scientific reports

OPEN

High-throughput materials screening algorithm based on first-principles density functional theory and artificial neural network for high-entropy alloys

Meena Rittiruam[1,2,3], Jakapob Noppakhun[1,2,3], Sorawee Setasuban[1,2,4], Nuttanon Aumnongpho[1,2,3], Attachai Sriwattana[1,2,4], Suphawich Boonchuay[1,2,4], Tinnakorn Saelee[1,2,4], Chanthip Wangphon[1,2,4], Annop Ektarawong[5,6], Patchanee Chammingkwan[7], Toshiaki Taniike[7], Supareak Praserthdam[1,2✉] & Piyasan Praserthdam[2]

This work introduced the high-throughput phase prediction of PtPd-based high-entropy alloys via the algorithm based on a combined Korringa-Kohn-Rostoker coherent potential approximation (KKR-CPA) and artificial neural network (ANN) technique. As the first step, the KKR-CPA was employed to generate 2,720 data of formation energy and lattice parameters in the framework of the first-principles density functional theory. Following the data generation, 15 features were selected and verified for all HEA systems in each phase (FCC and BCC) via ANN. The algorithm exhibited high accuracy for all four prediction models on 36,556 data from 9139 HEA systems with 137,085 features, verified by R^2 closed to unity and the mean relative error (MRE) within 5%. From this dataset comprising 5002 and 4137 systems of FCC and BCC phases, it can be realized based on the highest tendency of HEA phase formation that (1) Sc, Co, Cu, Zn, Y, Ru, Cd, Os, Ir, Hg, Al, Si, P, As, and Tl favor FCC phase, (2) Hf, Ga, In, Sn, Pb, and Bi favor BCC phase, and (3) Ti, V, Cr, Mn, Fe, Ni, Zr, Nb, Mo, Tc, Rh, Ag, Ta, W, Re, Au, Ge, and Sb can be found in both FCC and BCC phases with comparable tendency, where all predictions are in good agreement with the data from the literature. Thus, the combination of KKR-CPA and ANN can reduce the computational cost for the screening of PtPd-based HEA and accurately predict the structure, i.e., FCC, BCC, etc.

High-entropy alloys (HEAs) are classified by configurational entropy of mixing (ΔS)[1], in which the criteria are $\Delta S \geq 1.36R$ and $\Delta S \geq 1.50R$ for tetra and penta-metallic alloys[2], respectively. This material has been employed in various applications due to its promising properties, especially catalytic[3–8] and mechanical properties[9–12]. Nowadays, discovering new formulae of HEA via experimental techniques requires a high cost of chemicals and characterization[13], where the phase and atomic composition are challenges for HEA materials[14]. Hence, the prediction of the possible phase formation based on computational techniques plays a crucial role in reducing

[1]High-Performance Computing Unit (CECC-HCU), Center of Excellence on Catalysis and Catalytic Reaction Engineering (CECC), Chulalongkorn University, Bangkok 10330, Thailand. [2]Center of Excellence on Catalysis and Catalytic Reaction Engineering (CECC), Chulalongkorn University, Bangkok 10330, Thailand. [3]Rittiruam Research Group, Chulalongkorn University, Bangkok 10330, Thailand. [4]Saelee Research Group, Chulalongkorn University, Bangkok 10330, Thailand. [5]Extreme Conditions Physics Research Laboratory and Center of Excellence in Physics of Energy Materials (CE:PEM), Department of Physics, Faculty of Science, Chulalongkorn University, Bangkok 10330, Thailand. [6]Chula Intelligent and Complex Systems, Faculty of Science, Chulalongkorn University, Bangkok 10330, Thailand. [7]Graduate School of Advanced Science and Technology, Japan Advanced Institute of Science and Technology, 1-1 Asahidai, Nomi, Ishikawa 923-1292, Japan. ✉email: supareak.p@chula.ac.th
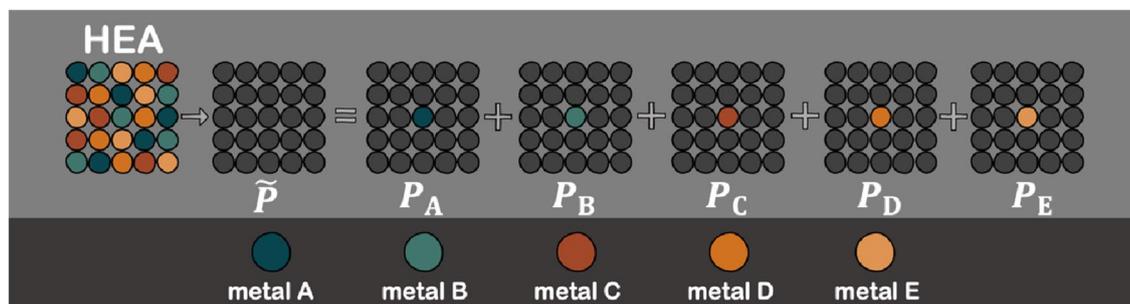
Figure 1. The coherent potential approximation (CPA) for the penta-metallic high-entropy alloy (HEA) recreated from Tian et al.[53].

HEA screening costs. It was demonstrated that a system with high ΔS in a multi-component alloy tended to form a single-phase HEA, implying that the system would be less likely to segregate[15].

During these several years, machine learning (ML) techniques have been employed to predict structural properties and discover unknown materials[16]. However, a large amount of data is required for ML to predict such properties accurately. This will be the main obstacle if one would like to screen the materials based on experimental techniques. Complementing the experimental data if it is not enough; one should employ the first-principles techniques, e.g., density functional theory (DFT), to help generating enough training data for accurate ML prediction[17,18]. This coupling of ML and DFT can tremendously reduce the cost of materials discovery compared to the complete experimental materials screening, where only the expensive screening via high-throughput experimentation would be enough for ML[19]. Therefore, various researchers employed the ML-based method to explore new HEA in recent years, as Kaufmann and Vecchio[20] reported, which employed the random forest technique to predict the single and multiple phases of binary, ternary, quaternary, and quinary alloys. They found that all predicted results agreed well with the validation from experimental data and the CALPHAD program. Moreover, Jin et al.[21] exhibit the coupled DFT-ML technique to screen the phases of multi-component alloys. The training dataset of binary alloys was generated by DFT random alloy method, known as the Korringa-Kohn-Rostoker coherent potential approximation (KKR-CPA). As a result, their prediction accuracy was up to 80.56% for multi-component alloys and, interestingly, up to 84.20% for HEA materials[21]. Apart from such an algorithm, numerous techniques, i.e., the gradient boosting model, trained with 1,807 datasets, demonstrated high accuracy of 96.41% for predicting single-phase and non-single-phase refractory HEAs (RHEAs)[22]. Other methods also exhibit high accuracy prediction, e.g., a combined ML and CALPHAD technique, an artificial neural network technique (ANN) coupled with experimental data[9,10,12,23–28]. In addition to phase formation of HEAs, machine learning was recently employed to predict the mechanical properties of HEA bulk materials, including microhardness[10,27], yield strength[12,23], dislocation density[12], elastic modulus[29], Young's modulus[30], hardness[11,31], and elastic constant[32]. These shed light on the ML-based high-throughput screening of HEA materials.

As in the aspect of the use of HEA as catalytic materials, the PtPd-based HEAs are among high potential candidates that can be utilized in CO_2 and CO reduction reaction[7,8], oxygen evolution reaction[33], oxygen reduction reaction[34], and hydrogen evolution reaction[35]. However, from the literature, it was found that most experimental works involving PtPd-based HEAs focused mainly on the report of limited formulae of HEA in terms of novel synthesis and characterization techniques[26,36–43]. Therefore, a gigantic set of possible HEA formulae is required to understand and utilize their promising properties fully. Hence, a firm protocol for screening HEA properties, such as their phase and electronic properties, must be studied and established. The phases and structural properties should be the first screening stage, in which bulk HEAs' thermodynamic stability can be extracted through the formation energy (ΔE_f) at the ground-state configuration at 0 K[1]. Regarding such calculations, Miedema's scheme model[44] is one of the simple and fast methods used to obtain ΔE_f, where the experimental technique must be used to obtain the individual energy for a monometallic component first. Yet, apart from the energy term, the structural information of a specific phase of alloys or compounds cannot be determined. To overcome such a problem, the periodic DFT was employed to determine both the ΔE_f and the structural information of a system is reported being successfully utilized in binary systems[45]. Nevertheless, to design the HEA, many possible atomic configurations have to be guessed and verified through thermodynamically optimized structures. The routine calculation via DFT alone would be intensive, although all information can be obtained. Also, suppose a combined DFT-ML technique is to be used for HEA discovery. In that case a huge configurational space must be generated to accurately determine the most favorable atomic configuration at a given operating temperature, pressure, and chemical composition. Thus, to reduce the computational cost, even more, we propose the combined KKR-CPA with ML technique for a rapid and low-cost HEA materials screening[46–49]. This method was proven by Jin et al.[21] to be a successful tool for the property prediction of multi-metallic alloys. So far, such a method has not been applied to HEA screening. The KKR method employed is based on Green's function method of multiple scattering theory to calculate the Green's function of a system without knowing its eigenvalue of DFT[50–52]. Whereas the CPA method solves the problem of atomic configuration through the effective medium with the weighted average of Green's function.

Accordingly, the CPA method is the most suitable for the multi-component system, such as the penta-metallic HEA comprising five metals: A, B, C, D, and E, as illustrated in Fig. 1. The CPA assumes the potential of each element (P_A, P_B, P_C, P_D, P_E) without the effect of the local environment, where the atomic component is placed in
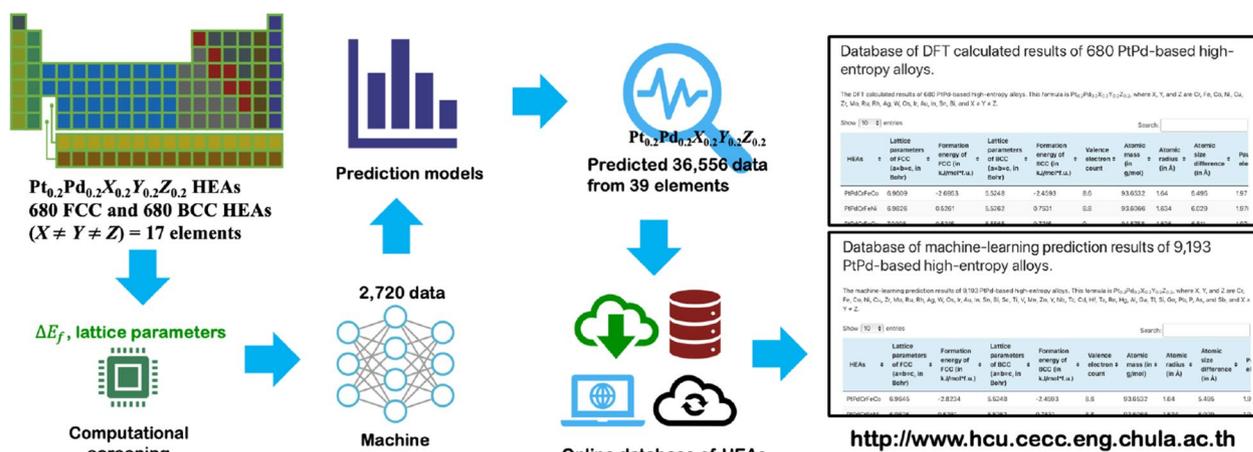
Figure 2. Our workflow based on first-principles density functional theory (DFT) and machine learning (ML) for the prediction of lattice parameter (*a*) and formation energy (ΔH_f) of HEA systems involving the utilization of our public database at http://www.hcu.cecc.eng.chula.ac.th/hea-database.

the system in the effective medium $(\widetilde{P})$[53]. $\widetilde{P}$ is determined selfconsistently with including all single site scattering events due to the respective atoms placed in the effective medium. It was previously reported that the KKR-CPA results were comparable to those obtained from a combined DFT and SQS (special quasi-random supercell) technique[54]. The validation of the KKR-CPA method for multi-component alloys is extensively explained in Sect. 1 of the Supporting Information based on the Akai-KKR software. Tables S1 and S2 demonstrate that the Akai-KKR package can indicate the stable phase for FCC and BCC alloys with a good agreement with available literature[54–65].

This work presents the ML prediction on the thermodynamically stable PtPd-based HEAs via ΔE_f. The protocol is illustrated in Fig. 2. HEA formulae in focus are the $Pt_{0.2}Pd_{0.2}X_{0.2}Y_{0.2}Z_{0.2}$ with *X*, *Y*, and *Z* being the components from a pool of 39 elements, i.e., Sc, Ti, V, Cr, Mn, Fe, Co, Ni, Cu, Zn, Y, Zr, Nb, Mo, Tc, Ru, Rh, Ag, Cd, Hf, Ta, W, Re, Os, Ir, Au, Hg, Al, Si, P, Ga, Ge, As, In, Sn, Sb, Tl, Pb, and Bi, where $X \neq Y \neq Z$. The training and test datasets of lattice parameters (*a*) and ΔE_f are obtained from the calculated results of 680 FCC and BCC HEA systems, constructed from 17 out of 39 elements mentioned above. The details of the DFT calculation and ML prediction model can be found in Sect. 2 of the Supporting Information. Consequently, the *a*(FCC), *a*(BCC), ΔE_f(FCC), and ΔE_f(BCC) of 9,139 HEAs (39 elements) are predicted via machine learning. The relative stability between FCC and BCC for a given HEA is obtained from the formation energy evaluated with respect to their constituent elements. Besides, 36,556 predicted data of 9139 HEAs with a total of 137,085 features are published on our public online database, as shown in Fig. 2. Finally, the screening rules for FCC and BCC-phase PtPd-based HEA are discussed and proposed.

Theoretical methods
DFT calculation. The ΔE_f of PtPd-based HEAs can be calculated as Eq. (1).

$$\Delta E_f = E_{total}^{HEA}(PtPdXYZ) - \sum_i (c_i E_{total}^{element}),$$
(1)

where $E_{total}^{HEA}(PtPdXYZ)$ is the total energy for the HEA system, and $E_{total}^{element}$ is the total energy for each element calculated from the natural form, e.g., FCC Pt, BCC Mn, HCP Co, etc., The c_i is the element concentration.

The total energies and lattice parameters of FCC and BCC PtPd*XYZ* HEAs were calculated using the Green's function method implemented in Akai-KKR (Machikaneyama) package[51,52]. The KKR-CPA was employed for calculating the electronic structure of the random alloy system comprising *n* components. The crystal potential was approximated by using the muffin-tin potential with the atomic-sphere approximation. The Perdew–Wang-91 generalized gradient approximation (GGA91)[66] was used for an exchange–correlation functional. The self-consistent calculation was performed using the criteria of $8 \times 8 \times 8$k-point mesh in the first Brillouin-zone. The electron density was calculated from the imaginary part of the Green's function evaluated on the complex energy contour whose width is 1.0 Ry from the Fermi energy. The iteration was performed until the difference between input and output potential becomes 10^{-6}. The maximum angular momentum for the expansion of Green's function was set to 2. The scalar relativistic approximation (SRA) was used for the relativistic treatment.

Machine learning. The 2,720 data of *a* and ΔH_f were collected from the DFT calculation of FCC and BCC PtPd*XYZ* HEAs ($X \neq Y \neq Z =$ Cr, Fe, Co, Ni, Cu, Zr, Mo, Ru, Rh, Ag, W, Os, Ir, Au, In, Sn, Bi) including 680 data of each FCC and BCC HEA for making each ML prediction model viz., *a*(FCC), *a*(BCC), ΔH_f(FCC), and ΔH_f(BCC). Fifteen features from the physical and chemical properties of each element, including valence electron count (VEC), atomic mass (*M*), atomic radius (r_{atomic}), atomic size difference (δ), Pauling electronegativity (χ), electronegativity difference ($\Delta\chi$), electron affinity (*EA*), density (ρ), molar volume (V_{mol}), melting point (T_{mel}),
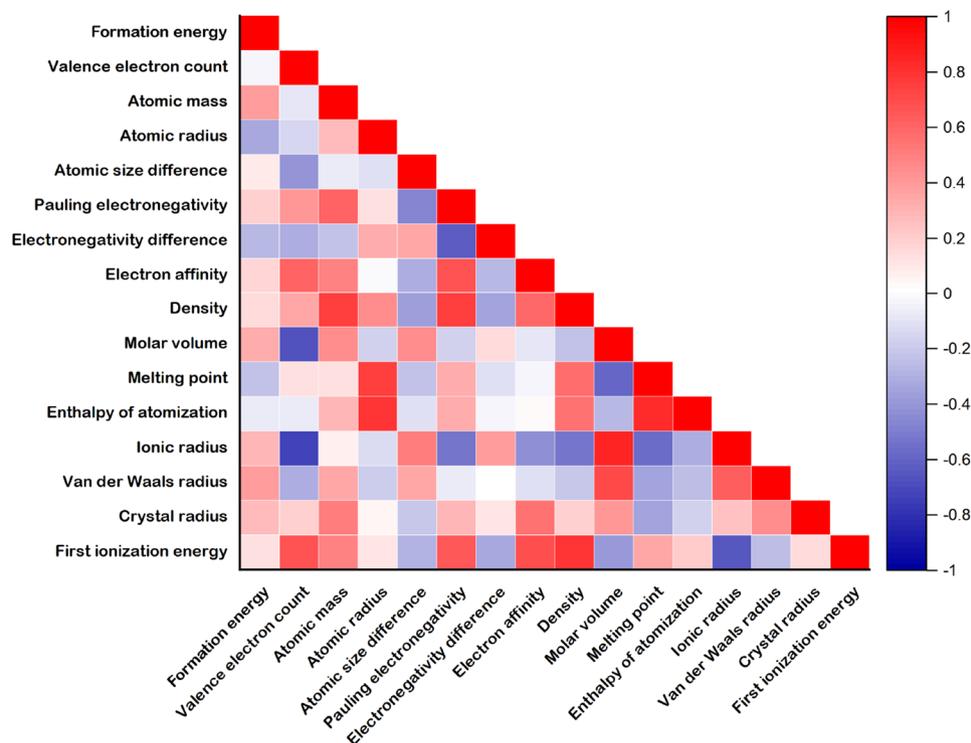
Figure 3. Matrix representation of Pearson correlation coefficient of features. The formation energy corresponds to ΔE_f(FCC) of HEAs.

enthalpy of atomization $\Delta H_{f,\text{atom}}$, ionic radius (r_{ionic}), Van der Waals radius (r_{VdW}), crystal radius (r_{crystal}), and first ionization energy (IE_1), were considered to be the descriptor of HEAs for making ML. The δ and $\Delta\chi$ were made using Eqs. (2) and (3). Other features can be built as Eq. (4).

$$\delta = 100\sqrt{\sum_{i=1}^{n} c_i \left(1 - \frac{r_i}{\bar{r}}\right)^2}, \tag{2}$$

$$\Delta\chi = \sqrt{\sum_{i=1}^{n} c_i (\chi_i - \overline{\chi})^2}, \tag{3}$$

$$\text{Feature} = \sum_{i=1}^{n} c_i (\text{Feature})_i, \tag{4}$$

where r_i, χ_i, and c_i are the r_{atomic}, χ, and concentration of element i. Weight-averaged features, $\bar{r}$ and $\overline{\chi}$ can be calculated using Eq. (4). The Pearson correlation coefficient of the features for ΔE_f(FCC) is illustrated in Fig. 3. The results for ΔH_f(BCC), a(FCC), and a(BCC) are shown in Figs. S1–S3. ΔE_f(FCC) and ΔE_f(BCC) demonstrated a weak relationship with the features. The a(FCC) and a(BCC) showed strong relation with VEC, V_{mol}, r_{ionic}, and r_{VdW}. However, the overview of correlation between features and ΔE_f (and a) displayed both weak and strong values. Hence, the artificial neural network (ANN) was chosen for the supervised ML method because it can accomplish the complex and nonlinear features of the ML prediction model[12,67]. The 680 data were randomly split into 70% for the training set (476 data) and 30% for the test set (204 data points). The ANN consisted of four hidden layers, where each hidden layer contained thirteen. The Levenberg–Marquardt backpropagation[68] was employed to train all models. The mean relative error (MRE) loop was used to improve the performance of prediction models, which can be calculated by Eq. (5).

$$\text{MRE} = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{x_i - x_{i,\text{predicted}}}{x_i} \right| \times 100\%, \tag{5}$$

where the MRE cutoff is set to 5%. The prediction accuracy was also evaluated by the mean square error (MSE), mean absolute error (MAE), and coefficient of determination (R^2) as illustrated in Eqs. (6)–(8).
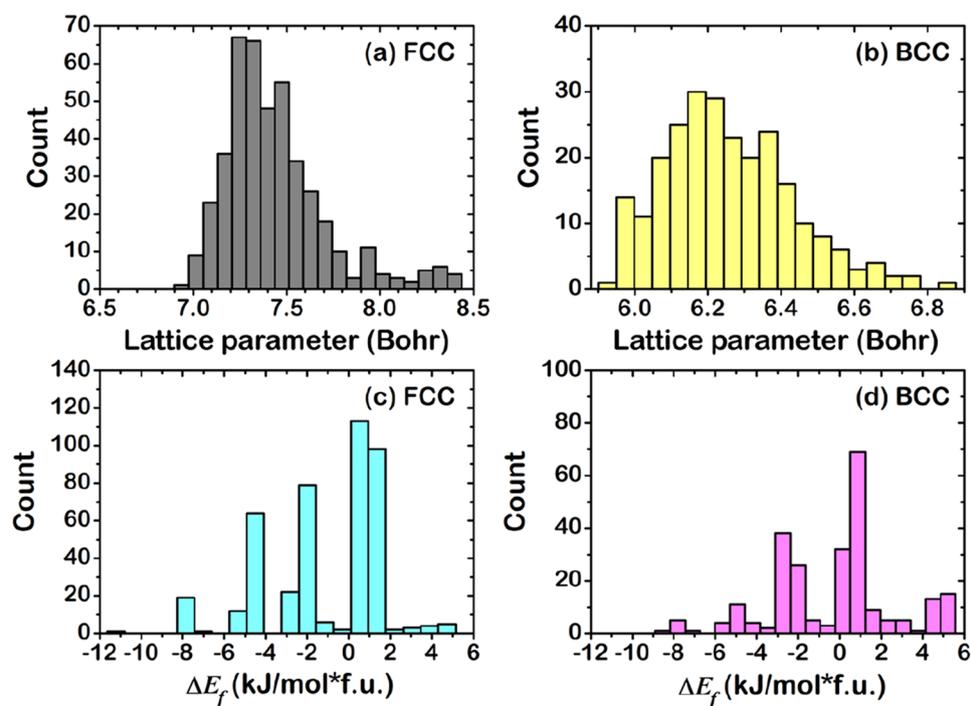
Figure 4. Distribution of the formation energy and lattice parameter given by DFT: (**a**) a(FCC), (**b**) a(BCC), (**c**) ΔE_f(FCC), and (**d**) ΔE_f(BCC).

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (x_i - x_{i,\text{predicted}})^2, \tag{6}$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} \left| x_i - x_{i,\text{predicted}} \right|, \tag{7}$$

$$\text{R}^2 = \frac{\sum (x_{i,\text{predicted}} - \overline{x})^2}{\sum \left(x_{i,\text{predicted}} - \overline{x} \right)^2 + \sum \left(x_i - x_{i,\text{predicted}} \right)^2}. \tag{8}$$

Results and discussion

DFT results. Figure S4 shows distribution data of DFT-calculated results of a(FCC), a(BCC), ΔE_f(FCC), and ΔE_f(BCC), while Fig. 4 chooses the more plausible phase. Each calculated HEA was classified into FCC or BCC phase by considering the ΔE_f to investigate the stable structure in thermodynamics[54,69]. For example, the DFT-calculated ΔE_f of PtPdCrFeCo HEA that is −2.6953 kJ/mol f.u. for FCC and −2.4593 kJ/mol f.u. for BCC indicates that FCC is thermodynamically favored over BCC due to the lowest ΔE_f. Figure 4 shows the data distribution, retaining only the thermodynamically stable phase for each HEA. As a result, 680 HEAs include 431 FCC (~ 63%) and 249 BCC phase (~ 37%). Based on Hess's law[70], the positive and negative values of ΔE_f can be denoted to endothermic (Endo) and exothermic (Exo) reaction, respectively, composing 220 Endo (and 211 Exo) for FCC and 129 Endo (and 120 Exo) for BCC phases. This DFT-calculated data demonstrates the ratio of Endo:Exo close to 50%:50%, which is suitably balanced for creating ML prediction models. Both Endo and Exo HEAs were analyzed in terms of components. Figure 5 reveals the distribution of each element counted from Endo and Exo HEAs. Also, color-mapping was employed to separate the data group. The high concentration of each data is ordered as follows: blue, light blue, cyan, green, yellow, orange, and red. It was found that the Exo FCC HEAs have Co and Ru as the most frequent element (Fig. 5a). Similarly, Zr is a good candidate in Exo BCC HEAs (Fig. 5b). For Endo HEAs, Cr, Fe, Ni, Cu, Mo, Rh, Ag, W, Os, Ir, and Au hinder forming in the FCC phase, while Sn and Bi are the remarkable elements in the BCC phase.

It is noted that the KKR-CPA uses Green's function method to calculate the electronic structure of alloys. In the KKR-CPA, the configuration average of the electronic structure is calculated[50–52], hence, there is no detail of atomic configuration in the KKR-CPA method. Thus optimized lattice constant should be regarded as average lattice constant. Figure 4a, b reveal the histogram of a(FCC) and a(BCC). The lattice parameter of most FCC HEAs is distributed in 7.3–7.5 Bohr, while that of BCC HEAs shows a large distribution from 6.0 to 6.5 Bohr.
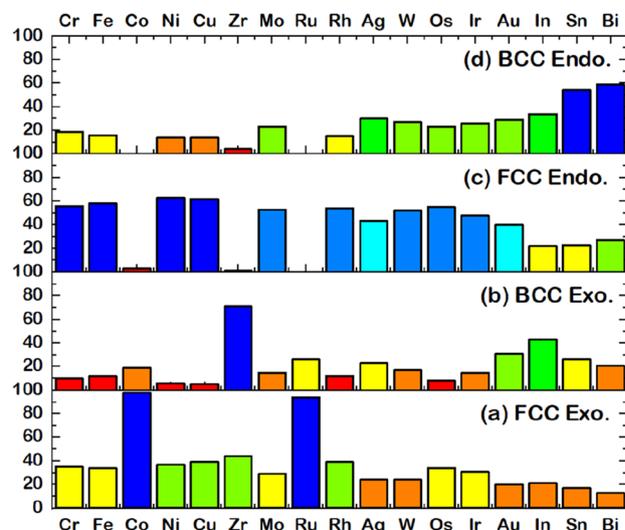
Figure 5. Distribution of elements counted from ΔE_f of PtPdXYZ HEAs ($X \neq Y \neq Z =$ Cr, Fe, Co, Ni, Cu, Zr, Mo, Ru, Rh, Ag, W, Os, Ir, Au, In, Sn, Bi) through DFT-calculated results, including (**a**) FCC exothermic cases, (**b**) BCC exothermic cases, (**c**) FCC endothermic cases, and (**d**) BCC endothermic cases. The color classification in each group is based on the frequency of the system found to have either FCC Exo, BCC Exo, FCC Endo, or BCC Endo.
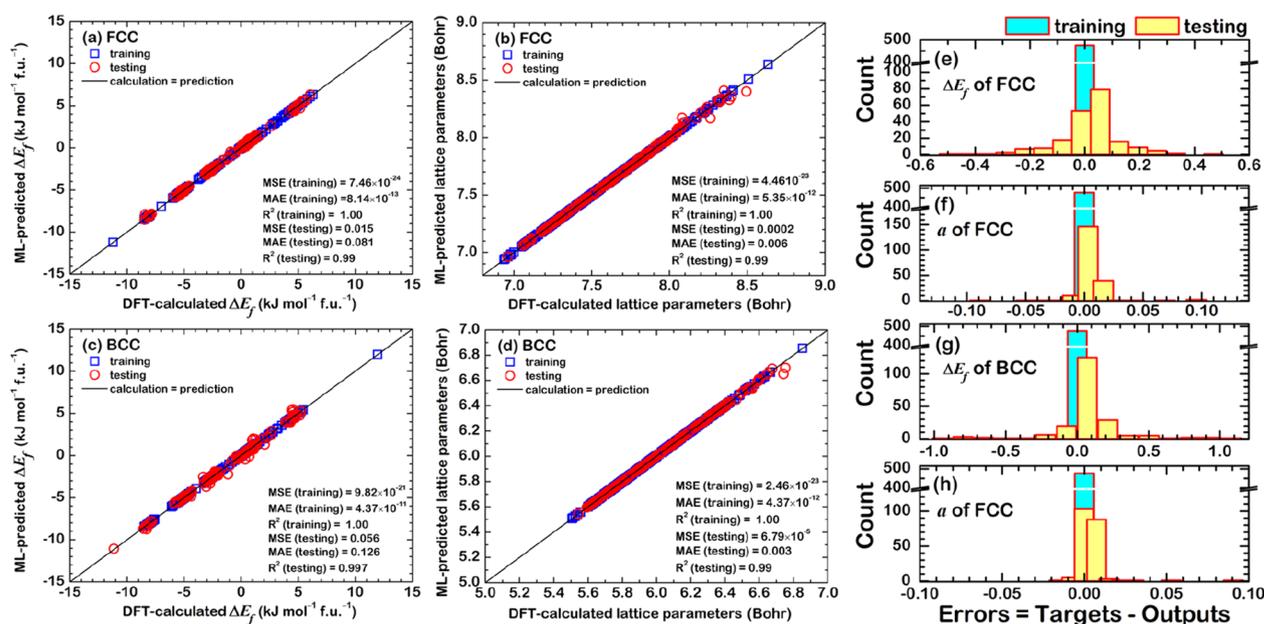
Figure 6. ML-predicted results compared with DFT-calculated data for (**a**) ΔE_f(FCC), (**b**) ΔE_f(BCC), (**c**) a (FCC), and (**d**) a(BCC). Error distribution in the testing is shown in (**e**)–(**h**).

ML-prediction models. The regression results are illustrated in Fig. 6, where DFT-calculated results were regressed through ANN. The MRE is 0.0237%, 0.0129%, 3.5554%, and 4.7429% for a(FCC), a(BCC), ΔE_f(FCC), and ΔE_f(BCC) prediction model, respectively. In the training results, the MSE and MAE values were close to zero and the R² equal to 1 in all the cases, indicating that the ANN with the MRE-loop can improve the accuracy of the training. This also helped increase the accuracy of testing, which can be confirmed by 0.99 of R². Figure 6e, g revealed the error values in the testing were ± 0.5 kJ/mol*f.u. for ΔE_f(FCC) and ± 1.0 kJ/mol*f.u. for ΔE_f(BCC). The error values of a (Fig. 6f, h) were close to zero and seemed more accurate than the ΔE_f because of the strong correlation with features (see Figs. S2 and S3).

The built ML models are employed to predict the 9,139 HEAs as the formulae of Pt$_{0.2}$Pd$_{0.2}$X$_{0.2}$Y$_{0.2}$Z$_{0.2}$ ($X \neq Y \neq Z$), where X, Y, and Z are considered from 39 elements: Sc, Ti, V, Cr, Mn, Fe, Co, Ni, Cu, Zn, Y, Zr, Nb, Mo, Tc, Ru, Rh, Ag, Cd, Hf, Ta, W, Re, Os, Ir, Au, Hg, Al, Si, P, Ga, Ge, As, In, Sn, Sb, Tl, Pb, and Bi. The ML-predicted results of a(FCC), a(BCC), ΔE_f(FCC), and ΔE_f(BCC) without distinguishing the phase are shown in Fig. S5. The
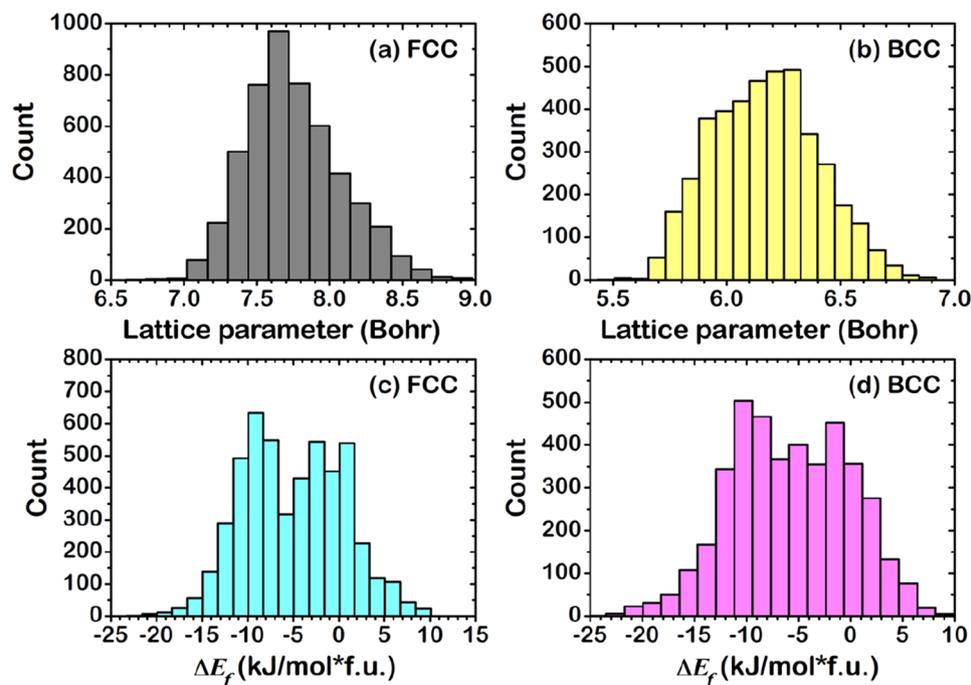
Figure 7. Distribution of ML-predicted results for (**a**) a(FCC), (**b**) a(BCC), (**c**) ΔE_f(FCC), and (**d**) ΔE_f(BCC).

thermodynamically stable phase of 9,139 HEAs is illustrated in Fig. 7. Figure 7a,b reveal enlarged distribution: 7.4–8.0 Bohr for a(FCC) and 5.9–6.5 Bohr for a(BCC) due to the variety of atomic radius of elements. Similarly, ΔE_f is distributed between about −25.0 kJ/mol f.u. and + 10.0 kJ/mol f.u. for both FCC and BCC. The number of FCC and BCC HEAs is 5002 and 4137, respectively. The number of Exo HEAs is 4140 for FCC and 3567 for BCC phase, more than the Endo HEAs, 862 for FCC and 570 for BCC phase.

To understand the role of each element involving Endo and Exo forms, Fig. 8 illustrates the element distribution obtained from 9139 PtPdXYZ HEAs ($X \neq Y \neq Z$ = Sc, Ti, V, Cr, Mn, Fe, Co, Ni, Cu, Zn, Y, Zr, Nb, Mo, Tc, Ru, Rh, Ag, Cd, Hf, Ta, W, Re, Os, Ir, Au, Hg, Al, Si, P, Ga, Ge, As, In, Sn, Sb, Tl, Pb, Bi). Overall, all elements of choice except Tl and Pb are found to form Endo HEAs as the ratio FCC:BCC of 50%:50%. Co, Zr, Zn, Al, Ga, Si, Ge, P, and As are hardly involved in Endo HEAs, contrary to Tl, which favors forming the FCC Endo HEAs (Fig. 8c). For the Exo HEAs, the green bar in the FCC phase and the cyan bar in the BCC phase indicate high distribution. In Exo FCC HEAs (Fig. 8a), Co, Zn, Y, Hg, P, and As can be totaled more than 400 HEAs, followed by Sc, Ti, V, Cu, Zr, Nb, Ru, Ag, Cd, Ta, Os, Ir, Si, Ge, and Sb counted by more than 300 HEAs (yellow bar). For the Exo BCC HEAs shown in Fig. 8b, the elements found in more than 300 HEAs are Ti, V, Mn, Y, Zr, Nb, Tc, Hf, Ta, Si, Ga, Ge, In, Sn, and Bi. Among these, Ga, the only element that naturally forms as an orthorhombic structure, plays a significant role in the BCC phase found in more than 500 Exo HEAs (the cyan bar in Fig. 8b). The phase of HEAs is summarized in Fig. 9. Among these data, 15 and 6 elements favor forming FCC and BCC, respectively, while the others almost count both FCC and BCC phases.

The literature on PtPd-based HEAs is hardly found because they are new group materials. Based on the available data, the phase classification using ΔE_f is demonstrated in Table 1. PtPd-based HEAs of five principal elements reported in the literature showed the predominance of the FCC phase. The ΔE_f values given by ANN models predicted the thermodynamic propensity of the FCC formation of these PtPd-based HEAs, thus agreeing with the available experimental data. In addition to FCC HEAs, BCC HEAs experimentally reported in literature such as CrFeCoNiAl[56,58], MnFeNiSiGa[56], NbMoTaW[54,62,63], VNbMoTaW[54,62,63], TiVCrFeCoNiCuAl[64], TiZrNbHfTa[61,65], and TiZrNbMo[58] were employed to validate the KKR-CPA method (Tables S1 and S2). In this case, calculated ΔE_f values successfully categorized the correct phase.

The prediction on the HEAs containing untrained elements such as Sc, Ti, V, Mn, Zn, Y, Nb, Tc, Cd, Hf, Ta, Re, Hg, Al, Si, P, Ga, Ge, As, Sb, Te, and Pb was tested by the calculated ΔE_f and a of PtPdCrXY, PtPdNiSnX, and PtPdXYZ. This aims to examine whether these prediction models are accurate. Because the DFT calculation of 9,139 HEAs includes 36,556 data for ΔE_f(FCC), ΔE_f(BCC), a(FCC), and a(BCC), the selected HEAs composing untrain elements are used to reduce computational time. The formulae PtPdNiSnX, PtPdCrXY, and PtPdXYZ are represented the HEAs containing one, two, and three untrained elements. The elements X and Y in PtPdNiSnX and PtPdCrXY are Sc, Ti, V, Mn, Zn, Y, Nb, Tc, Cd, Hf, Ta, Re, Hg, Al, Si, P, Ga, Ge, As, Sb, Te, Pb. The elements X, Y, and Z in PtPdXYZ are Sc, Ti, V, Mn, Zn, Al, Si, Sb. The 309 HEAs for validation are implemented in Table S3. The regression plot between DFT-calculation and ML-prediction is illustrated in Fig. 10a,b. The MAE, MSE, and R² are listed in Table 2. Although the MSE and MAE of the untrained data (Table 2) are higher than that of the trained data (Fig. 6a,c), the R² of all ΔE_f from the untrained data is still as high as 0.99. The predicted ΔE_f of the selected HEAs has an error ± 4 and ± 3 kJ mol⁻¹ f.u.⁻¹ for FCC and BCC prediction model, respectively (Fig. 10c,d). Figure 10b shows the regression plot of a that seems less accuracy. Figure 10e,f indicate the error
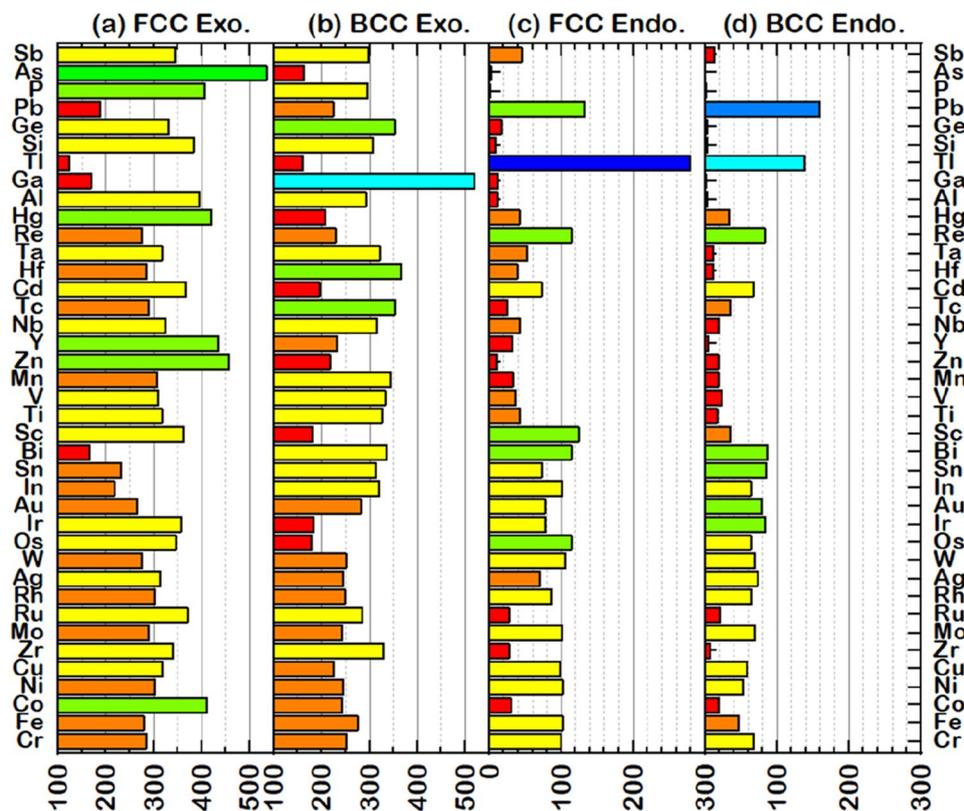
Figure 8. Distribution of elements counted from ΔE_f of PtPdXYZ HEAs ($X \neq Y \neq Z =$ Sc, Ti, V, Cr, Mn, Fe, Co, Ni, Cu, Zn, Y, Zr, Nb, Mo, Tc, Ru, Rh, Ag, Cd, Hf, Ta, W, Re, Os, Ir, Au, Hg, Al, Si, P, Ga, Ge, As, In, Sn, Sb, Tl, Pb, Bi) through ML-predicted results: (**a**) FCC exothermic cases, (**b**) BCC exothermic cases, (**c**) FCC endothermic cases, and (**d**) BCC endothermic cases. The color classification in each group is based on the frequency of the system found to have either FCC Exo, BCC Exo, FCC Endo, or BCC Endo.

Figure 9. Predicted role of individual elements in the HEA formation energy and phase.

value of predicted a less than ± 0.4 Å. This error in a is in an acceptable range for the bulk structure, although the R^2 of untrained a is less than that of the trained a (Fig. 6b,d). In addition to the R^2, the MSE and MAE values as well as the error value in the prediction should be considered when evaluating the performance of prediction. As a result, the MSE and MAE of untrained a are in the same range of that the trained a (Fig. 6b,d). Although there are some errors found on the prediction of selected HEAs, the performance of prediction models is accepted.

Conclusion

A combination of DFT and ANN was employed to predict the possible formulae of penta-metallic high-entropy alloys. The formation energy and lattice parameter of each system were determined via the KKR-CPA method. The ANN was employed to construct the prediction models and speed up the material screening. The training-to-testing data are 70%:30%, derived from DFT-calculated data of 680 HEAs from 17 elements. 15 features were used in such an algorithm are VEC, M, r_{atomic}, δ, χ, $\Delta\chi$, E_{ea}, ρ, V_{mol}, T_{mel}, $\Delta H_{f,\text{atom}}$, r_{ionic}, r_{VdW}, r_{crystal}, and IE$_1$. The built models possessed high accuracy in the testing, accompanied with the R^2 values close to unity and MRE within 5%. Based on the prediction models, 9,139 PtPd-based HEA systems created from a pool of 39 elements were classified into 5,002 FCC and 4,137 BCC systems; the HEA screening rules can be summarized as follows.

HEAs	ΔE_f (FCC)	a (FCC)	ΔE_f (BCC)	a (BCC)	Lowest ΔE_f	Predicted phase	References
PtPdFeCoNi	−2.6973	3.6713	−2.4330	2.9140	−2.6973	FCC	FCC[26,36] $a = 3.73$[36]
PtPdFeCoIr	−2.6859	3.7440	−2.376	2.9715	−2.6859	FCC	FCC[37]
PtPdFeRhIr	0.2036	3.8130	0.4338	3.0264	0.2036	FCC	FCC[38] $a = 3.834$[38]
PtPdCoNiCu	−2.6009	3.6831	−2.3553	2.9233	−2.6009	FCC	FCC[39]
PtPdCuAgAu	0.0882	3.9351	0.1317	3.1233	0.0882	FCC	FCC[7,8,33] $a = 3.936$[7]
PtPdRuRhIr	−5.2674	3.8669	−4.9991	3.0692	−5.2674	FCC	FCC[34,35,42,43] $a = 3.8560$[35] $a = 3.82$[36] $a = 3.84 - 3.96$[42]
PtPdRuAgIr	−5.0594	3.9276	−4.9444	3.1173	−5.0594	FCC	FCC[34,41]
PtPdRuRhAu	−5.2834	3.9165	−5.1907	3.1085	−5.2834	FCC	FCC[40]

Table 1. Phase classification of PtPd-based HEAs by ΔE_f (in kJ mol^{-1} f.u.$^{-1}$) with a (in Å) validated with literature data.
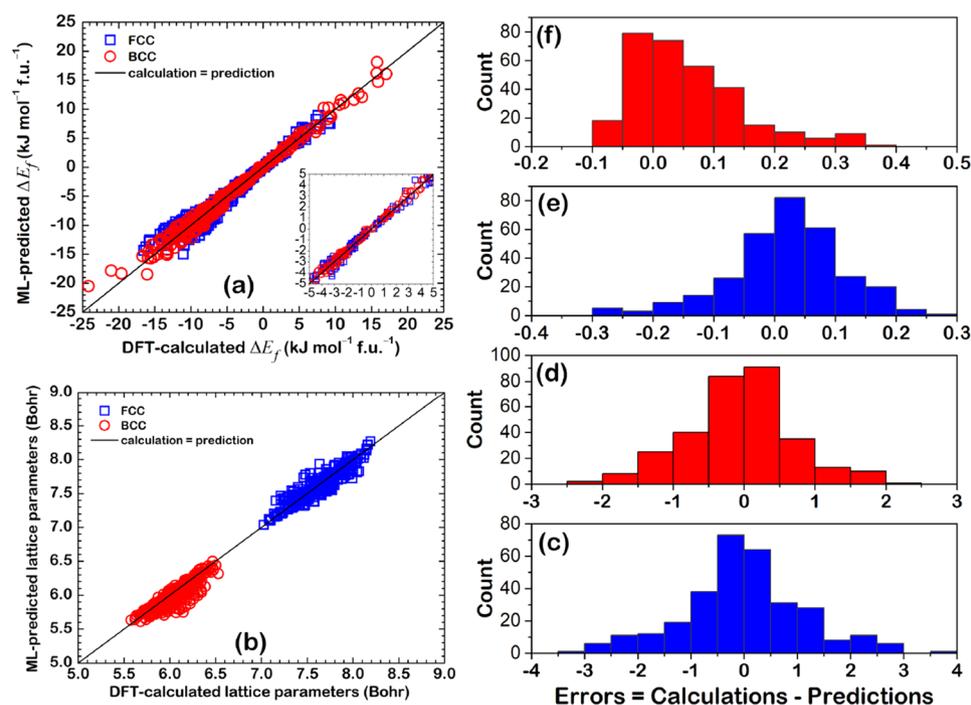
Figure 10. Validation of PtPdCr*XY*, PtPdNiSn*X*, and PtPd*XYZ* HEAs including regression plots of (**a**) formation energy, (**b**) lattice parameters, (**c**) error of formation energy in FCC data, (**d**) error of formation energy in BCC data, (**e**) error of lattice parameters in FCC data, (**f**) error of lattice parameters in BCC data.

	ΔE_f (FCC)	a (FCC)	ΔE_f (BCC)	a (BCC)
MSE	1.31	0.009(2)	0.58	0.011(5)
MAE	1.31	0.005(4)	0.58	0.000(2)
R^2	0.98(1)	0.86(7)	0.98(7)	0.75(9)

Table 2. MSE, MAE, and R^2 of formation energy (ΔE_f) and lattice parameters (a) for the predicted PtPdCr*XY*, PtPdNiSn*X*, and PtPd*XYZ*.

i. HEAs with the component of Sc, Co, Cu, Zn, Y, Ru, Cd, Os, Ir, Hg, Al, Si, P, As, and Tl have a high tendency to form in the FCC phase.

ii. HEAs with the component Hf, Ga, In, Sn, Pb, and Bi have a high tendency to form in the BCC phase.

iii. Ti, V, Cr, Mn, Fe, Ni, Zr, Nb, Mo, Tc, Rh, Ag, Ta, W, Re, Au, Ge, and Sb have a comparable tendency to form in either FCC or BCC phase.

These screening rules applied in this work provide the fundamental for the discovery of bulk HEA, where the development of algorithms for the screening of HEA in terms of stable surface configuration is the outlook for future work.

Data availability
The authors declare that relevant data are within the manuscript.

References
1. Sheng, G. & Liu, C. T. Phase stability in high entropy alloys: Formation of solid-solution phase or amorphous phase. *Prog. Nat. Sci. Mater. Int.* **21**(6), 433–446 (2011).
2. Xin, Y. *et al.* High-entropy alloys as a platform for catalysis: Progress. *Challenges Opportun. Acs Catal.* **10**(19), 11280–11306 (2020).
3. Roy, D., Mandal, S. C. & Pathak, B. Machine learning-driven high-throughput screening of alloy-based catalysts for selective CO_2 hydrogenation to methanol. *ACS Appl. Mater. Interfaces.* **13**(47), 56151–56163 (2021).
4. Li, X. *et al.* A transferable machine-learning scheme from pure metals to alloys for predicting adsorption energies. *J. Mater. Chem. A* **10**, 872 (2022).
5. Wang, S. *et al.* High entropy alloy/C nanoparticles derived from polymetallic MOF as promising electrocatalysts for alkaline oxygen evolution reaction. *Chem. Eng. J.* **429**, 132410 (2022).
6. Feng, D., Dong, Y., Nie, P., Zhang, L. & Qiao, Z.-A. CoNiCuMgZn high entropy alloy nanoparticles embedded onto graphene sheets via anchoring and alloying strategy as efficient electrocatalysts for hydrogen evolution reaction. *Chem. Eng. J.* **430**, 132883 (2022).
7. Nellaiappan, S. *et al.* High-entropy alloys as catalysts for the CO2 and CO reduction reactions: Experimental realization. *ACS Catal.* **10**(6), 3658–3663 (2020).
8. Pedersen, J. K., Batchelor, T. A., Bagger, A. & Rossmeisl, J. High-entropy alloys as catalysts for the CO2 and CO reduction reactions. *ACS Catal.* **10**(3), 2169–2176 (2020).
9. Li, J. *et al.* Chemical-element-distribution-mediated deformation partitioning and its control mechanical behavior in high-entropy alloys. *J. Mater. Sci. Technol.* **120**, 99–107 (2022).
10. Bundela, A. S. & Rahul, M. Machine learning-enabled framework for the prediction of mechanical properties in new high entropy alloys. *J. Alloy. Compd.* **908**, 164578 (2022).
11. Li, S., Li, S., Liu, D., Zou, R. & Yang, Z. Hardness prediction of high entropy alloys with machine learning and material descriptors selection by improved genetic algorithm. *Comput. Mater. Sci.* **205**, 111185 (2022).
12. Peng, J. *et al.* Vacancy dependent mechanical behaviors of high-entropy alloy. *Int. J. Mech. Sci.* **218**, 107065 (2022).
13. Yeh, J.-W. Alloy design strategies and future trends in high-entropy alloys. *Jom* **65**(12), 1759–1771 (2013).
14. Jiang, H. *et al.* A new strategy to design eutectic high-entropy alloys using simple mixture method. *Mater. Des.* **142**, 101–105 (2018).
15. Gao, M. C. *et al.* High-entropy functional materials. *J. Mater. Res.* **33**(19), 3138–3155 (2018).
16. Ryan, K., Lengyel, J. & Shatruk, M. Crystal structure prediction via deep learning. *J. Am. Chem. Soc.* **140**(32), 10158–10168 (2018).
17. Oliynyk, A. O. *et al.* High-throughput machine-learning-driven synthesis of full-Heusler compounds. *Chem. Mater.* **28**(20), 7324–7331 (2016).
18. Takahashi, K. *et al.* Catalysis gene expression profiling: Sequencing and designing catalysts. *J. Phys. Chem. Lett.* **12**(30), 7335–7341 (2021).
19. Sorkin, V. *et al.* First-principles-based high-throughput computation for high entropy alloys with short range order. *J. Alloy. Compd.* **882**, 160776 (2021).
20. Kaufmann, K. & Vecchio, K. S. Searching for high entropy alloys: A machine learning approach. *Acta Mater.* **198**, 178–222 (2020).
21. Jin, T., Park, I., Park, T., Park, J. & Shim, J. H. Accelerated crystal structure prediction of multi-elements random alloy using expandable features. *Sci. Rep.* **11**(1), 1–9 (2021).
22. Yan, Y., Lu, D. & Wang, K. Accelerated discovery of single-phase refractory high entropy alloys assisted by machine learning. *Comput. Mater. Sci.* **199**, 110723 (2021).
23. Klimenko, D., Stepanov, N., Li, J., Fang, Q. & Zherebtsov, S. Machine learning-based strength prediction for refractory high-entropy alloys of the Al-Cr-Nb-Ti-V-Zr system. *Materials* **14**(23), 7213 (2021).
24. Nassar, A. & Mullis, A. Rapid screening of high-entropy alloys using neural networks and constituent elements. *Comput. Mater. Sci.* **199**, 110755 (2021).
25. Hou, S. *et al.* A hybrid prediction frame for HEAs based on empirical knowledge and machine learning. *Acta Mater.* **228**, 117742 (2022).
26. Ragone, M. *et al.* Deep learning for mapping element distribution of high-entropy alloys in scanning transmission electron microscopy images. *Comput. Mater. Sci.* **201**, 110905 (2022).
27. Liu, F., Xiao, X., Huang, L., Tan, L. & Liu, Y. Design of NiCoCrAl eutectic high entropy alloys by combining machine learning with CALPHAD method. *Mater. Today Commun.* **30**, 103172 (2022).
28. Zhou, X.-Y. *et al.* Machine learning assisted design of FeCoNiCrMn high-entropy alloys with ultra-low hydrogen diffusion coefficients. *Acta Mater.* **224**, 117535 (2022).
29. Zhang, J., Cai, C., Kim, G., Wang, Y. & Chen, W. Composition design of high-entropy alloys with deep sets learning. *npj Comput. Mater.* **8**(1), 1–11 (2022).
30. Zhang, S. & Wang, G. Predicting mechanical properties of high entropy alloys with face centered cubic structure from first principles calculations. *Mater. Today Commun.* **32**, 104059 (2022).
31. Wen, C. *et al.* Machine learning assisted design of high entropy alloys with desired property. *Acta Mater.* **170**, 109–117 (2019).
32. Bhandari, U., Ghadimi, H., Zhang, C., Yang, S. & Guo, S. Predicting elastic constants of refractory complex concentrated alloys using machine learning approach. *Materials* **15**(14), 4997 (2022).
33. Katiyar, N. K. *et al.* Electrooxidation of hydrazine utilizing high-entropy alloys: Assisting the oxygen evolution reaction at the thermodynamic voltage. *ACS Catal.* **11**(22), 14000–14007 (2021).

34. Pedersen, J. K. et al. Bayesian optimization of high-entropy alloy compositions for electrocatalytic oxygen reduction. *Angew. Chem.* **60**(45), 24144–24152 (2021).
35. Wu, D. et al. On the electronic structure and hydrogen evolution reaction activity of platinum group metal-based high-entropy-alloy nanoparticles. *Chem. Sci.* **11**(47), 12731–12736 (2020).
36. Yao, Y. et al. High-throughput, combinatorial synthesis of multimetallic nanoclusters. *Proc. Natl. Acad. Sci.* **117**(12), 6316–6322 (2020).
37. Gao, S. et al. Synthesis of high-entropy alloy nanoparticles on supports by the fast moving bed pyrolysis. *Nat. Commun.* **11**(1), 1–11 (2020).
38. Baba, K., Ishizu, N., Nishizaki, T. & Kitagawa, J. Magnetic and transport properties of new dual-phase high-entropy alloy FeRhIr-PdPt. *Materials* **14**(11), 2877 (2021).
39. Chen, Y. et al. Synthesis of monodisperse high entropy alloy nanocatalysts from core@ shell nanoparticles. *Nanoscale Horizons* **6**(3), 231–237 (2021).
40. Okejiri, F. et al. Ultrasound-driven fabrication of high-entropy alloy nanocatalysts promoted by alcoholic ionic liquids. *Nano Res.* **15**, 1–7 (2021).
41. Batchelor, T. A. et al. Complex-solid-solution electrocatalyst discovery by computational prediction and high-throughput experimentation. *Angew. Chem. Int. Ed.* **60**(13), 6932–6937 (2021).
42. Broge, N. L., Bondesgaard, M., Søndergaard-Pedersen, F., Roelsgaard, M. & Iversen, B. B. Autocatalytic formation of high-entropy alloy nanoparticles. *Angew. Chem.* **132**(49), 22104–22108 (2020).
43. 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).
44. Takeuchi, A. & Inoue, A. Mixing enthalpy of liquid phase calculated by Miedema's scheme and approximated with sub-regular solution model for assessing forming ability of amorphous and glassy alloys. *Intermetallics* **18**(9), 1779–1789 (2010).
45. Troparevsky, M. C., Morris, J. R., Kent, P. R., Lupini, A. R. & Stocks, G. M. Criteria for predicting the formation of single-phase high-entropy alloys. *Phys. Rev. X* **5**(1), 011041 (2015).
46. Soven, P. Coherent-potential model of substitutional disordered alloys. *Phys. Rev.* **156**(3), 809 (1967).
47. Gyorffy, B. Coherent-potential approximation for a nonoverlapping-muffin-tin-potential model of random substitutional alloys. *Phys. Rev. B* **5**(6), 2382 (1972).
48. Stocks, G., Temmerman, W. & Gyorffy, B. Complete solution of the Korringa-Kohn-Rostoker coherent-potential-approximation equations: Cu-Ni alloys. *Phys. Rev. Lett.* **41**(5), 339 (1978).
49. Faulkner, J. & Stocks, G. Calculating properties with the coherent-potential approximation. *Phys. Rev. B* **21**(8), 3222 (1980).
50. Akai, H. & Dederichs, P. Local moment disorder in ferromagnetic alloys. *Phys. Rev. B* **47**(14), 8739 (1993).
51. Akai, H. Fast Korringa-Kohn-Rostoker coherent potential approximation and its application to FCC Ni-Fe systems. *J. Phys. Condens. Matter* **1**(43), 8045 (1989).
52. Akai, H. *AkaiKKR (Machikaneyama)*. http://kkr.issp.u-tokyo.ac.jp.
53. Tian, F. A review of solid-solution models of high-entropy alloys based on ab initio calculations. *Front. Mater.* **4**, 36 (2017).
54. Hu, Y. et al. First-principle calculation investigation of NbMoTaW based refractory high entropy alloys. *J. Alloy. Compd.* **827**, 153963 (2020).
55. Troparevsky, M. C. et al. Beyond atomic sizes and Hume-Rothery rules: Understanding and predicting high-entropy alloys. *Jom* **67**(10), 2350–2363 (2015).
56. Huang, E.-W. et al. Machine-learning and high-throughput studies for high-entropy materials. *Mater. Sci. Eng. R. Rep.* **147**, 100645 (2022).
57. Lucas, M. et al. Absence of long-range chemical ordering in equimolar FeCoCrNi. *Appl. Phys. Lett.* **100**(25), 251907 (2012).
58. Zhang, Y., Yang, X. & Liaw, P. Alloy design and properties optimization of high-entropy alloys. *Jom* **64**(7), 830–838 (2012).
59. Lucas, M. et al. Magnetic and vibrational properties of high-entropy alloys. *J. Appl. Phys.* **109**(7), 07E307 (2011).
60. Wu, Z., Bei, H., Otto, F., Pharr, G. M. & George, E. P. Recovery, recrystallization, grain growth and phase stability of a family of FCC-structured multi-component equiatomic solid solution alloys. *Intermetallics* **46**, 131–140 (2014).
61. Cantor, B., Chang, I., Knight, P. & Vincent, A. Microstructural development in equiatomic multicomponent alloys. *Mater. Sci. Eng. A* **375**, 213–218 (2004).
62. Senkov, O., Wilks, G., Miracle, D., Chuang, C. & Liaw, P. Refractory high-entropy alloys. *Intermetallics* **18**(9), 1758–1765 (2010).
63. Senkov, O. N., Wilks, G., Scott, J. & Miracle, D. B. Mechanical properties of Nb25Mo25Ta25W25 and V20Nb20Mo20Ta20W20 refractory high entropy alloys. *Intermetallics* **19**(5), 698–706 (2011).
64. Yeh, J.-W. et al. Formation of simple crystal structures in Cu-Co-Ni-Cr-Al-Fe-Ti-V alloys with multiprincipal metallic elements. *Metall. Mater. Trans. A* **35**(8), 2533–2536 (2004).
65. Senkov, O. et al. Microstructure and elevated temperature properties of a refractory TaNbHfZrTi alloy. *J. Mater. Sci.* **47**(9), 4062–4074 (2012).
66. Perdew, J. P. et al. Erratum: Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation. *Phys. Rev. B* **48**(7), 4978 (1993).
67. Huang, W., Martin, P. & Zhuang, H. L. Machine-learning phase prediction of high-entropy alloys. *Acta Mater.* **169**, 225–236 (2019).
68. Aldrich, C. *Exploratory Analysis of Metallurgical Process Data with Neural Networks and Related Methods* (Elsevier, 2002).
69. Rittiruam, M., Yangthaisong, A. & Seetawan, T. Enhancing the thermoelectric performance of self-defect TiNiSn: A first-principles calculation. *J. Electron. Mater.* **47**(12), 7456–7462 (2018).
70. Chakrabarty, D. K. *An Introduction to Physical Chemistry* (Alpha Science International Ltd, 2001).

Acknowledgments

Author contributions

M.R. designed the high-throughput screening of PtPd-based HEAs. M.R. calculated total energies and lattice parameters of FCC and BCC HEAs using the Akai-KKR package. J.N. and S.S. corrected and analyzed all enthalpy of formation energies and lattice parameters. J.N. and N.A. performed and analyzed machine learning data. J.N., S.S., N.A., A.S., S.B., and C.W. created all features of 9139 HEAs. M.R. performed validation of the DFT-calculated and ML-predicted results. All authors reviewed and discussed all results.

Competing interests

The authors declare no competing interests.

Additional information