ISSN: 2321-2152

IJMECE

International Journal of modern electronics and communication engineering

E-Mail
editor.ijmece@gmail.com
editor@ijmece.com

www.ijmece.com

ISSN2321-2152
www.ijmece .com
Vol 11, Issue.4 Oct 2023

MACHINE LEARNING OF MECHANICAL PROPERTIES OF STEELS

1. CHINTA BHARATH,2. VENGALA UPENDER,3. BOMMU SATISH REDDY

ABSTRACT

Knowledge of the mechanical properties of structural materials is essential for their practical applications. In the present work, three-hundred and sixty data samples on four mechanical properties of steels—fatigue strength, tensile strength, fracture strength and hardness—were selected from the Japan National Institute of Material Science database, comprising data on carbon steels and low-alloy steels. Five machine learning algorithms were used to predict the mechanical properties of the materials represented by the three-hundred and sixty data samples, and random forest regression showed the best predictive performance. Feature selection conducted by random forest and symbolic regressions revealed the four most important features that most influence the mechanical properties of steels: the tempering temperature of steel, and the alloying elements of carbon, chromium and molybdenum. Mathematical expressions were generated via symbolic regression, and the expressions explicitly predicted how each of the four mechanical properties varied quantitatively with the four most important features. This study demonstrates the great potential of symbolic regression in the discovery of novel advanced materials.

INTRODUCTION

The identification of structure-property relationships is fundamental to the discovery of new materials. However, the ability to comprehensively understand and manipulate structure-property relationships of materials is very challenging, due to the diversity and complexity of materials. As a result, data-driven discovery of novel advanced materials requires the use of advanced techniques such as big data and artificial intelligence, data mining and machine learning (ML) to accelerate research and development [1–8]. Materials 2 data and ML provide the foundation of this data-driven materials discovery paradigm, which integrates materials domain knowledge and artificial intelligence technology to form the new research field of materials informatics. In this new field, the Materials Genome Initiative aims to halve the cost and time from discovery to development to deployment of advanced materials [9]. This integrated approach applies materials data to explore structure-property relationships and to develop models and guidance for synthesis of new materials. For example, Homer et al. [10] and Zhu et al. [11] used ML tools to investigate grain boundaries in polycrystalline materials, and Raccuglia et al. [12] demonstrated a ML strategy to elucidate how to classify successful and failed synthesis conditions with the use of historically accumulated experimental data. Agrawal et al. [13,14] used ML algorithms to predict the fatigue strength of steels, which substantially improved the understanding of fatigue behavior. However, their ML predictions did not result in explicit mathematic expressions linking features and output properties, which are desirable for materials research, design, development and deployment. The purpose of this study was to predict the four mechanical properties of steels using five ML algorithms, especially using random forest (RF) regression and symbolic regression (SR). The performances of the five algorithms were assessed, revealing that RF performing the best, and explicit mathematical expressions were obtained from SR.

1.PG SCHOLAR.2. ASSOSCIATE PROFESSOR &HOD,3. ASSOSCAITE PROFESSOR
DEPARTMENT OF MECHANICAL ENGINEERING, ABDUL KALAM INSTITUTE OF TECHNOLOGY & SCIENCE, KOTHAGUDEM, TELANGANA 507101

The production of HSS is different from that of other steel types in two main aspects, the steel chemical composition and the heat treatment process that are adopted to achieve their higher strengths. There are four main processing routes for the production of high strength steel plates, namely: Thermo-Mechanical (TM) rolling, Thermo-Mechanical rolling with Accelerated Cooling (TM-ACC), Thermo-Mechanical rolling with Direct Quenching (TM-DQ) and Conventional route for QT plates . The different chemical composition and heat treatment processes employed in the production of HSS by these routes lead to material stress–strain behaviours that are different from those of conventional mild steels at both room and elevated temperatures. At room temperature, the ductility and the ratio of the ultimate-to-yield strength of HSS reduce with increasing strength. At elevated temperatures, high strength steels, like all steels, lose strength and stiffness with increasing temperatures, but the rate of degradation is influenced by the chemical composition and the heat treatment process.

The microstructure of mild steel can be either pearlite, bainite, or combination of both depending on the rate of cooling of austenite steel . Bainite and pearlite transformations are stable phases, and do not transform to other phases without reheating to form austenite (at about 900 °C). On the other hand, the microstructure of HSS is martensite which forms when the austenite is rapidly cooled, or quenched, to a relatively low temperature in order to prevent carbon diffusion and hence the formation of pearlite or bainite. Martensite is not an equilibrium phase because the quenching process prevents diffusion of carbon out of the austenite structure. It tends to transform to stable ferrite and cementite phases when it is heated to a temperature (higher than 250 °C) that initiates diffusion of trapped carbon in the lattice . The strength of HSS is controlled by the amount of developed cementite which is primarily a function of the chemical composition of the steel, the attained temperature and the duration of fire. Therefore, a combination of all of these factors, in particular the effect of chemical composition, need to be considered when predicting the elevated temperature mechanical properties of HSS materials.

LITERATURE REVIEW:

Steel contains important mechanical properties like tensile strength, yield strength and elongation. Oneof the basic traditional test conducted is uniaxial tensile test which is done for many reasons. In engineering applications, tests of tensile are used to select materials. To assure quality in the materials tensile properties are used. When making strides, modern materials estimation of ductile properties is included so that unmistakable materials can be related. In steel material, the protection across the load is a function of a cross section and mechanical properties. To figure out the mechanical properties of steel like tensile strength, yield strength, and elongation tensile test is performed.Yield Strength of a substance gives the stress when deformation exceeds the limit of plastic. Yield Strength is permanent when deformation is higher and results to stress. Yield Strength of a material is measured in Pascal.The results obtained are plotted on a curve called stress-strain curve.To identify the points from this curve is bit difficult. The Yield Strength of the material is identified at the point where stress is deviated from original point of the curve. Tensile strength is the highest point plotted on the stress-strain curve after the test has been performed.If the temperature varies then the tensile strength of a material varies proportion to it. Elongation is the point to which a substance may be developed or shorten before it shatter.It plays an important role during the manufacturing process and measures the amount of bending and shaping a material without any breaks.In traditional process manpower and time required are more. So in this project the proposed method integrates with the machine learning algorithms which reduces manpower, time and improves the efficiency. 4. PROPOSED METHOD The proposed approach will remove lot of manpower and time and finds a better way for prediction of steel mechanical properties using the machine learning algorithms. Machine Learning algorithms are combined with the material sciences. Here algorithms like Random Forest, Decision Trees, Naive Bayes, and Logistic Regression are used. The dataset required for this research is collected from the standards resources. In this paper, different standards are taken into consideration along with the carbon content, sectional size and temperature. In Machine Learning algorithms supervised methods are used for prediction as it can be trained with both input and output values. This approach will give better results to predict tensile strength, yield strength and elongation of steel with different standards. Figure1:System Architecture for Prediction Data Collection: The “Steel Prediction” dataset is collected from various sources and merged together with different parameters. It consists of seven attributes namely standards, carbon content, thickness, temperature, tensile strength, yield strength,

elongation. In the above attributes four are independent variables (standards, carbon content, thickness, temperature) and three are dependent variables(tensile strength, yield strength, elongation). Data Preprocessing: The second stage after collection of data is the preprocessing of data. The missing values are handled using the nan function. After filling the missing values split the dataset into two categories training set and testing set.In training set the algorithm will be able to learn the behavior of the system and predicts the output using testing samples. It is the process of preparing data for analysis by removing data that is incorrect, incomplete, duplicate, and irrelevant and it also includes standardizing dataset by correcting mistakes such as empty fields, missing values using. After cleaning the dataset validate the accuracy.

EXISTING SYSTEM:

The mechanical properties of steels include various characteristics such as yield strength, tensile strength, hardness, fatigue resistance, and fracture toughness. These properties are essential for designing materials that can withstand different loads, temperatures, and environmental conditions.To apply machine learning to steel properties, researchers gather extensive datasets that include information about the composition of steel alloys, the manufacturing process, and the results of mechanical tests conducted on the materials. This data serves as the foundation for building predictive mode. Feature engineering involves selecting and transforming relevant data attributes (features) that will be used as input variables for the machine learning models. These features can include the chemical composition of the steel, heat treatment parameters, microstructure details, and more. Researchers use various machine learning algorithms, such as linear regression, decision trees, random forests, support vector machines, and neural networks, to develop predictive models. These models learn to map the input features to the desired mechanical properties. The models are trained on a portion of the dataset, and their performance is evaluated using validation techniques to ensure they can make accurate predictions. Cross-validation and hold-out validation are common methods. Once trained and validated, the machine learning models can predict the mechanical properties of steel materials based on input data. This capability is valuable for material scientists, engineers, and industries that rely on steel for various applications. Beyond prediction, machine learning can assist in optimizing steel compositions and manufacturing processes to achieve desired mechanical properties. This is particularly valuable for developing advanced materials tailored to specific applications. Challenges in machine learning of steel properties include data quality, interpretability of models, and ethical considerations regarding the use of AI in material design and engineering. Ensuring model transparency and fairness . The application of machine learning in predicting steel properties has a profound impact on industries like automotive, aerospace, construction, and manufacturing. It accelerates materials research, reduces development costs, and enhances the performance and sustainability of steel-based products.

In summary, "Machine Learning of Mechanical Properties of Steels" represents a cutting-edge approach to materials science and engineering. It leverages the power of AI and data-driven insights to improve our understanding of steel behavior, optimize material design, and drive innovation across a wide range of industries.

PROPOSED SYSTEM:

A. CHEMICAL COMPOSITION OF HOT ROLLED ALLOY STEEL:

Most of the chemical composition of alloy steel consist of iron (Fe), carbon (C), manganese (Mn), silicon (Si), phosphorus (P) and sulfur (S). After reheating, roughing rolling, finishing rolling, laminar cooling and down coiler, a steel slab becomes a coil of a thin sheet. The mechanical properties can be adjusted with different chemical composition or process parameters. Complex interactions exist between the chemical composition and mechanical properties. For example, when the ratio of C in the steel is below 0.8Wt%, the YS and TS of the steel increase dramatically with the increases of C content, but the EL of steel decreases [24]. Furthermore, the content of S and P must be strictly controlled. S will reduce the hot workability and strength of steel and P will reduce the plasticity and toughness of the steel [25]. Mn has a strong ability of deoxidation and desulfurization, can greatly improve the hot workability and strength of steel. In steel production, molten iron is often mixed with oxygen (O2) and nitrogen (N2), which is detrimental to the mechanical properties of steel. For this reason, Mn, Si, aluminum (Al), vanadium (V), titanium (Ti) are often added as deoxidants to the steel [10]. The combination of Si, molybdenum (Mo), and chromium (Cr) not only makes the steel highly resistant to oxidation and corrosion but also enhances the strength and hardness of the steel. The addition of Al, niobium (Nb), and V to the steel can reduce the negative effects of N. The copper (Cu) in steel can improve the strength and

toughness and resistance to atmospheric corrosion of steel. Boron(B) can improve the compactness and hot rolling properties of steel. Nickel (Ni) not only can significantly increase the strength and toughness of steel, Ni and Cr are also the main alloying elements of stainless steel [26].

B. HOT ROLLED PROCESSING OF ALLOY STEEL:

The hot rolled processing of steel slab can be divided into five steps, reheating, roughing rolling, finishing rolling, laminar cooling, and down coiling, Firstly, the steel slab is reheated in furnace to a high temperature about 1100◦ to 1250◦ . Secondly, the hot steel slab is transfer to edger mill and roughing mill to reduce the width and thickness. The steel slab would become longer and thinner. Thirdly, the steel slab will pass through the finishing mill to control the thickness with high precision. Fourthly, the steel strip will pass through the laminar cooling region to reduce the temperature quickly. Finally, the steel strip is coiled by the down coiler. In fact, a series of complex microstructure changes occur in the manufacturing processes, which could determine the mechanical properties of the alloy steel. First, the reheating process provides a uniform temperature to the slab to provide a uniform initial austenite grain size Then, the roughing and finishing processes refine austenite by dynamic and static recrystallization. Furthermore, the steel sheet is continuously cooled by the laminar cooling system for refining transformed ferrite and pearlite grain [27]. The size and volume fraction of these grains determine the mechanical properties of the steel. These heat treatment temperatures play an important role in mechanical properties prediction. Therefore, the furnace temperature (FT), the roughing rolling temperature (RRT), the finishing rolling temperature (FRT) and the coiling temperature (CT) have a significant influence on the mechanical properties of hot rolled steel.

C. MECHANICAL PROPERTIES OF HOT ROLLED ALLOY STEEL :

The mechanical properties of alloy steel are TS, YS, and EL. TS is defined as the maximum tensile stress that the steel can withstand before breaking, and the YS is defined as the maximum stress of the steel can withstand before plastic deformation begins. EL is defined as the percentage of stretched length to the original length after the steel is broken. Fig. 2 shows the relationship between TS, YS, and EL of alloy steel. III. PROPOSED APPROACH This section describes the proposed prediction method based on CNN for mechanical properties of hot rolled alloy steel.

D. PROPOSED CNN METHOD FOR STEEL MECHANICAL PROPERTIES PREDICTION

The proposed CNN-based prediction model is composed of two parts, feature extraction part and prediction part. As shown in Fig. 4, the feature extraction part is composed of an input layer and several feature extractors. The feature extractor is stacked by convolutional layer, batch normalization (BN) operation, nonlinear activation layer, and pooling layer. Each feature extractor will extract its input features to get a feature map which will become the input data of the subsequent feature extractor. The prediction part contains two fully connected (FC) layers and an output layer. The feature maps output by the last feature extractor will be transferred to the FC layers and perform the prediction task. Finally, the predicted steel mechanical properties value will be output by the output layer. The input layer receives the converted two-dimension data image, and the convolutional layer uses the convolution kernel which consists of a weight matrix with the same size as the receptive field on the input layer and a bias value to establish a local connection. As the receptive field slides from the top left of the input layer to the bottom right, the convolutional layer obtains the feature map of the input layer filtered by the convolution kernel, which can be expressed as Equation (3). $y_{jk} = \left(\sum_{p=1}^{F} \sum_{q=1}^{F} w_{pq} x_{(p+j*s)(q+k*s)} \right) + b), 0 \le j \le H - F S , 0 \le k \le W - F S$ (3) y_{jk} represents the value of the node positioned at (j, k) on the feature map, F represents the height and width of the receptive field, H and W represent the height and width of the input data, S indicates the stride of the receptive field. $x_{(p+j*s)(q+k*s)}$ represents the input data with the coordinate at $(p + j * s, q + k * s)$, w_{pq} and b denote the weight located at (p, q) on the weight matrix and the bias, respectively. To obtain sufficient characteristics of the mechanical properties, a set of convolutional kernels are used to perform the convolution operation. By the convolution operation, each convolution kernel can get a feature map, and different nodes on the feature map correspond to different receptive fields when the convolution kernel sweeps across the input layer. The connection pattern that each node on the same feature map connects to its receptive field by the same convolution kernel is called parameter sharing. The local connection and the parameter sharing are two important characteristics of CNN, which can reduce the number of parameters, extract the features in raw data

effectively, and enhance the generalization ability of the model [28]. For the lth convolutional layer with Pl convolutional kernels, the output can be denoted as Equation (4). u i l = XPl−1 k=1 Wi l x k l + b i l , i ∈ [1, Pl] (4) x k l and u i l represent the input matrix and output matrix of the lth layer respectively, where i and k denote the channel index in the lth convolutional layer and the l-1th convolutional layer, respectively. The weight matrix and bias contained in the kth convolution kernel of the lth convolutional layer are denoted by Wi l and b i l respectively. The size of the feature map will shrink after convolution operation. The zeropadding method is applied to keep the size of the output feature map, which centers the output feature map and adds zero values at all the edges of the output feature map. BN operation is added after each convolutional layer to improve training speed and achieve higher prediction accuracy, which allows us to be less careful about the initialization method and use much higher learning rates [29]. First, each dimension of the input is normalized into a stable distribution with the mean of zero and variance of one. Then, the normalized value is scaled and shifted by a pair of learnable variables γ and β to restore the data distribution that should be learned in the previous layer. The BN operation can be expressed as Equation (5). y (k) = γ (k) x (k) − E x (k) q Var x (k) + + β (k) , k ∈ [1,N] (5) VOLUME 7, 2019 470 Z.-W. Xu et al.: Mechanical Properties Prediction for Hot Rolled Alloy Steel Using CNN where N represents the total dimension of the input, x (k) and y (k) represent the kth dimension of the input and output of the layer respectively, E(x (k)) and Var x (k) represent the mean value and variance of the kth dimension of the input respectively. A very small real number is added, to avoid the denominator is zero. The learnable variables for scaling and shifting the normalized value are represented by γ (k) and β (k) respectively. Rectified linear unit (ReLU) is applied to the proposed CNN model as the nonlinear activation function, which can prevent the vanishing gradient and exploding gradient problems in the neural network and enhance the training speed [30]. Let max denote the function to select the larger value between x and zero, the ReLU activation function can be indicated as Equation (6). ReLU(x) = max(x, 0) (6) In the pooling layer, downsampling is applied to scale down and concentrate feature map to obtain the most significant features in the input feature map. The max-pooling method is used as the pooling method by selecting the maximum value in the pooling field. The information of the concentrated feature maps obtained from the final pooling layer is transmitted to the prediction part by flattening the condensed feature maps into a dense vector. Each node on the first FC layer is connected to the dense vector and the output is passed to the second FC layer and finally transmitted to the output layer. The dropout [31] which can effectively avoid the proposed model from overfitting by randomly selecting some nodes to skip weight updates in each iteration of training is used at all the FC layers to enhance the generalization ability of the model. The predicted value is obtained by using feature extractions to gain the features from raw data and applying FC layers to process the feature information. The model output can be expressed as Equation (7). yˆ = wfc2σ(wfc1fl(pool(σ(bn(XPL−1 k=1 Wi L x k L + b i L)))) + bfc1) + bfc2 (7) where the weight and the bias of the first FC layer are denoted by wfc1 and bfc1 respectively, wfc2 and bfc2 are the weight and the bias of the second FC layer, σ represents the nonlinear activation function ReLU, fl denotes the flattening operation which flattens features into a dense vector, pool denotes the max-pooling method, bn denotes the BN operation, L represents the total number of convolutional layers, PL−1 denotes the number of convolutional kernels in the L-1th convolutional layer, the channel index of the Lth and L-1th convolutional layer are denoted by i and k respectively and x k L represents the input of the Lth layer. The mean squared error (MSE) is applied as the loss function to measure the distance between predicted values and actual values. Minimizing MSE is taken as the training goal of the proposed model and the minimum MSE is achieved by continuously adjusting the weight and bias of each neuron. The MSE can be expressed as Equation (8). MSE = 1 M XM i=1 yˆi − yi 2 (8) where yˆi represents the predicted value, yi represents the actual value, and M represents the sample size in the data sets.

Results and Discussion

ML models with all features Four ML algorithms – RF, linear least-square (LLS), k-nearest neighbors (KNN) and architecture-neural network (ANN) – were conducted on the dataset comprising all 16 features (termed ‘All’). The performances of these algorithms were evaluated by ten-fold cross-validation, in which the data were divided into ten parts (nine parts for training data and one part for testing data) and the training and testing were cycled ten times to allow the use of all data in testing. The predictive power of an ML algorithm on the testing

data was measured by the correlation coefficient (R) and the relative root-mean-square errors (RRMSE), which are defined by 5 ( )( ) ∑ ( ) ∑ ( ) ∑ = = = − − − − = n i n i i i n i i i y y y y y y y R 1 1 2 2 1 ˆ ˆ ˆ ˆ , (1) ( ) y y y n RRMSE n i ∑ i i = − = 1 2 ˆ 1 (2) Figure 1. The R and RRMSE values of the RF (random forest), LLS (linear least-square), KNN (knearest neighbors) and ANN (architecture-neural network) models using all 16 features: (a) fatigue strength of all models, and the performance of the best model, ANN@All; (b) tensile strength and the performance of the best model, ANN@All; (c) fracture strength and the performance of the best model, RF@All; and (d) hardness and the performance of the best model ANN@All. 6 where n is the number of testing data, and y, y ˆ and y denote the actual value, the predicted value and the average value, respectively. R lies between 0 and 1, and a value of 1 indicates a perfect prediction. An RRMSE value of zero indicates a perfect fit. In general, a higher value for R and a lower value for RRMSE indicate a better ML algorithm [16]. Figure 1 shows the R and RRMSE values of the four ML algorithms and compares the best predicted values from one of the ML algorithms with the measured values for each of the four mechanical properties. As can be seen, the RF has the greatest predictive power for fracture strength (R = 0.9725, RRMSE = 23.56%), whilst the ANN algorithm gives the best results for fatigue strength (R = 0.9699, RRMSE = 24.49%), tensile strength (R = 0.9857, RRMSE = 16.89%) and hardness (R = 0.9836, RRMSE = 18.13%).

Feature selection

Feature selection is crucial in ML; given the fact that ML algorithms such as RF and SR have feature selection functions, these algorithms are emphasized here. The importance of the features computed by RF is denoted RFI, and that of the features computed by SR is called SRI. Figures 2 (a-b) show the RFI and SRI values, respectively, for each original feature. The RFI values indicate that the four most important features are the presence of Mo and Cr, the normalizing temperature and the tempering temperature, whilst the SRI values indicate that the four most important features are the tempering temperature, and the presence of C, Cr and Mo, which correspondingly yield two feature subsets of RFI (NT, TT, Cr, Mo) and SRI (TT, C, Cr, Mo). Figure 2. Normalized (a) random forest importance (RFI) and (b) symbolic regression importance (SRI) of the 16 features for fatigue strength, tensile strength, fracture strength, and hardness. 7 The four ML algorithms were conducted using the RFI features (NT, TT, Cr, Mo) and the SRI features (TT, C, Cr, Mo). Figure 3 shows the cross-validation R values and the predicted values of the best model against the measured value for each of the four features. The results illustrate that the RF algorithm with the feature subset SRI (TT, C, Cr, Mo) outperforms the other algorithms. The RF models with the feature subset SRI (TT, C, Cr, Mo) predict the four target properties with high predictive accuracy (R > 0.9550, RRMSE < 30.00%). Figure 3. R values of the RF (random forest), LLS (linear least-square), KNN (k-nearest neighbors) and ANN (architecture-neural network) models with the selected RFI and SRI feature subsets: (a) 8 for fatigue strength, and the performance of the best model, RF@SRI; (b) for tensile strength, and the performance of the best model, RF@SRI; (c) for fracture strength and the performance of the best model, RF@SRI; and (d) for hardness, and the performance of the best model RF@SRI.

Mathematical expressions

With SRI features (TT, C, Cr, Mo), SR gave the following mathematical expressions for fatigue strength (FaS) (MPa), tensile strength (TS) (MPa), fracture strength (FrS) (MPa), and hardness (H) (HV). FaS= -0.8685TT+316.7C+367.6Cr- 227.5Cr +708.6Mo +785.0 2 2 (3) TS= -1.827TT-119.7 C+643.2Cr-379.9Cr +1514Mo +2122 2 2 (4) FrS= -1.176TT- 46.12 C+695.4Cr- 415.3Cr +1461Mo +2267 2 2 (5) H = -0.5839TT-38.41 C+191.2Cr-113.3Cr +104.0Mo+681.9 2 (6) where all elements are expressed in wt.% and TT is expressed in (°C). The equations had strong predictive power (R > 0.9425, RRMSE < 33.30%), as shown in Figure 4. Equations (3-6) each include a minus sign with the tempering temperature, which indicates that lower tempering temperatures are should improve the strength and hardness of steels. The alloying elements of C, Cr and Mo are also strengthening elements.

CONCLUSION

The research here ensures that Machine Learning techniques will benefit in terms of accuracy and firmness for predicting the mechanical properties of steel metal. The dataset here is implemented with different machine learning algorithms like Random Forest, Decision Tree, Naive Bayes and Logistic Regression to achieve the good performance. Here 70% of the data is used for training and 30% data is used for testing.The results obtained are validated with the test data to ensure the results obtained are correct. The implemented algorithms gave better results for predicting tensile strength, yield strength and elongation of steel. The accuracy obtained here is more than 90% for different machine learning

algorithms. The work presented here can be future enhanced for different metals with their input processing parameters and can predict the mechanical properties of other metals. It plays an important role in many of the applications which removes lot of manpower and time.

REFERENCES

1 Ramprasad R, Batra R, Pilania G, et al. Machine learning in materials informatics: Recent applications and prospects. npj Comput Mater, 2017, 3: 54.

2 Xue D, Xue D, Yuan R, et al. An informatics approach to transformation temperatures of NiTi-based shape memory alloys. Acta Mater, 2017, 125: 532–541. 3 Ward L, Agrawal A, Choudhary A, et al. A general-purpose machine learning framework for predicting properties of inorganic materials. npj Comput Mater, 2016, 2: 1–7. 4 Senderowitz H, Barad H-N, Yosipof A, et al. Data Mining and Machine Learning Tools for Combinatorial Material Science of All-Oxide Photovoltaic Cells. Mol Inform, 2015, 34: 367–379.

5 Agrawal A, Choudhary AN. Perspective: Materials informatics and big data: Realization of the “fourth paradigm” of science in materials science. APL Mater, 2016, 4: 53208.

6 Xiong J, Shi SQ, Zhang TY. A machine-learning approach to predicting and understanding the properties of amorphous metallic alloys. Mater Des, 2020, 187: 108378

7 Takahashi K, Tanaka Y. Material synthesis and design from first principle calculations and machine learning. Comput Mater Sci, 2016, 112: 364–367.

8 Seshadri R, Wolverton C, Hill J, et al. Materials science with large-scale data and informatics: Unlocking new opportunities. MRS Bull, 2016, 41: 399–409.

9 Green ML, Choi CL, Hattrick-Simpers JR, et al. Fulfilling the promise of the materials genome initiative with high-throughput experimental methodologies. Appl Phys Rev, 2017, 4: 011105.

10 Huber L, Hadian R, Grabowski B, et al. A machine learning approach to model solute grain boundary segregation. npj Comput Mater, 2018, 4: 64.