Scripta Materialia 208 (2022) 114335

Contents lists available at ScienceDirect

Scripta Materialia

journal homepage: www.elsevier.com/locate/scriptamat

Acceleration of phase diagram construction by machine learning incorporating Gibbs' phase rule

Kei Terayama[a,b,c,d,*], Kwangsik Han[e], Ryoji Katsube[f], Ikuo Ohnuma[e], Taichi Abe[e,g], Yoshitaro Nose[h], Ryo Tamura[d,g,i,j,*]

[a] Graduate School of Medical Life Science, Yokohama City University, 1-7-29, Suehiro-cho, Tsurumi-ku, Kanagawa 230-0045, Japan
[b] Graduate School of Medicine, Kyoto University, 53 Shogoin-Kawaharacho, Sakyo-ku, Kyoto 606-8507, Japan
[c] Medical Sciences Innovation Hub Program, RIKEN Cluster for Science, Technology and Innovation Hub, Tsurumi-ku, Kanagawa 230-0045, Japan
[d] RIKEN Center for Advanced Intelligence Project, 1-4-1 Nihonbashi, Chuo-ku, Tokyo 103-0027, Japan
[e] Research Center for Structural Materials, National Institute for Materials Science, 1-2-1 Sengen, Tsukuba, Ibaraki 305-0047, Japan
[f] Department of Materials Science and Engineering, Kyoto University, Yoshida Honmachi, Sakyo-ku, Kyoto 606-8501, Japan
[g] Research and Services Division of Materials Data and Integrated System, National Institute for Materials Science, 1-1 Namiki, Tsukuba, Ibaraki 305-0044, Japan
[h] Department of Materials Science and Engineering, Kyoto University, Yoshida Honmachi, Sakyo-ku, Kyoto 606-8501, Japan
[i] International Center for Materials Nanoarchitectonics (WPI-MANA), National Institute for Materials Science, 1-1 Namiki, Tsukuba, Ibaraki 305-0044, Japan
[j] Graduate School of Frontier Sciences, The University of Tokyo, 5-1-5 Kashiwa-no-ha, Kashiwa, Chiba 277-8561, Japan

ARTICLE INFO

Article history:
Received 1 June 2021
Revised 4 October 2021
Accepted 7 October 2021

Keywords:
Phase diagram
Machine learning
Alloys
Analytical methods

ABSTRACT

To efficiently construct phase diagrams of alloy systems, a machine learning-based method advanced by thermodynamics on phase equilibria is proposed. With the use of uncertainty sampling in active learning, the next point to be synthesized or measured can be recommended to efficiently draw the phase diagram. For appropriate recommendations, two ingenuities are introduced in the machine learning method: training data preparation when the multiphase coexisting region is detected and search space reduction based on the Gibbs' phase rule. We demonstrate the construction of ternary phase diagrams using our machine learning method by incorporating these ingenuities. The complicated phase diagram of alloy systems could be effectively plotted even when knowing only the information of single-component systems in the initial step. The recommendation made by our machine learning method can help reduce the number of experiments required to construct a phase diagram to approximately 1/8 compared with random sampling.

Phase diagrams are indispensable for the development of materials. In alloy systems, phase diagrams help understand the dependency of a microstructure in specific phases on the temperature, pressure, and composition. Knowing the phase diagram of a target alloy system, the microstructures of the alloy can be controlled, and desired material properties can be obtained for industrial applications. To date, many binary phase diagrams have been investigated, and there is much information regarding the phase equilibria of binary systems [1]. However, even for binary systems, determining the phase equilibria requires a significant amount of time and cost. Further, the superalloys [2,3] and aluminum alloys [4,5] in practical use are becoming more complicated, and low-, medium-, and high-entropy alloys contain many types of elements [6–8]. For such complicated alloys, experimental data are limited, and there is no clarity regarding their phase diagram. It is difficult to experimentally determine the phase diagram without specific guidelines, particularly for a complicated alloy system.

In recent years, the material science field has witnessed increasing application of machine learning [9–18]. For example, the black-box optimization method represented by Bayesian optimization has been extensively applied, and materials with desired properties have been prepared with as few experiments as possible [19–21]. In this optimization scheme, a candidate material is iteratively recommended through machine-learning predictions. Inspired by these advancements, some machine learning methods have been proposed for phase diagrams [22–26]. As a black-box optimization scheme for phase diagrams, we developed a method

* Corresponding authors at: Graduate School of Medical Life Science, Yokohama City University, 1-7-29, Suehiro-cho, Tsurumi-ku, Kanagawa 230-0045, Japan and International Center for Materials Nanoarchitectonics (WPI-MANA), National Institute for Materials Science, 1-1 Namiki, Tsukuba, Ibaraki 305-0044, Japan.
E-mail addresses: terayama@yokohama-cu.ac.jp (K. Terayama), tamura.ryo@nims.go.jp (R. Tamura).

https://doi.org/10.1016/j.scriptamat.2021.114335

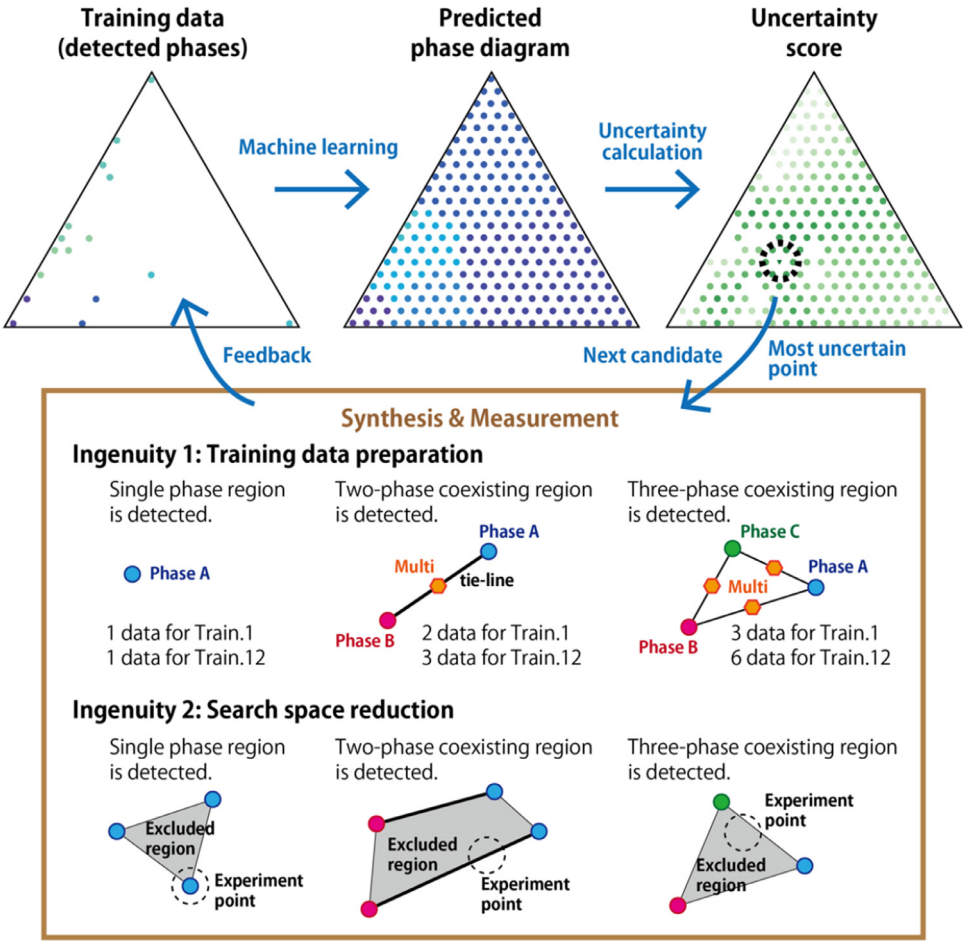

Fig. 1. Flow of phase diagram construction using machine learning. The next candidate condition with the highest uncertainty is recommended by our machine learning method. At the synthesis and measurement parts of the selected condition, the two ingenuities are introduced for alloy systems. The first is the training data preparation when multiphases are detected. The second is the search space reduction based on the degree of freedom evaluated by the phase rule.

using uncertainty sampling (US) called phase diagram construction (PDC) [22].

In this method, the most uncertain point in the phase diagram is recommended for the subsequent experiment (see Fig. 1). In the following, we detail the steps involved in this method. (i) Initialization: The phase diagram to be drawn is discretized, and the position vectors of discretized points are input as $\{\mathbf{x}_i\}_{i=1,\dots,N}$. Among them, some points are randomly selected as initial data. Each phase of the selected points is identified by performing experiments, and $p = 1, \dots, P$ is the label of phases at each point when P types of phases are found. These are used as training data $D = \{\mathbf{x}_j, p_j\}_{j=1,\dots,M}$ when the number of initial data is M. (ii) Phase estimation: The probability distribution of the phases at each point is estimated by semi-supervised learning using label propagation (LP) or label spreading (LS) methods as $P(p|\mathbf{x})$ from the training data D. (iii) Uncertainty score: The uncertainty score is evaluated using $P(p|\mathbf{x})$; for example, one of the definitions is $u(\mathbf{x}) = 1 - \max_p P(p|\mathbf{x})$. The most uncertain point which maximizes the uncertainty score in the phase diagram is selected as the next candidate for experiments. (iv) Experiment: Experiments are performed according to the selected condition, and the phase is identified. If the identified phase is known, the label of this point is selected from $1, \dots, P$, while if a new phase is detected, a new label of $P + 1$ is attached. The training dataset is increased as $D = \{\mathbf{x}_j, p_j\}_{j=1,\dots,M+1}$. Steps (ii)-(iv) are iterated to obtain a better phase diagram. In general, the uncertain points in the phase diagram are concentrated near the phase boundary or are far from the already synthesized points. In Ref. [22], for the well-known diagrams such as liquidus projection, the performance of our method was demonstrated, and we confirmed that the phase diagram could be effectively obtained. Furthermore, the US method was applied to construct new diagrams to obtain Zn-Sn-P films by molecular beam epitaxy [27] and to design block copolymers by the self-consistent field theory [28]. Thus, the US method is a proven tool for drawing phase diagrams. The code of PDC is available at https://github.com/tsudalab/PDC.

On the contrary, equilibrium phase diagrams of alloy systems are more complicated due to regions of multiphase (phase coexistence) in addition to single-phase regions, and then how to deal with such regions is key when phase diagrams are successful constructed using PDC method. Therefore, in this study, we address a strategy to prepare the training data when multiphase coexisting region is detected (see Fig. 1). In addition, a phase diagram is constructed according to the Gibbs' phase rule [29]. Therefore, the search space can be narrowed down by utilizing the degree of freedom based on the phase rule, resulting in an accurate phase diagram can be obtained with fewer cycles. In this study, we only focused on ternary phase diagrams with a fixed temperature and pressure as the first step. These ideas can be applied to high-dimensional phase diagrams in a straightforward manner.

In this study, we introduce two ingenuities to efficiently construct the phase diagram of alloy systems. The first is training data preparation when multiphases are detected. The second is search space reduction based on the phase rule. First, we consider the

first ingenuity. Depending on whether a single-phase region or a multiphase coexisting region is detected, the treatment of new data points added to the training dataset should be changed. If a single-phase region is detected, the type of single phase at the recommended point can only be obtained experimentally. Therefore, only one training data point is added. On the other hand, when a multiphase coexisting region is detected, information in addition to the recommended point is acquired. For example, in the case of two-phase coexistence, the endpoints of the tie line are the corresponding single phases (see Fig. 1), and a two-phase coexisting region appears along the tie line. In a ternary phase diagram, the three-phase coexisting region forms a triangle, with its vertices representing the three different single phases (see Fig. 1). The sides of the triangle are two-phase coexisting regions. Thus, there are several ways in which the data points from the rich information can be added to the training data. In this study, two strategies were considered.

The first strategy is focused on only type of single phase for the training data set, and phase estimation is performed under the condition that regions of multiphase are excluded. This is based on the idea that an effective search of the three-phase coexisting regions is key to drawing the phase diagram. These regions will be located at the uncertain point where the possibilities of three regions of single phase are stacked, and appropriate selection will be realized when only single phases are considered. The proposed approach, called single-phase training ("Train.1"), for preparing the training dataset is described as follows. (i) *Detecting single-phase region:* The type of single phase at the recommended point is added to the training dataset. Thus, one data point is added. (ii) *Detecting two-phase coexisting region:* The types of single phases at the end points of the tie-line are added to the training dataset. Thus, two data points are added. (iii) *Detecting three-phase coexisting region:* The types of single phases at the vertices of the three-phase coexisting region are added to the training dataset. Thus, three data points are added.

In the second strategy, information regarding the multiphase is included in the training. The representative points in the two-phase coexisting region are used as the training data, and the reason is as follows. If all the points in the detected coexisting regions are added to the training dataset, the number of the training data points corresponding to the multiphase will be too high. In such cases, from the characteristics of the LP and LS methods, the coexisting region will be too large, and the regions of single-phase are erased from the predicted diagram (see Supplementary note A). This would be inefficient for appropriate recommendations. Thus, the following approach, called single and multiphase training ("Train.12"), is considered for preparing the training data. (i) *Detecting single-phase region:* The type of single phase at the recommended point is added to the training data. Thus, one data point is added. (ii) *Detecting two-phase coexisting region:* The types of single phases at the end points of the tie-line are added to the training dataset. In addition, the midpoint of the tie line is assigned to the multiphase in the training dataset (e.g., the cross point in Fig. 1). Here, regardless of the type of multiphases, the same label is attached. Thus, three data points are added. (iii) *Detecting three-phase coexistence:* The types of single phases at the vertices of the three-phase coexisting region are added to the training dataset. In addition, the midpoints of the sides of the triangular region are assigned to the multiphase (e.g., cross points in Fig. 1). Here, regardless of the types of multiphases, the same label is attached, which is common for detecting the two-phase coexisting region. Thus, six data points are added. In the training data, the types of detected single phases and one label of multiphase are included.

Next, we explain the idea of reducing the search space in the phase diagram based on the Gibbs' phase rule. For the equilibrium phase diagram, the phase rule was introduced by Gibbs, and the degree of freedom at each point when the temperature and pressure are fixed can be defined by

$$F = C - P, \tag{1}$$

where C is the number of components, and P is the number of phases. In this study, we focused on a ternary phase diagram, in which case $C = 3$. Thus, the degrees of freedom are as follows: $F = 2$ in the single-phase region ($P = 1$), $F = 1$ in the two-phase coexisting region ($P = 2$), and $F = 0$ in the three-phase coexisting region ($P = 3$). In the three-phase coexisting region, a new phase cannot appear. Thus, it is not necessary to search in this region. However, the degree of freedom remains in the single-phase and two-phase coexisting regions, which cannot be excluded from the search space because the phase boundary is not determined. However, to rapidly construct an outline of the phase diagram, the appropriate exclusion of the search space should be made effectively in our machine learning approach. In this study, we consider the following three types of exclusion strategies based on the phase rule. (i) *Exclusion of three-phase coexisting region ("Ex.3")*: The inside of the three-phase coexisting region is excluded from the search space (see Fig. 1). This excluded area has $F = 0$. (ii) *Exclusion of two- and three-phase coexisting regions ("Ex.23")*: In addition to (i), the region surrounded by the same tie lines is excluded (see Fig. 1). These excluded areas have $F = 0$ or 1. (iii) *Exclusion of single-phase, and two- and three-phase coexisting regions ("Ex.123")*: In addition to (ii), the region surrounded by the straight lines connecting to the same types of single-phase points is excluded (see Fig. 1). These excluded areas have $F = 0$, 1, and 2. Here, the possibility of new phases appearing in the excluded region is the highest in (iii) and is zero in (i). Conversely, the area of the excluded region will increase in the order of (iii), (ii), and (i).

We demonstrate our machine learning technique by applying it to construct the ternary phase diagrams of Cr–Co–Ni, Co–Cr–Mn, Fe–Co–Cr, and Cr–Mn–Ni systems. These phase diagrams were obtained using the Thermo-Calc software with the TCHEA3 database [30] and are shown in Fig. 2. The numbers of phases are summarized in Table 1. Cr–Co–Ni has the simplest one, whereas Cr–Mn–Ni has the most complicated one. Note that these ternary systems are subsystems of high-entropy CrMnFeCoNi (Cantor) alloys [31,32].

To make the construction of phase diagrams more challenging, we consider the case in which only the phase information of single-component systems is known in the initial step. In other words, the information for the binary alloys is not used. To evaluate the performance, the metric of the remaining area D_r is introduced, defined as the area remaining after subtracting the area by Ex.123 from the entire area. D_r is normalized, and the value is 1 when the experimental results are nonexistent and 0 if all the phases are filled from the viewpoint of Ex.123. Fig. 3 shows the value of D_r with respect to the iteration step. Fig. 4 shows the number of three-phase coexisting regions detected with respect to the iteration step. Here, we compare the results where combinations of two training strategies (Train.1 and Train.12) and three exclusion strategies (Ex.3, Ex.23, and Ex.123) were used. Furthermore, the results of random sampling, where the next recommendation is randomly generated in each iteration without training or exclusion of the search space, are also shown. In these figures, the LP method is used for training, and the uncertainty score is evaluated using the least confident (LC) method. In Supplementary note B, the definitions of the uncertainty score are given, and the results of other methods are shown for Cr–Co–Ni (simplest) and Cr–Mn–Ni (most complicated) cases.

In all the cases, D_r is more rapidly decreased by Train.12 than Train.1. This means that information regarding the multiphase is required in the training to efficiently construct the phase diagram. As shown in Supplementary Movies 1–4, by Train.1, the recom-

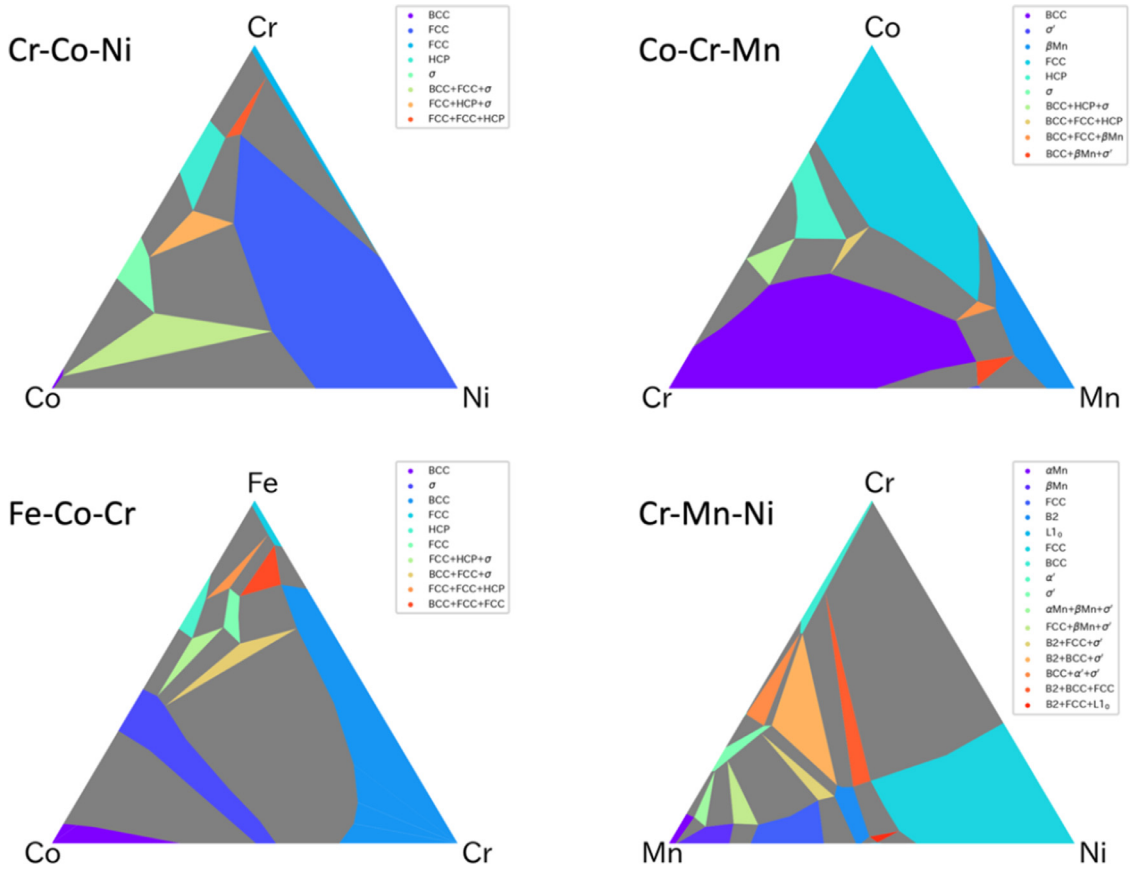

Fig. 2. Target ternary phase diagrams obtained using the Thermo-Calc software. The gray areas represent the two-phase coexisting regions.

Table 1
Numbers of single-phase region two- and three-phase coexisting regions in the respective ternary phase diagrams of alloys.

	Cr-Co-Ni	Co-Cr-Mn	Fe-Co-Cr	Cr-Mn-Ni
# of single-phase region	5	6	6	9
# of two-phase coexisting region	7	9	9	15
# of three-phase coexisting region	3	4	4	7
# of total phases	15	19	19	31

mended points are intensively selected in the two-phase coexisting regions, and it is difficult to detect new phases. Contrarily, even if Train.1 is used, the performance is dramatically improved when exclusion strategies Ex.23 and Ex.123 are applied instead of Ex.1.

The most natural exclusion strategy is Ex.3, and if Tran.12 is adopted as the training strategy, the efficiency is higher than that of random sampling in terms of D_r. By focusing on the number of detected three-phase coexisting regions, a satisfactory performance can be achieved, except in the case of Fe–Co–Cr system. Thus, we conclude that our machine learning method can be used for constructing phase diagrams if appropriate training and exclusion strategies are implemented. However, to rapidly determine the outline, the strategies Train.12 & Ex.23 and Train.12 & Ex.123 are more suitable. If these strategies are adopted, a complete phase diagram can be obtained in approximately 40 experiments with the help of machine learning. When the random sampling is adopted, over 300 experiments are necessary. Thus, the recommendation made by our machine learning method can help reduce the number of experiments required to construct a phase diagram to approximately 1/8 compared with random sampling. Notice that if the information regarding the binary systems is included as initial data, the results can be much faster. For Cr–Co–Ni and Cr–Mn–Ni systems, the phase diagram can be obtained in less than 15 experiments (see Supplementary note C).

In conclusion, the consideration of multiphase for training phase diagrams was discussed for an efficient recommendation by machine learning. The appropriate exclusion of the search space based on the Gibbs' phase rule was addressed. We demonstrated that if appropriate training for the multiphase and exclusion strategies is used, machine learning can help effectively construct the phase diagram of complex alloy systems. In a previous study [22], we demonstrated the performance of our method for the phase diagrams with only single phase regions. In this case, the exclusion strategy based on the phase rule should be effective, and the results can be obtained with fewer cycles.

Our approach can be extended to multicomponent alloys and temperature-dependent phase diagrams in a straightforward manner. When such difficult phase diagram will be constructed, including Gibbs' phase rules, it will be useful to incorporate knowledge of various rules and equations related to phase diagrams such as Clausius-Clapeyron equation when generating the training data. Furthermore, our approach can serve as a powerful tool for difficult-to-conceive higher dimensions. We believe that our ma-

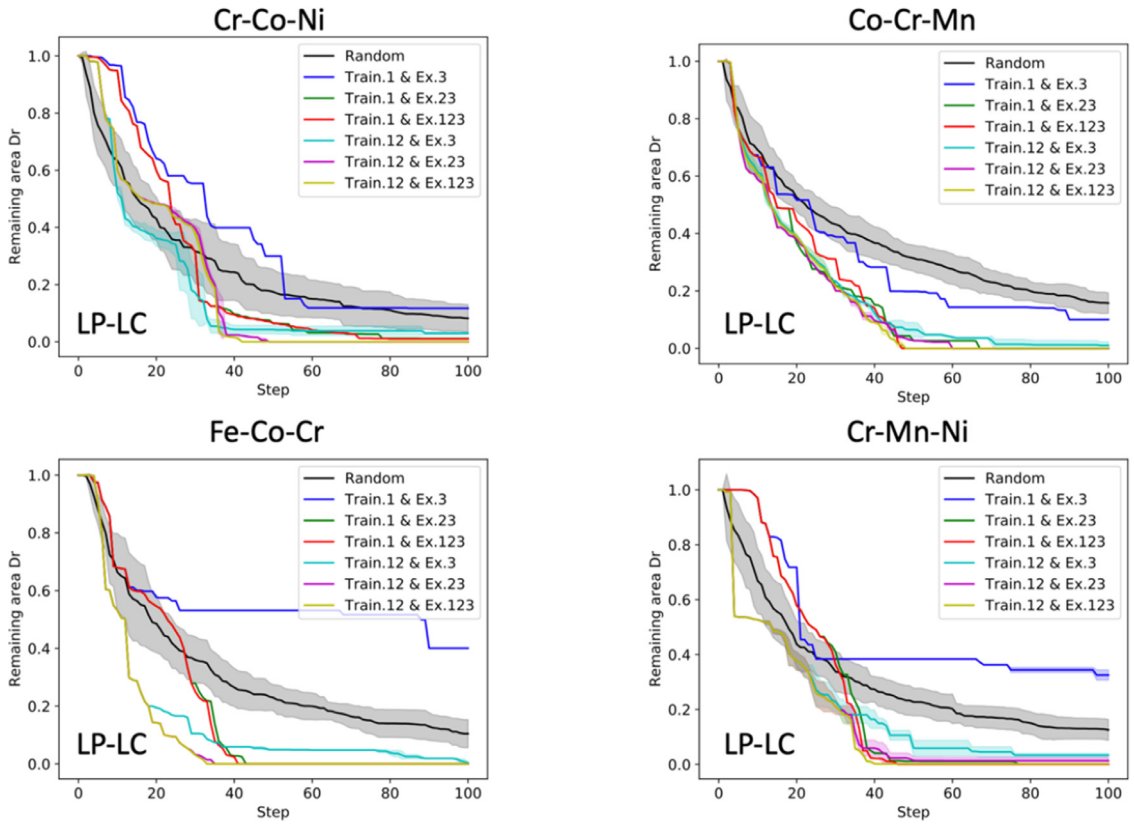

Fig. 3. Remaining area D_r plotted against the iteration step for the two training strategies and three exclusion strategies. For the training strategy, single-phase training (Train.1) and single-phase and multiphase training (Train.12) are considered. To exclude the search space based on the phase rule, exclusion of three-phase regions (Ex.3); exclusion of two- and three-phase regions (Ex.23); exclusion of single-, two-, and three-phase regions (Ex.123) are used. Training is performed by LP, and the uncertainty score is evaluated by LC.

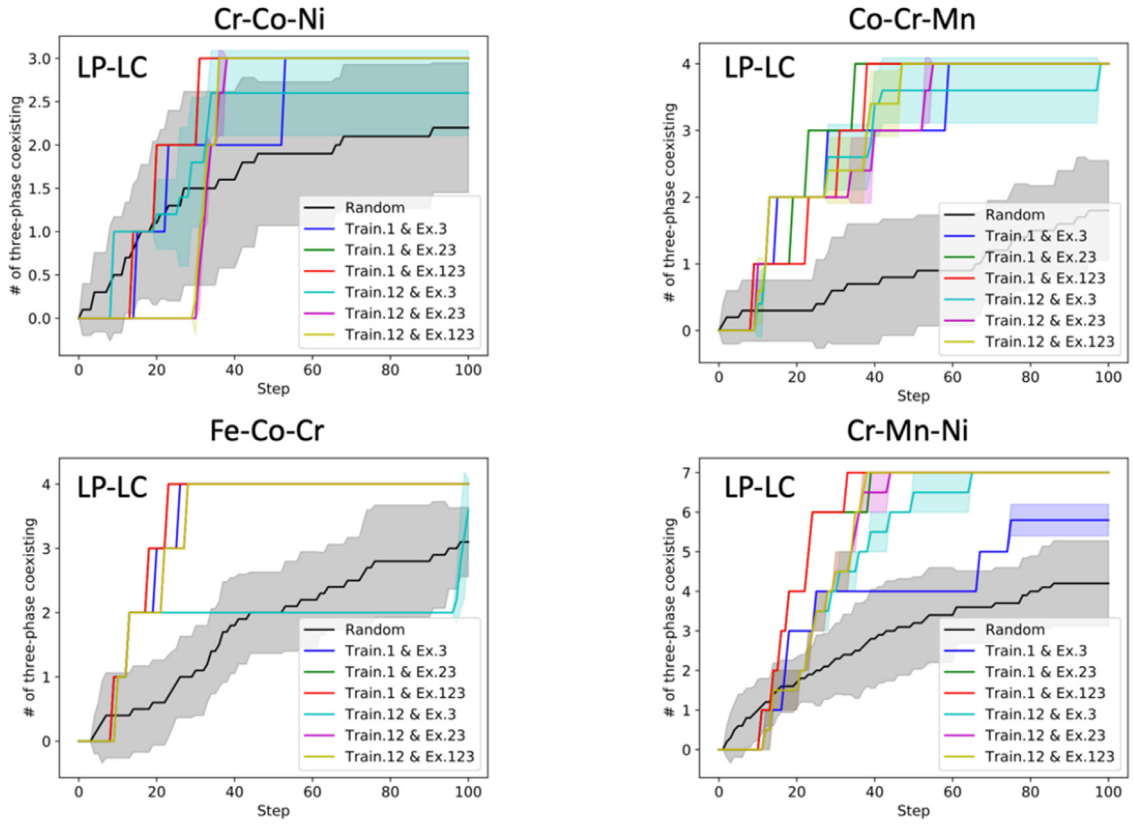

Fig. 4. Number of detected three-phase coexisting regions with respect to the iteration step when using the two training strategies and three exclusion strategies.

chine learning technique can help researchers create phase diagrams for uninvestigated alloy systems.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

We would like to acknowledge Koji Tsuda, Fumiyasu Oba, Hiroki Taniguchi, and Hidenori Hiramatsu for fruitful discussions. This study was supported by a project subsidized by Core Research for Evolutional Science and Technology (CREST) (Grant No. JPMJCR17J2) from the Japan Science and Technology Agency (JST). We would like to thank Editage (www.editage.com) for English language editing.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.scriptamat.2021.114335.

References

[1] T.B. Massalski, H. Okamoto, Binary Alloy Phase Diagrams, ASM International, 1990.
[2] Y.F. Gu, T. Osada, T. Yokokawa, H. Harada, J. Fujioka, D. Nagahama, M. Okuno, in: Superalloys, John Wiley & Sons, Ltd, 2016, pp. 209–216.
[3] W.E. Frazier, J. Mater. Eng. Perform. 23 (2014) 1917–1928.
[4] J.R. Davis, Aluminum and Aluminum Alloys, ASM International, 1993.
[5] J. Hirsch, B. Skrotzki, G. Gottstein, Aluminium Alloys: The Physical and Mechanical Properties, John Wiley & Sons, 2008.
[6] M.H. Tsai, J.W. Yeh, Mater. Res. Lett. 2 (2014) 107–123.
[7] D.B. Miracle, O.N. Senkov, Acta Mater. 122 (2017) 448–511.
[8] E.P. George, D. Raabe, R.O. Ritchie, Nat. Rev. Mater. 4 (2019) 515–534.
[9] L.M. Ghiringhelli, J. Vybiral, S.V. Levchenko, C. Draxl, M. Scheffler, Phys. Rev. Lett. 114 (2015) 105503.
[10] A. Seko, A. Togo, H. Hayashi, K. Tsuda, L. Chaput, I. Tanaka, Phys. Rev. Lett. 115 (2015) 205901.
[11] S. Ju, T. Shiga, L. Feng, Z. Hou, K. Tsuda, J. Shiomi, Phys. Rev. X 7 (2017) 021024.
[12] J.E. Gubernatis, T. Lookman, Phys. Rev. Mater. 2 (2018) 120301.
[13] V. Stanev, C. Oses, A.G. Kusne, E. Rodriguez, J. Paglione, S. Curtarolo, I. Takeuchi, NPJ Comput. Mater. 4 (2018) 1–14.
[14] T. Fukazawa, Y. Harashima, Z. Hou, T. Miyake, Phys. Rev. Mater. 3 (2019) 053807.
[15] M. Todorović, M.U. Gutmann, J. Corander, P. Rinke, NPJ Comput. Mater. 5 (2019) 1–7.
[16] R. Tamura, M. Watanabe, H. Mamiya, K. Washio, M. Yano, K. Danno, A. Kato, T. Shoji, Sci. Technol. Adv. Mater. 21 (2020) 540–551.
[17] K. Kitai, J. Guo, S. Ju, S. Tanaka, K. Tsuda, J. Shiomi, R. Tamura, Phys. Rev. Res. 2 (2020) 013319.
[18] K. Terayama, M. Sumita, R. Tamura, K. Tsuda, Acc. Chem. Res. 54 (2021) 1334–1346.
[19] K. Homma, Y. Liu, M. Sumita, R. Tamura, N. Fushimi, J. Iwata, K. Tsuda, C. Kaneta, J. Phys. Chem. C 124 (2020) 12865–12870.
[20] R. Tamura, T. Osada, K. Minagawa, T. Kohata, M. Hirosawa, K. Tsuda, K. Kawagishi, Mater. Des. 198 (2021) 109290.
[21] I. Ohkubo, Z. Hou, J.N. Lee, T. Aizawa, M. Lippmaa, T. Chikyow, K. Tsuda, T. Mori, Mater. Today Phys. 16 (2021) 100296.
[22] K. Terayama, R. Tamura, Y. Nose, H. Hiramatsu, H. Hosono, Y. Okuno, K. Tsuda, Phys. Rev. Mater. 3 (2019) 033802.
[23] K. Terayama, K. Tsuda, R. Tamura, Jpn. J. Appl. Phys. 58 (2019) 098001.
[24] P. Nguyen, T. Tran, S. Gupta, S. Rana, M. Barnett, S. Venkatesh, in: Proceedings of the SIAM International Conference on Data Mining (SDM), Society for Industrial and Applied Mathematics, 2019, pp. 549–557.
[25] C. Dai, S.C. Glotzer, J. Phys. Chem. B 124 (2020) 1275–1284.
[26] Y. Tian, R. Yuan, D. Xue, Y. Zhou, Y. Wang, X. Ding, J. Sun, T. Lookman, Adv. Sci. 8 (2021) 2003165.
[27] R. Katsube, K. Terayama, R. Tamura, Y. Nose, ACS Mater. Lett. 2 (2020) 571–575.
[28] S. Zhao, T. Cai, L. Zhang, W. Li, J. Lin, ACS Macro Lett. 10 (2021) 598–602.
[29] K.B. Krauskopf, D.K. Bird, Introduction to Geochemistry, McGraw-Hill, 1995.
[30] https://Thermocalc.com/.
[31] B. Cantor, I.T.H. Chang, P. Knight, A.J.B. Vincent, Mater. Sci. Eng. A 375–377 (2004) 213–218.
[32] J.W. Yeh, S.K. Chen, S.J. Lin, J.Y. Gan, T.S. Chin, T.T. Shun, C.H. Tsau, S.Y. Chang, Adv. Eng. Mater. 6 (2004) 299–303.