that the Cr term is the only positive term in Supplementary Equation (2). Combining all the component effects, we found that the hardest HEAs may have compositions close to $Co_{0.2-0.3}Cr_{0.3-0.5}Ti_{0.05-0.1}Mo_{0.2-0.35}W_{0.05}$. Generally speaking, the hard HEAs have Mo and Co slightly higher than equimolar 0.2, much more Cr, but less Ti and least W.

Table 2 (left columns) lists the alloy compositions of the top ten hardest HEAs synthesized in the two experiment stages, compared with the ML prediction for these alloys with the same compositions. Nine of the ten hardest HEAs prepared had HV > 900, larger than the reported hard cast HEAs with HV = ~850[29]. Moreover, the ten hardest HEAs predicted by the best ML model [SVM_rbf/AD_8 ML-2 (H138)] are also

Table 2. Compositions of top ten hardest HEAs with experiment hardness (left) and hardness predicted by the best ML model [ML-2 (H138) SVM_rbf/AD_8, right] with experimental and ML predicted hardness (HV)

Composition	<exp.>	ML	Composition	ML	<exp.>
$Co_{0.25}Cr_{0.35}Ti_{0.15}Mo_{0.2}W_{0.05}$	**963**	934	$Co_{0.3}Cr_{0.35}Ti_{0.05}Mo_{0.25}W_{0.05}$	951	846
$Co_{0.2}Cr_{0.4}Ti_{0.15}Mo_{0.2}W_{0.05}$	**958**	929	$Co_{0.25}Cr_{0.4}Ti_{0.1}Mo_{0.2}W_{0.05}$	950	**971**
$Co_{0.3}Cr_{0.3}Ti_{0.15}Mo_{0.2}W_{0.05}$	**958**	925	$Co_{0.25}Cr_{0.4}Ti_{0.05}Mo_{0.25}W_{0.05}$	949	865
$Co_{0.25}Cr_{0.25}Ti_{0.2}Mo_{0.25}W_{0.05}$	**924**	885	$Co_{0.25}Cr_{0.35}Ti_{0.1}Mo_{0.25}W_{0.05}$	949	**943**
$Co_{0.2}Cr_{0.25}Ti_{0.2}Mo_{0.3}W_{0.05}$	**918**	870	$Co_{0.3}Cr_{0.35}Ti_{0.1}Mo_{0.2}W_{0.05}$	948	**905**
$Co_{0.3}Cr_{0.35}Ti_{0.15}Mo_{0.15}W_{0.05}$	**913**	927	$Co_{0.25}Cr_{0.35}Ti_{0.05}Mo_{0.3}W_{0.05}$	946	876
$Co_{0.5}Cr_{0.2}Ti_{0.05}Mo_{0.1}W_{0.15}$	**911**	822	$Co_{0.3}Cr_{0.3}Ti_{0.05}Mo_{0.3}W_{0.05}$	944	888
$Co_{0.5}Cr_{0.3}Ti_{0.05}Mo_{0.05}W_{0.1}$	**909**	776	$Co_{0.3}Cr_{0.3}Ti_{0.1}Mo_{0.25}W_{0.05}$	944	**917**
$Co_{0.3}Cr_{0.55}Ti_{0.05}Mo_{0.05}W_{0.05}$	**903**	815	$Co_{0.3}Cr_{0.4}Ti_{0.05}Mo_{0.2}W_{0.05}$	943	847
$Co_{0.25}Cr_{0.25}Ti_{0.25}Mo_{0.2}W_{0.05}$	890	857	$Co_{0.2}Cr_{0.4}Ti_{0.1}Mo_{0.25}W_{0.05}$	940	**913**

HEAs: High-entropy alloys; ML: machine learning.

listed in Table 2 (right columns). For validation, we further synthesized these alloys predicted by ML and measured the hardness of at least 12 samples for each composition. The averaged results show that five of the ten predicted alloys had HV > 900, among which $Co_{0.25}Cr_{0.4}Ti_{0.1}Mo_{0.2}W_{0.05}$ had the highest hardness of HV = 971 in this work. In total, we discovered 14 hard HEAs with HV > 900 in the experiments guided by ML (values in bold in Table 2).

Moreover, we found some empirical rules for the hardest HEAs based on the analysis of the experiment compositions: (1) Co + Cr + Mo = ~0.8 and W + Ti = ~0.2; (2) Cr/W= ~5-8 and Mo/W = ~4-6, and Co/W= ~4-6. If Mo changes, we need to adjust Co and Cr to balance the total sum of Co + Cr + Mo to be ~0.8. The W content should be as low as possible and the sum of W + Ti is ~0.2 (e.g., W = 0.05 and Ti = ~0.15 or vice versa). The top ten hardest HEAs predicted by the ML-2 (H138) model also obeyed such empirical composition rules, e.g., Co + Cr + Mo = ~0.9 and W + Ti = 0.1-0.15. These composition design rules and the compositions of the hardest HEAs are consistent with the optimal composition analyses of the full CH map discussed earlier.

Descriptor-hardness correlation map
In addition to the composition-hardness relations discussed above, we projected the hardness onto the ML descriptor spaces to create descriptor-hardness (DH) maps, which have many advantages. Many descriptors have physical meanings that can help us to understand what physical properties have large influences on the hardness beyond chemical compositions. Some physical descriptors may be generally applicable to the design of other alloy systems with different elements not studied yet, though this is beyond the scope of this work. The 2D descriptor representations of the 5D component space of the HEAs are also compact and easy for visualization, as well as providing rapid analysis, prediction and design.

We made 2D DH contour maps where the two descriptors were chosen from the six important descriptors (VDSHOT in Supplementary Table 1) related to the phase classification. There are 15 DH maps with various combinations of the two out of six descriptors. The DH maps of five pairs of descriptors containing T_m with VEC, ΔH, ΔS, δ and Ω are shown in Figure 8A-E, while the rest are shown in Supplementary Figure 7 of the SM. Each type of DH map has three plots. The first one [e.g., Figure 8(x.1), x = A-E] is a plot of the experimental hardness of the H138 composition dataset. The second one [e.g., Figure 8(x.2), x = A-E] is the hardness of the same H138 dataset predicted by the single best ML-2 (H138) model, SVM_rbf/AD. The
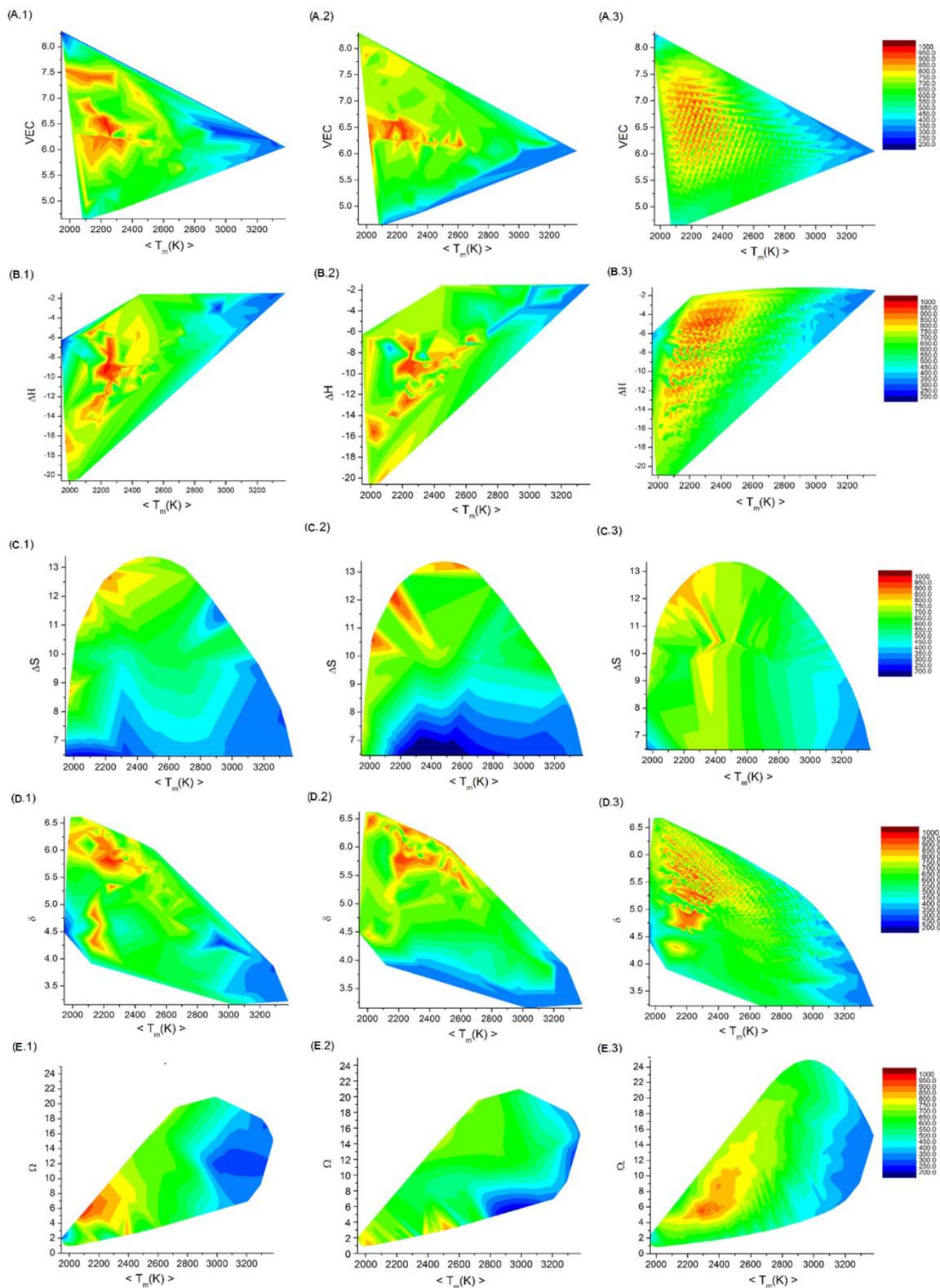
Figure 8. DH contour maps for (A.1)-(E.1) experimental hardness (H138), (A.2)-(E.2) H138 dataset and (A.3)-(E.3) H3876+H48 dataset predicted by the best ML-2 model (H138), SVM_rbf/AD_8. The five pairs of descriptors (A)-(E) are T_m with VEC, ΔH, ΔS, δ, and Ω. ML: Machine learning; VEC: valence electron concentration.

third one [e.g., Figure 8(x.3), x = A-E] is the predicted hardness values for the 3876 compositions (H3876) plus the 48 compositions designed in exp-1.1 (H48). The first two plots can be used to compare the prediction directly with the experiments for the same H138 dataset, while the third plot provides predictions on the completest H3876 + H48 dataset, including all the hypothetical alloy compositions. Both the predicted DH maps agreed reasonably well with the experimental DH maps.

Though quantitative differences may exist, the hardness distributions present qualitatively similar patterns under the 2D descriptor representations. In applications, the predictions become more reliable if confirmed by multiple DH maps. Among these DH maps, the large and low hardness regions were separated most obviously in the T_m - Ω maps [Figure 8(E.1)-(E.3)], indicating that the hardness had the most localized distributions under the T_m and Ω descriptor representations. Specifically, the largest hardness appeared when the T_m was the lowest and the Ω was slightly smaller than their median. The low T_m corresponds to the low refractory W compositions of the hardest HEAs. The small Ω may be caused by the low entropy of mixing due to the large deviation of W and Ti composition from the equimolar system in the hardest HEAs. These features of the T_m and Ω descriptors are consistent with the composition design rules discussed earlier and may be used to design hard HEAs rapidly.

MICROSTRUCTURES OF COCRTIMOW HEA

To understand the hardening mechanism of how the components contribute to the hardness, we examined the microstructures of the cast equimolar CoCrTiMoW HEA. Figure 9A-C show the XRD pattern, SEM images, EDS analysis and SAED pattern of the CoCrTiMoW alloy. Table 3 lists the chemical compositions of the dendrite trunk, interdendritic regions and precipitated phases analyzed by EDS. The fast solidification of the HEA led to the common dendrite morphology of multicomponent alloys. On the basis of the combined XRD, SEM, EDS and SAED analyses, the dendrite trunks (spot 1/2) were the BCC W-Mo solid solutions containing Cr/Ti/Co elements. The refractory components W and Mo solidified earlier than the others and first formed the dendrite due to the anisotropic crystal growth associated with the different solute diffusion behaviors. The later solidification formed the interdendritic regions mainly consisting of Co-Cr-Ti intermetallic compounds (spot 5/6) with some Co-Ti microsegregation (spot 7/8), plus a few α-Ti precipitated phases (spot 3/4).

The hardening mechanisms may be attributed to both solid-solution and second-phase strengthening. The W-Mo solid solutions are strengthened by Cr/Ti/Co element substitutions in the BCC lattice. The dislocations or cracks initiated from the W-Mo dendrites would be hindered at the interdendritic regions by the hard Co-Cr-Ti/Co-Ti intermetallic compounds or α-Ti precipitated phases. Mo and W contributed to the hardness by mainly forming the BCC dendrites. It was found that W dissolved into the interdendritic regions less than Mo, so W is less effective than Mo as the second-phase hardening element in the studied HEAs. The Cr element was distributed into both the solid solutions and the intermetallic compounds, confirming the important roles of Cr in increasing hardness, as suggested by the ML models. The Co element formed the intermetallic compounds to strengthen the interdendritic regions. The 20 at.% Ti content was obviously too much so that the excess Ti formed the pure precipitates after distributing into the solid solutions and interdendritic regions. The pure Ti precipitates with a hardness between ~120 and 160 HV may be too soft to hinder dislocations or cracks, so the Ti content should be reduced. The distribution of the alloying elements in microstructures helps to understand the roles of the components on the hardness predicted by the ML models.

CONCLUSIONS

In this work, we optimized the compositions of non-equimolar hard HEAs ($Co_xCr_yTi_zMo_uW_v$) via HTEs
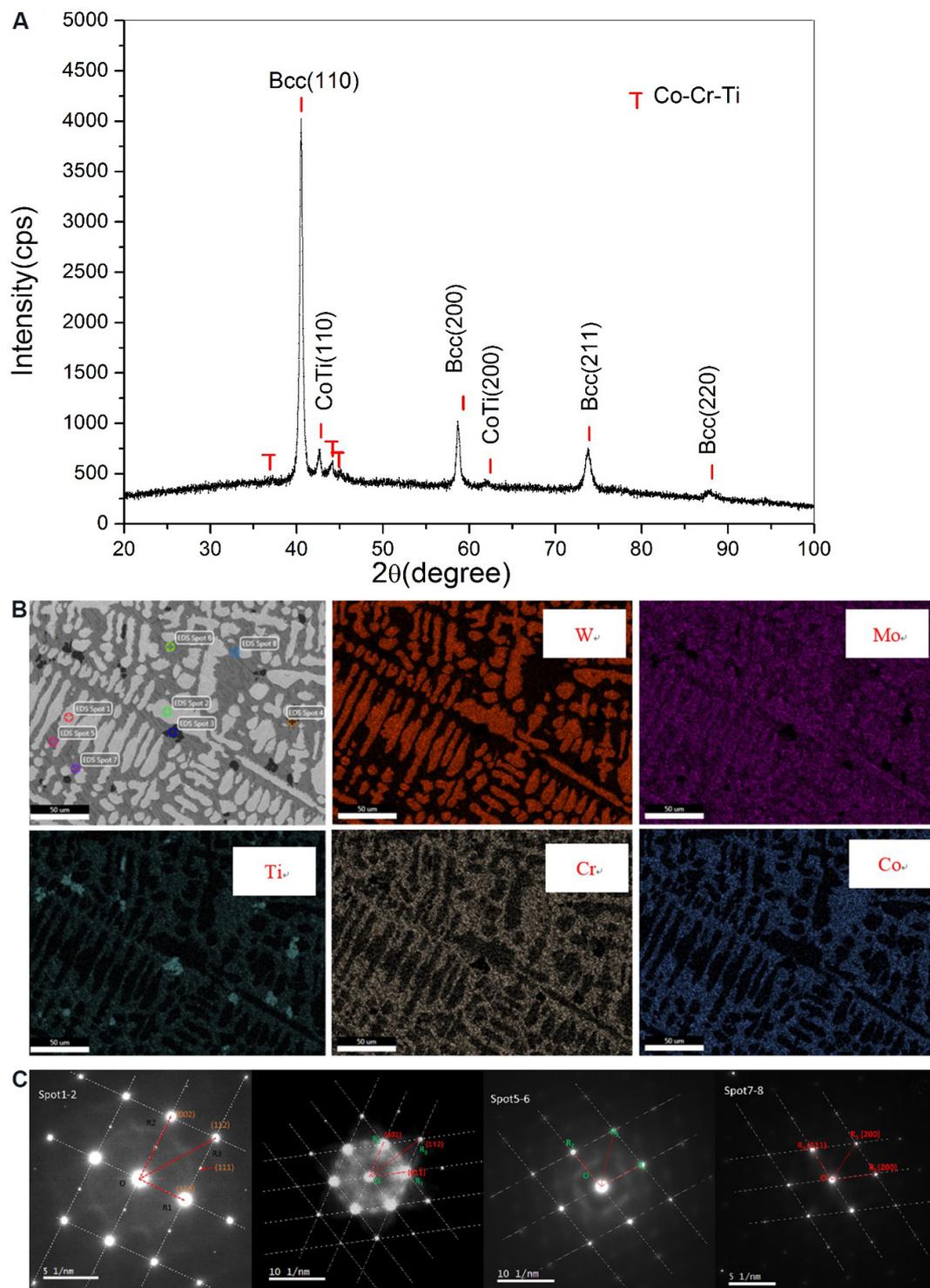
Figure 9. (A)The XRD pattern of CoCrTiMoW as cast. (B) BSE micrograph and EDS analysis of CoCrTiMoW as cast. (C) Selected area diffraction pattern with the spot corresponding to BSE section. Spot 1-2 shows the BCC phase, Spot 3-4 shows the α-Ti, spot 5-6 unknown, spot 7-8 shows CoTi. XRD: X-ray diffraction; EDS: energy dispersive spectroscopy; BCC: body-centered cubic.

guided by the predictions of ML. A two-stage process was proposed to develop hard alloys. In the first stage, we designed the HEA compositions by varying the critical elements (48 compositions) and using important

Table 3. Chemical compositions of different regions of CoCrTiMoW analyzed by EDS (at.%)

Element	Spot 1	Spot 2	Spot 3	Spot 4	Spot 5	Spot 6	Spot 7	Spot 8
W	41.31	41.97	0.13	0.13	2.15	2.23	0.5	0.55
Mo	26.74	27.79	0.1	0.11	8.45	8.3	2.52	2.73
Ti	12.08	12.7	99.1	99.15	23.33	23.41	40.7	39.38
Cr	14.5	15.82	0	0	30.59	30.56	6.94	8.3
Co	5.37	1.73	0.66	0.62	35.47	35.51	49.34	49.04

EDS: Energy dispersive spectroscopy.

phase descriptors (63 compositions). Based on these 111 experimental data, ML-1 models were constructed using supporting vector machine and random forest algorithms, guiding the design of 27 compositions in the second experiment stage. Finally, ML-2 models were constructed using all 138 experimental data, predicting the hardness with the errors comparable with the experimental ones. In total, we synthesized 14 superhard HEAs (~10% of real experiments) with HV > 900 via ML-guided HTEs.

Moreover, the ML models were used to predict the hardness of hypothetical 3876 alloys covering the whole composition range of the five-component alloys, revealing the complete composition-hardness correlations. The hardness distributions of the 5D component space were then projected onto 2D descriptor space to reveal the important physical descriptors based on descriptor-hardness correlations. The hardening mechanisms can be elaborated by both solid-solution and second-phase strengthening according to the typical cast microstructures of equimolar CoCrTiMoW HEA: BCC Mo-W dendritic solid solutions and interdendritic regions consisting of Co-Cr-Ti/CoTi intermetallic compounds and Ti precipitates.

In summary, the complete composition-hardness relationships were obtained by carrying out ~1/28 experiments of the full composition space guided by ML. Such a data-driven experimental development process was further accelerated by more than tenfold using all-process high-throughput alloy synthesis equipment. This work demonstrated that all-process high-throughput experiments guided by ML methods could accelerate the composition optimization of multicomponent alloys by several hundred times at a fraction of the cost. A single target material property was used in this work to prove the concept of the ML-HTE design strategy, but a similar methodology can be extended to optimize multiple target material properties and processing conditions if a large amount of high-quality experimental data are available.

DECLARATIONS

Authors' contributions
Made substantial contributions to conception, supervision, and design of the study and performed data analysis and interpretation as well as manuscript editing and review: Liu Y
Performed major experiments and machine learning modeling and data analysis and interpretation as well as draft writing: Wang J
Performed the part of machine learning modeling and data analysis: Xiao B
Performed the part of verification experiments and data analysis: Shu J

Availability of data and materials
The supplementary materials are available including machine learning descriptors, linear regression results, composition-hardness maps, and descriptor-hardness maps. The source codes of the machine learning models together with the experimental and predicted data are deposited onto the repository of Github (https://github.com/xiaobin9107/CoCrTiMoW-HEA-data-and-ML-scripts) for free download.

Conflicts of interest
All authors declared that there are no conflicts of interest.

Ethical approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Copyright

REFERENCES

1. Xiang XD, Sun X, Briceño G, et al. A combinatorial approach to materials discovery. *Science* 1995;268:1738-40. DOI PubMed
2. Xiang X, Wang G, Zhang X, Xiang Y, Wang H. Individualized pixel synthesis and characterization of combinatorial materials chips. *Engineering* 2015;1:225-33. DOI
3. Wang J, Yoo Y, Gao C, et al. Identification of a blue photoluminescent composite material from a combinatorial library. *Science* 1998;279:1712-4. DOI PubMed
4. Zhao L, Zhou Y, Wang H, et al. High-throughput synthesis and characterization of a combinatorial materials library in bulk alloys. *Metall Mater Trans A* 2021;52:1159-68. DOI
5. Liu Y, Hu Z, Suo Z, et al. High-throughput experiments facilitate materials innovation: a review. *Sci China Technol Sci* 2019;62:521-45. DOI
6. Bligaard T, Jóhannesson GH, Ruban AV, Skriver HL, Jacobsen KW, Nørskov JK. Pareto-optimal alloys. *Appl Phys Lett* 2003;83:4527-9. DOI
7. Reed R, Tao T, Warnken N. Alloys-by-design: application to nickel-based single crystal superalloys. *Acta Mater* 2009;57:5898-913. DOI
8. Jong M, Chen W, Geerlings H, Asta M, Persson KA. A database to enable discovery and design of piezoelectric materials. *Sci Data* 2015;2:150053. DOI PubMed PMC
9. Ong SP, Wang L, Kang B, Ceder G. Li-Fe-P-O$_2$ phase diagram from first principles calculations. *Chem Mater* 2008;20:1798-807. DOI
10. Ong SP, Richards WD, Jain A, et al. Python materials genomics (pymatgen): a robust, open-source python library for materials analysis. *Comput Mater Sci* 2013;68:314-9. DOI
11. Brun F, Yoshida T, Robson J, Narayan V, Bhadeshia H, Mackay D. Theoretical design of ferritic creep resistant steels using neural network, kinetic, and thermodynamic models. *Mater Sci Technol* 2013;15:547-54. DOI
12. Sourmail T, Bhadeshia HKDH, Mackay DJC. Neural network model of creep strength of austenitic stainless steels. *Mater Sci Technol* 2013;18:655-63. DOI
13. Joo M, Ryu J, Bhadeshia HKDH. Domains of steels with identical properties. *Mater Manuf Process* 2009;24:53-8. DOI
14. Khalaj G, Nazari A, Pouraliakbar H. Prediction of martensite fraction of microalloyed steel by artificial neural networks. *NNW* 2013;23:117-30. DOI
15. Khalaj G. Artificial neural network to predict the effects of coating parameters on layer thickness of chromium carbonitride coating on pre-nitrided steels. *Neural Comput & Applic* 2013;23:779-86. DOI
16. Faizabadi MJ, Khalaj G, Pouraliakbar H, Jandaghi MR. Predictions of toughness and hardness by using chemical composition and

tensile properties in microalloyed line pipe steels. *Neural Comput & Applic* 2014;25:1993-9. DOI

17. Pouraliakbar H, Khalaj M, Nazerfakhari M, Khalaj G. Artificial neural networks for hardness prediction of HAZ with chemical composition and tensile test of X70 pipeline steels. *J Iron Steel Res Int* 2015;22:446-50. DOI

18. Narimani N, Zarei B, Pouraliakbar H, Khalaj G. Predictions of corrosion current density and potential by using chemical composition and corrosion cell characteristics in microalloyed pipeline steels. *Measurement* 2015;62:97-107. DOI

19. Conduit B, Jones N, Stone H, Conduit G. Design of a nickel-base superalloy using a neural network. *Mater Des* 2017;131:358-65. DOI

20. Conduit B, Jones N, Stone H, Conduit G. Probabilistic design of a molybdenum-base alloy using a neural network. *Scr Mater* 2018;146:82-6. DOI

21. Conduit B, Illston T, Baker S, et al. Probabilistic neural network identification of an alloy for direct laser deposition. *Mater Des* 2019;168:107644. DOI

22. Islam N, Huang W, Zhuang HL. Machine learning for phase selection in multi-principal element alloys. *Comput Mater Sci* 2018;150:230-5. DOI

23. Arisoy YM, Özel T. Machine learning based predictive modeling of machining induced microhardness and grain size in Ti-6Al-4V alloy. *Mater Manuf Process* 2014;30:425-33. DOI

24. Khalaj G, Pouraliakbar H. Computer-aided modeling for predicting layer thickness of a duplex treated ceramic coating on tool steels. *Ceram Int* 2014;40:5515-22. DOI

25. Isayev O, Oses C, Toher C, Gossett E, Curtarolo S, Tropsha A. Universal fragment descriptors for predicting properties of inorganic crystals. *Nat Commun* 2017;8:15679. DOI PubMed PMC

26. de Jong M, Chen W, Notestine R, et al. A statistical learning framework for materials science: application to elastic moduli of k-nary inorganic polycrystalline compounds. *Sci Rep* 2016;6:34256. DOI PubMed PMC

27. Correa-baena J, Hippalgaonkar K, van Duren J, et al. Accelerating materials development via automation, machine learning, and high-performance computing. *Joule* 2018;2:1410-20. DOI

28. Cantor B, Chang I, Knight P, Vincent A. Microstructural development in equiatomic multicomponent alloys. *Materials Science and Engineering: A* 2004;375-377:213-8. DOI

29. Yeh J, Chen S, Lin S, et al. Nanostructured high-entropy alloys with multiple principal elements: novel alloy design concepts and outcomes. *Adv Eng Mater* 2004;6:299-303. DOI

30. Ranganathan S. Alloyed pleasures: multimetallic cocktails. *Curr Sci* 2003;85:1404-6.

31. Fazakas É, Zadorozhnyy V, Varga L, et al. Experimental and theoretical study of $Ti_{20}Zr_{20}Hf_{20}Nb_{20}X_{20}$ (X=V or Cr) refractory high-entropy alloys. *Int J Refract Metals Hard Mater* 2014;47:131-8. DOI

32. Liu Y, Zhang Y, Zhang H, et al. Microstructure and mechanical properties of refractory $HfMo_{0.5}NbTiV_{0.5}$Six high-entropy composites. *J Alloys Compd* 2017;694:869-76. DOI

33. Chang C, Titus MS, Yeh J. Oxidation Behavior between 700 and 1300 °C of refractory TiZrNbHfTa high-entropy alloys containing aluminum. *Adv Eng Mater* 2018;20:1700948. DOI

34. Cao P, Ni X, Tian F, Varga LK, Vitos L. Ab initio study of $Al_xMoNbTiV$ high-entropy alloys. *J Phys Condens Matter* 2015;27:075401. DOI PubMed

35. Senkov O, Wilks G, Miracle D, Chuang C, Liaw P. Refractory high-entropy alloys. *Intermetallics* 2010;18:1758-65. DOI

36. Senkov O, Wilks G, Scott J, Miracle D. Mechanical properties of $Nb_{25}Mo_{25}Ta_{25}W_{25}$ and $V_{20}Nb_{20}Mo_{20}Ta_{20}W_{20}$ refractory high entropy alloys. *Intermetallics* 2011;19:698-706. DOI

37. Zou Y, Ma H, Spolenak R. Ultrastrong ductile and stable high-entropy alloys at small scales. *Nat Commun* 2015;6:7748. DOI PubMed PMC

38. Wu Y, Cai Y, Chen X, et al. Phase composition and solid solution strengthening effect in TiZrNbMoV high-entropy alloys. *Mater Des* 2015;83:651-60. DOI

39. Melnick A, Soolshenko V. Thermodynamic design of high-entropy refractory alloys. *J Alloys Compd* 2017;694:223-7. DOI

40. Han Z, Luan H, Liu X, et al. Microstructures and mechanical properties of Ti NbMoTaW refractory high-entropy alloys. *Materials Science and Engineering: A* 2018;712:380-5. DOI

41. Zhang W, Liaw PK, Zhang Y. Science and technology in high-entropy alloys. *Sci China Mater* 2018;61:2-22. DOI

42. Dasari S, Chaudhary V, Gwalani B, et al. Highly tunable magnetic and mechanical properties in an $Al_{0.3}CoFeNi$ complex concentrated alloy. *Materialia* 2020;12:100755. DOI

43. Chaudhary V, Chaudhary R, Banerjee R, Ramanujan R. Accelerated and conventional development of magnetic high entropy alloys. *Mater Today* 2021;49:231-52. DOI

44. Sun Y, Lu Z, Liu X, et al. Prediction of Ti-Zr-Nb-Ta high-entropy alloys with desirable hardness by combining machine learning and experimental data. *Appl Phys Lett* 2021;119:201905. DOI

45. Wen C, Wang C, Zhang Y, et al. Modeling solid solution strengthening in high entropy alloys using machine learning. *Acta Mater* 2021;212:116917. DOI

46. MTI Corporation. Available from: https://www.mtixtl.com/ [Last accessed on 23 Mar 2022].

47. Kube SA, Sohn S, Uhl D, Datye A, Mehta A, Schroers J. Phase selection motifs in High Entropy Alloys revealed through combinatorial methods: Large atomic size difference favors BCC over FCC. *Acta Mater* 2019;166:677-86. DOI

48. Gao MC, Yeh JW, Liaw PK, Zhang Y. High-entropy alloys: fundamentals and applications. Switzerland: Springer International

Publishing; 2016. DOI

49.   Müller KR, Mika S, Rätsch G, Tsuda K, Schölkopf B. An introduction to kernel-based learning algorithms. *IEEE Trans Neural Netw* 2001;12:181-201. DOI PubMed

50.   Chang C, Lin C. LIBSVM: a library for support vector machines. *ACM Trans Intell Syst Technol* 2011;2:1-27. DOI

51.   Breiman L. Random forests. *Mach Learn* 2001;45:5-32. DOI

52.   Scikit-learn in Python. Available from: https://scikit-learn.org/stable/ [Last accessed on 23 Mar 2022].

53.   Zhang Y, Zhou Y, Lin J, Chen G, Liaw P. Solid-solution phase formation rules for multi-component alloys. *Adv Eng Mater* 2008;10:534-8. DOI

54.   Yang X, Zhang Y. Prediction of high-entropy stabilized solid-solution in multi-component alloys. *Mater Chem Phys* 2012;132:233-8. DOI

55.   Guo S, Ng C, Lu J, Liu CT. Effect of valence electron concentration on stability of FCC or BCC phase in high entropy alloys. *J Appl Phys* 2011;109:103505. DOI

56.   Database on properties of chemical elements. Available from: http://phases.imet-db.ru/elements/main.aspx [Last accessed on 23 Mar 2022].