A review of machine learning for structural materials

Taylor D. Sparks[1], Steven K. Kauwe,[1] Marcus E. Parry,[1] Aria Mansouri Tehrani,[2] and Jakoah Brgoch[2]

[1]Department of Materials Science and Engineering, University of Utah, Salt Lake City, United States, 84112
[2]Department of Chemistry, University of Houston, Houston, United States, 77204

Xxxx. Xxx. Xxx. Xxx. YYYY. AA:1–25

https://doi.org/10.1146/((please add article doi))

Keywords

machine learning, high entropy alloys, structural materials, bulk metallic glasses, fatigue, materials informatics

Abstract

The development of structural materials with outstanding mechanical response have long been sought for innumerable industrial, technological, and even biomedical applications. However, these compounds tend to achieve their fascinating properties from a myriad of interactions spanning multiple scales, from localized chemical bonding to macroscopic interactions between grains. This has limited the ability for researchers to develop new materials on a reasonable timeline. Fortunately, the advent of machine learning in materials science has provided a new approach to analyze high-dimensional space and identify correlations among the structure-composition-property-processing relationships that may have been previously missed. In this review, we examine some successful examples using data science to improve known structural materials by analyzing fatigue and failure as well as discuss approaches to develop entirely new classes of structural materials in complex compositions spaces including high entropy alloys and bulk metallic glasses. Highlighting the recent advancement in this field demonstrates the power of data-driven methodologies that will hopefully lead to the production of market-ready structural materials.

Contents

1. INTRODUCTION

Over the course of history humanity has engineered their surroundings into the tools and materials they needed. Perhaps the best example of this has been the evolution of structural materials over time. The fundamental role of any structural material is their load bearing nature. Consider the homes and buildings in which we live, from the most humble hut made of natural materials to the modern skyscraper made exclusively from engineered materials like modern steel, concrete, and glass.

This progression from natural to engineered materials certainly included improvement to load bearing aspects such as strength, hardness, and toughness. Nevertheless, the application of structural materials has evolved with modern technology to incorporate additional constraints and materials requirements. Consider the hubris of the engineer who thought to design an airplane that could carry 838 passengers over almost 15,000 kilometers, more than a third the circumference of the earth. The material requirements for this go far beyond the natural world. Indeed, the aluminum alloys making up the skin of modern day aviation have incredible strength to density ratios, just to name one aspect. In the case of Al 319, this achievement required eleven different elements, about a tenth of the usable periodic table (1). In addition to this absurd compositional complexity, the weight fraction requirements extend into 4 significant digits (2). Now compare that to the "original" aviation alloy, the Wright Flyer crankcase (3). In this alloy 8% copper was simply added to aid in processing but had the unexpected result of strengthening via precipitate hardening (4). It took over 100 years for the modern Al 319 to emerge with its incredibly specific and complex composition, but the result is an alloy that vastly supersedes previous innovation.

Even if flying is a rare event, consider the daily exercise of charging your cellphone or even turning on the lights. The typical assumption is that reliable, inexpensive electrical power will be available everywhere we go. Making this possible is the unseen innovation of the gas turbine engine which provides over a third of the electrical power of the United States. The materials operating in these turbines are exposed to incredibly severe environments. For example, combustion gases reach up to 2000°C. System innovations such as thermal barrier coatings (5) and micro cooling channels (6) can do much to address problems associated with thermal stability. Other challenges require materials innovation

to address oxidation and creep resistance. In response to the failure of steel and other state-of-the-art materials, engineers had to develop superior materials such as single-crystal nickel superalloys for the turbine blades (7). Today we enjoy unprecedented energy efficiency.

Given the complexity of modern aviation and turbines, it is easy to understand why these applications would require similarly complex structural materials. At the same time, the human body itself is far more complex than any engineered system and so it too requires advanced materials for biomaterials. Within the last 160 years, the average life expectancy in the United States went from 35 to 76 years old (8). Similar trends are occurring worldwide. This sets the stage for an interesting biomedical engineering challenge. Parts of the human body that experience high wear and cyclic loading such as knees, hips, valves, and more are wearing out and eventually failing. Today, artificial prosthesis to improve quality of life at advanced ages and in the cases of accidents or defects are widespread and increasingly common. For example, there were over 300,000 hip replacements in the United States in 2010 alone (9)! Demand is greater than ever for effective structural biomaterials to serve as joints, stents, screws, valves, and more. Unique biomaterials such as titanium alloys and Nitinol have been developed as structural materials to address unique challenges in biocompatibility and corrosion, but they also require mechanical properties such as high hardness for wear resistance and low elastic modulus to better mechanically couple to biological structural materials like cortical bone (10, 11).

These vignettes show how structural materials are making possible incredible advances in aviation, energy, and biomaterials but also unfortunately serve as reminders of how slow materials discovery, development, and deployment can be. Indeed, the typical timeline for this process can span decades (12).

The postwar era of the 1950s and 60s saw incredible technological advances in energy, telecommunication, computing, materials and much more. Thiel and Masters defined this era as *definite optimism* (13) since the society in this era imagined a specific future unlimited by current technological constraints. Consider the 1954 speech by the chairman of the United States Atomic Energy Commission:

> It is not too much to expect that our children will enjoy in their homes electrical energy too cheap to meter, will know of great periodic regional famines in the world only as matters of history, will travel effortlessly over the seas and under them and through the air with a minimum of danger and at great speeds, and will experience a lifespan far longer than ours, as disease yields and man comes to understand what causes him to age (14).

Even though we have not yet reached the aspirations of 1954, today we nevertheless have ambitious expectations for structural materials of the future as we hope for hypersonic or interstellar flight, blast and impact resistant armor, clothing and fabrics that never wear out, energy conversion approaching Carnot efficiency, rapid drilling, polishing, and grinding at low cost, extremely efficient transportation and buildings, and much more. The realization of mankind's ambitions will rely on the continued improvement and development of advanced structural materials and there exists greater need than ever to accelerate this the production of market-ready materials. Fortunately, the boom in materials informatics, or the application of data science to materials research, could lead to greatly accelerated production of novel materials (15, 16, 17, 18, 19, 20).

Structural materials is a broad class and includes structural glasses, ceramics, concrete and cements, alloys, and more. For this review, we restrict our summary to materials

informatics-based research on investigating fatigue and microstructure of alloys, properties of high entropy alloys, and the glass forming ability of bulk metallic glasses. These materials each play critical roles in transportation, energy, and biomedical engineering; however, there remains much to be done and understood. For example, fatigue is the leading cause of failure and remains a critical problem in aviation and many other applications. A great deal of research has gone into quantifying the relationships between fatigue and microstructure for alloys and exploring ways to model the entire fatigue process with machine learning (ML) approaches. High entropy alloys are poised to displace state-of-the-art structural materials in a variety of applications featuring extreme environments. However, modeling the behavior has been challenging considering their chemical makeup and structural complexity. Bulk metallic glasses feature a unique combination of biocompatibility, high hardness, high strength, and low stiffness. Not all compositions form bulk metallic glasses, therefore, a great deal of data-driven research has been directed towards better understanding the glass forming ability of the compositions and to predict potential new compositions.

This review article will assume a basic understanding of machine learning principles and shall instead focus on describing the application of data science on structural materials. However, as a general introduction to describe the essential processes involved in a materials informatics-based approach, consider the following case study on the accelerated discovery of superhard materials.

1.1. Case study: Accelerated discovery of superhard materials via materials informatics

Hardness, defined as the surface resistance of a material to indentation, is potentially one of the main mechanical characteristics that is sought after in a multitude of applications for wear resistance. Consequently, the quest to achieve materials with higher hardness has entertained scientists for decades. Interestingly, a naturally occurring metastable form of carbon, diamond, is one of the hardest known materials with a Vickers hardness of approximately 96 GPa surpassed only by wurtzite BN and lonsdaleite (21). Diamond itself forms under extreme temperatures and pressures and even synthetic, polycrystalline diamond requires temperatures and pressures of $1400°C$ and 6 GPa that limit its applications due to costly synthesis (22).

Materials are deemed "superhard" if their Vickers hardness exceeds 40 GPa. Over the years, only a handful of signle phase compounds, such as c-BN and BC_2N, prepared at high temperatures and pressures have been reported to exhibit superhardness. More recently, it was proposed that the incorporation of light main group elements in the structure of 5d-transition metals could enhance their hardness by creating strong localized bonds within the metallic framework (23, 24). ReB_2 and WB_4 are the best example of such materials surpassing the superhard region at low indentation loads while only requiring high temperature for synthesis. Despite such a breakthrough, the progress of finding new intrinsic superhard materials has remained challenging.

The compositional space for new candidate superhard materials is vast and the search for new materials would benefit from a screening technique to reduce the number of potential candidates. Materials informatics methods such as machine learning can be implemented to reduce the amount of trial-and-error necessary, and, therefore, accelerate the discovery rate. Recently, Mansouri Tehrani and coworkers employed just such a materials informatics approach and successfully identified two new superhard materials (25).

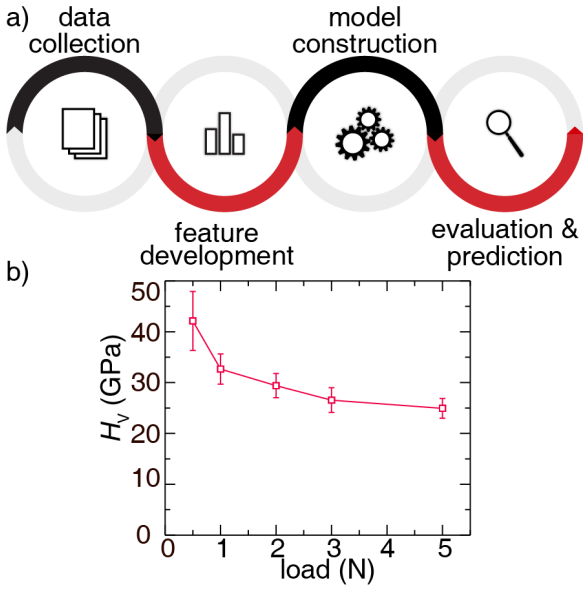


Figure 1

a) Overview of a typical machine learning work-flow. Figure reproduced from (26). b) Vickers hardness measurement of $Mo_{0.9}W_{1.1}BC$ at various indentation loads. Figure reproduced from (25).

The first step to any materials informatics approach is to identify, collect, and organize data, Figure 1a. Unfortunately, compiling a comprehensive list of compositions with experimentally measured hardness to be used as training data for machine learning algorithms is quite arduous made worse by the fact that reported hardness data are often inconsistent; for example not reporting the indentation load, or using various indentation techniques compounds the problem. Alternatively, researchers have utilized bulk modulus and shear modulus as proxies for hardness since these properties are related (27). More importantly, the source of the moduli can be from density functional perturbation theory (DFPT) calculations, which in this work stem from the Materials Project database for 2572 inorganic compositions.

Once data is identified, the next step is to featurize the data. Materials informatics approaches typically rely on a composition-based or structure-based feature vectors. In the work by Mansouri Tehrani and coworkers, both composition and structure-based features were used for predictions with 71 unique features being used for bulk modulus and 55 features being used for shear modulus (25). After the data were assembled and featurized the next step is to select an algorithm and then train and validate a model (Figure 1a). Many different algorithms exist such as linear regressions, support vector machine (SVM), random forests, and neural networks. K-fold cross-validation and grid searches are employed to optimally train a model and then validation is done on a completely separate test data set. This worked that an SVM was able to predict bulk modulus with a root mean square error (RMSE) of 17.21 GPa (R^2=0.94) and shear modulus with a RMSE of 16.35 GPa (R^2=0.84) suggested excellent predictive power of their model.

Machine learning based on this model can then be employed to rapidly screen candidate

materials from known structural databases. In this example, the researchers were able to predict 15770 binary, 56266 ternary, and 46251 quaternary compounds. In addition to rapid predictions ($\sim 30\,s$ of wall time) the algorithm was able to generate predictions for many disordered structures and f-electron containing compounds that would not have been easily possible to screen with DFT alone. The team then analyzed the predicted data set for the highest predicted bulk and shear modulus values. Among these were two compositions, $Mo_{0.9}W_{1.1}BC$ and $ReWC_{0.8}$ that stood out and were therefore selected for experimental investigation. These compounds were synthesized and Vickers hardness measurements were carried out. Interestingly, both of these compositions exhibited high hardness even reaching the 40 GPa limit of superhardness at low indentation loads. Figure 1b shows the load-dependent Vickers hardness measurement of $Mo_{0.9}W_{1.1}BC$ confirming high hardness even at the asymptotic region. The team then followed up with studies to confirm the predicted bulk modulus values, examine lattice strain and texture formation upon loading (28), and explain the chemical origin of the high hardness and surprising ductility in $Mo_{0.9}W_{1.1}BC$ (29) and other compounds in the chemical series (30). This case study is an excellent example of the powerful role that machine learning can play in not only the identification of new structural materials, but it also shows how a research project that begins with machine learning can develop into a grander endeavour.

2. FATIGUE AND MICROSTRUCTURE OF ALLOYS

In this section we cover machine learning approaches focused on steels and alloys. These materials are critical to modern society and represent incredible amounts of scientific and engineering effort over the past century. Although the literature in this field encompasses a staggering amount of material, the application of machine learning on these systems has largely focused on two critical areas. First, there has been a focus on the prediction of fatigue properties, and second the generation of quantitative representations of microstructure (QRM) is a major research effort. We look at these two areas in detail by examining the progress these fields have made and highlight many of the influential works.

2.1. Fatigue properties

Fatigue is when materials weaken upon cyclic loading conditions. Weakening can result in fracture and failure after a low or high number of cycles. Since fatigue represents the most common form of failure of metals and alloys, it is of particular interest of study for materials scientists, mechanical engineers, and civil engineers alike. Not all materials are equally susceptible to fatigue. In fact, some steel and titanium alloys exhibit a fatigue limit after which, theoretically, a material can be cyclically loaded indefinitely without fatigue failure. Other materials weaken continuously until failure upon cycling. The application of machine learning techniques for predicting properties related to fatigue extends back more than two decades.

The majority of machine learning applications for predicting steel and alloy fatigue properties can be considered within the realm of two different applications. The first application focuses on learning from large macroscopic features in relation to the structural failure of steel and alloy components. This topic is more thematically appropriate for civil engineering tasks and is not covered here. Instead, we focus on works that predict fatigue behavior as a materials property using techniques and data that are reminiscent to tradi-

tional fatigue characterization and modeling but approached from a data-driven direction. Among these works are a particular focus on predicting fatigue crack growth and fatigue strength, which we cover below.

2.1.1. Fatigue crack growth. Traditional approaches to predicting fatigue crack growth often assume a mathematical model, *e.g.*, the Paris-Erdogan model, to relate stress intensity factor range and fatigue crack growth rate. Although these fracture mechanics concepts have served well in guiding the structural decisions of many materials scientists, there have been a number of machine learning approaches that have been developed to predict fatigue crack growth in a data-driven manner. Among these approaches, some predict crack growth directly, while others predict useful parameters pertaining to traditional fracture mechanics concepts used to predict crack growth rate.

One of the early works for the prediction of fatigue crack growth rate was done by Bhadeshia et al. using Bayesian neural networks on nickel-based superalloys (31). This work used a database of almost 2000 instances of fatigue crack growth with 51 inputs to the network including the stress intensity factor range, the chemical composition, the temperature, the grain size, the condition of heat treatment, frequency, loading condition, atmosphere, r-ratio, load waveform, sample thickness and yield strength. They showed effective ability to predict fatigue crack growth rates across a large variety of superalloys.

In 1997, shortly after the work of Bhadeshia, Kang et al. explored fatigue crack opening load using machine learning. This subsequent work showed that a neural network approach could take the differential displacement signal curve as an input, and return the crack opening load while avoiding the use of empirical parameters (32). To accomplish this outcome, simulated data for generating and learning on these curves was used. They then showed their approach was capable of predicting the experimental crack opening load at acceptable values for 2024-T351 aluminum alloy even when trained on the simulated single phase data. Neural network based learning was also investigated by Iacovillo et al. as a way to build a nonlinear model the could remarkably learn *beyond* the Paris region (see Figure 2). They trained their models on experimental results using a duplex austeniticferritic stainless steel obtained by powder metallurgy routes using AISI 316 LHC (70%) and AISI 434 LHC stainless steels powders and showed very good agreement between the network performance and the held-out experimental data.

In contrast to purely data-driven modeling, Mohanty et al. were able to demonstrate the symbiotic nature of physical and statistical modeling by incorporating a Gaussian process and state-space model in conjunction for crack growth predictions as a function of cycles (33). This was achieved by training the Gaussian process to predict crack length given the number of cycles. This output was then used to calculate a constraint factor which is used in the state-space model to generate a final prediction. This hybrid approach was more successful than a pure physics or data-driven approach. They theorized this success was due to the Gaussian process accounting for uncertainty in the crack length, avoiding the issue of integrating error associated with a purely state-space model.

More recently, Wang et al. performed a comparison study where they made predictions on the fatigue crack growth rate using three different machine learning algorithms (34). For this task, they used three different neural net architectures: extreme learning machines (ELM), radial basis function networks, and genetic algorithm optimized back propagation. Each network used the stress intensity factor range (ΔK) and the stress ratio as inputs to predict the crack growth rate. They demonstrated that ELM was the only algorithm

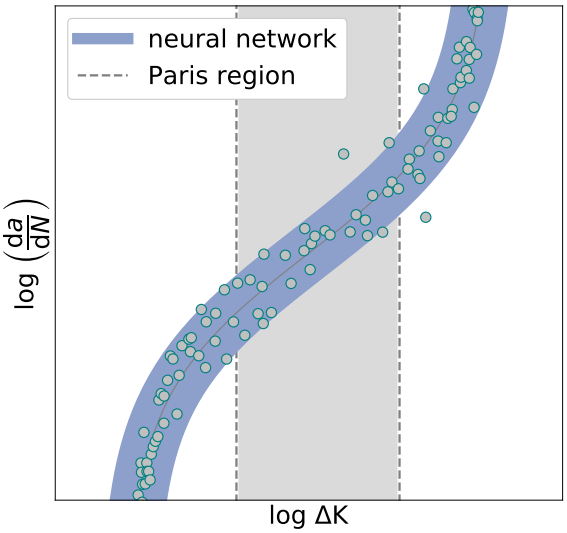


Figure 2

A neural network-based modeling approach allows researchers to model the crack growth rate through all regimes. Contrast this to the common Paris model which only captures the linear behavior, hence the name "Paris region".

which succeeded at both interpolation, and extrapolation beyond the training data. Using these three models, they also showed that the algorithms were more robust than the two-parameter k^* method (35), with the advantage of being able to predict different sample types including 7050-T7451 aluminum alloy, Ti6Al4V titanium alloy, ADB610 steel and D16 aluminum. They suggest the use of a data-driven approach for predicting crack growth if one has access to sufficient data.

Machine learning applications in fatigue prediction have not been limited to so called "abstract" or "black-box" modeling. Physics-inspired materials models have also been developed for describing the small fatigue crack driving force in polycrystalline materials. The work of Rovinelli *et al.* employed two Bayesian networks, one trained on crack propagation rate and the other trained on crack propagation direction. In this work, directly generating predictions is not the primary goal of the network. Instead, the learned parameters are ultimately used to ascertain a deterministic driving force mechanism behind small fatigue crack growth. Although multiple micromechanical variables and features had already been identified in influencing the propagation of small fatigue cracks, they show the Bayesian network prediction outperforms the fatigue metrics available in the literature. Furthermore, the proposed expression, though worse than the Bayesian network itself also outperforms previous literature metrics and also provides a meaningful explanation as to the driving forces behind the small fatigue crack propagation.

2.1.2. Fatigue strength. Studies have not been limited to fatigue crack propagation. Indeed, there has also been interest in predicting the fatigue strength of materials. One of the earlier works, by Canyurt *et al.* predicted the fatigue strength of bonded joints by employing a genetic algorithm in combination with cubic and quadratic fits to the data (36). The

genetic algorithm considered the design parameters of surface roughness, bonding clearance, interference fit, temperature and adherent material and devised proper coefficients to fit the experimental results. They concluded that a genetic algorithm could be an alternate technique for discovering near-optimal solutions in regards to fatigue strength estimations for bonded materials.

Fatigue strength has also been to subject of a series of publications showing how machine learning tools can be used to explore the effects of composition and processing parameters on the fatigue strength of steel. Work done by Agrawal *et al.* makes use of the open access materials database hosted by Japans National Institute for Materials Science. In their first work, they focused on creating a framework for exploring materials informatics by highlighting techniques for effective preprocessing, establishing effective feature selection, applying machine learning techniques to the data, and effective evaluation to avoid overfitting (37). Their followup works then establish improvements to their predictions using various model ensembling (38), and finally, they publish the tool online democratizing the ability to make highly accurate steel fatigue strength predictions (12).

Overall, we see a positive application of machine learning techniques for the prediction of fatigue properties. In general, machine learning models have been shown to be particularly potent for specific experimental or simulation setups. We have also seen exhibits of minor generality with effective predictions outside of the material group used for training. However, the use of these models is limited to the training data and, to our knowledge, there has been no attempt for creating a general model for all structural systems.

2.2. Microstructure

Microstructural predictions using machine learning take a number of different forms based on a general theme: the generation of statistically accurate representations and/or reconstructions of the microstructure. Methods regarding this task have focused on generating lower dimensional representation of statistical descriptors, and end-to-end learning frameworks.

2.2.1. Reduced-order statistical representation. The classification and reconstruction of microstructure using statistical methods is a well established practice (39, 40, 41). Machine learning has been consistently used in this domain for reducing the feature space, therefore increasing the efficiency of the representation. An early example of this was shown by Sundararaghavan et al. who extracted planar image statistics from the 3D microstructures (42). They then showed that principle component analysis (PCA) could be applied to these statistics and an SVM could be used to classify microstructures showing the effectiveness of PCA for creating a reduced-order representation of microstructure. In 2015 Xu et al. demonstrated how a more meaningful reduction of statistical features could be developed by using a machine-learning based method to eliminate redundant microstructure descriptors while preserving physical relevance (43).

Concurrently, DeCost and Holm used a "bag of visual features" to create generic microstructural signatures for classifying different microstructures (44) showing that a machine learning-based process could preserving accurate representation, and therefore function for classification of microstructures. This same machine learning functionality was demonstrated by Liu et al. who used effective microstructure classification to reduce microstructural search space for the optimization of magnetoelastic Fe-Ga alloy microstructure (45).

In 2017 DeCost et al. also showed how computer visions systems could be used to extract features by processing micrograph images, which could then be fed to a SVM to classify powder feedstocks (46).

2.2.2. End-to-end learning. Although many works use statistical methods for featurizing images, featurization can take the form via purely data-driven methods. For example, in 2016, Cang et al. showed that a 5-layer deep belief network, using orientation-invariant filters, could create microstructure features that drastically reduce dimensionality while creating visually plausible reconstructions across various systems corresponding to Ti-6Al-4V alloy (47). Moreover, these representations were also shown to preserve the original critical fracture force of the non-reconstructed system. Chawdhury et al. also explored featurization using neural networks in 2016 (48). However, instead of using a deep belief network trained on micrograph images, they passed the micrographs into a pre-trained caffe model(49) and used the output from the fully connected 6th layer as features for learning (Illustrated in Figure 3). Interestingly, they showed that this featurization step performs better than texture and shape statistics or a visual bag of words despite never being trained on micrograph images.

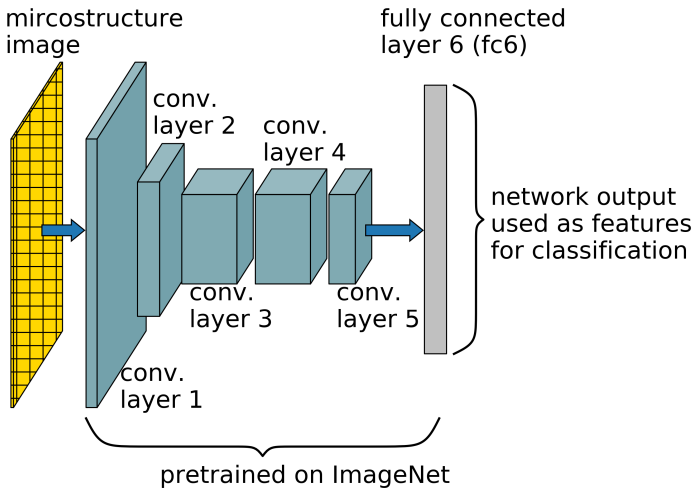


Figure 3

Effective classification of microstructure can be done with relatively few data. This is achieved using transfer learning where the convolutional filters from a neural network, pre-trained on natural images, can be applied to extract features from the micrographs. These features can then be used with statistical learning tools to achieve excellent results.

Lubbers et al. also used a pre-trained neural network for microstructural identification (50). In their work, they use multidimensional scaling to compare the vector representation from a pre-trained VGG (51) network and standard two-point correlations by reconstructing a 1D manifold of images with fixed noise amplitude and scale parameter, but varying angle. They show that the CNN-based texture representation outperform 2-point statistics and is applicable to small synthetic datasets. Indeed, the use of a pre-trained

VGG network connected to a learning pipeline even demonstrates successful classification in the in-718 dataset which is difficult even for expert humans (52). Using an approach similar to Lubbers, Cecen et al. explored how a neural network and 2-point spatial correlation based features compare when used to explain structure-property linkages for effective elastic properties of high contrast composites (53). Using a database of 5900 3D microstructures, they were able to train a 3D convolutional neural network (CNN) to generate features that were then input to a linear regression to predict the effective elastic properties. 2-points statistics were used on these same volumes and PCA was used to get a reduced representation before being used in a linear regression. They show that although the CNN produces better results, the optimal model is built using both 2-point and CNN filters for the input to the linear regression.

3. HIGH ENTROPY ALLOYS

High entropy alloys (HEAs), or multi-principal element alloys (MPEAs), are an emerging, intriguing class of materials which appeared in the early 2000s. Initial study of HEAs was motivated by two separate research groups; one intrigued by the immense, unexplored compositional phase space of multicomponent alloys (54), and another interested in controlling configurational entropy to favor single-phase solid solutions over intermetallic phases (55). HEAs typically consist of five or more principal elements, each with a concentration between 5 at.% and 35 at.%, but may also contain minor elements at lower concentrations. The unique compositional makeup of HEAs fundamentally deviates from historically conventional structural alloys, in which a material is comprised of a single base element with appealing properties, while small amounts of alloying elements may be added for property optimization. As illustrated in Figure 4, the large majority of conventional structural alloys contain a dominant base element of at least 80 wt.%, while all but few sophisticated superalloys and austenitic nickel alloys contain a base element of at least 50 wt.%. (56, 57) HEAs present the opportunity to explore unfamiliar compositional phase space and for novel materials discovery.

Due to entropic effects, HEAs exhibit a high degree of solid-solid solubility and can possess exceptional properties. For instance, MPEAs have illustrated extraordinary strength, ductility, and fracture toughness at cryogenic temperatures (58), as well as exceptional mechanical properties, thermal properties, and irradiation resistance at high temperatures (59). The vast compositional phase space and potential to tailor hardness, strength, oxidation and corrosion resistance, irradiation resistance, and thermal stability drive the appeal for HEAs as structural materials of the future. However, complex chemistries, data limitations, and the near infinite compositional range of HEAs present challenges to alloy design. Materials informatics techniques are key to combat these challenges and explore the vast phase space. In this section, we turn our focus towards some of the data driven approaches taken to achieve accelerated HEA design, while the many unanswered questions and untapped potential relating to HEAs are reviewed elsewhere (57, 60, 61, 62, 63).

3.1. Phase formation

The discovery of novel single-phase solid solutions is a primary motivation in HEA research because single-phase solid solutions have promising mechanical properties for structural applications, such as increased strength, ductility, and the potential for solid solution hard-

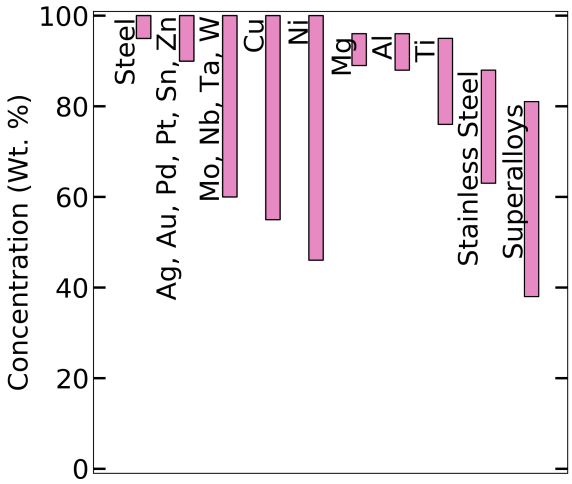


Figure 4

The concentration of the base element in commercially and industrially relevant alloys and steels is portrayed. The large majority have a base element concentration greater than 80 wt.%, and all have a base element concentration above 50 wt.% aside from a select few superalloys and austenitic nickel alloys. Data extracted from (56); figure reproduced from (57).

ening. MPEAs with intermetallic phases are also of interest as strengthening may occur via precipitate hardening. Several groups have utilized computational strategies to predict the formation of single-phase HEAs, implementing first-principles DFT calculations and calculated phase diagram (CALPHAD)-based approaches to assess phase formation and stability (62, 64, 65, 66, 67, 68). Though useful, the computational expense of these methods limits their utility in the accelerated exploration of unfamiliar HEA compositional space. The recent implementation of data-driven techniques confronts this issue in an attempt to efficiently predict phase formation in HEAs.

For example, Islam et al. employed a neural network architecture on a data set of 118 equiatomic or near-equiatomic, as-cast MPEAs to predict phase formation; amorphous, intermetallic, or solid solution (69). The model established correlations between one categorical feature, describing the phase, and five numerical features, including valence electron concentration (VEC), difference in Pauling electronegativities ($\Delta\chi$), atomic size difference (δ), mixing enthalpy (ΔH_{mix}), and mixing entropy (ΔS_{mix}). Cross-validation revealed a predictive accuracy of 83% on this limited data set, indicative of the potential aptitude for machine learning techniques to predict phase formation as the field progresses and data becomes more readily available. Furthermore, Islam and coworkers ranked the importance of the features, reporting that VEC was the most influential to phase formation, while ΔS_{mix} illustrated the least relative importance. These findings corroborate a previous experimental study which explored the relationship between the phase stability of an HEA to the physicochemical and thermodynamic properties of its constituent elements (70). Guo et al. determined that the ΔH_{mix} is key to the formation of solid solutions, but VEC decisively establishes the stability of face-centered cubic (FCC) and body-centered cubic (BCC) phases in HEAs. These results are invaluable to the future design of HEAs with tailored ductility.

Following up to the study by Islam and coworkers, Huang et al. comparatively examined three common machine learning frameworks to assess the predictive accuracy of the phase selection of HEAs (71). A K-nearest neighbors algorithm, SVM, and artificial neural network were implemented on an expanded experimental dataset (63) comprised of 401 HEAs, partitioned into 174 solid solution phases, 54 intermetallic phases, and 173 combinations of solid solution and intermetallic phases. A supervised multi-layer feed-forward neural network outperformed an unsupervised neural network model, as well as the K-nearest neighbors and SVM models. Additionally, Huang and coworkers illustrated that the supervised neural network more accurately classified phase selection than parametric methods. The relative importance of the five input features was again evaluated, illustrating that the VEC and δ features are significantly more important in determining phase selectivity than $\Delta\chi$, ΔH_{mix}, and ΔS_{mix}. The architecture of the supervised neural network is illustrated in Figure 5, consisting of an input layer, three hidden layers with bias nodes, and an output layer. Three hidden layers were chosen to optimize testing accuracy while minimizing computational cost. The hidden layers, combined with sigmoid activation functions, are key to the effectiveness of the framework as they learn correlations to solve the nonlinearity problems within the data set.

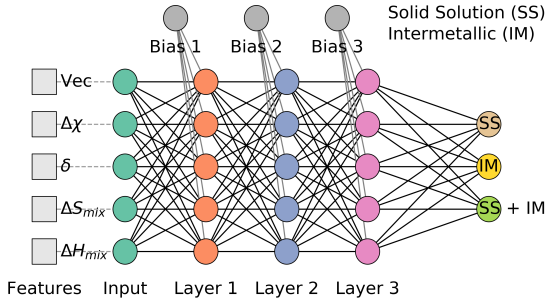


Figure 5

A supervised multi-layer feed forward artificial neural network illustrating the five numerical input features, input layer, three hidden layers, bias nodes, and output layer. Figure reproduced from (71).

3.2. Mechanical properties

Although informatics-based approaches in HEA research have primarily focused on phase selection, several studies have applied machine learning techniques to predict hardness and other mechanical properties of HEAs directly. Due to the limited availability of hardness data across the vast compositional phase space for HEAs, this task proves difficult. In one study, Chang et al. employed artificial neural network-based methods to suggest high-hardness HEAs within a non-equimolar, six-component AlCoCrFeMnNi system, maintaining the fraction of each element between 5 at.% and 35 at.% (72). To optimize the concentration of each element for hardness, a simulated annealing algorithm was implemented, iteratively adjusting composition from an initial $Al_{0.3}$CoCrFeMnNi alloy. The optimization

process is illustrated in Figure 6, with compositional variations revealing the importance of Co, Cr, and Al in achieving high hardness in contrast to Ni, Mn, and Fe. Three compositionally distinctive candidates with predicted high hardness were selected for experimental study. Each alloy exhibited Vickers hardness > 600 HV, while one composition was measured at 650 HV, greater than the highest literature-reported hardness value (539 HV) for this alloy system at the time (73). Hardness results were also correlated to structural and phase information, providing insight to inform data frameworks in the future. This study illustrates the potential to efficiently explore unknown compositional space to optimize hardness using a limited database, though only within a specific HEA system consisting of a specified set of elements.

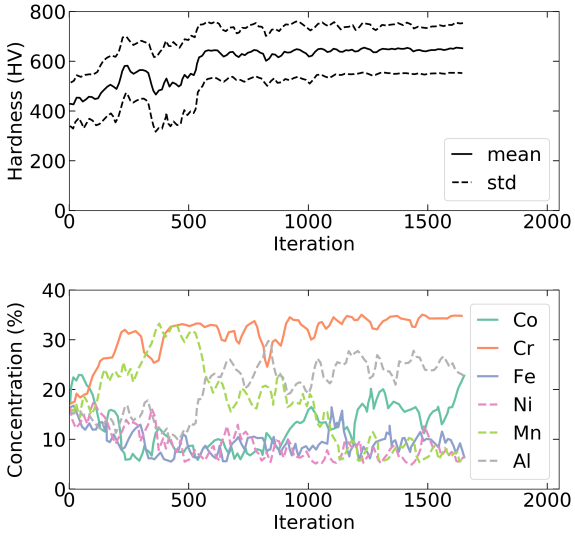


Figure 6

Alloy composition is iteratively optimized (bottom) to achieve high predicted hardness (top). Figure reproduced from (72).

A similar study was performed on an AlCoCrCuFeNi system, in which Wen and coworkers applied a machine learning surrogate model to identify high-hardness HEAs within the specified system (74). In addition, the study considered experimental design and feedback in an iterative optimization strategy, resulting in improved model performance with each iteration. Utility functions were used to select experimental candidates for fabrication and testing, then to reinform the training data. Of the 42 alloys synthesized, 35 exhibited hardness values greater than the highest value in the training dataset, while 17 alloys saw an increase of more than 10% when compared to the maximum hardness value in the training data. The authors suggested that a similar design strategy be applied to the search for optimized lightweight HEAs with high strength, and to extend the framework to the study of bulk metallic glasses and superalloys.

While the studies mentioned above demonstrate the ability to optimize hardness of a particular HEA system comprised of a specific set of elements, predicting entirely novel HEA compositions in the broader phase space remains a challenge. In the search for new, useful HEAs with high strength and high hardness, Rickman et al. employed a canonical corre-

lation analysis combined with a genetic algorithm optimization strategy to systematically explore HEA chemistry and compositional space, rapidly screening alloys to accelerate the identification of HEAs with targeted properties (75). A dataset of 82 HEAs with experimentally measured Vickers Hardness values was utilized as a training set, while seven quantities were considered as predictor variables; Radius asymmetry (δ), ΔH_{mix}, ΔS_{mix}, mean melting temperature (T_m), ratio of entropy/enthalpy (Ω), Young's modulus asymmetry (ϵ), and VEC. Four outcome variables were considered; the presence of BCC solid solution(s) only, the presence of FCC solution(s) only, the presence of intermetallic phases (IM), and the Vickers hardness. It was found that ΔS_{mix}, ΔH_{mix}, and VEC were most influential to hardness. Compositional disorder (ΔS_{mix}) was positively correlated to hardness enhancement. Conversely, ΔH_{mix} scaled inversely to hardness, illustrating that phase separation results in decreased hardness. Furthermore, decreased VEC encourages the formation of a BCC phase, consistent with an increase in strength and hardness (76). To evaluate the method, Rickman et al. selected seven candidate alloys suggested by the genetic algorithm for fabrication and indentation testing, demonstrating with 90% confidence that the process appropriately suggested alloys with enhanced hardness and illustrating the capabilities of machine learning to jointly suggest mechanical properties and phase information of HEAs.

Another research group has performed related investigations in the search for low density, high-hardness HEAs (77, 78). Menou et al. implemented a multi-objective genetic algorithm, allowing for the concurrent optimization of density, single-phase stability, and solid solution hardening, while permitting the efficient screening of non-equimolar HEA compositions (77). The proposed method was applied to the design of FCC HEAs, resulting in the selection of an $Al_{10}Co_{17}Fe_{34}Mo_5Ni_{34}$ (at.%) alloy with appealing predicted properties for further exploration. Characterization of the alloy revealed superior hardness, yield stress, and tensile strength at a lower or comparable density to other FCC HEAs previously examined. In a related study, Menou et al. employed a multi-objective genetic algorithm to explore light and strong BCC HEA compositions (78), as BCC phases may exhibit improved solid solution hardening at lower densities when compared to FCC phases (63). From more than 3000 suggested compositions with optimal phase stability, solid solution hardening, and density, a BCC $Al_{31}Cr_{37}Mn_7Mo_6Ti_{19}$ (at.%) alloy was selected for experimental characterization. Menou et al. observed exceptional solid solution hardening, a hardness of 658 HV, and a density below 5.5 g/cm^3, making the selected alloy one of the hardest materials yet reported for such low density.

4. BULK METALLIC GLASSES

Bulk metallic glasses (BMGs) are an exciting class of amorphous materials that display a varied range of physical properties due to their vast potential compositional space and lack of long-range crystallographic order (79). These glassy alloys specifically have some of the unique mechanical properties including near-theoretical strength and large elastic strain limits (80, 81, 82, 83). Researchers are therefore continually working to not only identify novel BMGs but they are also developing tools to predict glass forming ability *a priori*. However, the multicomponent nature of these systems has generally limited the ability to predict the novel BMGs. As a result, most of the BMGs identified thus far have been discovered by trial and error, using simple empirical rules, or through considerable experimental effort (84).

One method that researchers have considered is using materials screening by searching

for correlations among independent mechanical properties. For example, as discussed in Section 1.1, hardness is known to be correlated to shear modulus (G). Indeed, a correlation between hardness (H_V) and G has already been established for crystalline materials, $H_v = 0.151G$, and also used to analyze the intrinsic hardness of 37 BMGs. The results reveal reasonable agreement between the predicted hardness and the experimentally measured values. This simple empirical equation is an excellent starting point that can be utilized in predicting the mechanical properties of BMGs (84). High-throughput methods are another way researchers have tackled exploring the glass forming ability of BMGs as well as to analyze the composition-based functional properties such as hardness (85, 86, 87). Employing a systematic approach using a laser additive manufacturing techniques to deposit the library of continuously graded Cu-Zr-Ti alloys produced an extensive number of compositions while subsequent nanoindentation measurements revealed a relationship between the hardness and ductility that was dependent on the Cu content with a notable improvement in the ductility for lower Cu concentrations (87). Other BMG compositions such as $Zr_{50}Cu_{50}$ have also demonstrated an incredible plasticity, strain stiffening, and enhanced ductility (88). Additionally, high toughness and ductility with simultaneous high strength can be achieved through bulk metallic glass matrix composites of Ti-based systems (89). These observations are all particularly interesting because BMGs are generally considered brittle materials. The massive chemical tunability of BMGs provides an incredible avenue to acquire unanticipated structural functionality that are not readily available in crystalline compounds.

Accelerating the development of BMGs require two fundamental steps. First, the prior glass forming ability of the constituent elements should be predicted with high accuracy and ease. Second, the mechanical attribute-composition relationship should be realized to only target compositions with high potential. Machine-learning methods are a natural pathway that have already been pursued by researchers with ample opportunity remaining.

4.1. Descriptors for bulk metallic glasses

Machine-learning models are constructed based on a combination of training data represented in vector formats that can be described as linear coefficients of features capable of capturing the systems essential physics or chemistry. These coefficients are then optimized during the training process until reaching a reasonable degree of predictive power for instances outside of the training data set. One of the biggest limitations for machine learning, in particular for materials science, are the limited amount of data that can be represented by the model. A solution to acquire a statistically robust model even with a small data set is to engineer simple, yet highly-correlated proxies that correlate to the desired properties but are simple to estimate. For example, critical cooling rate is an essential factor for the formation of metallic glasses but these data are not readily available. The unavailability as well as the difficulty in attaining the critical cooling rate limits its suitability to be included as a feature despite the high relevance. The critical cooling rate is, however, dependent on some thermodynamic and kinetic properties that are more easily accessible. The entropy of fusion and kinetic fragility are shown to demonstrate high correlation to the critical cooling rate for the formation of metallic glasses which can therefore be implemented in subsequent machine learning analysis as important physical descriptors (90).

To correlate attributes to the glass forming ability of BMGs, research has postulated that the existence of many competing meta-stable crystalline structures with different structures

are crucial for the formation of BMGs. This competition is likely to hinder the growth of nucleation before reaching the critical radius that results in crystallization. Therefore, utilizing high-throughput DFT calculations within the framework of AFLOW, a simple descriptor is constructed, called entropic factor. The entropic factor indicates the frustration of the formation of a single homogeneous crystal structure. High values of the entropic factor correlates positively with the glass forming ability of a particular structure. Based on this analysis, 17% of binary alloys are predicted to be able to form as glasses which demonstrate the breadth of possible glassy materials that are unexplored (85). Another example of work in this field, has turned to use genetic programming to find correlation to the critical diameter of BMGs. The critical diameter which is defined as the maximum diameter at which metallic glass can be formed is reported to be a valid indicator of glass forming ability. A parameter (G_p) based on glass transition temperature, onset crystallization temperature, and offset temperature of melting is then introduced showing a high correlation coefficient, 0.67. G_p is later implemented in using principle component analysis (PCA) to develop a predictive model for glass forming ability of BMGs (86, 91). Construction of such descriptors, as opposed to exhaustive and random selection of descriptors, provide an opportunity to build strong predictive models with even limited number of training data that is commonly observed in materials science informatics problems.

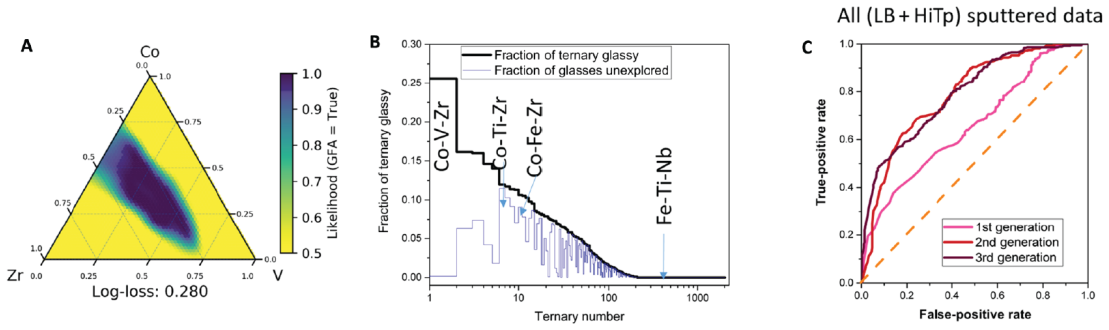
4.2. ML-prediction of glass forming ability

A majority of research applying machine learning to BMGs has centered on predicting glass forming ability. For example, PCA has been used based on 594 BMGs defined by eleven descriptors leading to classification of the clusters of BMGs (92). A general machine-learning framework applicable to the prediction of glass forming ability of BMGs was then constructed. This research shows that partitioning the inputs into similar groups can results in restriction of the degrees of freedom that are needed to capture a phenomenon and therefore enhancing the outcome. For BMGs, a random forest classifier is implemented and trained on 5369 compositions derived from melt-spinning experiments. 90% accuracy is achieved using a 10-fold cross validation and the method is further verified by confirming the accurate reproduction of the experimental Zr-Al-Ni ternary phase diagram (93).

Support vector machine (SVM) algorithms were also employed to predict the glass forming ability of BMGs on a small data set based only on compositional information. To distinguish compositions that can be formed as glasses and the ones that cannot, both positive and negative instances of such compositions are required in the training set. Unfortunately, it is usually not straightforward to find negative reports within the literature. Regardless, 31 binary alloys are collected with their composition ranging from 5 to 95% with 1% increments used as training data. Due to the limited number of training data, all data are reserved for training instead of setting a portion of the data aside as a test set. Instead, two new groups of data are introduced as the test set. A set called the "Target" which includes 339 binaries that form metallic glasses using the melt-spun technique. The other set, called "All", consists of 1131 possible binary compositions with available input data. The performance of the model is evaluated by its capability to distinguish between these two groups, as glass former and non-glass former. Initially, eleven descriptors were utilized; however, exhaustively exploring the influence of all the possible combination of these descriptors revealed that the best performing model is obtained when only two descriptors, fictive liquidus temperature and the difference in liquidus temperature, are included with

the latter being the most influential descriptor (94).

A combined machine learning and high-throughput experimental method has shown a significantly accelerated rate for discovering metallic glasses. Initially, an ML model was created using a random forest algorithm to describe the glass forming ability of compositions. Two different theories are then added as descriptors to potentially enhance the models. Even though these two descriptors do not clearly improve the overall statistics, as they most likely are already covered within the rest of the variables, the nuances that they make turns out to be of merit later when comparing machine learning prediction to experimental results. Synthetic routes are also taken into consideration in this research as they are imperative to the formation of BMGs. This initial model shows high accuracy when assessed based on grouping cross-validation as well as the subsequent high-throughput experimentation on the Co-V-Zr ternary phase space leading to new metallic glasses. Following the experimentation in the Co-V-Zr phase space, a down-sized set of the new data are included in training set to improve the model accordingly. The down-sizing is performed to avoid the introduction of bias by the addition of many data-points within the Co-V-Zr composition-space. The inclusion of these data are particularly useful as they contain many negative instances. As shown by Figure 7a, the second-generation of ML model is capable of accurately predicting the glass forming ability of the Co-V-Zr with log-loss of only 0.28 compared to 1.75 of the first generation. Further, the ternaries are ranked in a decreasing order as shown by the black line in Figure 7b whereas the blue line indicate the fraction of the compositions that have not been experimentally measured. Therefore, two more ternaries systems with high glass forming ability and low number of known compounds have been selected for further analysis (Co-Ti-Zr and Co-Fe-Zr) while Fe-Ti-Nb is picked as a system with an extremely low predicted glass forming ability. Subsequent experimentation confirm the accuracy of the ML predictions for all the selected systems. Finally, Figure 7c demonstrate the ROC curve of the first, second, and the third generation of ML models highlighting how this iterative experimental/ML approach can lead to more accurate models and hence discovery of new compounds (95).



Revised predictions for Co-V-Zr ternary - 2nd iteration

Predictions of ternaries with largest GFR from 2nd generation ML model

Comparison of ROC curves for 1st, 2nd, and 3rd generation ML model for sputtered synthesis

Figure 7

a) Glass forming ability (GFA) using the second iteration ML model b) Prediction and selection of ternaries with the highest glass forming ability that are also the most unexplored. c) Comparing ROC between the first, second, and third ML generations cross-validated against all sputtered co-deposited values.

4.3. Beyond glass forming ability

A major challenge in this field is that identifying deep eutectic points in simple systems is difficult and information on quaternary or higher systems are rarely known. Therefore, researchers have gone beyond text recognition and have identified deep eutectic points of phase diagrams showing another viable platform to predict glass forming ability of BMGs (96). Research has also shown it is possible to predict the glass forming ability of BMGs, have trained machine learning models to predict critical casting diameter and supercooled liquid range of metals. These models are then used to predict these properties for 2.7 million adjustment of alloys and the top candidates have been experimentally made showing superior properties (97). Finally, a machine learning framework using a gradient boosted decision tree model is made to correlate plastic heterogeneity of BMGs to quench-in structural defects in Cu-Zr (98).

5. CHALLENGES AND FUTURE DIRECTIONS

In this review we have covered many of the current works regarding the data-driven study of structural materials. Researchers have used machine learning for high-accuracy modeling of fatigue behavior, and classification/reconstruction of micrographs. Researchers have also demonstrated the successful discovery of new HEAs, new superhard materials, as well as the ability to predict BMG formation. While this is encouraging, these are only first steps. Many of the limitations and challenges for structural materials mirror those of the broader materials informatics field (20, 99, 100). Future work is required to unlock the potential for data-enabled development of future structural materials. The most pressing need is for additional data. There is an enormous amount of experimental and computational data being generated at present. Despite this monumental effort, data is often unusable for data-driven research due to a lack of standardization at publication. As a result, materials properties are not located in databases that can be easily accessed for machine learning, visualization, and data analytics. Moreover, when data is available for learning, the community has not always adopted best practices. In particular, overly optimistic results have often been reported due to overfitting associated with improper use of training and test data. An emphasis on experimental validation and repeatability made possible by code archival is helping to overcome this challenge. As we work towards better practices regarding data and modeling, we expect the field to produce viable and novel results. There is outstanding potential to apply data science across materials science and it will also play a central role in the advancement of structural materials.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

M.E.P., S.K.K., and T.D.S. acknowledge funding and support from the National Science Foundation through NSF-CMMI 1562226 and NSF-DMR 1651668. A.M.T and J.B. thank the National Science Foundation through No. NSF-CMMI 15-62142, the R.A. Welch foundation (E1981), the Texas Center for Superconductivity at the University of Houston, and

Seed Funding for Advanced Computing (SeFAC) at the University of Houston for for generous financial support of this research.

LITERATURE CITED

1. F. C. R. Hernandez, J. M. H. Ramírez, and R. Mackay, "Applications in the automotive and aerospace industries," in *Al-Si Alloys*, pp. 163–171, Springer, 2017.
2. P. Voorhees. personal communication.
3. F. W. Gayle and M. Goodway, "Precipitation hardening in the first aerospace aluminum alloy: the wright flyer crankcase," *Science*, vol. 266, no. 5187, pp. 1015–1017, 1994.
4. L. S. Hobbs, "Wright brothers' engines and their design," *Smithsonian Annals of Flight*, 1971.
5. D. R. Clarke, M. Oechsner, and N. P. Padture, "Thermal-barrier coatings for more efficient gas-turbine engines," *MRS bulletin*, vol. 37, no. 10, pp. 891–898, 2012.
6. C.-P. Lee, R. Darolia, and R. E. Schafrik, "Turbine blade tip having thermal barrier coating-formed micro cooling channels," Oct. 8 2002. US Patent 6,461,107.
7. P. Caron and T. Khan, "Evolution of ni-based superalloys for single crystal gas turbine blade applications," *Aerospace Science and Technology*, vol. 3, no. 8, pp. 513–523, 1999.
8. M. Roser, "Life expectancy," *Our World in Data*, 2019. https://ourworldindata.org/life-expectancy.
9. H. M. Kremers, D. R. Larson, C. S. Crowson, W. K. Kremers, R. E. Washington, C. A. Steiner, W. A. Jiranek, and D. J. Berry, "Prevalence of total hip and knee replacement in the united states," *The Journal of bone and joint surgery. American volume*, vol. 97, no. 17, p. 1386, 2015.
10. H. Li and Y. Zheng, "Recent advances in bulk metallic glasses for biomedical applications," *Acta biomaterialia*, vol. 36, pp. 1–20, 2016.
11. D. V. Louzguine-Luzgin and A. Inoue, "Bulk metallic glasses: Formation, structure, properties, and applications," in *Handbook of magnetic materials*, vol. 21, pp. 131–171, Elsevier, 2013.
12. A. Aspuru-Guzik and K. Persson, "Materials acceleration platform: Accelerating advanced energy materials discovery by integrating high-throughput methods and artificial intelligence.," 2018.
13. P. A. Thiel and B. Masters, *Zero to one: Notes on startups, or how to build the future*. Broadway Business, 2014.
14. L. L. Strauss, 9 1954. Remarks by Chairman Lewis L. Strauss at the Founder's Day Dinner, National Association of Science Writers.[Accessed: 2019 10 26].
15. K. Rajan, "Materials informatics," *Materials Today*, vol. 8, no. 10, pp. 38–45, 2005.
16. A. Agrawal and A. Choudhary, "Perspective: Materials informatics and big data: Realization of the fourth paradigm of science in materials science," *Apl Materials*, vol. 4, no. 5, p. 053208, 2016.
17. S. R. Kalidindi, D. B. Brough, S. Li, A. Cecen, A. L. Blekh, F. Y. P. Congo, and C. Campbell, "Role of materials data science and informatics in accelerated materials innovation," *Mrs Bulletin*, vol. 41, no. 8, pp. 596–602, 2016.
18. M. W. Gaultois, A. O. Oliynyk, A. Mar, T. D. Sparks, G. J. Mulholland, and B. Meredig, "Perspective: Web-based machine learning models for real-time screening of thermoelectric materials properties," *APL Materials*, vol. 4, no. 5, p. 053213, 2016.
19. T. D. Sparks, M. W. Gaultois, A. Oliynyk, J. Brgoch, and B. Meredig, "Data mining our way to the next generation of thermoelectrics," *Scripta Materialia*, vol. 111, pp. 10–15, 2016.
20. R. Seshadri and T. D. Sparks, "Perspective: interactive material property databases through aggregation of literature data," *APL Materials*, vol. 4, no. 5, p. 053206, 2016.
21. Z. Pan, H. Sun, Y. Zhang, and C. Chen, "Harder than diamond: superior indentation strength of wurtzite bn and lonsdaleite," *Physical Review Letters*, vol. 102, no. 5, p. 055503, 2009.
22. A. M. Tehrani, L. Ghadbeigi, J. Brgoch, and T. D. Sparks, "Balancing mechanical properties

and sustainability in the search for superhard materials," *Integrating Materials and Manufacturing Innovation*, vol. 6, no. 1, pp. 1–8, 2017.
23. H.-Y. Chung, M. B. Weinberger, J. B. Levine, A. Kavner, J.-M. Yang, S. H. Tolbert, and R. B. Kaner, "Synthesis of ultra-incompressible superhard rhenium diboride at ambient pressure," *Science*, vol. 316, no. 5823, pp. 436–439, 2007.
24. J. B. Levine, S. H. Tolbert, and R. B. Kaner, "Advancements in the search for superhard ultra-incompressible metal borides," *Advanced Functional Materials*, vol. 19, no. 22, pp. 3519–3533, 2009.
25. A. Mansouri Tehrani, A. O. Oliynyk, M. Parry, Z. Rizvi, S. Couper, F. Lin, L. Miyagi, T. D. Sparks, and J. Brgoch, "Machine learning directed search for ultraincompressible, superhard materials," *Journal of the American Chemical Society*, vol. 140, no. 31, pp. 9844–9853, 2018.
26. A. Mansouri Tehrani and J. Brgoch, "Hard and superhard materials: A computational perspective," *Journal of Solid State Chemistry*, vol. 271, pp. 47–58, mar 2019.
27. X.-Q. Chen, H. Niu, D. Li, and Y. Li, "Modeling hardness of polycrystalline materials and bulk metallic glasses," *Intermetallics*, vol. 19, no. 9, pp. 1275–1281, 2011.
28. M. Parry, S. Couper, A. M. Tehrani, A. O. Oliynyk, J. Brgoch, L. Miyagi, and T. D. Sparks, "Lattice strain and texture analysis of superhard mo 0.9 w 1.1 bc and rewc 0.8 via diamond anvil cell deformation," *Journal of Materials Chemistry A*, 2019.
29. A. M. Tehrani, A. Lim, and J. Brgoch, "Mechanism for unconventional nonlinear elasticity," *Physical Review B*, vol. 100, no. 6, p. 060102, 2019.
30. A. Mansouri Tehrani, A. O. Oliynyk, Z. Rizvi, S. Lotfi, M. Parry, T. D. Sparks, and J. Brgoch, "Atomic substitution to balance hardness, ductility, and sustainability in molybdenum tungsten borocarbide," *Chemistry of Materials*, vol. 31, no. 18, pp. 7696–7703, 2019.
31. H. Fujii, M. DJC, and B. HKDH, "Bayesian neural network analysis of fatigue crack growth rate in nickel base superalloys," *Isij International*, vol. 36, no. 11, pp. 1373–1382, 1996.
32. J.-Y. Kang and J.-H. Song, "Neural network applications in determining the fatigue crack opening load," *International journal of fatigue*, vol. 20, no. 1, pp. 57–69, 1998.
33. S. Mohanty, R. Teale, A. Chattopadhyay, P. Peralta, and C. Willhauck, "Mixed gaussian process and state-space approach for fatigue crack growth prediction," in *International workshop on structural heath monitoring*, vol. 2, pp. 1108–1115, Citeseer, 2007.
34. H. Wang, W. Zhang, F. Sun, and W. Zhang, "A comparison study of machine learning based algorithms for fatigue crack growth calculation," *Materials*, vol. 10, no. 5, p. 543, 2017.
35. S. Dinda and D. Kujawski, "Correlation and prediction of fatigue crack growth for different r-ratios using kmax and δk+ parameters," *Engineering Fracture Mechanics*, vol. 71, no. 12, pp. 1779–1790, 2004.
36. O. E. Canyurt, "Estimation of welded joint strength using genetic algorithm approach," *International Journal of Mechanical Sciences*, vol. 47, no. 8, pp. 1249–1261, 2005.
37. A. Agrawal, P. D. Deshpande, A. Cecen, G. P. Basavarsu, A. N. Choudhary, and S. R. Kalidindi, "Exploration of data science techniques to predict fatigue strength of steel from composition and processing parameters," *Integrating Materials and Manufacturing Innovation*, vol. 3, no. 1, p. 8, 2014.
38. A. Agrawal and A. Choudhary, "A fatigue strength predictor for steels using ensemble data mining: Steel fatigue strength predictor," in *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, pp. 2497–2500, ACM, 2016.
39. D. T. Fullwood, S. R. Niezgoda, and S. R. Kalidindi, "Microstructure reconstructions from 2-point statistics using phase-recovery algorithms," *Acta Materialia*, vol. 56, no. 5, pp. 942–948, 2008.
40. S. R. Kalidindi, S. R. Niezgoda, and A. A. Salem, "Microstructure informatics using higher-order statistics and efficient data-mining protocols," *Jom*, vol. 63, no. 4, pp. 34–41, 2011.
41. S. Niezgoda, D. Fullwood, and S. Kalidindi, "Delineation of the space of 2-point correlations in a composite material system," *Acta Materialia*, vol. 56, no. 18, pp. 5285–5292, 2008.

42. V. Sundararaghavan and N. Zabaras, "Classification and reconstruction of three-dimensional microstructures using support vector machines," *Computational Materials Science*, vol. 32, no. 2, pp. 223–239, 2005.
43. H. Xu, R. Liu, A. Choudhary, and W. Chen, "A machine learning-based design representation method for designing heterogeneous microstructures," *Journal of Mechanical Design*, vol. 137, no. 5, p. 051403, 2015.
44. B. L. DeCost and E. A. Holm, "A computer vision approach for automated analysis and classification of microstructural image data," *Computational materials science*, vol. 110, pp. 126–133, 2015.
45. R. Liu, A. Kumar, Z. Chen, A. Agrawal, V. Sundararaghavan, and A. Choudhary, "A predictive machine learning approach for microstructure optimization and materials design," *Scientific reports*, vol. 5, p. 11551, 2015.
46. B. L. DeCost, H. Jain, A. D. Rollett, and E. A. Holm, "Computer vision and machine learning for autonomous characterization of am powder feedstocks," *Jom*, vol. 69, no. 3, pp. 456–465, 2017.
47. R. Cang and M. Y. Ren, "Deep network-based feature extraction and reconstruction of complex material microstructures," in *ASME 2016 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, pp. V02BT03A008–V02BT03A008, American Society of Mechanical Engineers, 2016.
48. A. Chowdhury, E. Kautz, B. Yener, and D. Lewis, "Image driven machine learning methods for microstructure recognition," *Computational Materials Science*, vol. 123, pp. 176–187, 2016.
49. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems 25* (F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, eds.), pp. 1097–1105, Curran Associates, Inc., 2012.
50. N. Lubbers, T. Lookman, and K. Barros, "Inferring low-dimensional microstructure representations using convolutional neural networks," *Physical Review E*, vol. 96, no. 5, p. 052111, 2017.
51. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
52. A. R. Kitahara and E. A. Holm, "Microstructure cluster analysis with transfer learning and unsupervised learning," *Integrating Materials and Manufacturing Innovation*, vol. 7, no. 3, pp. 148–156, 2018.
53. A. Cecen, H. Dai, Y. C. Yabansu, S. R. Kalidindi, and L. Song, "Material structure-property linkages using three-dimensional convolutional neural networks," *Acta Materialia*, vol. 146, pp. 76–84, 2018.
54. B. Cantor, I. Chang, P. Knight, and A. Vincent, "Microstructural development in equiatomic multicomponent alloys," *Materials Science Engineering A*, vol. 375-377, no. 1-2, pp. 213–218, 2004.
55. J. . Yeh, S. . Chen, S. . Lin, J. . Gan, T. . Chin, T. . Shun, C. . Tsau, and S. . Chang, "Nanostructured highentropy alloys with multiple principal elements: Novel alloy design concepts and outcomes," *Advanced Engineering Materials*, vol. 6, no. 5, pp. 299–303, 2004.
56. J. Davis, "Metals handbook desk edition 2 nd," *Materials Park, OH: American Society for Metals*, 1998.
57. S. Gorsse, J.-P. Couzini, and D. B. Miracle, "From high-entropy alloys to complex concentrated alloys," *Comptes rendus - Physique*, vol. 19, no. 8, pp. 721–736, 2018.
58. B. Gludovatz, A. Hohenwarter, K. V. S. Thurston, H. Bei, Z. Wu, E. P. George, and R. O. Ritchie, "Exceptional damage-tolerance of a medium-entropy alloy crconi at cryogenic temperatures," *Nature Communications*, vol. 7, no. 1, 2016.
59. S.-Q. Xia, Z. Wang, T.-F. Yang, and Y. Zhang, "Irradiation behavior in high entropy alloys," *Journal of Iron and Steel Research International*, vol. 22, no. 10, pp. 879–884, 2015.

60. J.-W. Yeh, "Alloy design strategies and future trends in high-entropy alloys," *JOM*, vol. 65, no. 12, pp. 1759–1771, 2013.
61. M.-H. Tsai and J.-W. Yeh, "High-entropy alloys: A critical review," *Materials Research Letters*, vol. 2, no. 3, pp. 107–123, 2014.
62. D. Miracle, J. Miller, O. Senkov, C. Woodward, M. Uchic, and J. Tiley, "Exploration and development of high entropy alloys for structural applications," *Entropy*, vol. 16, no. 1, pp. 494–525, 2014.
63. D. Miracle and O. Senkov, "A critical review of high entropy alloys and related concepts," *Acta Materialia*, vol. 122, pp. 448–511, 2017.
64. O. N. Senkov, J. D. Miller, D. B. Miracle, and C. Woodward, "Accelerated exploration of multi-principal element alloys with solid solution phases," *Nature communications*, vol. 6, no. 1, p. 6529, 2015.
65. O. Senkov, J. Miller, D. Miracle, and C. Woodward, "Accelerated exploration of multi-principal element alloys for structural applications," *Calphad*, vol. 50, pp. 32 – 48, 2015.
66. M. Troparevsky, J. Morris, P. Kent, A. Lupini, and G. Stocks, "Criteria for predicting the formation of single-phase high-entropy alloys," *Physical Review X*, vol. 5, no. 1, 2015.
67. M. C. Gao, B. Zhang, S. Yang, and S. M. Guo, "Senary refractory high-entropy alloy hfnbtativzr," *Metallurgical and Materials Transactions A*, vol. 47, pp. 3333–3345, Jul 2016.
68. I. Toda-Caraballo, J. Wrbel, D. Nguyen-Manh, P. Prez, and P. Rivera-Daz-del Castillo, "Simulation and modeling in high entropy alloys," *JOM*, vol. 69, no. 11, pp. 2137–2149, 2017.
69. N. Islam, W. Huang, and H. L. Zhuang, "Machine learning for phase selection in multi-principal element alloys," *Computational Materials Science*, vol. 150, pp. 230–235, 2018.
70. S. Guo, C. Ng, J. Lu, and C. T. Liu, "Effect of valence electron concentration on stability of fcc or bcc phase in high entropy alloys," *Journal of Applied Physics*, vol. 109, no. 10, p. 103505, 2011.
71. W. Huang, P. Martin, and H. L. Zhuang, "Machine-learning phase prediction of high-entropy alloys," *Acta Materialia*, vol. 169, pp. 225–236, 2019.
72. Y.-J. Chang, C.-Y. Jui, W.-J. Lee, and A.-C. Yeh, "Prediction of the composition and hardness of high-entropy alloys by machine learning," *JOM*, vol. 71, no. 10, p. lt;xocs:firstpage xmlns:xocs=quot;quot;/gt;, 2019.
73. J. He, W. Liu, H. Wang, Y. Wu, X. Liu, T. Nieh, and Z. Lu, "Effects of al addition on structural evolution and tensile properties of the feconicrmn high-entropy alloy system," *Acta Materialia*, vol. 62, no. 1, pp. 105–113, 2014.
74. C. Wen, Y. Zhang, C. Wang, D. Xue, Y. Bai, S. Antonov, L. Dai, T. Lookman, and Y. Su, "Machine learning assisted design of high entropy alloys with desired property," *Acta Materialia*, vol. 170, pp. 109–117, 2019.
75. J. Rickman, H. Chan, M. Harmer, J. Smeltzer, C. Marvel, A. Roy, and G. Balasubramanian, "Materials informatics for the screening of multi-principal elements and high-entropy alloys," *Nature Communications*, vol. 10, no. 1, 2019.
76. R. Chen, G. Qin, H. Zheng, L. Wang, Y. Su, Y. Chiu, H. Ding, J. Guo, and H. Fu, "Composition design of high entropy alloys using the valence electron concentration to balance strength and ductility," *Acta Materialia*, vol. 144, pp. 129–137, 2018.
77. E. Menou, I. Toda-Caraballo, P. E. J. R.-D. del Castillo, C. Pineau, E. Bertrand, G. Ramstein, and F. Tancret, "Evolutionary design of strong and stable high entropy alloys using multi-objective optimisation based on physical models, statisticsandthermodynamics," *Materials Design*, vol. 143, pp. 185 – 195, 2018.
78. E. Menou, F. Tancret, I. Toda-Caraballo, G. Ramstein, P. Castany, E. Bertrand, N. Gautier, and P. E. J. Rivera Daz-Del-Castillo, "Computational design of light and strong high entropy alloys (hea): Obtainment of an extremely high specific solid solution hardening," *Scripta Materialia*, vol. 156, pp. 120–123, 2018.
79. J. F. Löffler, "Bulk metallic glasses," *Intermetallics*, vol. 11, pp. 529–540, jun 2003.

80. M. M. Trexler and N. N. Thadhani, "Mechanical properties of bulk metallic glasses," *Progress in Materials Science*, vol. 55, pp. 759–839, nov 2010.
81. M. Ashby and A. Greer, "Metallic glasses as structural materials," *Scripta Materialia*, vol. 54, pp. 321–326, feb 2006.
82. J. J. Si, T. Wang, Y. D. Wu, Y. H. Cai, X. H. Chen, W. Y. Wang, Z. K. Liu, and X. D. Hui, "Cr-based bulk metallic glasses with ultrahigh hardness," *Applied Physics Letters*, vol. 106, p. 251905, jun 2015.
83. D. Pan, A. Inoue, T. Sakurai, and M. W. Chen, "Experimental characterization of shear transformation zones for plastic flow of bulk metallic glasses.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, pp. 14769–72, sep 2008.
84. X.-Q. Chen, H. Niu, D. Li, and Y. Li, "Modeling hardness of polycrystalline materials and bulk metallic glasses," *Intermetallics*, vol. 19, pp. 1275–1281, sep 2011.
85. E. Perim, D. Lee, Y. Liu, C. Toher, P. Gong, Y. Li, W. N. Simmons, O. Levy, J. J. Vlassak, J. Schroers, and S. Curtarolo, "Spectral descriptors for bulk metallic glasses based on the thermodynamics of competing crystalline phases," *Nature Communications*, vol. 7, p. 12315, nov 2016.
86. M. K. Tripathi, S. Ganguly, P. Dey, and P. P. Chattopadhyay, "Evolution of glass forming ability indicator by genetic programming," *Computational Materials Science*, vol. 118, pp. 56–65, jun 2016.
87. P. Tsai and K. M. Flores, "High-throughput discovery and characterization of multicomponent bulk metallic glass alloys," *Acta Materialia*, vol. 120, pp. 426–434, may 2016.
88. M. Chen, A. Inoue, W. Zhang, and T. Sakurai, "Extraordinary Plasticity of Ductile Bulk Metallic Glasses," *Physical Review Letters*, vol. 96, p. 245502, jun 2006.
89. D. C. Hofmann, J.-Y. Suh, A. Wiest, M.-L. Lind, M. D. Demetriou, and W. L. Johnson, "Development of tough, low-density titanium-based bulk metallic glass matrix composites with tensile ductility," *Proceedings of the National Academy of Sciences*, vol. 105, pp. 20136–20140, dec 2008.
90. L. M. Wang, Y. Tian, R. Liu, and W. Wang, "A "universal" criterion for metallic glass formation," *Applied Physics Letters*, vol. 100, p. 261913, jun 2012.
91. M. K. Tripathi, P. P. Chattopadhyay, and S. Ganguly, "A predictable glass forming ability expression by statistical learning and evolutionary intelligence," *Intermetallics*, vol. 90, pp. 9–15, nov 2017.
92. M. K. Tripathi, P. P. Chattopadhyay, and S. Ganguly, "Multivariate analysis and classification of bulk metallic glasses using principal component analysis," *Computational Materials Science*, vol. 107, pp. 79–87, sep 2015.
93. L. Ward, A. Agrawal, A. Choudhary, and C. Wolverton, "A general-purpose machine learning framework for predicting properties of inorganic materials," *npj Computational Materials*, vol. 2, p. 16028, nov 2016.
94. Y. T. Sun, H. Y. Bai, M. Z. Li, and W. H. Wang, "Machine Learning Approach for Prediction and Understanding of Glass-Forming Ability," *Journal of Physical Chemistry Letters*, vol. 8, pp. 3434–3439, jul 2017.
95. F. Ren, L. Ward, T. Williams, K. J. Laws, C. Wolverton, J. Hattrick-Simpers, and A. Mehta, "Accelerated discovery of metallic glasses through iteration of machine learning and high-throughput experiments," *Science Advances*, vol. 4, p. eaaq1566, apr 2018.
96. A. Dasgupta, S. R. Broderick, C. Mack, B. U. Kota, R. Subramanian, S. Setlur, V. Govindaraju, and K. Rajan, "Probabilistic Assessment of Glass Forming Ability Rules for Metallic Glasses Aided by Automated Analysis of Phase Diagrams," *Scientific Reports*, vol. 9, p. 357, dec 2019.
97. 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 Materialia*, vol. 159, pp. 102–111, oct 2018.

98. Q. Wang and A. Jain, "Linking plastic heterogeneity of bulk metallic glasses to quench-in structural defects with machine learning," pp. 1–35, apr 2019.
99. T. Mueller, A. G. Kusne, and R. Ramprasad, "Machine learning in materials science: Recent progress and emerging applications," *Reviews in Computational Chemistry*, vol. 29, pp. 186–273, 2016.
100. R. Ramprasad, R. Batra, G. Pilania, A. Mannodi-Kanakkithodi, and C. Kim, "Machine learning in materials informatics: recent applications and prospects," *npj Computational Materials*, vol. 3, no. 1, pp. 1–13, 2017.