High-throughput and data-driven machine learning techniques for discovering high-entropy alloys

Check for updates

Lu Zhichao[1,2], Ma Dong[2], Liu Xiongjun[1,3] ✉ & Zhaoping Lu [1,3] ✉

High-entropy alloys (HEAs) have attracted extensive attention in recent decades due to their unique chemical, physical, and mechanical properties. An in-depth understanding of the structure–property relationship in HEAs is the key to the discovery and design of new compositions with desirable properties. Related to this, materials genome strategy has been increasingly used for discovering new HEAs with better performance. This review paper provides an overview of key advances in this fast-growing area, along with current challenges and potential opportunities for HEAs. We also discuss related topics, such as high-throughput preparation, characterization, and computation of HEAs, and data-driven machine learning for accelerating alloy development. Finally, future research directions and perspectives for the materials genome-assisted design of HEAs are proposed and discussed.

High-entropy alloys (HEAs), also called multi-principal element alloys[1–3], are chemically disordered but topologically ordered with the formation of random solid-solution (SS) structures, such as face-centered cubic (FCC), body-centered cubic (BCC), or hexagonal-close-packed (HCP). Understanding the composition–structure–properties relationship has long been a topic of great interest in HEAs. Thus, extensive studies have been carried out on various HEAs, and many attractive properties have been achieved in the last two decades. These properties include good plasticity, high strength and hardness, outstanding high-temperature-softening resistance, and unique electrical and magnetic properties. In the past few years, besides metallic systems, high entropy materials have expanded to ceramics made of carbides, borides, or nitrides of IV and V group transition metals, which have remarkable properties[4–6]. Due to these unique properties and large composition space, high entropy materials have promising potential applications under extreme conditions, such as, in high-temperature structural components, corrosion-resistant parts, coatings, and nuclear materials[7].

However, with regard to the property-oriented designs of HEAs, some challenges remain to be solved. (1) Owing to the chemically disordered structure, HEAs are not necessarily equimolar compositions; that is, many potential elements in the periodic table can conceivably be incorporated into HEAs via microalloying or principal element substitution. Therefore, an essentially infinite number of HEAs are available. Since the compositions of HEAs can be continuously adjustable, the properties of interest can be optimized. Conceptually, this poses a serious challenge—How can potential HEAs with properties of interest be fine-tuned efficiently in such a large composition space rather than in a conventional "trial and error" manner[8]?

(2) Coupled with the fact that fully understanding the complicated interplay between constituents and properties is a prerequisite when designing new HEAs, How can the intrinsic relationship in a vast and complex database be uncovered? To date, inspired by the Materials Genome Initiative (MGI), high-throughput techniques (preparation, characterization, and calculation) and the data-driven machine learning (ML) method have been adopted by synergistically combining experiment, theory, and computation in a tightly integrated and high-throughput manner, and to predict and optimize HEAs at an unparalleled scale and in an effective way[9]. These tools can be used to screen extensive composition space for a desired property and simultaneously pinpoint specific alloys with the desired properties. Specifically, high-throughput techniques are able to bridge the gap between experiments and ML modeling; that is, high-throughput approaches can provide valuable materials information for the following ML, and vice versa, ML can provide intelligent feedback to the experiments[10–12]. Through continuing efforts to integrate experiment, computation, and data-driven ML, the underlying structure–property relationships to the materials genome can be revealed and thus seed a new generation of advanced HEAs[13].

This review aims to present a brief state-of-the-art overview of the materials genome strategy (MGS) applied in HEAs and provide a timely focus on key developments, including challenges and opportunities, in this interdisciplinary area. Specifically, we will give a brief introduction to the development of HEAs and the application of MGI in this field. Additionally, some challenges will also be listed in a brief manner in "Introduction". In section "High-throughput preparation and characterization of HEAs", the main high-throughput preparation and characterization techniques for

[1]Beijing Advanced Innovation Center for Materials Genome Engineering, State Key Laboratory for Advanced Metals and Materials, University of Science and Technology Beijing, Beijing, China. [2]Songshan Lake Materials Laboratory, Dongguan, China. [3]Institute for Materials Intelligent Technology, Liaoning Academy of Materials, Shenyang, China. ✉e-mail: xjliu@ustb.edu.cn; luzp@ustb.edu.cn

HEAs will be discussed in detailed and critical issues needed to be solved will also be proposed. In section "High-throughput computing for HEAs", we will present and discuss applications of high-throughput computation method in accelerating the development of HEAs. An in-depth discussion about data-driven ML strategy for HEAs will be provided in section "Data-driven machine learning strategies ". Finally, in "Outlook" section, we will give an outlook of potential research activities to be exploited and main scientific challenges to be addressed in the future. The core purpose underlying the brief review is to provide an important opportunity to advance the understanding of MGS employed in HEAs and to offer researchers a platform to foster new ideas.

High-throughput preparation and characterization of HEAs

The design of HEAs poses a significant challenge when exploring the phase structure and desirable properties through the vast potential multi-component compositional space available[14]. As such, unconventional high-throughput preparation techniques are crucially important, particularly for effectively narrowing down the alloys in a wide composition space. Among these, HEAs exploit a variety of preparation techniques, such as, combinatorial thin film deposition, laser additive manufacturing (LAM), rapid alloying prototype, diffusion multiples, and those based on welding. In what follows, we will give an overview of the different high-throughput techniques that were used to prepare multi-component HEAs and point out some critical issues that needed to be resolved.

High-throughput preparation techniques for HEAs

LAM. Combinatorial LAM endows the process with both high heating and cooling rates, and has been used as an efficient method for the synthesis of HEAs. Among various LAM methods, laser metal deposition (LMD) is the preferred technique used to make HEA combinational libraries. During the LMD process, the feedstock nozzles convey the raw material powder to a rapidly moving melt pool formed by a laser through an inert gas flow. Apparently, LMD is more suitable for high-throughput synthesis owing to the advantage of its real-time and variable feeding system, which applies two or more hoppers with different powder feeders to permit changes in the deposited powder compositions[15–21].

Combinatorial laser deposition of compositionally graded complex alloys has been regarded as an attractive approach for assessing the composition–microstructure–property relationships of HEAs. LMD is quite capable of synthesizing refractory HEAs that are difficult to make[19]. Melia et al. prepared a MoNbTaW alloy system by additive manufacturing with commercial refractory elemental powders, which have good spherical morphology, leveraging the additive manufacturing process and mechanical testing to enable rapid alloy exploration, as shown in Fig. 1. In the steady state, there was an evident linear spatial trend in the composition and a significantly variation of hardness, with composition dominated by solution strengthening (Fig. 1d)[19]. Compared to other mechanical properties (i.e., strength, plasticity, toughness, etc.), hardness is the simplest one that can be obtained effectively by mechanical testing automatically in areas with different compositions of small samples. In view of the hardness–strength relationship ($H_v \approx \frac{3\sigma_y (MPa)}{9.81}$)[22], hardness allows for indirect and efficient evaluations of mechanical properties.

Borkar et al. studied the compositionally graded $Al_xCrCuFeNi_2$ ($0 < x < 1.5$) HEAs produced by laser deposition from a blend of elemental powders, using a double powder feeder with two hoppers containing $CrCuFeNi_2$ and $Al_2CrCuFeNi_2$ powders, respectively. The sample of a cylindrical geometry was deposited with a smooth change of alloy composition in height[15]. Additionally, an identical laser deposition processing method, laser-engineered net shaping (LENS), was also applied to construct the compositional and microstructural libraries of $Al_xCoCrFeNi$ in a high-throughput manner[18]. The discrepancy between LENS and the above-mentioned case was that the substrate (CoCrFeNi plate) for LENS was priorly made by an arc-melting and copper mold-casting method, while in

Borkar's work, a blend of powders of a nominal composition of $CrCuFeNi_2$ was used. During the LENS process, the laser power and moving speed remained unchanged, and the feeding rate of Al powder for each monolayer patch increased in certain increments. The entire deposition process includes the addition of Al and two subsequent remelting processes perpendicular to the deposition direction, to improve the mixing and compositional homogeneity of the alloyed region[18].

In fact, the design and parameter adjustment of the LMD process has an important effect on sample preparation. For example, the substrate greatly influences the composition and microstructure of the deposited alloys, which can be improved by increasing the stack thickness or a reasonable experimental design. The former will not only increase the preparation cost, but will also affect the microstructure uniformity. Selecting the main component of the alloy as the substrate material, depositing the sample in the thickness direction with less affection for the substrate, and a controlled composition gradient could form a reasonable experimental design[17,23].

Combinatorial deposition of thin film materials libraries. Combinatorial thin film synthesis by sputtering using multiple deposition sources is a state-of-the-art route for constructing of materials libraries that are composed of a wide range of gradually changed alloy compositions[24,25]. Continuous preparation of multiple gradients can be achieved by adjusting the processing parameters, such as the compositions of the targets, the power and angle of each gun, and the material and rotation of the substrate. For HEAs, several approaches based on sputtering have been employed for alloy design by tweaking these parameters[25–30]. Since HEAs contain more than three principal elements, the preparation of thin film samples is an important part of high-throughput experiments. The composition library can be prepared by stacking multilayers, which were deposited by coordinating the single sputtering source with a removable mask. Due to the characteristics of the system alloys, the homogenization process requires a higher temperature with a long heat treatment duration. The substrate material and the experimental process are desired for higher requirements.

The co-sputtering method, that is, using several targets to sputter simultaneously to obtain a composition library through the adjustment of processing parameters, is an ideal way to realize the compositional design of HEAs. However, the number of targets available is often insufficient to meet the number of constituent elements. Considering the mechanical coordination and characterization accuracy, this issue can be effectively resolved by using targets of multiple elements. At present, the synthesis of combinatorial films is widely used in the preparation of a sample library of HEAs. Ludwig et al. reported quinary Ru–Rh–Pd–Ir–Pt composition spread thin-film materials libraries that were co-sputtered from a load-lock-equipped combinatorial magnetron sputtering system. Fig. 2a shows the coverage of a ternary and a quaternary library, and Fig. 2b illustrates co-deposition from five deposition sources and compositional gradients of a co-sputtered quinary materials library. Six composition spread materials libraries were synthesized via target arrangement permutations, each of which comprises composition gradients in different subsections of the quinary composition space. Forty percent of all possible Ru-Rh-Pd-Ir-Pt HEA compositions (defined by 10–35 at.% variations of individual elements) were covered by the materials libraries[31]. Additionally, Zhang et al. reported the Ti–Al–Cr–Fe–Ni composition library prepared via magnetron co-sputtering using Al, Ti, and CrFeNi targets[29]. They considered that the Cr, Co, and Fe atomic sizes are similar, and the inclusion of two magnetic components can avoid the interference of the magnetic field. Elements with great differences in sputtering power cannot be placed on the same target simultaneously. Therefore, the components in the alloy are divided into Al, Ti, and CrFeNi targets, with a 1:1:1 composition ratio of the CrFeNi target. A similar method is also applied to the (Cr, Fe, V)–(Ta, W) alloy system[28], which is regarded as a pseudo-binary alloy composed of transition group elements Cr, Fe, V, and refractory elements Ta, W. Samples fabricated using

Fig. 1 | Analysis of the additive manufacturing processed $(MoTaW)_x(Nb)_{1-x}$ compositionally graded part cross-section. a An optical image. **b** {100} Pole figure oriented parallel to the build direction with maximum intensity of 5.19 MRD. **c** IPF map. **d** The composition and hardness gradients along the height of the part. The arrows in **d** show the two axes of the hardness data[19] (adapted with permission from ref. 19. Copyright 2020 Elsevier).
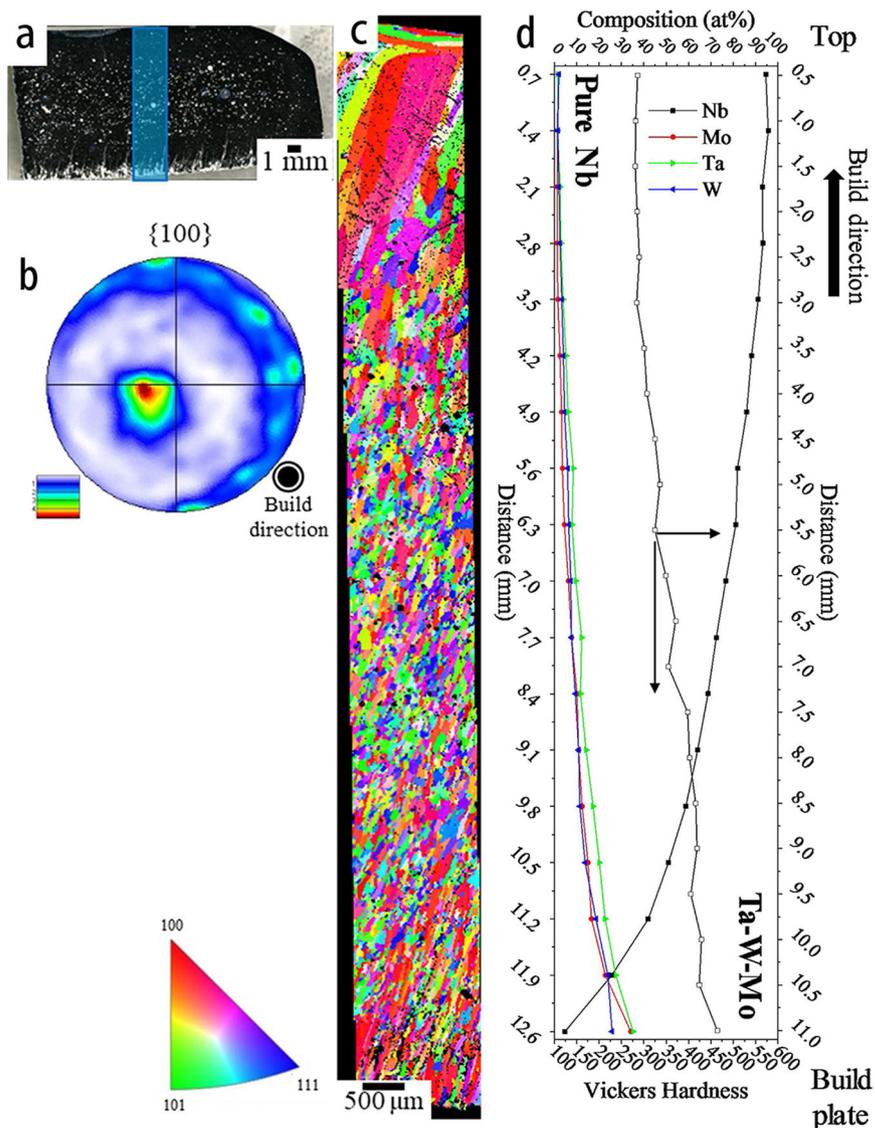
this method possess a large compositional range, with the interdependence of composition changes at the same time.

As mentioned above, the number of targets required for the synthesis of HEAs can be met by either mixing the elements or using a single target made of separate element segments. How to achieve the alloy composition gradient is hence an important component of a high-throughput experiment, which will greatly affect the subsequent characterization. A materials library composed of continuously changed alloy compositions can be achieved by adjusting the angle of the sputtering guns with respect to the substrate normally kept at a certain angle. Patches of alloy units closer to a particular source exhibit a higher concentration of the corresponding elements, with a lower density of the elements in patches farther away. The co-sputtering of multiple element targets can, in principle, make up a complete composition space of the system, preparing multiple alloy compositions in one experiment[25,26,30]. Moreover, the composition gradient of elements was controlled by tuning the sputtering power in some actual cases due to the differences in deposition efficiency and the desired composition range of the constituted elements. In the development of HEAs, the change of a single element in content has a great effect on the formation of solid solutions, which depends on the power of the single sputtering target. Marshal et al. reported the influence of Al content (from 3.5 to 54 at.%) on the phase

formation and magnetic properties in the FeMnCoCrAl thin film libraries[32]. Five compositionally graded FeMnCoCrAl thin films were deposited to study the influence of Al by varying the power density from 2 to 7.5 W/cm^2 to obtain multiple Al concentration gradients, which facilitated the investigation of the effect of Al. The sample also contains both Co and Cr component gradients, and multiple explications can be obtained through nondestructive characterization techniques.

Since the thin film samples obtained by sputtering exhibit a smooth composition change spatially, it is necessary to divide the composition into several sample units to facilitate subsequent characterization[26,30,33,34]. One way is to directly cover the substrate with a physical metal mask of micron thickness during deposition to generate sample units[26,35]. An improved method is the use of a micro-machined Si deposition mask to mitigate deposition-shadowing effects by the creation of angled walls[36]. The sample units provide enough workspace for the subsequent characterization, which improves the accuracy of the experiment and reduces the difficulty of operation. In addition, sample units can be partitioned by the proper operation of the substrate. The substrate was divided into small squares by laser beam cutting to facilitate the characterization of the material library prior to deposition[28,29]. Cutting is done on the back side of the substrate to mark the location of each sample, and semipermeable cutting is required to ensure the integrity of the front surface of the substrate.

High-throughput diffusion multiples. A diffusion multiple exposes three or more different metal blocks to high temperatures and ensures their close interfacial contact to create solid solutions or intermetallic compounds by thermal diffusion, allowing for the effective evaluation of the phase formation kinetics and several materials phenomena. Unlike traditional diffusion couples and diffusion triples[37], diffusion multiples are usually prepared via hot isostatic pressing, by which different multiples can be assembled into a single sample. In recent years, Zhao et al. have conducted preliminary studies on high-throughput diffusion multiples[38–41] and the technique has recently been extended to the development of HEAs[42–45]. To reduce the interference of other factors on the diffusion of elements, the synthesis of diffusion multiple assemblies has more stringent requirements on the constituents and experimental parameters; for example, the assemblies usually need to be fixed in a vacuum and the blocks melted by arc melting using high-purity base metals to avoid the contamination of the interstitial elements. The diffusion temperature or annealing temperature usually varies from 0.6 to 0.8 T_m, with multiple temperature control groups to achieve a firm connection at the diffusion interface to reduce the pressure on the

contact interface defects[45]. Hot isostatic pressing is applied in a vacuum diffusion welding furnace or vacuum hot-pressing apparatus to strengthen the bonds between the different blocks. Additionally, the diffusion time for diffusion multiples should be long enough, usually in the tens of hours, to produce sufficient diffusion layers facilitating subsequent characterization.

For HEAs, a large diffusion region can be obtained by high temperature with a long diffusion time. There are several typical cases for diffusions multiples in HEAs, whose difference is primarily in the design of the diffusion units of metal blocks. As shown in Fig. 3a, the diffusion multiple consists of three metal blocks, with the two equiatomic binary alloys (Fe–50Mn and Co–50Ni) bonded together, followed by bonding Cr on top of the surface[42]. Quinary alloys in diffusive contact were created at the triple-junction interface with two ternary alloy systems (FeMnCr and CoNiCr) and one quaternary FeMnCoNi alloy system prepared simultaneously at the contact surface. For a high-throughput determination of inter diffusivity matrices in the Co–Cr–Fe–Mn–Ni alloy system[43], a solid diffusion multiple was made by the hot-pressing technique, using quinary alloy blocks prepared by arc melting under an Ar atmosphere (see Fig. 3b). The blocks were
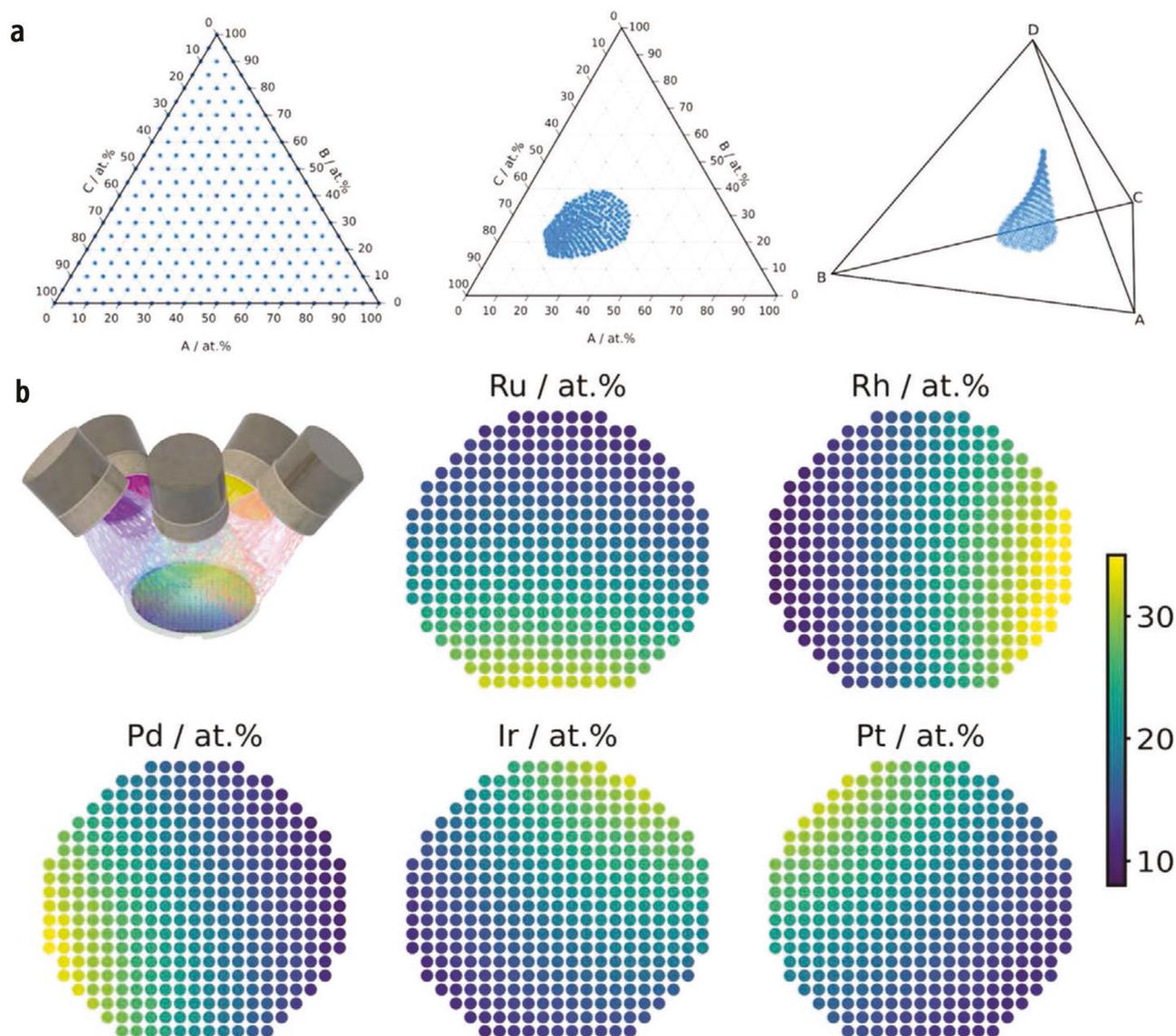
Fig. 2 | Visualization of the compositional coverage of continuous composition spread materials libraries co-sputtered from three to five sources forming ternary to quinary (HEA) systems. a Coverage of a ternary and a quaternary library. **b** Illustration of co-deposition from five deposition sources and compositional gradients of a co-sputtered quinary materials library[31] (adapted with permission from ref. 31. Copyright 2022 John Wiley and Sons).

cut from buttons of four different quinary alloys, with the surface contact blocks having a composition difference in a certain constituent element. The diffusion multiple was fabricated by a self-assembly vacuum hot-pressing apparatus, after which the Co–Cr–Fe–Mn–Ni diffusion multiple was sealed into an evacuated quartz tube for annealing. The sample has a composition gradient in the directions perpendicular to the four contact surfaces, which serves as a carrier for studying multi-component diffusion.

Current studies mainly focus on the Co–Cr–Fe–Mn–Ni system with medium-sized atoms. Due to the sluggish diffusion effect of HEAs[46,47], the diffusion multiple preparation requires a long time and a high temperature. However, for refractory alloys, especially those containing W, Nb, Mo, and Ta, the high melting temperatures and lower diffusion coefficients limit the high-throughput preparation under the current experimental conditions. In summary, the diffusion multiple techniques are effective in making some HEAs and is well suited for determining their diffusion coefficients and phase precipitation kinetics.

Gradient alloying via welding. To create alloy libraries, a concept based on welding has been used, which serves as a screening tool in the study of HEAs[48,49]. The continuous movement of the molten pool and the large heat-affected zone during welding are the main factors for high-throughput preparation using this technique. Due to the complex compositions of HEAs, it is crucial to homogenize the composition in the molten pool-covering area. Friction stir welding (FSW) is a solid-state high-temperature severe plastic deformation process using the heat generated by the friction between a high-speed rotating welding tool and the workpiece. Friction stir processing has been used to realize the continuous preparation of alloy libraries via the existing welding technique with changing workpiece composition[50–52]. Using welding methods as a high-throughput technique allows for faster creation of material property libraries, and some researchers have used these techniques to screen HEAs.

As shown in Fig. 4a, the FSP technique for high-throughput compositional screening was used to study the effect of gradient variations of Cu on the phases (ε-HCP and γ-FCC) and mechanical properties of a vacuum arc melted $Fe_{40}Mn_{20}Co_{20}Cr_{15}Si_5$ (at.%) base material[52]. To achieve compositional gradients in the FSW sample, the tapered section of the pure copper piece was modified in the base alloy tank to have similar dimensions, creating precise dimensional control through milling. The assembled region was then subjected to a friction stir process, with the tapered portions of Cu and HEA mounted in the grooves of the substrate. A Cu backing plate was placed below for cooling, using a W-Re tool for alloying a continuous increase of the additional element. Figure 4b demonstrates the nanoindentation response of FSW-alloyed samples along the gradient alloying direction, in terms of the load-displacement curve, and elastic modulus values can be calculated from the load-displacement curve. Nanoindentation behavior makes it clear that modulus values can be altered by changing the alloy chemistry[52]. Similarly, FSW was used to explore the possibility of introducing a BCC transformation domain in a γ-FCC dominated $Fe_{38.5}Mn_{20}Cr_{15}Co_{20}Si_5Cu_{1.5}$ (at.%) alloy by the vanadium addition, in which the V strips of different widths and controlled channels on base metal fit precisely for the following process[51]. The high temperature from frictional heating and deformation due to tool rotation during the FSW leads to uniformity of microstructure and composition. However, the composition adjustment of the alloy prepared by this FSW is limited to a certain constituent element, and the composition diversity needs to be improved.

Radiofrequency inductively coupled plasma (RF-ICP). Although the co-sputtering method has demonstrated great potential for high-throughput synthesis of HEAs, their properties can be significantly distinct from those of their bulk counterparts due to the smaller grain size in the film sample and the size effect. Therefore, a fast and high-throughput synthesis method for potentially bulk HEAs is highly desirable. In this regard, RF-ICP has been utilized for high-throughput alloy preparation by making use of a plasma arc high energy density beam as a heat source,
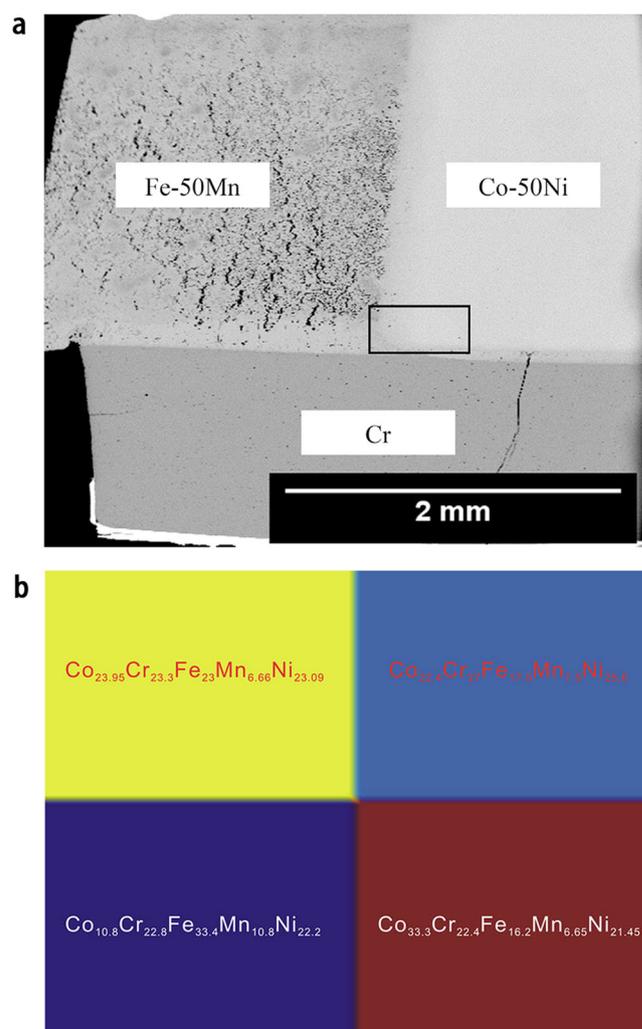
Fig. 3 | Different combination modes for the synthesis of HEAs by diffusion multiple methods. a Micrograph of the hot-pressed appearance of "tri-junction" diffusion multiple sample[42] (adapted with permission from ref. 42. Copyright 2016 Elsevier). b The schematic diagram for the Co–Cr–Fe–Mn–Ni diffusion multiple[43] (adapted with permission from ref. 43. Copyright 2017 Springer Nature).

together with high heating/cooling rates[53,54]. The system includes an ICP torch, an RF generator, and a water-cooled copper crucible that contains mixed pure meal powder. The melting process was under the RF-ICP torch for less than 40 s and argon was kept running to cool down the alloy and protect the samples from the air after the plasma was off. RF-ICP was used for rapid synthesis of the FeCoNiCu bulk HEAs (Fig. 5)[54]. The fast synthesis method by RF-ICP shows the potential of the high-throughput preparation of HEAs for their accelerated discovery. As seen, the automatic platform can process 20 samples in one run (~10 min). Moreover, this platform is highly scalable by incorporating more RF-ICP torches, leading to considerable improvement in the efficiency of HEA synthesis.

Additionally, when combined with an automatic powder mixing and blending system, the fast and combinatorial synthesis of bulk HEAs can be achieved in a much more efficient manner. The high-throughput capacity of this RF-ICP method enables efficient verification of the new alloys predicted by computational simulations or the ML algorithm. The RF-ICP method allows researchers to prepare 100 designed alloys within an hour, leading to higher efficiency than other typical methods for HEAs synthesis[54].

Structure characterization
After high-throughput preparation of HEAs, it is crucial to characterize the microstructure of these samples to evaluate their properties. The analysis of
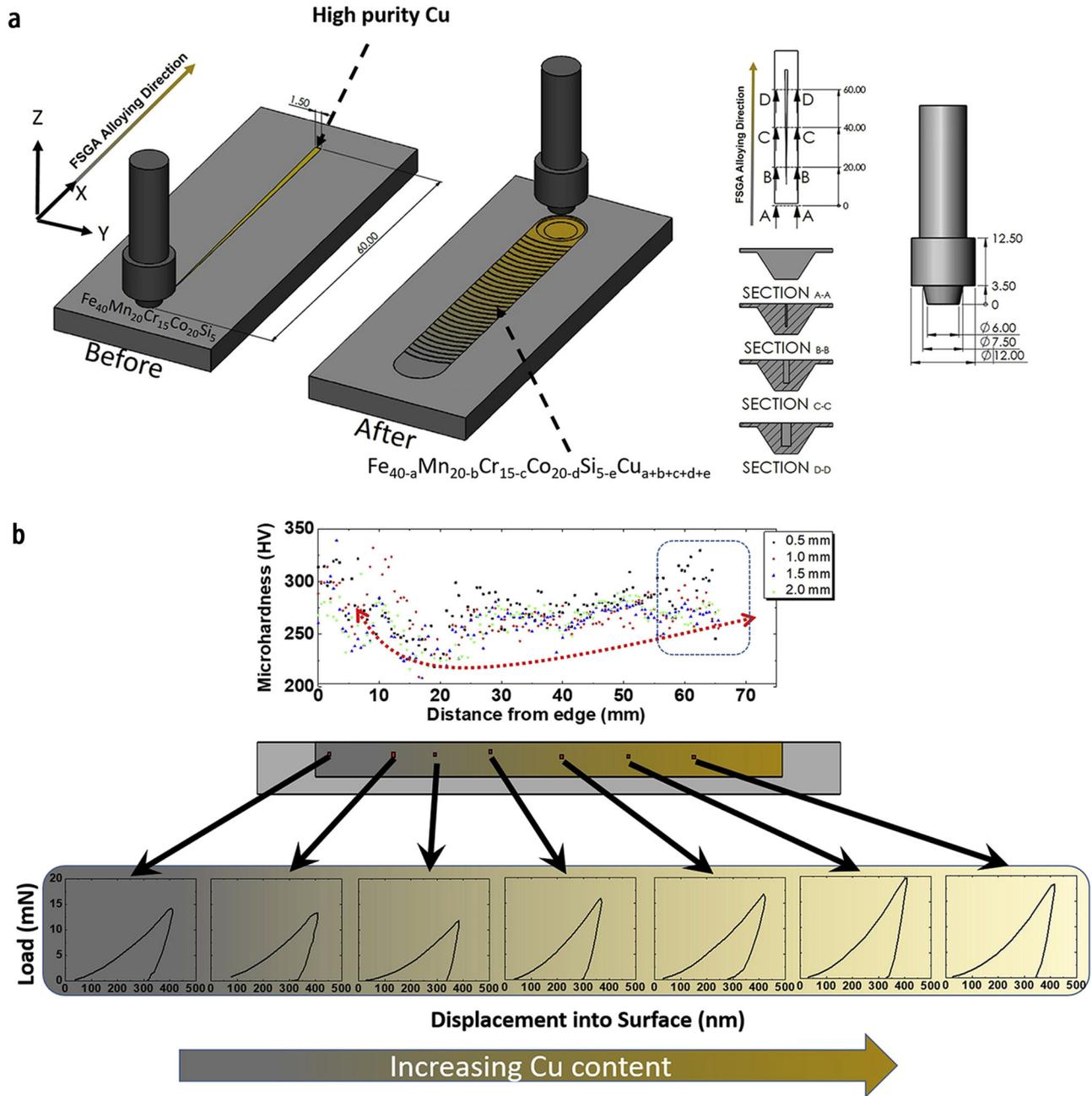
Fig. 4 | Alloy synthesis on welding technologies. a A schematic of friction stir gradient alloying process and corresponding tools. **b** Microhardness variation in gradient alloyed sample superimposed with the schematic of the FSW assembly and nanoindentation plots (load vs displacements) of different areas on the FSW assembly[52] (adapted with permission from ref. 52. Copyright 2020 Elsevier).

the composition and structure of complex materials libraries is the basic component of high-throughput characterization, which is usually carried out by electromagnetic spectrum in different bands such as X-ray, ultra-violet, and infrared. Energy dispersive X-ray (EDX) analysis is by far the most commonly used method to measure the composition of individual alloys in a library synthesized by high-throughput methods[30,32]. It is effective for the characterization of sputtered thin films with continuous component gradients. The theoretical component gradients were obtained by simulation and parameter adjustment, and then the compositions of different points in the prepared sample were measured and calibrated using a bulk sample of known compositions in the library. To improve measurement accuracy, the multiple iteration method is a useful choice when analyzing the composition of a library sample. It should be noted here that EDX performed on a large area cannot provide information about whether the

elements are truly incorporated into a single-phase lattice. In this regard, combining other methods such as scanning electron microscopy (SEM) and X-ray diffraction (XRD) is a routine strategy and is effective for high-throughput phase identification with high resolution by adjusting the magnification of the acquired images. For example, Fig. 6a shows a schematic representation of a combinatorial high-throughput $Al_x(CoCrFeNi)_{100-x}$ HEA synthesis setup via two individual magnetron sputtering sources. As indicated by grazing-incidence X-ray diffraction (GIXRD) and EDX results (Fig. 6b), the structures of thin films transform from FCC to BCC with increasing amounts of Al. The surface morphologies of the thin films investigated by SEM are shown in Fig. 6c. As seen, the thin film with low Al content (#1–#6) that possesses an FCC structure, has a relatively compact column structure with column diameters of about 20 nm, while for #7–#10 HEA thin film with a BCC structure, the smaller elongated
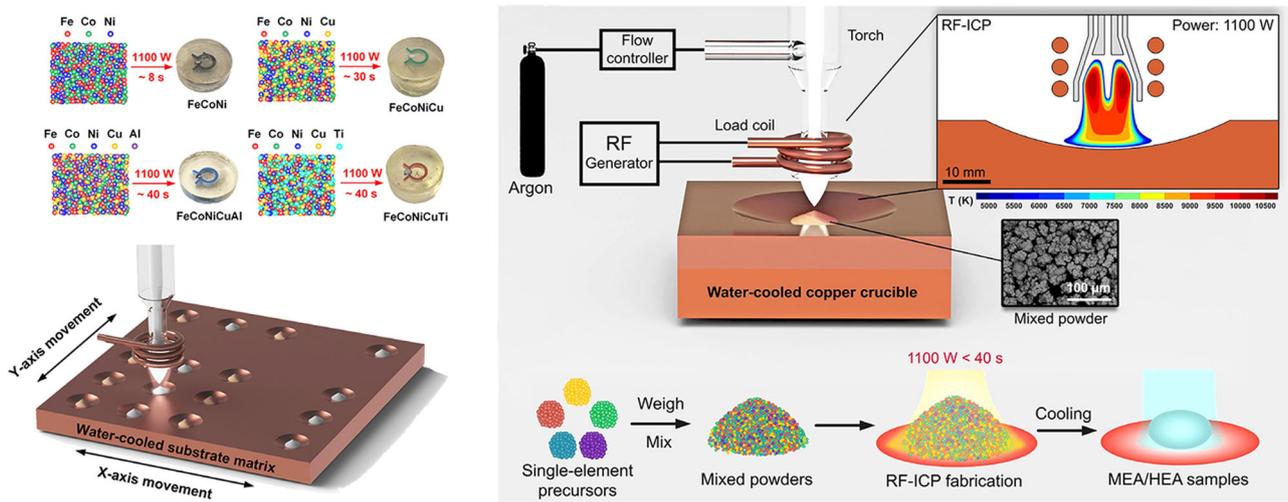
Fig. 5 | Schematic of the high-throughput experimental setup for synthesizing alloys with a large composition space via the RF-ICP system. The mixed pure metal powders, such as Fe, Co, Ni, Cu, and Al etc., are placed in a water-cooled copper crucible and melted under the RF-ICP torch. Argon is kept running to cool down the samples and prevent them from the air[54] (adapted with permission from ref. 54. Copyright 2020 John Wiley and Sons).
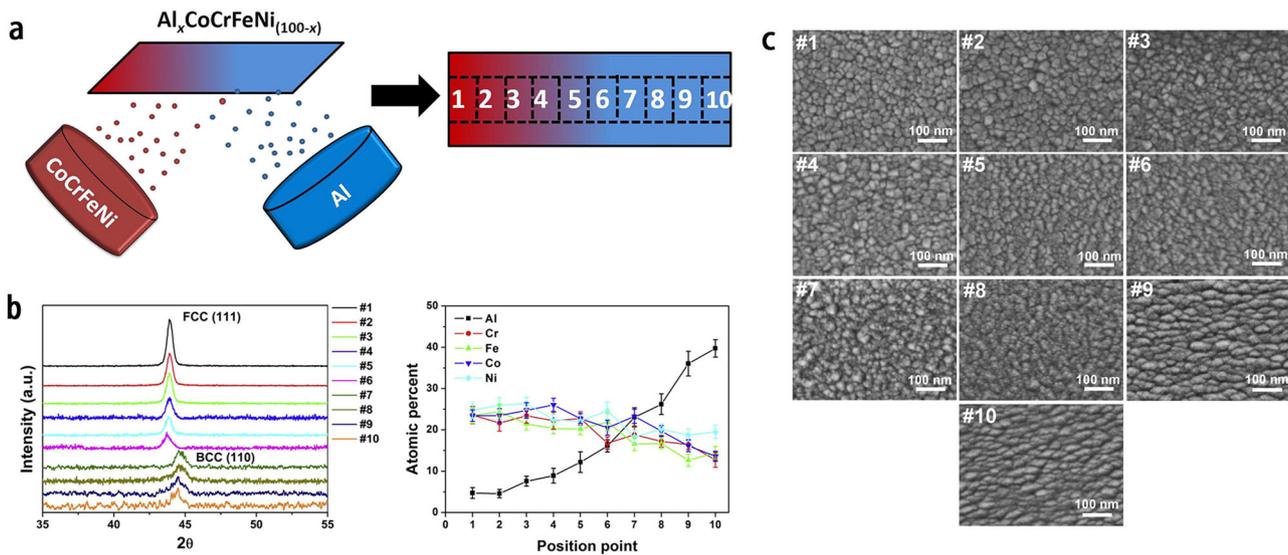
Fig. 6 | $Al_x(CoCrFeNi)_{100−x}$ HEA thin films synthesized using the magnetron sputtering method. a Schematic representation of a combinatorial high-throughput HEA synthesis setup using two individual magnetron sputtering sources, and the top-surface view of the substrate containing the HEA library with a constant concentration gradient. **b** GIXRD patterns of diffraction intensity vs diffraction angle 2θ, and EDX atomic percent vs sample number analysis. **c** SEM images of $Al_x(CoCrFeNi)_{100−x}$ thin films along the Al concentration gradient[55] (adapted with permission from ref. 55. Copyright 2020 Elsevier).

columns form on the surface, and the elongated small columns grow to accumulate forming elliptical columns of larger sizes. The SEM results indicate that the column morphology changes are associated with the composition of the thin films[55]. In some cases, an optical microscope combined with a camera can take a series of images at short intervals, which can then be verified by SEM to measure structural evolution in situ[36]. Metallography reflects the microstructure information of alloy samples, to some extent, and can be used as an alternative method for high-throughput characterization.

Apart from SEM, electron backscattered diffraction (EBSD) is also an important method for high-throughput screening of materials libraries. The orientation information can be obtained by analyzing the symmetry of EBSD patterns. Moreover, when coupled with the spacing of the Kikuchi lines associated with EBSD, one can identify the phase information of materials libraries. Zhao et al. used this method to rapidly screen the local microstructure changes and clarify the phase formation mechanism

associated with interdiffusion in the Cu–Pt–Ru ternary libraries, which were synthesized by diffusion couples[40]. It should be emphasized that such an approach can be developed to study phase diagrams of diffusion couple-based combinatorial libraries when coupled with other elemental chemistry analytical techniques. For other combinatorial libraries made by LAM or multiple deposition sources, EBSD has also been implemented to gather detailed phase information of $Al_xCrCuFeNi_2$[15], $Al_xCoCrFeNi$[17], and CoCrFeMnNi Cantor alloys libraries[56], and the like. However, a major bottleneck of EBSD is the analysis of patterns, which requires human input to select potential phases for dictionary pattern matching. High-throughput and autonomous determination of crystal symmetry is the most important step in making EBSD into a high-throughput technique. In this case, a data-driven ML approach is developed for rapid and autonomous identification of the crystal symmetry from EBSD patterns. Typical algorithms, including convolutional neural networks[57], few-shot learning[58], and support vector machine[59], are adopted to autonomously determine phase structures of NiAl

and steels, indicating the feasibility and efficiency of the combined EBSD-ML approach for high-throughput characterization.

For multi-component alloys with complex thermodynamic processes and crystallization behavior, structural characterization is essential for the rapid screening of new materials. A detailed structural characterization has been conducted to characterize Fe, Ni, Co, and the correlated compound phases using a scanning microbeam X-ray diffractometer with a spatial resolution of 50–300 μm[60]. It utilizes a unique high-brilliance and high-intensity X-ray microfocus source with a 300-micron beam spot. Another optimized test mode is the use of automatic XRD detection equipment to measure the component points through a mask. The sampling spacing can depend on the accuracy of the device or the sample points with a specific component gradient, which provides great flexibility. With the development of XRD technology towards high brightness and micro-focus, the process and time consumption of high-throughput characterization will be simplified and reduced to some extent.

From a microdomain or in situ measurement perspective, synchrotron X-ray and neutron scattering techniques have great capabilities for high-throughput characterization. Synchrotron radiation sources achieve high brightness micro-focus in the full spectrum from infrared to hard X-ray, and meet the requirements in brightness and spatial resolution for rapid and accurate characterization of high-throughput material samples[34,61,62]. Neutron scattering has potential applications for the characterization of the magnetic structure of combinatorial materials libraries, as neutrons have deep penetration power and possess magnetic moments. The high temporal and spatial resolution enabled by synchrotron and neutron scattering facilities can break the flux bottleneck in high-throughput experiments.

Mechanical properties

High-throughput assessments of properties for HEAs have been carried out extensively over the years. Mechanical parameters, such as modulus, hardness, and so forth, can be used to predict and screen bulk materials with better mechanical properties. In recent years, a series of micromechanical testing methods have been developed, including testing for hardness[52], tension[63], compression[64], fatigue[65], thermal stress[66], and the small punch test (SPT)[67–69] etc., providing tools for high-throughput characterization of HEAs.

Hardness is the basic index for testing the mechanical properties of structural materials. Among the different mechanical properties, hardness is the simplest one that can be obtained in a high-throughput manner. An automated hardness tester with a constant loading mode generates a micro-hardness mapping along the set trace. The nanoindentation technique enables fully automated measurements of hardness and elastic modulus from a small region of a sample with a displacement resolution of 1 nm, which is useful for directly probing a single phase since the interaction volume is within a single grain. Coury et al. converted nano hardness and elastic modulus into yield strength using a revised Clausner–Ritcher equation, and the strain hardening coefficient was kept independent from stress–strain curves. In addition, the experimental and simulation results indicate that the strength is maximized when the atomic size mismatch is maximized. Moreover, it is necessary to consider the strain hardening of these alloys to accurately estimate their strength by nanoindentation[44]. Shukla et al. performed micro-hardness tests along four depth levels of a sample from the top surface, and each indent point was also 0.5 mm apart along the alloying path so that the systemic hardness and moduli were obtained. It is found that an increase in moduli and hardness values can be attributed to solute–matrix interaction. The as-cast ε phase-dominant microstructure showed ~153 GPa moduli, while the same for a completely γ microstructure with supersaturated Cu content reached up to ~224 GPa[52]. In measuring nano hardness and modulus of multicomponent samples, the setting route of nanoindentation is usually along the gradient direction of a specific alloying element, and the discrete points and micro-hardness can be analyzed for multiscale observation[70]. According to the hardness–strength relationship, one can efficiently select the potential HEA candidates with the

desired mechanical properties, such as higher yield strength in a relatively large composition space. Although there may exist discrepancies between the high-throughput made and bulk samples for the absolute value of mechanical properties, the variation trend of the composition and properties of interest can still provide effective information for the design of new HEAs.

Due to the small scale of the HEAs prepared by the above-mentioned high-throughput methods (i.e., addictive manufacturing, sputtering, etc.), it is usually difficult to cut bulk specimens from the thin layers or coatings. The SPT is an evolving small specimen test technique with the potential to extract the mechanical properties (ductility, elastic modulus, yield strength, ultimate tensile strength, fracture toughness, etc.) from small-volume HEA specimens prepared by high-throughput methods[67,69]. It should be noted here that a prerequisite for using this test is to establish correlations between SPT and conventional mechanical tests such as tensile testing for HEAs in priori. However, the SPT response is easily influenced by different test parameters, that is, for specimen shapes and thickness, test speed, ball diameter, and so on. It is therefore imperative to understand the effects of these parameters. This necessitates the optimization of test parameters to obtain nearly unique SPT responses, at least for a class of HEA materials. Thus, it is necessary to relate the conventional and SPT results by empirical and analytical relations.

Additionally, the cooling rates of commonly used addictive manufacturing and sputtering high-throughput methods are much higher than those of traditional casting methods used for the preparation of bulk HEAs. Notably, in some extreme cases, owing in part to the multi-principle nature of HEAs, the HEA coatings or layers via high-throughput methods can form amorphous structures, which make the mechanical properties quite different from the bulk HEAs. Thus, the optimization of preparation parameters to make the cooling rate agree with the casting method is significant for the formation of HEAs. In sum, although there are some discrepancies between thin and bulk HEA materials, the SPT methods can at least determine the mechanical properties and guide the researchers to develop better HEAs in a such large composition space.

Physical properties

As one of the typical physical properties, the magnetic properties of HEAs depend heavily on the size, microstructure, and preparation process of the sample. Many efforts have been made to measure and map magnetic properties at very high spatial resolution. Borkar et al. presented a new combinatorial approach, based on laser additive deposition of compositionally graded alloys, for rapid assessment of the composition–microstructure–magnetic relationships in $Al_xCrCuFeNi_2$ alloys ($0<x<1.5$ at.%) HEAs. Along the same alloy gradient, the microstructures are FCC solid solution, FCC/L12, mixed FCC/L12 + BCC/B2, and finally predominantly BCC/B2 with increasing Al content. Owing to the change of microstructures, the low Al-containing FCC/L12 regions are weakly ferromagnetic, while the BCC/B2 regions with higher Al contents are strongly ferromagnetic, exhibiting lower coercivity and higher saturation magnetization[15]. For the FeMnCoCrAl HEA system, Marshal et al. developed thin-film libraries for the combinatorial evaluation of the phase formation and magnetic properties combined with spatially resolved atom probe tomography and DFT simulation. It was found that the addition of Al can promote the formation of BCC structure, which exhibits soft ferromagnetic behavior. A further increase in the non-ferromagnetic Al content beyond 8 wt% decreased the overall saturation magnetization due to the substitution of ferromagnetic species by paramagnetic Al and lattice distortions, which was in agreement with DFT predictions[32]. As can be seen in these cases, high-throughput techniques are efficient in explaining the microalloying effects on the magnetic properties of HEAs and therefore have great potential for the future designs of soft magnetic HEAs with better performance. However, it should also be noted that the size effect and magnetocrystalline anisotropy caused by thin film may lead to some artifacts, which can be eliminated by increasing the thickness of as-prepared film/layer libraries
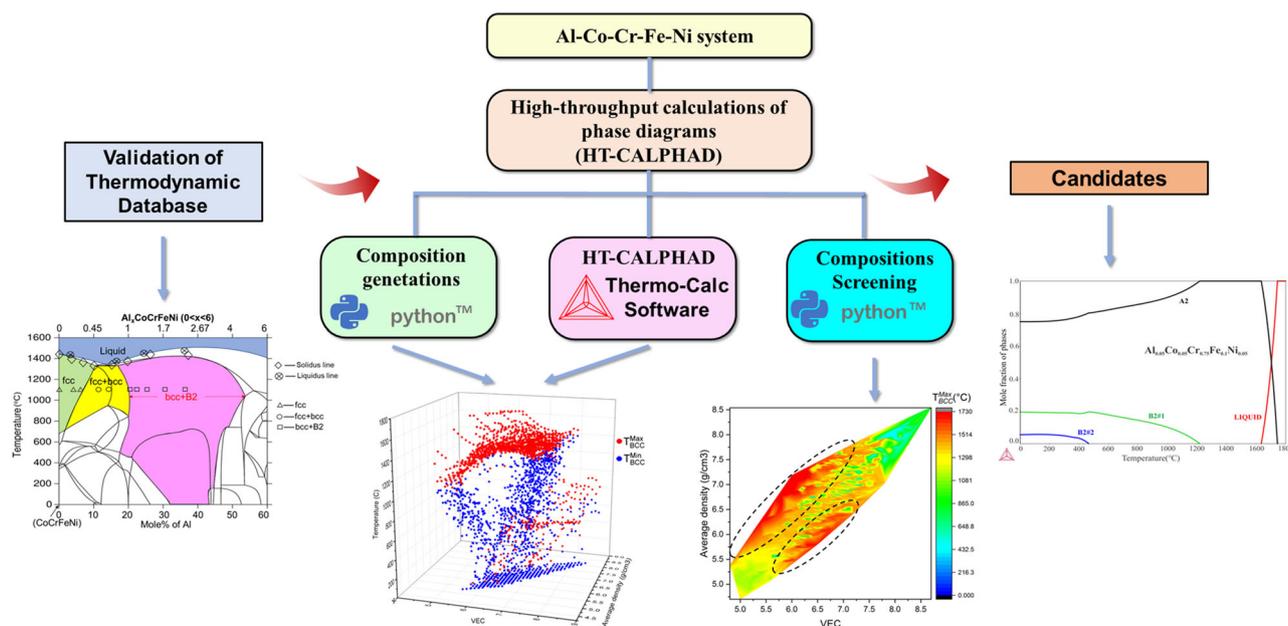
Fig. 7 | The schematic of discovering HEAs with a high-throughput CALPHAD approach. Al–Co–Cr–Fe–Ni quinary systems were used as the case study to investigate the reliability VEC rule and its application to the material design[94] (adapted with permission from ref. 94. Copyright 2020 Elsevier).

or changing the measurement direction when performing magnetic testing.

Besides saturation magnetization, studies of high-throughput techniques for other physical properties of HEAs are rather sparse. However, when expanding the scope to other materials synthesized via high-throughput techniques, there are different physical properties of interest. For example, useful combinatorial methods for examining magnetic properties include magnetic force microscopy and scanning magneto-optical Kerr effect imaging[40]. In addition, the Decay microwave probe microscope, with very high micro-region resolution, can measure magnetic properties, including susceptibility and spin resonance. Combined with automatic sample table control and data acquisition, it is possible to realize a high-throughput automatic electromagnetic measurement of the composite material chips[71,72]. Additionally, magnetic optical imaging with an indicator film is a useful approach to study magnetic phase diagrams and the composition dependence of Curie temperature[73]. Although these cases are based on other materials, one can still be enlightened by the above-mentioned techniques, which provide new ideas for the study of functional HEAs with better performance.

High-throughput computing for HEAs

Besides experimental methods, theoretical and calculation methods also play an increasingly important role in exploring HEAs. The combination of simulations and experiments is beneficial to understanding of the physical mechanism underlying the phase formation and the structure–properties relationship. Currently, some classic computational simulation methods have been developed to guide the design of new HEAs in a large composition space, such as molecular dynamics (MD)[74] and first-principle calculations based on density functional theory (DFT)[75]. With the advancement of computer hardware and software, the emergence of high-throughput calculation (HTC) has attracted much attention, which can significantly speed up material design and shorten the research and development cycle[76]. The core idea of HTC is "integration," which emphasizes the integration of calculation and data, and the integration of multi-scale simulations. Therefore, many integrated computing platforms have been developed, including Pymatgen[77], FireWorks[78] and Atomate[79], ALKEMIE[80], MatCloud[81], and Open Quantum Materials Database[82], which have been widely used in the community and make workflow and dataflow more maneuverable and transparent.

HTC has been widely used to investigate phase structure and its evolution with composition and temperature in HEAs. Among different simulation methods, due to their versatility and reliability, the development of high-throughput DFT methods for calculating the properties is of great interest in the HEAs community. The typical approach that has been developed to model HEAs is DFT based on the coherent potential approximation[83–85]. In addition, applying high-throughput first-principles calculations, Santodonato et al.[86] studied the temperature and composition-related phase evolution in HEAs, and focused on the aluminum-containing HEA with an enhanced multiphase microstructure. Additionally, the first-principles-based integrated software AFLOW[87] is exploited to high-throughput screening of the crystal structure of alloys. Lederer et al.[88] screened thousands of systems in the AFLOW library, and predicted a large number of previously unknown potential quaternary and quinary solid solution alloys, which provides a helpful guide for designing new HEAs with a solid solution structure.

The calculation of phase diagram (CALPHAD) approach aims to study the thermodynamic properties of various phases by developing thermodynamic models[89]. There are different software developed based on the CALPHAD approach, one of the typical commercial software is Thermo-Calc, which includes high-throughput modules such as TC-Python. Thermo-Calc users run batch calculations for many varied parameters in a high-throughput manner. Many attempts have been made to develop thermodynamic modeling in a variety of different alloy systems using the high-throughput CALPHAD method, including phase diagrams and thermodynamic properties[90–95]. Due to the limitations of the empirical VEC rule in different HEA systems, Zhong et al. recently proposed a data screening procedure to develop new HEAs via a high-throughput CALPHAD approach (as shown in Fig. 7)[94] and found the relationship between phase formation behavior and VEC. Additionally, Zhang et al.[90] reported a sufficiently large database of the Al–Co–Cr–Cu–Fe–Ni HEA system to calculate the primary solidification phase. Klaver et al.[93] used the Thermo-Calc to determine the phase evolution behavior of AlCrMnMoTi, AlCrMoNbTiV, AlCrMnNbTiV, and AlCrFeTiV alloys at different temperatures and found that AlCrMnNbTiV and AlCrMoNbTiV were better HEA formers. Gurao and Biswas[91] studied 1287 equiatomic quinary alloys using the CALPHAD method to find single-phase FCC and BCC HEAs. According to their calculation results, they achieved the optimized alloy composition just by preparing two FCC alloys and seven BCC alloys, which

dramatically increased the efficiency of alloy designing. In particular, CALPHAD can predict the phase diagram under extreme conditions, such as high temperatures and high pressures, which are difficult to explore for experimental studies.

As a newly emerging technology, HTC still faces critical challenges. First, most integrated calculation programs currently available are based on first-principles calculations; thus the material data are obtained from a few to dozens of atoms, which requires developing the HTC further on a larger scale. In this regard, combining ML and first principles to develop high-precision potential functions for MD simulations is a significant trial[96]. Second, the classification of the accumulated materials data is still vague, making it difficult to maintain a materials database in the future. It should be clearly divided according to an authoritative materials classification system. In addition, the data format should be strictly followed in the acquisition process. In terms of an in-depth understanding of HEAs, due to the multi-principal elements contained in HEAs and the metastable state in thermodynamics, there is an urgent need to develop a reliable thermodynamics database that contains a series of composition, temperature, and phase-equilibrium data for HEA systems. In this regard, the related binary and ternary systems should be gathered and assessed by implementing experiments and calculations on HEA systems.

Data-driven ML strategies

The enormous composition space for designing HEAs offers not only opportunities but also great challenges, requiring intelligent and efficient strategies for materials discovery. As a burgeoning branch of materials science, data-driven methods, such as ML, which are used to study a wealth of existing experimental and computational data, have become a very exciting area of research in materials science. ML refers to programs that automatically improve their ability to perform tasks by learning from experience in many scenarios. This automates the time-consuming knowledge acquisition process, which is essential to speed up computing and reduce the cost of developing data-based systems. With ML, when given enough data and a rule-discovery algorithm, computers can analyze the trends in datasets and further help one to understand the relationships between properties and different parameters, which is beneficial in guiding materials modeling. ML is most useful in situations in which human learning is impossible, such as when data and interactions within the data are too complicated and intractable for human understanding and conceptualization[97].

Datasets for HEAs

The first and most important step in ML is to generate robust datasets for training the ML model. The selection of suitable data can be deceptive in ML, which is why so much emphasis is placed on the visualization of the datasets[98]. The construction of a dataset is task-oriented; that is, the final prediction plays a decisive role in what type of data should be collected.

The study of ML in HEAs mainly focuses on the formation of single-phase solid solutions (i.e., BCC, FCC, and HCP), while some work has been carried out on mechanical properties such as hardness and modulus. Compared to traditional metallic materials, HEAs are newcomers that have been studied for only nearly two decades. To date, most HEA data have been collected from published experimental work or simulation methods. Miracle and Senkov's review summarizes a dataset containing 648 entries of HEAs in different systems[14]. Based on this dataset, Zhuang et al. constructed a dataset composed of 401 HEAs, which consists of 174 SS phases, 54 intermetallics (IM), and 173 SS + IM phases, by removing some multiple alloys with the same composition[99]. Later, in 2020, Gao et al. built a dataset consisting of 1252 samples—625 single-phase and 627 multi-phase alloys—covering binaries and multi-component systems[100]. Besides experimental data, computational methods, such as high-throughput ab initio and DFT-based approaches, are used alternatively to produce phase formation information. Curtarolo et al. developed a high-throughput ab initio method called LTVC (Lederer–Toher–Vecchio–Curtarolo) to predict the transition temperature of multi-component systems[88]. In this way, a dataset containing

a total of 1798 unique equiatomic compositions was constructed, consisting of 117 binaries, 441 ternaries, 1110 quaternaries, and 130 quinaries. Based on this dataset, Vecchio et al. built a data-driven workflow for predicting the composition–phase–structure relationship[101].

Besides the phase formation data, there are property datasets of HEAs. Using the integrated CALPHAD-ML approach, Sun and Lu et al. predicted the hardness of Ti–Zr–Nb–Ta refractory HEA, which included building a database of 100 quaternary alloys, training the ML model, hardness prediction, and experimental verification[102]. A database composed of alloy composition and hardness data for the Ti–Zr–Nb–Ta RHEAs was established by combining CALPHAD. To search for high-entropy ceramics, Vecchio et al. performed an ML framework on 56 previously reported entropy-formation ability values, including nine synthesized compositions, six single phase, and three multi-phases. The high-entropy ceramics in the dataset are mainly composed of eight carbide-forming metal elements (Hf, Nb, Ta, Ti, Mo, V, W, and Zr)[103]. Regarding the modulus, Chen et al. combined first principles and ML to predict the elasticity of severely lattice-distorted HEAs with experimental validation. The ML models were trained on 6826 ordered inorganic compounds from the Materials Project database to predict the Voigt–Reuss–Hill averages of bulk and shear modulus with log-normalization[104]. In the case of experimental data for modulus, Roy et al. compiled Young's modulus consisting of only 87 HEA entries from limited available experimental reports[105]. All the above-mentioned datasets are summarized in Table 1.

Despite substantial progress in the construction of datasets for HEAs, the data size improvement is still far from complete. As a result, the results of calculations and predictions based on these databases may deviate significantly from the experimental results. Moreover, when reporting their findings, researchers tend to publish only favorable data, while the bad data points are often dropped. This will lead to the dataset being unbalanced and will affect subsequent ML models' performance. Therefore, there is an urgent need to develop reliable and robust databases dedicated to HEAs. As such, high-throughput preparation and characterization, as well as HTC, would be a reliable approach to batch production of HEA libraries, including composition and property information.

Phase formation prediction

As a new paradigm for developing HEAs, the data-to-knowledge ML strategy has the potential to explore complex structures and property space in an efficient way. Additionally, it can also yield valuable insights into the key factors that determine macro-performance and thus guide the design of HEAs with enhanced properties. As mentioned above, ML in the field of HEAs relies on the availability of libraries of compositions, structures, and properties that have been assembled and scrutinized by experimental and computational methods. Considering the different data sizes, phase formation behaviors (i.e., single solid solution formation for HEAs) have attracted much attention from the academic community[105–120]. In addition, there are increasing studies on the physical or mechanical properties of HEAs. From the perspective of ML, the two cases above correspond to classification and regression issues, respectively. As such, in this section, we will review ML techniques and propose the possibility of further development of ML in HEAs.

Phase formation behavior is crucial to the performance of HEAs. While computer simulations, such as first-principles calculations and MD simulations, have become a commonly used tool for materials discovery, their computation expense limits their application in the accelerated exploration of potential HEAs. The recent implementation of data-driven techniques has provided a possible alternative for efficiently predicting phase formation in HEAs[109–112,117,119,121,122]. ML can recognize the inner data pattern and construct a model to make quick predictions for unseen samples. Based on very sparse data, Raabe et al. proposed an active learning framework, which includes three main steps—targeted composition generation, physics-informed screening, and experimental feedback—to accelerate the design of high-entropy Invar alloys in an almost infinite compositional space (see Fig. 8). Compared with the conventional design approach, which requires years

and many experiments, this ML workflow requires only a few months to develop HEAs with desirable properties[121]. Wu et al. used ML to successfully predict eutectic HEAs with excellent mechanical properties in the Al–Co–Cr–Fe–Ni HEA system, and analyzed the key elements for forming eutectic HEAs[117]. Islam et al. established a neural network model to predict the formation of the HEA phase. Cross-validation revealed a predictive accuracy of 83% on this limited data set[109]. Amitava et al. used more algorithms to establish multiple prediction models and forecast the different structures of the solid solution (FCC, BCC). The prediction accuracy is over

90%, which is attributed to the fact that the random forest model has overwhelming advantages in dealing with small datasets compared to the artificial neural network algorithm[111]. Thus, understanding and applying multiple ML algorithms is necessary for the prediction of HEA phase formation. Moreover, to solve the data shortage problem of HEAs, Lee utilized a conditional generative adversarial network to find a model distribution that emulates the distribution of known HEAs, then augmented realistic samples based on feature representation, and finally realized the expansion of the original dataset[119]. The results show that the accuracy of the model is significantly improved due to data augmentation.

Compared with the original ML modeling method, using feature engineering to construct a new descriptor can effectively determine the structure–performance relationship[123]. Material descriptors and models determine the robustness of the ML prediction. Pei et al. carried out the ML modeling analysis of many parameters and the link between the phases, and identified the physical parameters that are crucial to the formation of solid solutions[100], such as volume modulus, melting temperature, etc. Dai et al. used feature engineering and the ML strategy to extend the descriptor dimension from a low dimension originally to a high dimension[114]. Due to the uniqueness of different algorithm constructions, the best performance model depends on the effective combination of datasets, descriptors, and algorithms. In this regard, Zhang et al. proposed a systematic framework that utilized a genetic algorithm (GA) to efficiently select the ML model and materials descriptors from a huge number of alternatives and demonstrated its efficiency on two-phase formation problems in HEAs[114]. Generally, the prediction accuracy of the model can be improved through hyperparameter optimization, such as increasing the number of hidden layers and neurons in the neural network[107]. Overfitting and underfitting are the common problems that any ML may encounter[113], and there is no exception in the study of predicting HEA phases by ML. Huang et al. found the overfitting phenomenon using ML phase projection. By adjusting the super parameters involved in the training process, training accuracy can always be improved to a higher level[99]. Wen et al. proposed ML models to predict the solid solution strength/hardness of HEAs[123]. Figure 9 shows the prediction error for the hardness of HEAs by five-fold cross-validation with possible

Table 1 | Datasets of phase and mechanical properties for HEAs

Item		Data size	Year	Ref.
Phase	Single phase	648	2017	14
	174 SS + 54 IM + 173 SS and IM	401	2019	99
	625 SS + 627 multiphases	1252	2020	100
	117 Binaries, 441 ternaries, 1110 quaternaries, and 130 quinaries	1798 (DFT)	2020	88
	Single phase	70	2023	153
Mechanical properties	Hardness	100	2021	102
	Hardness	162	2021	123
	Hardness	85	2021	22
	Hardness	557	2022	154
	Hardness	290	2021	155
	Modulus	6826	2019	104
	Modulus	87	2020	105
	Yield strength	122	2021	156
	Yield strength	501	2023	157
	Thermal expansion coefficient	699	2022	121
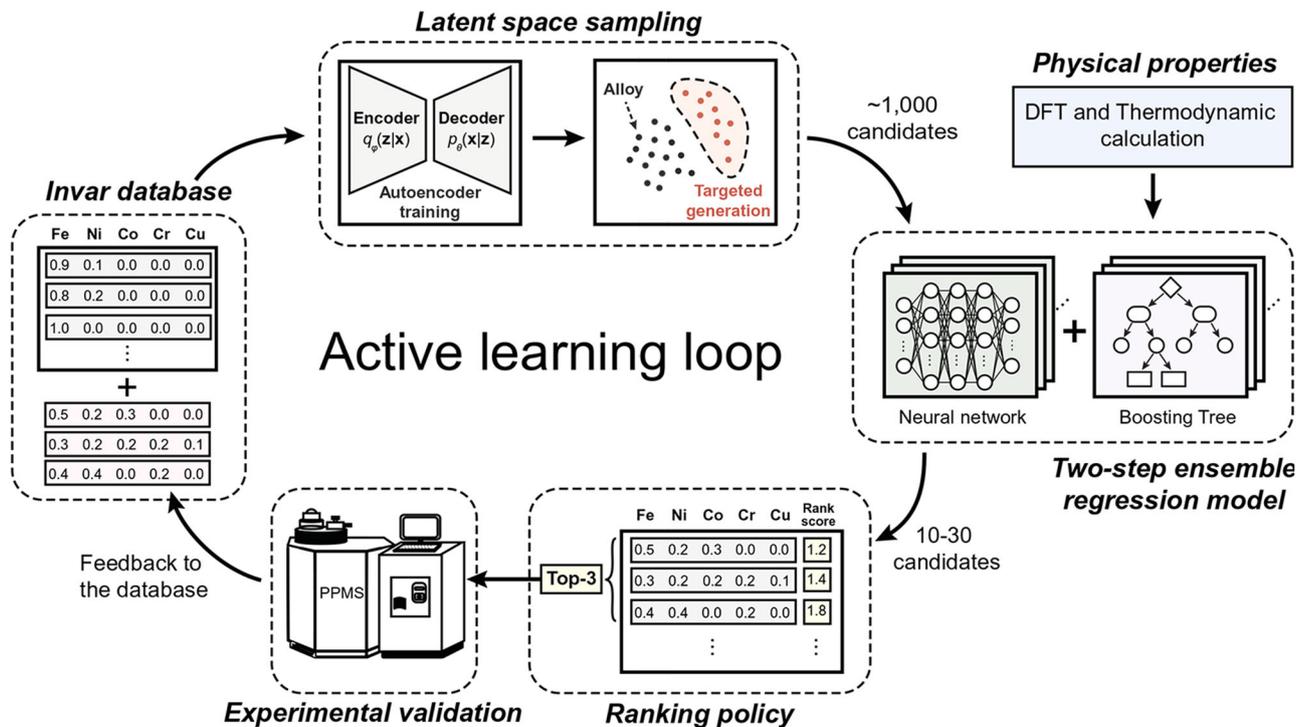
Fig. 8 | Schematic flow chart of the active learning framework. This framework aims to design the composition of HEAs, combining ML models, DFT calculations, thermodynamic simulations, and experimental feedback[121] (adapted with permission from ref. 121. Copyright 2022 AAAS).
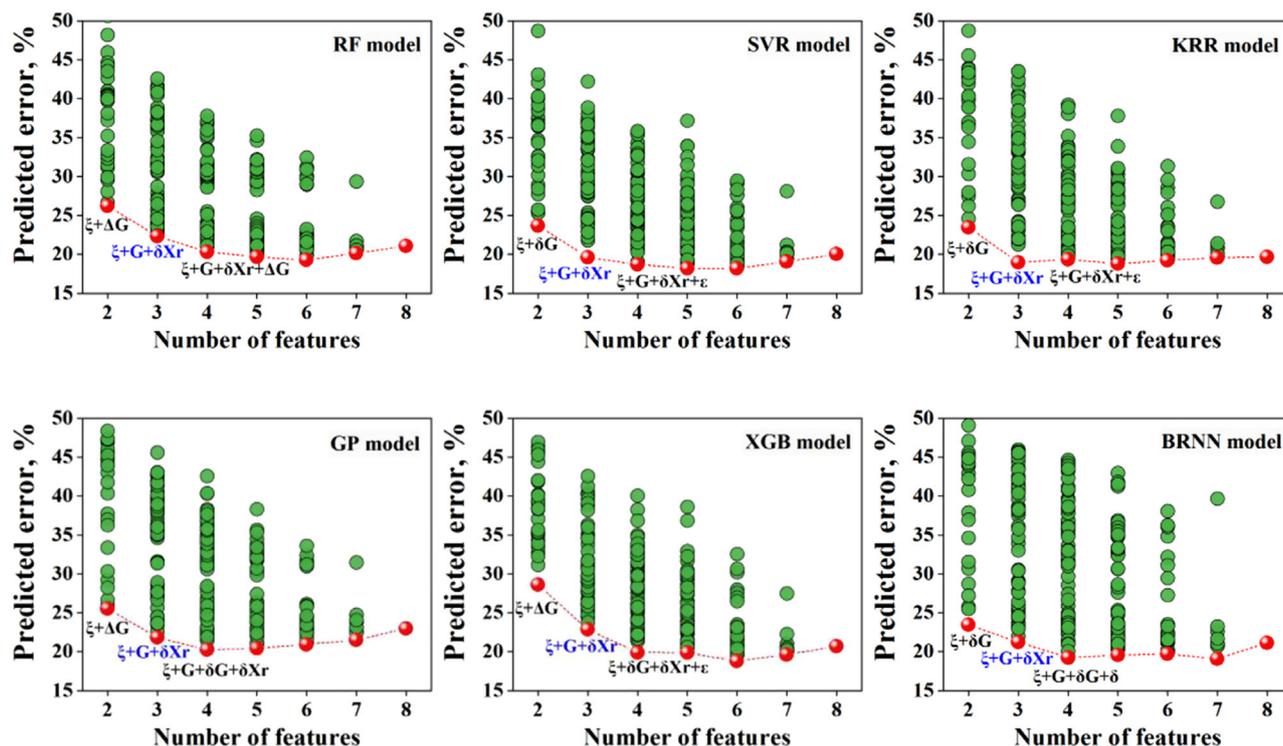
Fig. 9 | Feature selection based on combinations of features from different ML algorithms. The predicted error of each model contains a subset of the eight features in the data set[123] (adapted with permission from ref. 123. Copyright 2020 Elsevier).

combinations of different features (ξ, δXr, and ε, etc.). All ML models, including random forests (RF), support vector regression (SVR), kernel ridge regression (KRR), Gaussian process (GP), extreme gradient boosting (XGB), and Bayesian regularized neural networks (BRNN), show a basin-like tendency, indicating that too many or too few features will reduce the accuracy. According to "Occam's razor" principle, simplicity, and interpretability with a minimum number of features are necessary for adequate accuracy. Using more features complicates the interpretation of the model and risks overlearning.

In the absence of unified evaluation criteria, excessive optimism is often reported[116] as a result of overfitting and the use of inappropriate training and test data. It is necessary to propose new standard criteria that can be used to evaluate the true accuracy and performance of ML models. An emphasis on experimental validation and repeatability through code archiving also helps overcome this challenge. The regularization method can be incorporated into the ML model to improve the generalizability of the model[119]. The hyperparameters of the model can also be optimized by the Bayesian optimization method to obtain good generalizability under the condition of high accuracy. In addition, constructing new rules with strong interpretability and universality through ML is desirable, which can be explored using conformable regression. Therefore, combining experimental results with theoretical guidance to analyze specific target characteristics is imperative to screen new HEAs with good performance[115].

Prediction of mechanical properties
As a new kind of structural material that can serve under extreme environments, HEAs exhibit unique mechanical properties, such as high strength and hardness, and low moduli. These properties are generally used as selection parameters in the search for new alloys. This raises the question of whether ML algorithms can be readily used to the search for candidate alloys with better mechanical properties in such a large composition space.

As one of the most typical mechanical properties of HEAs, hardness has strong correlations with other properties, which requires an in-depth

understanding. For instance, based on a reliable hardness–strength relationship, complex mechanical tests can be replaced to some extent by efficient and inexpensive hardness tests for a fast and comprehensive assessment of mechanical properties. Hence, developing data-driven methods, in addition to experimental methods, is essential to effectively calculate, predict, and evaluate the hardness of HEAs. In this regard, several studies have attempted to explore the possibility of ML as an aid in hardness assessment. For example, using the integrated CALPHAD-ML approach, Sun and Lu et al. predicted the hardness of Ti–Zr–Nb–Ta refractory HEAs, which included building a database of 100 quaternary alloys, training the ML model, hardness prediction, and experimental verification, as shown in Fig. 10[102]. Menou et al. used a multi-objective optimization GA, together with solid solution hardening and thermodynamic modeling (CALPHAD), to design HEAs with high hardness[124]. Combining the radial basis function neural network algorithm and first-principles calculations, Zhu et al. found the key role of Al and its significant influence on hardness in modeling the Al–Cr–Fe–Ni system[125]. In a similar Al–Co–Cr–Cu–Fe–Ni system, Su et al. formulated a property-orientated materials design strategy combining ML, design of experiment, and feedback from experiment to search for HEAs with high hardness[126]. On this basis, they further proposed ML models, including feature engineering and physical models, to provide insights for predicting the hardness of these HEAs.

In recent years, there have been several studies on the moduli of HEAs. Recent developments in the field of HEAs have sparked interest in using ML to predict moduli. Balasubramanian et al. implemented gradient boost algorithms to predict Young's modulus (E) as well as the phase structure of low-, medium-, and HEAs composed of refractory elements. The ML result was in good agreement with the experiments and revealed that the melting temperature and the enthalpy of mixing are the key features determining the E of refractory HEAs[105]. Fewer studies have evaluated the role of ML in the plasticity or strength of HEAs compared to other mechanical properties (e.g., hardness and modulus). A principal reason is that the plasticity and strength data are very sensitive to the preparation process and sample sizes,
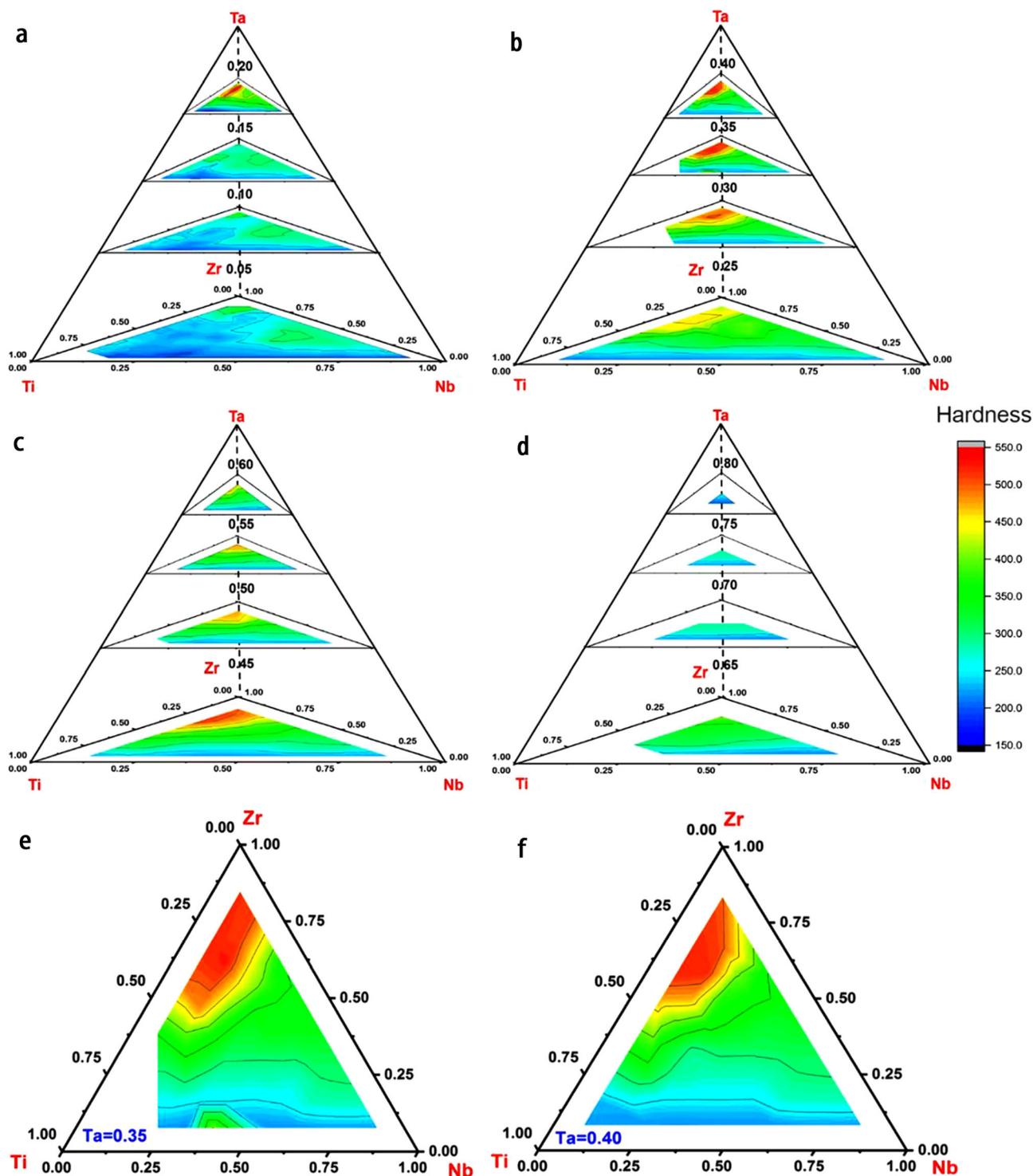
Fig. 10 | Hardness distributions as functions of the Ta content. a 0.05–0.2 at.% Ta. **b** 0.25–0.4 at.% Ta. **c** 0.45–0.6 at.% Ta. **d** 0.65–0.8 at.% Ta. **e** Hardness values were predicted using the ML model for RHEAs with Ta contents of 0.35. **f** Hardness values predicted using the ML model for RHEAs with Ta contents of 0.4[102] (adapted with permission from ref. 102. Copyright 2021 AIP Publishing).

leading to the poor quality of the original input dataset. Despite the obstacles, some attempts have been made to investigate the possibility of an ML framework for predicting the plasticity and strength of disordered alloys. Recently, Liu et al. constructed a data set through high-throughput preparation of solid solutions using powder metallurgy with Zr–Ti–Nb–O alloys as target materials[127]. Their study provides an enlightening idea for enhancing the plasticity of HEAs by tailoring key features via tuning the element content.

ML force fields

MD simulations are normally conducted with classic interatomic potentials. As these potentials often scale linearly with the number of atoms, they are computationally inexpensive, and the loss in accuracy is ignored to facilitate longer simulations or simulations with large-scale systems that include hundreds of thousands of atoms. However, the construction of force fields and tight-binding parameters is not straightforward. Given this, ML methods can provide a useful option for creating a reliable potential energy

Fig. 11 | Polycrystalline models obtained via simulation method. a The same polycrystalline model after random initialization with equimolar quantities of Nb, Mo, W, and Ta[144] (adapted with permission from ref. 144. Copyright 2020 Springer Nature). **b** Snapshot of polycrystalline model after hybrid Monte Carlo/MD simulations. **c**, **d** Defect evolution during annealing[146] (adapted with permission from ref. 146. Copyright 2021 American Physical Society).
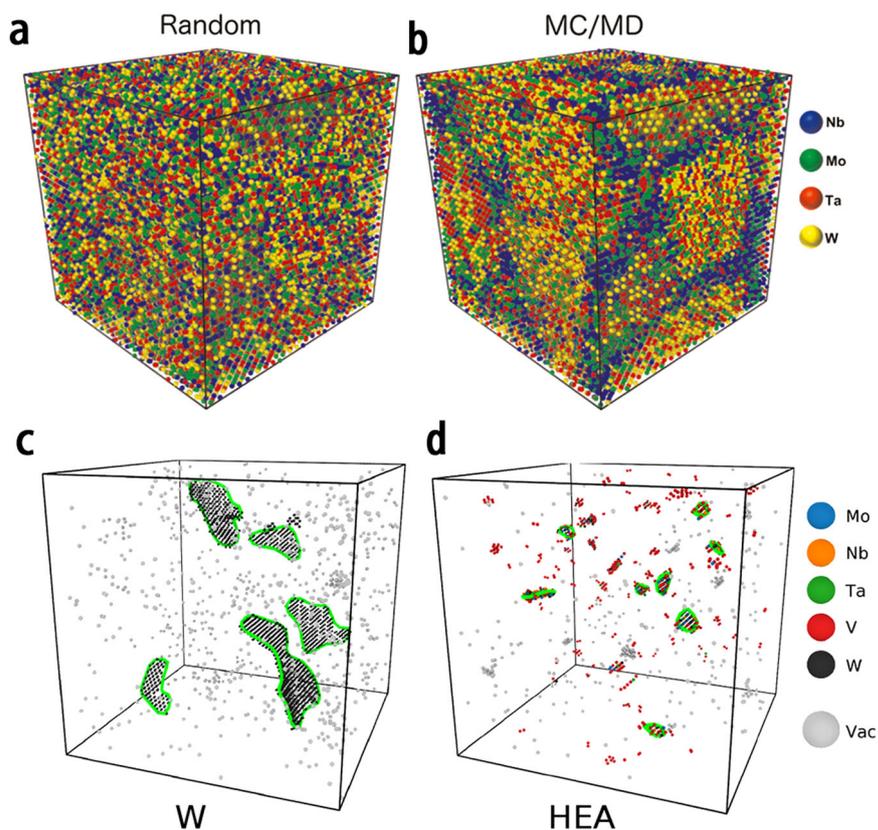
representation. Machine learning potentials (MLPs) are mathematical representations of the multidimensional potential-energy surface as a function of atomic positions. Unlike traditional potentials, reference databases of MLPs are usually generated by DFT calculations without experimental information. The other two ingredients required for MLPs are local structural descriptors, such as atom-centered symmetry function descriptors[128], the smooth overlap of atomic positions[129], and spectral neighbor analysis potential descriptors[130–132] etc., representing atomic configurations and supervised learning models to obtain reliable relations between structure and energy, force, or stress tensor[133–135].

MLPs have greatly promoted the studies of structure, thermodynamics, and mechanical properties of HEAs. Short-range ordering (SRO) refers to local chemical/structural ordering, which is a common structural feature in HEAs. It arises from the chemical interactions of constituent elements and significantly affects structural stability, and magnetic and mechanical properties[136–138]. Meshkov et al. used a low-rank potential in combination with MC simulations to investigate chemical SRO in the equiatomic fcc CoCrFeNi HEA, and demonstrated that Fe and Cr form sublattices[139]. Similar schemes were also employed to study the phase stability, phase transitions, and chemical SRO of the bcc NbMoTaW HEA by Kostiuchenko et al.[140] They claimed that if local lattice distortions are introduced, the single phase stabilizes instead of separating into sublattices until it drops to room temperature. Later on, a new algorithm combining the thermodynamic integration method with moment tensor potentials was developed by Grabowski et al. to study the anharmonic free energy of a five-component VNbMoTaW refractory HEA, which achieved DFT-level accuracy[141]. DeepMD was also applied to molten TiZrHfNb using ab initio molecular dynamics (AIMD) trajectories[142]. Structural analyses of a VZrNbHfTa melt via partial RDFs and SRO parameters were exploited using high-dimensional neural network potential, indicating that vanadium atoms are repulsed by other types of atoms[143]. Another NbMoTaW potential, adopting the SNAP model, was applied to study the complex strengthening mechanisms by modeling Nb segregations to the grain boundaries.

Applying the SNAP model, polycrystalline models with and without Monte Carlo/MD simulations were obtained, as shown in Fig. 11a–b[144]. Byggmästar et al. developed a set of Gaussian approximation potentials that were used to study segregation and radiation damage of the bcc refractory VNbMoTaW HEA[145,146]. The potentials show good accuracy and transferability in terms of elasticity, thermal stability, liquid and defect structure, and surface properties[145]. Figure 11c, d shows that the final defect structure of irradiated VNbMoTaW contains only smaller dislocation loops with respect to the pure W. In conclusion, the reduction of interstitial migration, the immovable dislocation loops, and the increase of vacancy mobility together promote the recombination of defects rather than clustering in HEAs[146]. In addition, there are some MLPs for medium entropy alloys[147–149] and high entropy ceramics[4–6]. For example, Pak et al. used Canonical Monte Carlo simulations with the ML interatomic potentials to determine the temperature conditions for the formation of single-phase and multi-phase high-entropy ceramics and claimed that for $TiZrNbHfTaC_5$ produced with electric arc discharge, the single-phase formation temperature was as high as 2000 K[6].

In general, interatomic potentials based on ML help to address the longstanding dilemma between efficiency and accuracy in MD simulations, but there are still some challenges in this field. First, the completeness of databases organized for the potentials of multicomponent chemically disordered systems is complicated and non-standardized, which is further exacerbated by short- or medium-range orders. Additionally, it is difficult to apply MLPs out of databases due to better flexibility but less extrapolation. Another concern is that MLPs are not based on physical information[150]. While active learning approaches[151] and physically informed MLPs[152] may be the solutions, further development is still needed.

Outlook

This paper presents a concise review covering several aspects of this rapidly growing field over the past two decades, from high-throughput experiments and computations to the data-driven ML of HEAs. To inspire and spur new

ideas, we present some perspectives and possible research directions in HEAs.

High-throughput characterization techniques and high-quality data acquisition for HEAs

To keep pace with continuous advancements in high-throughput material preparation methods, it is crucial to develop high-throughput characterization techniques that offer high resolution, efficiency, and affordability. From a microdomain or in situ measurement perspective, synchrotron X-ray techniques possess exceptional capabilities for high-throughput characterization of a vast array of material samples due to their remarkable brightness, and high temporal and spatial resolution, thereby alleviating the flux bottleneck in high-throughput experiments. In addition, subsequent data crafting with high quality remains an ongoing challenge. Manually extracting data with expert knowledge is a time-consuming task for thousands of articles. Thus, it is increasingly necessary to develop methods for automated data extraction that are both rapid and accurate. Techniques such as web-crawler, natural language processing, or pattern recognition could potentially facilitate the automatic extraction of information from articles or patterns such as SEM, EBSD synchrotron XRD, and others.

Metastable state of HEAs

Due to the multi-principal elements contained in HEAs and the metastable state, there is an urgent need to understand the nonequilibrium thermodynamics of HEAs from both experimental and calculation perspectives. The cooling rates of some high-throughput methods are much higher than those of traditional casting methods used for the preparation of bulk HEAs. In some extreme cases, owing in part to the multi-principal nature of HEAs, the combinational materials libraries made using high-throughput methods can form amorphous structures, which make the properties quite different from bulk HEAs. In terms of high-throughput CALPHAD, to develop a reliable thermodynamic database for HEA systems, the related binary and ternary systems should be gathered and assessed by implementing experiments and calculations.

Analysis of SRO in HEAs

To understand comprehensively the correlations between SRO and properties, and to facilitate the development of innovative alloys, it is imperative to scientifically describe and quantitatively characterize SRO in these compositionally complex alloys. However, the multi-principal element nature of HEAs poses significant challenges for direct experimental observation and accurate description of the SRO. Detailed chemical ordering information can be obtained by combining ML techniques with AIMD simulations or reverse Monte Carlo refinement methods.

Evaluation criteria and interpretability of ML methods for HEAs

In the absence of unified evaluation criteria, excessive optimism is frequently observed, resulting from overfitting and the use of unsuitable training and test data. It is essential to propose new standardized criteria to properly assess the true accuracy and performance of ML models. Prioritizing experimental validation and repeatability through code archiving can also help mitigate this issue. Additionally, the interpretability of ML models remains limited and necessitates bridging existing gaps. There is a need to develop new rules with robust interpretability and universality through ML exploration using appropriate algorithms. Techniques such as partial dependence plots, individual conditional expectation, permutation feature importance, global surrogate, local surrogate (LIME), and SHAP (SHapley Additive exPlanations) exhibit varying technical characteristics that enhance interpretability.

In summary, the future studies of high-throughput experiments, computations, and data-driven ML in HEAs will focus on a comprehensive workflow design, incorporating rational experimental design, automated high-throughput synthesis, fundamental principles of high-throughput materials characterization, computational modeling, and data mining

techniques. This multidisciplinary approach will offer a robust framework for the rational design and discovery of materials.

References

1. Zhang, Y. et al. Microstructures and properties of high-entropy alloys. *Prog. Mater. Sci.* **61**, 1–93 (2014). ***A comprehensive revier of high entropy alloys.***
2. George, E. P., Raabe, D. & Ritchie, R. O. High-entropy alloys. *Nat. Rev. Mater.* **4**, 515–534 (2019).
3. Cantor, B. Multicomponent high-entropy Cantor alloys. *Prog. Mater. Sci.* **120**, 100754 (2021).
4. Dai, F.-Z. et al. Theoretical prediction on thermal and mechanical properties of high entropy $(Zr_{0.2}Hf_{0.2}Ti_{0.2}Nb_{0.2}Ta_{0.2})C$ by deep learning potential. *J. Mater. Sci. Technol.* **43**, 168–174 (2020).
5. Dai, F.-Z. et al. Temperature dependent thermal and elastic properties of high entropy $(Ti_{0.2}Zr_{0.2}Hf_{0.2}Nb_{0.2}Ta_{0.2})B_2$: molecular dynamics simulation by deep learning potential. *J. Mater. Sci. Technol.* **72**, 8–15 (2021).
6. Pak, A. Y. et al. Machine learning-driven synthesis of $TiZrNbHfTaC_5$ high-entropy carbide. *npj Comput. Mater.* **9**, 7 (2023).
7. Li, H. et al. Fe-based bulk metallic glasses: glass formation, fabrication, properties and applications. *Prog. Mater. Sci.* https://doi.org/10.1016/j.pmatsci.2019.01.003 (2019).
8. Ye, Y. et al. High-entropy alloy: challenges and prospects. *Mater. Today* **19**, 349–362 (2016).
9. Jain, A. Commentary: the materials project: a materials genome approach to accelerating materials innovation. *APL Mater.* **1**, 011002 (2013).
10. Ramprasad, R. et al. Machine learning in materials informatics: recent applications and prospects. *npj Comput. Mater.* **3**, 1–13 (2017).
11. Liu, X. et al. Machine learning-based glass formation prediction in multicomponent alloys. *Acta Mater.* **201**, 182–190 (2020).
12. Liu, Y. et al. Machine learning in materials genome initiative: a review. *J. Mater. Sci. Technol.* **57**, 113–122 (2020).
13. de Pablo, J. J. et al. New frontiers for the materials genome initiative. *npj Comput. Mater.* **5**, 1–23 (2019).
14. Miracle, D. B. & Senkov, O. N. A critical review of high entropy alloys and related concepts. *Acta Mater.* **122**, 448–511 (2017).
15. Borkar, T. et al. A combinatorial assessment of $Al_xCrCuFeNi_2$ $(0 < x < 1.5)$ complex concentrated alloys: microstructure, microhardness, and magnetic properties. *Acta Mater.* **116**, 63–76 (2016). ***This article discusses a novel combinatorial approach for assessing composition–microstructure–microhardness–magnetic property relationships of laser deposited compositionally graded $Al_xCrCuFeNi_2$ $(0 < x < 1.5)$ complex concentrated alloys.***
16. Knoll, H. et al. Combinatorial alloy design by laser additive manufacturing. *Steel Res. Int.* **88**, 1600416 (2017).
17. Li, M. et al. Evaluation of microstructure and mechanical property variations $Al_xCoCrFeNi$ high entropy alloys produced by a high-throughput laser deposition method. *Intermetallics* **95**, 110–118 (2018).
18. Li, M. & Flores, K. M. Laser processing as a high-throughput method to investigate microstructure–processing–property relationships in multiprincipal element alloys. *J. Alloys Compd.* **825**, 154025 (2020).
19. Melia, M. A. et al. High-throughput additive manufacturing and characterization of refractory high entropy alloys. *Appl. Mater. Today* **19**, 100560 (2020).
20. Moorehead, M. et al. High-throughput synthesis of Mo–Nb–Ta–W high-entropy alloys via additive manufacturing. *Mater. Des.* **187**, 108358 (2020).

21. Pegues, J. W. et al. Exploring additive manufacturing as a high-throughput screening tool for multiphase high entropy alloys. *Addit. Manuf.* **37**, 101598 (2021).

22. Huang, X. et al. Machine learning assisted modelling and design of solid solution hardened high entropy alloys. *Mater. Des.* **211**, 110177 (2021).

23. Tsai, P. & Flores, K. M. High-throughput discovery and characterization of multicomponent bulk metallic glass alloys. *Acta Mater.* **120**, 426–434 (2016).

24. Kelly, P. J. & Arnell, R. D. Magnetron sputtering: a review of recent developments and applications. *Vacuum* **56**, 159–172 (2000).

25. Ding, S. et al. Combinatorial development of bulk metallic glasses. *Nat. Mater.* **13**, 494–500 (2014).

26. Liu, Y. et al. Combinatorial development of antibacterial Zr–Cu–Al–Ag thin film metallic glasses. *Sci. Rep.* **6**, 1–8 (2016).

27. Kauffmann, A. et al. Combinatorial exploration of the high entropy alloy system Co–Cr–Fe–Mn–Ni. *Surf. Coat. Technol.* **325**, 174–180 (2017).

28. Xing, Q. et al. High-throughput screening solar-thermal conversion films in a pseudobinary (Cr, Fe, V)–(Ta, W) system. *ACS Comb. Sci.* **20**, 602–610 (2018).

29. Zhang, Y. et al. Compositional gradient films constructed by sputtering in a multicomponent Ti–Al–(Cr, Fe, Ni) system. *J. Mater. Res.* **33**, 3330–3338 (2018).

30. Li, M.-X. et al. High-temperature bulk metallic glasses developed by combinatorial methods. *Nature* **569**, 99–103 (2019).

31. Banko, L. et al. Unravelling composition–activity–stability trends in high entropy alloy electrocatalysts by using a data-guided combinatorial synthesis strategy and computational modeling. *Adv. Energy Mater.* **12**, 2103312 (2022). ***A strategy for effective extensions of high-dimensional composition spaces for the exemplary Ru–Rh–Pd–Ir–Pt system covered by combinatorial synthesis was demonstrated***.

32. Marshal, A. et al. Combinatorial evaluation of phase formation and magnetic properties of FeMnCoCrAl high entropy alloy thin film library. *Sci. Rep.* **9**, 1–11 (2019).

33. Ren, F. et al. Accelerated discovery of metallic glasses through iteration of machine learning and high-throughput experiments. *Sci. Adv.* **4**, eaaq1566 (2018).

34. Kube, S. A. et al. Phase selection motifs in high entropy alloys revealed through combinatorial methods: large atomic size difference favors BCC over FCC. *Acta Mater.* **166**, 677–686 (2019).

35. Datye, A. et al. Accelerated discovery and mechanical property characterization of bioresorbable amorphous alloys in the Mg–Zn–Ca and the Fe–Mg–Zn systems using high-throughput methods. *J. Mater. Chem. B* **7**, 5392–5400 (2019).

36. Ding, S. et al. Solidification of Au–Cu–Si alloys investigated by a combinatorial approach. *J. Appl. Phys.* **111**, 114901 (2012).

37. Zhao, J.-C., Jackson, M. & Peluso, L. Determination of the Nb–Cr–Si phase diagram using diffusion multiples. *Acta Mater.* **51**, 6395–6405 (2003).

38. Zhao, J.-C. et al. A diffusion multiple approach for the accelerated design of structural materials. *MRS Bull.* **27**, 324–329 (2002).

39. Zhao, J.-C., Zheng, X. & Cahill, D. G. High-throughput diffusion multiples. *Mater. Today* **8**, 28–37 (2005).

40. Zhao, J.-C. Combinatorial approaches as effective tools in the study of phase diagrams and composition–structure–property relationships. *Prog. Mater. Sci.* **51**, 557–631 (2006).

41. Zhao, J.-C., Zheng, X. & Cahill, D. G. High-throughput measurements of materials properties. *JOM* **63**, 40–44 (2011).

42. Wilson, P., Field, R. & Kaufman, M. The use of diffusion multiples to examine the compositional dependence of phase stability and hardness of the Co–Cr–Fe–Mn–Ni high entropy alloy system. *Intermetallics* **75**, 15–24 (2016).

43. Chen, W. & Zhang, L. High-throughput determination of interdiffusion coefficients for Co–Cr–Fe–Mn–Ni high-entropy alloys. *J. Phase Equilib. Diffus.* **38**, 457–465 (2017).

44. Coury, F. G. et al. High-throughput solid solution strengthening characterization in high entropy alloys. *Acta Mater.* **167**, 1–11 (2019).

45. Ding, W. et al. Diffusion bonding of copper to titanium using CoCrFeMnNi high-entropy alloy interlayer. *Intermetallics* **129**, 107027 (2021).

46. Tsai, K.-Y., Tsai, M.-H. & Yeh, J.-W. Sluggish diffusion in Co–Cr–Fe–Mn–Ni high-entropy alloys. *Acta Mater.* **61**, 4887–4897 (2013).

47. Kucza, W. et al. Studies of "sluggish diffusion" effect in Co–Cr–Fe–Mn–Ni, Co–Cr–Fe–Ni and Co–Fe–Mn–Ni high entropy alloys; determination of tracer diffusivities by combinatorial approach. *J. Alloys Compd.* **731**, 920–928 (2018).

48. Wang, T. et al. Effect of reactive alloy elements on friction stir welded butt joints of metallurgically immiscible magnesium alloys and steel. *J. Manuf. Processes* **39**, 138–145 (2019).

49. Wang, T. et al. Towards heterogeneous Al$_x$CoCrFeNi high entropy alloy via friction stir processing. *Mater. Lett.* **236**, 472–475 (2019).

50. Sinha, S. et al. Immiscible nanostructured copper–aluminum–niobium alloy with excellent precipitation strengthening upon friction stir processing and aging. *Scr. Mater.* **164**, 42–47 (2019).

51. Agrawal, P. et al. Friction stir gradient alloying: a high-throughput method to explore the influence of V in enabling HCP to BCC transformation in a γ-FCC dominated high entropy alloy. *Appl. Mater. Today* **21**, 100853 (2020).

52. Shukla, S. et al. Friction stir gradient alloying: a novel solid-state high throughput screening technique for high entropy alloys. *Mater. Today Commun.* **23**, 100869 (2020).

53. Tong, L. & Reddy, R. G. Synthesis of titanium carbide nano-powders by thermal plasma. *Scr. Mater.* **52**, 1253–1258 (2005).

54. Zhu, B. et al. Fast and high-throughput synthesis of medium-and high-entropy alloys using radio frequency inductively coupled plasma. *Adv. Eng. Mater.* **23**, 2001116 (2021).

55. Shi, Y. et al. High-throughput synthesis and corrosion behavior of sputter-deposited nanocrystalline Al$_x$(CoCrFeNi) 100 − x combinatorial high-entropy alloys. *Mater. Des.* **195**, 109018 (2020).

56. Haase, C. et al. Combining thermodynamic modeling and 3D printing of elemental powder blends for high-throughput investigation of high-entropy alloys—towards rapid alloy screening and design. *Mater. Sci. Eng. A* **688**, 180–189 (2017).

57. Kaufmann, K. et al. Crystal symmetry determination in electron diffraction using machine learning. *Science* **367**, 564–568 (2020).

58. Kaufmann, K. et al. Efficient few-shot machine learning for classification of EBSD patterns. *Sci. Rep.* **11**, 8172 (2021).

59. Tsutsui, K. et al. Microstructural diagram for steel based on crystallography with machine learning. *Comput. Mater. Sci.* **159**, 403–411 (2019).

60. Yoo, Y. K. et al. Identification of amorphous phases in the Fe–Ni–Co ternary alloy system using continuous phase diagram material chips. *Intermetallics* **14**, 241–247 (2006).

61. Hui, J. et al. High-throughput investigation of crystal-to-glass transformation of Ti–Ni–Cu ternary alloy. *Sci. Rep.* **9**, 1–8 (2019).

62. Joress, H. et al. A high-throughput structural and electrochemical study of metallic glass formation in Ni–Ti–Al. *ACS Comb. Sci.* **22**, 330–338 (2020).

63. Haque, M. & Saif, M. A review of MEMS-based microscale and tensile and bending testing. *Exp. Mech.* **43**, 248–255 (2003).

64. Uchic, M. D. et al. Sample dimensions influence strength and crystal plasticity. *Science* **305**, 986–989 (2004).

65. McCluskey, P. J. et al. Precipitation and thermal fatigue in Ni–Ti–Zr shape memory alloy thin films by combinatorial nanocalorimetry. *Acta Mater.* **59**, 5116–5124 (2011).

66. Kim, H.-J. et al. High-throughput analysis of thin-film stresses using arrays of micromachined cantilever beams. *Rev. Sci. Instrum.* **79**, 045112 (2008).

67. Figiel, H., Zogał, O. & Yartys, V. Effect of iron content on the microstructure evolution, mechanical properties and wear resistance of FeXCoCrNi high-entropy alloy system produced via MA-SPS Parisa. *J. Alloys Compd.* **404**, 1 (2005).

68. Arunkumar, S. Overview of small punch test. *Met. Mater. Int.* **26**, 719–738 (2020).

69. Cai, Y. et al. Fracture and wear mechanisms of FeMnCrNiCo + x(TiC) composite high-entropy alloy cladding layers. *Appl. Surf. Sci.* **543**, 148794 (2021).

70. Marshal, A. et al. Combinatorial synthesis of high entropy alloys: introduction of a novel, single phase, body-centered-cubic FeMnCoCrAl solid solution. *J. Alloys Compd.* **691**, 683–689 (2017).

71. Wei, T. et al. Scanning tip microwave near-field microscope. *Appl. Phys. Lett.* **68**, 3506–3508 (1996).

72. Gao, C., Duewer, F. & Xiang, X.-D. Quantitative microwave evanescent microscopy. *Appl. Phys. Lett.* **75**, 3005–3007 (1999).

73. Turchinskaya, M. et al. Rapid constructing magnetic phase diagrams by magneto-optical imaging of composition spread films. *J. Mater. Res.* **19**, 2546–2548 (2004).

74. Plimpton, S. Fast parallel algorithms for short-range molecular dynamics. *J. Comput. Phys.* **117**, 1–19 (1995).

75. Kresse, G. & Furthmüller, J. Efficiency of ab initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput. Mater. Sci.* **6**, 15–50 (1996).

76. Raicu, I. (ed) *Many-task Computing: Bridging the Gap Between High-Throughput Computing and High-performance Computing* (University of Chicago, 2009).

77. Ong, S. P. et al. Python materials genomics (pymatgen): a robust, open-source python library for materials analysis. *Comput. Mater. Sci.* **68**, 314–319 (2013).

78. Jain, A. et al. FireWorks: a dynamic workflow system designed for high-throughput applications. *Concurrency Comput. Pract. Exper.* **27**, 5037–5059 (2015).

79. Mathew, K. et al. Atomate: a high-level interface to generate, execute, and analyze computational materials science workflows. *Comput. Mater. Sci.* **139**, 140–152 (2017).

80. Wang, G. et al. ALKEMIE: an intelligent computational platform for accelerating materials discovery and design. *Comput. Mater. Sci.* **186**, 110064 (2021).

81. Yang, X. et al. MatCloud: a high-throughput computational infrastructure for integrated management of materials simulation, data and resources. *Comput. Mater. Sci.* **146**, 319–333 (2018).

82. Kirklin, S. et al. The open quantum materials database (OQMD): assessing the accuracy of DFT formation energies. *npj Comput. Mater.* **1**, 1–15 (2015).

83. Soven, P. Coherent-potential model of substitutional disordered alloys. *Phys. Rev.* **156**, 809 (1967).

84. Tian, F. A review of solid-solution models of high-entropy alloys based on ab initio calculations. *Front. Mater.* **4**, 36 (2017).

85. Aitken, Z. H., Sorkin, V. & Zhang, Y.-W. Atomistic modeling of nanoscale plasticity in high-entropy alloys. *J. Mater. Res.* **34**, 1509–1532 (2019).

86. Santodonato, L. J. et al. Predictive multiphase evolution in Al-containing high-entropy alloys. *Nat. Commun.* **9**, 4520 (2018).

87. C, S. C. A. B. et al. AFLOWLIB.ORG: a distributed materials properties repository from high-throughput ab initio calculations. *Comput. Mater. Sci.* **58**, 227–235 (2012).

88. Lederer, Y. et al. The search for high entropy alloys: a high-throughput ab initio approach. *Acta Mater.* **159**, 364–383 (2018).

89. Kaufman, L. & Bernstein, H. (eds) *Computer Calculation of Phase Diagrams. With Special Reference to Refractory Metals* (Academic Press, 1970).

90. Zhang, C. et al. Computational thermodynamics aided high-entropy alloy design. *JOM* **64**, 839–845 (2012).

91. Gurao, N. & Biswas, K. In the quest of single phase multi-component multiprincipal high entropy alloys. *J. Alloys Compd.* **697**, 434–442 (2017).

92. Chen, H.-L., Mao, H. & Chen, Q. Database development and Calphad calculations for high entropy alloys: challenges, strategies, and tips. *Mater. Chem. Phys.* **210**, 279–290 (2018).

93. Klaver, T., Simonovic, D. & Sluiter, M. H. Brute force composition scanning with a CALPHAD database to find low temperature body centered cubic high entropy alloys. *Entropy* **20**, 911 (2018).

94. Yang, S. et al. Revisit the VEC rule in high entropy alloys (HEAs) with high-throughput CALPHAD approach and its applications for material design—a case study with Al–Co–Cr–Fe–Ni system. *Acta Mater.* **192**, 11–19 (2020).

95. Feng, R. et al. High-throughput design of high-performance lightweight high-entropy alloys. *Nat. Commun.* **12**, 4329 (2021).

96. Cubuk, E. D. et al. Identifying structural flow defects in disordered solids using machine-learning methods. *Phys. Rev. Lett.* **114**, 108001 (2015).

97. Wang, A. Y.-T. et al. Machine learning for materials scientists: an introductory guide toward best practices. *J. Chem. Mater.* **32**, 4954–4965 (2020).

98. Jablonka, K. M. et al. Big-data science in porous materials: materials genomics and machine learning. *Chem. Rev.* **120**, 8066–8129 (2020).

99. Huang, W., Martin, P. & Zhuang, H. Machine-learning phase prediction of high-entropy alloys. *Acta Mater.* **169**, 225–236 (2019).

100. Pei, Z. et al. Machine-learning informed prediction of high-entropy solid solution formation: beyond the Hume–Rothery rules. *npj Comput. Mater.* **6**, 1–8 (2020).

101. Kaufmann, K. & Vecchio, K. S. Searching for high entropy alloys: a machine learning approach. *Acta Mater.* **198**, 178–222 (2020). ***A novel high-throughput approach called "ML-HEA" was proposed to predict the solid solution forming ability by coupling thermodynamic and chemical features with a random forest machine learning model.***

102. Sun, Y. et al. Prediction of Ti–Zr–Nb–Ta high-entropy alloys with desirable hardness by combining machine learning and experimental data. *Appl. Phys. Lett.* **119**, 201905 (2021). ***This work combines a machine learning (ML) model with phase diagram calculations (CALPHAD) to design Ti–Zr–Nb–Ta refractory HEAs with a desirable hardness.***

103. Kaufmann, K. et al. Discovery of high-entropy ceramics via machine learning. *npj Comput. Mater.* **6**, 9 (2020).

104. Kim, G. et al. First-principles and machine learning predictions of elasticity in severely lattice-distorted high-entropy alloys with experimental validation. *Acta Mater.* **181**, 124–138 (2019).

105. Roy, A. et al. Machine learned feature identification for predicting phase and Young's modulus of low-, medium-and high-entropy alloys. *Scr. Mater.* **185**, 152–158 (2020).

106. Guo, S. et al. Effect of valence electron concentration on stability of fcc or bcc phase in high entropy alloys. *J. Appl. Phys.* **109**, 103505 (2011).

107. Crisci, C., Ghattas, B. & Perera, G. A review of supervised machine learning algorithms and their applications to ecological data. *Ecol. Modell.* **240**, 113–122 (2012).

108. Najafabadi, M. M. et al. Deep learning applications and challenges in big data analytics. *J. Big Data* **2**, 1–21 (2015).

109. Islam, N., Huang, W. & Zhuang, H. Machine learning for phase selection in multi-principal element alloys. *Comput. Mater. Sci.* **150**, 230–235 (2018).

110. Jha, R. et al. Combined machine learning and CALPHAD approach for discovering processing-structure relationships in soft magnetic alloys. *Comput. Mater. Sci.* **150**, 202–211 (2018).

111. Choudhury, A. et al. Structure prediction of multi-principal element alloys using ensemble learning. *Eng. Comput.* **37**, 1003–1022 (2020).

112. Zhou, Z. et al. Machine learning guided appraisal and exploration of phase design for high entropy alloys. *npj Comput. Mater.* **5**, 1–9 (2019).

113. Bu, C. & Zhang, Z. Research on overfitting problem and correction in machine learning. *J. Phys. Conf. Ser.* https://doi.org/10.1088/1742-6596/1693/1/012100 (2020).

114. Dai, D. et al. Using machine learning and feature engineering to characterize limited material datasets of high-entropy alloys. *Comput. Mater. Sci.* **175**, 109618 (2020).

115. Li, R. et al. High-throughput calculations for high-entropy alloys: a brief review. *Front. Mater.* **7**, 290 (2020).

116. Sparks, T. D. et al. Machine learning for structural materials. *Annu. Rev. Mater. Res.* **50**, 27–48 (2020).

117. Wu, Q. et al. Uncovering the eutectics design by machine learning in the Al–Co–Cr–Fe–Ni high entropy system. *Acta Mater.* **182**, 278–286 (2020).

118. Krishna, Y. V., Jaiswal, U. K. & Rahul, M. Machine learning approach to predict new multiphase high entropy alloys. *Scr. Mater.* **197**, 113804 (2021).

119. Lee, S. Y. et al. Deep learning-based phase prediction of high-entropy alloys: optimization, generation, and explanation. *Mater. Des.* **197**, 109260 (2021).

120. Machaka, R. Machine learning-based prediction of phases in high-entropy alloys. *Comput. Mater. Sci.* **188**, 110244 (2021).

121. Rao, Z. et al. Machine learning-enabled high-entropy alloy discovery. *Science* **378**, 78–85 (2022).

122. Pei, Z. et al. Toward the design of ultrahigh-entropy alloys via mining six million texts. *Nat. Commun.* **14**, 54 (2023).

123. Wen, C. et al. Modeling solid solution strengthening in high entropy alloys using machine learning. *Acta Mater.* **212**, 116917 (2021).

124. Menou, E. et al. Computational design of light and strong high entropy alloys (HEA): obtainment of an extremely high specific solid solution hardening. *Scr. Mater.* **156**, 120–123 (2018).

125. Qiao, L. et al. Modelling and prediction of hardness in multi-component alloys: a combined machine learning, first principles and experimental study. *J. Alloy. Compd.* **853**, 156959 (2021).

126. Wen, C. et al. Machine learning assisted design of high entropy alloys with desired property. *Acta Mater.* **170**, 109–117 (2019).

127. Si, S. et al. Study on strengthening effects of Zr–Ti–Nb–O alloys via high throughput powder metallurgy and data-driven machine learning. *Mater. Des.* **206**, 109777 (2021).

128. Behler, J. Atom-centered symmetry functions for constructing high-dimensional neural network potentials. *J. Chem. Phys.* **134**, 074106 (2011).

129. Bartók, A. P., Kondor, R. & Csányi, G. On representing chemical environments. *Phys. Rev. B* **87**, 184115 (2013).

130. Thompson, A. P. et al. Spectral neighbor analysis method for automated generation of quantum-accurate interatomic potentials. *J. Comput. Phys.* **285**, 316–330 (2015).

131. Chen, C. et al. Accurate force field for molybdenum by machine learning large materials data. *Phys. Rev. Mater.* **1**, 043603 (2017).

132. Li, X.-G. et al. Quantum-accurate spectral neighbor analysis potential models for Ni–Mo binary alloys and fcc metals. *Phys. Rev. B* **98**, 094104 (2018).

133. Behler, J. & Parrinello, M. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.* **98**, 146401 (2007).

134. Bartók, A. P. et al. Gaussian approximation potentials: the accuracy of quantum mechanics, without the electrons. *Phys. Rev. Lett.* **104**, 136403 (2010).

135. Shapeev, A. V. Moment tensor potentials: a class of systematically improvable interatomic potentials. *Multiscale Model. Simul.* **14**, 1153–1173 (2016).

136. Lei, Z. et al. Enhanced strength and ductility in a high-entropy alloy via ordered oxygen complexes. *Nature* **563**, 546–550 (2018).

137. Ding, Q. et al. Tuning element distribution, structure and properties by composition in high-entropy alloys. *Nature* **574**, 223–227 (2019).

138. Zhang, L. et al. The effect of randomness on the strength of high-entropy alloys. *Acta Mater.* **166**, 424–434 (2019).

139. Meshkov, E. et al. Sublattice formation in CoCrFeNi high-entropy alloy. *Intermetallics* **112**, 106542 (2019).

140. Kostiuchenko, T. et al. Impact of lattice relaxations on phase transitions in a high-entropy alloy studied by machine-learning potentials. *npj Comput. Mater.* **5**, 1–7 (2019). ***This work proposed an efficient computational method based on machine-learning potentials and combined Monte Carlo simulations to study phase stability, phase transitions, and chemical short-range order of HEAs.***

141. Grabowski, B. et al. Ab initio vibrational free energies including anharmonicity for multicomponent alloys. *npj Comput. Mater.* **5**, 1–6 (2019).

142. Balyakin, I. & Rempel, A. Machine learning interatomic potential for molten TiZrHfNb. *AIP Conf. Proc.* **2313**, 030037 (2020).

143. Balyakin, I. et al. Ab initio molecular dynamics and high-dimensional neural network potential study of VZrNbHfTa melt. *J. Phys. Condens. Matter* **32**, 214006 (2020).

144. Li, X.-G. et al. Complex strengthening mechanisms in the NbMoTaW multi-principal element alloy. *npj Comput. Mater.* **6**, 1–10 (2020).

145. Byggmästar, J., Nordlund, K. & Djurabekova, F. Gaussian approximation potentials for body-centered-cubic transition metals. *Phys. Rev. Mater.* **4**, 093802 (2020).

146. Byggmästar, J., Nordlund, K. & Djurabekova, F. Modeling refractory high-entropy alloys with efficient machine-learned interatomic potentials: defects and segregation. *Phys. Rev. B* **10**, 104101 (2021).

147. Jafary-Zadeh, M. et al. Applying a machine learning interatomic potential to unravel the effects of local lattice distortion on the elastic properties of multi-principal element alloys. *J. Alloys Compd.* **803**, 1054–1062 (2019).

148. Kostiuchenko, T. et al. Short-range order in face-centered cubic VCoNi alloys. *Phys. Rev. Mater.* **4**, 113802 (2020).

149. Zhao, L. et al. Anomalous dislocation core structure in shock compressed bcc high-entropy alloys. *Acta Mater.* **209**, 116801 (2021).

150. Behler, J. Perspective: machine learning potentials for atomistic simulations. *J. Chem. Phys.* **145**, 170901 (2016).

151. Podryabinkin, E. V. & Shapeev, A. V. Active learning of linearly parametrized interatomic potentials. *Comput. Mater. Sci.* **140**, 171–180 (2017).

152. Pun, G. P. et al. Physically informed artificial neural networks for atomistic modeling of materials. *Nat. Commun.* **10**, 1–10 (2019).

153. Meng, H. et al. Formation ability descriptors for high-entropy diborides established through high-throughput experiments and machine learning. *Acta Mater.* **256**, 119132 (2023).

154. Jaafreh, R. et al. Machine learning guided discovery of super-hard high entropy ceramics. *Mater. Lett.* **306**, 130899 (2022).

155. Xiong, J., Shi, S.-Q. & Zhang, T.-Y. Machine learning of phases and mechanical properties in complex concentrated alloys. *J. Mater. Sci. Technol.* **87**, 133–142 (2021).

156. Bhandari, U. et al. Yield strength prediction of high-entropy alloys using machine learning. *Mater. Today Commun.* **26**, 101871 (2021).

157. Wang, J. et al. A neural network model for high entropy alloy design. *npj Comput. Mater.* **9**, 60 (2023).

Acknowledgements

Author contributions

Zhichao L. and Zhaoping L. conceived the idea and concept, wrote the initial manuscript, and validated the discussion. X.L. revised the paper. D.M. and Zhaoping L. supervised the work, led the project, and contributed to the final writing. All authors discussed and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Liu Xiongjun or Zhaoping Lu.

Peer review information *Communications Materials* thanks Vineeth Venugopal, Ben Breitung, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editors: Eun Soo Park and John Plummer.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.