

Machine-learning for designing nanoarchitected materials by dealloying

C. Zhao

To be published in "Communications Materials"

November 2022

Photon Sciences

Brookhaven National Laboratory

U.S. Department of Energy









USDOE Office of Science (SC), Basic Energy Sciences (BES) (SC-22)

Notice: This manuscript has been authored by employees of Brookhaven Science Associates, LLC under Contract No.DE-SC0012704 with the U.S. Department of Energy. The publisher by accepting the manuscript for publication acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, nor any of their contractors, subcontractors, or their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or any third party's use or the results of such use of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof or its contractors or subcontractors. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

Machine-learning for designing nanoarchitected materials by dealloying

Chonghang Zhao^{1,2}[✉], Cheng-Chu Chung¹, Siying Jiang³, Marcus M. Noack⁴, Jiun-Han Chen⁵, Kedar Manandhar⁶, Joshua Lynch², Hui Zhong⁷, Wei Zhu³, Phillip Maffettone², Daniel Olds², Masafumi Fukuto², Ichiro Takeuchi⁶, Sanjit Ghose², Thomas Caswell², Kevin G. Yager⁸ & Yu-chen Karen Chen-Wiegart^{1,2}[✉]

Machine learning-augmented materials design is an emerging method for rapidly developing new materials. It is especially useful for designing new nanoarchitected materials, whose design parameter space is often large and complex. Metal-agent dealloying, a materials design method for fabricating nanoporous or nanocomposite from a wide range of elements, has attracted significant interest. Here, a machine learning approach is introduced to explore metal-agent dealloying, leading to the prediction of 132 plausible ternary dealloying systems. A machine learning-augmented framework is tested, including predicting dealloying systems and characterizing combinatorial thin films via automated and autonomous machine learning-driven synchrotron techniques. This work demonstrates the potential to utilize machine learning-augmented methods for creating nanoarchitected thin films.

¹Department of Materials Science and Chemical Engineering, Stony Brook University, Stony Brook, NY 11794, USA. ²National Synchrotron Light Source II, Brookhaven National Laboratory, Upton, NY 11973, USA. ³Department of Applied Mathematics and Statistics, Stony Brook University, Stony Brook, NY 11794, USA. ⁴Applied Mathematics and Computational Research Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. ⁵Independent Researcher, Taipei 106, Taiwan. ⁶Department of Materials Science and Engineering, University of Maryland, College Park, MD 20742, USA. ⁷Department of Joint Photon Science Institute, Stony Brook University, Stony Brook, NY 11794, USA. ⁸Center for Functional Nanomaterials, Brookhaven National Laboratory, Upton, NY 11973, USA. ✉email: czhao@bnl.gov; Karen.Chen-Wiegart@stonybrook.edu

Nanoarchitecture materials such as nanoporous metals are versatile due to their unique properties including high surface-area-to-volume ratios, light weights, and high thermal and electrical conductivities^{1–3}. Dealloying is a promising method for fabricating nanoporous metals. During dealloying of a parent alloy (A-B), one or more components (a metal/alloy B) are removed with a dealloying agent (a metal/solution C), and the remaining components form a bicontinuous structure through a self-organizing process⁴. Recently, liquid metal dealloying (LMD), where liquid metal is used as the dealloying agent (metal C), was reintroduced to fabricate less noble nanoporous materials⁵, including stainless steel^{6,7}, silicon^{8,9}, magnesium¹⁰, graphite¹¹, α -titanium^{5,12}, β -titanium¹³, and TiVNbMoTa high entropy alloys¹⁴. Solid-state metal dealloying (SSMD), or solid-state interfacial dealloying (SSID), has been introduced to fabricate nanoporous Fe, Fe-Cr, and α -Ti with a finer ligament, which can be used to overcome the limitation of high fabrication temperatures and handle liquid metal difficulties in LMD^{15–18}. However, the fundamental mechanisms leading to metal-agent dealloying remain unclear, thereby creating challenges in defining a strategy for materials design using metal-agent dealloying.

The material design principle of LMD proposed by Wada et al.⁵ is primarily focused on the thermodynamic quantities of the differences in the mixing enthalpies between the parent alloy A-B (ΔH_{AB}^m) and the dealloying agent (C) mixed with one of the components from the parent alloy (ΔH_{AC}^m or ΔH_{BC}^m). The general guideline is to have a negative mixing enthalpy of BC ($\Delta H_{BC}^m < 0$) and $\Delta H_{BC}^m < \Delta H_{AB}^m$ to provide a driving force to dealloy B from AB, while utilizing a positive ΔH_{AC}^m so that mixing between AC would not be preferred. However, while such criteria have been widely used to design new material systems, they have not been consistent with experimental observations. LMD also occurs when the mixing enthalpy between the elements in parent alloys is more negative than the mixing enthalpy between the soluble element and the dealloying agent ($0 > \Delta H_{BC}^m > \Delta H_{AB}^m$), such as in TiVNbMoTa-Ni, C-Mn, Ti-Cu, and Nb-Ni systems^{11,14,19}. SSID was even reported in some systems with a positive mixing enthalpy between the soluble element in the parent alloy and the dealloying agent ($\Delta H_{BC}^m > 0$ and $\Delta H_{AB}^m > 0$)²⁰.

These contradictions could have resulted from one of the three factors or a combination of them. The first such factor is the inaccuracy of the mixing enthalpy value available. The widely used mixing enthalpy was proposed and calculated by the Miedema model²¹. Although the Miedema model provides a way to calculate the mixing enthalpies in binary alloy systems, its accuracy is limited when the mixing enthalpy is close to zero²². In binary systems containing nonmetals or semimetals, a large transformation energy calibration was needed²³. The second factor is that in considering the relative phase stability, the mixing enthalpy may not be the right parameter to use. Instead, analysis in the literature has shown that the thermodynamic stability of a material should be defined by its Gibbs energy of decomposition²⁴. The third factor is the entropy contribution to the Gibbs free energy. Conventionally, mixing enthalpy has been treated as the dominant quantity for determining Gibbs free energy. However, new studies indicate that configurational and vibrational entropies can contribute to stabilizing/destabilizing the solid solution²⁵. Overall, the conventional criteria of using the mixing enthalpy difference calculated from the Miedema model to determine if a system can be dealloyed should be re-examined to critically enable the fabrication of a wider range of elements based on sound design principles.

Using machine learning (ML) algorithms to facilitate new materials design is a rapidly growing research area. Complementing computational simulation and theoretical modeling

techniques, ML methods promise to efficiently predict the presence of novel materials. Recent developments in the field include a general ML framework for predicting inorganic materials²⁶, with a wide range of materials predicted. Depending on their composition, crystal structure, and microstructures, ML methods have been used to predict materials, such as high entropy alloys²⁷, shape memory alloys²⁸, and thin-film metallic glasses²⁹. Combinatorial thin-film synthesis has been used to achieve the preparation of a great number of material compositions that could not be realized previously. By combining combinatorial sample synthesis with synchrotron characterization, the correlation of material properties with composition to realize rapid materials discovery becomes feasible³⁰. A combination of iterative ML prediction and combinatorial synthesis has also been applied for the discovery of metallic glasses²⁹. Recently, an autonomous decision-making algorithm-driven scientific instrument was developed; it minimized the number of measurements needed, prevented the collection of redundant information, and thus optimized the utilization of experimental and computing resources^{31–33}. The Gaussian-process-based decision-making algorithm (gpCAM) was independently developed by CAMERA³⁴. The term “autonomous experiment” describes the combination of automatic decision-making, analysis, instrumentation, and communication tools to create a closed-loop that can make intelligent decisions during an experiment without human interaction^{33,35}. Here, during a synchrotron diffraction experiment, gpCAM makes decisions based on the diffraction intensity of the phase(s) of interest from the collected data, and autonomously drives the next diffraction measurements to probe regions with the highest uncertainty. Closed-loop concepts with machine learning prediction and an autonomous experiment were demonstrated³⁶, and the development of carbon nanotube growth was realized³⁷. The ability to design self-driving materials provides a promising pathway toward autonomously designing and synthesizing novel materials³⁸. In the field of dealloying research, data mining and automated image analysis were introduced to identifying coarsening mechanism³⁹.

In this study, we established a workflow for designing nanopores/nanocomposites, simultaneously exploiting the advances in ML prediction, combinatorial sample preparation, and autonomous synchrotron characterization. We applied several ML methods to predict new ternary dealloying systems by employing training and testing ML methods on randomly split published dealloying data based on systems that exhibit successful dealloying behaviors. We used ML ranked important variables to analyze the underlying materials design principles in the metal-agent dealloying method. In total, the ML models predicted 132 ternary dealloying systems from the selected 16 metal elements. A proof of concept for validating the ML method through thin-film SSID experiments was tested by dealloying a Ti-Cu alloy with Mg. Because the Ti-Cu/Mg system was reported previously⁴⁰ with an established preparation and characterization methods, it is thus an ideal system to demonstrate new materials design and characterization methodologies here. A combinatorial Ti-Cu thin film was prepared to cover a wide range of parent alloy compositions. Other processing parameters were analyzed, including the dealloying time, dealloying temperature, and amount of dealloying agent. We applied synchrotron X-ray diffraction (XRD) characterization to combinatorial dealloyed thin films to explore the kinetics in Ti-Cu/Mg dealloying systems, including the phase transition, crystallization, and parent alloy composition threshold for dealloying. In addition, the crystallographic companion agent (XCA) was used to automate the XRD analysis, and gpCAM for autonomous experimental control was used to drive XRD characterization at the synchrotron beamline for faster materials discovery. This helped in overcoming the complexity and

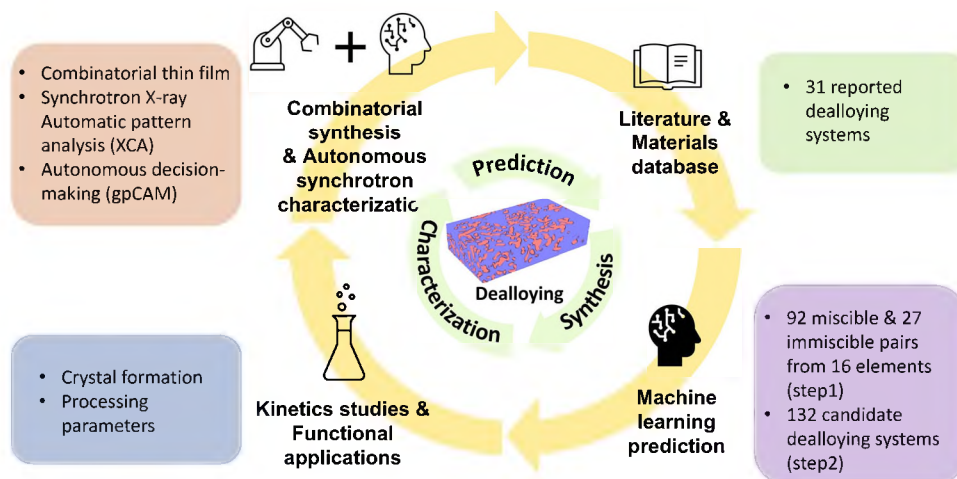


Fig. 1 Workflow of an ML-augmented framework to design SSID. The workflow includes the ML prediction of new ternary SSID systems (Xenonpy⁷⁶, pymatgen⁸², matminer⁷⁷, and literatures⁷⁸), combinatorial thin-film deposition of predicted systems, and autonomously driven synchrotron X-ray characterization (XCA⁵⁰, gpCAM³⁴, and bluesky⁸¹ libraries). The characterization result is returned to enrich the ML training datasets.

increasing the efficiency of the time-consuming validation methods for new materials with a large parameter space.

Results and discussion

Workflow of an ML-augmented framework to design SSID. The workflow includes the ML prediction of new ternary SSID systems, combinatorial thin-film deposition of predicted systems, and autonomously driven synchrotron X-ray characterization, as shown in Fig. 1.

For ML prediction, ML methods are first trained based on published dealloying systems from the literature and are then employed to predict new dealloying systems. The prediction involves classifying the elemental pairs into two categories: miscible and immiscible, composing the elemental pairs into potential ternary dealloying systems, and selecting the parent alloy and the dealloying agent from each of the potential ternary systems. Different ML methods lead to different results, and the majority vote from the three ML methods with equal weight is used towards ensemble prediction, which is then validated by the experiment. The variables selected by the ML methods are used to elucidate the design criteria in SSID thin films. Note that the current training set was built on both SSID and LMD methods, considering both methods are based on the selective dissolution using a metallic dealloying agent. In the future, further refined models that train these two sets of data separately can be developed with more experimental results available in both methods.

In the experiment, we used combinatorial thin-film deposition to prepare dealloying systems and characterized them with decision-making algorithm-driven autonomous synchrotron X-rays. Efficient data collection enables the analysis of the kinetics in multiple SSID thin-film systems, validation of the ML prediction results, and improvement in the ML methods by enriching the ML training sets with the experimental results. The potential functional application is also expected to be tested at the synchrotron after validating the dealloying in the predicted systems.

The experiment and subsequent enrichment of the ML training sets with experimental results are crucial for the following reasons: (1) the reported dealloying systems were mostly designed based on the mixing enthalpy difference calculated by Miedema's

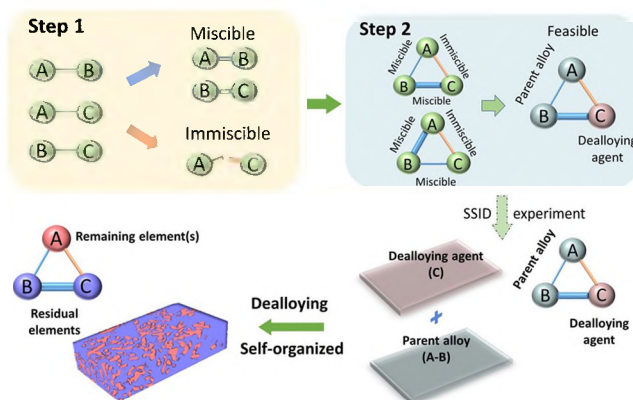


Fig. 2 Schematic of ML predictions of miscible/immiscible pairs and ternary dealloying systems. The two steps for predicting dealloying systems by the ML method. The first step is to classify the miscible and immiscible pairs and compose potential ternary dealloying systems. The second step is to determine the parent alloy and the dealloying agent in each of the ternary systems that are prescreened in the first step.

method, where testing and including more systems that did not fit the mixing enthalpy difference would help in exploring the design criteria, (2) to easily etch away the residual dealloying agent, the reported dealloying agent was mostly limited to Mg and Cu, leading to a bias in the training sets that may limit the prediction accuracies of the ML methods, and (3) the reported dealloying systems were primarily focused on successfully dealloyed systems, whereas systems that cannot be dealloyed are still useful for learning design criteria and training ML methods.

Machine learning prediction of ternary metal-agent dealloying systems. The ML predictions geared towards answering two research questions were broken down into two steps, as shown in Fig. 2. In the first step, we trained the ML methods to identify the miscible/immiscible pairs and then compose ternary dealloying systems with two miscible pairs and one immiscible pair. In the second step, we differentiated the parent alloy from the dealloying agent in each ternary dealloying system that was composed in the first step. In each of these two steps, three different ML methods

were applied for classification. The selected ternary systems are expected to form bi-continuous composite after isothermal treatment at an elevated temperature to introduce solid-state interfacial dealloying, as shown in Fig. 2. Here a two-step classification was employed. This was because the literature only reports ternary systems that can introduce dealloying, without information on systems that cannot introduce dealloying. This hindered us to label ternary systems directly in a classification process with supervised learning method. With the help of the first classification step to construct the miscible and immiscible pairs first, the plausible dealloying systems can be better predicted with a more accurately defined searching space.

In the first step, we trained three different ML methods to classify miscible and immiscible elemental pairs. The normalized confusion matrices from each ML method on the testing sets after variable selections are summarized in Supplementary Fig. 1. The three methods showed good classification performance on the testing sets; the accuracies of the random forest, XGBoost, and SVM were 1.00, 1.00, and 0.84, respectively. Note that the high accuracy is also related to the limited sample size, but the good performance on testing sets showed no overfitting issue. Future experiments can help to expand the dataset and achieve a more robust model. As a comparison, the accuracy of determining whether the mixing enthalpy is positive or negative to classify the miscible/immiscible pairs from reported dealloying systems is 0.828. The top 10 most important variables ranked by SHAP values for each of the three ML methods are shown in Supplementary Fig. 2. It is not surprising that the mixing enthalpy is the top, most important variable ranked by both ML methods, as the reported systems from the literature were mostly designed based on the mixing enthalpy. The solubility, which is commonly used in determining dealloying systems, was also highly ranked. It is clear that the high mixing enthalpy and low solubility will lead to immiscibility. Other high-ranking variables, such as the heat of formation, the energy at the ground state, the formation energy, and the decomposition energy as the energy above the convex hull of stable points, are common contributors in the determination of the phase stability. The high rankings of these features indicated that the ML methods rely heavily on thermodynamic stability to determine miscible/immiscible pairs, which is consistent with the scientific understanding of phase stability and signifies the reliability of these ML methods.

The overlap of the new miscible pairs predicted by the three ML methods is shown in Fig. 3a. Since each binary pair can only be classified as a miscible or immiscible pair, the two classification results are in a complementary relationship. We selected the pairs that are commonly predicted from the majority (two or three) ML methods as the classification results. Therefore, a total of 94 miscible and 27 immiscible pairs were selected to generate 169 potential ternary dealloying systems.

In the second step, we separately trained three different ML methods to distinguish the parent alloy from the dealloying agent using the training dataset with a 70% random split from the 31 reported ternary dealloying systems. The normalized confusion matrices from the three ML methods are summarized in Supplementary Fig. 3. All three ML methods showed good classification performances in determining the parent alloys and dealloying agents in the testing dataset, which is the remaining 30% of the random split, and the accuracies of the random forest, SVM, and XGBoost methods in classifying the parent alloy were 1.00, 0.95, and 1.00, respectively. In comparison, the conventional method based on the mixing enthalpy difference between the parent alloy and sacrificial element with a dealloying agent showed an accuracy of only 0.516 for the 31 reported ternary dealloying systems. We then removed the overlap of the reported ternary systems from the classified A-B-C systems and

focused on the newly predicted ternary dealloying systems. The overlap of classified ternary A-B-C systems with A-B as parent alloy from mixtures A-B or B-C as parent alloy in the first step by the three ML methods is shown in Fig. 3b. All the predicted ternary dealloying systems are summarized in Supplementary Table 1. The systems colored gray were voted higher by all three ML methods, while other colored systems were voted higher by only one of the two ML methods. Among the 132 predicted systems, only 59 systems satisfy the mixing enthalpy criteria, which are labeled in green and summarized in Supplementary Table 2, and the rest 73 systems that did not satisfy the mixing enthalpy criteria are labeled in orange in Supplementary Table 2. Note that using the mixing enthalpy to determine whether a system can be dealloyed is still valuable, but there are additional cases to be considered and explored. Compared to the training dataset in which 20 out of 31 (64.5%) of the dealloying systems consist of Mg as the dealloying agent, there are only 26 out of the 132 (19.7%) of the predicted dealloying systems using Mg as the dealloying agent. It shows the potential of using ML methods to explore new dealloying systems.

After removing the highly correlated variables, the top 12 most important variables selected by three ML methods by SHAP values, are shown in Fig. 3c, and the corresponding descriptions of each variable are summarized in Table 1. The variables are ranked along the vertical axis by their influence on differentiating the dealloying agent from the parent alloy. The horizontal axis shows the SHAP values for classifying the feasible and infeasible ternary dealloying systems. The color scale represents the feature value from high (red) to low (blue). Here the solubility difference between two elemental pairs was used to explain the meaning of SHAP values. The system with a relatively low solubility difference corresponds with a positive SHAP value, implies that it tends to be able to form a ternary dealloying system. The top 10 most important variables ranked by SHAP values from all three ML methods for variables selection are shown in Supplementary Fig. 4a–c. The correlation matrix showing the correlation coefficients between each pair of variables is summarized in Supplementary Fig. 4d. Different from the variables ranked in Step 1 to classify miscible/immiscible pairs, in Step 2 the mixing enthalpy was no longer ranked highly by any of ML methods.

Other thermodynamic variables, such as the energy at the ground state, heat of fusion, and heat of formation, were ranked highly. The energy at the ground state is related to the relative stabilities of the compounds⁴¹. The melting point is then associated with the elemental diffusivity²⁰. Furthermore, the atomic volume was also ranked high by the ML methods; this is consistent with our previous suggestion that the entropic contribution should be included in the design criteria of the metal-agent dealloying method⁴⁰. Conventionally, the mixing enthalpy has been treated as a dominant quantity due to the identification of stable alloys, and the contribution of the mixing entropy term for determining the total mixing Gibbs free energy was considered to be less important for its relatively small value, particularly in ordered structures²⁵. In a solid solution, the entropic contribution has a significant impact on the alloy phase stability. Because we included both reported intermetallic and solid solution dealloying systems in our training sets, the ML methods were able to rank the important variables for both alloy systems. Therefore, the limitation regarding the conventional design criteria for limited intermetallic alloys can be resolved by using the ML methods.

The number of predicted ternary dealloying systems were organized by element A, where Fig. 3d depicts ternary systems A-B-C. After removing the ternary system which cannot be distinguished by ML methods, as well as the systems that have been previously reported, we obtained a total of 132 ternary systems that

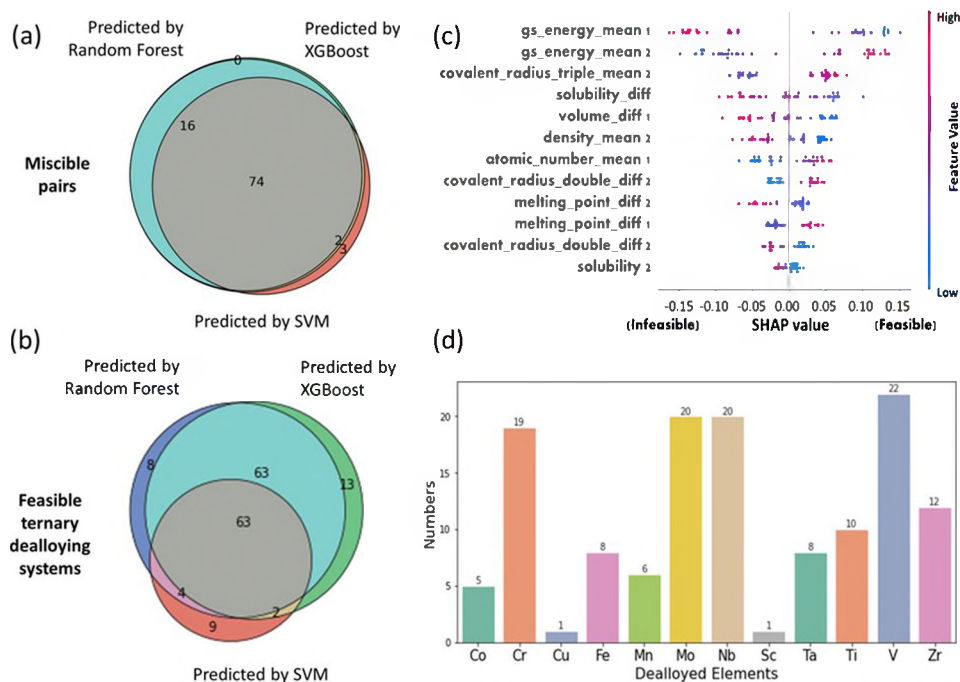


Fig. 3 The ML prediction result and corresponding SHAP value analysis. **a** The overlap of the miscible and immiscible pairs predicted by the three ML methods, namely random forest, XGBoost, and SVM. **b** The overlap of predicted ternary dealloying systems from the three ML methods. **c** The 12 variables which are ranked by their impacts on differentiating the dealloying agent from the parent alloy, as sorted by the random forest method. Variables ending with 1 represent the properties of the first two elements in a ternary system, and those ending with 2 represent the properties of the last two elements. **d** The number of predicted ternary dealloying systems organized by element A, where the parent alloy is A-B. The results are voted higher by at least two ML methods.

| Table 1 Description of the final 12 most important variables selected by three ML methods, using for determining parent alloy (Step 2). | |
|---|--|
| Variables used in ML methods | Description |
| gs_energy_mean1 | Average of the DFT energy per atom (raw VASP value) for the $T = 0$ K ground state from the first two elements in the ternary system |
| gs_energy_mean2 | Average of the DFT energy per atom (raw VASP value) for the $T = 0$ K ground state from the last two elements in the ternary system |
| Solubility_diff | The difference in the maximum equilibrium solubilities between the first two elements and last two elements in the same ternary system |
| density_mean2 | Average of the density at 295 K of the last two elements in the ternary system |
| atomic_number_mean1 | Average of the number of protons found in the nucleus of an atom, from the first two elements in the ternary system |
| covalent_radius triple_mean2 | Average of the number in the triple-bond covalent radii, as defined by Pyykko et al., from the last two elements in the ternary system |
| solubility 2 | The maximum equilibrium solubility between the last two elements |
| melting_point_diff 1 | The difference in the melting point between the first two elements in the ternary system |
| melting_point_diff 2 | The difference in the melting point between the last two elements in the ternary system |
| covalent_radius double_diff 2 | The difference in the double-bond covalent radii, as defined by Pyykko et al., from the last two elements in the ternary system |
| covalent_radius double_diff 1 | The difference in the double bond covalent radii, as defined by Pyykko et al., from the first two elements in the ternary system |
| icsd_volume_diff 1 | The difference in the atomic volumes from ICSD database between the first two elements in the ternary system |

Variables ending in 1 represent the variables of the first two elements in a ternary system, and those ending in 2 represent variables of the last two elements in the same ternary system. The variables were collected from Xenonpy⁷⁶, pymatgen⁸², matminer⁷⁷, and literatures⁷⁸. The names of the variables follow the convention as shown in these sources.

have been voted highly by at least two ML methods. Interestingly, the predicted elements can be dealloyed by a large number of dealloying systems that correspond to elements such as Cr and Nb that have been reported and fabricated based on the metal-agent dealloying method⁴². ML methods can not only provide an approach to explore metal elements that have not been explored but also enable the design of bicontinuous nanocomposites with wider elemental compositions for the explored elements.

Combinatorial thin-film SSID and synchrotron XRD characterization. Combinatorial thin-film deposition is used to prepare a range of parent alloy compositions efficiently, which is critical for analyzing the composition-dependent structural/morphological evolution and validating the ML method predictions. With its high brilliance X-ray, the synchrotron enables the characterization of a large number of samples in a short time⁴³. These two components are critical in the materials discovery for metal-agent dealloying.

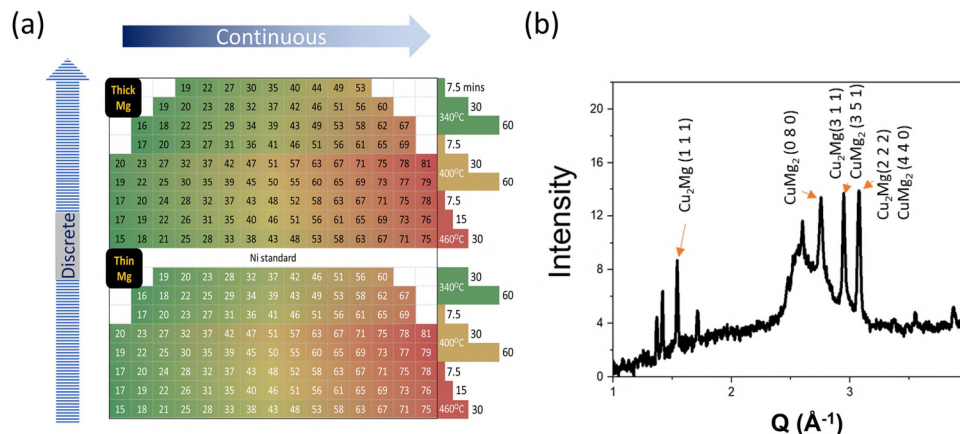


Fig. 4 Sample layout for XRD measurement and XRD result. **a** Sample layout for a series of combinatorial thin-film samples with four controlled parameters: parent alloy composition (Ti ~20–80 at.% in a Ti-Cu parent alloy), Mg dealloying agent film thickness (250 and 450 nm), dealloying time (7.5, 15, 30, and/or 60 min, depending on the dealloying temperature) and temperature (340, 400 and 460 °C). The numbers in each grid indicate atomic concentration of Ti in the parent alloy film. **b** XRD pattern from a sample dealloyed at 400 °C for 60 min, after subtracting the glass background.

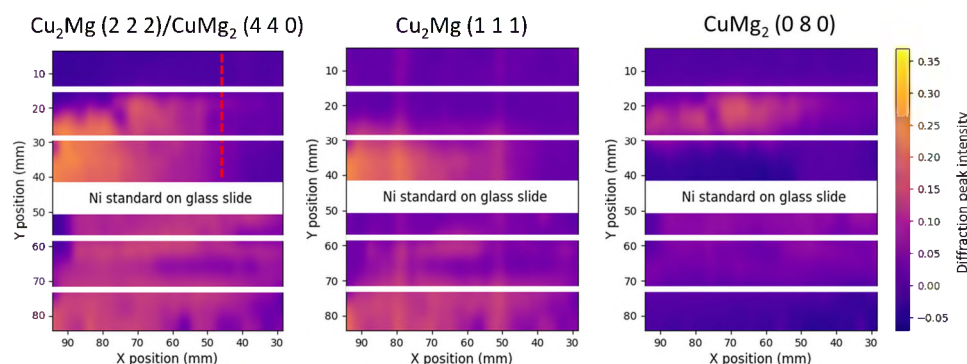


Fig. 5 Selected diffraction peaks and interpolated diffraction peak intensities from Cu_2Mg (1 1 1), CuMg_2 (0 8 0), and Cu_2Mg (2 2 2) overlapped with CuMg_2 (4 4 0). The white color lines separate the regions with different dealloying temperature.

Here, the reason for choosing Ti-Cu (the A-B parent alloy) dealloyed by Mg (the C dealloying agent) with the SSID method is twofold. The first reason for this choice is that Ti-Cu dealloyed by Mg has been reported only in bulk samples with discrete compositions, while the precise dealloying composition threshold is missing. Another reason for this choice is that the Ti-Cu-Mg system does not adhere to the conventional mixing enthalpy difference criteria, which states that the mixing enthalpy of Cu-Mg is less negative than that of Ti-Cu²¹. In addition, the sample preparation procedure for the Ti-Cu/Mg system has been well-established. We can validate the new characterization methods in the current work against the known the crystalline phases and use the prior knowledge in the XCA method. Thus, the dealloying behavior of the Ti-Cu-Mg system through the thin-film SSID method was studied, also serving as a proof of concept for establishing a method for future experiments with a wider range of elements to (1) experimentally validate the ML method predictions, and (2) acquire new datasets for ML method training and testing, as illustrated by the workflow in Fig. 1. In the future, the accuracy of each ML method can be validated with experiments to determine the ML method with the highest accurate probability, or to combine the prediction from the ML methods each with a different weight factor based on the accuracy. In addition, experimenters may choose a dealloying system based on different design goals, such as elemental compositions, cost of the materials, and properties of the materials, etc.

A combinatorial Ti-Cu thin film was prepared by the cosputtering method⁴⁴. Compared to conventional samples with discrete compositions, combinatorial samples provide an effective way to study samples with compositional gradients. In addition, we considered the influence of the processing parameters on the dealloying combinatorial samples. A total of four parameters were considered, including 3 dealloying times (7.5, 15, 30 and/or 60 min, depending on the dealloying temperature), 3 dealloying temperatures (340, 400, and 460 °C), 2 thicknesses of the dealloying agent, a Mg film, (250 and 450 nm), and a continuous parent alloy composition (Ti ~20–80 at.% in a Ti-Cu alloy). The layout with the conditions for combinatorial samples is shown in Fig. 4a, and the setup of samples measured at the XPD beamline is shown in Supplementary Fig. 5. The representative XRD result after subtracting the background from glass substrate is shown in Fig. 4b.

The scanning step size along the horizontal (X) axis was 2 mm, and that along the vertical (Y) axis was 4.75 mm. A total of $34 \times 18 = 612$ data points were collected upon scanning an 80.64 mm by 66 mm area. The interpolated diffraction peak intensities from Cu_2Mg (1 1 1), CuMg_2 (0 8 0), and Cu_2Mg (2 2 2) overlap with CuMg_2 (4 4 0) in grid scanning and are shown in Fig. 5. Grid-scanned mapping was used to analyze the kinetics in the Ti-Cu dealloying system and was also used as the ground truth to validate the automatic XRD pattern analysis and autonomous algorithm-driven characterization result, which will be discussed in the next section.

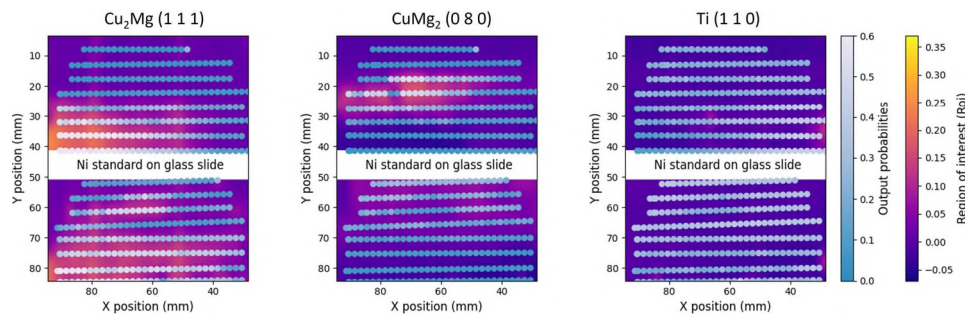


Fig. 6 The distribution of the XCA-generated phase probability, represented by round points in a gray-to-blue color scheme, overlaid on the interpolated intensity of the designated diffraction peak in the grid-scanned result, represented by a yellow-to-purple color scheme. The diffraction analysis was based on Cu_2Mg (1 1 1), CuMg_2 (0 8 0), and Ti (1 1 0).

The variation in the dealloying-generated Cu_2Mg and CuMg_2 phases with dealloying temperature is shown in Fig. 5. XRD from the CuMg_2 phase can be detected only below 460 °C, and XRD from the Cu_2Mg phase can be detected only when dealloying at 460 °C with thick Mg. Such a phase distribution is consistent with the phase transition in Cu and Mg interdiffusion. Arcot et al. showed that with excess Cu in the system, CuMg_2 transformed to Cu_2Mg when the annealing temperature was >380 °C⁴⁵. When a small amount of oxygen is distributed in the sample, crystal phase formation can be increased over a range of more than 200 °C. Knowing the dealloying-generated phase distribution, we used CuMg_2 and Cu_2Mg phases together to determine the dealloying progress.

The parting limit is the composition threshold of the parent alloy; when the concentration of the sacrificial component (B) falls below this threshold, the parent alloy cannot be fully dealloyed. While the sacrificial element composition in the parent alloy is above the parting limit, an atomic-scale network of sacrificial components (B) runs through the entire structure of the parent alloy for B element dissolution. This dissolution process is called a percolation dissolution mechanism in dealloying⁴⁶. The theoretical values of the site percolation threshold for face-centered cubic (FCC) and hexagonal close packed (HCP) structures have been determined to be ~20 at.%, where ~24.5 at.% has been reported for body-centered cubic (BCC) structures^{47,48}. Although few systems, such as the Cu-Zn system, showed a parting limit that is close to the theoretical percolation limit of ~20%, more common systems, such as Au-Ag and Au-Cu, showed a much higher parting limit of ~55%.

We used the boundary of the dealloying-generated crystalline phase to search for the parting limit in the thin-film SSID Ti-Cu system. Since only crystalline Cu_2Mg and CuMg_2 phases were detected from the XRD pattern, their distributions were used to find the parting limit. In the samples dealloyed with a thin Mg layer, the Cu_2Mg and CuMg_2 phases can be found based on samples with a sacrificial element Cu of 20–80 at.% in the parent alloy. Such a composition range matches the theoretical percolation limit. It is interesting that in the samples dealloyed by the thicker Mg films, a clear decrease in diffraction peak intensity can be found when the sacrificial element composition is below ~35 at.% in the parent alloy, as indicated by vertical red dashed lines in Fig. 5. Since thick Mg layer samples provide more material for dealloying, the difference cannot be induced by limited diffusion materials. In addition, the percolation threshold is related to the atomic coordination environment⁴⁹, and a relatively thin layer is not expected to change the environment. The differences in dealloying-generated crystalline phases in thinner vs. thicker Mg films may be attributed to geometric differences in the thin films, although further kinetics analysis is needed.

ML-augmented synchrotron characterization: ML-informed automated (XCA) and ML-driven autonomous experiments (gpCAM). A neural-network-based automated XRD pattern analysis method, XCA⁵⁰, and a Gaussian process regression-based algorithm, gpCAM³², were tested separately for ML-informed automated and autonomous characterization. The goal here is to develop a methodology for rapidly validating the ML prediction results through experiments and enriching the training datasets for ML methods.

The working principle of XCA is that an ensemble model is trained from synthetic datasets by inputting the CIFs of expected phases and subsequently generates the existence probability of a given phase given the diffraction pattern. The trained XCA model can be applied to analyze the collected XRD patterns in real time. In our study, the XCA method first generated a synthetic dataset from the CIFs of the expected Ti, Cu_2Mg , CuMg_2 , and Mg phases with a range of experimental variations, such as the diffraction instrument parameters, peak shape, sample texturing, and offset. The ensemble of CNNs were trained based on this synthetic dataset. Then, the trained model was applied to analyze the grid-scanned results on combinatorial Ti-Cu-Mg samples and the output existence probability of Ti, Cu_2Mg and CuMg_2 from each input diffraction pattern. To analyze the accuracy of the XCA method, we overlapped the XCA-generated existence probability of each phase with the interpolated intensity of a key diffraction peak interest for each phase from the grid scan, as shown in Fig. 6. The layout discrepancy between the XCA-generated probability distribution and grid-scanned data points is related to the sample layout. The XCA focused only on the samples with calibrated compositions, while grid scans covered the full dimension of the samples, as shown in Fig. 5. Overall, the distribution of the high probabilities of Cu_2Mg and CuMg_2 phases matches well with the distribution of strong diffraction intensity from the grid-scanned result. Crystalline Ti was barely detected in the dealloyed sample, but XCA generated relatively high existence probabilities among a wide compositional range of the dealloyed samples. This deviation stems from the discrete probability distribution being confined to only the input phases and no other possible impurities. The challenge of classifying out-of-distribution data, and small amounts of phases in samples need to be considered carefully in the future. Overall, the accuracy in detecting a large amount of Cu_2Mg and CuMg_2 phases demonstrated that the XCA is a powerful tool for analyzing the evolution of the phase during the dealloying processes. In addition, the XCA holds great advantages in considering the preferred orientation, peak shifting, and phase mixtures, which are critical for characterizing thin-film materials⁵⁰. The XCA method provides an additional tool for future experimental validation and is thus valuable to be included in the current materials discovery workflow as a proof-of-concept. The data-

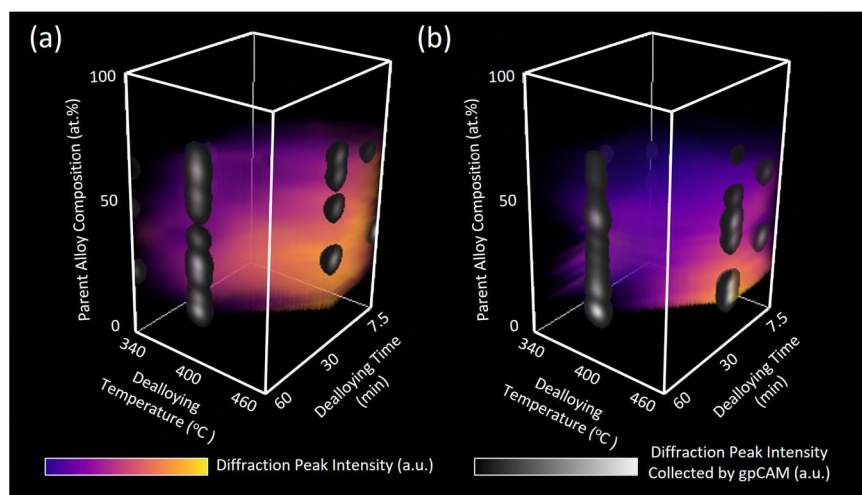


Fig. 7 The data-collection distribution of the autonomous experiment driven by the gpCAM algorithm overlaid on the interpolation of the diffraction peak intensities from Cu_2Mg (3 1 1) and CuMg_2 (3 5 1). **a** The sample with thin Mg layer, **b** The sample with thick Mg layer. Note that the colormap in grid-scan and gpCAM are different for better visualization. (For video sequence of diffraction peak intensity analysis see Supplementary Movie 1–2).

collection points in autonomous characterization driven by an XCA-generated probability are summarized in Supplementary Fig. 6.

We also tested the gpCAM algorithm, which autonomously decides on future measurements and is critical in driving experiments in the multidimensional parameter space. gpCAM was developed based on Gaussian-process regression and we augmented the GP posterior variance with local information and measurement costs. In this experiment, gpCAM considers the absolute value of the gradient to find regions where the diffraction intensity rapidly changes as local information and the total time required to acquire a new datapoint as the measurement cost. Here, gpCAM aims to search the phase boundary based on the intensities of the diffraction peaks of dealloying-generated Cu_2Mg and CuMg_2 . The selected diffraction peak was previously characterized as an overlapping peak from Cu_2Mg (3 1 1) and CuMg_2 (3 5 1), and the peak in the q range is 3.047 \AA^{-1} – 3.106 \AA^{-1} , where q is the momentum transfer of the scattered wave vector to the incident wave vector. By comparing the peak intensity distributions from Cu_2Mg (1 1 1) and CuMg_2 (0 8 0), we were able to determine that this diffraction peak is mainly attributed to the Cu_2Mg phase, thus corresponding to Cu_2Mg (3 1 1).

Compared to the grid scan in which 612 points were collected, in the gpCAM-driven autonomous experiment, only 133 points were collected. With only ~21.7% of the data points, the autonomously driven analysis was able to successfully determine the boundary of the Cu_2Mg phase by means of the key processing parameters. The intensity and distribution of gpCAM data- and mapping of the Cu_2Mg (3 1 1) diffraction peak intensity in the grid scan are shown in Fig. 7. The collection trajectory, intensity of the Cu_2Mg (3 1 1) diffraction peak in the gpCAM-driven characterization is shown in Supplementary Fig. 7. A large number of collected points were distributed at the boundary between 400°C for 60 min and 460°C for 7.5 min in both thin and thick dealloying agent samples. This result indicates that gpCAM determined the Cu_2Mg phase transition distributed between 400°C and 460°C , matching the grid scan results. Overall, this match of the phase boundary from grid-scanned mapping with gpCAM collected points proves the capability of gpCAM in analyzing the phase transition in SSID.

It should be noted that while gpCAM determines the boundary based on the diffraction peak intensity calculated from a fixed q range, it could be replaced by the phase probability result from

the XCA method. From each collected XRD pattern, the XCA generates the probability of an existing phase, which can be used to replace the diffraction peak intensity in the gpCAM algorithm. gpCAM searches through the phase distribution based on the probability of a given phase existing at each location. In such circumstances, we can prevent XRD pattern variation in thin-film samples, such as a strain-induced phase shift, calculate the compositional variation according to Vegard's law, accurately determine the dealloying-generated phase intensity, and efficiently analyze the dealloying results. To fully realize a closed-loop approach, future test on the systems predicted in this work could consider different design criteria and cost functions based on the goal of the materials design. For instance, other dealloying agents beyond Cu and Mg could be prioritized to create novel systems. Alloys with compositions not following the conventional mixing enthalpy differences would be of fundamental research interests. For particular applications, the cost of materials and processing constraints such as temperature can be considered.

Conclusion

An ML approach to dealloying and designing nanoporous/nanocomposite materials is first introduced. The proposed ML ensemble pipeline consisting of three ML methods was able to predict 132 ternary dealloying systems from 16 selected metal elements after training with published dealloying systems. In addition to elements that have never been reported by the dealloying method, more elemental combinations have been introduced to pave the way for the development of nanocomposites and nanoporous materials with a broader range of applications. The important variables ranked by the ML methods, such as energy at the ground state, fusion heat, and heat formation, indicate that thermodynamic stability is the key to designing dealloying systems. In addition, atomic volume differences that contributed to entropy changes were also ranked highly by the ML methods. The ML methods have demonstrated their potential to overcome the limitations of conventional design criteria in intermetallic alloys and can be readily applied to a wider range of alloys.

The kinetics in Ti-Cu-Mg dealloying systems was studied using combinatorial thin-film deposition and synchrotron XRD characterization. The relation between the phase transition of dealloying-generated Cu_2Mg and CuMg_2 phases and the dealloying conditions was discussed. Relation between the dealloying-

generated phase and the film thickness and parent alloy composition was analyzed.

A neural-network-based automated XRD pattern analysis method, XCA, and a Gaussian process regression-based algorithm using the gpCAM software were validated through the characterization of Ti-Cu-Mg dealloying systems. The accuracy of XCA analysis in identifying the Cu_2Mg and CuMg_2 phases generated from dealloying and the efficiency of finding the Cu_2Mg phase boundary through gpCAM were demonstrated. In the future, validated ML-informed automated pattern analysis and the ML-augmented decision-making algorithm can be combined to further the development of the autonomous experimental methods, where a range of automatic decision making, analysis, instrumentation, and communication tools are combined to create a closed-loop that can make intelligent decisions during an experiment without human interaction.

The ML-augmented workflow built in this work will significantly improve data-acquisition accuracy and rate, which will pave the way for rapidly validating the ML predictions, enriching the ML training set with experimental results, improving the accuracy of the ML methods, and developing nanoporous/nanocomposite materials. Moreover, such a workflow can be used to design a wider range of solid-state materials with limited published results.

Methods

Machine learning prediction of ternary systems for metal-agent dealloying.

To construct a ternary system where the dealloying of a parent alloy by a dealloying agent can occur, two questions must be answered. First, among all possible ternary systems, how do we identify promising ternary systems that may display dealloying behavior with two miscible pairs and one immiscible pair of elements? Here we simplify the question and exclude the potential ternary systems that composed of three miscible pairs or composed of one miscible pair and two immiscible pairs. Second, within a ternary system, how do we discern which element will act as a dealloying agent in the immiscible pair? Note that the other element is then the remaining element left within the parent alloy. While for both questions, thermodynamics can provide guidelines for identifying possible answers, these two questions are essentially classification problems that can also be addressed using ML methods.

Based on the two questions outlined above, we divided the ML-based classification process for identifying dealloying systems into two steps. In Step 1, ML methods were trained to classify the published miscible and immiscible pairs. Once trained, the first-step ML classifier can be applied to classify new elemental pairs into miscible and immiscible pairs. Two miscible pairs and one immiscible pair were then combined to form ternary systems. In Step 2, ML methods were trained to distinguish ternary systems that will lead to dealloying process. This was done by identifying systems with the A-B combination as a parent alloy from a mixture of ternary systems where either A-B or B-C combination is the parent alloy; both new systems satisfy the criterion of Step 1 in that they contain a composition of two immiscible pairs and one miscible pair. However, only the systems with A-B as parent alloys will lead to the dealloying process. Once trained, the second-step ML classifier can identify the parent alloy and dealloying agent from new ternary systems.

To establish ML methods to address the above classification problems, training and testing datasets were collected from a total of 21 different peer-reviewed papers, including 13 discussing the LMD method^{8,10,11,14,19,51–63}, 4 papers addressing the SSD method^{17,20,64,65}, and 4 papers examining the aqueous solution dealloying (ASD) method^{66–69}, listed in Supplementary Table 3. These data were then used to classify miscible/immiscible pairs. Excluding the ASD systems, only the training and testing datasets from metal-agent dealloying systems were used for classifying parent alloys and dealloying agents. Each of the ternary dealloying systems is composed of three elements, A, B, and C, where the parent alloy elements A and B are miscible. In addition, the sacrificial element (B) and the dealloying agent (C) are miscible, while the remaining element (A) and the dealloying agent (C) are immiscible. We organized all the reported dealloying systems to follow the A-B-C sequence, specifically to distinguish dealloying agent C from the remaining element A. To identify miscible/immiscible pairs, the organized A-B-C ternary systems were separated into A-B and B-C miscible pairs, as well as A-C immiscible pairs. Some reported systems have more than two elements in the parent alloy. Following the miscible/immiscible relation, these systems ($A_1A_2\ldots A_n$)-B-C were separated into multiple miscible pairs (A_1B , $A_2B\ldots A_nB$, BC) and immiscible pairs (A_1C , $A_2C\ldots A_nC$) and placed into the miscible/immiscible dataset. In addition, we also separated ($A_1A_2\ldots A_n$)-B-C into A_1 -B-C, A_2 -B-C, ..., A_n -B-C ternary systems, which were included in the training and testing datasets for classifying parent alloys and dealloying agents.

We first conducted the training and testing of the ML methods for both steps using the data from the literature. In the first step, we classified all reported dealloying systems into miscible and immiscible pairs. Two types of elemental pairs are labeled as miscible pairs: (1) the parent alloy elemental pair, and (2) the pair of sacrificial element and dealloying agent. The pair of the residual element and the dealloying agent is labeled as immiscible pair. In our dataset, we summarized 64 pairs, including 43 miscible and 21 immiscible pairs. To train each ML method, the dataset was randomly divided, where 70% was allocated for training and 30% was allocated for testing classifier performance. To increase the reliability of the ML predictions, we used three ML methods from the scikit-learn library⁷⁰: random forest⁷¹, support vector machine (SVM)⁷², and XGBoost⁷³. We trained the ML methods with all variables and used the SHAP (SHapley Additive exPlanations) value, based on a game theoretic approach for explaining outputs of ML models to select variables⁷⁴. The union set of the top five important variables from each ML method was selected, leading to a total of 10 variables. Then we removed two highly correlated variables (correlation coefficients $r > 0.8$) based on the correlation matrix. The final 8 variables selected were used to train the three ML methods with one more iteration, listed in Supplementary Table 4. The trained ML classifier was then applied to predict new miscible and immiscible pairs generated from 16 selected elements, including Mg, Al, Sc, Ti, V, Cr, Mn, Fe, Co, Ni, Cu, Zn, Zr, Nb, Mo, and Ta. The fabrication of half of the selected elements has not been reported based on the metal-agent dealloying method, according to McCue et al.'s report⁴². Although all the three ML methods showed great performance in both training and testing process, they showed different prediction results on new pairs. We used the majority vote from the three ML classifiers with equal weight as the ensemble prediction result⁷⁵ so that the common predictions from at least two ML classifiers were included, while the prediction from only one ML classifier without intersections with the other ML classifiers' prediction was not included in the final classification result. The newly classified miscible and immiscible pairs derived from the ensemble prediction results were used to build new potential ternary dealloying systems A-B-C, where A-B and B-C are two miscible pairs and A-C is an immiscible pair.

In Step 2, we then introduced the second classification step to identify the parent alloy and dealloying agent in the potential ternary systems, as classified in Step 1. The parent alloy and dealloying agent were not identified in the new ternary A-B-C dealloying system in the first step. In Step 1, we did not determine the relative strength of the miscibility. In Step 2, we then identified whether A-B or B-C had a stronger tendency to mix; in other words, we classified whether A-B or B-C was the parent alloy. From the published result, we knew that the C element had a stronger tendency to combine with the B element so that the A element could not dealloy the B-C pair. In Step 2, we classified the ternary alloy systems where B-C combination is the parent alloy as systems that cannot be dealloyed. A total of 62 systems, including 31 feasible systems, were reported as feasible ternary systems with A-B combination as parent alloy; the 31 corresponding infeasible ternary systems with B-C combination as the parent alloy were randomly split, resulting in 70% of the system allocated for training and 30% for testing. The elemental distribution in training and testing dataset in the step 1 and 2 are summarized in Supplementary Fig. 8. We then separately trained 3 ML methods (random forest, XGBoost, and SVM) to classify feasible and infeasible systems. The variables selection based on the SHAP values were conducted following the similar procedure as in step 1, except that the top eight important variables were selected from each ML method, leading to a total of 18 in the union set. After removing six values with $r > 0.8$, a total of 12 final variables were selected (Table 1). The trained classifiers were then applied to differentiate ternary systems with A-B or B-C as parent alloy that were composed in the first step.

The variables that were input into the ML methods to classify and predict new dealloying systems were sets of quantitative attributes describing the materials. We selected 55 relevant element-level properties and electronic structure attributes available in the Python library XenonPy⁷⁶, including 21 calculated elemental properties developed for a general machine learning framework for inorganic materials²⁶. To avoid bias from training sets that restricted dealloying agents to specific elements (Cu and Mg), we focused on the relationships within elemental pairs rather than the individual element properties. We computed the means and absolute differences of each pair's 52 attributes and used them as variables in the ML methods. In addition, we included two key variables, mixing enthalpy and maximum equilibrium solubility in binary systems, that are conventionally used to classify dealloying systems. The mixing enthalpies of all elemental pairs were calculated based on Miedema's method⁴¹ and are available from the Python library Matminer⁷⁷. The maximum equilibrium solubilities of the first 83 elements in the periodic table were collected from the literature⁷⁸. We also included thermodynamic attributes to determine the compound stabilities, including the formation energy and energy above the convex hull (approximate as decomposition enthalpy)²⁴, which are available in the Matminer library⁷⁷. Since the compositions of the dealloying-generated B-C phases were not reported in all the literature, we did not consider a specific composition in each A-B-C system. We generalized the dealloying for compositional attribution and included only the minimum and maximum values of the formation energy and energy above the hull among each binary pair here. All variables were summarized in Supplementary Table 5.

To select the optimal ML methods, we evaluated the performances using fivefold cross-validation for the three different ML methods in the scikit-learn

library⁷⁰, including random forest, SVM, and XGBoost. Hyperparameters, including the number of estimators and maximum depth in XGBoost and random forest, and C values in SVM were selected through a fivefold cross-validated randomized search⁷⁹. The confusion matrix were used to assess each method's classification performance. Although the confusion matrix from three ML methods all showed a good classification performance, and the majority of the predictions from three ML methods overlapped, there is still a small number of predictions from each ML method disagrees with each other. Because of the limited data availability, and clustered data at Ti, Cu and Mg elements, we further validated the trained ML methods with the leave-one-element-out method in both steps. The *cross_val_score* function available in the *scikit-learn* package was applied with 30 fold cross-validation, and the output accuracies of trained ML methods were averaged and collected. In step 1, the averaged accuracy for the random forest is 0.90, the accuracy for XGBoost is 0.89, and the accuracy for SVM is 0.92. In step 2, the accuracy for the random forest is 0.92, the accuracy for XGBoost is 0.87, and the accuracy for SVM is 1.00. The result indicates that ML method is reliable.

Combinatorial thin-film deposition and synchrotron characterization. Borosilicate glass slides (TedPella) with an initial area of $114 \times 159 \text{ mm}^2$ and a thickness of $\sim 200 \mu\text{m}$ were cut down to $72.5 \times 72.5 \text{ mm}^2$ in size for the deposition substrate. Before deposition, the glass slides were cleaned with isopropyl alcohol and deionized water, followed by treatment in an oxygen-plasma environment. A Ta film (99.95% purity, 3" diameter, and 0.125" thickness target from Kurt J. Lesker) was deposited by direct current (DC) sputtering as a barrier layer. A combinatorial film with a compositional gradient of $\text{Ti}_x\text{Cu}_{1-x}$ ($x = 20\text{--}80 \text{ at.}\%$) was prepared by cosputtering Ti and Cu targets (2" diameter and 0.125" thickness, Kurt J. Lesker) at the University of Maryland⁴⁴. A Mg sputtering target (99.95% purity, 3" diameter, and 0.25" thickness from Kurt J. Lesker) was used as the dealloying agent. The deposition of the homogeneous barrier layer Ta film and dealloying agent Mg film was conducted at the Center for Functional Nanomaterials (CFN) in Brookhaven National Laboratory (BNL). For the Ta and Mg sputtering targets, a cleaning protocol that involved sputtering the target for 5–10 min with the sputtering shutter closed was conducted to remove the surface oxides. Ta, Ti-Cu, and Mg films were sequentially sputtered onto borosilicate glass slides. Two types of samples with different thicknesses of Mg were prepared: one with a 250 nm thin Mg layer and another with a 450 nm thick Mg layer. The thickness of the rest layer was consistent in both types of samples: 80 nm for Ta and $220 \pm 35 \text{ nm}$ for Ti-Cu.

After deposition, the samples were heated by rapid thermal processing (RTP-600S, Modular Process Technology Corp.) for an isothermal heat treatment to introduce dealloying. All heat treatment processes were conducted in a reduced gas atmosphere (4 vol.% hydrogen and 96 vol.% argon) to prevent oxidation during the heat treatment. The samples were heated from room temperature to the designated dealloying temperature in 30 s and kept at the dealloying temperature for a designated duration of time. The samples were then cooled to room temperature in $\sim 150 \text{ s}$. The heating temperature and time were determined based on the estimated diffusion length between Cu and Mg, as calculated based on the diffusion data in the literature⁸⁰.

X-ray diffraction analysis was conducted at the X-ray powder diffraction (XPD) beamline 28-ID-2 at the National Synchrotron Light Source II (NSLS-II) in BNL. The incident X-ray beam energy was 66.16 keV, with a corresponding X-ray wavelength of 0.1874 \AA . The beam size was $0.5 \text{ mm} \times 0.5 \text{ mm}$. A large-area X-ray detector with 2048×2048 pixels was used to measure the XRD patterns, where the size of each pixel was $200 \times 200 \mu\text{m}^2$. The distance from the sample to the detector was first calibrated with a Ni standard and determined to be 1356.038 mm . The exposure time for collecting each XRD pattern was 60 s. Phase identification based on the XPD results was carried out by comparing the peak locations to those of reference compounds using the commercial software package Jade (Materials Data, Inc. Jade 9). The analysis of diffraction peak intensity was based on the integral of peak intensity after removing the background. The background was fitted by linear interpolating the averaged values of three points on both sides of a given diffraction peak. For instance, for the Cu_2Mg (311) peak, we integrated the peak intensity with the corresponding q range from 2.925 to 2.974 \AA^{-1} . The diffraction background was removed based on a linear interpolation of the peak intensity at 2.925 and 2.974 \AA^{-1} , where values were averaged from three data points at the immediately left of 2.925 \AA^{-1} and right of 2.974 \AA^{-1} respectively.

Automatic XRD pattern analysis and autonomous characterization. A ensemble of feed forward convolutional neural networks were applied to automatically analyze the XRD pattern⁵⁰. These models were built using the Crystallography Companion Agent (XCA) package, and trained on synthetic datasets by inputting the crystallographic information file (CIF) of the expected phases and creating an output of probabilistic classifications. The XCA input included Mg, Cu_2Mg , CuMg_2 , and Ti CIFs collected from the materials project. The autonomous experiment was driven by gpCAM to autonomously detect phase transformation and drive the XRD data-collection process³¹.

Bluesky, a library for experimental control and scientific data collection, was used for synchrotron characterization⁸¹. This library supports the collection and analysis of data in real time, enabling the incorporation of XCA and gpCAM into the ML-augmented materials design process.

In the autonomous characterization process, the experimental control was coordinated based on two Python programs: bluesky control of the XPD beamline and gpCAM implementation of Gaussian-process-based optimization for decision making. Although XCA performed automated XRD pattern analysis separately in this experiment, it demonstrated the feasibility of combining this step with autonomous experimental control; such control involves bluesky controlling the beamline, XCA analyzing the pattern, and gpCAM deciding the measurement point from the XCA-generated probability.

Data availability

Data contained in this manuscript are available from the corresponding authors upon reasonable request.

Code availability

The codes that support the findings of this study are available from the corresponding author upon reasonable request.

Received: 22 December 2021; Accepted: 10 October 2022;

Published online: 09 November 2022

References

- Liu, X. et al. Power generation by reverse electrodialysis in a single-layer nanoporous membrane made from core-rim polycyclic aromatic hydrocarbons. *Nat. Nanotechnol.* **15**, 307–312 (2020).
- Zugic, B. et al. Dynamic restructuring drives catalytic activity on nanoporous gold-silver alloy catalysts. *Nat. Mater.* **16**, 558–564 (2017).
- Shi, S., Li, Y., Ngo-Dinh, B.-N., Markmann, J. & Weissmüller, J. Scaling behavior of stiffness and strength of hierarchical network nanomaterials. *Science* **371**, 1026–1033 (2021).
- Erlebacher, J., Aziz, M. J., Karma, A., Dimitrov, N. & Sieradzki, K. Evolution of nanoporosity in dealloying. *Nature* **410**, 450–453 (2001).
- Wada, T., Yubuta, K., Inoue, A. & Kato, H. Dealloying by metallic melt. *Mater. Lett.* **65**, 1076–1078 (2011).
- Zhao, C. H. et al. Three-dimensional morphological and chemical evolution of nanoporous stainless steel by liquid metal dealloying. *ACS Appl. Mater. Interfaces* **9**, 34172–34184 (2017).
- Wada, T. & Kato, H. Three-dimensional open-cell macroporous iron, chromium and ferritic stainless steel. *Scripta Materialia* **68**, 723–726 (2013).
- Wada, T. et al. Bulk-nanoporous-silicon negative electrode with extremely high cyclability for lithium-ion batteries prepared using a top-down process. *Nano Lett.* **14**, 4505–4510 (2014).
- Zhao, C. H. et al. Imaging of 3D morphological evolution of nanoporous silicon anode in lithium ion battery by X-ray nano-tomography. *Nano Energy* **52**, 381–390 (2018).
- Okulov, I. V. et al. Nanoporous magnesium. *Nano Res.* **11**, 6428–6435 (2018).
- Yu, S. G., Yubuta, K., Wada, T. & Kato, H. Three-dimensional bicontinuous porous graphite generated in low temperature metallic liquid. *Carbon* **96**, 403–410 (2016).
- Tsuda, M., Wada, T. & Kato, H. Kinetics of formation and coarsening of nanoporous alpha-titanium dealloyed with Mg melt. *J. Appl. Phys.* **114**, 113503–113501–113508 (2013).
- Wada, T., Setyawan, A. D., Yubuta, K. & Kato, H. Nano- to submicro-porous beta-Ti alloy prepared from dealloying in a metallic melt. *Scripta Materialia* **65**, 532–535 (2011).
- Joo, S. H. et al. Beating thermal coarsening in nanoporous materials via high-entropy design. *Adv. Mater.* **32**, <https://doi.org/10.1002/adma.201906160> (2020).
- Wada, T., Yubuta, K. & Kato, H. Evolution of a bicontinuous nanostructure via a solid-state interfacial dealloying reaction. *Scripta Materialia* **118**, 33–36 (2016).
- Shi, Y., Lian, L. X., Liu, Y. & Xing, N. Y. A novel solid-state dealloying method to prepare ultrathin ligament nanoporous Ti. *Appl. Phys. a-Mater. Sci. Processing* **125**, <https://doi.org/10.1007/s00339-019-2984-z> (2019).
- Zhao, C. H. et al. Bi-continuous pattern formation in thin films via solid-state interfacial dealloying studied by multimodal characterization. *Mater. Horizons* **6**, 1991–2002 (2019).
- Zhao, C. et al. Kinetics and Evolution of Solid-State Metal Dealloying in Thin Films with Multimodal Analysis. *Acta Materialia* **242**, 118433 (2022).
- Kim, J. W. et al. Optimizing niobium dealloying with metallic melt to fabricate porous structure for electrolytic capacitors. *Acta Materialia* **84**, 497–505 (2015).

20. McCue, I. & Demkowicz, M. J. Alloy design criteria for solid metal dealloying of thin films. *Jom* **69**, 2199–2205 (2017).
21. Miedema, A. R., Dechatel, P. F. & Deboer, F. R. Cohesion in alloys - fundamentals of a semi-empirical model. *Phys. B C* **100**, 1–28 (1980).
22. Zhang, R. F. et al. An informatics guided classification of miscible and immiscible binary alloy systems. *Sci. Rep.* **7**, <https://doi.org/10.1038/s41598-017-09704-1> (2017).
23. Takeuchi, A. & Inoue, A. Classification of bulk metallic glasses by atomic size difference, heat of mixing and period of constituent elements and its application to characterization of the main alloying element. *Mater. Trans.* **46**, 2817–2829 (2005).
24. Bartel, C. J. et al. A critical examination of compound stability predictions from machine-learned formation energies. *Npj Comput. Mater.* **6**, <https://doi.org/10.1038/s41524-020-00362-y> (2020).
25. Manzoor, A., Pandey, S., Chakraborty, D., Phillpot, S. R. & Aidhy, D. S. Entropy contributions to phase stability in binary random solid solutions. *Npj Comput. Mater.* **4**, <https://doi.org/10.1038/s41524-018-0102-y> (2018).
26. Ward, L., Agrawal, A., Choudhary, A. & Wolverton, C. A general-purpose machine learning framework for predicting properties of inorganic materials. *Npj Comput. Mater.* **2**, <https://doi.org/10.1038/npjcompumats.2016.28> (2016).
27. Kaufmann, K. & Vecchio, K. S. Searching for high entropy alloys: a machine learning approach. *Acta Materialia* **198**, 178–222 (2020).
28. Xue, D. Z. et al. Accelerated search for materials with targeted properties by adaptive design. *Nat. Commun.* **7**, <https://doi.org/10.1038/ncomms11241> (2016).
29. Ren, F. et al. Accelerated discovery of metallic glasses through iteration of machine learning and high-throughput experiments. *Sci. Adv.* **4**, <https://doi.org/10.1126/sciadv.aag1566> (2018).
30. Ludwig, A. Discovery of new materials using combinatorial synthesis and high-throughput characterization of thin-film materials libraries combined with computational methods. *Npj Comput. Mater.* **5**, <https://doi.org/10.1038/s41524-019-0205-0> (2019).
31. Noack, M. M., Doerk, G. S., Li, R. P., Fukuto, M. & Yager, K. G. Advances in kriging-based autonomous X-ray scattering experiments. *Sci. Rep.* **10**, <https://doi.org/10.1038/s41598-020-57887-x> (2020).
32. Noack, M. M. et al. Autonomous materials discovery driven by Gaussian process regression with inhomogeneous measurement noise and anisotropic kernels. *Sci. Rep.* **10**, <https://doi.org/10.1038/s41598-020-74394-1> (2020).
33. Noack, M. M. et al. Gaussian processes for autonomous data acquisition at large-scale synchrotron and neutron facilities. *Nat. Rev. Phys.* <https://doi.org/10.1038/s42254-021-00345-y> (2021).
34. gpCAM (Version 7.3.4) (2022).
35. Noack, M. M. & Sethian, J. A. *Autonomous Discovery in Science and Engineering*. Medium: ED (United States, 2021).
36. Stein, H. S. & Gregoire, J. M. Progress and prospects for accelerating materials science with automated and autonomous workflows. *Chem. Sci.* **10**, 9640–9649 (2019).
37. Nikolaev, P. et al. Autonomy in materials research: a case study in carbon nanotube growth. *Npj Comput. Mater.* **2**, <https://doi.org/10.1038/npjcompumats.2016.31> (2016).
38. Szymanski, N. J. et al. Toward autonomous design and synthesis of novel inorganic materials. *Mater. Horiz.* <https://doi.org/10.1039/d1mh00495f> (2021).
39. McCue, I., Stuckner, J., Murayama, M. & Demkowicz, M. J. Gaining new insights into nanoporous gold by mining and analysis of published images. *Sci. Rep.* **8**, <https://doi.org/10.1038/s41598-018-25122-3> (2018).
40. Zhao, C. H. et al. Design nanoporous metal thin films via solid state interfacial dealloying. *Nanoscale* **13**, <https://doi.org/10.1039/d1nr03709a> (2021).
41. Lu, Z. W., Wei, S. H., Zunger, A., Frotapessoa, S. & Ferreira, L. G. 1st-principles statistical-mechanics of structural stability of intermetallic compounds. *Phys. Rev. B* **44**, 512–544 (1991).
42. McCue, I., Karma, A. & Erlebacher, J. Pattern formation during electrochemical and liquid metal dealloying. *Mrs Bulletin* **43**, 27–34 (2018).
43. Kusne, A. G. et al. On-the-fly closed-loop materials discovery via Bayesian active learning. *Nat. Commun.* **11**, <https://doi.org/10.1038/s41467-020-19597-w> (2020).
44. Koinuma, H. & Takeuchi, I. Combinatorial solid-state chemistry of inorganic materials. *Nat. Mater.* **3**, 429–438 (2004).
45. Arcot, B. et al. Intermetallic formation in copper/magnesium thin-films - kinetics, nucleation and growth, and effect of interfacial oxygen. *J. Appl. Phys.* **76**, 5161–5170 (1994).
46. Chen, Q. & Sieradzki, K. Mechanisms and morphology evolution in dealloying. *J. Electrochem. Soc.* **160**, C226–C231 (2013).
47. Lorenz, C. D., May, R. & Ziff, R. M. Similarity of percolation thresholds on the HCP and FCC lattices. *J. Stat. Phys.* **98**, 961–970 (2000).
48. McCue, I., Gaskey, B., Geslin, P. A., Karma, A. & Erlebacher, J. Kinetics and morphological evolution of liquid metal dealloying. *Acta Materialia* **115**, 10–23 (2016).
49. Artymowicz, D. M., Erlebacher, J. & Newman, R. C. Relationship between the parting limit for de-alloying and a particular geometric high-density site percolation threshold. *Philos. Mag.* **89**, 1663–1693 (2009).
50. Maffettone, P. M. et al. Crystallography companion agent for high-throughput materials discovery. *Nat. Comput. Sci.* **1**, 290–297 (2021).
51. Joo, S.-H., Wada, T. & Kato, H. Development of porous FeCo by liquid metal dealloying: Evolution of porous morphology and effect of interaction between ligaments and melt. *Mater. Des.* **180**, <https://doi.org/10.1016/j.matdes.2019.107908> (2019).
52. McCue, I. et al. Size effects in the mechanical properties of bulk bicontinuous Ta/Cu nanocomposites made by liquid metal dealloying. *Adv. Eng. Mater.* **18**, 46–50 (2016).
53. Okulov, I. V. et al. Surface functionalization of biomedical Ti-6Al-7Nb alloy by liquid metal dealloying. *Nanomaterials* **10**, <https://doi.org/10.3390/nano10081479> (2020).
54. Greenidge, G. & Erlebacher, J. Porous graphite fabricated by liquid metal dealloying of silicon carbide. *Carbon* **165**, 45–54 (2020).
55. Mokhtari, M. et al. Cold-rolling influence on microstructure and mechanical properties of NiCr-Ag composites and porous NiCr obtained by liquid metal dealloying. *J. Alloys Compd.* **707**, 251–256 (2017).
56. Mokhtari, M. et al. Mechanical properties of FeCr-based composite materials elaborated by liquid metal dealloying towards bioapplication. *Advanced Eng. Mater.* **22**, <https://doi.org/10.1002/adem.202000381> (2020).
57. Okulov, I. V. et al. Open porous dealloying-based biomaterials as a novel biomaterial platform. *Mater. Sci. Eng. C Mater. Biol. Appl.* **88**, 95–103 (2018).
58. Wada, T., Yamada, J. & Kato, H. Preparation of three-dimensional nanoporous Si using dealloying by metallic melt and application as a lithium-ion rechargeable battery negative electrode. *J. Power Sources* **306**, 8–16 (2016).
59. Gaskey, B., McCue, I., Chuang, A. & Erlebacher, J. Self-assembled porous metal-intermetallic nanocomposites via liquid metal dealloying. *Acta Materialia* **164**, 293–300 (2019).
60. Okulov, I. V., Weissmuller, J. & Markmann, J. Dealloying-based interpenetrating-phase nanocomposites matching the elastic behavior of human bone. *Sci. Rep.* **7**, 20-21-27, (2017).
61. Okulov, A. V., Volegov, A. S., Weissmuller, J., Markmann, J. & Okulov, I. V. Dealloying-based metal-polymer composites for biomedical applications. *Scripta Materialia* **146**, 290–294 (2018).
62. Okulov, I. V., Okulov, A. V., Volegov, A. S. & Markmann, J. Tuning microstructure and mechanical properties of open porous TiNb and TiFe alloys by optimization of dealloying parameters. *Scripta Materialia* **154**, 68–72 (2018).
63. Park, W. Y., Wada, T., Joo, S. H., Han, J. H. & Kato, H. Novel hierarchical nanoporous graphene nanoplatelets with excellent rate capabilities produced via self-templating liquid metal dealloying. *Mater. Today Commun.* **24**, <https://doi.org/10.1016/j.mtcomm.2020.101120> (2020).
64. Zhang, F. M. et al. Preparation of nano to submicro-porous TiMo foams by spark plasma sintering. *Adv. Eng. Mater.* **19**, <https://doi.org/10.1002/adem.201600600> (2017).
65. McCue, I. et al. Self-organized high-hardness thermal spray coatings. <https://doi.org/10.2139/ssrn.3793930> (2021).
66. Hayes, J. R., Hodge, A. M., Biener, J., Hamza, A. V. & Sieradzki, K. Monolithic nanoporous copper by dealloying Mn-Cu. *J. Mater. Res.* **21**, 2611–2616 (2006).
67. Chen, L.-Y., Yu, J.-S., Fujita, T. & Chen, M.-W. Nanoporous copper with tunable nanoporosity for SERS applications. *Adv. Funct. Mater.* **19**, 1221–1226 (2009).
68. Xu, C. X., Li, Y. Y., Tian, F. & Ding, Y. Dealloying to nanoporous silver and its implementation as a template material for construction of nanotubular mesoporous bimetallic nanostructures. *Chemphyschem* **11**, 3320–3328 (2010).
69. Rahman, M. A., Zhu, X. & Wen, C. Fabrication of nanoporous Ni by chemical dealloying Al from Ni-Al alloys for lithium-ion batteries. *Int. J. Electrochem. Sci.* **10**, 3767–3783 (2015).
70. Pedregosa, F. et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
71. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
72. Cortes, C. & Vapnik, V. Support-vector networks. *Mach. Learn.* **20**, 273–297 (1995).
73. Chen, T. Q., Guestrin, C. & Assoc Comp, M. XGBoost: A Scalable Tree Boosting System. Kdd'16: Proc. 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining, 785–794, <https://doi.org/10.1145/2939672.2939785> (2016).
74. Lundberg, S. M. & Lee, S. I. In 31st Annual Conference on Neural Information Processing Systems (NIPS). (2017).
75. Polikar, R. Ensemble based systems in decision making. *IEEE Circuits Syst. Mag.* **6**, 21–45 (2006).
76. Yamada, H. et al. Predicting materials properties with little data using shotgun transfer learning. *Acs Central Sci.* **5**, 1717–1730 (2019).

77. Ward, L. et al. Matminer: an open source toolkit for materials data mining. *Computat. Mater. Sci.* **152**, 60–69 (2018).
78. Goodall, R. Data of the maximum solid solubility limits of binary systems of elements. *Data Brief* **26**, <https://doi.org/10.1016/j.dib.2019.104515> (2019).
79. Kohavi, R. In *IJCAI'95: Proc. 14th International Joint Conference on Artificial Intelligence*. 1137–1143 (1995).
80. Wu, H., Mayeshiba, T. & Morgan, D. High-throughput ab-initio dilute solute diffusion database. *Sci. Data* **3**, <https://doi.org/10.1038/sdata.2016.54> (2016).
81. Arkilic, A. et al. Towards integrated facility-wide data acquisition and analysis at NSLS-II. *Synchrotron Radiation News* **30**, 44–45 (2017).
82. Ong, S. P. et al. Python Materials Genomics (pymatgen): a robust, open-source python library for materials analysis. *Comput. Mater. Sci.* **68**, 314–319 (2013).

Acknowledgements

This material is based on work supported by the National Science Foundation under Grant No. DMR-1752839. Y.-c.K.C.-W. acknowledges the support provided via the Faculty Early Career Development Program (CAREER) program and the Metals and Metallic Nanostructures Program of the National Science Foundation. This research used resources and the X-ray powder diffraction beamline (XPD, 28-ID-2) of the National Synchrotron Light Source II, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Brookhaven National Laboratory under Contract No. DE-SC0012704. This research used the Materials Synthesis and Characterization, and Nanofabrication Facilities of the Center for Functional Nanomaterials (CFN), which is a U.S. Department of Energy Office of Science User Facility, at Brookhaven National Laboratory under Contract No. DE-SC0012704. C.Z. and Y.-c.K.C.-W. thank the Joint Photon Science Institute at Stony Brook University, which provides partial support for C.Z. via a student fellowship that was jointly proposed by Y.-c.K.C.-W. as principal investigator (PI), Yong Chu as co-PI and Juergen Thieme and Wah-Keat Lee as collaborators. We also thank Varun Kankanallu for proofreading the manuscript. The work by M.M.N. was funded through the Center for Advanced Mathematics for Energy Research Applications (CAMERA), which is jointly funded by the Advanced Scientific Computing Research (ASCR) and Basic Energy Sciences (BES) within the Department of Energy's Office of Science, under Contract No. DE-AC02-05CH11231.

Author contributions

C.Z., K.G.Y., S.G., T.C., and Y.-c.K.C.-W. and developed the research idea. C.Z. and S.J. developed the ML prediction workflow with input from W.Z.; C.Z. developed the machine learning prediction code with support from S.J. J.-H.C. helped review the ML-related methodology. Y.-c.K.C.-W. and C.Z. wrote the user proposals for the use of XPD beamlines at NSLS-II and equipment at CFN with the inputs from K.Y. and S.G. C.Z. conducted the Ta and Mg thin-film deposition; K.M. conducted the combinatorial Ti-Cu

thin-film deposition under the direction of I.T.; C.Z. conducted a RTP heating process. J.L. and T.C. incorporated the XCA and gpCAM into the XRD data collection process, with inputs from M.N., P.M., D.O.M.F., and K.G. Y.C.Z., J.L., T.C., S.G., and Y.-c.K.C.-W. conducted the synchrotron diffraction experiments with inputs from K.G.Y., M.F.P.M., and D.O.; C.Z. and C.-C.C. conducted the XRD diffraction analysis under the guidance of H.Z. and Y.-c.K.C.-W. C.-C. conducted the XCA and gpCAM data analysis under the guidance of T.C., P.M., M.N., C.Z., and Y.-c.K.C.-W. wrote the manuscript, with inputs from all co-authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s43246-022-00303-w>.

Correspondence and requests for materials should be addressed to Chonghang Zhao or Yu-chen Karen Chen-Wiegart.

Peer review information *Communications Materials* thanks Helge Stein, Logan Ward and the other, anonymous, reviewers for their contribution to the peer review of this work. Primary Handling Editor: John Plummer.

Reprints and permission information is available at <http://www.nature.com/reprints>

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022