

Review

Open Access



Machine learning-guided design and development of metallic structural materials

Jinxin Yu^{1,2}, Shengkun Xi¹, Shaobin Pan², Yongjie Wang², Qinghua Peng², Rongpei Shi¹, Cuiping Wang², Xingjun Liu^{1,2,3}

¹School of Materials Science and Engineering, and Institute of Materials Genome and Big Data, Harbin Institute of Technology, Shenzhen 518055, Guangdong, China.

²College of Materials and Fujian Provincial Key Laboratory of Materials Genome, Xiamen University, Xiamen 361005, Fujian, China.

³State Key Laboratory of Advanced Welding and Joining, Harbin Institute of Technology, Shenzhen 518055, Guangdong, China.

Correspondence to: Prof. Xingjun Liu, State Key Laboratory of Advanced Welding and Joining, Harbin Institute of Technology, Taoyuan street, Shenzhen 518055, Guangdong, China. E-mail: xjliu@hit.edu.cn; Prof. Cuiping Wang, College of Materials and Fujian Provincial Key Laboratory of Materials Genome, Xiamen University, Siming South Road 422, Xiamen 361005, Fujian, China. E-mail: wangcp@xmu.edu.cn; Rongpei Shi, School of Materials Science and Engineering, and Institute of Materials Genome and Big Data, Harbin Institute of Technology, Taoyuan street, Shenzhen 518055, Guangdong, China. E-mail: shirongpei@hit.edu.cn.

How to cite this article: Yu J, Xi S, Pan S, Wang Y, Peng Q, Shi R, Wang C, Liu X. Machine learning-guided design and development of metallic structural materials. *J Mater Inf* 2021;1:9. <https://dx.doi.org/10.20517/jmi.2021.08>

Received: 10 Sep 2021 **First Decision:** 12 Oct 2021 **Revised:** 25 Nov 2021 **Accepted:** 7 Dec 2021 **Published:** 24 Dec 2021

Academic Editors: Yan-Jing Su, Xiaoyan Song **Copy Editor:** Yue-Yue Zhang **Production Editor:** Yue-Yue Zhang

Abstract

In recent years, the advent of machine learning (ML) in materials science has provided a new tool for accelerating the design and discovery of new materials with a superior combination of mechanical properties for structural applications. In this review, we provide a brief overview of the current status of the ML-aided design and development of metallic alloys for structural applications, including high-performance copper alloys, nickel- and cobalt-based superalloys, titanium alloys for biomedical applications and high strength steel. We also present our perspectives regarding the further acceleration of data-driven discovery, development, design and deployment of metallic structural materials and the adoption of ML-based techniques in this endeavor.

Keywords: Structural materials, metallic alloys, materials informatics, machine learning, composition-processing-microstructure-property relationships



© The Author(s) 2021. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, 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.



INTRODUCTION

Materials have propelled the development of society over the course of history^[1]. Nowadays, it has been recognized that the rate of materials advances to a large extent underpins the rate at which our modern society advances. The development of structural materials with outstanding mechanical properties has long been sought for a wide spectrum of industrial, technological and biomedical applications. Nevertheless, the search for new or alternative materials with appropriate processing routes, whether through experiments or simulations, is often a slow and arduous task. These endeavors are punctuated by infrequent and, often, unexpected discoveries, each of which then prompts a flurry of intensive studies to better understand the underlying science that govern the behavior of these materials. Historically, the identification of correlations among the composition-processing-structure-property (CPSP) relationships in such a fashion has often relied on an iterative, labor-intensive process of trial and error^[2]. Such processes are reliable but, of course, expensive and slow (sometimes may take up to even decades), and consequently, the ability of researchers to develop new materials on a reasonable timeline has been severely limited.

Microstructure is one of the key challenges that limits the design and development of new materials^[2]. A large majority of the structural materials employed in advanced technologies embody hierarchical internal structures with rich details spanning across multiple length and/or structure scales (from the atomic to macroscale). Collectively, these features of the internal structure of a material are frequently referred to as the microstructure or simply the structure. Structural properties, such as hardness, strength, ductility, fracture, creep, fatigue and corrosion resistance, are extremely sensitive to both composition and microstructure, therefore adding a large number of orthogonal dimensions of complexity and significant difficulty to the evaluation and quantification of CPSP relationships, which is however required by the rational design of structural materials.

Considering the complexity of modern aviation and turbines^[3], in addition to the need for superior combinations of multiple properties, it is clear why these applications require comparably complex structural materials. The continuous increase in both compositional and microstructural space has made the development and application of effective methods for decoding CPSP relationships imperative. The last decade has seen a major movement in materials science with regards to the emerging but rapid revolution in combinatorial and high-throughput experimental characterization^[4,5] and calculations (e.g., integrated computational materials engineering^[6] and multi-scale simulations^[7]) of materials in general and structural materials specifically. In particular, artificial intelligence and machine learning (ML) add new dimensions for quantifying CPSP relationships. Importantly, the Materials Genome Initiative in the United States, Materials Genome Engineering in China^[8] and numerous analogous movements elsewhere around the world have allowed for the fusion of ongoing efforts in experimental tools, computations and the aggressive use of digital data and realizing synergistic interactions among them.

Indeed, the dramatically increasing data generation rates from both experiments and simulations necessitate the use of informatics on large datasets for retrieving knowledge. The recent decade has seen an explosion in ML utilized as a powerful tool to quantify CPSP linkages for materials spanning an array of classes and applications. The data-driven approach, to extract and curate materials knowledge from available datasets, is also frequently referred to as materials informatics^[2,9-14].

ML utilizes computer systems that do not require explicit programming for learning about the task(s) they are fulfilling or completing. The critical steps in a ML workflow broadly include [Figure 1]: (1) data collection and cleaning; (2) feature generation and selection (also referred to as featurization or feature engineering); (3) algorithm selection, training and hyperparameter optimization; (4) model uncertainty

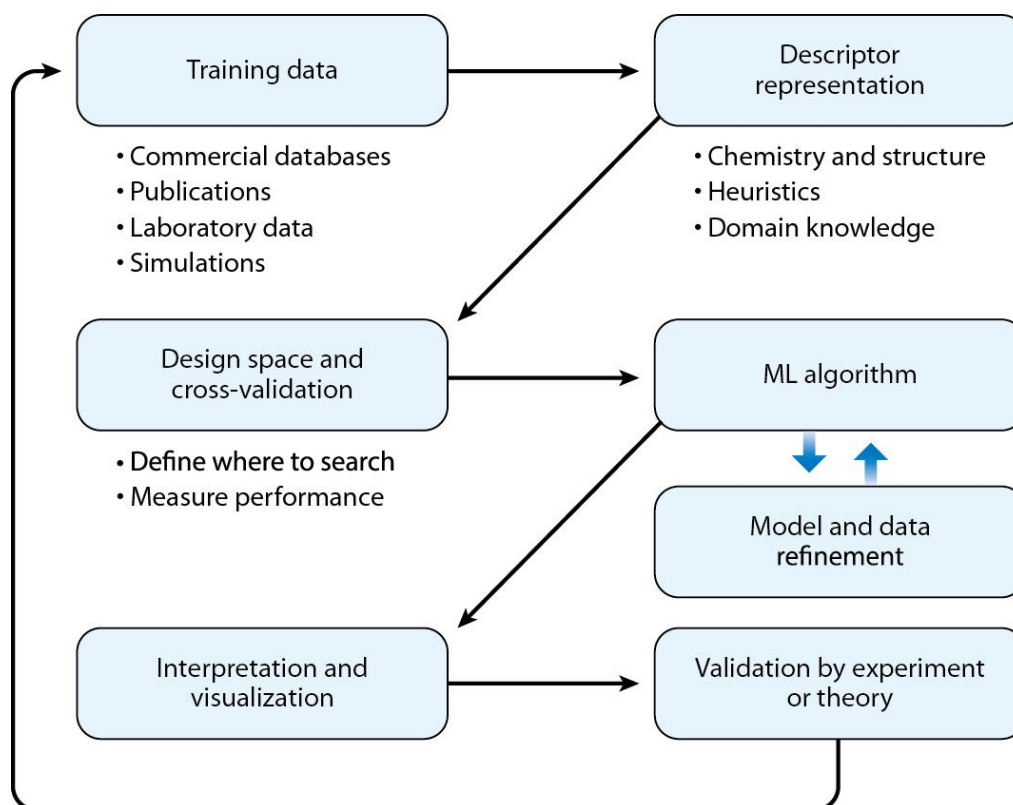


Figure 1. Schematic of a materials informatics pipeline common to validated ML studies^[15]. ML: Machine learning.

assessment (e.g., performance on test dataset) and domain applicability. The workflow begins with identifying, collecting and organizing a materials dataset for training, as well as a set of materials descriptors to extend the data with the available physical information. Composition- or structure-based feature vectors have been frequently selected as a descriptor or descriptors by materials informatics approaches. After being established and featurized, the dataset is then used to train a ML model and then to make a prediction of novel materials or properties for validation. A rich variety of algorithms exist that can be selected for training and making predictions, such as linear regressions, support vector machines, random forests (RFs), K-nearest neighbors, neural networks and so on. A simple method is used to classify learning algorithms used for ML^[8]. If one trains an algorithm with data that also contain the answer(s), then it is known as supervised learning. If one trains an algorithm with data and wants the machine to determine the patterns, then it is unsupervised learning. If one gives any algorithm a goal and expects the machine, through trial and error, to achieve the goal, then it is reinforcement learning.

To date, two active and highly related research areas in the application of ML in materials science and engineering are property predictions and materials discovery and design^[10-13,15-28]. As such, ML algorithms involved in property prediction naturally fall in the category of supervised learning, in view of the nature of the dataset. By providing a new understanding or key physical relationships governing a property or a combination of properties of interest, ML has built on its strength and gained an increasing momentum to enable the discovery, design and development of novel materials spanning an array of applications and materials classes.

Despite the fact that many aspects have been seen in ML and it is already changing how materials discovery and design are conducted, their interaction are generally believed to still be in the nascent stages^[10,13], with the full power of their synergy still far from being fully realized. Thus, the purpose of this review is to introduce the emerging data-driven capabilities for the optimization, design and discovery of structural materials. It should be noted that materials for structural applications or structural materials comprise a broad class that includes metallic alloys, ceramics, glasses, concrete and cements. We restrict our review to the ML-assisted design and discovery of metallic structural materials that have lagged behind other classes of materials in the use of data science for rapid optimization and development. Highlighting recent advancements in this field demonstrates the power of data-driven methodologies that will hopefully lead to the production of market-ready structural materials in an accelerated fashion with reduced costs. We also focus on the supervised learning models for property prediction, which is one of the most frequent uses of ML in materials science and engineering in general and metallic structural materials in specific.

Through this review, we aim to provide a brief discussion of major ML applications in the areas of metallic structural materials to help guide researchers that attempt to understand the landscape and perhaps take their first steps into the area. The remainder of the review is organized as follows. In CURRENT STATUS OF MATERIALS DESIGN ASSISTED BY MACHINE LEARNING, we present several successful examples of validated ML-driven materials discovery across a wide variety of known metallic alloys, including high-performance copper (Cu) alloys, nickel (Ni)- and cobalt (Co)-based superalloys and titanium (Ti) alloys for biomedical applications and high strength steel. The selection of these five classes of materials lies in a simple fact that these materials have numerous industrial, technological and even biomedical applications and their applications will continue to increase for many years to come. Of course, this is a representative but not exhaustive list of recent studies on the application of ML in metallic structural materials. For example, high-entropy alloys and bulk metallic glasses are excluded from the current review since they have been reviewed elsewhere^[11,13]. For each class of metallic alloy, we first identify the scientific questions and challenging issues, and then briefly summarize the recent progress in the application of data science to tackle these challenges.

Through these case studies, we identify several common themes, such as the use of domain knowledge to inform ML models and offer some more speculative thoughts on longer-term future opportunities and challenges, such as the issues regarding the interpretabilities of ML-based models, interactions between ML and experiments, characterizations and between ML and multiscale simulations, as well as integrating ML training and education in materials science and engineering programs.

CURRENT STATUS OF MATERIALS DESIGN ASSISTED BY MACHINE LEARNING

Property-oriented design for high-performance Cu alloys

Cu alloys are widely used in electronic components, connector components, resistance welding electrodes and lead frames due to their high hardness and electrical conductivity (EC)^[29-31]. Cu-Ni-Si-based alloys are particularly promising candidate materials that could meet the requirements of next-generation integrated circuits. As such, extensive research studies have been carried out to optimize existing Cu-Ni-Si-based alloys and to design new ones with an excellent combination of high strength and EC^[32,33]. However, these advances have been hindered due to the low efficiency of traditional experimental trial-and-error processes that are both labor intensive and time-consuming. Advances in ML, which can be applied to the identification of materials with target functionalities^[18,21,27], open up new avenues for accelerating the optimization and discovery of Cu alloys. However, with the exception of studies by Wang *et al.*^[27] and Zhang *et al.*^[21], as discussed below, only a few such efforts have been reported so far.

As shown in [Figure 2](#), the ML design system proposed by Wang *et al.*^[27] consists of three subsystems, namely, the model training, the compositional design and the property prediction. The system builds upon two kinds of back propagation neural network (NN) models that can learn the relationship between properties (e.g., mechanical strength and EC) and compositions of Cu alloys. More specifically, the C2P (i.e., composition \rightarrow property) model predicts the properties of alloys from their compositions, while the P2C (i.e., property \rightarrow composition) model predicts the compositions of alloys based on the targeted property requirements. Both NN models are first trained using the existing database of composition/property relationships for Cu alloys, followed by the validation of model predictions for both accuracy and reliability. As shown in [Figure 2](#), for a targeted property requirement, the authors relied on the iterative interaction between these two models to conduct a fast and robust alloy composition screening. The basic idea is to employ the C2P model to evaluate the alloy composition design schemes provided by the P2C model and to establish the confidential interval (i.e., the relative error between the target performance and the predicted one) for the composition design. The employment of this ML design system allowed for the rapid and successful design of several high-performance Cu alloys with an ultimate tensile strength of 600-950 MPa and an EC of 50.0% [i.e., International Annealed Copper Standard (IACS)]. It is expected that this new approach for realizing property-oriented composition design via ML is applicable to a wide range of metallic materials beyond Cu alloys.

In a more recent study from the same research group, in addition to alloy compositions, Zhang *et al.*^[21] further accounted for the contributions from the physical and chemical features of the elements to the properties of interest in their ML strategy. As shown in [Figure 3](#), the alloy design efforts for targeted property requirements are carried out by screening key alloy factors, through correlation screening, recursive elimination and exhaustive screening, followed by iterative composition design through Bayesian optimization. Using precipitation-strengthened Cu alloys as an example, the ML framework first obtained five and six kinds of alloy factors that most significantly affect hardness and EC, respectively. Based on the learned influence of the key element features on properties, the subsequent Bayesian optimization together with iterative optimization experiments resulted in the design of a new alloy (Cu-1.3Ni-1.4Co-0.56Si-0.03Mg, wt.%), with an excellent combination of desired mechanical (e.g., tensile strength of 858 MPa) and electrical properties (e.g., EC of 47.6% IACS). It is noteworthy that improving these two conflicting properties of Cu alloys has always been challenging using traditional experimental trial-and-error efforts alone but has been made possible with the aid of ML.

Even though the ML strategies discussed above have achieved tremendous success in property-oriented alloy design based on existing datasets for composition-property relationships, it should be noted that among these efforts, the processing routes have been intentionally excluded for making ML model training easier. As a matter of fact, the screening of data was made by limiting them to a well-defined processing route when constructing the training dataset for composition-property relationships. Consequently, such ML efforts do not permit a further improvement in materials properties through the optimization and design of processing routes through ML predictions.

To overcome this drawback, a recent study by Pan *et al.*^[18] established a composition-processing-property database in their ML model for assisting alloy design. The inclusion of processing features enables the trained ML framework to predict the property-oriented alloy composition and the corresponding processing routes. The workflow of the ML-assisted design process is schematically shown in [Figure 4](#). Using a Cu-Ni-Co-Si-based alloy as an example, the framework can be divided into five critical steps:

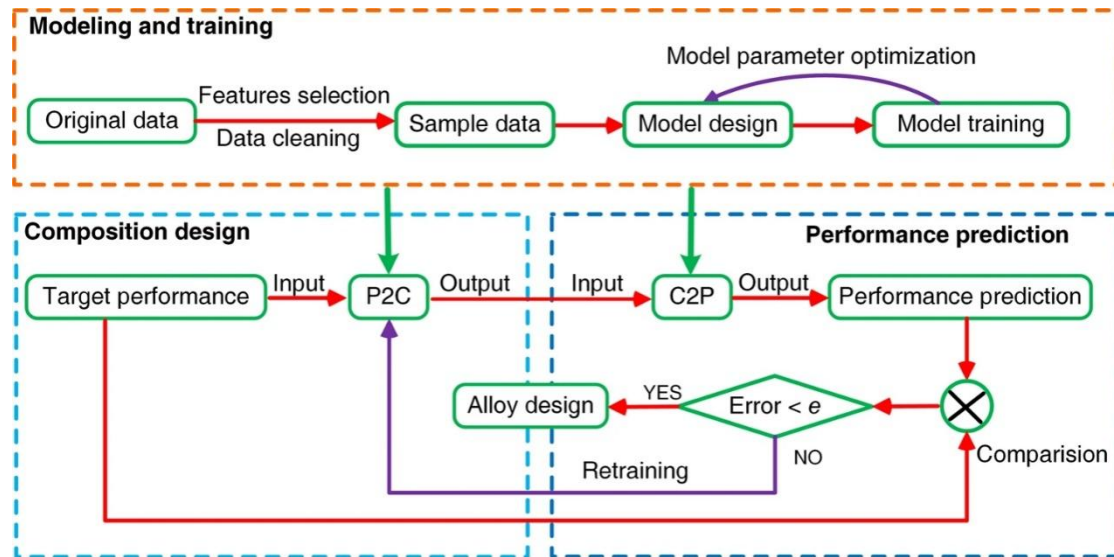


Figure 2. Schematic of a machine learning (ML)-assisted design framework for the rapid and accurate compositional design of Cu alloys^[27]. P2C: Performance to composition; C2P: composition to performance.

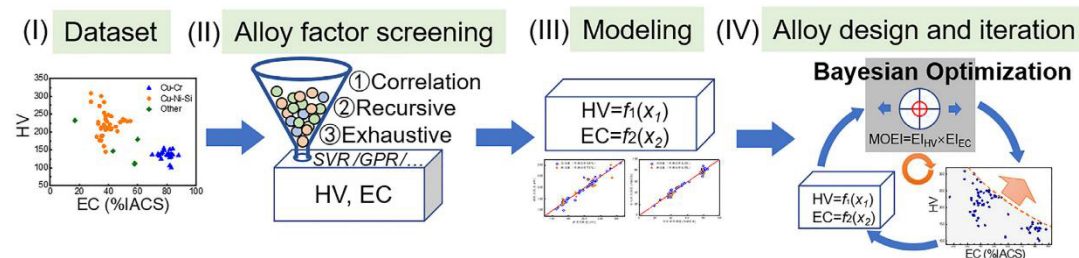


Figure 3. Schematic of an alloy composition design strategy that combines alloy factor screening and Bayesian optimization^[21]. IACS: International Annealed Copper Standard; EC: electrical conductivity; HV: hardness according to Vickers.

- (1) Establish a composition-processing-property database;
- (2) Establish a prediction model;
- (3) Construct a multi-dimensional search space;
- (4) Optimization;
- (5) Experimental testing and validation.

In step (1), six features (three alloying elements and three process parameters, such as the amount of cold deformation, aging temperature and time) and two attributes (microhardness and EC) are considered when constructing the database. In step (2), three different learning algorithms are trained using the collected dataset and the one (i.e., a RF in this study) with the highest accuracy is then selected for making predictions.

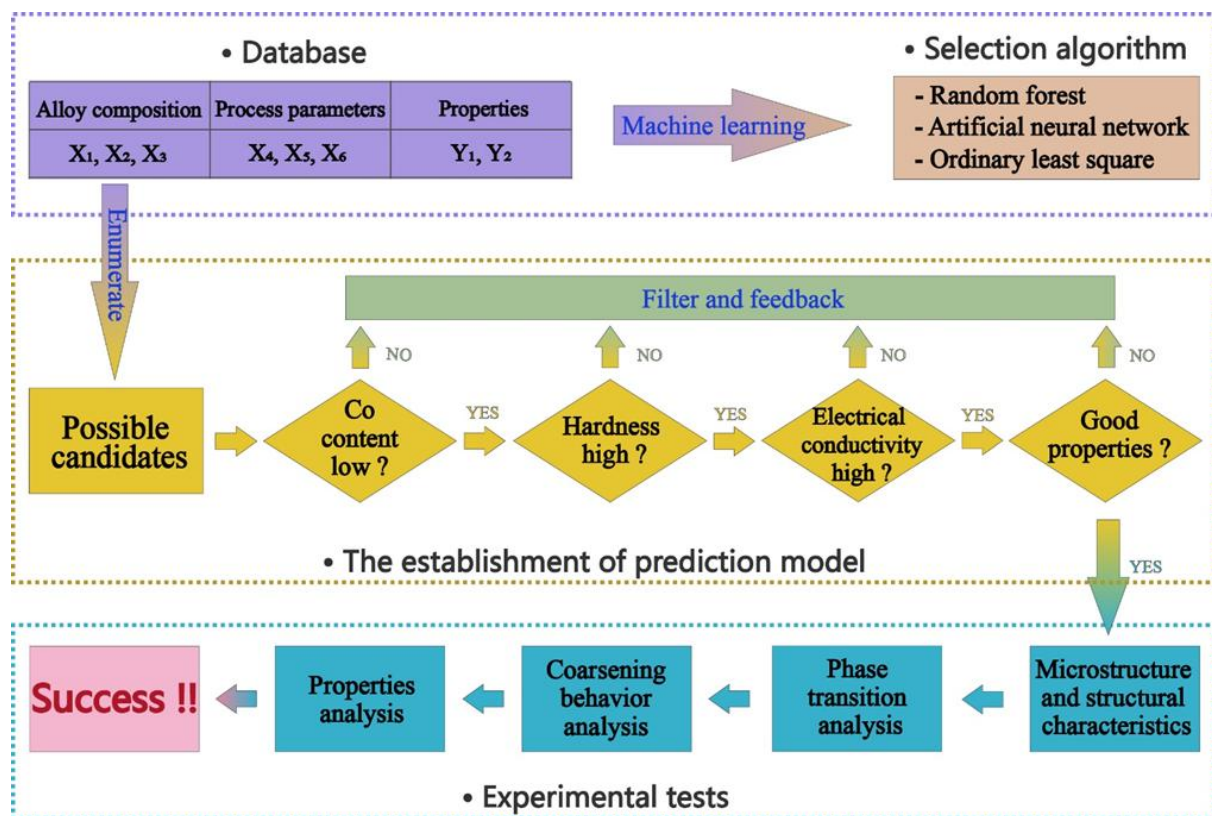


Figure 4. Schematic workflow of machine learning (ML)-assisted design process for Cu alloys with consideration of processing variables^[18].

In step (3), using the trained algorithm, predictions are then made in a multi-dimensional parameter space constructed using the six features. The boundary of the search space is sought based on the database and domain knowledge. The property-oriented search is then made with a constraint for the Co composition (0–1 wt.%) over 8640 possible candidate combinations of composition and processing, which leads to the identification of four groups of composition-processing combinations, all having the combination of high microhardness (≥ 248 HV) and high EC ($> 37\%$ IACS). Beyond predictions, three designed alloys are further prepared based on the predicted processing routes for experimental validation. Among them, the designed alloy (Cu-2.3Ni-0.7Co-0.7Si wt.%) prepared by the predicted processing route was identified with the best performance through experimental measurements. Moreover, the underlying mechanism for the best performance was also clarified through a comprehensive analysis of both the microstructure and phase transition behavior during processing.

The results in this study demonstrated that ML is capable of effectively and efficiently guiding and thus accelerating the development of Cu-based alloys with targeted properties. The determined linkage among CPSP relationships guided by ML predictions, in turn, enriches our knowledge of Cu alloys. There is no doubt that the strategy possesses a higher fidelity than others relying on composition-performance relationship alone. It is noteworthy that, however, the consideration of one additional degree of freedom in the database may increase the fidelity of prediction at the expense of high computational cost.

Alloy optimization and design of Ni-based superalloys

Ni-based superalloys have been widely used in demanding high-temperature structural applications (such as the hot zone of aircraft gas turbine engines) due to their unique combination of highly desirable properties, such as high-temperature strength and excellent resistance to thermal/mechanical fatigue, oxidation and creep^[34-37]. These excellent properties are mainly attributed to the chemistry and microstructure of the alloys, with the latter predominantly consisting of ordered, Ni₃Al-based, L1₂-structured, γ' -(Ni₃Al) precipitates that are coherently embedded in a disordered solid solution of a γ matrix^[38,39]. Besides processing and heat treatment, alloying also plays an important role in improving the mechanical properties of Ni-based superalloys, in view of their high sensitivity to alloy composition variation.

The overarching goal of property optimization and improvement through alloy design, however, remains challenging through experimental trial and error alone. The relatively large number of alloying elements (more than eight in general) makes it increasingly difficult to evaluate the relationships between composition, phases and properties^[37]. Thus, computational approaches have been widely employed to assist alloy composition optimization and design. Although these computational approaches (such as CALPHAD and first principles calculations) have achieved tremendous success in alloy design and are still very important, a new tool, namely, ML, is now emerging and fueling the next generation of superalloy discovery. This is primarily the result of the advent of materials databases^[36], including (1) rich, highly populated experimental databases through high-throughput characterization; (2) computational thermodynamic and kinetic databases for multicomponent systems; and (3) dynamically expanding databases populated with materials properties through automated first-principles calculations.

One such effort through alloy design is the aim of suppressing the formation of topologically close-packed (TCP) and geometrically close-packed (GCP) phases^[35], which are always regarded as being detrimental to mechanical properties. As a prime example, Qin *et al.*^[25] first established a database describing the relationship between composition and the formation propensity of these two detrimental phases in multicomponent Ni-based superalloys through combinatorial and high-throughput experiments, which employed advanced detection equipment for the automatic quantitative collection of compositional and microstructural information^[5]. After a pre-processing of the dataset for redundant feature removal, six different ML algorithms were trained using the selected dataset, through which a robust one was finally screened with area under curve values of 0.88 and 0.92 for the TCP and GCP phases, respectively. [Figure 5](#) shows the schematic workflow of the proposed method. A model with sound predictability may find immediate application for the discovery of Ni-based superalloys within undiscovered composition ranges that may have been missed by conventional approaches. In particular, the combination of high-throughput experiments for rapid materials database construction and ML models for data mining could be extensively used for other multicomponent alloys beyond Ni-based superalloys.

The γ' solvus temperature is a critical parameter when evaluating the stability of Ni-based superalloys under high-temperature working environments and designing alloy heat treatment schedules^[40]. ML has also been applied for predicting the γ' solvus temperature of Ni-based superalloys^[41-44]. In such a prediction, a database is first established to describe the relationship between alloy composition and γ' solvus temperature by collecting data in the published literature and from CALPHAD calculations.

Based on the collected dataset, several ML algorithms (ordinary least squares, support vector machine regression, decision tree regression and RF algorithm) are trained for predicting the γ' solvus temperature. The prediction results from each algorithm are shown in [Figure 6](#), which demonstrates that the RF algorithm exhibits the best performance and is thus selected for predicting the γ' solvus temperature. In

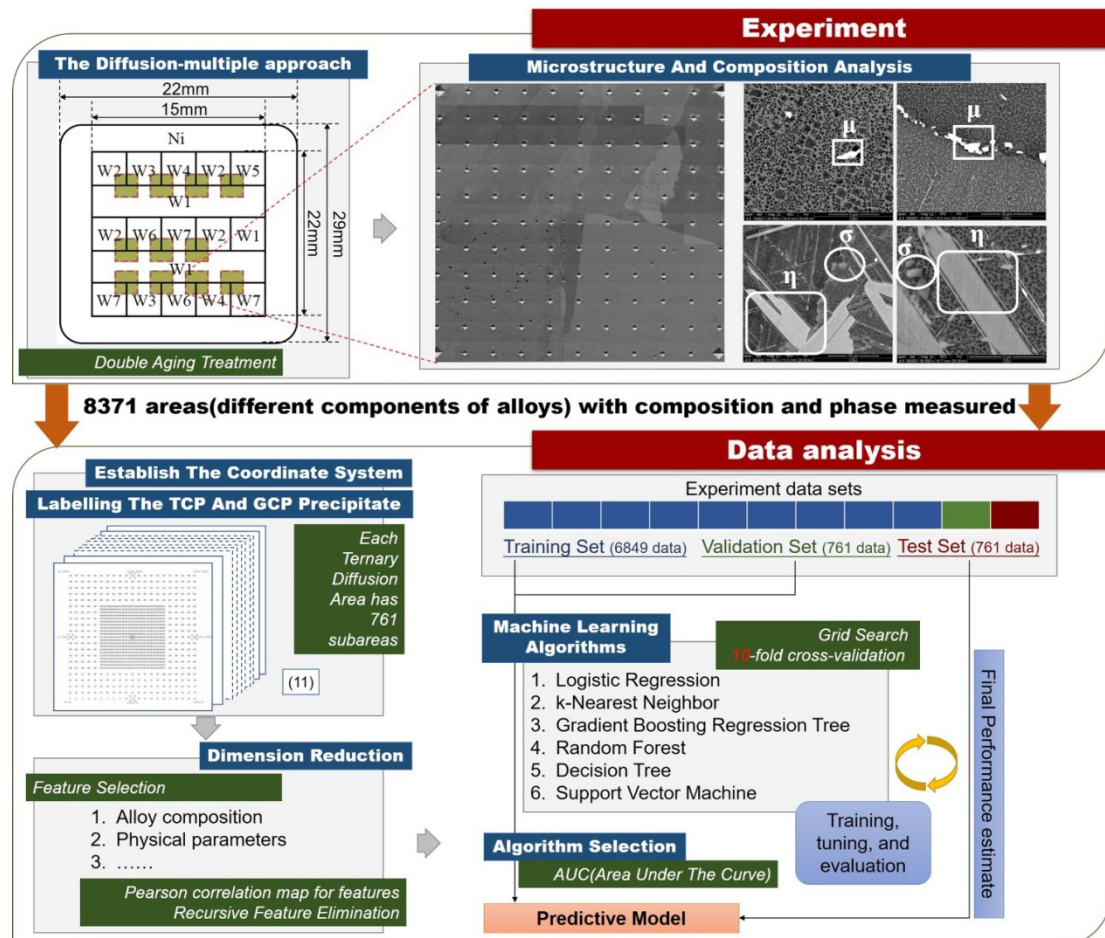


Figure 5. Schematic workflow of proposed method combining high-throughput experiments for database construction and machine learning models for data mining^[25].

order to evaluate the feasibility of this ML model, the predicted γ' solvus temperatures for seven Ni-based superalloys published recently by Chen *et al.*^[42], Cormier^[43], Rame *et al.*^[44], Kim *et al.*^[45] were compared with the experimental data and CALPHAD predictions (see Figure 7). It could be found that the RFR model outperforms the traditional CALPHAD calculations in terms of prediction accuracy.

Other examples include, but are not limited to, Tamura *et al.*^[24], who optimized the processing parameters of Ni-Co powder alloys based on Bayesian classification and Kim *et al.*^[45], who analyzed the oxidation resistance of Ni-based superalloys based on an artificial NN, through which they accelerated the establishment of the relationships between alloy composition and oxidation resistance.

Optimization of novel Co-based superalloys

Like Ni-based superalloys, Co-based superalloys also derive their strength from a coherent $L1_2$ ordered γ' precipitate in a disordered face-centered cubic γ matrix but represent a new opportunity for high-temperature alloy development, in view of their higher solidus temperature (1494 °C) compared to Ni-based superalloys (1455 °C)^[46,47]. The discovery of the γ/γ' microstructure with the addition of W to a Co-Al alloy has also opened up new possibilities for high-temperature alloy development^[47,48]. In the last five years, the emergence of Co-based superalloys with low density and high specific yield strength at elevated temperatures has further fueled the research and development efforts in the field^[48]. This shift could be

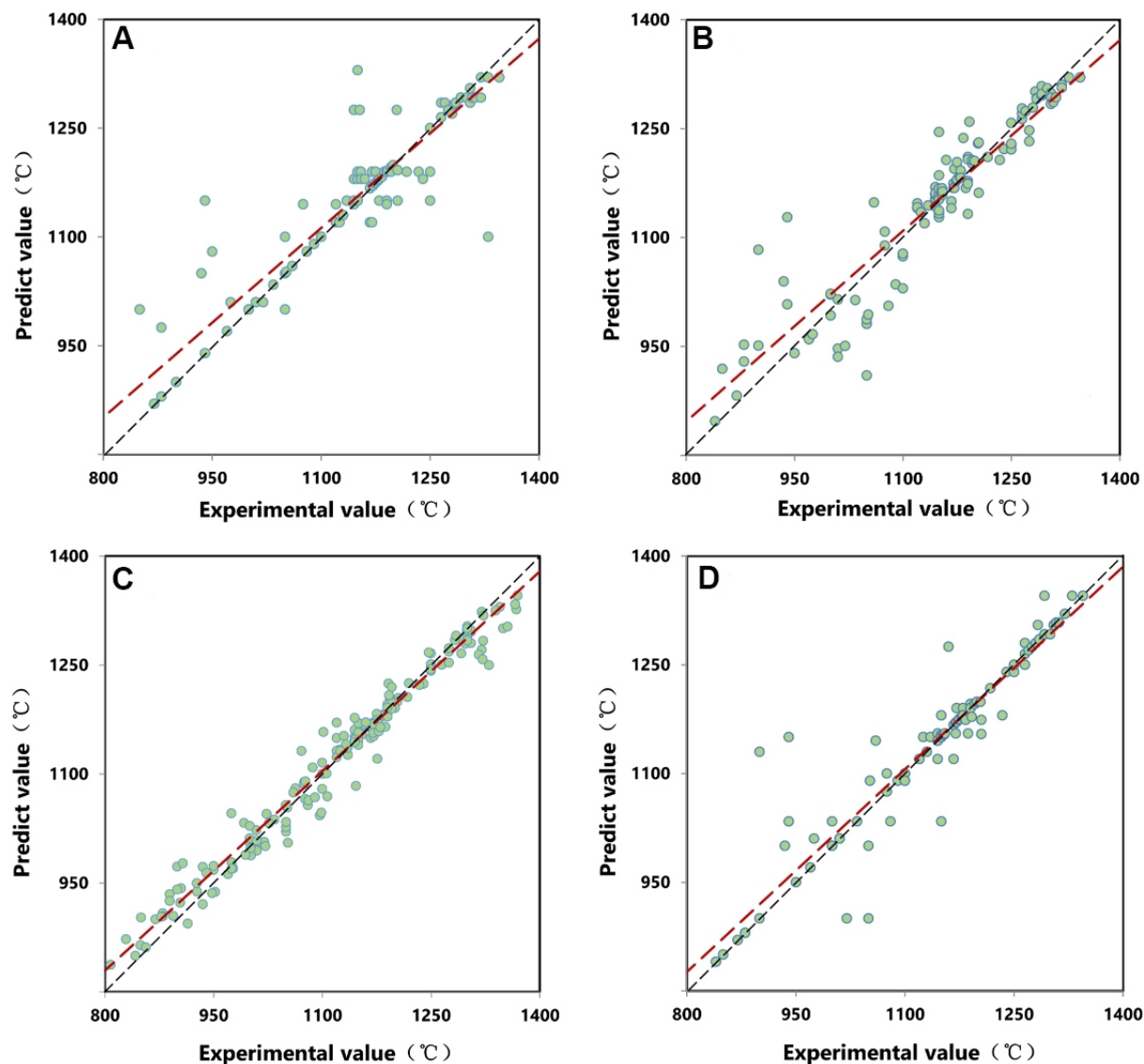


Figure 6. Predicting γ' solvus temperature based on different machine learning algorithms: (A) ordinary least squares; (B) decision tree regression; (C) RF regression; (D) support vector machine regression.

ascribed to the fact that the stabilization of the γ/γ' microstructure can be achieved in the Co-Al-Mo system, without W addition, by a small amount of Nb or Ta. These alloys have thus been referred to as W-free Co-based superalloys. These new generations of alloys with multiple alloying elements to form basic ternary, quaternary and even higher component alloys have steadily expanded the property envelopes, raising the likelihood of finding a modern class of superalloys with higher temperature capabilities. The wide spectrum of alloy element candidates (e.g., Ni, Ti, Ta, Mo and Cr), however, has made the search for new Co-based alloys, with a superior combination of high-temperature mechanical properties, γ' phase stability, oxidation resistance and lightweight (mass densities of 7.8-8.6 g/cm³), labor intensive and time-consuming.

Fortunately, the advent of ML in materials sciences has opened up a new avenue for analyzing high-dimensional space and identifying correlations among composition-processing-microstructure-property relationships that may have been previously missed by conventional trial-and-error approaches^[22]. For

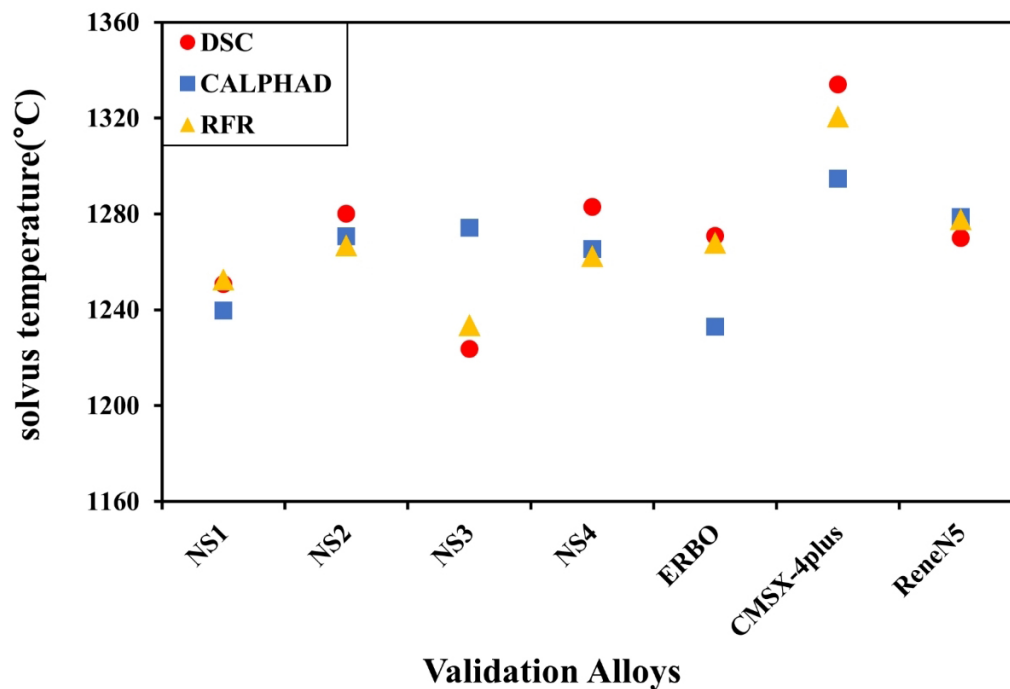


Figure 7. Comparison between experimental measurements of γ' solvus temperature^[42-45], predictions based on a ML RF algorithm and the CALPHAD approach. DSC: Differential scanning calorimetry; CALPHAD: CALculation of PHase diagram; RFR: RF regression.

example, an accelerated design strategy based on ML was formulated for Co-based superalloys, through which new superalloys with high γ' solvus temperature, area fraction and Cr content were successfully designed^[16,28].

The workflow of such a design process is schematically illustrated in [Figure 8](#). First, it begins with collecting relevant datasets from published journal articles, reports and our in-house database related to Co-based superalloys to establish the relevant database for the ML study. All the data were collected from experimental measurements. Second, the prediction models of four important properties of the superalloys are trained and established separately based on the established database. In particular, the presence of the γ' (Y_1) and other phases (Y_2) are predicted by classification methods, while the solvus temperature (Y_3) and area fraction (Y_4) of γ' are predicted using regression methods. Third, four single-property prediction models are combined for screening candidate alloys based on the comprehensive design requirements of different properties. Finally, the prediction results are experimentally validated. If the experimental results exhibit relatively significant differences from the predictions, the model parameters would then be adjusted using the new experimental datasets for improving the accuracy of prediction. Consequently, superalloys with the desired combination of required properties could be successfully designed and experimentally validated on a reasonable timeline, despite the enormous choice of alloy elements and their combinations.

We now briefly describe the application of such a ML-based strategy for designing new L_{1_2} -strengthened Co-based superalloys that fulfill all of the following four specific requirements: (1) γ' precipitates exist; (2) the existence of other phases is limited; (3) the γ' solvus temperature is as high as possible; and (4) the γ' area fraction is as high as possible provided that requirement (3) is met.

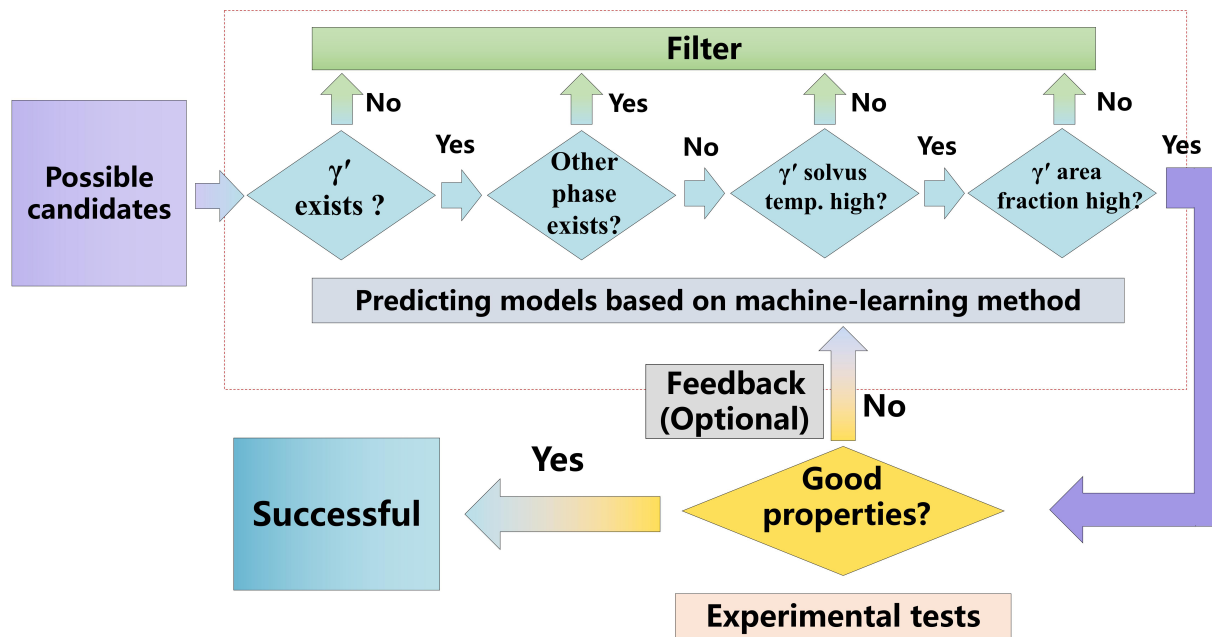


Figure 8. Framework of design process of L₁₂-strengthened Co-based superalloys^[16,28].

Co-Al-W-based, Co-V-Ta-based and other alloys have been successfully optimized. Figure 9 schematically illustrates the optimization process of Co-Al-W-based alloys. Six candidate alloy compositions were selected from 363,000 possible combinations. The subsequent experimental and prediction results agreed well with each other^[16].

The most recent application of a ML-based framework has led to the successful design of a series of novel W-free high-strength Co-V-Ta-based alloys. The designed alloys all exhibit a unique combination of low mass densities (8.67-8.86 g/cm³) and high γ' solvus temperatures (up to 1044 °C) and strength, which is more intriguing compared with that of traditional and novel γ' -reinforced Co-based superalloys (as shown in Figure 10)^[17].

In a similar study by Liu *et al.*^[22], the microstructural stability, γ' solvus temperature, γ' volume fraction, density, processing window, freezing range and oxidation resistance were also optimized by a ML method with active learning (as shown in Figure 11). The most promising candidate alloy, Co-36Ni-12Al-2Ti-4Ta-1W-2Cr (at.%), possesses a high γ' solvus temperature (1266 °C) without the precipitation of any deleterious phases, a volume fraction γ' of up to 74.5% after aging for 1000 h at 1000 °C, a density of 8.68 g/cm³ and excellent high-temperature oxidation resistance at 1000 °C due to the formation of a protective alumina layer.

Development of Ti alloys for biomedical applications

Owing to their unique combination of biocompatibility, excellent corrosion resistance, low Young's modulus and high strength, titanium (Ti) and its alloys have been finding increasingly widespread biomedical applications, such as in bone fixation devices^[49,50]. Ti-6Al-4V (wt.%) alloys have long been used as biomedical implant and prosthesis materials. However, their Young's modulus ($E \sim 110$ GPa) is much higher than that of human bone ($E < 35$ GPa). The mismatch in Young's modulus results in stress shielding, which causes abnormal degradation of the bone near the implantations, such as bone loss, loosening and even failure of the implant^[51]. Moreover, the major alloying elements, aluminum (Al) and vanadium (V),

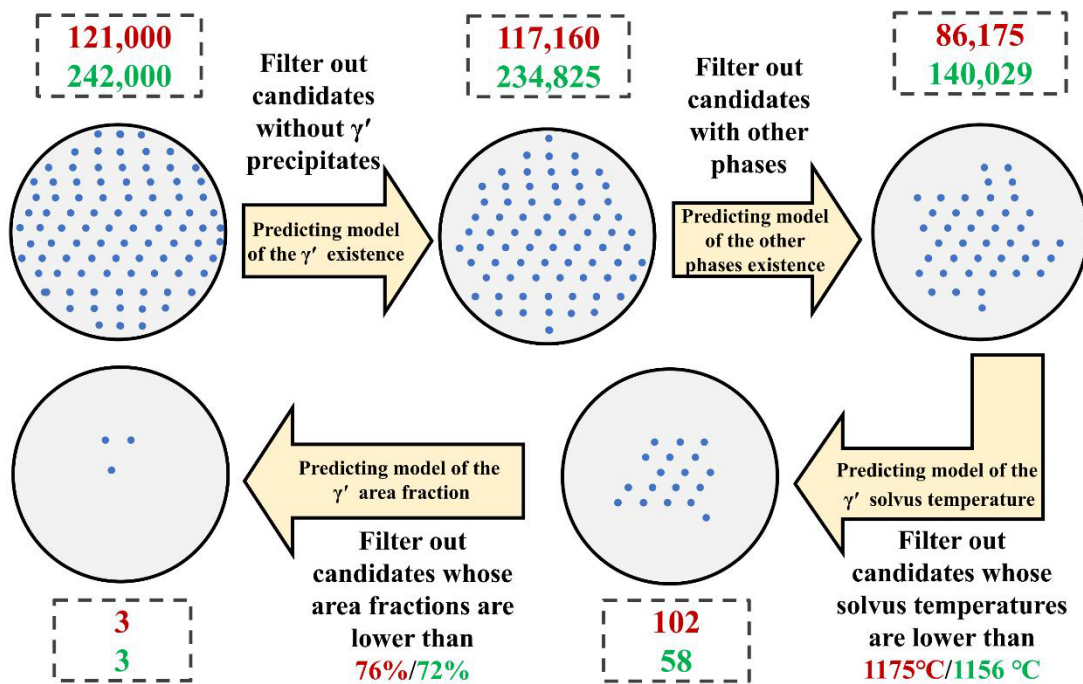


Figure 9. Schematic of optimization process of two search spaces^[16].

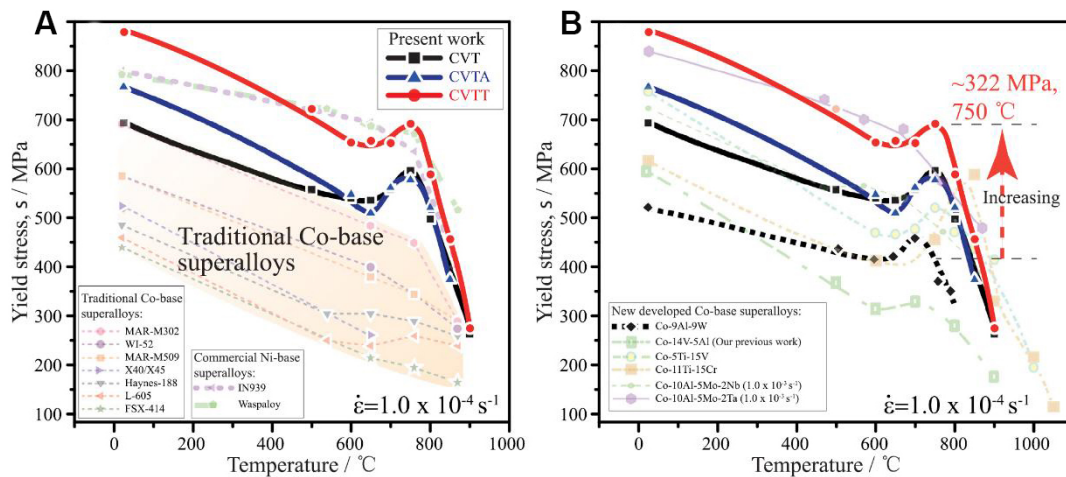


Figure 10. Temperature dependence of yield strengths of designed Co-V-Ta-based alloys (CVT, CVTT and CVTA)^[17]. For comparison, the yield strengths of the traditional Co- and commercial Ni-based superalloys and newly developed Co-based alloys are plotted in (A) and (B). The alloys were homogenized at 1100 °C for 24 h before annealing.

have been found to be toxic or irritating^[52,53]. Therefore, these drawbacks of Ti-6Al-4V warrant the development of new alloys containing vanadium replacements, such as niobium (Nb), iron (Fe) and molybdenum (Mo), and aluminum replacements, such as tantalum (Ta), hafnium (Hf) and zirconium (Zr).

Compared with commercially pure Ti, α or near α Ti, $\alpha + \beta$ Ti (e.g., Ti-6Al-4V) and β -Ti alloys have attracted increasing attention due to their prominent and highly desirable properties, such as high strength, low Young's modulus and corrosion resistance. New alloys, such as Ti-15Mo-5Zr-3Al, Ti-12Mo-6Zr-2Fe, Ti-30Nb, Ti-30Ta, Ti-13Nb-13Zr, Ti-35Nb-5Ta-7Zr and Ti-24Nb-4Zr-7.9Sn, have been developed and

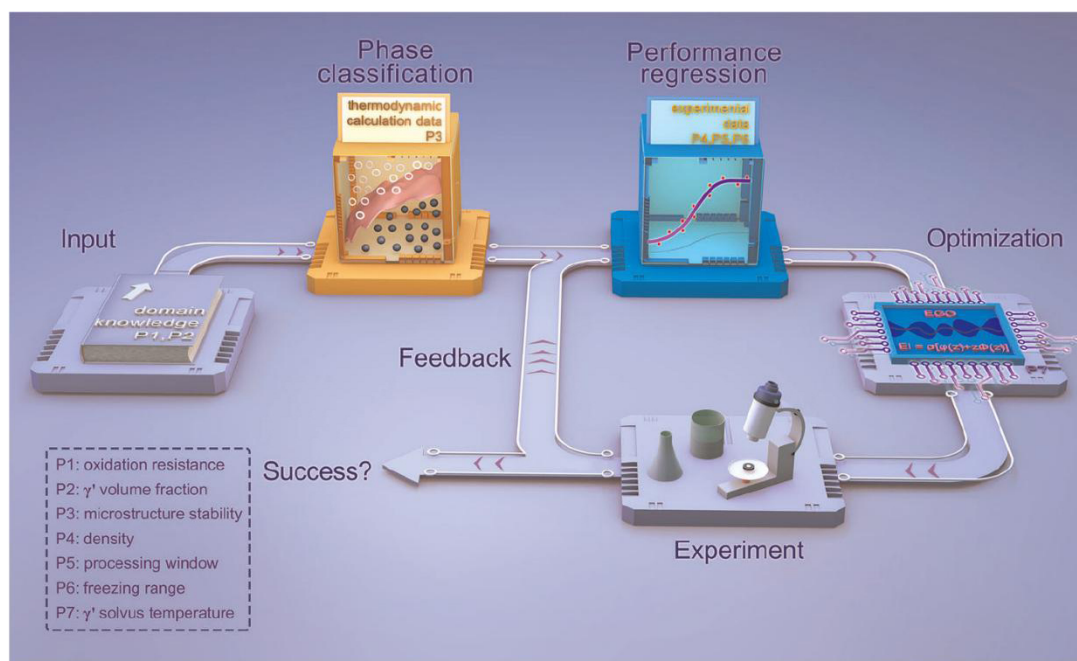


Figure 11. Schematic of multi-performance optimization approach^[22].

become popular choices for medical applications^[53].

In the pursuit of new alloys with low moduli, it has been generally accepted that it is difficult to achieve an excellent synergy between low Young's modulus and β -phase stability in simple alloy systems and low-modulus β -Ti alloys with E between 40 and 65 GPa are generally only realized in multicomponent systems^[23]. However, the influence of alloying elements on the Young's modulus, which have to be determined with mechanical tests, remains complicated and largely undetermined. As such, advances made in the discovery of low-modulus Ti alloys for biomedical implant applications have been limited and time-consuming.

Recently, ML has been finding widespread applications in the design of new β -Ti alloys with low moduli and cost. For instance, Yang *et al.*^[19] proposed a property-oriented ML assisted alloy design strategy to find low-modulus β -Ti alloys in the Ti-Mo-Nb-Zr-Sn-Ta system [Figure 12]. In particular, the ML algorithms are augmented with characteristic parameters, such as the Mo-equivalent parameter (Mo_{eq}) and cluster-composition formula. In the forward predictions that map the alloy composition to the Young's modulus, ML algorithms are encoded with Mo_{eq} that characterizes the β phase stability. In the reverse design of alloy composition(s) for a specific E value, the ML model is integrated with a cluster formula, which reflects the interactions among elements and is expressed with a composition formula to guide the selection of alloy compositions, since a given E value corresponds to a wide spectrum of compositions. Such an inverse design strategy led to the discovery of a new alloy (Ti-2.8Mo-13.5Nb-4.4Zr-8Sn-8.8Ta) with the lowest E of 46 GPa and several other alloys with a specific E , all of which were further experimentally validated. This example demonstrates that, when combined with domain knowledge, ML can be more effective and efficient in accelerating the discovery and deployment of new β -Ti alloys with properties desirable for biomedical applications.

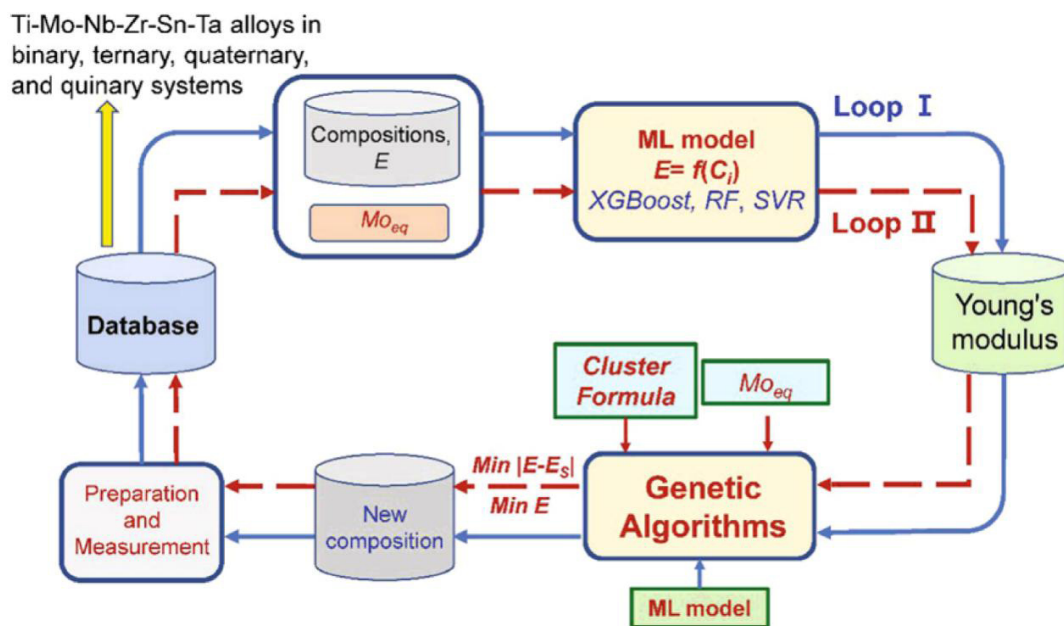


Figure 12. Schematic of ML-assisted design framework for multicomponent β -Ti alloys with low E ^[19].

With the aim of discovering and designing β -Ti alloys with low costs and moduli for biomedical applications, Wu *et al.*^[23] formulated a ML-assisted design framework based on an artificial neural network (ANN) approach. As shown in Figure 13, the framework consists of two ANNs, which are trained for predicting the martensitic transformation starting temperature (M_s) and E , respectively, as a function of alloy composition. The ML machinery, namely, β Low, then combines the two trained ANNs for recommending alloy compositions that simultaneously enable $M_s < 300$ K and a prescribed E value. Consequently, a compositional space is established within which the desired properties are realized. In particular, the ML-assisted approach provides not only high-throughput predictions but also flexibility to meet other specific requirements as demanded. For instance, in the search for low-modulus and low-cost β -Ti alloys, two constraints can be implemented straightforwardly for further composition screening, i.e., $[\text{Nb}] + [\text{Mo}] + [\text{Ta}] < 20$ (in wt.%) and $E < 50$ GPa. Based on the predictions, an alloy (Ti-12Nb-12Zr-12Sn, in wt.%) was discovered with a relatively low E of 42 GPa, as experimentally validated.

It is noted here that the ML strategy is assisted by two additional domain knowledges: one is related to the metastability of the β phase at room temperature and the other one is related to the cost of alloying elements. Moreover, the experimental datasets, collected from published studies, for training two ANN models are relatively small: 164 different alloys for E_s and 112 for M_s , respectively, with only six coincidences having both E and M_s for the same composition. The efficiency and accuracy of β Low thus break the barrier for alloy discovery. It is also worth emphasizing that the compositional space for low-modulus alloys is much wider than that covered by the existing data, which has been obtained with significant efforts over years. Surprisingly, the new alloy (Ti-12Nb-12Zr-12Sn) is not expected from conventional approaches since its composition is located outside and far away from the compositional space covered by existing alloys.

Besides predicting alloy compositions for β -Ti alloys with desired properties, ML is also capable of analyzing the relative importance of different input features. For example, Xiong *et al.*^[20] first used ML techniques informed by density functional theory calculations to guide the selection of Nb- and Zr-based binary Ti alloys with a Young's modulus below 40 GPa. They further used the multivariate decision tree method as an

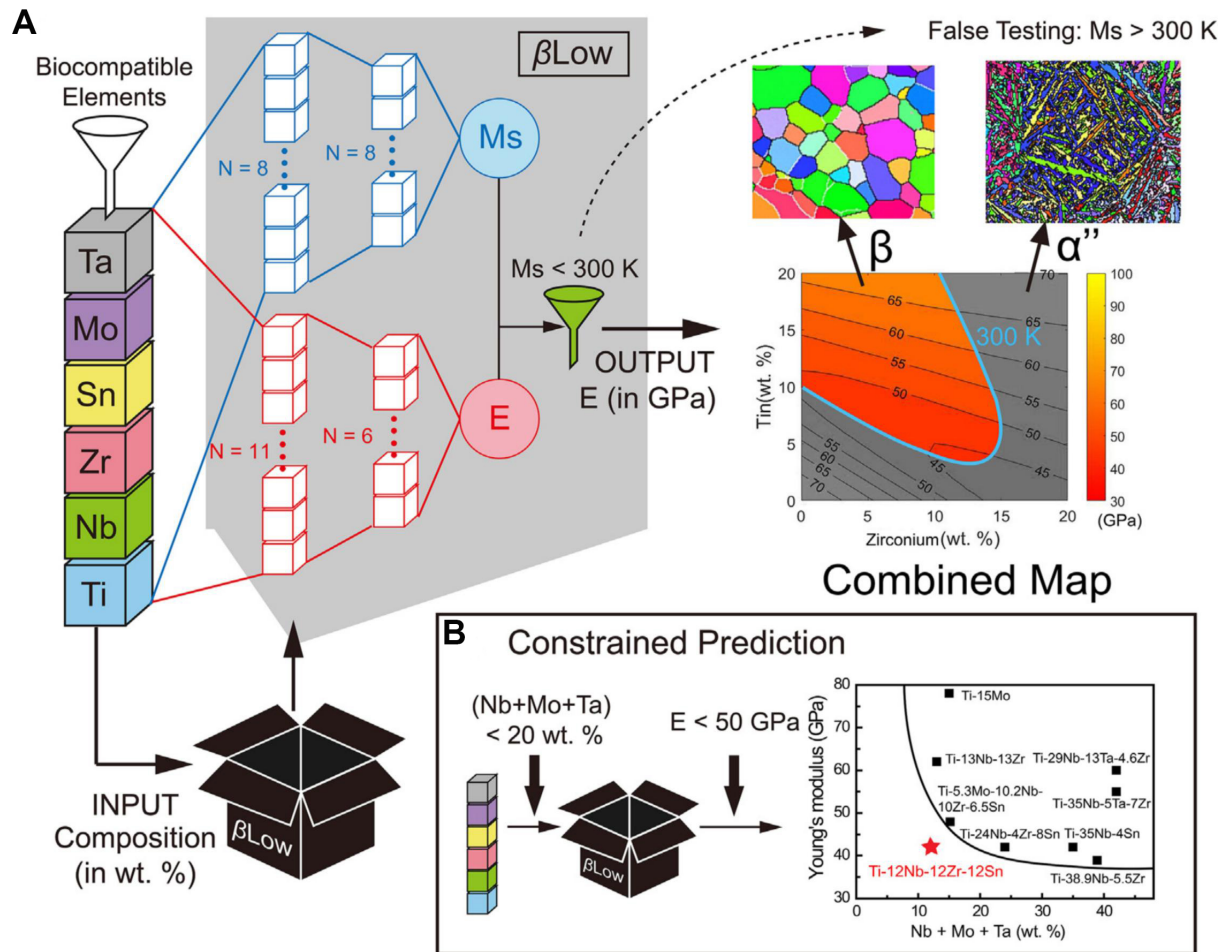


Figure 13. ML-assisted development of low-modulus and low-cost Ti alloys^[23]. (A) Detailed procedures for establishing artificial neural network (ANN)-based machine learning prediction model, namely, β Low. (B) Property-oriented alloy composition prediction with constraint using β Low.

efficient surrogate model for identifying structure variables having relatively high influences on the phase stability and Young's modulus. As shown in Figure 14, in the Ti-Nb and Ti-Zr systems, four key alloying parameters, d_{Bond} (the interatomic length for the nearest neighbors), V , $N(E_f)$ and R carry the importance to E_f (the formation energy per atom) and E from the greatest to the least in order and outweigh other features. Thus, the results suggest that a priority should be given to these four parameters when designing alloys.

The examples in this section demonstrate that ML is a promising tool that is capable of assisting scientists and researchers in finding low-modulus Ti alloys for biomedical applications in a reasonable timescale while at low cost. In particular, the new alloys discovered by ML-assisted approaches may have been previously missed using conventional trial-and-error approaches.

Optimization of steels

Despite the widespread application of the non-ferrous metallic alloys mentioned above, steels remain humankind's most important structural materials and have enabled technological breakthroughs in various fields, ranging from energy and transportation to safety and infrastructure^[11]. With the improvement of data collection technology, the production data for steel can be collected more efficiently. The data could include the contents of more than 20 elements and complicated process parameters^[54,55]. Another example is from

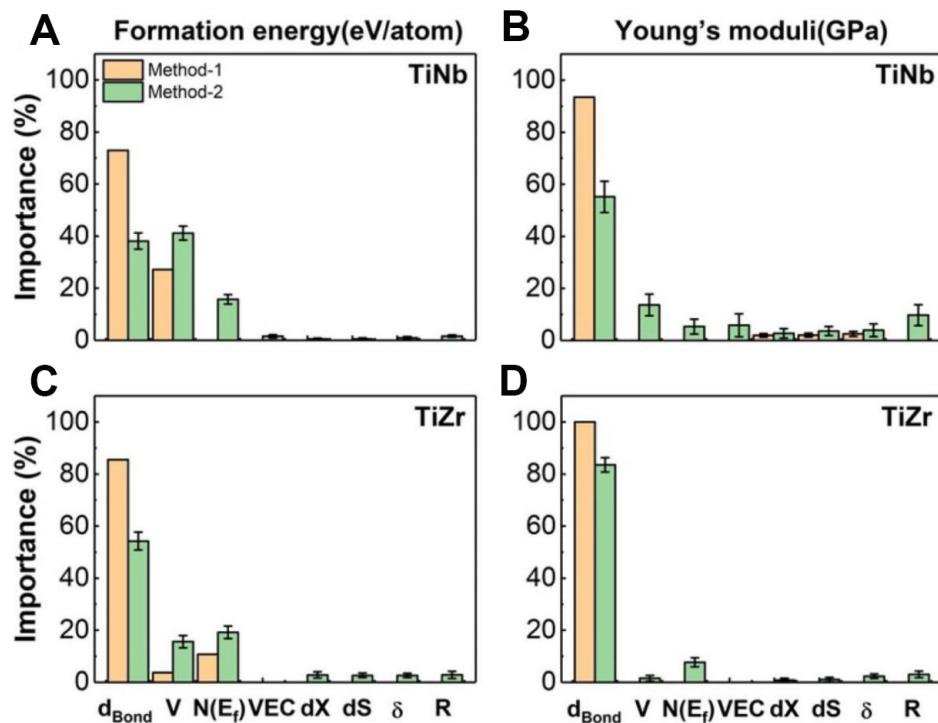


Figure 14. Ranking of importance for different feature parameters^[20] to (A) formation energy and (B) Young's modulus in Ti-Nb system, and (C) formation energy and (D) Young's modulus in Ti-Zr system.

Agrawal *et al.*^[56], who made use of the open access materials database hosted by Japan's National Institute for Materials science. In this work, the authors created a framework (see Figure 15) for exploring materials informatics and highlighted techniques for effective preprocessing, establishing effective feature selection, applying ML techniques to the data and evaluating to avoid overfitting. In a follow-up study, they then conducted improvements to their predictions using various model ensembles^[57], and eventually, they shared the tool online with the community, democratizing the capability to make highly accurate steel fatigue strength predictions^[58].

Materials design and discovery often inherently involve multi-objective optimization, which is more complex than the single-objective optimization. In particular, the optimization of one property may have a negative influence on another. One typical example is the classical problem of strength-ductility trade-off, which remains a difficult challenge. ML also holds significant potential for tackling the multi-objective optimization challenge. As an example, Guo *et al.*^[26] proposed a ML model that transforms the multi-objective problem into a nonlinear programming one, in which each material property is treated as a constraint for another one. Using the model, the boundaries of three properties (e.g., yield strength and ductility) under different alloy contents (see Figure 16) were determined with the results in agreement with engineering experience and materials knowledge. In particular, the quantitative impact of carbon content on the strength of steel, which is learned from ML, is reasonably consistent with previous studies, indicating the reliability and practicability of the model^[26]. On this basis, the alloy composition and processing routes can be obtained that are necessary for achieving the desired properties in steel.

Summary of ML-assisted property prediction and materials discovery and design

All above ML studies for different classes of materials have common goals, namely, exploiting the collected materials datasets for ML model development, rapidly estimating the properties of hundreds to thousands

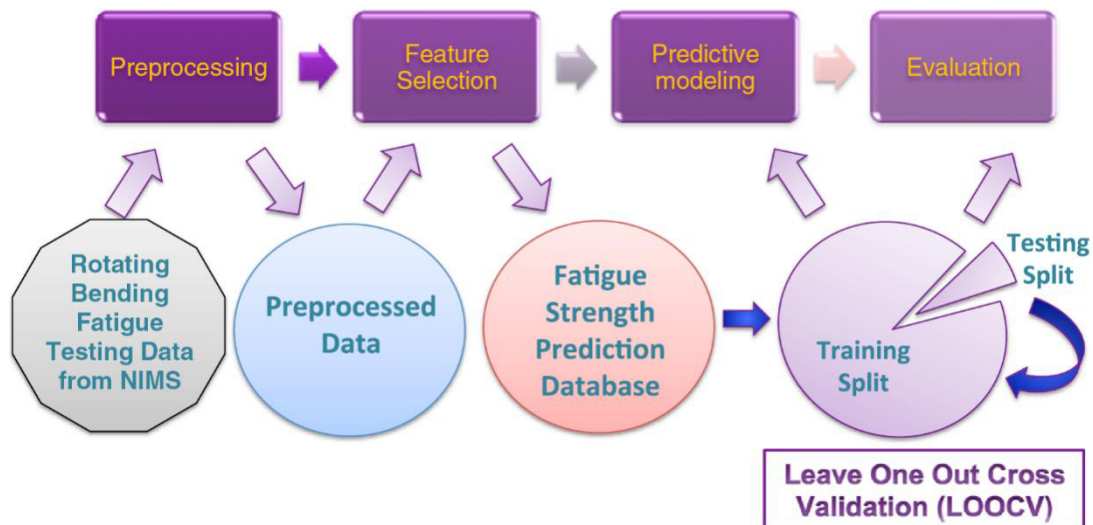


Figure 15. Schematic of data-driven approach to fatigue strength prediction of steels with knowledge discovery and data mining^[56].

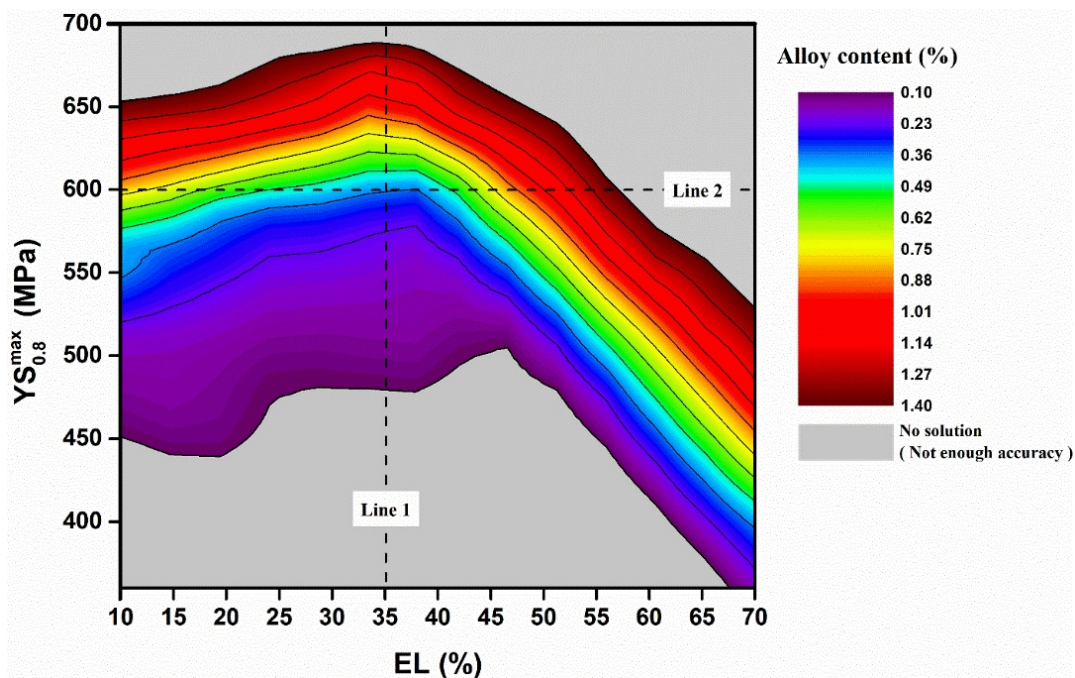


Figure 16. Predicted upper boundaries of steel properties with different alloy contents^[26]. YS: Yield strength; EL: elongation.

of new candidate materials using the trained ML model and virtually screening promising candidates for experimental validation. This has led to the accelerated materials discovery, with successful stories ranging from high-strength Cu alloys, Ni- and Co-based superalloys, low-modulus and low-cost Ti alloys to high-strength steels and other classes of materials not covered in this review.

It is noteworthy that many different algorithms exist, such as linear regressions, support vector machines, RFs and NNs, to name just a few. However, there is no general rule that one algorithm will outperform another in terms of model assessment by some statistics related to the differences between the predicted and

true data (e.g., the root mean squared error or the coefficient of dependence, R^2). Thus, the common practice so far is to implement several different algorithms on the dataset and then select the one with the highest prediction accuracy. Apparently, the best-performing algorithm varies case by case.

There exist two distinct screening pathways for materials discovery using ML models, namely, the forward and inverse designs. Most of the examples above solve the forward problem. To be specific, a “forward” model is built to predict the property of a given input material and using this model with an efficient search strategy in the materials space to screen those candidates that are predicted to meet the property targets. For example, ANN models were used to identify Ti candidates that meet the desired martensitic transformation starting temperature and Young’s modulus from a predefined list of alloy compositions. Another example is the use of a RFR model to screen Ni-based alloys with the desired γ' solvus temperature. In contrast, the inverse problem is tackled by direct generating materials candidates starting from the desired property objectives. For example, a P2C (i.e., property \rightarrow composition) model is developed for predicting the compositions of Cu-alloys based on the targeted properties requirement of hardness and EC.

Materials science generally focuses on multiple properties of the same material: for a new material to be successfully used in an application, it has to meet several property objectives. ML examples covered in this review use separate models for each task (i.e., property). For example, in the discovery of γ' -strengthened Co-based alloys, four separate models are developed to predict four important properties such as the presence of the γ' phase, the presence of the other phases, the solvus temperature and volume fractions of the γ' phase. These single-property models are then combined for alloy screening. ML methods such as multitask learning also exist (even though not covered in this review) that can be used to learn a shared vector representation for all tasks, and then use the shared representation for predicting different output properties collectively. For example, in the case of a NN, this can be achieved by having multiple output layers, one for each property, and a common loss function defined as the weighted sum of errors for each property prediction.

SUMMARY AND PERSPECTIVE

In summary, we have reviewed the current status regarding the data-driven development and design of metallic structural materials, including high-strength Cu alloys, Ni- and Co-based superalloys, biomedical Ti alloys and steels. It is evident that the advent of ML in materials science has provided a new tool for effectively and efficiently analyzing high-dimensional space and identifying correlations among CPSP relationships that may otherwise have been missed using conventional trial-and-error approaches. There is no doubt that ML is transforming the field of materials discovery and design and will continue to do so in decades to come. Through efficiently establishing the linkage among CPSP relationships, one can use ML as a tool, in principle, for attacking the problem of inverse design (i.e., the properties of a specific system are set by a designer and data and ML algorithms are then used to determine the composition and processing parameters needed to create the materials), as demonstrated by the examples of property-oriented design for a wide spectrum of structural materials discussed above. These data-driven approaches enable researchers to bypass the difficulties in tackling a myriad of interactions that underpin the fascinating properties of structural materials, but span across multiple scales, ranging from localized chemical bonding to macroscopic interactions between grains.

From a scientific perspective, the emerging trends shown by research and publications suggest that the full potential of ML-aided materials optimization and discovery is far from being fully realized^[2,9,10,12]. However, this is a rapidly-moving field, so we also provide an outlook and discuss some opportunities and challenges for the materials community for fully realizing the capabilities of ML. While the power and potentials of

data-driven methodologies have been well demonstrated, it is equally important to have a clear picture about the major limitations and challenges that are currently locking the potential for the data-enabled development of future metallic structural materials and beyond.

Improving the interpretability of ML models

As human beings, i.e., users who are asked to judge and/or trust the outcomes of ML-based predictions, we need to understand what model learns after training. Such an innate need is related to the notion of interpretability of the ML model^[59], which refers to the degree to which a ML user can understand and interpret the prediction made by a ML model^[60-62]. Despite the growing use of ML models in materials science and other domains, such as medicine and biology, researchers and/or users still find it difficult to trust and thus rely on these models in practice for different reasons. First, the available predictive models generally target a particular class of materials and rely largely on the domain knowledge of the researchers/users using the model. Second, most of the available ML models developed by data scientists primarily focus on the accuracy of model prediction as a performance metric but rarely explain their prediction in an interpretable way. The inability of ML users to interpret the outcomes of complex ML models hampers the adoption of these ML-based tools. Together with prediction accuracy, there is an increasing need for strategies to assess the correct model complexity and interpretability metrics. Comparing models in terms not only of their predictive accuracy but also of their interpretability metric should become a common practice^[10].

A significant number of ML interpretability techniques and studies have been developed, including global interpretability techniques (e.g., feature importance, partial dependence plots, individual conditional expectation, feature interaction and global surrogate models) and local interpretability techniques (e.g., local surrogate models and the Shapley value)^[62]. These techniques can be applied to present the role of the interpretability technique on assisting the users in different disciplines to obtain a better understanding and thus more trust of the ML-based predictions. We speculate that it is very likely in the future that ML interpretability techniques will form a substantial part of ML workflows and pipelines.

Mutual complementarity of machine learning and experiments

To date, for most materials informatics problems, relevant experimental data are relatively scattered across the literature and other sources, necessitating manual extraction when establishing datasets for model training. The manual construction of training datasets is, in general, a highly time and resource intensive endeavor. Due to the advent of combinatorial and high-throughput characterization techniques^[5,9], the ability to generate compositional and microstructural datasets are far exceeding the ability to analyze them. Nevertheless, a small amount of user input could become a laborious and time-consuming task. Thus, it is often the case in ML studies that one starts with a limited quantity of training data upon which to train an algorithm, searching for new materials in a vast design space. Consequently, the initial trained ML model will likely be incapable of describing the entirety of the design space with the desired accuracy.

An effective and promising mitigation strategy is to adopt the so-called active learning (see [Figure 17](#)^[63]), which is used as an integral guide for data collection efforts, in which the iterative design of experiments (or simulations) is carried out using ML property models and carefully tuned optimization approaches^[64,65]. Active learning begins with a rapid sampling of the potential regions of interest for constructing an initial ML model. It then identifies the ideal candidate experiments to achieve a goal, which is realized by balancing the exploitation of information contained in the existing data in the ML model (i.e., data points with the best predictions) and exploration of less-sample portions of the design space (i.e., data points likely having high model uncertainties). Upon executing these suggested experiments, the model is continually updated as the data from the new results flows in, such that the candidates in the next iteration would be

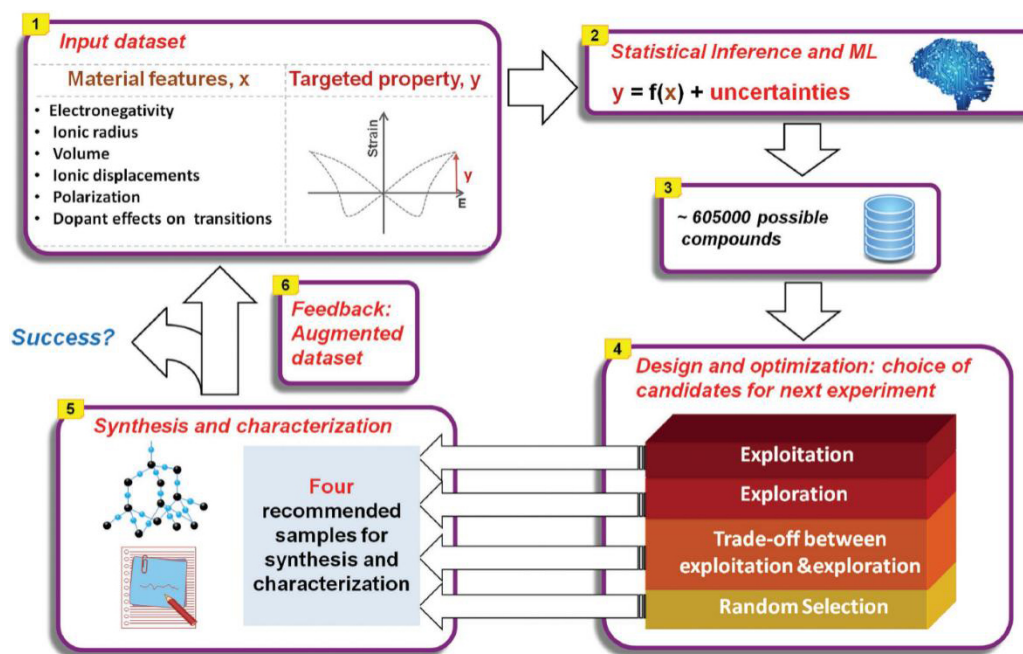


Figure 17. Schematic of an active learning workflow^[63].

likelier to represent improvements (i.e., closer or superior to the target). Active learning has yielded numerous successful outcomes, in which the number of experiments necessary to obtain a target outcome is significantly reduced. Successful examples include, but are not limited to, the discovery of new Pb (lead)-free piezoelectric materials with larger measured electrostrain in the BaTiO₃ family^[63] and new polymers with high glass transition temperatures. In particular, new polymers have been discovered by starting from a remarkably small training data set of only five materials^[66]. The power of active learning highlighted the recent advancements by the examples mentioned in refs.^[63,64,66], which hold significant potential in accelerating the development and optimization of metallic structural materials.

It is noteworthy that ML-based approaches at present do not and will not exclude experiments (even in the future). In fact, the main objective is not to replace experiments but to enhance their capabilities and to improve efficiencies. Moreover, ML-based approaches, or ML-driven materials discovery, rely on them for validating ML predictions and even improve the accuracy of the predictions (see Figure 18). Validation of the suggested experiments based on ML predictions stills need to be conducted. Upon execution of these suggested (either experimental or computational) experiments, the resulting data could further be used to retrain and improve the underlying ML, as being involved in active learning to tackle the small sample challenge ubiquitous in data-driven materials research. Materials discovery should be treated as an active learning problem and should not fall foul of one of the tenets of successful research (i.e., never stop learning), in which, models will be continually updated as new information flows in. Through the dynamic interactions between ML and experimental synthesis and characterization, a fundamental understanding of CSPP relationships can then be established in an accelerated fashion.

Mutual complementarity of machine learning and multiscale modeling

Metallic structural materials are known for having a complex hierarchy of structural and mechanical properties across spatial and temporal scales. Tackling such a complex materials problem, wherein the governing phenomena span different scales of materials behavior, requires multiscale approaches. In the

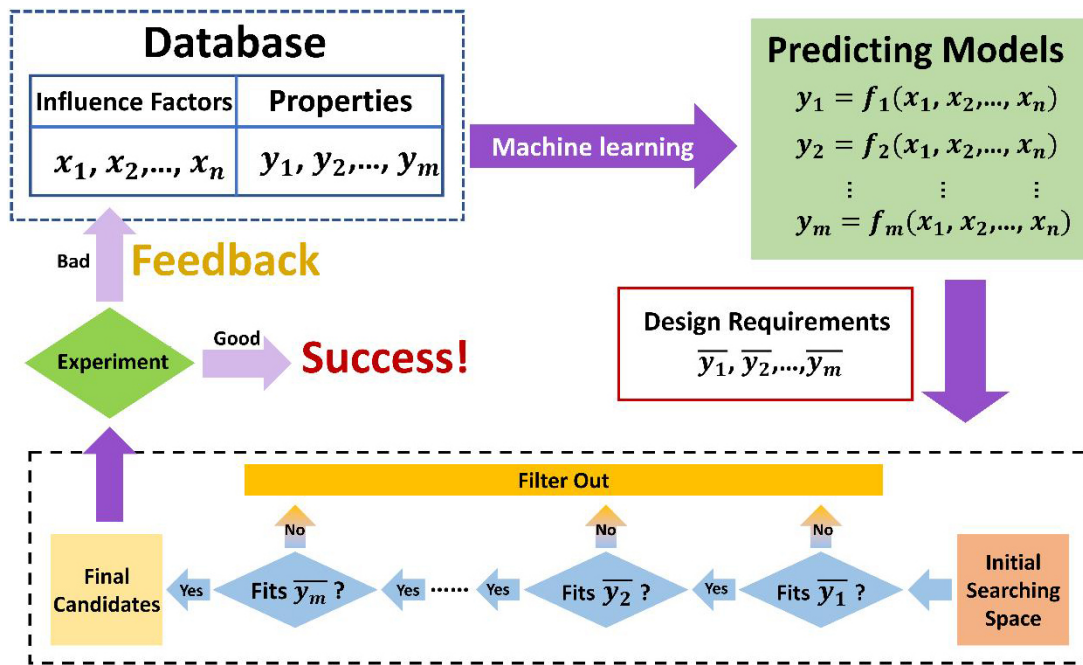


Figure 18. Schematic representation of a ML-based accelerated design strategy^[16].

past two decades, the modeling of these multiscale phenomena has been a point of attention that has advanced detailed deterministic models and their coupling across scales. Advances are as diverse as phase-field methods to model mesoscale behavior and molecular dynamics methods to deduce the fundamental atomic-scale dynamical processes governing materials responses. There can be no argument that multiscale modeling is now a successful strategy for integrating multiscale and multi-physics data, discovering the underlying mechanisms and has become an important tool for the discovery of new materials and materials phenomena and for gaining insight into the processes that govern materials behavior.

However, multiscale modeling alone often fails to efficiently combine large datasets from different sources (i.e., multi-modality) and different levels of resolution (i.e., multi-fidelity). Moreover, the interactions among different scales, both spatial and temporal, coupled with the various physical and chemical processes in different classes of structural materials systems, are complex with many unknown parameters. As a consequence, modeling structural materials in a multi-dimensional parametric space poses challenges of uncertainty quantification.

The recent rise of ML as a powerful tool to integrate multi-modality and multi-fidelity data and to uncover correlations between intertwined phenomena presents a unique opportunity to tackle the challenges facing by multiscale simulations in this regard. Nevertheless, one should be aware of the inherent limitations of ML techniques. For example, problems may arise when dealing with sparse or biased data. In these cases, the naive use of ML may result in ill-posed problems and generate non-physical predictions. Moreover, ML alone ignores the fundamental laws of physics and may result in non-physical solutions. Multiscale modeling, in contrast, has been a successful strategy to integrate multiscale, multi-physics data and elucidate the underlying mechanisms that explain the emergence of the governing equations.

It has become clear that ML and multiscale modeling can naturally complement and mutually benefit each other. By nature, ML is capable of exploring massive design space to identify correlations while multiscale

simulations allow for integrating physics-based knowledge to bridge the scales and efficiently pass information across temporal and spatial scales. Recent trends^[67] suggest that integrating ML and multiscale modeling is a promising strategy and path for better understanding in biological, biomedical and behavioral sciences and beyond, including the creation of robust predictive mechanistic models when dealing with sparse data, managing ill-posed problems and efficiently exploring massive design spaces for identifying correlations. For example, physics-informed ML (see the recent review by Karniadakis *et al.*^[68]) and physics-informed deep learning starts to play a crucial role at the front of ML and multiscale simulations fusion, through leveraging multi-modality and multi-fidelity data with any known physics. In turn, these data can then be exploited to discover the missing physics or unknown processes.

Integrating machine learning with materials science and engineering education

ML has shown tremendous success in areas well beyond metallic structural materials and this review has only touched the surface. As the community sees the emerging interdisciplinary field that harnesses data and combines modeling and simulations, experiments and ML technologies to create a new approach for materials discovery and manufacturing, there exists an increasing need for better education and training for materials scientists to appreciate the magnitude of data required to efficiently apply the techniques of data-driven research. Domain expert knowledge is essential to fully take advantage of the power of ML technologies since they are not directly transferrable to materials. As such, training programs should be tailored for the integration of materials science with ML and taught by experts in computational materials science and ML methods. Such training, at the graduate level and above, should include the best practice for data acquisition, curation, and sharing, and competency in tools that facilitate them. Such education program will train the future generation workforce and offer exciting opportunities for graduates in industries that have incorporated data-driven discovery and materials development.

DECLARATIONS

Authors' contributions

Made substantial contributions to conception and design of this review, writing and editing: Liu X, Wang C

Made substantial contributions to collation of literatures, figures preparation, and writing: Yu J, Xi S, Pan S, Wang Y, Peng Q

Performed data analysis, discussion and writing-review: Yu J, Shi R

Performed data acquisition and interpretation: Yu J, Shi R

Availability of data and materials

Not applicable.

Financial support and sponsorship

This study is supported by Key-area Research and Development Program of Guangdong Province (No. 2019B010943001) and Major-Special Science and Technology Project in Shandong Province (No. 2019JZZY010303).

Conflicts of interest

All authors declared that there are no conflicts of interest.

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Copyright

© The author(s) 2021

REFERENCES

1. Callister WD, Rethwisch DG. Materials science and engineering : an introduction. New York: Wiley; 2018. DOI
2. Rajan K. Materials informatics: the materials “Gene” and big data. *Annu Rev Mater Res* 2015;45:153-69. DOI
3. Williams JC, Starke EA. Progress in structural materials for aerospace systems | The Golden Jubilee Issue - selected topics in Materials Science and Engineering: past, present and future, edited by S. Suresh. *Acta Materialia* 2003;51:5775-99. DOI
4. Rajan K. Combinatorial materials sciences: experimental strategies for accelerated knowledge discovery. *Annu Rev Mater Res* 2008;38:299-322. DOI
5. Miracle DB, Li M, Zhang Z, Mishra R, Flores KM. Emerging capabilities for the high-throughput characterization of structural materials. *Annu Rev Mater Res* 2021;51:131-64. DOI
6. Horstemeyer MF. Integrated Computational Materials Engineering (ICME) for metals: using multiscale modeling to invigorate engineering design with science. John Wiley & Sons; 2012 DOI
7. Van Der Giessen E, Schultz PA, Bertin N, et al. Roadmap on multiscale materials modeling. *Model Simul Mat Sci Eng* 2020;28:043001. DOI PubMed
8. Su Y, Fu H, BAI Y, et al. Progress in materials genome engineering in China. *Acta Metall Sin* 2020;56:1313-23. DOI
9. Kalidindi SR, De Graef M. Materials data science: current status and future outlook. *Annu Rev Mater Res* 2015;45:171-93. DOI
10. Morgan D, Jacobs R. Opportunities and challenges for machine learning in materials science. *Annu Rev Mater Res* 2020;50:71-103. DOI
11. Sparks TD, Kauwe SK, Parry ME, Tehrani AM, Brgoch J. Machine learning for structural materials. *Annu Rev Mater Res* 2020;50:27-48. DOI
12. Suh C, Fare C, Warren JA, Pyzer-knapp EO. Evolving the materials genome: how machine learning is fueling the next generation of materials discovery. *Annu Rev Mater Res* 2020;50:1-25. DOI
13. Hart GLW, Mueller T, Toher C, Curtarolo S. Machine learning for alloys. *Nat Rev Mater* 2021;6:730-55. DOI
14. Ramakrishna S, Zhang T, Lu W, et al. Materials informatics. *J Intell Manuf* 2019;30:2307-26. DOI
15. Saal JE, Oliynyk AO, Meredig B. Machine learning in materials discovery: confirmed predictions and their underlying approaches. *Annu Rev Mater Res* 2020;50:49-69. DOI
16. Yu J, Wang C, Chen Y, Wang C, Liu X. Accelerated design of L12-strengthened Co-base superalloys based on machine learning of experimental data. *Materials & Design* 2020;195:108996. DOI
17. Ruan J, Xu W, Yang T, et al. Accelerated design of novel W-free high-strength Co-base superalloys with extremely wide γ/γ' region by machine learning and CALPHAD methods. *Acta Materialia* 2020;186:425-33. DOI
18. Pan S, Wang Y, Yu J, et al. Accelerated discovery of high-performance Cu-Ni-Co-Si alloys through machine learning. *Materials & Design* 2021;209:109929. DOI
19. Yang F, Li Z, Wang Q, et al. Cluster-formula-embedded machine learning for design of multicomponent β -Ti alloys with low Young's modulus. *npj Comput Mater* 2020;6:1-11. DOI
20. Xiong S, Li X, Wu X, et al. A combined machine learning and density functional theory study of binary Ti-Nb and Ti-Zr alloys: Stability and Young's modulus. *Computational Materials Science* 2020;184:109830. DOI
21. Zhang H, Fu H, Zhu S, Yong W, Xie J. Machine learning assisted composition effective design for precipitation strengthened copper alloys. *Acta Materialia* 2021;215:117118. DOI
22. Liu P, Huang H, Antonov S, et al. Machine learning assisted design of γ' -strengthened Co-base superalloys with multi-performance optimization. *npj Comput Mater* 2020;6:1-9. DOI
23. Wu C, Chang H, Wu C, et al. Machine learning recommends affordable new Ti alloy with bone-like modulus. *Materials Today* 2020;34:41-50. DOI
24. Tamura R, Osada T, Minagawa K, et al. Machine learning-driven optimization in powder manufacturing of Ni-Co based superalloy. *Materials & Design* 2021;198:109290. DOI
25. Qin Z, Wang Z, Wang Y, et al. Phase prediction of Ni-base superalloys via high-throughput experiments and machine learning. *Materials Research Letters* 2021;9:32-40. DOI
26. Guo S, Yu J, Liu X, Wang C, Jiang Q. A predicting model for properties of steel using the industrial big data based on machine learning. *Computational Materials Science* 2019;160:95-104. DOI
27. Wang C, Fu H, Jiang L, Xue D, Xie J. A property-oriented design strategy for high performance copper alloys via machine learning. *npj Comput Mater* 2019;5:1-8. DOI
28. Yu J, Guo S, Chen Y, et al. A two-stage predicting model for γ' solvus temperature of L12-strengthened Co-base superalloys based on machine learning. *Intermetallics* 2019;110:106466. DOI
29. Yi J, Jia Y, Zhao Y, et al. Precipitation behavior of Cu-3.0Ni-0.72Si alloy. *Acta Materialia* 2019;166:261-70. DOI
30. Cheng J, Tang B, Yu F, Shen B. Evaluation of nanoscaled precipitates in a Cu-Ni-Si-Cr alloy during aging. *J Alloys Compd*

- 2014;614:189-95. DOI
31. Hu T, Chen J, Liu J, Liu Z, Wu C. The crystallographic and morphological evolution of the strengthening precipitates in Cu-Ni-Si alloys. *Acta Materialia* 2013;61:1210-9. DOI
 32. Wei H, Chen Y, Zhao Y, Yu W, Su L, Tang D. Correlation mechanism of grain orientation/microstructure and mechanical properties of Cu-Ni-Si-Co alloy. *Mater Sci Eng A Struct Mater* 2021;814:141239. DOI
 33. Li J, Huang G, Mi X, Peng L, Xie H, Kang Y. Microstructure evolution and properties of a quaternary Cu-Ni-Co-Si alloy with high strength and conductivity. *Mater Sci Eng A Struct Mater* 2019;766:138390. DOI
 34. Pollock TM, Tin S. Nickel-based superalloys for advanced turbine engines: chemistry, microstructure and properties. *J Propuls Power* 2006;22:361-74. DOI
 35. Reed RC. The superalloys: fundamentals and applications. Cambridge university press; 2008. DOI
 36. Pollock TM. Alloy design for aircraft engines. *Nat Mater* 2016;15:809-15. DOI PubMed
 37. Smith TM, Esser BD, Antolin N, et al. Phase transformation strengthening of high-temperature superalloys. *Nat Commun* 2016;7:13434. DOI PubMed PMC
 38. Li Y, Ye X, Li J, Zhang Y, Koizumi Y, Chiba A. Influence of cobalt addition on microstructure and hot workability of IN713C superalloy. *Materials & Design* 2017;122:340-6. DOI
 39. Cloots M, Kunze K, Uggowitz PJ, Wegener K. Microstructural characteristics of the nickel-based alloy IN738LC and the cobalt-based alloy Mar-M509 produced by selective laser melting. *Mater Sci Eng A Struct Mater* 2016;658:68-76. DOI
 40. Huron ES, Bain KR, Mourer DP, et al. The influence of grain boundary elements on properties and microstructures of P/M nickel base superalloys. *Superalloys* 2004;2004:73-81. DOI
 41. Horst OM, Schmitz D, Schreuer J, et al. Thermoelastic properties and γ' -solvus temperatures of single-crystal Ni-base superalloys. *J Mater Sci* 2021;56:7637-58. DOI
 42. Chen J, Huo Q, Chen J, et al. Tailoring the creep properties of second-generation Ni-based single crystal superalloys by composition optimization of Mo, W and Ti. *Mater Sci Eng A Struct Mater* 2021;799:140163. DOI
 43. Cormier J. Thermal cycling creep resistance of Ni-based single crystal superalloys. Proceedings of the 13th International Symposium of Superalloys, Superalloys. 2016. p. 385-94. DOI
 44. Rame J, Utada S, Bortoluci Ormastroni LM, et al. Platinum-containing new generation nickel-based superalloy for single crystalline applications. In: Tin S, Hardy M, Clews J, Cormier J, Feng Q, Marcin J, O'brien C, Suzuki A, editors. Superalloys 2020. Cham: Springer International Publishing; 2020. p. 71-81. DOI
 45. Kim H, Park S, Seo S, Yoo Y, Jeong H, Jang H. Regression analysis of high-temperature oxidation of Ni-based superalloys using artificial neural network. *Corrosion Science* 2021;180:109207. DOI
 46. Suzuki A, Inui H, Pollock TM. L1₂-strengthened cobalt-base superalloys. *Annu Rev Mater Res* 2015;45:345-68. DOI
 47. Sato J, Omori T, Oikawa K, Ohnuma I, Kainuma R, Ishida K. Cobalt-base high-temperature alloys. *Science* 2006;312:90-1. DOI PubMed
 48. Makineni SK, Singh MP, Chattopadhyay K. Low-density, high-temperature Co base superalloys. *Annu Rev Mater Res* 2021;51:187-208. DOI
 49. Niinomi M, Liu Y, Nakai M, Liu H, Li H. Biomedical titanium alloys with Young's moduli close to that of cortical bone. *Regen Biomater* 2016;3:173-85. DOI PubMed PMC
 50. Wang Q, Dong C, Liaw PK. Structural Stabilities of β -Ti alloys studied using a new Mo equivalent derived from $[\beta/(\alpha + \beta)]$ phase-boundary slopes. *Metall and Mat Trans A* 2015;46:3440-7. DOI
 51. Sumitomo N, Noritake K, Hattori T, et al. Experiment study on fracture fixation with low rigidity titanium alloy: plate fixation of tibia fracture model in rabbit. *J Mater Sci Mater Med* 2008;19:1581-6. DOI PubMed
 52. Abdel-Hady Gepreel M, Niinomi M. Biocompatibility of Ti-alloys for long-term implantation. *J Mech Behav Biomed Mater* 2013;20:407-15. DOI PubMed
 53. Eisenbarth E, Velten D, Müller M, Thull R, Breme J. Biocompatibility of beta-stabilizing elements of titanium alloys. *Biomaterials* 2004;25:5705-13. DOI PubMed
 54. Peters H, Ebel A, Havkmann J, et al. Industrial data mining in steel industry. 30th Journées Siderurgiques Internationales (JSI); 2012 Dec; Paris. *Stahl und Eisen* 2012;132.
 55. Ordieres-Meré J, González-Marcos A, Castejón-Limas M, Martínez-de-Pisón FJ. Chapter 20: Data mining experiences in steel industry. Handbook of research on machine learning applications and trends: algorithms, methods and techniques. Capítulo de Libro; 2010. p. 427-39. DOI
 56. Agrawal A, Deshpande PD, Cecen A, Basavarsu GP, Choudhary AN, Kalidindi SR. Exploration of data science techniques to predict fatigue strength of steel from composition and processing parameters. *Integr Mater Manuf Innov* 2014;3:90-108. DOI
 57. Agrawal A, Choudhary A. A fatigue strength predictor for steels using ensemble data mining: steel fatigue strength predictor. Proceedings of the 25th ACM International on Conference on information and knowledge management; 2016 Oct 24. 2016; p. 2497-500. DOI
 58. Agrawal A, Choudhary A. An online tool for predicting fatigue strength of steel alloys based on ensemble data mining. *Int J Fatigue* 2018;113:389-400. DOI
 59. Poon AIF, Sung JJY. Opening the black box of AI-Medicine. *J Gastroenterol Hepatol* 2021;36:581-4. DOI PubMed
 60. Hakkoum H, Idri A, Abnane I. Assessing and Comparing Interpretability Techniques for Artificial Neural Networks Breast Cancer Classification. *Comput Methods Biomech Biomed Eng Imaging Vis* 2021;9:587-99. DOI
 61. Linardatos P, Papastefanopoulos V, Kotsiantis S. Explainable AI: a review of machine learning interpretability methods. *Entropy*

- (*Basel*) 2020;23:18. [DOI](#) [PubMed](#) [PMC](#)
62. Elshawi R, Al-Mallah MH, Sakr S. On the interpretability of machine learning-based model for predicting hypertension. *BMC Med Inform Decis Mak* 2019;19:146. [DOI](#) [PubMed](#) [PMC](#)
 63. Yuan R, Liu Z, Balachandran PV, et al. Accelerated discovery of large electrostrains in BaTiO₃-based piezoelectrics using active learning. *Adv Mater* 2018;30:1702884. [DOI](#) [PubMed](#)
 64. Lookman T, Balachandran PV, Xue D, Yuan R. Active learning in materials science with emphasis on adaptive sampling using uncertainties for targeted design. *npj Comput Mater* 2019;5:1-17. [DOI](#)
 65. Lookman T, Balachandran PV, Xue D, Hogden J, Theiler J. Statistical inference and adaptive design for materials discovery. *Curr Opin Solid State Mater Sci* 2017;21:121-8. [DOI](#)
 66. Kim C, Chandrasekaran A, Jha A, Ramprasad R. Active-learning and materials design: the example of high glass transition temperature polymers. *MRS Communications* 2019;9:860-6. [DOI](#)
 67. Peng GCY, Alber M, Tepole AB, et al. Multiscale modeling meets machine learning: what can we learn? *Arch Comput Methods Eng* 2021;28:1017-37. [DOI](#) [PubMed](#) [PMC](#)
 68. Karniadakis GE, Kevrekidis IG, Lu L, Perdikaris P, Wang S, Yang L. Physics-informed machine learning. *Nat Rev Phys* 2021;3:422-40. [DOI](#)