Evidence-based recommender system for high-entropy alloys

Minh-Quyet Ha, Duong-Nguyen Nguyen, Viet-Cuong Nguyen, Takahiro Nagata, Toyohiro Chikyow, Hiori Kino, Takashi Miyake, Thierry Denœux, Van-Nam Huynh, Hieu-Chi Dam

Evidence-based recommender system for high-entropy alloys

Minh-Quyet Ha,[1] Duong-Nguyen Nguyen,[1] Viet-Cuong Nguyen,[2] Takahiro Nagata,[3] Toyohiro Chikyow,[4] Hiori Kino,[4] Takashi Miyake,[5] Thierry Denœux,[6] Van-Nam Huynh,[1] and Hieu-Chi Dam[1]

[1)] *Japan Advanced Institute of Science and Technology, 1-1 Asahidai, Nomi, Ishikawa 923-1292, Japan*

[2)] *HPC SYSTEMS Inc., 3-9-15 Kaigan, Minato, Tokyo 108-0022, Japan*

[3)] *Research Center for Functional Materials, National Institute for Materials Science, 1-2-1 Sengen, Tsukuba, Ibaraki 305-0044, Japan*

[4)] *Research and Services Division of Materials Data and Integrated System, National Institute for Materials Science, 1-2-1 Sengen, Tsukuba, Ibaraki 305-0044, Japan*

[5)] *Research Center for Computational Design of Advanced Functional Materials, National Institute of Advanced Industrial Science and Technology, 1-1-1 Umezono, Tsukuba, Ibaraki 305-8568, Japan*

[6)] *Université de technologie de Compiègne, CNRS, UMR 7253 Heudiasyc, Compiègne, France*

(*Electronic mail: dam@jaist.ac.jp)

(Dated: 5 June 2021)

Existing data-driven approaches for exploring high-entropy alloys (HEAs) face three challenges: numerous element-combination candidates, designing appropriate descriptors, and limited and biased existing data. To overcome these issues, here we show the development of an evidence-based material recommender system (ERS) that adopts the Dempster–Shafer theory, a general framework for reasoning with uncertainty. Herein, without using material descriptors, we model, collect and combine pieces of evidence from data about the HEA phase existence of alloys. To evaluate the ERS, we compared its HEA-recommendation capability with those of matrix-factorization- and supervised-learning-based recommender systems on four widely known data sets of up-to-five-component alloys. The k-fold cross-validation on the data sets suggested the ERS outperforms all the competitors. Furthermore, the ERS shows good extrapolation capabilities in recommending quaternary and quinary HEAs. We experimentally validated the most strongly recommended Fe-Co-based magnetic HEA, namely FeCoMnNi, and confirmed that its thin film shows a body-centered cubic structure.

I. INTRODUCTION

Multi-principle element alloys (MPEAs, among which alloys with ≥ 5 elements are also called high-entropy alloys, HEAs) are a new alloy development concept[1–3], whereby the alloys comprise multiple elements and form highly disordered solid-solution phases. Since their discovery, MPEAs and HEAs have attracted the interest of the scientific community owing to their promising properties and potential applications[4,5]. Such alloys show high strength-to-weight ratios, tensile strengths, and corrosion and oxidation resistances. For consistency with the published data used in this study, we use the term HEA to refer to random alloys comprising multiple equiatomically combined elements and forming solid-solution phase. From the materials development perspective, specific element combinations that will most likely form single-phase HEAs must necessarily be recommended for experimental validation. Deductive and inductive approaches are both used to accomplish this task, and are based on entirely different concepts.

In the deductive approach, it is necessary to understand the HEA formation mechanisms or begin with the quantum-mechanics equations derived based on numerous first-principles calculations. In previous HEA research, it was hypothesized that HEA constituent elements form a single-phase solid solution owing to configurational-entropy-induced stabilization. However, this hypothesis is correct only for some multicomponent alloys, most of which have been experimentally demonstrated to form multiple phases[6]. Although much attention has been devoted to the formation mechanism driving HEA stability, the key factors governing the formation of single-phase HEAs remain unknown[7]. Constructing phase diagrams for multicomponent alloys by first-principles calculations can also directly predict which alloys will form solid solutions[8]. However, this method involves energy calculations for many configurations and the implementation of statistical mechanical models for estimating thermodynamic properties, both of which are computationally demanding[9]. Therefore, it is imperative to search for HEAs by first-principles calculations.

Several inductive screening methods have been developed using descriptors created from knowledge of condensed matter theory, with parameters fitted to the available experimental data to predict the possible HEA[10,11] or their structure phases[12–14]. However, applying the inductive approach requires sufficient and balanced data to ensure prediction accuracy, usually not available with experimental material data that is either lacking or heavily biased

toward positive results[15,16]. In addition, although it would be desirable to quantitatively evaluate the prediction uncertainty even if a high prediction accuracy cannot be obtained, this has not yet been achieved. Another challenge is to design suitable material descriptors to represent the data of alloys comprising different numbers of elements. Descriptors calculated from the atomic properties of the constituent elements (e.g., mean, variance, and difference of atomic sizes) are often adopted[14,17–21]. However, it is mathematically difficult to accurately assess the similarity or dissimilarity between alloys with different numbers of compositions; and there are inevitable limits to the results obtained by data-driven approaches using these descriptors[17,22]. A solution for this problem is to describe the alloy using one-hot vectors of constituent elements; however, this approach raises another difficulty, which is designing a proper metric in this vector space[23].

To overcome these issues and focus on predicting whether the HEA phase exists for particular combinations of elements, we adopted the Dempster–Shafer theory[24–26], referred to as the evidence theory, to develop a descriptor-free recommender system, called evidence-based recommender system (ERS), for exploring potential HEAs.

The Dempster–Shafer theory can be considered as a generalization of the Bayesian approach for dealing with situations of incomplete information and imperfect data, and is deemed suitable for solving material data problems. Given a set Ω of possibilities (called the *frame of discernment*), evidence theory assigns non-negative weights (summing to one) to *subsets* of Ω, instead of assigning them to elements of Ω as in the Bayesian approach. By adopting the evidence theory, we can model, collect and combine pieces of evidence from multiple alloy data without using material descriptors. Consequently, the proposed system can suggest HEAs by learning from multiple data of alloys with fewer constituent elements.

The proposed recommender system is based on the elemental substitution method widely used to synthesize various materials. This method is used to replace the element or group of elements with a counterpart showing similar chemical functions, such that the properties of the target material are not affected. The difficulty in this approach is the proper assessment of the similarity between the chemical functions of the alloy metal combinations to discover potential HEAs. To address this issue, we consider each pair of observed alloys as a piece of evidence to compare the contribution of their constituent elements or a combination thereof to the target property (forming HEA phase). The obtained similarity evidence is then used to generate evidence for hypothesizing whether the substituted alloys are HEAs. The ERS

consists of three main steps (Supplementary Figure 1):

1. **Measure the similarity between element combinations:** All the pieces of evidence obtained from the data are modeled and combined to conclude the similarity between the element combinations by using evidence theory.

2. **Evaluate the hypothesis on the properties of the substituted alloys:** The pieces of evidence for the substituted alloys are modeled and combined to evaluate the hypothesis about the target property (forming HEA phase) by using evidence theory.

3. **Rank substituted alloys:** The substituted alloys are ranked according to various criteria based on the combined evidence of their target properties to recommend potential HEAs.

II. RESULTS

A. Evidence-based recommender methodology

Each alloy A in data set $\mathcal{D}$ is represented by a set of its components. The property of interest y_A for the alloy A, which can be either HEA or $\neg HEA$ (not HEA), indicates whether the HEA phase exists for the alloy A. We first measure the similarity between element combinations by adapting the evidence theory to model and combine all pieces of evidence obtained form the data set $\mathcal{D}$.

The similarity between objects appears in various forms[27]: ratings of pairs, sortings of objects, communality betweeen associations, substitutability, and correlation between occurrences. Here, the solid-solution formability for combinations of elements are discussed, along with the measure of similarity, in terms of substitutability between the elements combinations. Each nondisjoint pair of alloys A_i and A_j in $\mathcal{D}$ is a source of evidence for measuring the substitutability between element combinations $C_t = A_i - (A_i \cap A_j) = A_i - A_j$ and $C_v = A_j - (A_i \cap A_j) = A_j - A_i$ (Figure 1 a). The nonempty intersection set $A_i \cap A_j$ is considered as the context for the similarity measurement. If $y_{A_i} = y_{A_j}$ then C_t and C_v are substitutable, otherwise C_t and C_v are not substitutable.

To model evidence about the similarity between any pair of element combinations, we first define a frame of discernment[25] $\Omega_{sim} = \{similar, dissimilar\}$ containing all possible values.

The evidence collected from alloys A_i and A_j is then represented by a mass function[25] (or a basic probability assignment), $m_{A_i,A_j}^{C_t,C_v}$, which assigns probability masses to all the nonempty subsets of Ω_{sim} (i.e., $\{similar\}$, $\{dissimilar\}$, and $\{similar, dissimilar\}$), as follows:

$$m_{A_i,A_j}^{C_t,C_v}(\{similar\}) = \begin{cases} \alpha & \text{if } y_{A_i} = y_{A_j} \\ 0 & \text{otherwise} \end{cases}, \tag{1}$$

$$m_{A_i,A_j}^{C_t,C_v}(\{dissimilar\}) = \begin{cases} \alpha & \text{if } y_{A_i} \neq y_{A_j} \\ 0 & \text{otherwise} \end{cases}, \tag{2}$$

$$m_{A_i,A_j}^{C_t,C_v}(\{similar, dissimilar\}) = 1 - \alpha \tag{3}$$

Note that the masses assigned to $\{similar\}$ and $\{dissimilar\}$ indicate the degrees of belief exactly committed to A_i and A_j to support the similarity and dissimilarity between C_t and C_v, respectively. The weight assigned to subset $\{similar, dissimilar\}$ expresses the degree of belief that A_i and A_j provide no information about the similarity (or dissimilarity) between C_t and C_v. Here, the parameter α is determined by an exhaustive search ($0<\alpha<1$) for the best cross-validation score (Section IV C). We retain some degree of uncertainty ($1-\alpha$) about the similarities collected from each piece of evidence for dealing with the inconsistencies in the data set. The sum of the masses assigned to all three nonempty subsets of Ω_{sim} is 1.

Suppose that we can collect multiple pieces of evidence from $\mathcal{D}$ to compare two element combinations C_t and C_v, all obtained mass functions corresponding to those pieces of evidence are then combined using the Dempster rule of combinations[24] to assign the final mass $m_{\mathcal{D}}^{C_t,C_v}$ (Section IV A). Similar analyses are performed for all pairs of element combinations of interest to obtain a symmetric matrix M consisting of all the similarities between them ($M[t,v] = M[v,t] = m_{\mathcal{D}}^{C_t,C_v}(\{similar\})$).

For hypothesizing whether a potential alloy A_{new} forms an HEA phase, we apply the substitution method using the obtained matrix M. We replace a combination of elements, C_t, in an existing alloy, A_k, ($C_t \subset A_k$) with a combination of elements, C_v, adequate to obtain alloy A_{new} showing a property (label $y_{A_{new}}$) similar to that of A_k (label y_{A_k}). On the basis of the label of A_k and the similarity between C_t and C_v, the basic beliefs on the label of A_{new} are quantified (Figure 1 b). If C_t and C_v are substitutable (non-substitutable), this serves as a piece of evidence that the labels of A_{new} and A_k are the same (different).

To model evidence about existence of HEA phase in a particular alloy, we first define a frame of discernment[25] $\Omega_{HEA} = \{HEA, \neg HEA\}$.The evidence collected from A_k, C_t, and C_v is then represented by the mass function $m^{A_{new}}_{A_k,C_t\leftarrow C_v}$, which assigns probability masses to all the nonempty subsets of Ω_{HEA} (i.e., $\{HEA\}$, $\{\neg HEA\}$, and $\{HEA, \neg HEA\}$), as follows:

$$m^{A_{new}}_{A_k,C_t\leftarrow C_v}(\{HEA\}) = \begin{cases} M[t,v] & \text{if } y_{A_k} = HEA \\ 0 & \text{otherwise} \end{cases}, \tag{4}$$

$$m^{A_{new}}_{A_k,C_t\leftarrow C_v}(\{\neg HEA\}) = \begin{cases} M[t,v] & \text{if } y_{A_k} = \neg HEA \\ 0 & \text{otherwise} \end{cases}, \tag{5}$$

$$m^{A_{new}}_{A_k,C_t\leftarrow C_v}(\{HEA, \neg HEA\}) = 1 - M[t,v], \tag{6}$$

Note that the masses assigned to $\{HEA\}$ and $\{\neg HEA\}$ reflect the levels of confidence whereby A_k and the substitution of C_v for C_t support the probabilities that A_{new} is or is not an HEA, respectively. The mass assigned to subset $\{HEA, \neg HEA\}$, expresses the probability that A_k, C_t, and C_v provide no information about the property of A_{new}. The sum of the probability masses assigned to all three nonempty subsets of Ω_{HEA} is 1.

We assume that for a specific hypothetical alloy, A_{new}, we can collect pieces of evidence about its properties from $\mathcal{D}$ (pair of A_{host} and the corresponding substitution to obtain A_{new} from A_{host}). The obtained mass functions for A_{new} are then combined using the Dempster rule[24] to obtain a final mass function $m^{A_{new}}$ (Section IV A). Similar analyses are performed for all the possible alloys (A_{new}) that are not included in the observed data. We then use the final value of $m^{A_{new}}_{\mathcal{D}}(\{HEA\})$ for sorting the ranking of recommendation. The proposed recommender system considers the alloys with a higher value of $m^{A_{new}}_{\mathcal{D}}(\{HEA\})$ to have the greater potential of having HEA phases.

B. Experimental settings

We use eight data sets (Table I) consisting of binary, ternary, quaternary, and quinary alloys comprising multiple equiatomically combined elements to evaluate the proposed system for recommending HEAs and revealing the HEA formation mechanisms. The alloys contained in the data sets comprise $\mathcal{E} =$ { Fe, Co, Ir, Cu, Ni, Pt, Pd, Rh, Au, Ag, Ru, Os,

Si, As, Al, Tc, Re, Mn, Ta, Ti, W, Mo, Cr, V, Hf, Nb, and Zr}. Any alloy contained in the following data sets is predicted as an HEA if its order-disorder transition temperature is below its melting temperature. All the data sets are shown in detail in Supplementary Section II.

It should be noted that our system has the capability to collect and combine evidence from multiple data sets to reasonably draw the final conclusions. However, in the evaluation of HEA-recommendation capability, each data set comes from a different experiment or calculation method; therefore, we evaluate the proposed method with each data set separately to ensure the consistency between the training and test sets.

We compare the HEA-recommendation performance of the proposed ERS with those of matrix-based recommender systems[32] previously developed using nonnegative matrix factorization (NMF)[33] and singular-value decomposition (SVD)[34]. To use the matrix-based recommender systems for exploring potential HEAs, we apply two types of rating-matrix representations. In addition, the performances of recommender systems based on supervised-learning methods (support vector machines[35] (SVM), logistic-regression[36], decision tree[37], and Naïve-Bayes[38]) are compared with that of the ERS. We apply a compositional descriptor to employ the SVM- and logistic-regression-based recommender systems. The binary elemental descriptor is used to represent the alloys in our system and in the decision-tree- and Naïve-Bayes-based recommender systems. The material descriptors are shown in detail in the Methods section (IV B).

C. Learning about the similarity between elements

By applying the proposed ERS to the $\mathcal{D}_{\text{ASMI16}}$, $\mathcal{D}_{\text{CALPHAD}}$, $\mathcal{D}_{\text{AFLOW}}$, and $\mathcal{D}_{\text{LTVC}}$ data sets (Table I), we assess the similarity between the $\mathcal{E}$ elements and all the possible binary combinations obtained therein. Figure 2 (a, b, c, and d) show the M_{ASMI16}, M_{CALPHAD}, M_{AFLOW}, and M_{LTVC} similarity matrices obtained for all the $\mathcal{E}$ elements in the first four experiments. These similarity matrices are then properly transformed into distance matrices to which Ward's hierarchical agglomerative clustering[39] can be applied to construct the corresponding hierarchically clustered structures of these elements (Figure 2 e, f, g, and h).

The similarity matrix M_{ASMI16} reveals three distinct element groups (Figure 2 e) consisting of Ti, V, Cr, Mn, Zr, Nb, Mo, Hf, Ta, and W; Fe, Co, Ni, Cu, Rh, Pd, Ir, Pt, and Au;

and Al, Ag, Tc, Si, Ru, As, Re, and Os, where the first two groups correspond to the early and late transition metals, respectively. Given the similar physical and chemical properties of these elements, the high degree of similarity between the elements within the same group, as revealed by the ERS, is rational. Interestingly, the matrix M_{ASMI16} shows a remarkable similarity between Mn (an earlier transition metal) and Au (a late transition metal). Furthermore, the similarity matrix M_{ASMI16} indicates none of the belief about the similarity among the elements in the third group and between the elements of the third group and the other two groups because the binary alloys contained in $\mathcal{D}_{\mathrm{ASMI16}}$ do not contain these elements (Supplementary Figure 2 a). Therefore, no evidence of similarities can be collected from $\mathcal{D}_{\mathrm{ASMI16}}$ for these elements.

The similarity matrix M_{CALPHAD} also reveals three somewhat modified element groups (Figure 2 f) compared to those obtained from $\mathcal{D}_{\mathrm{ASMI16}}$. Because $\mathcal{D}_{\mathrm{CALPHAD}}$ contains some Tc- and Re-containing alloys, these elements join the group of early transition metals. Similarly, $\mathcal{D}_{\mathrm{CALPHAD}}$ contains more Ag- and Au-containing alloys, and these elements join the group of late transition metals. Therefore, only Al, Si, As, and Os remain in the third group. Although no evidence of any similarities between Si and As can be collected from $\mathcal{D}_{\mathrm{CALPHAD}}$ (Supplementary Figure 2 a), Os and Al are somewhat similar to the first and second groups, respectively.

In contrast, it is difficult to divide all the elements contained in $\mathcal{E}$ into groups according to the matrix M_{AFLOW}. However, some characteristic groups of metallic elements are distinct. Although two distinct groups of early or late transition metals are observed(Figure 2 g), there are some notable differences between these results and those obtained from $\mathcal{D}_{\mathrm{ASMI16}}$ and $\mathcal{D}_{\mathrm{CALPHAD}}$ (Supplementary Section III). In addition, the similarity matrix M_{AFLOW} does not show any similarity between Os and any of the other elements because very few Os-containing alloys are contained in the data set (Supplementary Figure 2 a). Furthermore, the similarity matrices M_{LTVC} and M_{AFLOW} are approximately similar. However, the hierarchically clustered structure constructed from $\mathcal{D}_{\mathrm{LTVC}}$ indicates that Cu, Ag, and Au form a distinct subgroup (Figure 2 h).

Figure 3 shows the correlation between the pairwise similarities learned from the $\mathcal{D}_{\mathrm{AFLOW}}$ and $\mathcal{D}_{\mathrm{LTVC}}$ data sets and the corresponding difference between the periodic-table group index obtained for each of the transition metal pairs contained in $\mathcal{E}$. Clearly, the elements showing the same periodic-table group index ($\Delta_{group} = 0$) tend to show high similarity scores (Figure

3 a and c) and low dissimilarity scores (Figure 3 b and d). Therefore, the elements in the same group similarly contribute to HEA formation and are substitutable for each other. However, it should be noted that several pairs of elements have a similarity with a low degree of belief even though they belong to the same groups, i.e. {(Ti, Zr), (Cu, Ag), (Fe, Ru)} in $\mathcal{D}_{\mathrm{AFLOW}}$ and {(Ti, Zr), (Mn, Re), (Ni, Pd)} in $\mathcal{D}_{\mathrm{LTVC}}$ (Figure 2 c and d).

Furthermore, as the difference in the group index increases from 0 to 4, the similarity between the elements decreases ($\Delta_{group} : 0 \rightarrow 4$). The results learned from the $\mathcal{D}_{\mathrm{AFLOW}}$ and $\mathcal{D}_{\mathrm{LTVC}}$ data sets both show that the elements are the least similar when the difference between their group indices is three or four. However, the elements become slightly more similar as Δ_{group} increases from 5 to 7, which is consistent with the domain knowledge about the differences between early and late transition metals.

D. Evaluation of recommendation capability by cross-validation

We apply k-fold cross-validation to the $\mathcal{D}_{\mathrm{ASMI16}}$, $\mathcal{D}_{\mathrm{CALPHAD}}$, $\mathcal{D}_{\mathrm{AFLOW}}$, and $\mathcal{D}_{\mathrm{LTVC}}$ data sets to assess the HEA-recommendation capabilities of the ERS, the four matrix-based recommender systems (NMF and SVD, each one with two types of matrix representations)[32]. These two matrix representations (*type 1* and *type 2*) decompose an alloy into two elementary components A and B with different sizes (Section IV B). We also compare the ERS with the four supervised-learning-method-based (i.e., decision tree, Naïve-Bayes, logistic-regression, and SVM) recommender systems.

The learned similarity matrix is used to rank all the alloys contained in the test set and all the possible combinatorial alloys other than those used to train the similarity matrix. The resulting alloy rankings are then used to evaluate the HEA-recommendation performance. We designed a virtual experiment that sequentially identifies the alloys on the basis of the order in which they were previously ranked. To evaluate the HEA-recommendation capability of the proposed ERS, we monitor the rank of HEAs in the test set and the HEA recall depending on the number of trials required to identify all possible HEAs. The detailed experimental conditions are shown in the section IV D.

Figure 4 (a–d) illustrate the distributions of the HEA ranks of the test set recommended by the different systems. The HEAs in the test set are generally recommended with higher rank using the ERS (i.e., the ERS rank distributions are on the left of the curves for the

other systems). Consequently, the ERS can significantly reduce the number of trials required to recover the HEAs in the test set compared to the competitor systems. Only in the experiment with $\mathcal{D}_{\text{ASMI16}}$, the distributions of the rank using the ERS and NMF (*type 2*) are somewhat similar (Fig. 4 a). We also monitor the dependence of the HEA recall ratio on the number of trials required to measure the HEA-recommendation performance of the ERS quantitatively. In summary, the ERS outperforms the other systems in recalling one-half and three-quarters of the HEAs in the test set (Supplementary Section IV A). However, the ERS cannot reliably recall the remaining one-quarter of the HEAs because insufficient evidence is available in the training data to make inferences about the remaining HEAs. Interestingly, in the $\mathcal{D}_{\text{ASMI16}}$ and $\mathcal{D}_{\text{CALPHAD}}$ experiments, the supervised-method-based recommender systems either approximately randomly selected possible HEAs (Naïve Bayes and decision tree) or could not rank any at all (logistic regression and SVM) because these data sets contain only positively labeled HEAs.

E. Evaluation of recommendation capability by extrapolation

The cross-validation experiments show the recommendation systems based on supervised learning methods (SVMs[35], logistic regression[36], decision trees[37], and Naïve-Bayes[38]) have much lower recommendation performance. These results are attributed to the inappropriate assessment of the similarity between alloys with different numbers of compositions (Section IV B). Therefore, to evaluate the HEA-recommendation capability when extrapolating the number of components, we focus on comparing the performances of the ERS with those of matrix-based recommender systems. The detailed experimental settings are shown in the Methods section (IV E).

Figure 4 (e–h) illustrate the distributions of the recommended HEA rank of the quaternary and quinary HEAs in the test set that are extrapolated using recommender systems. The obtained results show that the ERS outperforms the capability of the competitor systems for recommending quaternary HEAs (Fig. 4 e and f) and substantially outperforms the capability of the other systems for recommending quinary HEAs (Fig. 4 g and h). Interestingly, in the experiments with $\mathcal{D}^{\text{quinary}}_{\text{LTVC}}$ and $\mathcal{D}^{\text{quinary}}_{\text{AFLOW}}$, the numbers of quinary HEAs in the test set, and those found in the top 100 and top 1,000 HEA candidates recommended by the ERS, are much larger than those predicted by the competitor systems. These numbers

are very high because the two data sets only contain quinary alloys of the early transition metals. Much of the evidence of the similarities between these element combinations can be collected from the corresponding data sets containing binary, ternary, and quaternary alloys (Supplementary Figure 2 b). Moreover, to recall 50 and 75% of the quinary HEAs from these data sets, approximately 10-100 fewer trials are required by the ERS than by the NMF and SVD-based recommender systems. The results of experiments monitoring the dependence of the HEA recall ratio on the number of trials required are listed in detail in Supplementary Section IV B. In the absence of sufficient evidence, the answer of the system, regarding a mixture of many types of elements, will retain a large degree of uncertainty ($m(\{HEA, \neg HEA\}) \approx 1$).

F. Synthesis of recommended FeMnCo-based HEAs

Fe-Co-based film soft-magnetic materials have attracted interest from device community and will be applied to improve the performance of next-generation high-power devices[40]. Therefore, we focus on Fe-Co-based quaternary alloys containing the first transition-series elements. We combine all evidence collected from all the data sets to recommend quaternary Fe-Co-based HEAs for experimental validation.

Figure 5 a shows the recommended possible magnetic quaternary HEAs containing Fe, Mn, and Co. Clearly, FeMnCoNi is the only HEA candidate recommended with a belief higher than 0.5. Although FeMnCoCr and FeMnCoCu are HEA candidates recommended with the next highest belief, some uncertainty still remains as to their potential as HEAs. Therefore, we chose FeMnCoNi as the target HEA candidate for the experimental validation (Figure 5, see the Methods section IV F for further information).

Figure 5 c shows a 2D-XRD image of a region of the $Fe_{0.25}Co_{0.25}Mn_{0.25}Ni_{0.25}$ alloy annealed at 400°C. A reflection attributed to the (110) plane of the BCC crystal structure appears in the ring pattern at $2\theta = 44.7°$ (PDF 03-065-7519[41]). Note that out-of-plane XRD measurements were also performed to identify the crystal structure in more detail, as shown in Supplementary Figure 6(a), indicating the formation of a polycrystalline film. Reportedly, the BCC crystal structure of the FeCoMn alloy is stable[42], and previous reports have mentioned that FeCoMnNi alloy has an face-centered cubic FCC structure in high temperature synthesized bulk; however, detailed information is still not available[43,44].

Therefore, to investigate the stability of the crystal structure, the effect of Ni doping on the crystal structure was analyzed based on the heat map generated from the X-ray diffraction patterns of FeCoMn films prepared with various Ni contents (Fig 5 d). For an Ni content above 0.3, the FCC structure is also observed at $2\theta = 43.5°$, corresponding to the (111) reflection [Supplementary Figure 6(b)] (PDF 03-065-5131[41]). These results suggest that the $Fe_{0.25}Co_{0.25}Mn_{0.25}Ni_{0.25}$ HEA shows a BCC structure. In our experiment, the BCC structure of the starting material, FeCoMn, is considered as an essential reason for which the thin films produced by this method tend to be in the BCC phase.

III. DISCUSSION

Application of inductive approach usually requires sufficient and balanced data to ensure prediction accuracy. However, material data is usually lacking or heavily biased toward positive results (Table I). It is very challenging to build prediction model using such small data and a very heavy skew toward positive results. In addition, conflicts within and between data sets of materials are also challenges that inductive approaches must overcome. Therefore quantitative assessment of the uncertainty of the prediction itself is indispensable. The ERS has the advantage in dealing with these situations. Instead of forcibly merging data from multiple data sets, our system rationally consider each data set as a source of evidence and combine the evidence to reasonably draw the final conclusions for recommending HEA, where the uncertainty can be quantitatively evaluated.

To serve the purpose of screening the elements combination forming HEA phases, the ERS focuses on fundamental question of whether the HEA phase exists. We design a frame of discernment $\Omega_{HEA} = \{HEA, \neg HEA\}$ to model the existence of HEA phases with mass functions. Consequently, the ERS has not answered essential questions regarding the structure and other properties of the HEAs. However, by redesigning the frame of discernment reflecting the additional properties of interest, we can also construct a model that can recommend the potential alloys forming the HEA phases with the desirable properties. Furthermore, in the experimental validation, detailed quantitative investigation of the secondary phases in the synthesized FeCoMnNi alloy thin film was not done due to technical difficulties.

REFERENCES

[1]Yeh, J.-W. *et al.* Nanostructured high-entropy alloys with multiple principal elements: Novel alloy design concepts and outcomes. *Adv. Eng. Mater.* **6**, 299–303 (2004).

[2]Cantor, B., Chang, I., Knight, P. & Vincent, A. Microstructural development in equiatomic multicomponent alloys. *Mater. Sci. Eng. A* **375**, 213 – 218 (2004).

[3]Senkov, O. N., Miller, J. D., Miracle, D. B. & Woodward, C. Accelerated exploration of multi-principal element alloys with solid solution phases. *Nat. Commun.* **6**, 6529 (2015).

[4]Rickman, J. M. *et al.* Materials informatics for the screening of multi-principal elements and high-entropy alloys. *Nat. Commun.* **10**, 2618 (2019).

[5]Tsai, M.-H. & Yeh, J.-W. High-entropy alloys: A critical review. *Mater. Res. Lett.* **2**, 107–123 (2014).

[6]GUO, S. & LIU, C. Phase stability in high entropy alloys: Formation of solid-solution phase or amorphous phase. *Prog. Nat. Sci.* **21**, 433–446 (2011).

[7]Zhang, Y., Guo, S., Liu, C. T. & Yang, X. *Phase Formation Rules*, 21–49 (Springer International Publishing, Cham, 2016). URL `https://doi.org/10.1007/978-3-319-27013-5_2`.

[8]Huhn, W. P. & Widom, M. Prediction of A2 to B2 Phase Transition in the High-Entropy Alloy Mo-Nb-Ta-W. *JOM* **65**, 1772–1779 (2013).

[9]van de Walle, A. & Asta, M. Self-driven lattice-model monte carlo simulations of alloy thermodynamic properties and phase diagrams. *Model. Simul. Mater. Sci. Eng.* **10**, 521–538 (2002).

[10]Zhang, Y., Zhou, Y., Lin, J., Chen, G. & Liaw, P. Solid-solution phase formation rules for multi-component alloys. *Adv. Eng. Mater.* **10**, 534–538 (2008).

[11]Ye, Y., Wang, Q., Lu, J., Liu, C. & Yang, Y. Design of high entropy alloys: A single-parameter thermodynamic rule. *Scr. Mater.* **104**, 53–55 (2015).

[12]Tsai, M.-H. Three strategies for the design of advanced high-entropy alloys. *Entropy* **18**, 252 (2016).

[13]Tsai, M.-H., Tsai, R.-C., Chang, T. & Huang, W.-F. Intermetallic phases in high-entropy alloys: Statistical analysis of their prevalence and structural inheritance. *Metals* **9**, 247 (2019).

[14]Huang, W., Martin, P. & Zhuang, H. L. Machine-learning phase prediction of high-entropy

alloys. *Acta Mater.* **169**, 225–236 (2019).

[15]George, E. P., Raabe, D. & Ritchie, R. O. High-entropy alloys. *Nat. Rev. Mater.* **4**, 515–534 (2019).

[16]Konno, T. *et al.* Deep learning model for finding new superconductors. *Phys. Rev. B* **103**, 014509 (2021).

[17]Pham, T. L., Kino, H., Terakura, K., Miyake, T. & Dam, H. C. Novel mixture model for the representation of potential energy surfaces. *J. Chem. Phys.* **145**, 154103 (2016).

[18]Kobayashi, R., Giofré, D., Junge, T., Ceriotti, M. & Curtin, W. A. Neural network potential for al-mg-si alloys. *Phys. Rev. Materials* **1**, 053604 (2017).

[19]Tamura, T. *et al.* Fast and scalable prediction of local energy at grain boundaries: machine-learning based modeling of first-principles calculations. *Model. Simul. Mater. Sci. Eng.* **25**, 075003 (2017).

[20]Seko, A., Hayashi, H., Nakayama, K., Takahashi, A. & Tanaka, I. Representation of compounds for machine-learning prediction of physical properties. *Phys. Rev. B* **95**, 144110 (2017).

[21]Kobayashi, R. nap: A molecular dynamics package with parameter-optimization programs for classical and machine-learning potentials. *J. Open Source Softw.* **6**, 2768 (2021).

[22]Nguyen, D.-N. *et al.* Committee machine that votes for similarity between materials. *IUCrJ* **5**, 830–840 (2018).

[23]Pham, T. L. *et al.* Machine learning reveals orbital interaction in materials. *Sci. Technol. Adv. Mater.* **18**, 756–765 (2017).

[24]Dempster, A. P. A Generalization of Bayesian Inference. *J R STAT SOC B* **30**, 205–232 (1968).

[25]Shafer, G. *A Mathematical Theory of Evidence* (Princeton University Press, 1976). URL `https://doi.org/10.2307/j.ctv10vm1qb`.

[26]Denœux, T., Dubois, D. & Prade, H. Representations of uncertainty in artificial intelligence: Beyond probability and possibility. In Marquis, P., Papini, O. & Prade, H. (eds.) *A Guided Tour of Artificial Intelligence Research*, vol. 1, chap. 4, 119–150 (Springer Verlag, 2020). URL `https://doi.org/10.1007/978-3-030-06164-7_4`.

[27]Tversky, A. Features of similarity. *Psychol. Rev.* **84**, 327–352 (1977).

[28]Binary Alloy Phase Diagrams. In Okamoto, H., Schlesinger, M. & Mueller, E. (eds.) *Alloy Phase Diagrams*, vol. 3 (ASM International, 2016). URL `https://doi.org/10.31399/`

asm.hb.v03.a0006247.

[29]Alman, D. Searching for next single-phase high-entropy alloy compositions. *Entropy* **15**, 4504–4519 (2013).

[30]Zhang, F. *et al.* An understanding of high entropy alloys from phase diagram calculations. *CALPHAD* **45**, 1–10 (2014).

[31]Lederer, Y., Toher, C., Vecchio, K. S. & Curtarolo, S. The search for high entropy alloys: A high-throughput ab-initio approach. *Acta Mater.* **159**, 364–383 (2018).

[32]Seko, A., Hayashi, H., Kashima, H. & Tanaka, I. Matrix- and tensor-based recommender systems for the discovery of currently unknown inorganic compounds. *Phys. Rev. Mater.* **2**, 013805 (2018).

[33]Paatero, P., Tapper, U., Aalto, P. & Kulmala, M. Matrix factorization methods for analysing diffusion battery data. *J. Aerosol Sci.* **22**, S273–S276 (1991).

[34]Golub, G. H. & Reinsch, C. Singular value decomposition and least squares solutions. *Numer. Math.* **14**, 403–420 (1970).

[35]Hearst, M. A. Support vector machines. *IEEE Intell. Syst.* **13**, 18–28 (1998).

[36]LaValley, M. P. Logistic regression. *Circulation* **117**, 2395–2399 (2008).

[37]Quinlan, J. R. Induction of decision trees. *Mach. Learn.* **1**, 81–106 (1986).

[38]Yager, R. R. An extension of the naive bayesian classifier. *Inf. Sci.* **176**, 577–588 (2006).

[39]Murtagh, F. & Legendre, P. Ward's hierarchical agglomerative clustering method: Which algorithms implement ward's criterion? *J. Classif.* **31**, 274–295 (2014).

[40]Silveyra, J. M., Ferrara, E., Huber, D. L. & Monson, T. C. Soft magnetic materials for a sustainable and electrified world. *Science* **362** (2018).

[41]Gates-Rector, S. & Blanton, T. The powder diffraction file: a quality materials characterization database. *Powder Diffr.* **34**, 352–360 (2019).

[42]Snow, R. J., Bhatkar, H., N'Diaye, A. T., Arenholz, E. & Idzerda, Y. U. Large moments in bcc $Fe_xCo_yMn_z$ ternary alloy thin films. *Appl. Phys. Lett.* **112**, 072403 (2018).

[43]Wu, Z., Bei, H., Otto, F., Pharr, G. & George, E. Recovery, recrystallization, grain growth and phase stability of a family of fcc-structured multi-component equiatomic solid solution alloys. *Intermetallics* **46**, 131–140 (2014).

[44]Cui, P. *et al.* Effect of Ti on microstructures and mechanical properties of high entropy alloys based on CoFeMnNi system. *Mater. Sci. Eng. A* **737**, 198–204 (2018).

[45]Seko, A., Togo, A. & Tanaka, I. *Descriptors for Machine Learning of Materials*

Data, 3–23 (Springer Singapore, Singapore, 2018). URL https://doi.org/10.1007/978-981-10-7617-6_1.

[46]Nyshadham, C. *et al.* A computational high-throughput search for new ternary superalloys. *Acta Mater.* **122**, 438–447 (2017).

[47]Koinuma, H. & Takeuchi, I. Combinatorial solid-state chemistry of inorganic materials. *Nat. Mater.* **3**, 429–438 (2004).

[48]He, B. B. *Geometry and Fundamentals*, chap. 2, 29–55 (John Wiley & Sons, Ltd, 2018). URL https://doi.org/10.1002/9781119356080.ch2.

[49]Dam, H.-C., Kino, H. & Ha, M.-Q. High-entropy alloys data sets of evidence-based recommender system for combinatorial materials synthesis (2021). URL https://doi.org/10.5281/zenodo.4557463. https://doi.org/10.5281/zenodo.4557463.

[50]Dam, H.-C. & Ha, M.-Q. ERS Capsule: Source code for the paper "Evidence-based recommender system and experimental validation for high-entropy alloys" (2021). URL https://doi.org/10.24433/CO.5083650.v1. https://doi.org/10.24433/CO.5083650.v1.

ACKNOWLEDGMENTS

This work is supported by the Ministry of Education, Culture, Sports, Science, and Technology of Japan (MEXT) ESICMM Grant Number 12016013; the Program for Promoting Research on the Supercomputer Fugaku (DPMSD); the JST-Mirai Program "Development of Materials Design Workflow and Data Library for Materials Foundry," Grant Number JPMJMI18G5; and JSPS KAKENHI Grants 20K05301, JP19H05815 (Grants-in-Aid for Scientific Research on Innovative Areas Interface Ionics), and 20K05068, Japan.

AUTHOR CONTRIBUTIONS

M.-Q.H.: Conceived and designed the experiments, Performed the experiments, Analyzed the data, Contributed materials/analysis tools, Wrote the paper. **D.-N.N.**: Conceived and designed the experiments, Analyzed the data, Contributed materials/analysis tools. **V.-C.N.**: Performed the experiments. **T.N.**: Conceived and designed the experiments, Performed the experiments, Analyzed the data, Contributed materials/analysis tools, Wrote the paper. **T.C.**: Analyzed the data. **H.K.**: Conceived and designed the experiments, An-

alyzed the data, Wrote the paper. **T.M.**: Conceived and designed the experiments, Wrote the paper. **T.D.**: Conceived and designed the experiments, Writing – Review & Editing. **V.-N.H.**: Conceived and designed the experiments, Wrote the paper. **H.-C.D.**: Conceived and designed the experiments, Performed the experiments, Analyzed the data, Contributed materials/analysis tools, Wrote the paper.

COMPETING INTERESTS

The authors declare no competing interests.

FIGURE LEGENDS

Figure 1: An illustration of Evidence-based recommender methodology. Venn diagrams shows the logical relationships between alloys (A_i, A_j, A_k, and A_{new}) and element combinations (C_t and C_v), which are used to model evidence of (a) similarities between element combinations and (b) new alloys by element substitution method.

Figure 2: Visualization of similarities between elements. Top: Heat maps for similarity matrices (a) M_{ASMI16}, (b) M_{CALPHAD}, (c) M_{AFLOW}, and (d) M_{LTVC}. Each matrix element is the probability mass that the similarity mass function of the corresponding element pair is assigned to subset $\{similar\}$ of Ω_{sim}. These matrix elements indicate the degree of belief learned from the similarity data of the corresponding element pairs. In these figures, the degrees of belief are illustrated using colormap. Bottom: Hierarchically clustered structures of all elements in $\mathcal{E}$ constructed using hierarchical agglomerative clustering and these similarity matrices (e) M_{ASMI16}, (f) M_{CALPHAD}, (g) M_{AFLOW}, and (h) M_{LTVC} data sets. The blue, green, and gray regions indicate groups of early and late transition metals, and elements without similarity evidence, respectively.

Figure 3: Correlation between pairwise similarity and difference in group index (Δ_{group}) of elements. Sub-figures illustrate the distribution of pairwise similarities, which are obtained from (a, b) $\mathcal{D}_{\mathrm{AFLOW}}$ and (c, d) $\mathcal{D}_{\mathrm{LTVC}}$ data sets, according to the Δ_{group} of these element pair. Colormap illustrates the estimated density of the distribution of pairwise similarity.

Figure 4: Evaluation of HEA-recommendation capability. Probability density functions of the rank of the HEAs in the test sets in (a) $\mathcal{D}_{\mathrm{ASMI16}}$, (b) $\mathcal{D}_{\mathrm{CALPHAD}}$, (c) $\mathcal{D}_{\mathrm{AFLOW}}$, (d)

$\mathcal{D}_{\text{LTVC}}$, (e) $\mathcal{D}^{\text{quaternary}}_{\text{AFLOW}}$, (f) $\mathcal{D}^{\text{quaternary}}_{\text{LTVC}}$, (g) $\mathcal{D}^{\text{quinary}}_{\text{AFLOW}}$, and (h) $\mathcal{D}^{\text{quinary}}_{\text{LTVC}}$ experiments. The ranks of HEAs in the test sets are expressed on a base-10 logarithmic scale. The HEAs with higher ranking order are recommended materials with a firmer belief in the formation of the HEA phase.

Figure 5: Recommendation and experimental validation for thin film of FeCoMnNi HEA. (a) Recommended candidates for Fe-Co-based HEAs containing first-transition-series elements: FeMnCoTi, FeMnCoV, FeMnCoCr, FeMnCoNi, and FeMnCoCu. (b) Schematic illustration of the sample, which includes 200 cycles of 0.5 nm spread film, was fabricated on SiO_2/Si (100) substrate using the combinatorial method. Each spread film consists of a 0.25 nm FeCoMn sublayer and a 0.25 nm 1-x(FeCoMn)-xNi sublayer. (c) 2D-XRD image of $Fe_{0.25}Co_{0.25}Mn_{0.25}Ni_{0.25}$ thin film measured by changing the incident angle of X-rays. (d) Heat map shows the dependence of the X-ray diffraction intensity of 1-x(FeCoMn)-xNi films on Ni composition and diffraction angle θ.

TABLE I. Summary of alloy data sets used in evaluation experiments. No. alloys: number of alloys included in each data set; No. HEAs: number of the alloys confirmed or estimated to form HEA phase in each data set; No. candidates: number of possible alloys generated using the set of all elements in the data sets. The "HEA rate" is the ratio of No. HEA to No. alloys, whereas the "Observation rate" is the ratio of No. alloys to No. candidates.

[h!] *Data set*	No. alloys	No. HEAs	No. candidates	HEAs rate	Observation rate
$\mathcal{D}_{\text{ASMI16}}$[28]	45 binary alloys	45	351	100%	13%
$\mathcal{D}_{\text{CALPHAD}}$[3,29]	243 ternary alloys	243	2925	100%	9%
$\mathcal{D}_{\text{AFLOW}}$[30]	117 binary alloys	60	351	51%	33%
	441 ternary alloys	234	2925	53%	15%
$\mathcal{D}_{\text{LTVC}}$[31]	117 binary alloys	58	351	49%	33%
	441 ternary alloys	148	2925	33%	15%
$\mathcal{D}^{\text{quaternary}}_{\text{AFLOW}}$[30]	1,110 quaternary alloys	754	17,550	68%	6%
$\mathcal{D}^{\text{quaternary}}_{\text{LTVC}}$[31]	1,110 quaternary alloys	480	17,550	43%	6%
$\mathcal{D}^{\text{quinary}}_{\text{AFLOW}}$[30]	130 quinary alloys	129	80,730	99%	0.16%
$\mathcal{D}^{\text{quinary}}_{\text{LTVC}}$[31]	130 quinary alloys	91	80,730	70%	0.16%

IV. METHODS

A. Combining multiple pieces of evidence

We assume that we can collect q pieces of evidence from $\mathcal{D}$ to compare a specific pair of element combinations, C_t and C_v. If no evidence is found, the mass function m^{C_t,C_v}_{none} is initialized, which assigns a probability mass of 1 to subset $\{similar, dissimilar\}$. m^{C_t,C_v}_{none} models the condition under which no information about the similarity (or dissimilarity) between C_t and C_v is available. Any two pieces of evidence a and b modeled by the corresponding mass functions m^{C_t,C_v}_a and m^{C_t,C_v}_b can be combined using the Dempster rule[24] to assign the joint mass $m^{C_t,C_v}_{a,b}$ to each subset ω of Ω_{sim} (i.e. $\{similar\}$, $\{dissimilar\}$, or

$\{similar, dissimilar\}$) as follows:

$$\begin{aligned} m_{a,b}^{C_t,C_v}(\omega) &= \left(m_a^{C_t,C_v} \oplus m_b^{C_t,C_v}\right)(\omega) \\ &= \frac{\sum\limits_{\forall \omega_k \cap \omega_h = \omega} m_a^{C_t,C_v}(\omega_k) \times m_b^{C_t,C_v}(\omega_h)}{1 - \sum\limits_{\forall \omega_k \cap \omega_h = \emptyset} m_a^{C_t,C_v}(\omega_k) \times m_b^{C_t,C_v}(\omega_h)}, \end{aligned} \tag{7}$$

where ω, ω_k and ω_h are subsets of Ω_{sim}. Note that the Dempster rule is commutative and yields the same result by changing the order of $m_a^{C_t,C_v}$ and $m_b^{C_t,C_v}$. All the q obtained mass functions corresponding to the q collected pieces of evidence from $\mathcal{D}$ are then combined using the Dempster rule to assign the final mass $m_{\mathcal{D}}^{C_t,C_v}$ as follows:

$$m_{\mathcal{D}}^{C_t,C_v}(\omega) = \left(m_1^{C_t,C_v} \oplus m_2^{C_t,C_v} \oplus \cdots \oplus m_q^{C_t,C_v}\right)(\omega). \tag{8}$$

Multiple pieces of evidence about the label of each new alloy are combined using the similar manner. We assume that for a specific hypothetical alloy, A_{new}, we can collect pieces of evidence about its properties from $\mathcal{D}$ (pair of A_{host} and the corresponding substitution to obtain A_{new} from A_{host}). If no evidence is found, $m_{none}^{A_{new}}$ is initialized and a probability mass of 1 is applied to set $\{HEA, \neg HEA\}$. $m_{none}^{A_{new}}$ models the condition that no information about the label of A_{new} can be obtained from $\mathcal{D}$. The obtained mass functions for A_{new} are then combined using the Dempster rule[24] to obtain a final mass function $m_{\mathcal{D}}^{A_{new}}$ on Ω_{HEA}.

B. Materials descriptors

Descriptors, which are the representation of alloys, play a crucial role in building a recommender system to explore potential new HEAs. In this research, the raw data of alloys is represented in the form of elements combination. Several descriptors have been studied in materials informatics to represent the compounds[45]. To employ the data-driven approaches for this work, we applied compositional descriptor[20], rating matrix representation[32] and binary elemental descriptor[45].

Compositional descriptor represents an alloy by a set of 135 features composed of means, standard deviations, and covariance of established atomic representations that form the alloy. The descriptor can be applied not only to crystalline systems but also to molecular system. We adopted 15 atomic representations: (1) atomic number, (2) atomic mass,(3) period and (4) group in the periodic table, (5) first ionization energy, (6) second ionization energy,

(7) Pauling electronegativity, (8) Allen electronegativity, (9) van der Waals radius, (10) covalentradius, (11) atomic radius, (12) melting point, (13) boiling point, (14) density, and (15) specific heat. However, the compositional descriptor hardly distinguishes compounds which have different numbers of the atom because it is to regard the atomic representations of a compound as distributions of data. Therefore, the compositional descriptor cannot be applied in the case of having extrapolation in the number of components.

The rating matrix representation, which is a descriptor-free approach, shows a robust performance of recommendations for a wide variety of data sets in the Machine Learning community. Seko et al. adopted the representation to build a recommender system for exploring currently unknown chemically relevant compositions[32]. In that work, a composition data set needs to be transformed into just two feature sets, which corresponds to users and items in a user-item rating matrix. Ratings of missing elements are approximately predicted based on the similarity of features given by the representation. To build a recommender system for HEA, we first define the candidate alloys as AB, where A and B correspond to elementary components of the alloys. We introduce two kinds of matrix representations for the eight alloys data sets. An alloy is decomposed into two elementary components with the following number of elements.

- *Type 1*: $|A| \in \{1, 2\}$ and $|B| \in \{1, 2, 3\}$. The numbers of possible components A and B are respectively 378 and 3303. The size of the rating matrix is (378×3303).
- *Type 2*: $|A| = 1$ and $|B| \in \{1, 2, 3, 4\}$. The numbers of possible components A and B are 27 and 20853, respectively. The size of the rating matrix is (27×20853).

Binary elemental descriptors is simply a binary digit representing the presence of chemical elements. The number of binary elemental descriptors corresponds to the number of element types included in the training data. In this work, the alloys data sets are composed of 27 kinds of elements; Thus, an alloy is described by a 27-dimensional binary vector with elements of one or zero.

C. Tuning hyper-parameter of the ERS

Because data sets used in this work are the output of calculation prediction methods, we add some degree of uncertainty α in the mass function which models similarity evidence. In

each data set, we use grid search to determine the α that best reproduced the alloy labels in the data set (achieving best cross-validation score). Details of the cross-validation schemes are mentioned in Section IV D. The search space of α is from 0.01 to 0.9 with a step of 0.01. However, the relative magnitudes of (degree of belief HEA) and (degree of belief not HEA) are almost unchanged. In summary, the absolute value of alpha has little effect on the final result of the recommender system.

D. Experimental settings for cross-validation

Cross-validated testing accuracy rates of our method when considered as a supervised learning method are 80% and 75% in $\mathcal{D}_{\text{AFLOW}}$, and $\mathcal{D}_{\text{LTVC}}$ data sets, respectively, which are almost at the same level with those in the previous study[14]. However, our work pays more attention toward calculating the recall, which is the percentage of the total HEAs correctly classified. This recall value is a more appropriate evaluation measure compared to supervised learning accuracy for finding new combinations of elements having HEA phases.

Because the $\mathcal{D}_{\text{ASMI16}}$ data set only contains binary alloys, we can learn a similarity matrix between the elements from a training set sampled from $\mathcal{D}_{\text{ASMI16}}$. By applying the proposed process for recommending substituted alloys, we can rank all the possible binary alloys other than those in the training set. A total of 351 hypothetical binary alloys showing equivalent components can be generated from the 27 elements in $\mathcal{E}$, 45 of which are contained in $\mathcal{D}_{\text{ASMI16}}$. Because no information is available for the other 306 alloys, they are ranked by the constructed model. We apply 9-fold cross-validation to $\mathcal{D}_{\text{ASMI16}}$. A total of 40 out of the 45 alloys in $\mathcal{D}_{\text{ASMI16}}$ are used as the training set, and the remaining 5 alloys are used as the test set to evaluate the HEA recall rate. The model learned from the 40 alloys in the training set is then used to rank the other 311 alloys, including the 5 in the test set. This cross-validation is repeated 100 times so that the HEA-recommendation performance can be reliably calculated.

Because the $\mathcal{D}_{\text{CALPHAD}}$ data set only contains ternary alloys, we can learn a similarity matrix between the elements or binary combinations thereof from a training set sampled from $\mathcal{D}_{\text{CALPHAD}}$. We can build a model to rank all the possible ternary alloys other than those in the training set. There are 2,925 hypothetical ternary alloys showing equivalent components that can be generated from the 27 elements in $\mathcal{E}$, 243 of which are contained in

$\mathcal{D}_{\mathrm{CALPHAD}}$. Because no information is available for the other 2,682 alloys, they are ranked by the constructed model. We apply 9-fold cross-validation to $\mathcal{D}_{\mathrm{CALPHAD}}$ and use 216 of the 243 alloys in $\mathcal{D}_{\mathrm{CALPHAD}}$ as the training set. The remaining 27 alloys in $\mathcal{D}_{\mathrm{CALPHAD}}$ are used as the test set to evaluate the HEA recall rate. The model learned from the 216 alloys in the training set is used to rank the other 2,709 alloys, including the 27 in the test set. This cross-validation is also repeated 100 times to ensure the reliable evaluation of the HEA-recommendation performance.

In contrast, the $\mathcal{D}_{\mathrm{ASMI16}}$, $\mathcal{D}_{\mathrm{CALPHAD}}$, $\mathcal{D}_{\mathrm{AFLOW}}$, and $\mathcal{D}_{\mathrm{LTVC}}$ data sets contain both binary and ternary alloys. Owing to the information obtained from both types of alloys, we can learn a similarity matrix between the various elements, elements and binary combinations thereof, and binary element combinations obtained from the training set sampled from $\mathcal{D}_{\mathrm{AFLOW}}$ and $\mathcal{D}_{\mathrm{LTVC}}$. We can build a model to rank all the possible candidates for binary and ternary alloys other than those in the training set. There are 3,276 hypothetical binary and ternary alloys showing equivalent components that can be generated from the 27 elements in $\mathcal{E}$, 558 of which are contained in $\mathcal{D}_{\mathrm{AFLOW}}$. Because no information is available for the other 2,718 alloys, they are ranked by the constructed model. We apply 9-fold cross-validation to $\mathcal{D}_{\mathrm{AFLOW}}$ and use 496 of the 558 alloys in $\mathcal{D}_{\mathrm{AFLOW}}$ as the training set. The remaining 62 alloys in $\mathcal{D}_{\mathrm{AFLOW}}$ are used as the test set to evaluate the HEA recall rate. The model learned from the 496 alloys in the training set is used to rank the other 2,780 alloys including the 62 in the test set. The same evaluation method is applied to $\mathcal{D}_{\mathrm{AFLOW}}$.

A similar experiment is conducted with the $\mathcal{D}_{\mathrm{LTVC}}$ data set to evaluate the HEA-recommendation performance of the proposed ERS. Note that although the $\mathcal{D}_{\mathrm{LTVC}}$ data set contains the same alloys as the $\mathcal{D}_{\mathrm{AFLOW}}$ one, the target properties of the alloys are dissimilar because the values are estimated using different computation methods[31,46].

It should be noted that owing to the computational cost, these experiments do not use the selected alloys (i.e., those in the test set) to improve the accuracy of the HEA recommendation model for the next trial. A recommendation model based on the results of previous trials may work more accurately.

E. Experimental settings for evaluation of extrapolation capability

Because $\mathcal{D}_{\mathrm{AFLOW}}$ contains both binary and ternary alloys, we can learn the similarities between the various elements and binary combinations thereof. Consequently, we can apply the ERS to $\mathcal{D}_{\mathrm{AFLOW}}$ to rank the 17,550 quaternary alloys comprising the 27 elements contained in $\mathcal{E}$. Additionally, $\mathcal{D}_{\mathrm{AFLOW}}$ and $\mathcal{D}_{\mathrm{AFLOW}}^{\mathrm{quaternary}}$ are both used to build a recommender system that ranks all the possible candidates (i.e., 80,730 alloys) for synthesizing quinary HEAs. The 754 quaternary HEAs in $\mathcal{D}_{\mathrm{AFLOW}}^{\mathrm{quaternary}}$ and 129 quinary HEAs in $\mathcal{D}_{\mathrm{AFLOW}}^{\mathrm{quinary}}$ are used to monitor the HEA recall rate for recommending quaternary and quinary HEAs, respectively. Moreover, similar experiments are conducted on the $\mathcal{D}_{\mathrm{LTVC}}$, $\mathcal{D}_{\mathrm{LTVC}}^{\mathrm{quaternary}}$, and $\mathcal{D}_{\mathrm{LTVC}}^{\mathrm{quinary}}$ data sets to evaluate the HEA-recommendation performance of the ERS.

F. Synthesis of FeMnCoNi high entropy alloy thin film

As a case study, we fabricated a high entropy alloy film of $Fe_{0.25}Co_{0.25}Mn_{0.25}Ni_{0.25}$. A 100 nm thick-thermal oxidized SiO_2/Si (100) substrate was used. After the organic solvent and deionized water cleaning, the substrate was loaded in a combinatorial multi target RF-sputtering system (COMET inc., CMS-6400). To identify the stable crystal structure and its composition dependence, a composition spread film was fabricated by combinatorial method[47]. For the composition spread film, we used two targets of FeCoMn (1:1:1) and Ni (3N grade). The base pressure was below 1×10^{-5} Pa, and Ar gas pressure was set as 0.3 Pa. To adjust the deposition rate as 0.23 ± 0.01 nm/s, RF-sputtering powers of FeCoMn and Ni targets were set at 100 and 120 W, respectively. To enhance the crystallinity, the sample was annealed at 400°C for 30 min under a vacuum condition below 6×10^{-3} Pa (Advanced RIKO, MILA-3000).

Figure 5(b) shows the sample structure. The composition film layer consists of three layers. One is a single FeCoMn layer with a thickness of 0.25 nm. The other layers are composition spread film formed by FeCoMn and Ni layers. For the composition-spread film deposition, during the FeCoMn layer deposition, a mask moved 18.5 mm at constant speed from a point 1.5 mm from the edge of the substrate to another end where the film thickness gradually changed. After that, the targets were changed to Ni. The mask moved to the opposite direction during the Ni film deposition. The total thickness of one unit of the 1-

x(FeCoMn)-xNi composition spread layer/ FeCoMn stack structure is 0.5 nm. Alternating between the three deposition steps created composition-spread region with a width of 18.5 mm. The total film thickness in the composition-spread region was set to 100 nm. The composition spread was confirmed by an X-ray fluorescence spectrometer (XRF: Shimadzu, μEDX-1400) with a measuring spot diameter of 50 μm, as shown in Supplementary Figure 5.

The crystal structure was identified by X-ray diffraction (XRD). An XRD system with a 5-kW rotating anode Cu target x-ray source and a high-resolution 2D-detector (BRUKER AXS, D8 Discover Super Speed with GADDS) was used to determine the crystal structure. The 2D-detector system can detect part of the Debye–Scherrer ring rapidly and two-dimensionally[48]

In the evaluation of the phase separation temperature and magnetization properties of the other FeCoMn-X compositions, we found that the phase separation and inflection point were observed near 400°C. Therefore, we set the annealing temperature as 400°C. In the reported experiment, the annealing was performed at only 400°C; however, for FeCoMnNi, structural changes at higher temperatures are expected and are currently under investigation. Supplementary Figure 7 shows the XRD patterns of the sample as deposited and annealed. The BCC phase was confirmed for the annealed thin film sample at the equiatomical composition of FeCoMnNi (x=0.25). Even at room temperature, a weak peak of the BCC can be observed for the FeCoM-rich composition.

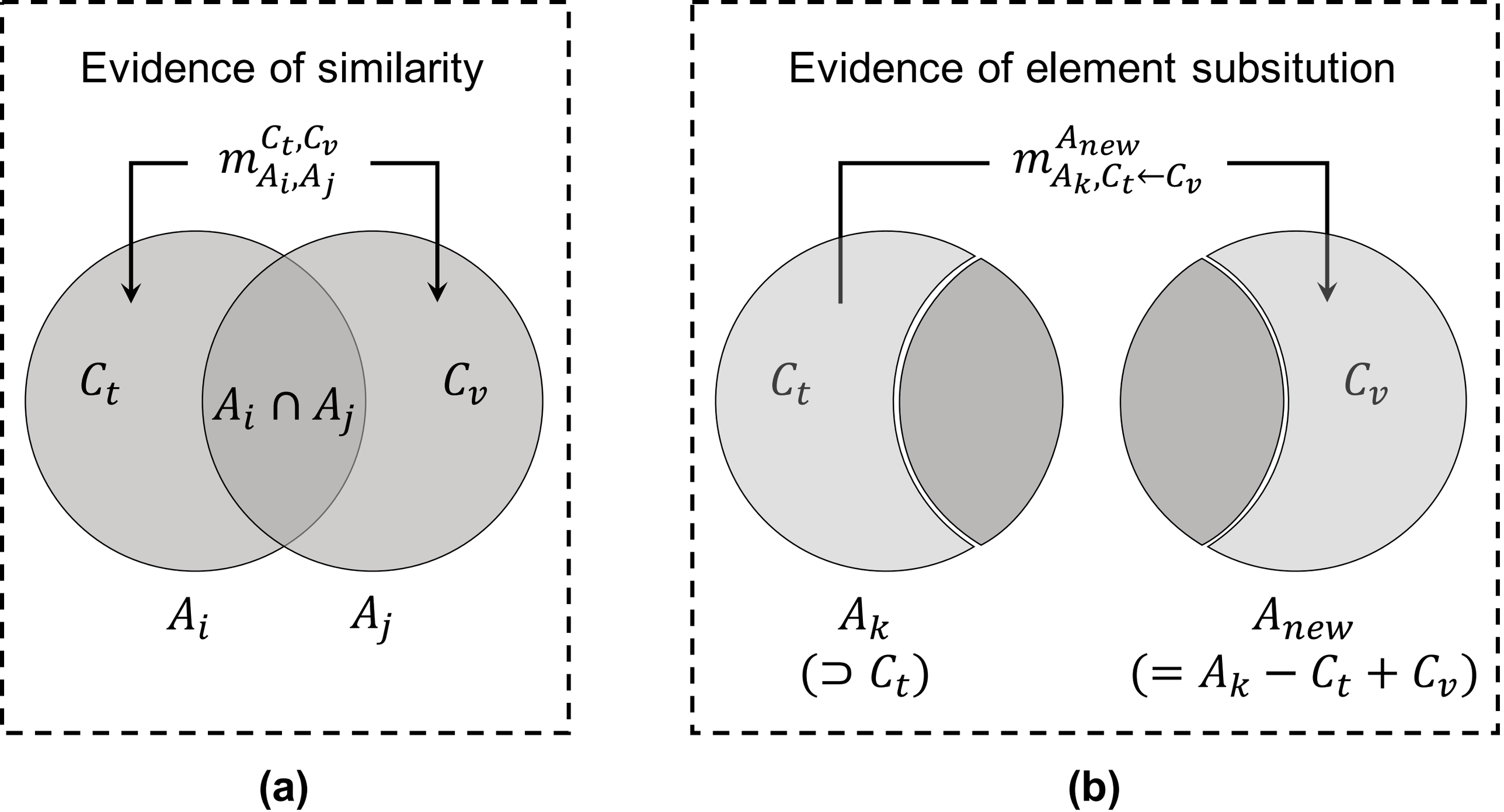
DATA AVAILABILITY

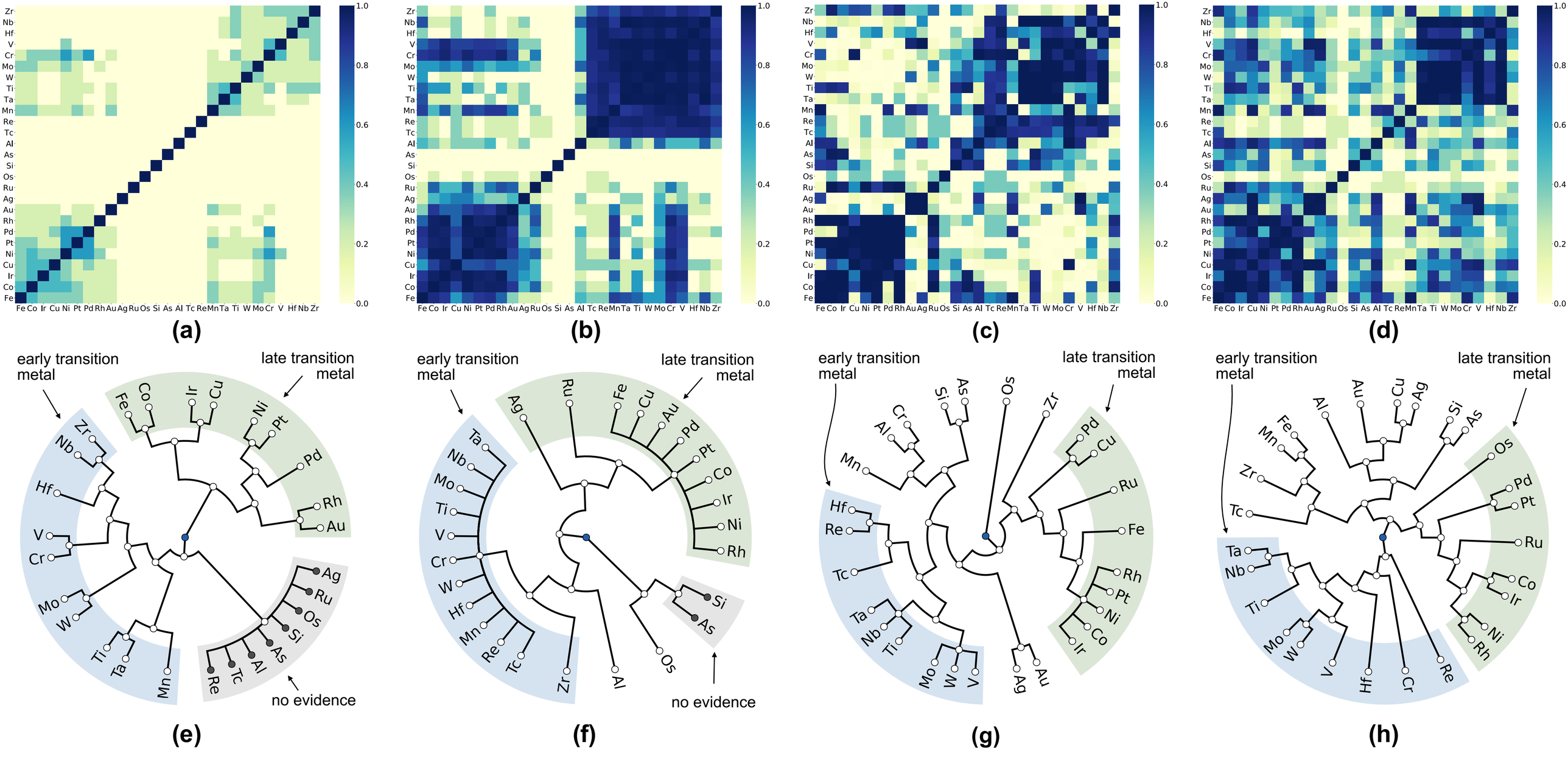
Supporting data for all data plotted in Figures 1–5 (as well as Supplementary Figures 1–7) are available as source data in spreadsheets and the Supplementary Information, respectively. Data sets related to this article are deposited to Zenodo repository[49].

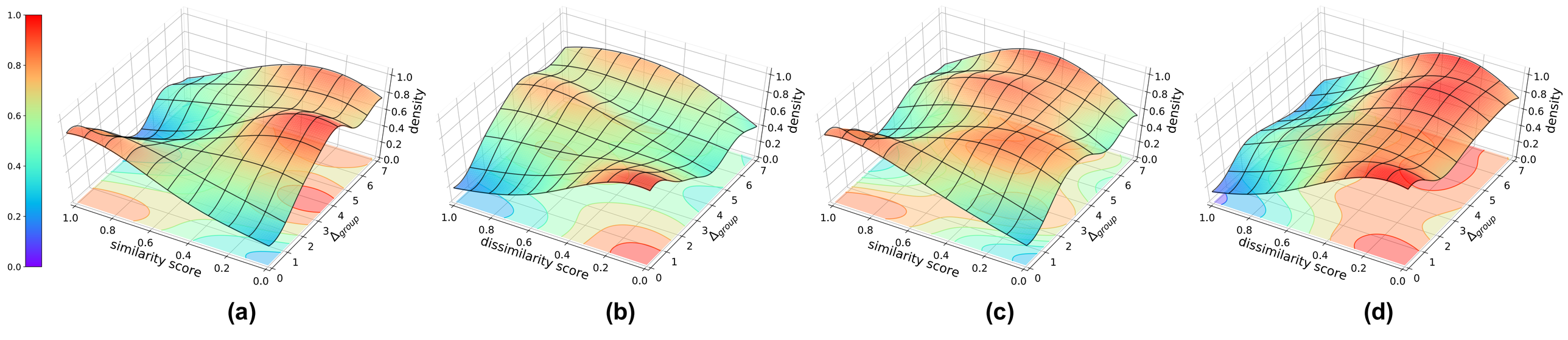
CODE AVAILABILITY

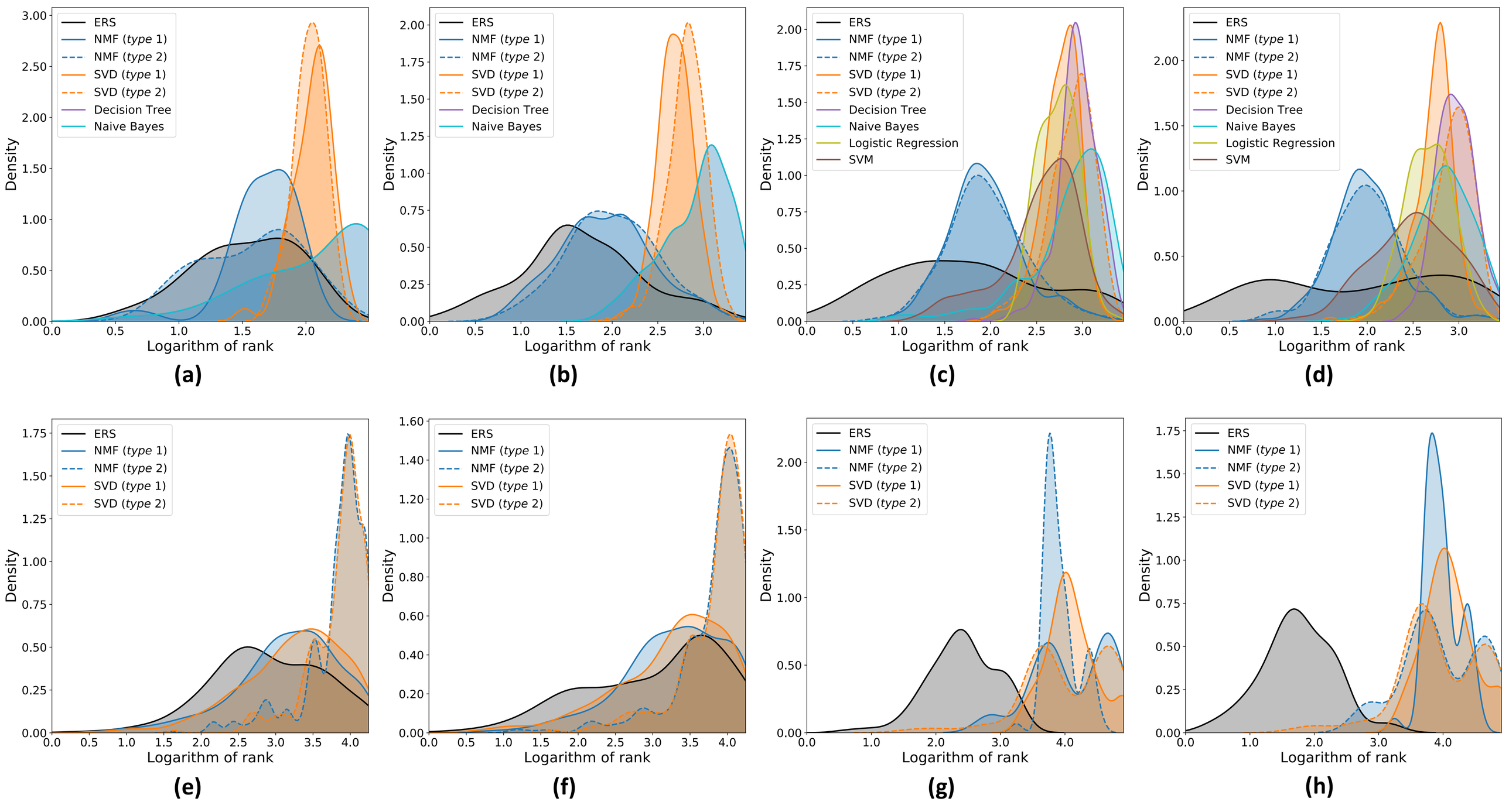
The full code, along with 1) a basic example to show the usage of commender system, 2) an example to illustrate the similarity measurement and 3) an example to explain the method to evaluate the recommender system using an experiment with k-folds cross-validation, have

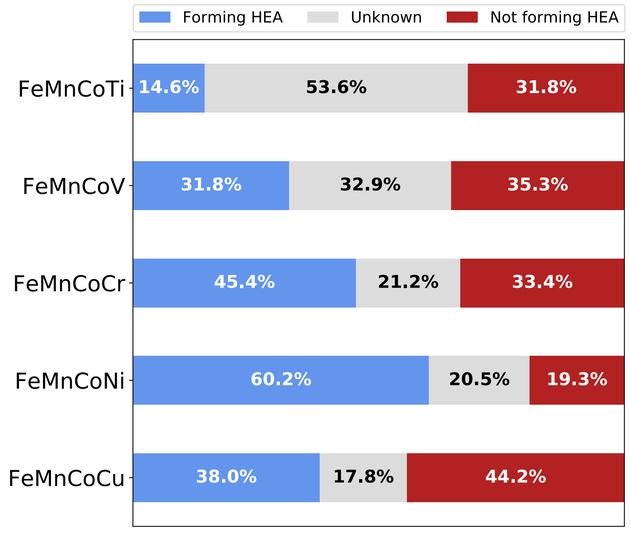
been deposited to Code Ocean repository[50].


Evidence of similarity
$m_{A_i,A_j}^{C_t,C_v}$
C_t
$A_i \cap A_j$
C_v
A_i
A_j
(a)
Evidence of element subsitution
$m_{A_k,C_t \leftarrow C_v}^{A_{new}}$
C_t
C_v
A_k
$(\supset C_t)$
A_{new}
$(= A_k - C_t + C_v)$
(b)

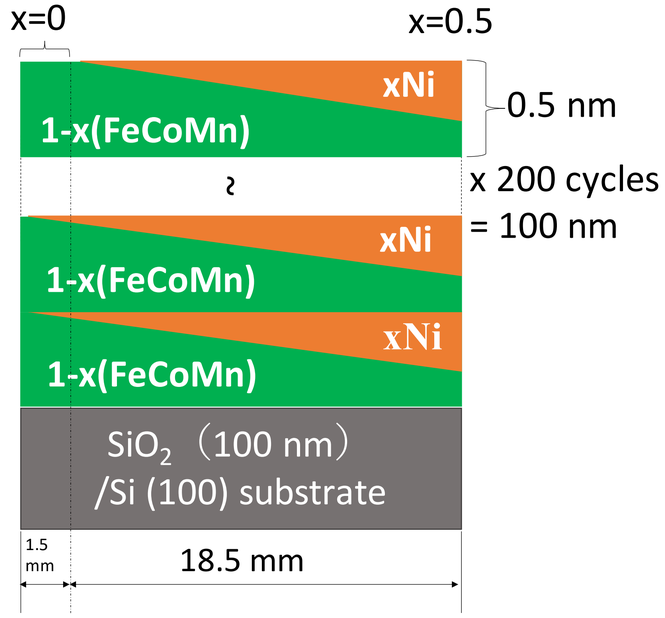

Zr Nb Hf V Cr Mo W Ti Ta Mn Re Tc Al As Si Os Ru Ag Au Rh Pd Pt Ni Cu Ir Co Fe
Fe Co Ir Cu Ni Pt Pd Rh Au Ag Ru Os Si As Al Tc Re Mn Ta Ti W Mo Cr V Hf Nb Zr
1.0 0.8 0.6 0.4 0.2 0.0
(a)
(b)
(c)
(d)
early transition metal
late transition metal
no evidence
(e)
(f)
(g)
(h)


1.0
0.8
0.6
0.4
0.2
0.0
density
similarity score
dissimilarity score
Δ_{group}
0
1
2
3
4
5
6
7
(a)
(b)
(c)
(d)

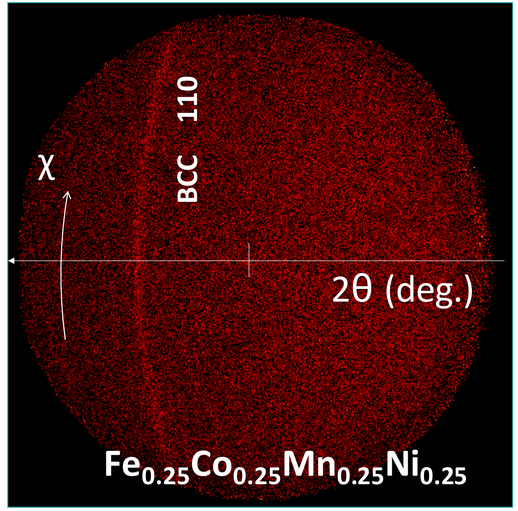

ERS
NMF (type 1)
NMF (type 2)
SVD (type 1)
SVD (type 2)
Decision Tree
Naive Bayes
Logistic Regression
SVM
Density
Logarithm of rank
(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)


Forming HEA
Unknown
Not forming HEA
FeMnCoTi
14.6%
53.6%
31.8%
FeMnCoV
31.8%
32.9%
35.3%
FeMnCoCr
45.4%
21.2%
33.4%
FeMnCoNi
60.2%
20.5%
19.3%
FeMnCoCu
38.0%
17.8%
44.2%

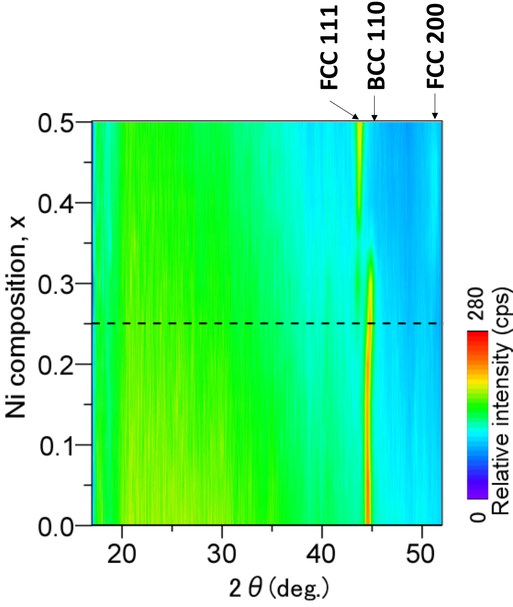
(a)


x=0
x=0.5
xNi
1-x(FeCoMn)
0.5 nm
x 200 cycles
= 100 nm
xNi
1-x(FeCoMn)
xNi
1-x(FeCoMn)
SiO2 （100 nm）
/Si (100) substrate
1.5 mm
18.5 mm

(b)


χ
BCC 110
2θ (deg.)
Fe0.25Co0.25Mn0.25Ni0.25

(c)


FCC 111
BCC 110
FCC 200
Ni composition, x
0.0
0.1
0.2
0.3
0.4
0.5
20
30
40
50
2 θ (deg.)
0
280
Relative intensity (cps)

(d)

Supplementary Information

Minh-Quyet Ha,[1] Duong-Nguyen Nguyen,[1] Viet-Cuong Nguyen,[2] Takahiro Nagata,[3] Toyohiro Chikyow,[4] Hiori Kino,[4] Takashi Miyake,[5] Thierry Denœux,[6] Van-Nam Huynh,[1] and Hieu-Chi Dam[1]

[1)] *Japan Advanced Institute of Science and Technology, 1-1 Asahidai, Nomi, Ishikawa 923-1292, Japan*

[2)] *HPC SYSTEMS Inc., Minato, Tokyo 108-0022, Japan*

[3)] *RCFM, National Institute for Materials Science, 1-2-1 Sengen, Tsukuba, Ibaraki 305-0044, Japan*

[4)] *MaDIS, National Institute for Materials Science, 1-2-1 Sengen, Tsukuba, Ibaraki 305-0044, Japan*

[5)] *CD-FMat, AIST, 1-1-1 Umezono, Tsukuba, Ibaraki 305-8568, Japan*

[6)] *Université de technologie de Compiègne, CNRS, UMR 7253 Heudiasyc, Compiègne, France*

(*Electronic mail: dam@jaist.ac.jp)

(Dated: 5 June 2021)

I. ILLUSTRATIVE EXAMPLES

The following examples provide explanations of how the evidence theory work to learn the similarity and infer the HEA formation for new element combinations, identifying equiatomic alloys.

Example 1: Suppose we have collected four pairs of alloys from experiments. Three of those pairs are alloys that both form HEA phase: $pair_1 = (\{A^1, B^1, C^1, D\}, \{A^1, B^1, C^1, E\})$; $pair_2 = (\{A^2, B^2, C^2, D\}, \{A^2, B^2, C^2, E\})$; and $pair_3 = (\{A^3, B^3, C^3, D\}, \{A^3, B^3, C^3, E\})$. The fourth pair $pair_4 = (\{A^4, B^4, C^4, D\}, \{A^4, B^4, C^4, E\})$ is different from the other three, in which $\{A^4, B^4, C^4, D\}$ forms HEA phase while $\{A^4, B^4, C^4, E\}$ does not form HEA phase. We consider each pair as a source of evidence support that $\{D\}$ is similar to $\{E\}$ in term of substitutability to form the HEA phase. Each evidence is modeled using mass function as follows:

$$
\begin{aligned}
&m_{pair_1}^{\{C\},\{D\}}(\{similar\}) = 0.1,\\
&m_{pair_1}^{\{C\},\{D\}}(\{dissimilar\}) = 0,\\
&m_{pair_1}^{\{C\},\{D\}}(\{similar, dissimilar\}) = 0.9
\end{aligned}
$$

$$
\begin{aligned}
&m_{pair_2}^{\{C\},\{D\}}(\{similar\}) = 0.1,\\
&m_{pair_2}^{\{C\},\{D\}}(\{dissimilar\}) = 0,\\
&m_{pair_2}^{\{C\},\{D\}}(\{similar, dissimilar\}) = 0.9
\end{aligned}
$$

$$
\begin{aligned}
&m_{pair_3}^{\{C\},\{D\}}(\{similar\}) = 0.1,\\
&m_{pair_3}^{\{C\},\{D\}}(\{dissimilar\}) = 0,\\
&m_{pair_3}^{\{C\},\{D\}}(\{similar, dissimilar\}) = 0.9
\end{aligned}
$$

$$
\begin{aligned}
&m_{pair_4}^{\{C\},\{D\}}(\{similar\}) = 0,\\
&m_{pair_4}^{\{C\},\{D\}}(\{dissimilar\}) = 0.1,\\
&m_{pair_4}^{\{C\},\{D\}}(\{similar, dissimilar\}) = 0.9
\end{aligned}
$$

The three pieces of evidence are combined using the Dempster' rule of combination to accumulate the believe that $\{D\}$ is similar to $\{E\}$:

$$m^{\{C\},\{D\}}(\{similar\}) = 0.25,$$
$$m^{\{C\},\{D\}}(\{dissimilar\}) = 0.075,$$
$$m^{\{C\},\{D\}}(\{similar, dissimilar\}) = 0.675$$

Next, if we observed (included in the data) that the HEA phase exists for alloy $\{G, H, I, D\}$, the ERS (which focuses on finding some chance for discovering new combination of elements that the HEA phase exist and ignores the belief regarding $\neg HEA$) will consider that there is some believe that the HEA phase also exists for $\{G, H, I, E\}$ (by substituting $\{D\}$ with $\{E\}$). The evidence is modeled using mass function as follows:

$$m^{\{G,H,I,E\}}_{\{G,H,I,D\},\{D\}\leftarrow\{E\}}(\{\neg HEA\}) = 0,$$
$$m^{\{G,H,I,E\}}_{\{G,H,I,D\},\{D\}\leftarrow\{E\}}(\{HEA\}) = m^{C,D}(\{similar\}) = 0.25,$$
$$m^{\{G,H,I,E\}}_{\{G,H,I,D\},\{D\}\leftarrow\{E\}}(\{HEA, \neg HEA\}) = 1 - m^{C,D}(\{similar\}) = 0.75$$

Example 2: In a same manner but for an extrapolative recommendation: if the HEA phases exist for all the alloys in the three following pairs: $pair_1 = (\{A^1, B^1, C\}, \{A^1, B^1, D, E\})$, $pair_2 = (\{A^2, B^2, C\}, \{A^2, B^2, D, E\})$, $pair_3 = (\{A^3, B^3, C\}, \{A^3, B^3, D, E\})$. In the fourth pair $pair_4 = (\{A^4, B^4, C\}, \{A^4, B^4, D, E\})$, $\{A^4, B^4, C\}$ forms HEA phase while $\{A^4, B^4, D, E\}$ does not form HEA phase. The algorithm will accumulate the believe that $\{C\}$ is similar to $\{D, E\}$ as follows:

$$m^{\{C\},\{D,E\}}(\{similar\}) = 0.25,$$
$$m^{\{C\},\{D,E\}}(\{dissimilar\}) = 0.075,$$
$$m^{\{C\},\{D,E\}}(\{similar, dissimilar\}) = 0.675$$

Consequently, if we observed (included in the data) that the HEA phase exists for $\{G, H, I, C\}$, the algorithm (which focuses on finding some chance for discovering new combination of elements that the HEA phase exist and ignores the belief regarding $\neg HEA$) will consider that there is some believe that the HEA phase also exists for $\{G, H, I, D, E\}$ (by substituting $\{C\}$ with $\{D, E\}$).

$$m^{\{G,H,I,D,E\}}_{\{G,H,I,C\},\{C\}\leftarrow\{D,E\}}(\{\neg HEA\}) = 0,$$
$$m^{\{G,H,I,D,E\}}_{\{G,H,I,C\},\{C\}\leftarrow\{D,E\}}(\{HEA\}) = m^{C,D}(\{similar\}) = 0.25,$$
$$m^{\{G,H,I,D,E\}}_{\{G,H,I,C\},\{C\}\leftarrow\{D,E\}}(\{HEA, \neg HEA\}) = 1 - m^{C,D}(\{similar\}) = 0.75$$

II. ALLOYS DATA SETS

In the evaluation experiments, we use eight data sets consisting of binary, ternary, quaternary, and quinary alloys comprising multiple equiatomically combined elements. The data sets consist of data from experiments and calculations. In this section, we will follow Ref. 35 to describe the data sets. The alloys contained in the data sets comprise $\mathcal{E} = \{$ Fe, Co, Ir, Cu, Ni, Pt, Pd, Rh, Au, Ag, Ru, Os, Si, As, Al, Tc, Re, Mn, Ta, Ti, W, Mo, Cr, V, Hf, Nb, and Zr$\}$. Supplementary Figure 2 shows the proportion of 27 elements in the data sets. Any alloy contained in the following data sets is considered as an HEA if its order-disorder transition temperature is below its melting temperature.

- $\mathcal{D}_{\text{ASMI16}}$: The order-disorder transition temperatures (T_c^{exp}) and melting temperatures (T_m^{exp}) of the alloys are both experimentally evaluated[1]. All of the alloys contained in $\mathcal{D}_{\text{ASMI16}}$ show an order-disorder transition temperature below their melting temperature ($T_c^{\text{exp}} < T_m^{\text{exp}}$).

- $\mathcal{D}_{\text{CALPHAD}}$: The order-disorder transition temperatures ($T_c^{\star}$) and melting temperatures ($T_m^{\star}$) of the alloys are both predicted using calculated-phase-diagram (CALPHAD) calculations[2–4] based on the temperatures for some binary alloys (three possible for each ternary alloy) found in the Thermo-Calc software SSOL5 database[5]. Similar to the $\mathcal{D}_{\text{ASMI16}}$ data set, the $\mathcal{D}_{\text{CALPHAD}}$ data set only contains the alloys satisfying $T_c^{\star} < T_m^{\star}$.

- $\mathcal{D}_{\text{AFLOW}}$, $\mathcal{D}_{\text{AFLOW}}^{\text{quaternary}}$, and $\mathcal{D}_{\text{AFLOW}}^{\text{quinary}}$: The order-disorder transition temperatures (T_c^{AFLOW}) of the alloys contained in these data sets are estimated using the automatic flow (AFLOW) convex-hull database[6]. The melting temperatures T_m^{exp} and $T_m^{\star}$ are applied to the binary and ternary alloys, respectively. The alloy is considered as an HEA if $T_c^{\text{AFLOW}} < T_m^{\text{exp}}$ for binary alloys and $T_c^{\text{AFLOW}} < T_m^{\star}$ for ternary, quaternary, and quinary alloys).

- $\mathcal{D}_{\text{LTVC}}$, $\mathcal{D}_{\text{LTVC}}^{\text{quaternary}}$, and $\mathcal{D}_{\text{LTVC}}^{\text{quinary}}$: These data sets contain the same alloys as those contained in data sets $\mathcal{D}_{\text{AFLOW}}$, $\mathcal{D}_{\text{AFLOW}}^{\text{quaternary}}$, and $\mathcal{D}_{\text{AFLOW}}^{\text{quinary}}$, respectively. However, the properties of the alloys contained in these data sets are predicted using the method of Lederer, Toher, Vecchio, and Curtarolo (LTVC)[7]. *Ab-initio* calculations are used

to estimate the order-disorder transition temperatures (T_c^{LTVC}) of the alloys contained in these data sets. In addition, the T_m values are the same as those of the alloys contained in the AFLOW data sets. Any alloy in these data sets is predicted as an HEA if $T_c^{\mathrm{LTVC}} < T_m^{\mathrm{exp}}$ for binary alloys and $T_c^{\mathrm{LTVC}} < T_m^{\star}$ for ternary, quaternary, and quinary alloys

Note that $\mathcal{D}_{\mathrm{ASMI16}}$ and $\mathcal{D}_{\mathrm{CALPHAD}}$ only contain confirmed and predicted HEAs, respectively. Therefore, although we assume that the properties of all the other binary or ternary alloys (not included in the data set) have not yet been confirmed, we do not assume that those alloys are not HEAs.

III. DIFFERENCES BETWEEN SIMILARITY MATRICES LEARNED FROM $\mathcal{D}_{\mathrm{CALPHAD}}$ AND $\mathcal{D}_{\mathrm{AFLOW}}$

There are some notable differences between these results obtained from experiments with $\mathcal{D}_{\mathrm{CALPHAD}}$ and $\mathcal{D}_{\mathrm{AFLOW}}$. The similarity matrix learned from $\mathcal{D}_{\mathrm{AFLOW}}$ shows that Au and Ag are very similar (Supplementary Figure 3 b). Furthermore, both are similar to V, Mn, and Al but not to other late transition metals (Supplementary Figure 3 a). Mn is also similar to Tc, Re, and Cr but not to the other early transition metals. However, Tc and Re are somewhat similar to the other early transition metals. Furthermore, Zr is somewhat similar to the late transition metals, but different from the early transition metals. Clearly, these results are different from that obtained from $\mathcal{D}_{\mathrm{CALPHAD}}$ owing to the difference between the predicted label (HEA or $\neg HEA$) for the Zr-containing alloys recommended based on CALPHAD and AFLOW calculations, as listed in Supplementary Table 1. Al, Si, and As are all similar to each other and to Fe and Co (Supplementary Figure 3 a). However, Al is similar to V, Cr, and Mn but not to Ti, whereas Si and As are very similar to Ti but not to V or Cr.

IV. MONITORING HEA RECALL RATIOS IN TEST SET

A. Evaluation of HEA-recommendation capability by cross-validation

In the experiment with $\mathcal{D}_{\text{ASMI16}}$, the result shows that the ERS can significantly reduce the number of trials required to recall all the HEAs in the test set compared to the competitor systems (Supplementary Figure 4 a). The proposed ERS requires less than 12, 25, and 80% of all the possible trials to recall one-half, three-quarters, and all the HEAs in the test set, respectively (Supplementary Table 2). In the $\mathcal{D}_{\text{CALPHAD}}$ experiment, the ERS requires less than 2 and 5% of all the possible trials to recall one-half and three-quarters of the HEAs in the test set, respectively, which are the fewest trials required among all the recommender systems (Supplementary Figure 4 b and Supplementary Table 2). Interestingly, in the $\mathcal{D}_{\text{ASMI16}}$ and $\mathcal{D}_{\text{CALPHAD}}$ experiments, the supervised-method-based recommender systems either approximately randomly selected possible HEAs (Naïve Bayes and decision tree) or could not rank any (logistic regression and SVM) at all because these data sets contain only positively labeled HEAs.

The result in $\mathcal{D}_{\text{AFLOW}}$ experiment demonstrates that the ERS also outperforms the competitor systems in recalling one-half of the HEAs in the test set. However, the ERS cannot reliably recall the one-quarter of the HEAs remaining in the test set because not enough evidence is available in the training data to make inferences about the remaining HEAs (Supplementary Figure 4 c and Supplementary Table 2). The $\mathcal{D}_{\text{LTVC}}$ and $\mathcal{D}_{\text{AFLOW}}$ experimental results are identical(Supplementary Figure 4 d). Although the ERS performs better than the other recommendation systems in recovering one-half of the test HEAs in the $\mathcal{D}_{\text{LTVC}}$ data set (requiring only less than 3% of the number of possible trials), it cannot reliably recover the remaining one-quarter of the test HEAs owing to the lack of evidence in the training data (Supplementary Table 2).

B. Evaluation of HEA-recommendation capability by extrapolation

In the $\mathcal{D}^{\text{quaternary}}_{\text{AFLOW}}$ experiment, the ERS performs significantly better than the NMF-based recommender system, requiring less than 5 and 19% of the total number of possible HEA candidates to recall 50 and 75% of the HEAs in the test set, respectively (Supplementary Table 3). In the $\mathcal{D}^{\text{quaternary}}_{\text{LTVC}}$ experiment, the ERS and competitor matrix-based system devel-

oped using the first type of matrix representation require 13 and 32% and 14 and 41% of the total number of possible HEA candidates to recall 50 and 75% of the HEAs in the test set, respectively (Supplementary Table 3). Further investigation indicates that the ERS hardly recommends any quaternary alloys in $\mathcal{D}_{\mathrm{LTVC}}^{\mathrm{quaternary}}$ because these alloys cannot be generated by substituting elements in any of the ternary alloys in $\mathcal{D}_{\mathrm{LTVC}}$ (Supplementary Table 4). Therefore, the properties of these alloys cannot be inferred from the evidence collected from $\mathcal{D}_{\mathrm{LTVC}}$. As a result, the rankings obtained for these alloys are significantly low; therefore, the HEA recall rate is even lower than those obtained for randomly recommended HEAs. The results obtained for $\mathcal{D}_{\mathrm{LTVC}}^{\mathrm{quinary}}$ and $\mathcal{D}_{\mathrm{AFLOW}}^{\mathrm{quinary}}$ both show that the ERS drastically outperforms the capability of the competitor systems for recommending quinary HEAs. To recall 50, 75, and 100% of the HEAs from these data sets, 10–100 times fewer trials are required using the ERS than are required using the matrix-based recommender systems (Supplementary Table 3).

Supplementary Table 1. Comparison of the properties of alloys containing Zr in the two datasets $\mathcal{D}_{\text{ASMI16}}$ and $\mathcal{D}_{\text{CALPHAD}}$ to those predicted in $\mathcal{D}_{\text{LTVC}}$ and $\mathcal{D}_{\text{AFLOW}}$.

	$\mathcal{D}_{\text{ASMI16}}$ 4 alloys	$\mathcal{D}_{\text{CALPHAD}}$ 19 alloys
#in agreement with $\mathcal{D}_{\text{AFLOW}}$	4	10
#disagreement with $\mathcal{D}_{\text{AFLOW}}$	0	9
#in agreement with $\mathcal{D}_{\text{LTVC}}$	3	10
#disagreement with $\mathcal{D}_{\text{LTVC}}$	1	9

Supplementary Table 2. Ratio of number of trials (out of total number of possible trials) required to recall 50, 75, and 100% of HEAs in test set.

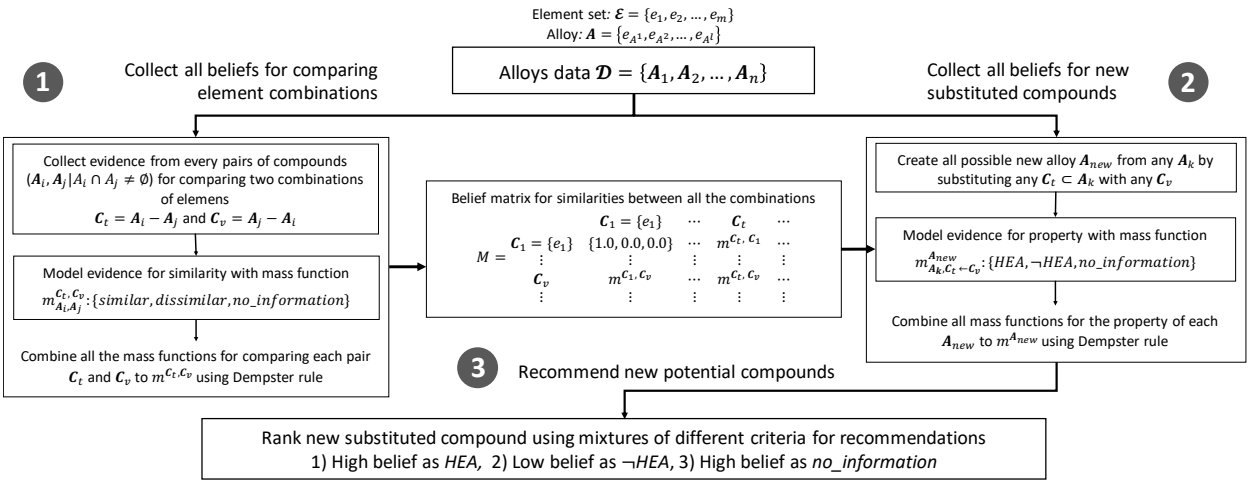
Data set	Model	Recall rates		
		Half	Three-quarters	Full
$\mathcal{D}_{\text{ASMI16}}$	ERS	**12%**	**25%**	**80%**
	NMF (type 1)	16%	25%	92%
	NMF (type 2)	13%	26%	98%
	SVD (type 1)	31%	68%	99%
	SVD (type 2)	23%	64%	99%
	Decision Tree	77%	90%	99%
	Naïve Bayes	77%	90%	99%
	Logistic Regression	-	-	-
	SVM	-	-	-
$\mathcal{D}_{\text{CALPHAD}}$	ERS	**2%**	**5%**	92%
	NMF (type 1)	3%	7%	**89%**
	NMF (type 2)	3%	8%	93%
	SVD (type 1)	14%	28%	94%
	SVD (type 2)	17	37%	93%
	Decision Tree	39%	52%	94%
	Naïve Bayes	39%	52%	94%
	Logistic Regression	-	-	-
	SVM	-	-	-
$\mathcal{D}_{\text{AFLOW}}$	ERS	**2%**	8%	97%
	NMF (type 1)	3%	**6%**	96%
	NMF (type 2)	3%	6%	**85%**
	SVD (type 1)	16%	35%	99%
	SVD (type 2)	20%	50%	99%
	Decision Tree	31%	51%	99%
	Naïve Bayes	33%	53%	99%
	Logistic Regression	20%	29%	93%
	SVM	15%	26%	99%
$\mathcal{D}_{\text{LTVC}}$	ERS	**3%**	23%	97 %
	NMF (type 1)	4%	**6%**	96%
	NMF (type 2)	4%	7%	**86%**
	SVD (type 1)	14%	33%	99%
	SVD (type 2)	19%	52%	99%
	Decision Tree	32%	48%	99%
	Naïve Bayes	26%	42%	99%
	Logistic Regression	17%	26%	89%
	SVM	12%	26%	99%

Supplementary Table 3. Ratio of number of trials (out of total number of possible trials) required to recall 50, 75, and 100% of HEAs in test set by extrapolating HEA-recommendation capability.

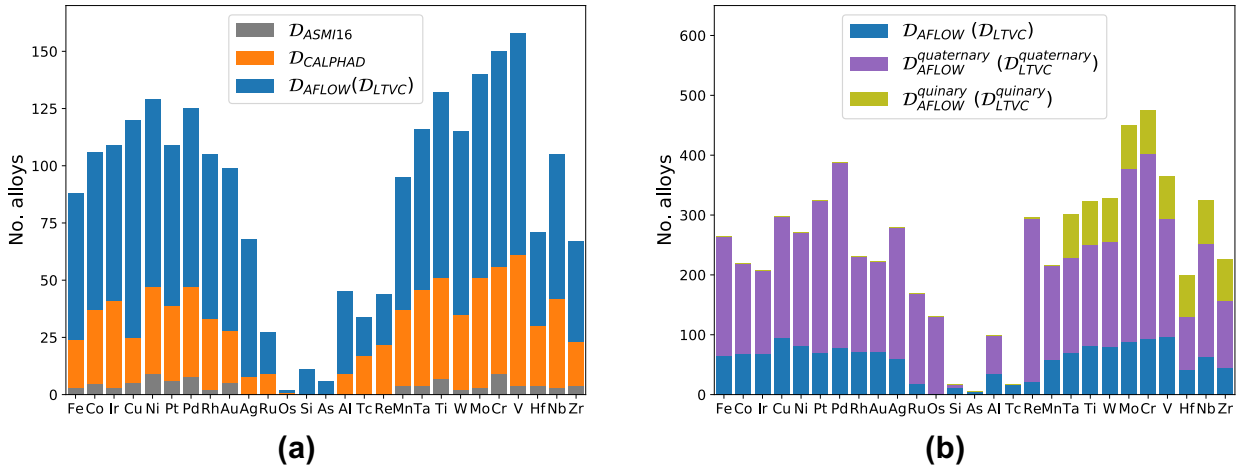
Data set	Model	Recall rates		
		Half	Three-quarters	Full
$\mathcal{D}_{\text{AFLOW}}^{\text{quaternary}}$	ERS	**5%**	**19%**	**99%**
	NMF (type 1)	10%	24%	99%
	NMF (type 2)	50%	67%	99%
	SVD (type 1)	13%	32%	99%
	SVD (type 2)	53%	67%	99%
$\mathcal{D}_{\text{AFLOW}}^{\text{quinary}}$	ERS	**0.4%**	**1%**	**3%**
	NMF (type 1)	10%	56%	98%
	NMF (type 2)	9%	14%	47%
	SVD (type 1)	15%	27%	99%
	SVD (type 2)	8%	57%	99%
$\mathcal{D}_{\text{LTVC}}^{\text{quaternary}}$	ERS	**13%**	**32%**	**99%**
	NMF (type 1)	14%	41%	99%
	NMF (type 2)	50%	71%	99%
	SVD (type 1)	15%	39%	99%
	SVD (type 2)	53%	71%	99%
$\mathcal{D}_{\text{LTVC}}^{\text{quinary}}$	ERS	**0.07%**	**0.2%**	**2%**
	NMF (type 1)	11%	16%	47%
	NMF (type 2)	10%	53%	93%
	SVD (type 1)	15%	27%	99%
	SVD (type 2)	7%	54%	93%

Supplementary Table 4. List of 63 quaternary HEAs in $\mathcal{D}_{\mathrm{LTVC}}^{\mathrm{quaternary}}$ that no evidence about their properties is found.

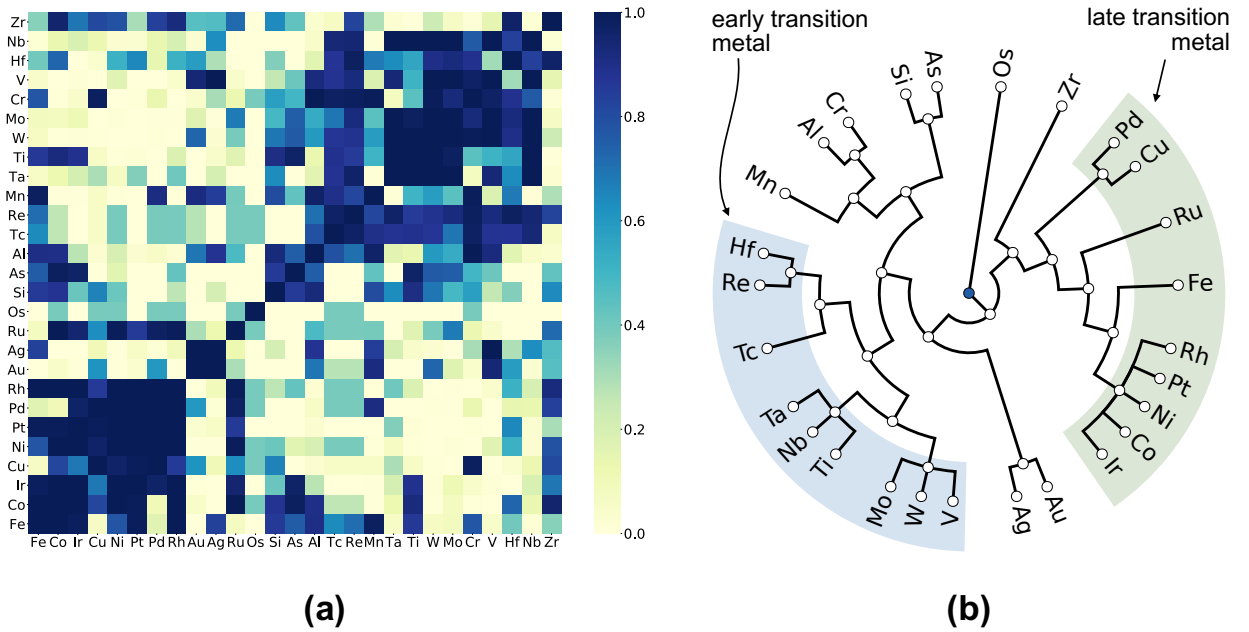
FeAuRePd	AuNiPdOs	NiRePtOs	RhRePtOs	RePtOsAg
FeNiRePd	AuRhRePd	NiPdRuOs	RhPdOsAg	PdRuOsAg
FeMoOsAg	AuRhPdOs	NiPdCuOs	CoRePdRu	PdCuOsCr
FeRhPdOs	AuRePdRu	NiPdOsCr	CoRePdCu	ReCuPtOs
FeRePdRu	AuRePdCu	NiPdOsAg	CoRePdOs	RhRePdAg
FeRePdCu	AuRePdOs	MoRhPdOs	CoRePdAg	NiRePdAg
FeRePdOs	AuRePdAg	MoRePdOs	CoReRuPt	AuNiReAg
FeReRuAg	NiMoPdOs	MoRePtOs	CoRePtOs	ReRuPtOs
FeReOsAg	NiRhRePd	MoReOsAg	CoPdRuOs	RhRePdOs
FePdRuOs	NiRhPdOs	MoPdRuOs	CoRuPtOs	NiRePdOs
FePdCuOs	NiCoRePd	MoRuPtOs	RePdRuOs	FeCuOsAg
FePdOsCr	NiRePdRu	RhCoPdOs	RePdPtOs	
FeRuOsAg	NiRePdCu	RhRePdCu	RePdOsAg	



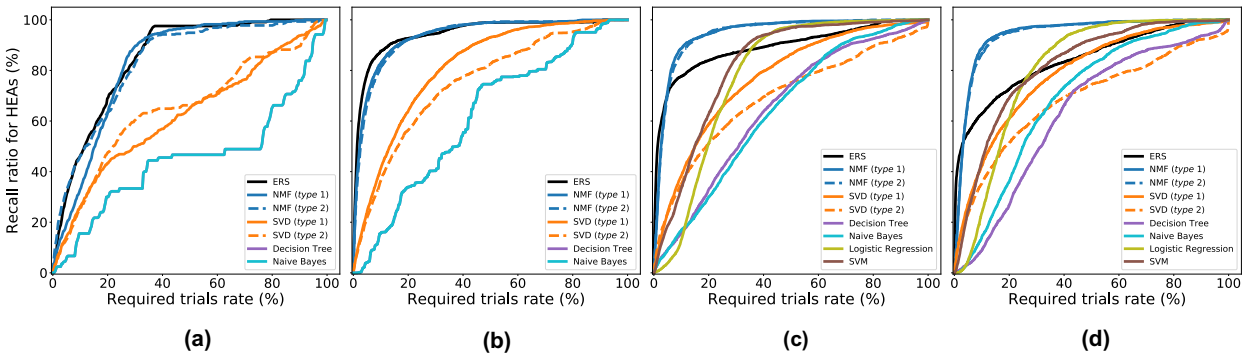
Supplementary Figure 1. Workflow chart illustrating three ERS stages required for recommending undiscovered HEAs. The data set $\mathcal{D}$ includes observed materials generated from a finite set $\mathcal{E}$ of elements.



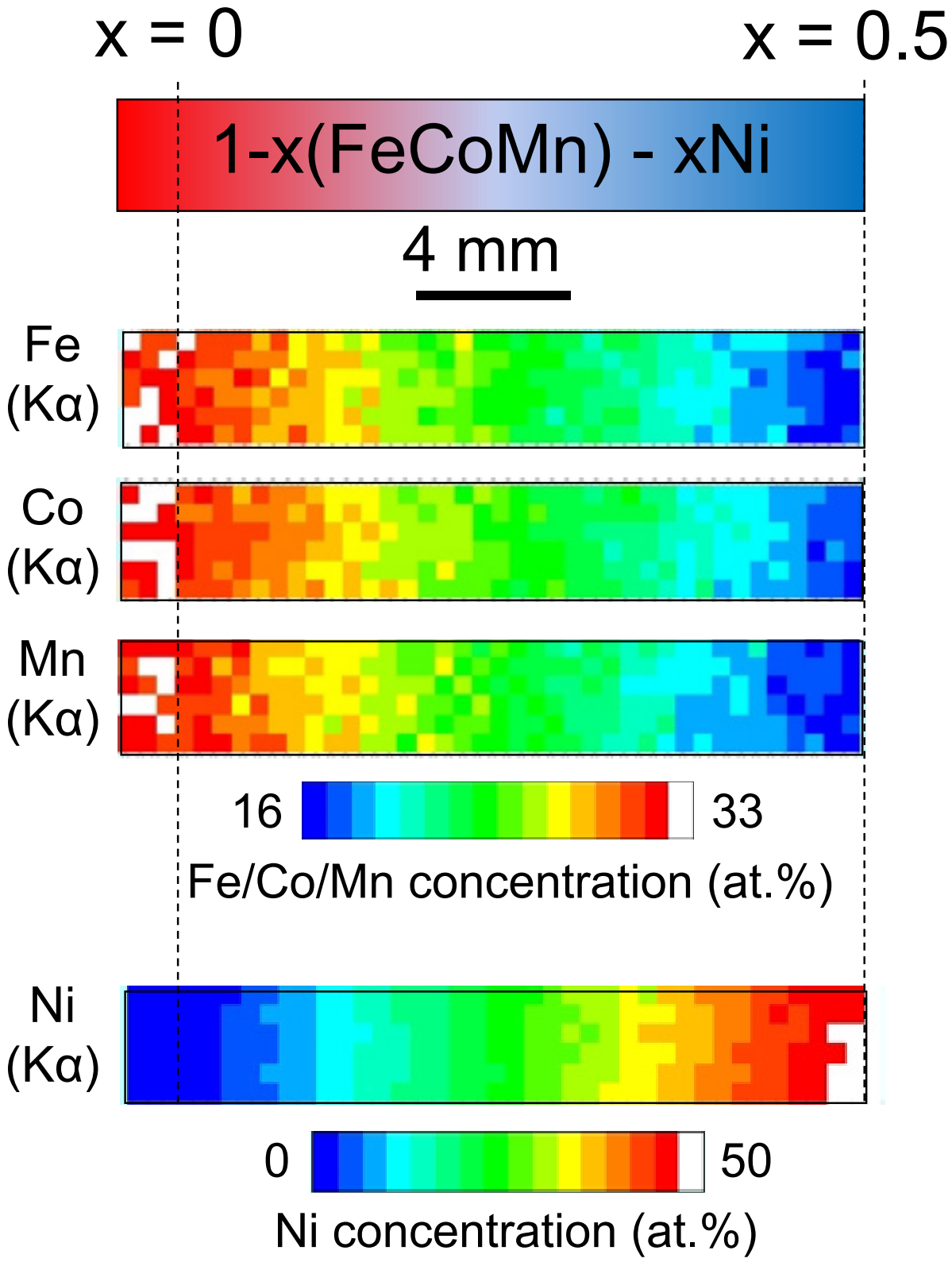
Supplementary Figure 2. Proportions of 27 elements in $\mathcal{D}_{\text{ASMI16}}$, $\mathcal{D}_{\text{CALPHAD}}$, $\mathcal{D}_{\text{AFLOW}}$, $\mathcal{D}^{\text{quaternary}}_{\text{AFLOW}}$, $\mathcal{D}^{\text{quinary}}_{\text{AFLOW}}$, $\mathcal{D}_{\text{LTVC}}$, $\mathcal{D}^{\text{quaternary}}_{\text{LTVC}}$, and $\mathcal{D}^{\text{quinary}}_{\text{LTVC}}$ data sets.



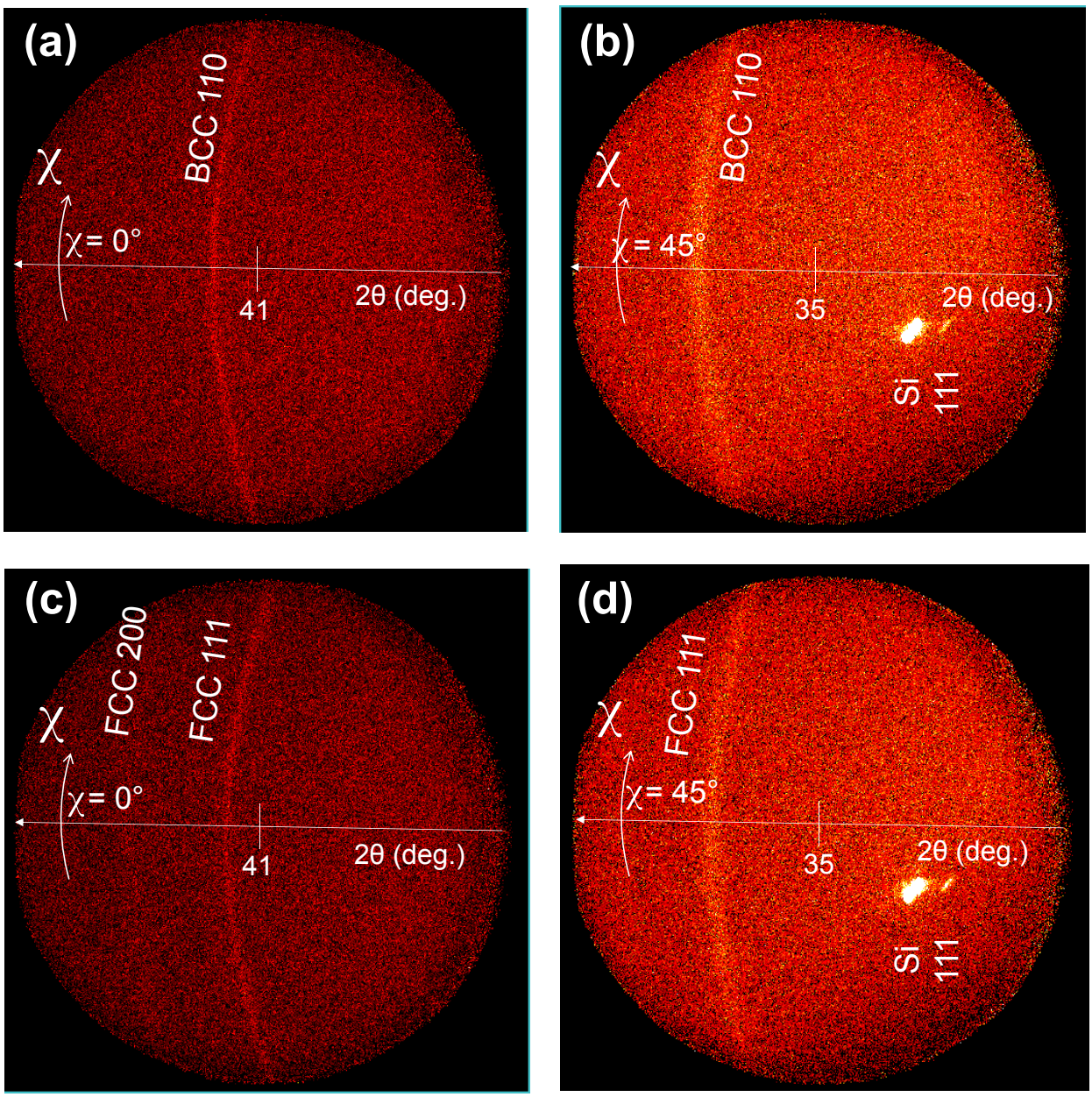
Supplementary Figure 3. (a) Heat maps for similarity matrices among 27 elements $\mathcal{E}$ obtained from$\mathcal{D}_{\text{AFLOW}}$ data set. (b) Hierarchically clustered structure of all elements in $\mathcal{E}$ constructed using the similarity matrix and hierarchical agglomerative clustering. Blue and green regions indicate groups of early and late transition metals, respectively.



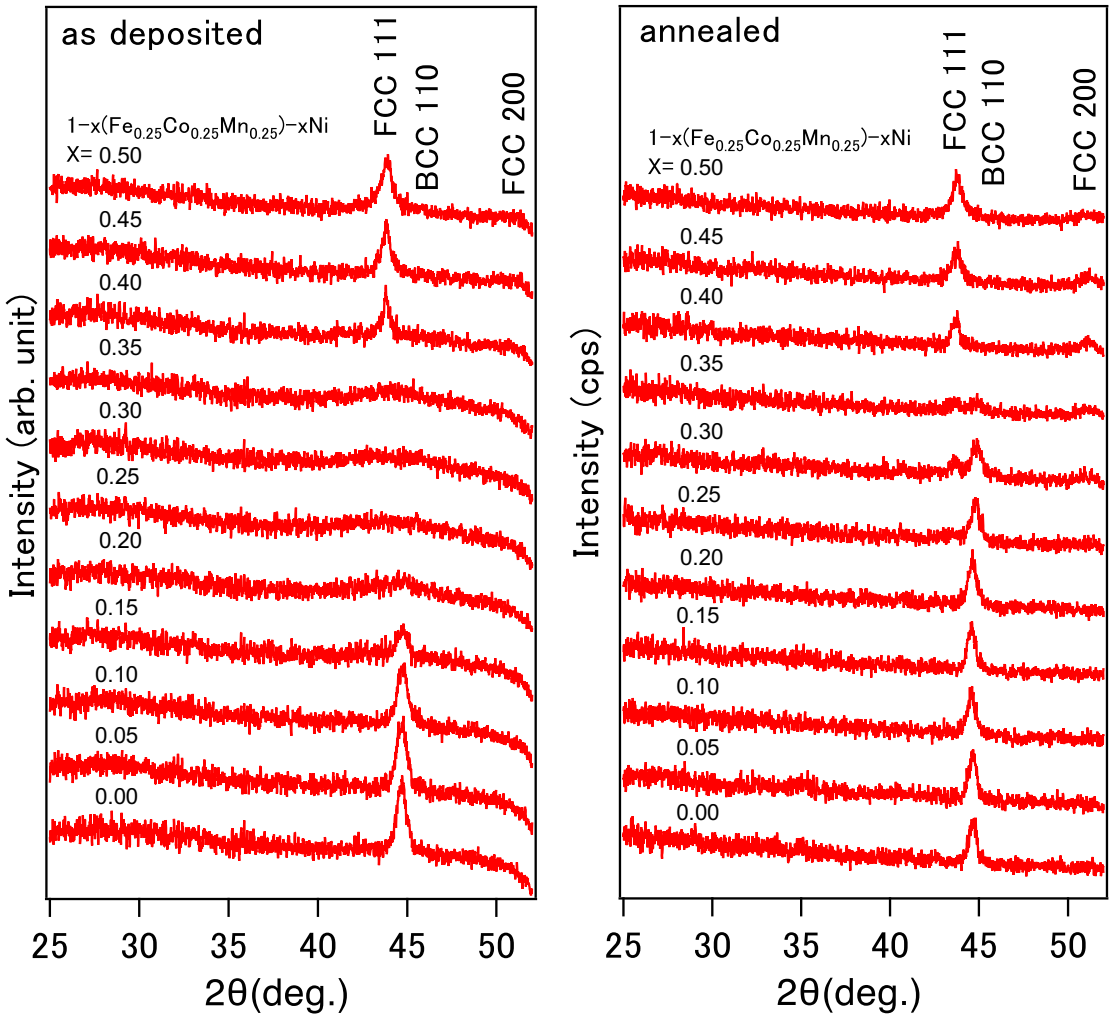
Supplementary Figure 4. Dependence of HEA recall ratio in the test sets on the number of trial required using k-fold cross-validation on (a) $\mathcal{D}_{\text{ASMI16}}$, (b) $\mathcal{D}_{\text{CALPHAD}}$, (c) $\mathcal{D}_{\text{AFLOW}}$, and (d) $\mathcal{D}_{\text{LTVC}}$ data sets.



Supplementary Figure 5. Heatmapping image of Fe, Co, Mn, and Ni concentration estimated by EDX analysis. Composition was estimated from the XRF intensity of balk target materials and single-phase films of FeCoMn and Ni.



Supplementary Figure 6. 2D-XRD images at center γ angles of 0 and 45 ∘of FeCoMn-Ni films with low- (a,b) and high- (c,d) Ni concentrations. According to the powder diffraction pattern data base[8] (PDF 03-065-7519 and PDF 03-065-5131), for BCC, except for the reflection from (110), the signal intensities from other plane are not enough high to detect them in film form. So, the reflection from (110) is only detected. For FCC, in addition to the reflection from (111), the second strongest signal from (200) can barely be detected. The signals do not show no γ angle dependence, meaning the films are polycrystals in disordered crystal orientation.



Supplementary Figure 7. XRD patterns of the as deposited and annealed at 400°C of FeCoMn-Ni film using an XRD system with a 5-kW rotating anode Cu target x-ray source. The BCC phase was confirmed for the annealed thin film sample at the equiatomical composition of FeCoMnNi (x=0.25)

REFERENCES

[1]"Binary Alloy Phase Diagrams," in *Alloy Phase Diagrams*, Vol. 3, edited by H. Okamoto, M. Schlesinger, and E. Mueller (ASM International, 2016).

[2]O. N. Senkov, J. D. Miller, D. B. Miracle, and C. Woodward, "Accelerated exploration of multi-principal element alloys with solid solution phases," Nat. Commun. **6**, 6529 (2015).

[3]D. Alman, "Searching for next single-phase high-entropy alloy compositions," Entropy **15**, 4504–4519 (2013).

[4]F. Zhang, C. Zhang, S. Chen, J. Zhu, W. Cao, and U. Kattner, "An understanding of high entropy alloys from phase diagram calculations," CALPHAD **45**, 1–10 (2014).

[5]J.-O. Andersson, T. Helander, L. Höglund, P. Shi, and B. Sundman, "Thermo-calc & dictra, computational tools for materials science," Calphad **26**, 273–312 (2002).

[6]C. Nyshadham, C. Oses, J. E. Hansen, I. Takeuchi, S. Curtarolo, and G. L. Hart, "A computational high-throughput search for new ternary superalloys," Acta Mater. **122**, 438–447 (2017).

[7]Y. Lederer, C. Toher, K. S. Vecchio, and S. Curtarolo, "The search for high entropy alloys: A high-throughput ab-initio approach," Acta Mater. **159**, 364–383 (2018).

[8]S. Gates-Rector and T. Blanton, "The powder diffraction file: a quality materials characterization database," Powder Diffr. **34**, 352–360 (2019).