Journal of
Applied Physics

Evidence-based data mining method to reveal similarities between materials based on physical mechanisms F

Cite as: J. Appl. Phys. **133**, 053904 (2023); https://doi.org/10.1063/5.0134999
Submitted: 15 November 2022 • Accepted: 17 January 2023 • Published Online: 06 February 2023

Minh-Quyet Ha, Duong-Nguyen Nguyen, Viet-Cuong Nguyen, et al.

COLLECTIONS

Note: This paper is part of the Special Topic on: Multi-Principal Element Materials: Structure, Property, and Processing.

F This paper was selected as Featured

AIP Publishing

J. Appl. Phys. **133**, 053904 (2023); https://doi.org/10.1063/5.0134999 **133**, 053904

Evidence-based data mining method to reveal similarities between materials based on physical mechanisms

Cite as: J. Appl. Phys. **133**, 053904 (2023); doi: 10.1063/5.0134999
Submitted: 15 November 2022 · Accepted: 17 January 2023 ·
Published Online: 6 February 2023

Minh-Quyet Ha,[1] Duong-Nguyen Nguyen,[1] Viet-Cuong Nguyen,[2] Hiori Kino,[3] Yasunobu Ando,[4] Takashi Miyake,[4] Thierry Denœux,[5] Van-Nam Huynh,[1] and Hieu-Chi Dam[1,a)]

AFFILIATIONS

[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 and Services Division of Materials Data and Integrated System, National Institute for Materials Science, 1-2-1 Sengen, Tsukuba, Ibaraki 305-0044, Japan
[4]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
[5]Heudiasyc–UMR CNRS 7253, Université de Technologie de Compiègne, Compiègne, France

Note: This paper is part of the Special Topic on: Multi-Principal Element Materials: Structure, Property, and Processing.

[a)]**Author to whom correspondence should be addressed:** dam@jaist.ac.jp

ABSTRACT

Measuring the similarity between materials is essential for estimating their properties and revealing the associated physical mechanisms. However, current methods for measuring the similarity between materials rely on theoretically derived descriptors and parameters fitted from experimental or computational data, which are often insufficient and biased. Furthermore, outliers and data generated by multiple mechanisms are usually included in the dataset, making the data-driven approach challenging and mathematically complicated. To overcome such issues, we apply the Dempster–Shafer theory to develop an evidential regression-based similarity measurement (eRSM) method, which can rationally transform data into evidence. It then combines such evidence to conclude the similarities between materials, considering their physical properties. To evaluate the eRSM, we used two material datasets, including $3d$ transition metal–$4f$ rare-earth binary and quaternary high-entropy alloys with target properties, Curie temperature, and magnetization. Based on the information obtained on the similarities between the materials, a clustering technique is applied to learn the cluster structures of the materials that facilitate the interpretation of the mechanism. The unsupervised learning experiments demonstrate that the obtained similarities are applicable to detect anomalies and appropriately identify groups of materials whose properties correlate differently with their compositions. Furthermore, significant improvements in the accuracies of the predictions for the Curie temperature and magnetization of the quaternary alloys are obtained by introducing the similarities, with the reduction in mean absolute errors of 36% and 18%, respectively. The results show that the eRSM can adequately measure the similarities and dissimilarities between materials in these datasets with respect to mechanisms of the target properties.

I. INTRODUCTION

The concept of machine learning has great potential for application in several areas of materials science, especially for discovering new materials. In materials science, a number of the problems addressed by data-driven approaches require the effective utilization of existing material data for predicting the properties of new materials and understanding the underlying physicochemical mechanisms.[1]

From an engineering point of view, developing a data-driven model that quickly and accurately predicts the physical properties of possible materials from accumulated data can reduce the time required for material development. By applying a data-driven model to screen materials *in silico*, we narrow down the candidates that require expensive calculations and experiments to verify. If there are sufficient independent supervised data from the distribution of the target material data, a model with high prediction accuracy can be built using state-of-the-art data-driven techniques. However, because materials research and development aim to develop materials that are superior to existing ones, the distribution of the target prediction data may be completely different from the distribution of the original training data. Therefore, there are concerns about whether data-driven models can accurately predict the physical properties of new materials.

On the contrary, considering the history of materials science, researchers have discovered various materials through a loop of hypothesis and verification based on their knowledge, experience, and serendipity. Particularly, hypothesizing relies heavily on describing, interpreting, and understanding the underlying physicochemical mechanisms of the observed physical phenomena of materials. Scientifically, applying a data-driven approach to extracting knowledge from existing complicated material data can accelerate the process of describing, interpreting, and understanding the physicochemical mechanisms underlying the observed physical phenomena of materials. This reduces the time required for material development. Hence, to be effectively applied to materials science, data-driven approaches that are interpretable and understandable to humans must be developed.

One of the most intuitive and interpretable data-driven approaches for humans is analogy-based inductive reasoning, which infers the properties of a new instance using the information of the observed instances that are most similar to it.[2–5] By applying analogy-based models, we can easily explain the reasoning process behind the predictions and reveal the physicochemical mechanisms rationalizing the observations.[6,7] Materials scientists have resolved different problems in materials science by systematizing information about analogies in composition or structure between materials that exhibit similar physicochemical properties.[8–11]

Especially, in a discipline based on fundamental principles, such as condensed matter physics, it is essential to elucidate the physical mechanisms and which materials are manifested through each of these physical mechanisms. However, despite several new materials and superior properties having been discovered, it is still difficult to appropriately quantify the similarities between materials to elucidate the underlying physicochemical mechanisms of these properties. Furthermore, this difficulty arises from the fact that the mechanisms of materials' properties are typically interpreted in terms of physicochemical concepts based on relative criteria.

The phenomenon of superconductivity in materials, which originates from the instability of metals, is a well-known example of the above difficulty. One of the most successful theories that describe the microscopic mechanisms is the Bardeen–Cooper–Schrieffer (BCS) theory for superconductivity,[12] the origin of which is electron–phonon interactions. However, there also exist other mechanisms. For example, one of the most plausive origins of superconductivity in the high-T_C cuprates is electron–electron interactions. Nevertheless, it is not easy to achieve a consensus of classifying the superconducting mechanism of materials among researchers as the origins. Although the emergence of superconductivity is basically due to the instability in the metallic phase, it is not easy to achieve the consensus because both the mentioned and other mechanisms can contribute cooperatively in increasing the T_C value, for example. Although it is challenging to classify individual materials when considering phenomena that cause such a situation, it is expected that the underlying physical mechanisms can be discovered if we can inductively quantify the similarities between the materials of interest and group similar materials using all observation data.

Incidentally, inductive reasoning with inefficient similarity assessment can lead to misidentification of outliers[13] and difficulty in explaining the underlying physicochemical mechanisms of datasets using single models. Therefore, regarding predefined material descriptors, an exhaustive examination of all possible hypotheses about the unknown physicochemical mechanisms is necessary to assess the similarity between the materials. Furthermore, similarity measures are usually context-dependent. Because the context changes, the similarity measure must be modified to adequately capture the phenomena under study.[14,15] Thus, a quantitative measure of similarity needs to consider the uncertainty arising from the context or the measurement itself, especially in situations where material data are often insufficient and heavily biased. Moreover, similarities from different contexts may not be directly comparable in the integration to draw conclusions about the similarity between materials. These reasons make it challenging to apply data-driven approaches to materials science.

To overcome such issues and efficiently extract knowledge from the data, we propose a new approach that shifts from measuring the similarity between materials to quantitatively measure the confidence in their similarities. We adopt the Dempster–Shafer theory,[16–18] referred to as the evidence theory, to develop an evidential regression-based similarity measurement (eRSM) for detecting subgroups of materials such that leaned models from the subgroups show high correlations between descriptors and the target property of the constituent materials. Further analysis of models describing the subgroups provides valuable information to extract, interpret, and understand physical mechanisms. The Dempster–Shafer theory can be regarded as a generalization of the Bayesian approach for solving the problem of incomplete and insufficient information. Moreover, it is suitable for solving material data problems.[19,20] The measure of similarity here refers to whether the observed physical properties of the materials under study are described using the same hidden mechanism that has not yet been revealed. In other words, we consider any pair of materials (in the dataset) as similar if their physical properties can be described by the same hidden mechanism; otherwise, the pair of materials is considered dissimilar. We then first generate numerous hypothetical mechanisms by randomly choosing subsets of data instances and constructing regression models for each subset. Each regression model is considered a source of evidence of the similarities between materials. Thereafter, the Dempster–Shafer theory,[16–18] which has a foundation for modeling and combining the uncertainty of evidence, is applied to integrate the collected pieces of evidence to draw conclusions about the similarities between materials. The eRSM consists of three main steps as follows:

1. *Collect sources of evidence:* Hypothetical mechanisms are collected from a dataset by applying regression analysis with single or mixture models and are used as sources of evidence to rationalize the similarity states of materials.
2. *Model similarity evidence:* An appropriate mass function is designed to model the obtained evidence within the framework of the evidence theory.
3. *Combine pieces of evidence:* Dempster's rule of combination is used to integrate the pieces of the evidence.

The steps of the eRSM are explained in detail in Sec. II. Regarding the framework of the evidence theory, the essential contributions of the eRSM are collecting sources of evidence about the similarities between materials from datasets and designing suitable mass functions to model the pieces of evidence rationally. The effectiveness of obtained similarities using the eRSM for subdividing alloys from datasets into homogenous subgroups is supported by experiments on (1) a dataset of binary alloys with their Curie temperature as a target property (Sec. III B) and (2) two datasets of quaternary alloys with their magnetization (Sec. III C) and Curie temperature (Sec. III D) as the target properties. Further analysis of the detected subgroups to interpret the underlying physical mechanisms is shown in Sec. III E.

II. METHODOLOGY

We consider a dataset $\mathcal{D}$ consisting of p data instances. We assume that a data instance with index i in $\mathcal{D}$ is described by n predefined descriptors and is represented by an n-dimensional numerical vector, $\boldsymbol{x}_i = \left(x_i^1, x_i^2, \ldots, x_i^n\right) \in \mathbb{R}^n$. The target property of the data instance $\boldsymbol{x}_i$ is $y_i \in \mathbb{R}$. Thereafter, the dataset $\mathcal{D} = \left\{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2) \ldots (\boldsymbol{x}_p, y_p)\right\}$ is represented using a $(p \times (n+1))$ matrix. In this study, we consider that $\mathcal{D}$ may contain pairs of data instances $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$, where $\boldsymbol{x}_i \approx \boldsymbol{x}_j$; however, the value of y_i is far from y_j.

A. Collecting sources of similarity evidence

We perform random subset sampling of the data instances without replacement to collect a large amount of evidence of the similarity between pairs of data instances in $\mathcal{D}$. Considering each sample, we obtain two datasets: the reference dataset, $\mathcal{D}_{ref}$, and the evaluation dataset, $\mathcal{D}_{eval}$ ($\mathcal{D}_{ref} \cap \mathcal{D}_{eval} = \emptyset$ and $\mathcal{D}_{ref} \cup \mathcal{D}_{eval} = \mathcal{D}$). Considering $\mathcal{D}_{ref}$, we can generate a single function or multiple reference functions $f_r : \mathbb{R}^n \rightarrow \mathbb{R}$ using a Gaussian process (GP)[21] or a mixture of Gaussian processes (MGP),[22] respectively. This study applies GP- or MGP-based models instead of other nonlinear regression models, such as kernel ridge regression,[23] random forest regression,[24] or artificial neural networks[25] because GP or MGP can quantify the uncertainty of its prediction without introducing any other statistical validation. The sampling ratios of $\mathcal{D}_{ref}$ from $\mathcal{D}$ are fixed at 0.3 and 0.7 for the experiments with GP and MGP, respectively. Each reference function f_r is considered a source to provide pieces of evidence for the similarity between $(\boldsymbol{x}_i, y_i)$ and $(\boldsymbol{x}_j, y_j)$ in $\mathcal{D}_{eval}$. The function f_r is not used to provide any information about the similarities between the data instances in $\mathcal{D}_{ref}$ or between a data instance in $\mathcal{D}_{ref}$ and a data instance in $\mathcal{D}_{eval}$. This is to exclude self-evaluation to ensure the objectivity of the evidence. Regarding a reference function f_r, we consider the state of the similarity between $(\boldsymbol{x}_i, y_i)$ and $(\boldsymbol{x}_j, y_j)$ as

- Similar: Both data instances can be considered to have been generated by the function f_r [Fig. 1(a)].
- Dissimilar: Only one of the data instances can be considered to have been generated by the function f_r [Fig. 1(b)].
- Uncertain: Neither of the data instances can be considered to have been generated by the function f_r [Fig. 1(c)]. The uncertain state indicates that f_r does not provide any information about the similarity between $(\boldsymbol{x}_i, y_i)$ and $(\boldsymbol{x}_j, y_j)$.

To quantitatively evaluate whether $(\boldsymbol{x}_i, y_i)$ can be considered to have been generated by the regression function f_r, we use the

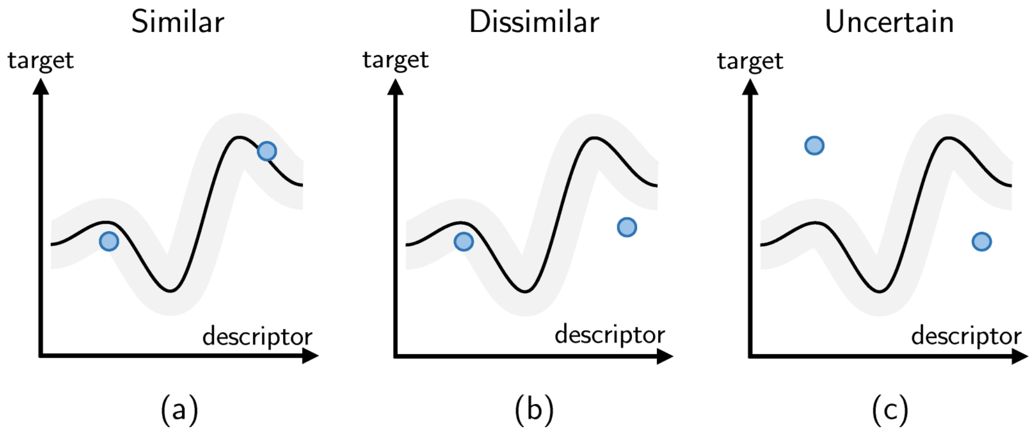

FIG. 1. Illustrative figures of the three possible similarity states between two data instances (blue circles), including similar (a), dissimilar (b), and uncertain (c), considering a referential regression model f_r (black line). The gray region is the interval that determines whether a data instance can be considered to have been generated by regression model f_r.

likelihood $p(O_i|f_r)$, the probability of event O_i that a data instance $(\boldsymbol{x}_i, y_i)$ is observed, considering f_r. The likelihood $p(O_i|f_r)$ is modeled using a normal distribution with mean and standard deviation depending on the predicted target value $\hat{y}_i = f_r(\boldsymbol{x}_i)$ and the corresponding standard error $\sigma_{\boldsymbol{x}_i}$ by f_r, respectively. This is expressed as

$$p(O_i|f_r) = \begin{cases} 1 & \text{if } \Delta_i \leq 3\,\bar{\sigma}, \\ 2 \times \int_{\Delta_i - 3\,\bar{\sigma}}^{+\infty} \mathcal{N}(u|0, \alpha\,\sigma_{\boldsymbol{x}_i})du & \text{otherwise,} \end{cases} \tag{1}$$

where $\Delta_i = |y_i - \hat{y}_i| = |y_i - f_r(\boldsymbol{x}_i)|$ is the deviation from the true to the predicted target values of data instance i using f_r, and $\bar{\sigma}$ is the average of the predictive standard error of all the data instances in $\mathcal{D}_{ref}$. α is the hyperparameter used to adjust the condition that restricts the data instances belonging to the function f_r. In other words, the interval that determines the probability that a data instance $(\boldsymbol{x}_i, y_i)$ belongs to f_r is $\alpha\,\sigma_{\boldsymbol{x}_i}$, and if the data instance falls outside this interval, it is determined that it does not belong to f_r. By increasing or decreasing the value of the parameter α, the condition for determining whether a data instance $(\boldsymbol{x}_i, y_i)$ belongs to f_r is relaxed or tightened, making $p(O_i|f_r)$ larger or smaller, respectively. Optimal values of α can be chosen using statistical criteria and appropriate validation methods; however, we set $\alpha = 2$ for all experiments in this work to reduce model complexity. We consider $p(O_i|f_r)$ as the probability that $(\boldsymbol{x}_i, y_i)$ is generated by f_r, and $p(\overline{O}_i|f_r) = 1 - p(O_i|f_r)$ is the probability that $(\boldsymbol{x}_i, y_i)$ is not generated by f_r. Figure 1 in the supplementary material illustrates the process of modeling the probability $p(O_i|f_r)$.

Events where $(\boldsymbol{x}_i, y_i)$ or $(\boldsymbol{x}_j, y_j)$ is generated by the function f_r are independent events. Therefore, considering the function f_r, we can evaluate the joint probabilities of observing

- both data instances:

$$p(O_i, O_j|f_r) = p(O_i|f_r) \times p(O_j|f_r); \tag{2}$$

- only one of the data instances:

$$\begin{aligned} &p(O_i, \overline{O}_j|f_r) + p(\overline{O}_i, O_j|f_r) \\ &\quad = p(O_i|f_r) \times p(\overline{O}_j|f_r) + p(\overline{O}_i|f_r) \times p(O_j|f_r); \end{aligned} \tag{3}$$

- neither of the data instances:

$$\begin{aligned} p(\overline{O}_i, \overline{O}_j|f_r) &= p(\overline{O}_i|f_r) \times p(\overline{O}_j|f_r) \\ &= 1 - p(O_i, O_j|f_r) - p(O_i, \overline{O}_j|f_r) - p(\overline{O}_i, O_j|f_r). \end{aligned} \tag{4}$$

B. Modeling evidence by mass functions

Considering the Dempster–Shafer theory framework,[16] we begin by defining the frame of discernment Ω. Let $\Omega = \{s, ds\}$ be the universal set representing the similarity states of any two data instances $(\boldsymbol{x}_i, y_i)$ and $(\boldsymbol{x}_j, y_j)$. s and ds denote the similarity and dissimilarity states between the two data instances, respectively.

According to the Dempster–Shafer theory, the evidence of the similarity states between these two data instances is represented by a mass function $m^{i,j}$ (or a basic probability assignment).[16] This assigns probability masses to all the nonempty subsets of Ω ($\mathcal{X} = \{\{s\}, \{ds\}, \{s, ds\}\}$). It is defined as follows:

$$m^{i,j} : \mathcal{X} \rightarrow [0, 1] \text{ with } \sum_{E \in \mathcal{X}} m(E) = 1. \tag{5}$$

The masses assigned to $\{s\}$ and $\{ds\}$ reflect the degrees of belief exactly committed to the evidence to support the similarity and dissimilarity between $(\boldsymbol{x}_i, y_i)$ and $(\boldsymbol{x}_j, y_j)$, respectively. The weight assigned to $\{s, ds\}$ expresses the degree of belief that the evidence provides no information about the similarity (or dissimilarity) between $(\boldsymbol{x}_i, y_i)$ and $(\boldsymbol{x}_j, y_j)$.

Therefore, the mass function $m_{f_r}^{i,j}$, which models a piece of evidence of the similarity between $(\boldsymbol{x}_i, y_i)$ and $(\boldsymbol{x}_j, y_j)$ collected from f_r, is defined as follows:

$$m_{f_r}^{i,j}(\{s\}) = \frac{p(O_i, O_j|f_r)}{\gamma_{i,j}}, \tag{6}$$

$$m_{f_r}^{i,j}(\{ds\}) = \frac{p(O_i, \overline{O}_j|f_r) + p(\overline{O}_i, O_j|f_r)}{\gamma_{i,j}}, \tag{7}$$

$$m_{f_r}^{i,j}(\{s, ds\}) = 1 - \frac{1}{\gamma_{i,j}} + \frac{p(\overline{O}_i, \overline{O}_j|f_r)}{\gamma_{i,j}}, \tag{8}$$

where $\gamma_{i,j} = \left(e^{\frac{\bar{\sigma}}{\Delta_y}} + 1\right) \times \left(\frac{\sigma_{\boldsymbol{x}_i}}{\bar{\sigma}} + 1\right) \times \left(\frac{\sigma_{\boldsymbol{x}_j}}{\bar{\sigma}} + 1\right)$ is a discounting factor,[16,26] which describes the unreliability of evidence about the similarity between $(\boldsymbol{x}_i, y_i)$ and $(\boldsymbol{x}_j, y_j)$ collected from a source of evidence f_r. Δ_y is the variation range of the target variable y in the dataset $\mathcal{D}$. The smaller the $\bar{\sigma}$ relative to Δ_y, the more reliable the learned regression function f_r. Also, when $\sigma_{\boldsymbol{x}_i}$ and $\sigma_{\boldsymbol{x}_j}$ are smaller than $\bar{\sigma}$, f_r can provide reliable evidence for the relationship between $(\boldsymbol{x}_i, y_i)$ and $(\boldsymbol{x}_j, y_j)$. By contrast, when $\sigma_{\boldsymbol{x}_i}$ and $\sigma_{\boldsymbol{x}_j}$ are large compared to $\bar{\sigma}$, f_r cannot provide reliable evidence for the relationship between $(\boldsymbol{x}_i, y_i)$ and $(\boldsymbol{x}_j, y_j)$. A detailed explanation of each component in $\gamma_{i,j}$ is provided in Sec. I of the supplementary material.

C. Dempster's rule in combining evidence

Assuming that we can collect q pieces of evidence from $\mathcal{F}_r = \{f_r^1, \ldots, f_r^q\}$, a set of q reference functions is generated from $\mathcal{D}$ to evaluate the similarity between a pair of data instances with indices i and j. According to the Dempster–Shafer theory framework, any two pieces of evidence collected from the reference functions f_r^l and f_r^k, which are modeled by the corresponding mass functions $m_{f_r^l}^{i,j}$ and $m_{f_r^k}^{i,j}$, respectively, can be combined using the Dempster rule of combination to assign the joint mass $m_{\{f_r^l, f_r^k\}}^{i,j}$ to each nonempty subset E of Ω as follows:

$$\begin{aligned} m_{\{f_r^l, f_r^k\}}^{i,j}(E) &= \left(m_{f_r^l}^{i,j} \oplus m_{f_r^k}^{i,j}\right)(E) \\ &= \frac{\sum_{E_t \cap E_v = E} m_{f_r^l}^{i,j}(E_t) \times m_{f_r^k}^{i,j}(E_v)}{1 - \sum_{E_t \cap E_v = \emptyset} m_{f_r^l}^{i,j}(E_t) \times m_{f_r^k}^{i,j}(E_v)}, \end{aligned} \tag{9}$$

where E, E_t, and E_v are nonempty subsets of Ω. Dempster's rule is commutative and associative.

Based on Dempster's rule, the obtained mass functions corresponding to the q pieces of evidence are combined to assign the final mass $m^{i,j}_{\mathcal{F}_r}$ as follows:

$$m^{i,j}_{\mathcal{F}_r}(E) = \left(m^{i,j}_{f^1_r} \oplus m^{i,j}_{f^2_r} \oplus \ldots \oplus m^{i,j}_{f^q_r}\right)(E). \qquad (10)$$

We perform similar analyses for all pairs of data instances in $\mathcal{D}$ to construct symmetric matrices M comprising the similarities $(M[i, j] = M[j, i] = m^{i,j}_{\mathcal{F}_r}(\{s\}))$ between them. Thereafter, the obtained matrix is applied for further unsupervised data mining analysis, such as clustering or data visualization.

III. EXPERIMENTS AND RESULTS

In this section, we perform three experiments to demonstrate the application of our similarity measurement in dealing with outliers and data generated by multiple mechanisms when designing material descriptors. We apply the eRSM to measure similarities between magnetic of three datasets for detecting subgroups of materials: (1) the experimentally observed Curie temperature dataset ($\mathcal{D}_{binary}$) of binary alloys for transitioning rare-earth metals, (2) the dataset of calculated magnetization of quaternary high-entropy alloys ($\mathcal{D}^{Mag}_{quaternary}$), and (3) the dataset of calculated Curie temperature of quaternary high-entropy alloys ($\mathcal{D}^{T_C}_{quaternary}$). Note that the datasets $\mathcal{D}^{Mag}_{quaternary}$ and $\mathcal{D}^{T_C}_{quaternary}$ contain similar alloys and differ only in the target properties.

A. Datasets

The details of the datasets investigated in this study are as follows.

- Binary alloys dataset $\mathcal{D}_{binary}$:[27] A material dataset containing 100 transition rare-earth metal binary alloys, comprising nickel (Ni), manganese (Mn), cobalt (Co), or iron (Fe), and the corresponding Curie temperatures (T_C). This dataset was collected from the Atomwork database of the National Institute for Materials Science.[28,29] Each binary alloy in $\mathcal{D}_{binary}$ is represented using seven descriptors: (1) and (2) the atomic number of transition metal (Z_T) and rare-earth (Z_R) constituents, (3) projection of the spin magnetic moment onto the total angular moment of the 4f elections ($J_{4f}(1 - g_j)$), (4) and (5) covalent radius (r_{covT}) and first ionization (IP_T) of the transition metal, and (6) and (7) concentration of the transition metal (C_T) and rare-earth metal (C_R). The selection of these seven descriptors has been discussed in detail in previous studies.[10,30]
- Quaternary high-entropy alloys datasets $\mathcal{D}_{quaternary}$:[27] A material dataset contains 990 equiatomic quaternary high-entropy alloys, which comprise 14 transition metals Ag, Cd, Co, Cr, Cu, Fe, Mn, Mo, Ni, Pd, Rh, Ru, Tc, Zn, and the corresponding calculated magnetizations and Curie temperatures in the BCC phase. The dataset was collected from an original dataset of 147 630 equiatomic quaternary high-entropy alloys calculated using the Korringa–Kohn–Rostoker coherent approximation method.[31] Each alloy in $\mathcal{D}_{quaternary}$ is represented using 135 compositional descriptors, including the means, standard deviations, and covariance of the atomic representations of their constituent elements[13] and four categorical features indicating the elements comprising the quaternary alloy. The feature selection process applied to this dataset has been discussed in detail in Sec. III of the supplementary material.

B. Assessment of the similarity between transition rare-earth metal binary alloys based on mechanisms of Curie temperature

In the first experiment, we show the versatility of the eRSM for detecting outliers and identifying a mixture of mechanisms. We

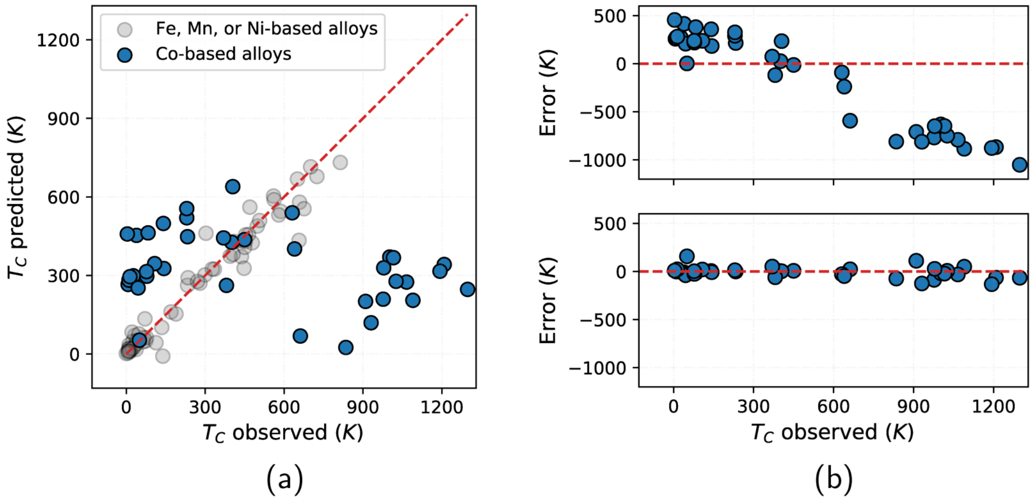

FIG. 2. (a) Observed and predicted Curie temperature of alloys in the dataset $\mathcal{D}_{binary}$ using model generated for nickel (Ni), iron (Fe), and manganese (Mn)-based alloys. The blue and gray points indicate cobalt (Co)-based alloys and alloys of other transition metals (Ni, Fe, Mn), respectively. (b) Prediction error of Co-based alloys when excluding (top) or including (bottom) data of other Co-based alloys to the training dataset.

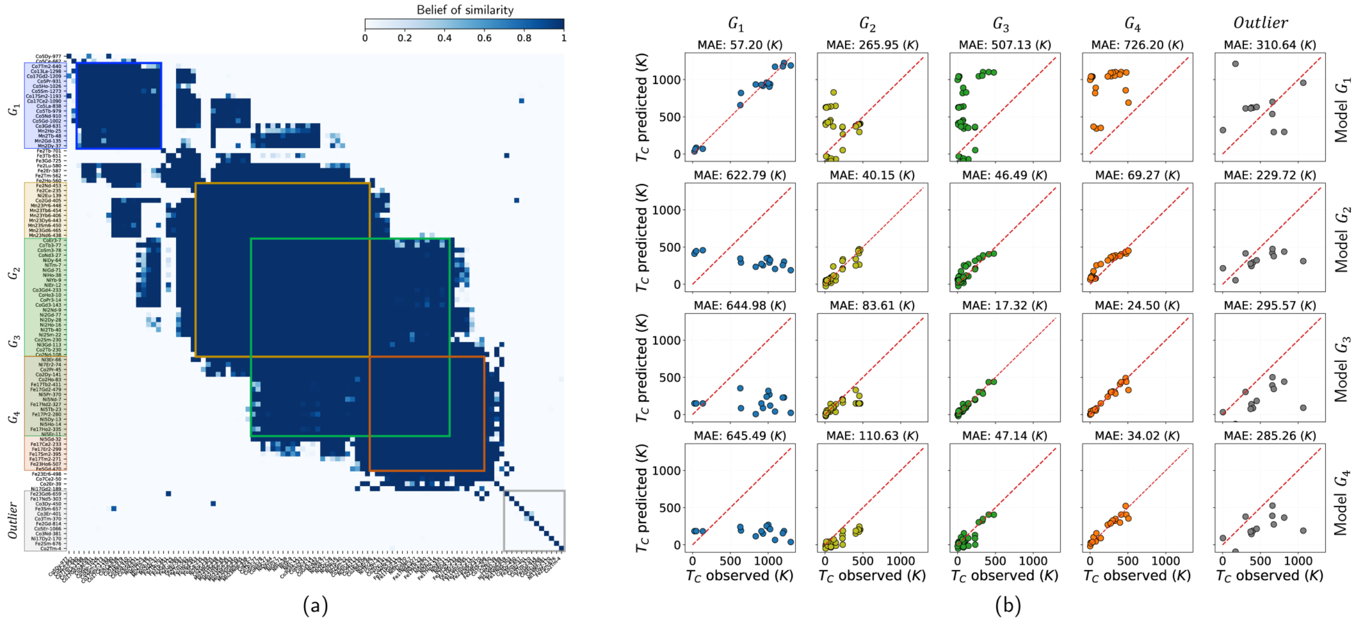

FIG. 3. (a) Heatmap illustrating the similarity matrix M_{binary} extracted for all the data instances in the $\mathcal{D}_{binary}$. (b) Confusion matrices measuring the regression-based similarities between alloys in four groups G_1–G_4 and the dissimilarities between the models generated for alloys in different groups.

apply the eRSM to assess the similarities between 100 transition rare earth metal binary alloys comprising nickel (Ni), manganese (Mn), cobalt (Co), or iron (Fe) in the dataset $\mathcal{D}_{binary}$ based on their Curie temperatures. We can construct a regression model using a Gaussian process by considering the data instances in $\mathcal{D}_{binary}$. This shows a high prediction accuracy with an R^2 score of 0.963 and a mean absolute error (MAE) of 40 (K) in tenfold cross-validation. However, such a nonparametric regression model does not guarantee the reliability of the model in the subsequent exploratory predictions. This is because the number of observable alloys is relatively small compared to the number of possible alloys.

Figure 2(a) shows the results of the exploratory prediction of the Curie temperature of the Co-based binary alloys in $\mathcal{D}_{binary}$ using a Gaussian process regression model constructed from the data of binary alloys of Ni, Mn, and Fe. The regression model constructed from the data of binary alloys of Ni, Mn, and Fe shows a high prediction accuracy in tenfold cross-validation [$R^2 = 0.946$ and $\text{MAE} = 35$ (K)]. Although the Co-based alloys with high Curie temperature tend to be underestimated by the model, the other Co-based alloys are often overestimated. The prediction error for the Co-based alloys is critically reduced when some data of the other Co-based alloys are included [Fig. 2(b)]. This observation supports the hypothesis that the underlying mechanisms are different between the Co-based alloys and alloys of other transition metals. This facilitates the use of the eRSM to clarify the mixture mechanism from this dataset.

By applying the eRSM on the dataset $\mathcal{D}_{binary}$, we obtain a similarity matrix M_{binary} with moderately high similarity values among the data instances [Fig. 3(a)]. Thus, approximately, all the data instances can be regressed by a relatively smooth function. This is consistent with the high prediction accuracy of tenfold cross-validation for all the alloys in the dataset. Considering the exploratory data analysis, to avoid false intuition or misunderstanding, the grouping of alloys in $\mathcal{D}_{binary}$ is done such that the similarities

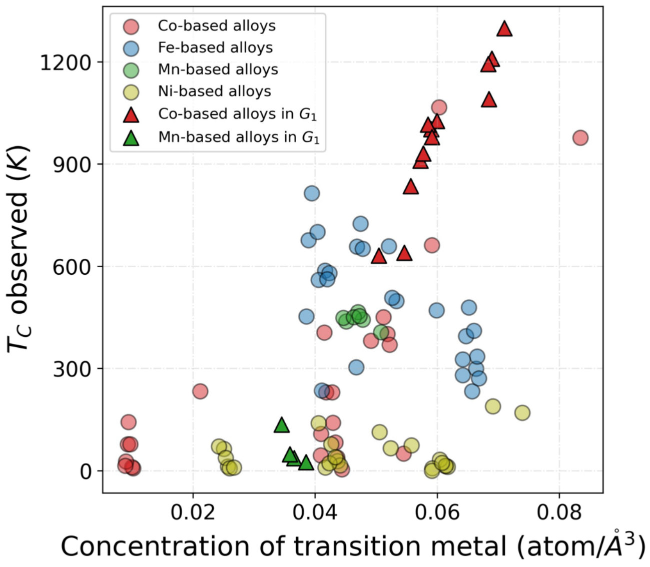

FIG. 4. Dependence of T_C on the concentration of the transition metal (C_T) in alloys. Red, blue, green, and yellow scatters indicate alloys containing cobalt (Co), iron (Fe), manganese (Mn), and nickel (Ni). Alloys in G_1 are highlighted by triangles.

between the alloys in each group are high. Moreover, one alloy can belong to more than one group simultaneously, or it can be in none of the groups. We apply a graph-based clustering method[32] to the extracted similarity matrix to detect overlapping subgroups of materials. As a result, we observe four groups of alloys, denoted as G_1, G_2, G_3, and G_4, which show high intragroup similarities, exceeding 0.7 [Fig. 3(a)]. Nevertheless, the similarity between the alloys in group G_1 and those in G_2, G_3, and G_4 is significantly dissimilar. In addition, a small group of alloys [Fig. 3(a), gray region] is approximately different from all the others and can be considered outliers. The remaining alloys are not assigned to any group to have confidence in the clustering analysis results.

To evaluate the validity of the analysis process quantitatively, we trained the regression models for T_C using data from each of the four groups G_1, G_2, G_3, and G_4. Moreover, we monitored their prediction accuracy on these groups. The confusion matrix summarizing the correlation between the observed and predicted T_C by the four learned regression models is shown in Fig. 4. The diagonal plots illustrate the cross-validation results of the models learned from the four groups of alloys. The off-diagonal plot shows the correlation between the observed T_C and the predictions made by the model learned from the alloys of the other groups. The obtained results confirm the intragroup similarity of the alloys in groups G_1, G_2, G_3, and G_4, respectively, dissimilarity between the five groups, and intra-group dissimilarity of the alloys considered outliers. This indicates that the obtained results suggest that the physical mechanisms of alloys in G_1 may be different from those of the alloys in G_2, G_3, and G_4. Nonetheless, it is difficult to determine the

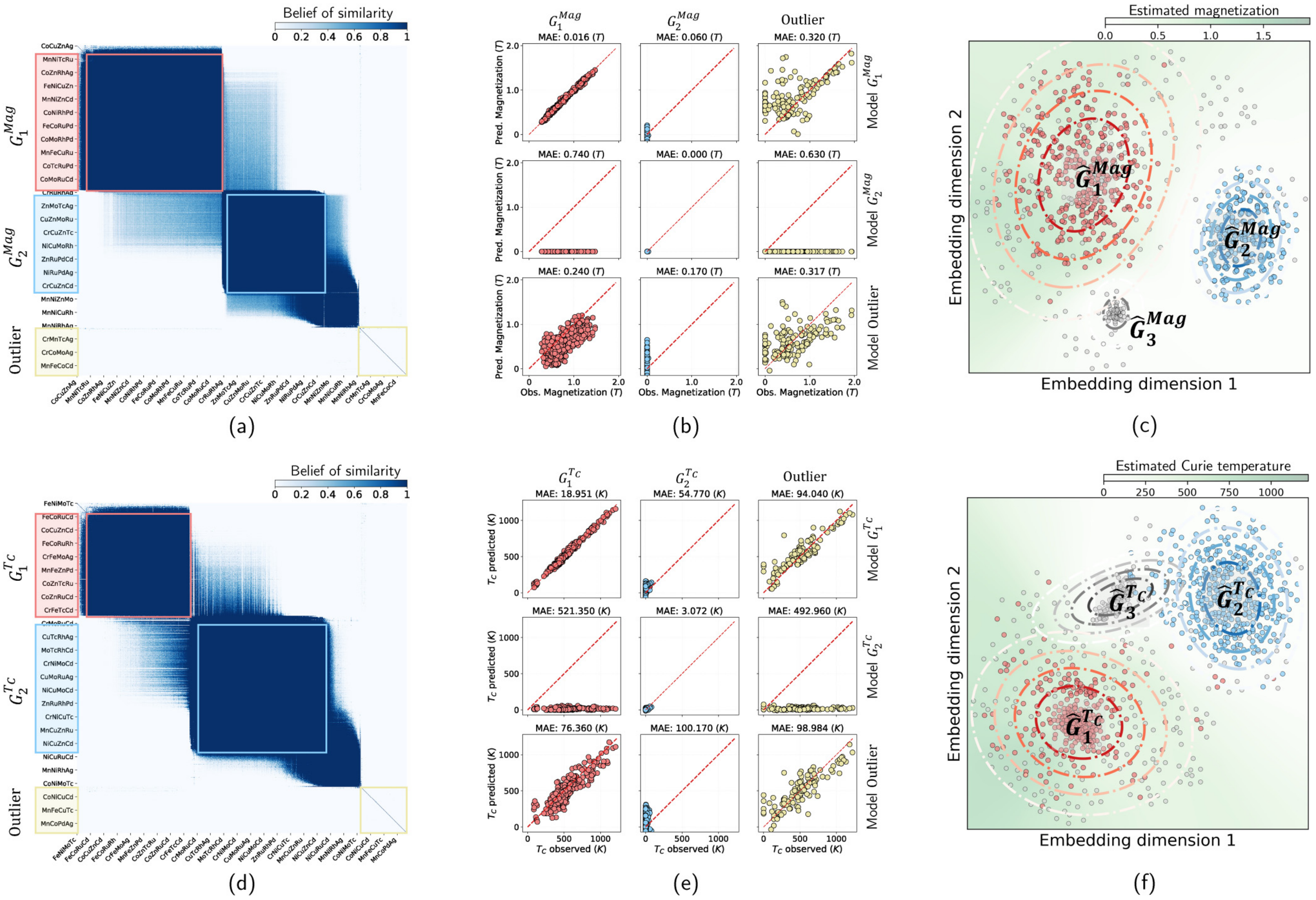

FIG. 5. (a) and (d) Heatmaps illustrating the similarity matrices $M^{Mag}_{quaternary}$ (a) and $M^{T_C}_{quaternary}$ (d) extracted from datasets $\mathcal{D}^{Mag}_{quaternary}$ and $\mathcal{D}^{T_C}_{quaternary}$, focusing on mechanisms of magnetization and T_C, respectively. (b) and (e) The confusion matrix summarizes the differences between the magnetization (b) or T_C (e) mechanisms of alloys in extracted groups. (c) and (f) Visualization of quaternary alloys in the two-dimensional embedding spaces constructed by applying the t-distributed stochastic neighbor embedding (t-SNE) to $M^{Mag}_{quaternary}$ (c) and $M^{T_C}_{quaternary}$ (f). Red, blue, and gray contours indicate gaussian models $\hat{G}_1^{Mag}$ $\left(\hat{G}_1^{T_C}\right)$, $\hat{G}_2^{Mag}$ $\left(\hat{G}_2^{T_C}\right)$, and $\hat{G}_3^{Mag}$ $\left(\hat{G}_3^{T_C}\right)$, respectively, learned by using the Gaussian mixture models[33] in the embedding space focusing on mechanisms of magnetization (T_C). In addition, red and blue points in sub-figures (b) and (c) [(e) and (f)] indicate the alloys in G_1^{Mag} $\left(G_1^{T_C}\right)$ and G_2^{Mag} $\left(G_2^{T_C}\right)$, respectively.

differences between the mechanisms of the T_C of alloys in G_2, G_3, and G_4.

Moreover, considering the alloys in G_1, there is a strong linear correlation between T_C and the concentration of transition metals in the alloys with a Pearson correlation coefficient of 0.95 (Fig. 4, triangle scatters). This result is consistent with the observation of the previous research[30] when considering all binary alloys of transition metals and rare-earth metals in $\mathcal{D}_{binary}$; the range of T_C is found to be correlated with the composition ratio of the transition metals. Furthermore, 13 of the 17 alloys in G_1 are Co-based alloys with high Curie temperatures ($T_C > 600$ K). By contrast, most of the other Co-based alloys in $\mathcal{D}_{binary}$ have lower Curie temperatures ($T_C < 500$ K) and are assigned to G_2, G_3, and G_4. These results are consistent with the observation that the regression model for Fe-, Mn-, and Ni-based alloys tends to underestimate the T_C of the Co-based alloys with high T_C and overestimates the T_C of the remaining Co-based alloys [Fig. 2(a)].

In addition, we examine the behavior of eRSM on toy datasets synthesized with outliers or multiple mechanisms to assess the efficiency of this similarity measure. Detailed results of these experiments are summarized in Sec. II of the supplementary material. Briefly, the eRSM demonstrates that it can effectively assess the

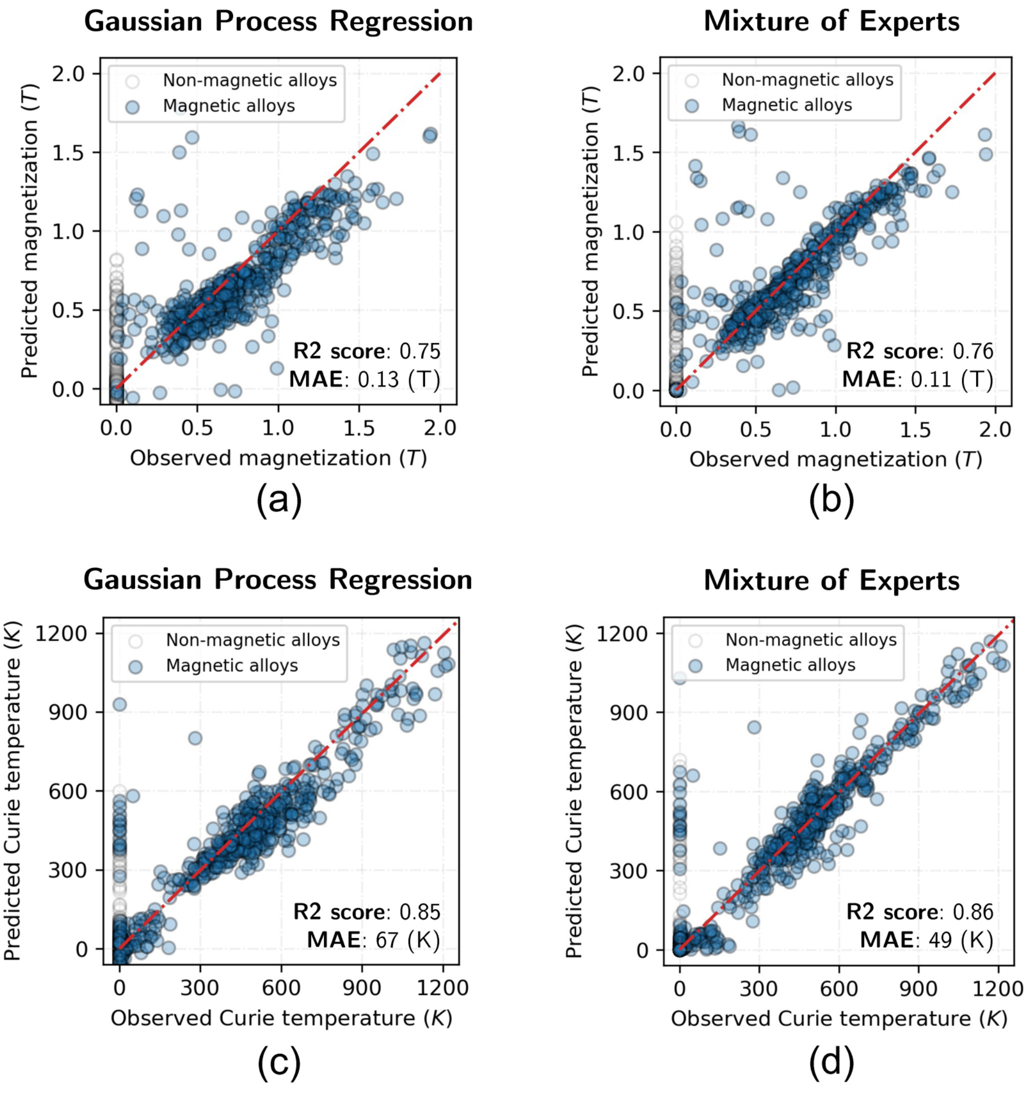

FIG. 6. Prediction accuracies for magnetization (a) and (b) and Curie temperature (c) and (d) of the alloys with tenfold cross-validations. Prediction validation results with single gaussian process regression models for magnetization and Curie temperature are shown in sub-figures (a) and (c), respectively. Prediction validation results with mixtures of expert models for magnetization and Curie temperature are shown in sub-figures (b) and (d), respectively. Blue and white circles indicate magnetic alloys (finite magnetization) and non-magnetic alloys (zero magnetization), respectively.

similarity between the data instances and use the similarity for detecting outliers and a mixture of mechanisms.

C. Assessment of the similarity between quaternary high-entropy alloys based on mechanisms of magnetization

The effectiveness of the eRSM in detecting outliers and identifying mixture mechanisms in the material dataset has been shown in the previous experiment. In the next two experiments, we show the potential of applying the measured similarity to design descriptors for materials.

Considering this experiment, we subsequently apply the eRSM to assess the similarities between 990 quaternary high-entropy alloys comprising 14 transition metals in the dataset $\mathcal{D}^{Mag}_{quaternary}$ based on their magnetization. To predict the magnetization of these alloys, we attempted to construct an optimal Gaussian process regression model using the designed descriptors. The Gaussian process can poorly regress the magnetization with an R^2 score of 0.75 and an MAE of 0.13 (T) in the tenfold cross-validation. The obtained results suggest that the magnetization of these alloys may not be described by a single model in the designed descriptor space. This indicates that the existence of outliers or mixture models of the magnetization properties of these alloys in the descriptor space should be considered in the analysis of this dataset.

Applying the eRSM, we obtain a similarity matrix $M^{Mag}_{quaternary}$ with two core groups of alloys denoted by G^{Mag}_1 and G^{Mag}_2, showing high intra-group similarities and exceeding 0.5 [Fig. 5(a)]. Some of the alloys in G^{Mag}_1 are similar to those in G^{Mag}_2; nonetheless, the rest show apparent dissimilarities. Furthermore, one small group of alloys [Fig. 5(a), yellow region] showed dissimilarities with the others and could be considered outliers. The remaining alloys in $\mathcal{D}^{Mag}_{quaternary}$ do not exhibit apparent similarities with alloys in groups G^{Mag}_1 and G^{Mag}_2. Therefore, they are not assigned to any group.

To validate the obtained results quantitatively, we trained three regression models using data from each group, G^{Mag}_1, G^{Mag}_2, and outliers. We monitored the prediction accuracy of the three learned regression models for data in all the groups. The confusion matrix summarizing the correlations between the observed and predicted values of the target variable using the learned regression models is shown in Fig. 5(c). The diagonal plots illustrate the tenfold cross-validation results of the models learned from these three groups of alloys. The off-diagonal plot shows the correlation between the observed magnetization and the predictions made by the model learned from the alloys of the other groups.

The obtained results confirm the intragroup similarity of the alloys in groups G^{Mag}_1 and G^{Mag}_2, respectively, the dissimilarity between the two groups, and the intra-group dissimilarity of the alloys considered as outliers. Specifically, we observe that group G^{Mag}_2 consists of ferrimagnetic alloys or alloys whose magnetization is relatively smaller [magnetization < 0.1 (T)] than the others in the group G^{Mag}_1. In contrast, using the data in G^{Mag}_1, we can construct a Gaussian process regression model with a high prediction accuracy with an R^2 score of 0.992 and an MAE of 0.016 (T) in the tenfold cross-validation.

Therefore, we can use the information of the constituent elements of each alloy to predict which group it belongs to in advance[20] and apply an appropriate regression model to improve prediction accuracy for the alloys. We combine the similarity measured by using the eRSM with the Jaccard similarity coefficient[34] and apply the t-distributed stochastic neighbor embedding[35] (t-SNE) to construct a two-dimensional embedding map [Fig. 5(c)]. Details of the combination method are shown in Sec. IV of the supplementary material. As a result, we can easily distinguish the alloys in groups G^{Mag}_1 (red) and G^{Mag}_2 (blue) when they form two separate regions with high density in the embedding space. We apply a Gaussian mixture model[33] (GMM) on the embedding space to identify groups and calculate the probability of an alloy belonging to a particular identified group. Alloys in different groups are treated differently by using a mixture of experts[36] (MoE) approach. Figures 6(a) and 6(b) show a reduction of the proposed mixture of experts in MAE of 18% compared with the result of the single model, from 0.13 (T) to 0.11 (T). Further analysis shows that applying the obtained similarities in MOE improves the prediction accuracy for magnetic alloys [Fig. 7(a) in the supplementary material].

D. Assessment of the similarity between the quaternary high-entropy alloys based on mechanisms of Curie temperature

Considering this experiment, the target data are the same as in Sec. III C ($\mathcal{D}_{quaternary}$); however, the physical property of interest is T_C. A regression model can be constructed using a Gaussian process. This shows a rather high prediction accuracy in tenfold cross-validation with an R^2 score of 0.85 and an MAE of 67 (K). We also observe two distinguishable groups of quaternary alloys in the dataset $\mathcal{D}^{T_C}_{quaternary}$ when applying the eRSM. Figure 5(d)

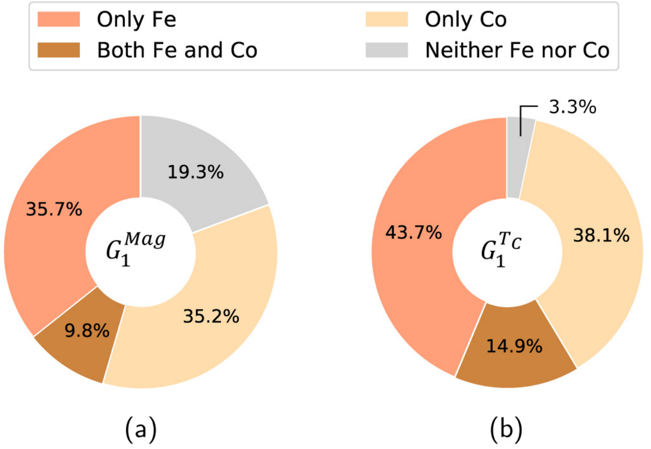

FIG. 7. Proportions of quaternary alloys containing Fe or Co in group G^{Mag}_1 (a) and $G^{T_C}_1$ (b).

illustrates the similarity matrix $M^{T_C}_{quaternary}$ with two groups of alloys denoted as $G_1^{T_C}$ and $G_2^{T_C}$, showing high intra-group similarities and exceeding 0.5. Some of the alloys in $G_1^{T_C}$ are similar to those in $G_2^{T_C}$. Nonetheless, the others exhibit apparent dissimilarities, which is consistent with the observation of two high-density regions (red) in the embedding map of $M^{T_C}_{quaternary}$ [Fig. 5(e)]. Furthermore, a small group of alloys [Fig. 5(d), yellow region] showed dissimilarities with all the others and could be considered outliers. The remaining alloys do not show apparent similarities with alloys in groups $G_1^{T_C}$ and $G_2^{T_C}$; thus, they are not assigned to any group.

Following the same analysis procedure as in Sec. III C, we trained regression models for Curie temperature using data from each of the three groups $G_1^{T_C}$, $G_2^{T_C}$, and outliers and monitored their prediction accuracy on these groups. Figure 5(f) shows the confusion matrix that summarizes the obtained results. The diagonal plots illustrate the tenfold cross-validation results of the models learned from these three groups of alloys. The off-diagonal plot shows the correlation between the observed Curie temperature and the predictions made by the regression model learned from the alloys of the other groups. We can also confirm the intra-group similarity of the alloys in groups $G_1^{T_C}$ and $G_2^{T_C}$, respectively, dissimilarity between the two groups, and intra-group dissimilarity of the alloys considered outliers.

Specifically, we observe that the Curie temperatures of approximately all the alloys in group $G_2^{T_C}$ have a low T_C, which is 0 (K) or

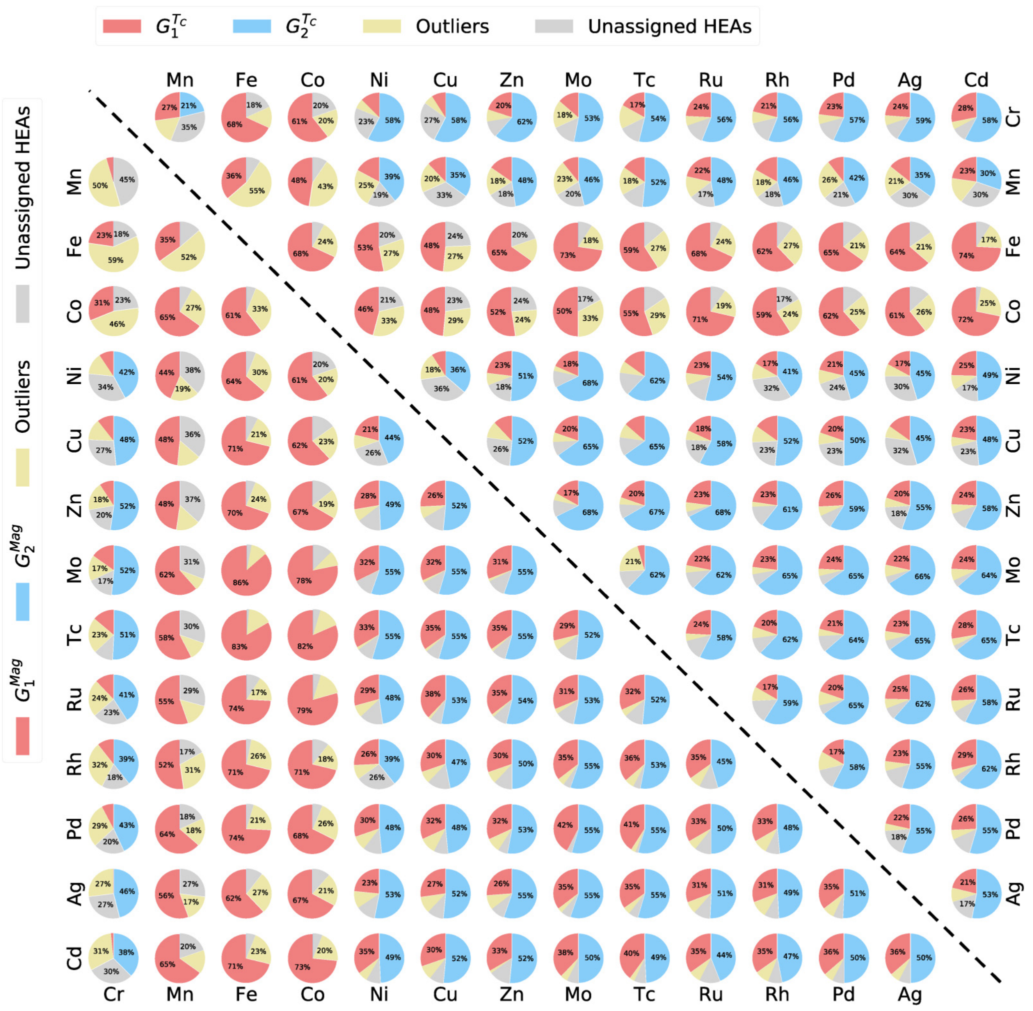

FIG. 8. Effect of coexistence of the 14 transition metals on magnetization and Curie temperature mechanisms. Each pie chart results from quaternary alloys containing the respective element pair. They show the percentages of alloys that follow the magnetization mechanisms (lower-left triangle) and Curie temperature mechanisms (upper-right triangle), as extracted by the eRSM. Red and blue areas indicate the percentages of alloys whose magnetization and T_C are finite $\left(G_1^{Mag} \text{ and } G_1^{T_C}\right)$ and zero $\left(G_2^{Mag} \text{ and } G_2^{T_C}\right)$, respectively. Yellow areas indicate the percentages of alloys that are detected as outliers. By contrast, gray regions indicate the fractions of alloys not assigned to the extracted groups.

relatively smaller than that of the other alloys. Furthermore, using the data in $G_1^{T_C}$, we can construct a Gaussian process regression model with a high prediction accuracy with an R^2 score of 0.985 and an MAE of 19 (K) in the tenfold cross-validation.

Therefore, we utilize the similarity information to design descriptors for quaternary alloys due to the effectiveness of the data for detecting the mixture of multiple mechanisms in the dataset. We apply similar methods as in the previous experiment to construct a two-dimensional embedding map [Fig. 5(f)] and then learn a mixture of experts to predict Curie temperature of quaternary alloys in the dataset $\mathcal{D}_{quaternary}^{T_C}$. The proposed mixture of models exhibits higher prediction accuracy than the single model in tenfold cross-validations [Figs. 6(c) and 6(d)]. The MAE of the proposed mixture of expert reduces approximately 36%, from 67 (K) to 49 (K).

E. Discussion of the obtained similarities between materials and the associated physical mechanisms

Regarding the experiments with the datasets $\mathcal{D}_{quaternary}^{Mag}$ and $\mathcal{D}_{quaternary}^{T_C}$ focusing on magnetization or T_C, the datasets seem to be a self-evident example where magnetization and T_C are cases sensitive to finite or zero. As we can see from the results described above (Secs. III C and III D and Sec. VI in the supplementary material), the prediction accuracy is low when considering a single regression model for the entire dataset. In this section, we pay attention to the analysis of the extracted alloys groups G_1^{Mag}, G_2^{Mag}, $G_1^{T_C}$, and $G_2^{T_C}$ to identify underlying patterns.

Figure 7 shows that Fe and Co, which have a large spin moment, ferromagnetic interactions with many elements and result in high magnetization or T_C, are dominant elements comprising alloys in two groups G_1^{Mag} (a) and $G_1^{T_C}$ (b). Furthermore, in the analysis that considers the proportion of the quaternary alloys fixing two of their four constituent elements concerning the extracted four groups G_1^{Mag}, G_2^{Mag}, $G_1^{T_C}$, and $G_2^{T_C}$, we observe that the proportion of Fe-containing and Co-containing alloys in two groups G_1^{Mag} (a) and $G_1^{T_C}$ is significantly larger than other groups (Fig. 8). Thus, the prediction models constructed from the data of the alloys in G_1^{Mag} or $G_1^{T_C}$ are more suitable to predict magnetization or T_C, respectively, of alloys containing these elements. The remaining Fe–X and Co–X (X denotes the other transition metals comprised in the alloys) alloys are considered outliers of the extracted mechanisms or unassigned HEAs, which are not assigned to any of these mechanisms. Conversely, Mn–X alloys exhibit similar behavior as Fe–X and Co–X when focusing on the magnetization mechanisms. However, for the Curie temperature, the Mn–X alloys are categorized in the group $G_2^{T_C}$ of low T_C besides the other groups. Especially among the Fe–X and Co–X alloys, the percentage of Fe–Mn and Co–Mn alloys considered outliers of the mechanisms extracted from $G_1^{T_C}$ is relatively higher, 55% and 43%, respectively (Fig. 8).

For further investigation, we organized the raw data of the quaternary alloys by focusing on the presence or absence of Mn. Figure 9 shows the correlation between magnetization and Curie temperature of 556 (56%) alloys with non-zero properties. The total number of data instances is 990, and the number of data instances where both T_C and magnetization are zero is 413 (42%), while there are 21 (2%) alloys with zero T_C but have finite magnetization. We found that the alloys containing all three elements, Mn, Fe, and Co, show high Curie temperatures [$T_C > 900$ K]. Conversely, the alloys containing either pairs of Mn–Fe or Mn–Co show moderate Curie temperatures. By contrast, the Mn-containing alloys without Fe or Co have low Curie temperatures [$T_C < 250$ K]. Furthermore, the trends of these three alloy groups do not offer any significant correlation between magnetization and Curie temperature. However, an apparent positive correlation between magnetization and Curie temperature can be observed for the group of Mn-free alloys.

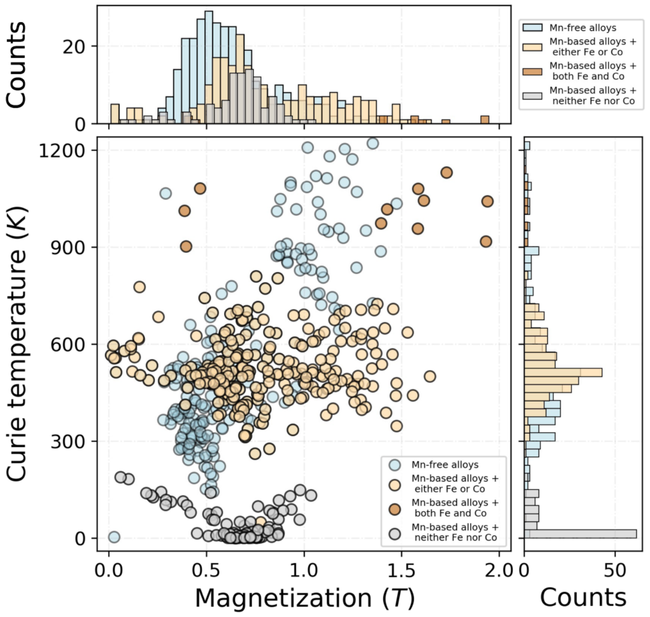

FIG. 9. Correlation between magnetization (T) and Curie temperature (K) of quaternary alloys with non-zero magnetization and non-zero Curie temperature in datasets $\mathcal{D}_{quaternary}^{Mag}$ and $\mathcal{D}_{quaternary}^{T_C}$. Marginal plots show a histogram of the properties of the alloys.

To interpret the results obtained, we considered a hypothesis of the origin of the observed data. The estimated magnetization is the sum of all the local magnetic moments divided by the unit volume. The local magnetic moments are determined by the spin configurations of atomic sites that stabilize the structure of alloys. Conversely, given a particular structure and spin configuration, the T_C can be estimated from the spin–spin exchange energy. First-principles calculations show that early transition metals and late transition metals often have antiferromagnetic interactions.[37] This interaction has also been confirmed in high-entropy alloys by using automatic exhaustive calculations.[31] Mn lies between early and late transition metals; thus, the estimation of the spin configuration (ferromagnetic or antiferromagnetic) in Mn-containing alloys should be cautiously considered in different situations, especially in high-entropy alloys whose elements can stochastically exist at the same atomic site. From this consideration, we can admit a hypothesis that the alloys containing Mn follow a different rule for magnetization than those grouped into G_2^{Mag}. Conversely, the alloys containing Mn may follow the same rules for T_C as the alloys grouped into $G_2^{T_C}$, albeit with a spin configuration that provides magnetization. The details are beyond the scope of this paper and will not be discussed here, but further analysis is promising.

IV. CONCLUSIONS

In this study, we developed a method that can be used to rationally transform material data from multiple sources into evidence of similarities between materials and combine the evidence to conclude the similarities between materials. The extracted similarity–dissimilarity information has significant potential for applications in the subgroup discovery of materials. The effectiveness of the eRSM in detecting homogenous subgroups of materials has been demonstrated by using two experiments on two datasets of magnetic materials. In addition, further analysis of the detected subgroups improves the existing knowledge of problems related to the applied datasets of magnetic materials. For example, we reveal the differences in the mechanisms of the Curie temperature of Co-based binary alloys when using our method to a dataset of 100 transition rare-earth metal binary alloys comprising Ni, Mn, Co, and Fe. Moreover, we explored the mechanisms of ferrimagnetic and low Curie temperature alloys from the magnetic dataset of calculated quaternary alloys. By measuring the similarity between materials with uncertainty, the method described herein is expected to extract valuable information for describing and interpreting the underlying physical mechanisms in material datasets.

SUPPLEMENTARY MATERIAL

See the supplementary material for the following additional information: (1) explanation of the formulation modeling uncertainty, (2) evaluation of the eRSM using the toy datasets, and (3) feature selection and pre-analysis in the dataset of quaternary high-entropy alloys.

ACKNOWLEDGMENTS

This work was supported by the Ministry of Education, Culture, Sports, Science, and Technology of Japan (MEXT) with the Program for Promoting Research on the Supercomputer Fugaku (DPMSD), JSPS KAKENHI grants 20K05301, JP19H05815 (Grants-in-Aid for Scientific Research on Innovative Areas Interface Ionics), 21K14396 (Grant-in-Aid for Early- Career Scientists), and 20K05068, Japan.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Minh-Quyet Ha: Conceptualization (equal); Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Validation (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal). **Duong-Nguyen Nguyen:** Conceptualization (equal); Formal analysis (supporting); Investigation (supporting); Methodology (supporting); Writing – original draft (supporting). **Viet Cuong Nguyen:** Funding acquisition (credit); Resources (credit); Software (credit). **Hiori Kino:** Data curation (lead); Formal analysis (supporting); Investigation (supporting); Methodology (equal); Validation (equal); Writing – original draft (equal); Writing – review & editing (equal). **Yasunobu Ando:** Formal analysis (supporting); Methodology (supporting); Writing – review & editing (supporting). **Takashi Miyake:** Formal analysis (supporting); Methodology (supporting); Validation (supporting); Writing – original draft (supporting); Writing – review & editing (supporting). **Thierry Denoeux:** Formal analysis (credit); Methodology (credit); Writing – review & editing (credit). **Van-Nam Huynh:** Conceptualization (supporting); Formal analysis (supporting); Investigation (supporting); Methodology (supporting); Writing – review & editing (supporting). **Hieu-Chi Dam:** Conceptualization (equal); Data curation (equal); Formal analysis (equal); Funding acquisition (equal); Investigation (equal); Methodology (equal); Project administration (equal); Resources (equal); Supervision (equal); Validation (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal).

DATA AVAILABILITY

The data that support the findings of this study are openly available in Zenodo at http://doi.org/10.5281/zenodo.7540840, Ref. 27.

REFERENCES

[1] B. Kailkhura, B. Gallagher, S. Kim, A. Hiszpanski, and T. Y.-J. Han, "Reliable and explainable machine-learning methods for accelerated material discovery," npj Comput. Mater. **5**, 108 (2019).

[2] J. Tenenbaum, "Learning the structure of similarity," Adv. Neural Inf. Process. Syst. **8**, 3–9 (1995).

[3] J. Tenenbaum, V. Silva, and J. Langford, "A global geometric framework for nonlinear dimensionality reduction," Science **290**, 2319–2323 (2000).

[4] Y. Yang, F. Liang, S. Yan, Z. Wang, and T. S. Huang, "On a theory of nonparametric pairwise similarity for clustering: Connecting clustering to classification," Adv. Neural Inf. Process. Syst. **27**, 145–153 (2014).

[5] C. Chen, O. Li, D. Tao, A. Barnett, C. Rudin, and J. K. Su, "This looks like that: Deep learning for interpretable image recognition," in *Advances in Neural Information Processing Systems*, edited by H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (Curran Associates, Inc., 2019), Vol. 32.

[6] B. Letham, C. Rudin, T. H. McCormick, and D. Madigan, "Interpretable classifiers using rules and Bayesian analysis: Building a better stroke prediction model," Ann. Appl. Stat. **9**, 1350–1371 (2015).

[7] C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," Nat. Mach. Intell. **1**, 206–215 (2019).

[8] B. R. Goldsmith, M. Boley, J. Vreeken, M. Scheffler, and L. M. Ghiringhelli, "Uncovering structure-property relationships of materials by subgroup discovery," New J. Phys. **19**, 013031 (2017).

[9] R. Ramprasad, R. Batra, G. Pilania, A. Mannodi-Kanakkithodi, and C. Kim, "Machine learning in materials informatics: Recent applications and prospects," npj Comput. Mater. **3**, 54 (2017).

[10] D.-N. Nguyen, T.-L. Pham, V.-C. Nguyen, T.-D. Ho, T. Tran, K. Takahashi, and H.-C. Dam, "Committee machine that votes for similarity between materials," IUCrJ **5**, 830–840 (2018).

[11] D.-N. Nguyen, T.-L. Pham, V.-C. Nguyen, H. Kino, T. Miyake, and H.-C. DAM, "Ensemble learning reveals dissimilarity between rare-earth transition binary alloys with respect to the Curie temperature," J. Phys.: Mater. **2**, 034009 (2019).

[12] J. Bardeen, L. N. Cooper, and J. R. Schrieffer, "Theory of superconductivity," Phys. Rev. **108**, 1175–1204 (1957).

[13] A. Seko, A. Togo, and I. Tanaka, "Descriptors for machine learning of materials data," in *Nanoinformatics*, edited by I. Tanaka (Springer Singapore, Singapore, 2018), pp. 3–23.

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

[15] R. L. Goldstone, D. L. Medin, and J. Halberstadt, "Similarity in context," Mem. Cognit. **25**, 237–255 (1997).

[16] G. Shafer, *A Mathematical Theory of Evidence* (Princeton University Press, 1976).

[17] T. Denœux, D. Dubois, and H. Prade, "Representations of uncertainty in artificial intelligence: Beyond probability and possibility," in *A Guided Tour of Artificial Intelligence Research*, edited by P. Marquis, O. Papini, and H. Prade (Springer-Verlag, 2020), Vol. 1, Chap. 4, pp. 119–150.

[18] A. P. Dempster, "Upper and lower probabilities induced by a multivalued mapping," Ann. Math. Stat. **38**, 325–339 (1967).

[19] N. Nu Thanh Ton, M.-Q. Ha, T. Ikenaga, A. Thakur, H.-C. Dam, and T. Taniike, "Solvent screening for efficient chemical exfoliation of graphite," 2D Mater. **8**, 015019 (2020).

[20] M.-Q. Ha, D.-N. Nguyen, V.-C. Nguyen, T. Nagata, T. Chikyow, H. Kino, T. Miyake, T. Denœux, V.-N. Huynh, and H.-C. Dam, "Evidence-based recommender system for high-entropy alloys," Nat. Comput. Sci. **1**, 470–478 (2021).

[21] C. Williams and C. Rasmussen, "Gaussian processes for regression," in *Advances in Neural Information Processing Systems 8*, Max-Planck-Gesellschaft (MIT Press, Cambridge, MA, 1996), pp. 514–520.

[22] M. Lázaro-Gredilla, S. Van Vaerenbergh, and N. D. Lawrence, "Overlapping mixtures of Gaussian processes for the data association problem," Pattern Recognit. **45**, 1386–1395 (2012).

[23] V. Vovk, "Kernel ridge regression," in *Empirical Inference* (Springer, 2013), pp. 105–116.

[24] L. Breiman, "Random forests," Mach. Learn. **45**, 5–32 (2001).

[25] A. Jain, J. Mao, and K. Mohiuddin, "Artificial neural networks: A tutorial," Computer **29**, 31–44 (1996).

[26] P. Smets, "Belief functions: The disjunctive rule of combination and the generalized Bayesian theorem," Int. J. Approx. Reason. **9**, 1–35 (1993).

[27] H.-C. Dam (2023). "Datasets of binary and quaternary alloys with Curie temperature and magnetization for the eRSM," Zenodo. http://doi.org/10.5281/zenodo.7540840.

[28] P. Villars, M. Berndt, K. Brandenburg, K. Cenzual, J. Daams, F. Hulliger, T. Massalski, H. Okamoto, K. Osaki, A. Prince, H. Putz, and S. Iwata, "The Pauling File, binaries edition," J. Alloys. Compd. **367**, 293–297 (2004).

[29] Y. Xu, M. Yamazaki, and P. Villars, "Inorganic materials database for exploring the nature of material," Jpn. J. Appl. Phys. **50**, 11RH02 (2011).

[30] H. C. Dam, V. C. Nguyen, T. L. Pham, A. T. Nguyen, K. Terakura, T. Miyake, and H. Kino, "Important descriptors and descriptor groups of Curie temperatures of rare-earth transition-metal binary alloys," J. Phys. Soc. Jpn. **87**, 113801 (2018).

[31] T. Fukushima, H. Akai, T. Chikyow, and H. Kino, "Automatic exhaustive calculations of large material space by Korringa-Kohn-Rostoker coherent potential approximation method applied to equiatomic quaternary high entropy alloys," Phys. Rev. Mater. **6**, 023802 (2022).

[32] Y.-Y. Ahn, J. P. Bagrow, and S. Lehmann, "Link communities reveal multiscale complexity in networks," Nature **466**, 761–764 (2010).

[33] D. Reynolds, "Gaussian Mixture Models," in *Encyclopedia of Biometrics*, edited by S. Z. Li and A. K. Jain (Springer, Boston, MA, 2015).

[34] A. H. Murphy, "The Finley affair: A signal event in the history of forecast verification," Weather Forecast. **11**, 3–20 (1996).

[35] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," J. Mach. Learn. Res. **9**, 2579–2605 (2008).

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

[37] H. Akai, M. Akai, S. Blügel, B. Drittler, H. Ebert, K. Terakura, R. Zeller, and P. H. Dederichs, "Theory of hyperfine interactions in metals," Prog. Theor. Phys. Suppl. **101**, 11–77 (1990).