IJMMM International Journal of Minerals, Metallurgy and Materials 矿物冶金与材料学报（英文版）

Recent progress in the machine learning-assisted rational design of alloys

Huadong Fu, Hongtao Zhang, Changsheng Wang, Wei Yong, and Jianxin Xie

Cite this article as:
Huadong Fu, Hongtao Zhang, Changsheng Wang, Wei Yong, and Jianxin Xie, Recent progress in the machine learning-assisted rational design of alloys, *Int. J. Miner. Metall. Mater.*, 29(2022), No. 4, pp. 635-644. https://doi.org/10.1007/s12613-022-2458-8

View the article online at SpringerLink or IJMMM Webpage.

International Journal of Minerals, Metallurgy and Materials
Volume 29, Number 4, April 2022, Page 635
https://doi.org/10.1007/s12613-022-2458-8

Invited Review

Recent progress in the machine learning-assisted rational design of alloys

Huadong Fu[1,2,3], *Hongtao Zhang*[1,2,3], *Changsheng Wang*[1,2,3], *Wei Yong*[2], *and Jianxin Xie*[1,2,3]✉

1) Beijing Advanced Innovation Center for Materials Genome Engineering, University of Science and Technology Beijing, Beijing 100083, China
2) Key Laboratory for Advanced Materials Processing (MOE), University of Science and Technology Beijing, Beijing 100083, China
3) Beijing Laboratory of Metallic Materials and Processing for Modern Transportation, University of Science and Technology Beijing, Beijing 100083, China
(Received: 8 January 2022; revised: 2 March 2022; accepted: 3 March 2022)

Abstract: Alloys designed with the traditional trial and error method have encountered several problems, such as long trial cycles and high costs. The rapid development of big data and artificial intelligence provides a new path for the efficient development of metallic materials, that is, machine learning-assisted design. In this paper, the basic strategy for the machine learning-assisted rational design of alloys was introduced. Research progress in the property-oriented reversal design of alloy composition, the screening design of alloy composition based on models established using element physical and chemical features or microstructure factors, and the optimal design of alloy composition and process parameters based on iterative feedback optimization was reviewed. Results showed the great advantages of machine learning, including high efficiency and low cost. Future development trends for the machine learning-assisted rational design of alloys were also discussed. Interpretable modeling, integrated modeling, high-throughput combination, multi-objective optimization, and innovative platform building were suggested as fields of great interest.

Keywords: machine learning; data mining; rational design; alloys

1. Introduction

The rapid development of high-end manufacturing fields, such as aerospace, energy and power, and electronic information, poses huge challenges to the research and property improvement of metallic materials [1–4]. Examples of these materials include corrosion-resistant ultrahigh-strength and high-toughness aluminum alloys used for the structural parts of large passenger aircraft [5], directional or single crystal superalloys used for the core hot-end parts of heavy-duty gas turbines [6], and high-strength and high-conductivity copper alloys used for high-end integrated circuit lead frames and electronic component connectors [7].

Traditional research and development methods for metallic materials mainly rely on plenty of experimental trials assisted by instructive theoretical calculations or simulations; this process requires long research cycles and high costs, and the high-efficiency and low-cost design requirements of high-end key metallic materials are difficult to meet. Approximately 10–20 years are needed to reach the final industrial application, and the iteration of new materials seriously lags behind the product design. To accelerate material research and meet the developmental needs of high-end manufacturing, the United States proposed the "Material Genome Initiative," and China launched the "Material Genome Engineering National Key Research and Development Program." These strategies aim to shorten the research cycle and reduce the research cost by developing common key technologies, such as high-throughput computation, high-throughput experiments, and big data technology [8].

In recent years, data-driven machine learning-assisted material design has rapidly developed, and the research of metallic material design based on machine learning prediction has achieved important breakthroughs in material structure and performance prediction, composition design, and process optimization [9–12]. In microstructure and property prediction, machine learning models are used to directly predict the microstructure and properties of metallic materials. For example, density functional theory (DFT) calculation data were used by Wang *et al.* [13] to build a model predicting the elastic constant of alloys and by Huber *et al.* [14] to build a model predicting the solute grain boundary segregation energy of alloys. In composition design, the type and content of alloying elements are considered, and the physical and chemical features of materials are used as input. Agrawal and Choudhary [15] applied elemental compositions and processing parameters as input and selected a suitable algorithm from 40 available ones to model and predict fatigue strength and optimize steel composition. Huang *et al.* [16] analyzed 401 sets of sample data obtained from literature, extracted empirical features, and established a classification model of the high-entropy alloy phase. Important progress in solving the time-consuming and laborious design and parameter optimization of alloys has also been achieved through

✉ Corresponding author: Jianxin Xie E-mail: jxxie@mater.ustb.edu.cn

machine learning-assisted design [9,10,17]. Chen *et al.* [17] proposed a machine learning-aided process parameter optimization strategy based on Pareto front analysis for Mg alloy aging treatment.

This article focused on the machine learning-assisted rational design of alloys and the related improvement on research efficiency. Combining relevant research works, this study introduced the basic strategy for the machine learning-assisted rational design of alloys and reviewed the research progress in the following three areas: property-oriented reversal design of alloy composition, screening design of alloy composition based on models established using element physical and chemical features or microstructure factors, and optimal design of alloy composition and process parameters based on iterative feedback optimization. Future development trends for the machine learning-assisted rational design of alloys were also discussed. This paper provides a reference for related scientific and technological workers interested in this field.

2. Basic strategy for material design via machine learning

The basic strategy of applying machine learning to metallic material design includes three aspects: data collection and processing, machine learning model construction and validation, and material design, as shown in Fig. 1. Dataset is the basis of machine learning modeling, and the amount and quality of data are the keys to determining the accuracy and generalization ability of the machine learning model. High accuracy is a prerequisite for the application of the machine learning model in the rational design of materials because this parameter affects the prediction reliability during material design. Machine learning models often require appropriate algorithms based on the characteristics of the research problem. Material design is the ultimate goal of machine learning prediction to find high-performance alloys that meet the design requirements.

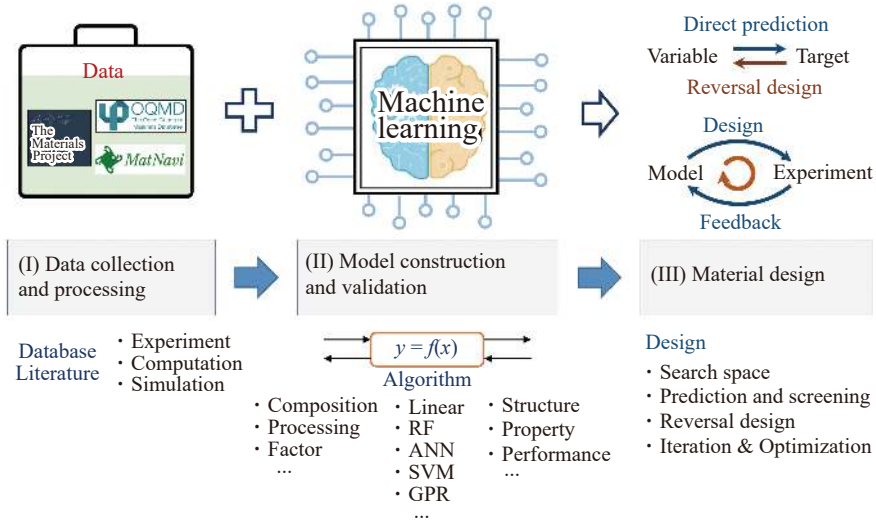

Fig. 1. Basic strategy for machine learning-assisted metallic material design. RF—Rondom forest; ANN—Artificial neural network; SVM—Support vector machine; GPR—Gaussian process regression.

2.1. Data collection and preparation

Data sources for metallic materials mainly include literature collection, database acquisition [11] (as shown in Table 1), simulation calculation, or experimental acquisition. The data types mainly include numerical data, character data, picture data, and voice or video.

Datasets collected from literature or databases usually have problems, such as data duplication, missing, and outliers, making them difficult to directly use in machine learning modeling. Hence, data preprocessing is required. Commonly used data preprocessing methods for metallic material design mainly include removing unique attributes, processing missing values, and standardizing data. The data scale is standardized by making the data obey a normal distribu-

Table 1. Databases for metallic material design

Name	Description	URL
AFLOWLIB	First principle high-throughput computed structures and properties of inorganic materials	http://aflowlib.org
Materials Project	Open web-based tool for first principle computed structures and properties of known and predicted materials	https://materialsproject.org
OQMD	Open Quantum Materials Database, first principle computed structures and properties	http://oqmd.org
ICSD	Inorganic Crystal Structure Database	https://icsd.fiz-karlsruhe.de
MatNavi	The NIMS Materials Database including polymer, inorganic materials, metallic materials, etc.	http://mits.nims.go.jp
MatWeb	The online materials information resource for various engineering materials, including thermoplastics, semiconductors, and fibers	http://matweb.com

tion to smoothly facilitate machine learning modeling [18].

2.2. Model construction

Machine learning modeling aims to establish a function $y = f_{model}(x)$ between input and output and make $f_{model}(x)$ as close to the real function relationship $f_{real}(x)$ as possible by optimizing the model parameters. Before the model is established, the input variable x and output variable y of the specific research problem must be clarified. In the research of metallic materials, y commonly includes the physical and chemical properties, macroscopic properties, and microstructure of the material. In the inverse design of alloy composition, the output variable y is the type and content of chemical elements. Related research [19–20] pointed out that in the study of the influence of composition on material properties, extracting the physical and chemical factors of the element and converting the composition into physical and chemical features as input can improve the accuracy and generalization ability of the model.

No single strategy has been established for machine learning modeling, and different research objects generally need different algorithms [21]. Therefore, a suitable machine learning algorithm must be selected according to the distribution of the dataset and the relationship between y and x when building a machine learning model. Commonly used machine learning algorithms mainly include linear algorithms, decision tree-based algorithms [22], artificial neural networks [23], support vector machines [24], and Bayesian-based algorithms [25]. Linear-based algorithms are suitable for handling simple relationships with simple operation and strong interpretability. Algorithms based on decision trees have strong interpretability and are suitable for processing high-dimensional data. For example, random forest decreases the influence of inputs that contribute less to the output when modeling; this algorithm is appropriate for problems with a large number of inputs and a strong correlation between input variables. An artificial neural network requires a certain amount of sample data because the model parameters are optimized through iteration. Support vector machine maps the data to high dimensions through the kernel function and solves the model parameters by convex function; the modeling speed is fast, and the algorithm is suitable for small datasets. Bayesian-based algorithms describe the problem in a probabilistic manner; for example, Gaussian process regression can quantify the prediction uncertainty through the a priori inference of the posterior and is suitable for datasets containing noise.

2.3. Material design and experimental verification

The ultimate goal of machine learning modeling is to guide the design and development of materials. After the mapping relationship between material features and target variables has been established by machine learning, the target variables (such as material properties) can be predicted in an unknown space to explore the optimal material composition and processing. In addition, machine learning modeling can also be combined with optimization algorithms to efficiently explore the unknown space globally and realize the optimal design of materials. After the optimal metallic material design is obtained using the above two ways, the reliability of the design is experimentally verified. In alloy composition optimization and effective design of process parameters, experimentation is an indispensable step for iterative design. The experimental conditions and methods used in experimental verification or iterative experiments should be consistent with the dataset to reduce errors generated by the experimental system.

3. Three typical rational designs based on machine learning

In recent years, machine learning-assisted material design has rapidly developed and has been widely studied [9–17]. To change the traditional experience-based trial and error research model and realize the rational design of alloys, we have developed three typical methods. One is to achieve property-oriented alloy composition design, that is, alloy composition inverse design with the required properties. The second is to explore the essential physical and metallurgical factors of composition that affect properties and create the alloy composition screening design on the basis of the models using these material features as input. The third is to optimize the design of alloy composition/process parameters based on iterative feedback optimization.

3.1. Property-oriented inverse design of alloys

Compared with the trial and error method guided by human experience, the use of data-driven machine learning methods for establishing the relationship among the composition, structure, and property of materials and guiding material research and development can significantly improve the efficiency and reduce the cost of material research, engineering, and application. In engineering applications, material property requirements are first proposed according to the specific application environment, and suitable materials are then designed according to the property requirements. This method is called inverse design [26], an effective way to break the existing material design experience and explore the unknown material space and a solution for "tailor-made" design materials to meet the application requirements. At present, studies on metallic materials have generated a large number of candidate component spaces according to the requirements of target property and designed materials that meet the performance requirements by using optimization algorithms, such as simulated annealing algorithm [27–29], genetic algorithm [29–30], and particle swarm algorithm [31]. Nevertheless, the inverse design of alloy composition remains difficult to achieve through reverse modeling or encoding–decoding.

With reference to the idea of generative adversarial nets, the authors [32] developed an alloy composition inverse design system Machine Learning Design System (MLDS)

using back propagation neural network modeling [33]. The system integrates two neural network modules of composition→property forward model prediction and property→composition design to realize the inverse design of alloy composition oriented to property requirements, as shown in Fig. 2. MLDS includes three subsystems: model training, composition design, and property prediction. After parameter optimization based on a given sample, a C2P model with high reliability (component→performance) and a P2C model with low reliability (performance→component) were obtained. The P2C model was then used to screen the potential alloy composition in the unknown composition space, and the selected alloy composition was employed as the input in the C2P model to predict the target property. The alloy composition provided by the P2C model was evaluated according to the target property using the prediction error of property as the criterion: if the prediction error exceeds a threshold, then the MLDS will be trained and designed again to screen out the appropriate alloy composition. On the basis of hundreds of collected sample data, a new high-strength and high-conductivity copper alloy Cu–3.00Ni–0.60Si–0.16Zn–0.15Cr–0.03P (ultimate tensile strength 775 ± 10 MPa, conductivity 48.0% ± 0.5% IACS, International Annealed Copper Standard) was developed as a candidate material for large-scale integrated circuit lead frames.

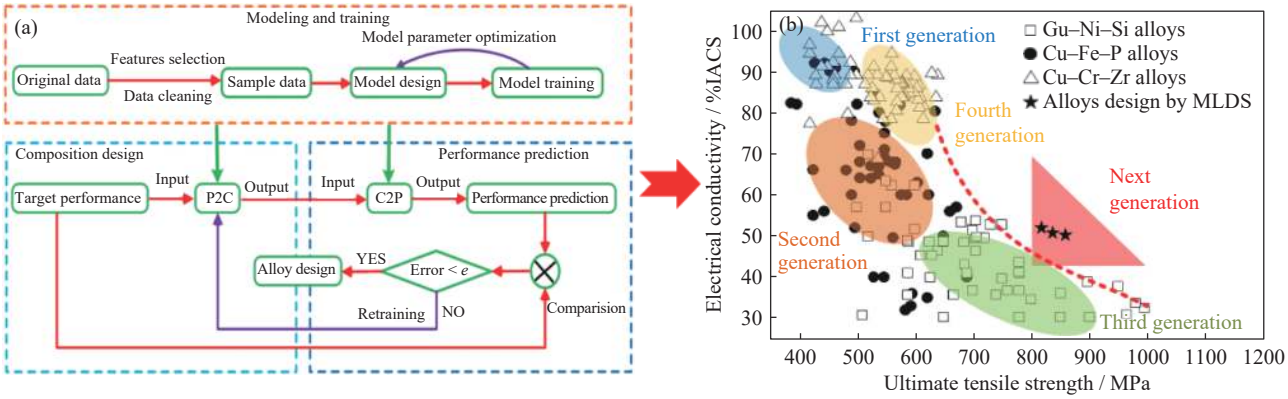

Fig. 2. MLDS system (a) and copper alloy design with high strength and high conductivity (b) [32].

Jiang *et al.* [34] used MLDS to design three groups of ultrahigh-strength, high-toughness aluminum alloys under ultimate tensile strength of 700–750 MPa, fracture toughness (K_{IC}) 33–35 MPa·m$^{1/2}$, and elongation (δ) 8%–10%. In each group of designed alloy, a typical alloy composition was selected for experimental verification. As shown in Fig. 3, the three typical aluminum alloys have achieved ultrahigh strength, high elongation, and high fracture toughness. This case study serves as a reference for the design of complex alloys with multi-objective property requirements and further confirms the feasibility of adopting MLDS for the inverse design of alloy composition.

3.2. Screening design of alloy compositions based on material features

During the initial application of machine learning in the field of metallic materials, researchers first tried to use machine learning methods to establish a model of the relationship between material composition and microstructure/properties and then to develop new materials [35–40]. However,

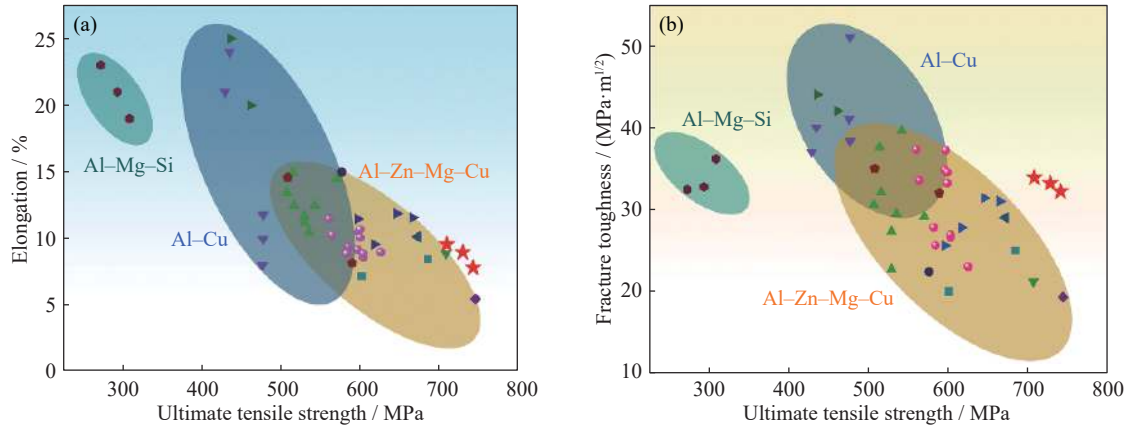

Fig. 3. Comprehensive property comparison between the ultrahigh-strength and high-toughness aluminum alloys designed by MLDS and the reported aluminum alloy: (a) ultimate tensile strength and elongation; (b) ultimate tensile strength and fracture toughness [34]. Reprinted from *J. Mater. Sci. Technol.*, 98, L. Jiang, C.S. Wang, H.D. Fu, J. Shen, Z.H. Zhang, and J.X. Xie, Discovery of aluminum alloys with ultra-strength and high-toughness via a property-oriented design strategy, 33-43, Copyright 2022, with permission from Elsevier.

the machine learning model based on alloy element content and property value is a model that knows the relationship but not the reason (labeled as a phenomenological model in this paper). This model cannot directly reveal the physical and chemical characteristics of the elements, the interactions and reactions between the elements, the types and volume fractions of the phases, and other physical and chemical mechanisms or metallurgical mechanisms that affect the properties of materials. Therefore, the screening and replacement design of alloying elements and the discovery of new alloying elements and new alloys will be difficult to realize with the phenomenological model. Machine learning modeling (mechanism model) based on the analysis of physicochemical features of alloy elements or material microstructure features (material features) is expected to solve these problems. This section introduces the research progress in metallic material selection and design based on material features from three aspects: construction, screening, and application of material features. Among these, material features can be constructed by applying the physicochemical parameters of elements, simulation calculation, and domain knowledge. Prior to modeling, a large number of material features must be screened to avoid the disaster of dimension and obtain the optimal subset of material features for material design.

3.2.1. Construction of material features

The informatic description of materials contains the following: description of alloy element types and physicochemical characteristics; direct description of material structures, such as cell constants; indirect description of organizations, such as phase volume fractions; and description of material system characteristics, such as entropy and enthalpy related to thermodynamics. These descriptions are collectively referred to as material features. Pearson manual [41] provides a variety of physicochemical characteristics corresponding to each element, including atomic size factor, electrochemical factor, thermodynamic factor, Mendeleev number factor, and cohesive energy factor. Some studies [42–45] used these elemental physicochemical characteristics and material composition to construct the factors representing the material by calculating the mean and standard deviation, known as "descriptors." The descriptors constructed by this method exhibit stability and strong interpretability. Stability is reflected as the ability to always map the components to the material feature space in the same dimension, regardless of changes in the element type. Interpretability is reflected as the intrinsic relationship between the basic features of materials and properties and can be used to intuitively analyze the mechanism of components affecting properties.

In addition to the physicochemical characteristics of elements used to construct descriptors, the features representing the inherent properties of materials can also be established by simulation calculation. For example, material features such as energy, electronic structure, stability of precipitates, coherent relationship with matrix and material elastic modulus can be obtained by first-principles calculations [46–48]. However, simulation calculation often requires large calculations and time costs. By contrast, the use of domain knowledge is a convenient and effective way to guide the selection of material features. For example, in the phase classification of high-entropy alloy, multiple empirical parameters are selected as material features on the basis of the ideal mixing entropy that can reflect the mixing degree of each phase or other related domain knowledge; good results have been achieved with this approach [49–51].

With the use of the above methods to construct material features, the obtained material features can also be combined with basic mathematical operators; the space of material features can be further expanded through nonlinear operations, which can be applied to the mining of other nonlinear relationships [52–53].

3.2.2. Screening of material features

With the continuous development of computer algorithms and material science knowledge, many material feature construction methods have been developed and applied to the rational design of alloys. However, dimension disaster arose from the huge set of material features. A solution is to establish an accurate model that efficiently and quickly screens the features with an important effect on the target parameters. Therefore, material feature screening methods based on mathematical statistics and machine learning algorithms have been developed rapidly. Some examples include filter screening, embedded screening, and wrapper screening.

Filter screening considers the information gain or Pearson correlation [54] between the material characteristics and target parameters. Features are sorted by importance, and those not linearly related to the target parameters are eliminated. Although this method can quickly eliminate a large number of material features, the nonlinear relationship between features and target parameters has not been taken into consideration. Embedded filtering aims to select material features while constructing classifiers; examples include least absolute shrinkage and selection operator [55] and random forest importance ranking [56]. The above two screening methods focus on the importance of the influence of features on target parameters, but the dependence between features is neglected. Considering the coupling effect between multiple material features, the wrapper screening method is often used in greedy search strategies, such as backward recursive feature elimination [57], which fully considers the impact of feature subsets on target parameters but has a relatively low screening efficiency. This method screens features by judging the accuracy of the model, and each feature is treated equally. Domain knowledge can also be filtered until the features are selected to retain the focus on features. For example, a data-driven multi-layer feature selection method incorporating domain expert knowledge (DML-FS_{dek}) was proposed [58], in which the domain expert knowledge is quantified by weighted scoring and integrated into the selection process to eliminate the risk of key features being deleted.

3.2.3. Application of material features

Clarifying the key material features that affect the target parameter is conducive to mining the impact mechanism of

composition on target parameters such as physicochemical properties and realizing a rational and efficient design through key feature modeling. By using material feature analysis combined with correlation screening, recursive elimination, and exhaustive method, the authors [59] found that the key features affecting the properties of solution-strengthened conductive copper alloys are absolute electronegativity, atomic radius, and nuclear electron distance. They then proposed a new method to design high-performance alloys according to the influence of key material features on alloy properties. A new alloy element, indium, which can significantly improve the tensile strength and conductivity of the alloy, was found by screening the possible solid solution elements in the periodic table. Four new solution-strengthened conductive copper alloys with indium content less than 0.7wt% were designed and prepared, and their comprehensive properties were significantly higher than those of the existing solution-strengthened conductive copper alloys, as shown in Fig. 4. The key material features were used to predict the properties of the solution-strengthened conductive aluminum alloy, and the prediction model with an error of less than 10% was established. This study shows that these material features are also the key intrinsic factors affecting the properties of solution-strengthened conductive aluminum alloys.

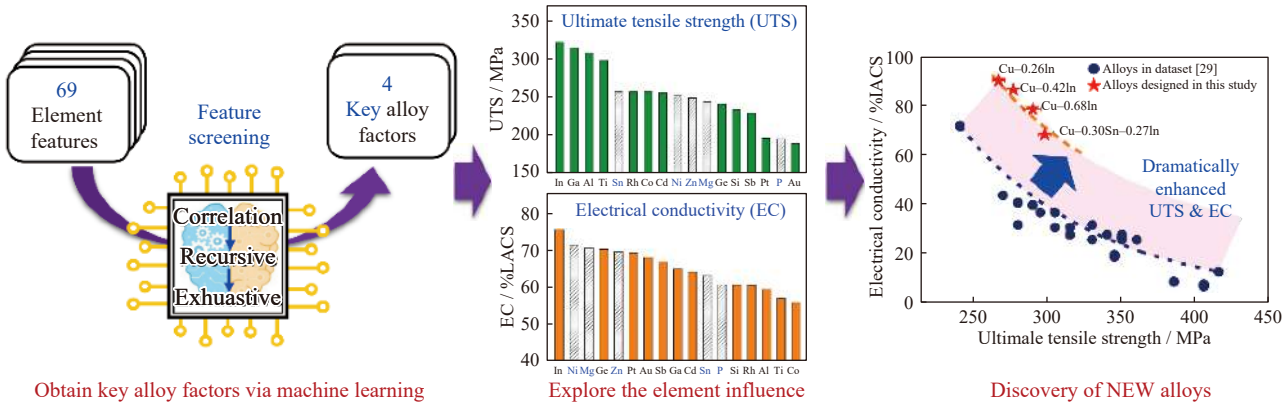

Fig. 4. Composition design and effect of high-performance copper alloy based on key material features [59]. Reprinted from *Acta Mater.*, 200, H.T. Zhang, H.D. Fu, X.Q. He, C.S. Wang, L. Jiang, L.Q. Chen, and J.X. Xie, Dramatically enhanced combination of ultimate tensile strength and electric conductivity of alloys via machine learning screening, 803-810, Copyright 2020, with permission from Elsevier.

Material features guided by domain knowledge have been widely used. Yeh [60] proposed the use of various empirical material features to find the rules controlling the phase stability of high-entropy alloys, including mixing enthalpy (ΔH_{mix}) and mixing entropy (ΔS_{mix}). Wen *et al.* [61] constructed the material features of solution-strengthened high-entropy alloys using domain knowledge and screened those features and the key parameters of the traditional solution strengthening models (S-, T-, and V-models) by applying a variety of machine learning algorithms. They found that poor atomic electronegativity is the control parameter of solution-strengthened high-entropy alloys. Finally, the mathematical expression of solution strengthening based on poor electronegativity was established. The model is simple, easy to use, and has improved accuracy compared with the traditional model.

3.3. Optimization design of alloys based on iterative feedback optimization

A machine learning model with high prediction accuracy and great generalization ability is difficult to establish using existing data, especially when a new alloy or a new process is being developed or when the sample data are difficult to obtain due to experimental difficulties or high costs. Hence, the active learning method that uses the experimental iterative feedback optimization method to reduce the number of required experiments has attracted attention. The main point of this technique is to design experimental points according to the predicted value and prediction uncertainty of the machine learning model and then iteratively optimize the next modeling and material design through the feedback of the experimental results until the model prediction meets the requirements [9]. The specific method includes four steps: data acquisition, machine learning model construction, experimental point design, experimental testing, and data feedback. Among these, the experimental point design strategy is the key to improving design efficiency.

Commonly used experimental point design methods mainly include selection according to model prediction and selection based on Bayesian optimization utility function. The Bayesian optimization method calculates the utility function through the model uncertainty and uses the value of the utility function to design experimental points, thereby balancing exploration and exploitation and exhibiting high optimization efficiency. One example is the widely used expected improvement (EI) utility function. EI obtains the predicted value μ and prediction uncertainty σ by employing modeling algorithms, such as Gaussian process regression, and then calculates the target expected improvement using the model predicted value and prediction uncertainty as

shown in Eq. (1).

$$\mathrm{EI} = \sigma[\varphi(z) + z\phi(z)] \quad (1)$$

where $\varphi(z)$ and $\phi(z)$ are the standard normal density and distribution functions (integral in $(z, +\infty)$), $z = (\mu - \mu^*)/\sigma$, and μ^* is benchmark target.

The authors [62] used the Bayesian optimization algorithm to balance global optimization and local optimization by calculating and comparing utility functions. As a result, the optimization efficiency was improved. Given its efficacy in solving the design problem of insufficient data, this method is often applied to the rapid optimization design of the composition and process parameters for complex alloys. In this paper, this method is labeled as optimization design.

3.3.1. Composition design

Experimental iterative active learning has less dependence on the amount of data and high design efficiency and has a substantially effect on the metallic material composition design. This method has been applied to the composition design of various metallic materials, such as shape memory alloys [63–64] and high-entropy alloys [65]. Xue *et al.* [63] used the active learning strategy to optimize the composition of Ti–Ni-based shape memory alloys with minimal thermal hysteresis of phase change. On the basis of 22 experimental data, a Support Vector Machine model was established using features such as valence electron concentration and electronegativity as inputs. Afterward, the EI values of candidate compositions were calculated, and the alloy composition with the largest EI value was selected for experimental verification. The verification value was fed back to the machine learning model for iteration. After six iterations of feedback, a new type of shape memory alloy, Ti–Ni–Cu–Fe–Pd, was successfully developed; this alloy has a phase change thermal hysteresis value (1.84 K) lower than that of the existing samples in the original data set.

Owing to the complexity of the composition–process–microstructure–property intrinsic relationship and application scenarios of metallic materials, attention must be paid to a variety of property indicators during design and preparation. Some of the properties often show conflicting relationships, such as the strength and plasticity of steel and the strength and conductivity of copper alloys. Therefore, traditional design methods cannot be applied to simultaneously improve the conflict properties of mechanical and electrical conductivity of complex alloys. The authors [62] proposed an alloy design strategy based on Bayesian optimization: combining key alloy factor screening with Bayesian optimization, constructing a multi-objective utility function to design alloy composition, and iteratively optimizing the properties of the alloy. One example is the aging precipitation-strengthened complex copper alloy. First, the combination of correlation screening, recursive elimination, and exhaustive screening was used. Five key alloy factors that affect alloy hardness and six key alloy factors that affect electric conductivity were screened. The relationship models for hardness–key alloy factor and electrical conductivity-key alloy factor with errors less than 7% and 9%, respectively, were then established. Finally, the Bayesian optimization algorithm was used to design the composition of the copper alloy. After iteration through experimental tests, the Cu–1.3Ni–1.4Co–0.56Si–0.03Mg alloy with excellent comprehensive properties was found. The measured tensile strength and electric conductivity reached 858 MPa and 47.6% IACS, respectively, which are better than those of the previously reported high-strength and medium-conductivity Cu–Ni–Co–Si alloy. Therefore, the simultaneous improvement of contradictory mechanical and electrical properties has been achieved, as shown in Fig. 5.

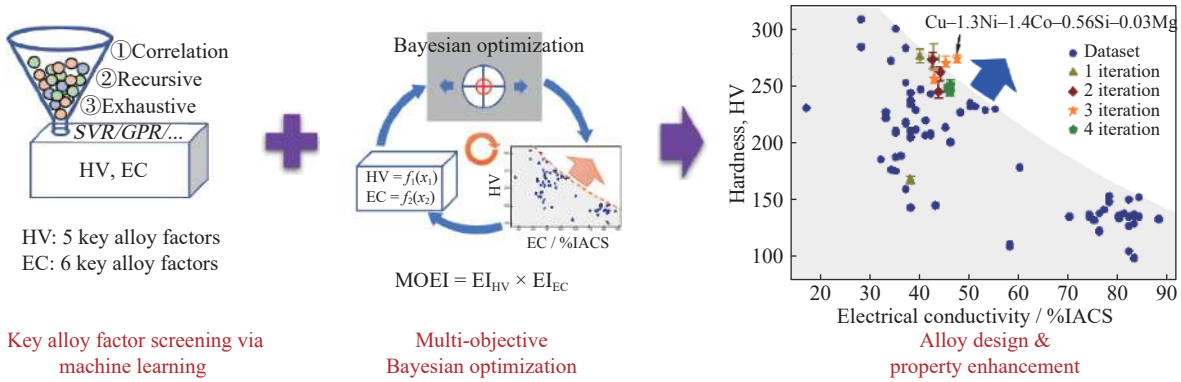

Fig. 5. Composition optimization design of copper alloy [62]. HV is hardness and EC is electrical conductivity. Reprinted from *Acta Mater.*, 215, H.T. Zhang, H.D. Fu, S.C. Zhu, W. Yong, and J.X. Xie, Machine learning assisted composition effective design for precipitation strengthened copper alloys, 117118, Copyright 2021, with permission from Elsevier.

3.3.2. Process parameter optimization

Process parameter optimization is another important application field of experimental iterative feedback optimization. The fabrication of metallic materials is complex, and the formulation and optimization of process parameters are time-consuming and laborious, especially for newly designed alloys that often lack sufficient data support. Therefore, reducing the number of process optimization experiments through experimental feedback active learning is important to rapidly achieve the design and optimization of process parameters. Liu *et al.* [66] used active learning to optimize the hardness of Mg–Al–Sn–Zn–Ca–Mn alloy and established an XGBoost machine learning model with variables such as composition, aging temperature, and aging time. The alloy com-

position and process parameters exceeding the highest hardness value in the initial dataset were found after only two experimental iterations.

The authors [67] established a dual-utility function optimization design based on the Bayesian optimization adaptive iterative algorithm by using mechanical and electrical properties as the optimization objectives. They quickly designed the deformation–aging process parameters for Cu–Ni–Co–Si alloy, as shown in Fig. 6. After four iterations, the process parameters that can simultaneously improve the mechanical and electrical properties were obtained (cold rolling deformation 90%, aging temperature 450°C, and aging time 1.25 h). The alloy hardness, tensile strength, and electric conductivity reached HV 285 ± 4, 872 ± 3 MPa, and 44.2% ± 0.7% IACS, respectively; the product of strength and conductivity increased by 7.8%; and the number of experiments was reduced by 99.4% compared with that in the traditional trial and error method. This research has overcome the problems of the long experiment design cycle, low efficiency, and high cost of alloy deformation–aging parameters and provides a new idea for the rapid design of material processing parameters.

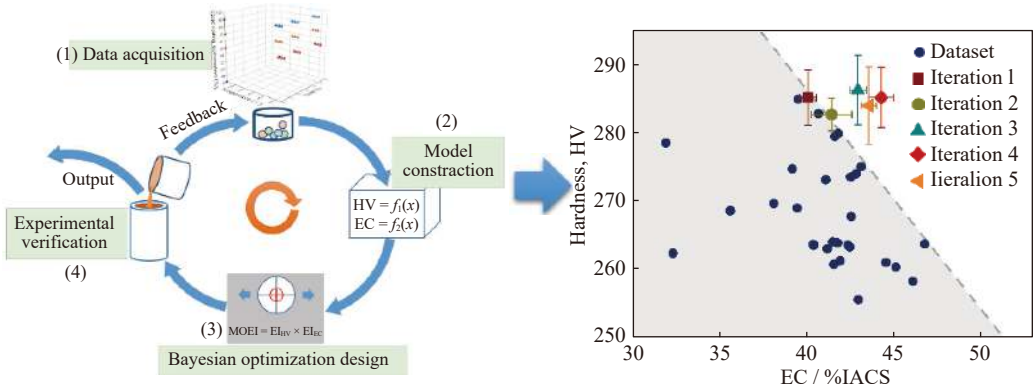

Fig. 6. Processing parameter optimization for copper alloy design (modified from [67]).

4. Summary and outlook

Data-driven machine learning has rapidly developed into a disruptive method for rational design and efficient research of metallic materials and effectively broke through the limitations of traditional experience-based trial and error methods. Through relevant research, this article reviewed the machine learning-assisted rational and efficient design of alloys. Research progress was reviewed by focusing on three aspects: the inverse design of alloy composition for property requirements, the screening design based on the physical and chemical features of alloy elements or the structural factors of materials, and the alloy composition and processing parameter optimization based on iterative feedback.

Future research on the machine learning-assisted rational design of alloys may focus on the following: (1) combining the mechanism of material science with data models to establish an interpretable machine learning model that reveals the intrinsic characteristics of material composition–process–microstructure–property relations and to achieve the alloy element screening design, alternative design, and rapid discovery of new alloys; (2) developing research methods that integrate computational material engineering and machine learning, break through the cross-scale modeling and design problems of alloys, and realize the integrated design of metallic material composition–process–property; (3) combining machine learning with material high-throughput calculations and experiments to quickly and efficiently discover and verify new alloys; (4) developing multi-objective collaborative optimization algorithms for engineering-applied alloys according to the comprehensive service performance requirements; and (5) building data-driven innovative research software and platforms for alloys and promoting the industrial application of machine learning.

Acknowledgements

This work was financially supported by the National Key Research and Development Program of China (No. 2021YFB3803101), National Natural Science Foundation of China (Nos. 51974028 and 52022011), and the Beijing Municipal Science and Technology Commission (No. Z191100001119125).

Conflict of Interest

The authors declare that they have no known competing interests.

References

[1] P.H. Abelson, Materials research and applications, *Science*, 274(1996), No. 5291, p. 1283.
[2] K. Lu, The future of metals, *Science*, 328(2010), p. 319.
[3] U. Bhandari, C.Y. Zhang, S.M. Guo, and S.Z. Yang, First-principles study on the mechanical and thermodynamic properties of MoNbTaTiW, *Int. J. Miner. Metall. Mater.*, 27(2020), No. 10, p. 1398.
[4] Z.S. Nong, H.Y. Wang, and J. C. Zhu, First-principles calculations of structural, elastic and electronic properties of $(TaNb)_{0.67}(HfZrTi)_{0.33}$ high-entropy alloy under high pressure, *Int. J. Miner. Metall. Mater.*, 27(2020), No. 10, p. 1405.
[5] J.P. Immarigeon, R.T. Holt, A.K. Koul, L. Zhao, W. Wallace,

and J.C. Beddoes, Lightweight materials for aircraft applications, *Mater. Charact.*, 35(1995), No. 1, p. 41.
[6] T.M. Pollock, Alloy design for aircraft engines, *Nat. Mater.*, 15(2016), No. 8, p. 809.
[7] Z. Li, Z. Xiao, Y.B. Jiang, Q. Lei, and J.X. Xie, Composition design, phase transition and fabrication of copper alloys with high strength and electrical conductivity, *Chin. J. Noferrous Met.,* 29(2019), No. 9, p. 2009.
[8] Y.J. Su, D.W. Zhang, Q. Feng, and J.X. Xie, A vision of materials genome engineering in China, *Engineering*, 2022. DOI: 10.1016/j.eng.2021.12.008
[9] J.X. Xie, Y.J. Su, D.Z. Xue, X. Jiang, H.D. Fu, and H.Y. Huang, Machine learning for materials research and development, *Acta Metall. Sinica*, 57(2021), No. 11, p. 1343.
[10] Z.H. Deng, H.Q. Yin, X. Jiang, C. Zhang, G.F. Zhang, B. Xu, G.Q. Yang, T. Zhang, M. Wu, and X.H. Qu, Machine-learning-assisted prediction of the mechanical properties of Cu–Al alloy, *Int. J. Miner. Metall. Mater.*, 27(2020), No. 3, p. 362.
[11] K.T. Butler, D.W. Davies, H. Cartwright, O. Isayev, and A. Walsh, Machine learning for molecular and materials science, *Nature*, 559(2018), No. 7715, p. 547.
[12] T. Lookman, P.V. Balachandran, D.Z. Xue, and R.H. Yuan, Active learning in materials science with emphasis on adaptive sampling using uncertainties for targeted design, *npj Comput. Mater.*, 5(2019), art. No. 21.
[13] J. Wang, X.Y. Yang, Z. Zeng, X.L. Zhang, X.S. Zhao, and Z.G. Wang, New methods for prediction of elastic constants based on density functional theory combined with machine learning, *Comput. Mater. Sci.*, 138(2017), p. 135.
[14] L. Huber, R. Hadian, B. Grabowski, and J. Neugebauer, A machine learning approach to model solute grain boundary segregation, *npj Comput. Mater.*, 4(2018), art. No. 64.
[15] A. Agrawal and A. Choudhary, An online tool for predicting fatigue strength of steel alloys based on ensemble data mining, *Int. J. Fatigue*, 113(2018), p. 389.
[16] W.J. Huang, P. Martin, and H.L. Zhuang, Machine-learning phase prediction of high-entropy alloys, *Acta Mater.*, 169(2019), p. 225.
[17] Y.F. Chen, Y. Tian, Y.M. Zhou, D.Q. Fang, X.D. Ding, J. Sun, and D.Z. Xue, Machine learning assisted multi-objective optimization for materials processing parameters: A case study in Mg alloy, *J. Alloys Compd.*, 844(2020), art. No. 156159.
[18] A. Famili, W.M. Shen, R. Weber, and E. Simoudis, Data preprocessing and intelligent data analysis, *Intell. Data Anal.*, 1(1997), No. 1-4, p. 3.
[19] P. Raccuglia, K.C. Elbert, P.D.F. Adler, C. Falk, M.B. Wenny, A. Mollo, M. Zeller, S.A. Friedler, J. Schrier, and A.J. Norquist, Machine-learning-assisted materials discovery using failed experiments, *Nature*, 533(2016), No. 7601, p. 73.
[20] R.H. Yuan, Z. Liu, P.V. Balachandran, D.Q. Xue, Y.M. Zhou, X.D. Ding, J. Sun, D.Z. Xue, and T. Lookman, Accelerated discovery of large electrostrains in $BaTiO_3$-based piezoelectrics using active learning, *Adv. Mater.*, 30(2018), No. 7, art. No. 1702884.
[21] D.H. Wolpert and W.G. Macready, No free lunch theorems for optimization, *IEEE Trans. Evol. Comput.*, 1(1997), No. 1, p. 67.
[22] J.R. Quinlan, Induction of decision trees, *Mach. Learn.*, 1(1986), p. 81.
[23] D.E. Rumelhart, G.E. Hinton, and R.J. Williams, Learning representations by back-propagating errors, *Nature*, 323(1986), No. 6088, p. 533.
[24] C. Cortes and V. Vapnik, Support-vector networks, *Mach. Learn.*, 20(1995), p. 273.
[25] I. Kononenko, Semi-naive Bayesian classifier, [in] *European Working Session on Learning*, Springer, Berlin, Heidelberg, 1991, p. 206.
[26] A. Zunger, Inverse design in search of materials with target functionalities, *Nat. Rev. Chem.*, 2(2018), art. No. 0121.
[27] S. Kunnikuruvan, A. Chakraborty, and D.T. Major, Monte Carlo- and simulated-annealing-based funneled approach for the prediction of cation ordering in mixed transition-metal oxide materials, *J. Phys. Chem. C*, 124(2020), No. 50, p. 27366.
[28] S.B. Roshan, M.B. Jooibari, R. Teimouri, G. Asgharzadeh-Ahmadi, M. Falahati-Naghibi, and H. Sohrabpoor, Optimization of friction stir welding process of AA7075 aluminum alloy to achieve desirable mechanical properties using ANFIS models and simulated annealing algorithm, *Int. J. Adv. Manuf. Technol.*, 69(2013), No. 5-8, p. 1803.
[29] N. Chakraborti, R. Sreevathsan, R. Jayakanth, and B. Bhattacharya, Tailor-made material design: An evolutionary approach using multi-objective genetic algorithms, *Comput. Mater. Sci.*, 45(2009), No. 1, p. 1.
[30] R. Sreevathsan, B. Bhattacharya, and N. Chakraborti, Designing ionic materials through multiobjective genetic algorithms, *Mater. Manuf. Process*es, 24(2009), No. 2, p. 162.
[31] T.D. Liu, L.Y. Xu, G.F. Shao, N.N. Tu, J.P. Tao, and Y.H. Wen, Structural optimization of Pt–Pd–Rh trimetallic nanoparticles using improved genetic algorithm, *J. Alloys Compd.*, 663(2016), p. 466.
[32] C.S. Wang, H.D. Fu, L. Jiang, D.Z. Xue, and J.X. Xie, A property-oriented design strategy for high performance copper alloys via machine learning, *npj Comput. Mater.*, 5(2019), art. No. 87.
[33] Y.J. Su, H.D. Fu, Y. Bai, X. Jiang, and J.X. Xie, Progress in materials genome engineering in China, *Acta Metall. Sin.*, 56(2020), No. 10, p. 1313.
[34] L. Jiang, C.S. Wang, H.D. Fu, J. Shen, Z.H. Zhang, and J.X. Xie, Discovery of aluminum alloys with ultra-strength and high-toughness via a property-oriented design strategy, *J. Mater. Sci. Technol.*, 98(2022), p. 33.
[35] H.X. Zong, G. Pilania, X.D. Ding, G.J. Ackland, and T. Lookman, Developing an interatomic potential for martensitic phase transformations in zirconium by machine learning, *npj Comput. Mater.*, 4(2018), art. No. 48.
[36] K. Tsutsui, H. Terasaki, T. Maemura, K. Hayashi, K. Moriguchi, and S. Morito, Microstructural diagram for steel based on crystallography with machine learning, *Comput. Mater. Sci.*, 159(2019), p. 403.
[37] S. Feng, H.Y. Zhou, and H.B. Dong, Using deep neural network with small dataset to predict material defects, *Mater. Des.*, 162(2019), p. 300.
[38] L. Ward, A. Agrawal, A. Choudhary, and C. Wolverton, A general-purpose machine learning framework for predicting properties of inorganic materials, *npj Comput. Mater.*, 2(2016), art. No. 16028.
[39] L. Ward, S.C. O'Keeffe, J. Stevick, G.R. Jelbert, M. Aykol, and C. Wolverton, A machine learning approach for engineering bulk metallic glass alloys, *Acta Mater.*, 159(2018), p. 102.
[40] N. Islam, W.J. Huang, and H.L. Zhuang, Machine learning for phase selection in multi-principal element alloys, *Comput. Mater. Sci.*, 150(2018), p. 230.
[41] P. Villars and L.D. Calvert, *Pearson's Handbook of Crystallographic Data for Intermetallic Phases*, ASM Int., Ohio, 1991.
[42] R.H. Yuan, D.Q. Xue, D.Z. Xue, J.S. Li, X.D. Ding, J. Sun, and T. Lookman, Knowledge-based descriptor for the compositional dependence of the phase transition in $BaTiO_3$-based ferroelectrics, *ACS Appl. Mater. Interfaces*, 12(2020), No. 40, p. 44970.
[43] S.Z. Li, H.R. Zhang, D.B. Dai, G.T. Ding, X. Wei, and Y.K. Guo, Study on the factors affecting solid solubility in binary alloys: An exploration by Machine Learning, *J. Alloys Compd.*, 782(2019), p. 110.
[44] Z.C. Lu, X. Chen, X.J. Liu, D.Y. Lin, Y. Wu, Y.B. Zhang, H. Wang, S.H. Jiang, H.X. Li, X.Z. Wang, and Z.P. Lu, Inter-

pretable machine-learning strategy for soft-magnetic property and thermal stability in Fe-based metallic glasses, *npj Comput. Mater.*, 6(2020), art. No. 187.

[45] K. Kaufmann and K.S. Vecchio, Searching for high entropy alloys: A machine learning approach, *Acta Mater.*, 198(2020), p. 178.

[46] V. Stanev, C. Oses, A.G. Kusne, E. Rodriguez, J. Paglione, S. Curtarolo, and I. Takeuchi, Machine learning modeling of superconducting critical temperature, *npj Comput. Mater.*, 4(2018), art. No. 29.

[47] H. Wu, A. Lorenson, B. Anderson, L. Witteman, H.T. Wu, B. Meredig, and D. Morgan, Robust FCC solute diffusion predictions from *ab-initio* machine learning methods, *Comput. Mater. Sci.*, 134(2017), p. 160.

[48] S. Kirklin, J.E. Saal, V.I. Hegde, and C. Wolverton, High-throughput computational search for strengthening precipitates in alloys, *Acta Mater.*, 102(2016), p. 125.

[49] Y. Zhang, C. Wen, C.X. Wang, S. Antonov, D.Z. Xue, Y. Bai, and Y.J. Su, Phase prediction in high entropy alloys with a rational selection of materials descriptors and machine learning models, *Acta Mater.*, 185(2020), p. 528.

[50] Z.Q. Zhou, Y.J. Zhou, Q.F. He, Z.Y. Ding, F.C. Li, and Y. Yang, Machine learning guided appraisal and exploration of phase design for high entropy alloys, *npj Comput. Mater.*, 5(2019), art. No. 128.

[51] J.M. Rickman, H.M. Chan, M.P. Harmer, J.A. Smeltzer, C.J. Marvel, A. Roy, and G. Balasubramanian, Materials informatics for the screening of multi-principal elements and high-entropy alloys, *Nat. Commun.*, 10(2019), art. No. 2618.

[52] X.M. Wang, Y.L. Xu, J. Yang, J.Y. Ni, W. Zhang, and W.H. Zhu, ThermoEPred-EL: Robust bandgap predictions of chalcogenides with diamond-like structure via feature cross-based stacked ensemble learning, *Comput. Mater. Sci.*, 169(2019), art. No. 109117.

[53] D.B. Dai, T. Xu, X. Wei, G.T. Ding, Y. Xu, J.C. Zhang, and H.R. Zhang, Using machine learning and feature engineering to characterize limited material datasets of high-entropy alloys, *Comput. Mater. Sci.*, 175(2020), art. No. 109618.

[54] J. Benesty, J.D. Chen, Y.T. Huang, and I. Cohen, *Noise Reduction in Speech Processing*, Springer, Berlin, Heidelberg, 2009.

[55] R. Tibshirani, Regression shrinkage and selection via the lasso, *J. R. Stat. Soc. Ser. B*, 58(1996), No. 1, p. 267.

[56] P.M. Granitto, C. Furlanello, F. Biasioli, and F. Gasperi, Recursive feature elimination with random forest for PTR-MS analysis of agroindustrial products, *Chemom. Intell. Lab. Syst.*, 83(2006), No. 2, p. 83.

[57] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, Gene selection for cancer classification using support vector machines, *Mach. Learn.*, 46(2002), p. 389.

[58] Y. Liu, J.M. Wu, M. Avdeev, and S.Q. Shi, Multi-layer feature selection incorporating weighted score-based expert knowledge toward modeling materials with targeted properties, *Adv. Theor. Simul.*, 3(2020), No. 2, art. No. 1900215.

[59] H.T. Zhang, H.D. Fu, X.Q. He, C.S. Wang, L. Jiang, L.Q. Chen, and J.X. Xie, Dramatically enhanced combination of ultimate tensile strength and electric conductivity of alloys via machine learning screening, *Acta Mater.*, 200(2020), p. 803.

[60] J.W. Yeh, Recent progress in high-entropy alloys, *Eur. J. Control*, 31(2006), No. 6, p. 633.

[61] C. Wen, C.X. Wang, Y. Zhang, S. Antonov, D.Z. Xue, T. Lookman, and Y.J. Su, Modeling solid solution strengthening in high entropy alloys using machine learning, *Acta Mater.*, 212(2021), art. No. 116917.

[62] H.T. Zhang, H.D. Fu, S.C. Zhu, W. Yong, and J.X. Xie, Machine learning assisted composition effective design for precipitation strengthened copper alloys, *Acta Mater.*, 215(2021), art. No. 117118.

[63] D.Z. Xue, P.V. Balachandran, J. Hogden, J. Theiler, D.Q. Xue, and T. Lookman, Accelerated search for materials with targeted properties by adaptive design, *Nat. Commun.*, 7(2016), art. No. 11241.

[64] D.Z. Xue, D.Q. Xue, R.H. Yuan, Y.M. Zhou, P.V. Balachandran, X.D. Ding, J. Sun, and T. Lookman, An informatics approach to transformation temperatures of NiTi-based shape memory alloys, *Acta Mater.*, 125(2017), p. 532.

[65] C. Wen, Y. Zhang, C.X. Wang, D.Z. Xue, Y. Bai, S. Antonov, L.H. Dai, T. Lookman, and Y.J. Su, Machine learning assisted design of high entropy alloys with desired property, *Acta Mater.*, 170(2019), p. 109.

[66] Y.W. Liu, L.Y. Wang, H. Zhang, G.M. Zhu, J. Wang, Y.H. Zhang, and X.Q. Zeng, Accelerated development of high-strength magnesium alloys by machine learning, *Metall. Mater. Trans. A*, 52(2021), No. 3, p. 943.

[67] H.T. Zhang, H.D. Fu, Y.H. Shen, and J.X. Xie, Rapid design of secondary deformation-aging parameters for ultra-low Co content Cu–Ni–Co–Si–X alloy via Bayesian optimization machine learning, *Int. J. Miner. Metall. Mater.*, 2022. DOI: 10.1007/s12613-022-2479-3