Journal of Materials Science & Technology 149 (2023) 237–246

Contents lists available at ScienceDirect

Journal of Materials Science & Technology

journal homepage: www.elsevier.com/locate/jmst

Research Article

Discovering a formula for the high temperature oxidation behavior of FeCrAlCoNi based high entropy alloys by domain knowledge-guided machine learning

Qinghua Wei[a], Bin Cao[a], Lucheng Deng[a], Ankang Sun[a], Ziqiang Dong[a,*], Tong-Yi Zhang[a,b,*]

[a] Materials Genome Institute, Shanghai University, Shanghai 200444, China
[b] Hong Kong University of Science and Technology (Guangzhou), Guangzhou 511458, China

ARTICLE INFO

Article history:
Received 19 September 2022
Revised 3 November 2022
Accepted 11 November 2022
Available online 5 February 2023

Keywords:
Domain knowledge
Interpretive formula
High entropy alloys
Oxidation

ABSTRACT

A mathematical formula of high physical interpretation, and accurate prediction and large generalization power is highly desirable for science, technology and engineering. In this study, we performed a domain knowledge-guided machine learning to discover high interpretive formula describing the high-temperature oxidation behavior of FeCrAlCoNi-based high entropy alloys (HEAs). The domain knowledge suggests that the exposure time dependent and thermally activated oxidation behavior can be described by the synergy formula of power law multiplying Arrhenius equation. The pre-factor, time exponent (m), and activation energy (Q) are dependent on the chemical compositions of eight elements in the FeCrAlCoNi-based HEAs. The Tree-Classifier for Linear Regression (TCLR) algorithm utilizes the two experimental features of exposure time (t) and temperature (T) to extract the spectrums of activation energy (Q) and time exponent (m) from the complex and high dimensional feature space, which automatically gives the spectrum of pre-factor. The three spectrums are assembled by using the element features, which leads to a general and interpretive formula with high prediction accuracy of the determination coefficient R^2=0.971. The role of each chemical element in the high-temperature oxidation behavior is analytically illustrated in the three spectrums, thereby the discovered interpretative formula provides a guidance to the inverse design of HEAs against high-temperature oxidation. The present work demonstrates the significance of domain knowledge in the development of materials informatics.

1. Introduction

1.1. Research background

The fast development of Artificial Intelligence (AI) and Machine Learning (ML) [1–8] promotes greatly materials informatics, the data-driven paradigm in materials innovation, and materials science and engineering. In the past decade, machine learning techniques have been successfully employed for accelerating the discovery and design of new/improved materials. For example, Xiong et al. [6] found the critical values of some important features when studying the hardness and ultimate tensile strength of complex concentrated alloys (CCAs) by ML. The critical feature value separates the feature Shapley additive explanation (SHAP) values into positive and negative regions, in which the feature improves/impairs the mechanical properties of CCAs, thereby providing a straightforward assessment in the design of high hardness and high ultimate tensile strength CCAs. ML [9] reveals a new thermal conductivity mechanism of nanoporous graphene that disordered structures may also enhance thermal conductivity. Kirman et al. [10] used ML to guide the sequence of ever-improved robotic synthetic trials for discovering new perovskite single crystals. Xue et al. [11] found some NiTi-based shape memory alloys with very low thermal hysteresis (minimum value is 1.84 K) through adaptive ML, and Zhong et al. [12] also used adaptive ML to accelerate the discovery of CO_2 electrocatalysts. These studies demonstrate that the discovery process of novel advanced materials could be significantly accelerated and simplified if machine learning techniques are properly employed.

AI and ML have achieved remarkable success in the field of materials science and engineering, but most AI and ML algorithms perform as “black-box” systems [13–19]. Recently, considerable domain knowledge-guided efforts are endeavored to en-

* Corresponding authors at: Materials Genome Institute, Shanghai University, Shanghai 200444, China.
E-mail addresses: zqdong@shu.edu.cn (Z. Dong), zhangty@shu.edu.cn (T.-Y. Zhang).

https://doi.org/10.1016/j.jmst.2022.11.040
1005-0302/

hance the interpretability of ML models, aiming at the development of "transparent-box" or "white-box" ML models. For example, Wang et al. [20] illustrated how to integrate statistical learning with domain knowledge to handle statistically unreliable small data of size- and notch length-dependent nominal strengths of concrete. With domain knowledge, they proposed four hypotheses and accordingly a size-dependent normal distribution model and a size- and pre-notch length- dependent normal distribution model to statistically analyze the small data. Chen et al. [21] used Symbolic Regression to establish a physical meaningful formula between the features and target of Charpy impact toughness, where the features including alloy composition, heat treatment parameters and atomic variables. Zhou et al. [22] proposed a data-driven and machine learning assisted prediction and optimization strategy to explore the prototype FeCoNiCrMn HEAs with low hydrogen diffusion coefficients. With domain knowledge, a quantitative relationship between the diffusion coefficient and chemical composition was proposed to guide the design of HEAs with low H diffusion coefficients. Cao et al. [1] proposed a domain knowledge-guided interpretive ML strategy to study the oxidation kinetics of ferritic-martensitic steels in supercritical water, developed a novel ML algorithm of Tree-Classifier for Linear Regression (TCLR), and introduced the oxidation Cr equivalent concentration to capture the correlations among steel compositions, testing environments and oxidation behaviors. Under the guidance of domain knowledge, a generalized Arrhenius oxidation formula is accomplished with high prediction accuracy and wide generality. Domain knowledge-guided interpretive ML is powerful in the data-driven discovery of physics-informed formulas. The present work follows the strategy to discovery a high interpretive formula for the high temperature oxidation behavior of FeCrAlCoNi based high entropy alloys.

High entropy alloys (HEAs), also known as multi-component materials, typically contain five or more metallic elements with more or less equal concentrations [23–25]. Such a HEA forms a HEA base or system and further alloying can be conducted on the base HEA to improve the material properties. HEAs are a highly attractive class of materials, exhibiting excellent thermal stability, radiation resistance, corrosion resistance, and mechanical properties [26–30]. Their excellent properties allow them to have wide applications, such as heat-resistant and wear-resistant coatings, thermal-resistant coatings, materials for nuclear industries, structural materials for transportation and energy industries, and many more [31,32]. For high-temperature applications, the oxidation behavior of HEAs is critical and crucial. Oxidation resistance is the key performance index of HEAs working for a long time at high temperatures, because oxidation can seriously deteriorate the mechanical properties of materials [33]. Therefore, many experimental investigations have been conducted to study the oxidation behavior of various HEAs [34–41]. The experimental results indicate that in general, the higher the exposure temperature is, the faster the oxidation rate of the alloy will be [34]. For instance, Kai et al. [35] experimentally investigated the kinetic curve of the oxidation behavior of the $Ni_2FeCoCrAl_x$ (where $x = 0$, 0.5, and 1.0) HEAs in dry air at temperatures ranging from 600 °C to 900 °C, where the oxidation rate monotonically increases with increasing temperature. Elements Al, Cr and Si can form dense oxides Al_2O_3, Cr_2O_3 and SiO_2, respectively, on the surface of a metallic material and the oxide film prevents further oxidation of the underneath material. For this reason, Al, Cr and Si are usually considered as the oxide protective elements [36,37]. Liu et al. [38] experimentally observed that an increasing Al content can enhance the oxidation resistance of HEAs due to the formation of a continuous Al_2O_3 scale and the addition of Si element favors the formation of Al_2O_3, making the Al_2O_3 layers more compact and thicker and hence improving the oxidation resistance. On the other hand, some previous experimental investigations [39–41] show that elements Mn, Zr and Hf are unfavorable to the high-temperature oxidation resistance of HEAs. Dewangan et al. [42] established an artificial neural network (ANN) model to predict HEAs' high temperature oxidation weight gain by taking alloying composition, exposure time and oxidation temperature as input features. The established ANN model has high prediction accuracy with a Pearson coefficient greater than 0.999, which demonstrates that the ANN is a powerful tool for predicting the oxidation weight gain of HEAs.

The high-temperature oxidation behavior of HEA is clearly a thermally activated process, and the associated kinetics is largely dependent on the materials and testing conditions. The correlation between oxidation weight gain and testing conditions is incorporated into the Arrhenius equation of $\Delta W = k_{\text{eff}} \exp(-\frac{Q}{nRT}) t^{\frac{1}{n}}$ [43], where ΔW per unit area (mg/dm^2) denotes the oxidation weight gain, k_{eff} (mg/dm^2/h) is the pre-factor of effective oxidation rate constant, Q is the activation energy of oxidation (J/mol), R is the gas constant (8.314 J/(mol K)), T is the absolute temperature (K), t is time (h) and $1/n$ is the exponent over time. The oxidation weight gain of HEAs is commonly determined by experiments and the measurements are costly and time-consuming due to long-term oxidation testing and difficult alloy manufacturing, especially when HEAs involve refractory elements, such as Nb, Mo, Ta and W [44]. Thus, each individual group usually carries out a few experiments by trial and error in the search for high oxidation-resistant HEAs. As a result, the collected data of high-temperature oxidation of the FeCrAlCoNi-based HEAs diversify greatly, as described later, while the development of high oxidation resistant HEAs demands clear and simple guidance. The present work aims at the discovery of a mathematical formula as such guidance from complex and diversified data collected from the literature. To achieve this goal, the strategy of a domain knowledge-guided interpretive ML is adopted here to discover a generalized formula for the high-temperature oxidation of HEAs in air.

1.2. Arrhenius equation

In the present work, we adopted a dimensionless Arrhenius equation of

$$\frac{\Delta w}{\Delta w_{00}} = \frac{\Delta w_0}{\Delta w_{00}} \left(\frac{t}{t_0}\right)^m \exp\left(-\frac{Q}{RT}\right) \tag{1a}$$

to express the oxidation weight gain of HEAs in air. The Arrhenius equation was proposed by Cao et al. to describe the oxidation behavior of ferritic-martensitic steels in supercritical water [1], where $\Delta w_{00} = 1$ mg/m^2 and $t_0 = 1$ h denote the reference weight gain and time, respectively, $\frac{\Delta w_0}{\Delta w_{00}}$ is the dimensionless pre-factor, and $\frac{t}{t_0}$ is the dimensionless time. With the dimensionless Arrhenius equation, the present work integrates the Least Square Linear Regression [45] and the Tree-Classifier for Linear Regression (TCLR) [1,46] to discover a generalized formula from high-temperature oxidation data of FeCrAlCoNi-based HEAs in air. The SHAP algorithm [39], developed based on the game theory, is also used to analyze the contribution of features in the trained extreme gradient boosting (XGBoost) model.

1.3. Workflow

In this work, the feature ranking with SHAP values is carried out to find out the crucial features and the results show that the two testing condition features of $-\frac{1}{RT}$ and $\ln(\frac{t}{t_0})$ are ranked at the top two, thereby verifying the rationality of the time power law associated Arrhenius equation proposed by the domain knowledge. Subsequently, the TCLR is conducted to extract the activation energy Q and time exponent m from the dataset by using $\ln(\frac{\Delta w}{\Delta w_{00}})$ versus $-\frac{1}{RT}$ and $\ln(\frac{\Delta w}{\Delta w_{00}})$ versus $\ln(\frac{t}{t_0})$, respectively. The extracted

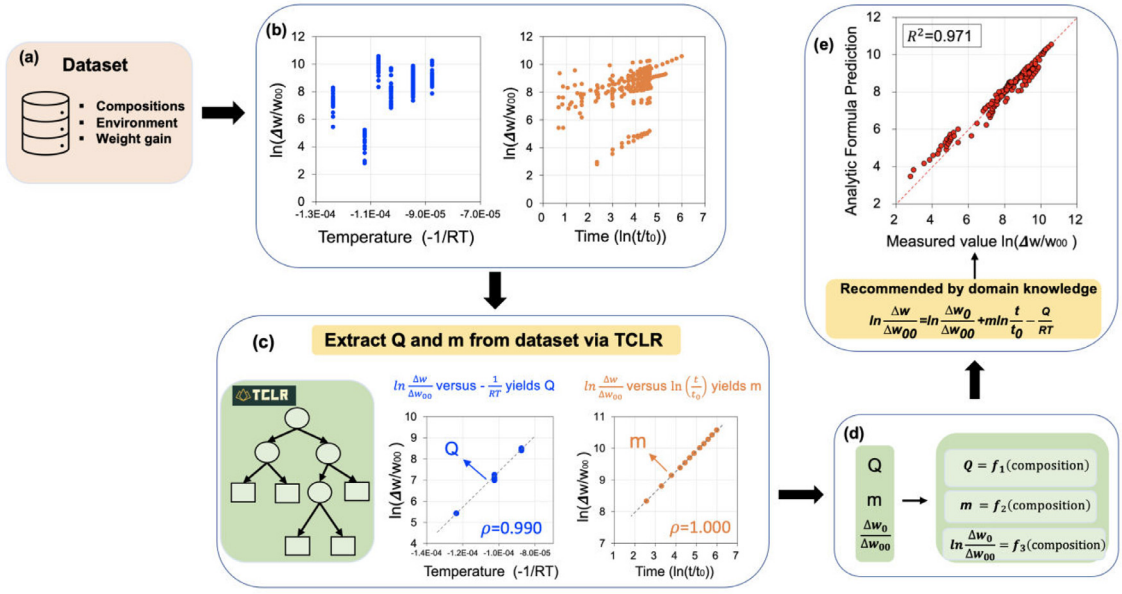

Fig. 1. Schematic diagram of the workflow. (a) The collection of high-temperature oxidation data of FeCrAlCoNi-based HEAs in air from the literature, (b) the diversity of the original data, (c) TCLR giving high linearity between $\ln(\frac{\Delta w}{\Delta w_{00}})$ and $-\frac{1}{RT}$ and between $\ln(\frac{\Delta w}{\Delta w_{00}})$ and $\ln(\frac{t}{t_0})$ for the data in each of leaves, (d) the three spectrums being functions of composition, and (e) the prediction of the final analytic formula.

values of Q and m depend on the alloy composition, and in turn, ML regression with quadratic function and Least Absolute Shrinkage and Selection Operator (LASSO) algorithm with Leave One Out Cross Validation (LOOCV) is done to obtain analytic relationships between Q and composition and between m and composition, and finally having an analytical formula for the oxidation behavior of all the collected data. Fig. 1 outlines the workflow of this study, where the dimensionless Arrhenius equation and the TCLR algorithm are the critical components in the domain knowledge-guided high interpretive ML strategy.

2. Methods

2.1. Data

A total of 205 high-temperature oxidation data of FeCrAlCoNi-based HEAs are collected from the literature [32,41,47–49,] and given in Table S1 in the Supplementary Information. All collected data of HEAs, which were synthesized by the arc-melting method and the test specimens were prepared in as-cast state, are generated via the same processing route. The high-temperature oxidation experiments were carried out in the air atmosphere and thus the experimental features are the testing temperature and exposure time. The eight alloy chemical composition features, two testing condition features of temperature and time, and one response of oxidation caused weight gain in units of mg/m^2 are shown in Table 1. In the present work, the collected oxidation data of HEAs only contain the eight elements of Al, Cr, Fe, Co, Ni, Cu, Ti, and Si. If HEAs contain refractory elements, the oxidation behavior of HEAs will be more complex and deserve more systematic investigations.

Table 1
The 10 original features and oxidation weight gain of 205 high entropy alloys data.

No.	Features	Description	Min	Max
1	Al	at.% of Aluminium	0	30.23
2	Cr	at.% of Chromium	16.67	23.26
3	Fe	at.% of Iron	16.67	23.26
4	Co	at.% of Cobalt	0	23.26
5	Ni	at.% of Niobium	16.67	23.26
6	Cu	at.% of Copper	0	20.00
7	Ti	at.% of Titanium	0	16.67
8	Si	at.% of Silicon	0	6.12
9	T	Test temperature (K)	973	1373
10	t	Exposure time (h)	2	401
	Δw	Oxidation weight gain (mg/m^2)	16.4	39,138.6

2.2. ML algorithms

The Extreme Gradient Boosting (XGBoost) algorithm is a powerful gradient-boosted decision tree (GBDT), and the details of the XGBoost algorithm are given in Supplementary Information. The present work uses the XGBoost algorithm in the open python library [50] to regress the oxidation weight gain of HEAs. The SHAP algorithm is developed based on the game theory that partitions each prediction into individual feature contributions [51]. The details of the SHAP algorithm are given in Supplementary Information.

Tree-Classifier for Linear Regression (TCLR) is a novel tree-based method to capture the functional relationships between features and targets [1,46], through partitioning the feature space into a set of sub-domains, and embodies a specific function in every subdomain. TCLR can automatically choose the splitting features and points with given criteria for splitting the tree. In each leaf, TCLR model the response as a function of one feature $x_{\hat{j}}(\hat{j} = 1, \cdots, \hat{m})\hat{m} < m$, $y = f(x_{\hat{j}})$ where f is called "form prior" determined by domain knowledge. If we implement a criterion of maximizing the Linearity Goodness (LG, but not linearity exclusive), then, finding the best partition in terms of maximum LG is generally computationally feasible. The LG metrics in TCLR include the Pearson correlation coefficient (R), the coefficient of determination (R^2) and maximal information coefficient (MIC). Starting with all the n data $\{x_{ij}\}(i = 1, \cdots, n)(j = 1, \cdots, m)$, consider a splitting variable j and splitting point i, and a child node pair of $R^l(x|x_j \le i)$ and $R^r(x|x_j > i)$. The optimal splitting variable j and splitting point i of a given dataset with positive linear correlation are determby:

$$\max_{j,i}\left[\frac{1}{2}\left(LG\left(x_{\hat{j}}, f\right)_{R^l} + LG\left(x_{\hat{j}}, f\right)_{R^r}\right) - LG\left(x_{\hat{j}}, f\right)_{R^l+R^r}\right]. \quad (2a)$$

TCLR binary partitions the space recursively according to Eq. (2a), until reaching the stop rules.

The Least Squares Linear Regression (LSLR), Least Absolute Shrinkage and Selection Operator (LASSO) algorithm with Leave One Out Cross Validation (LOOCV) are also used in the present work with the open source software scikit-learn [52], and the details of Linear Regression, LASSO and LOOCV are given in Supplementary Information. The predictive ability of an ML model is evaluated by the determination coefficient R^2, which is defined by

$$R^2 = 1 - \frac{\sum_{i=1}^{N}\left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{N}\left(y_i - \bar{y}\right)^2}, \tag{2b}$$

where N is the number of samples, y, $\hat{y}$, and $\bar{y}$ denote the actual value, the predicted value and the average value, respectively. The value of $R^2 = 1$ indicates a perfect prediction. If using the average value $\bar{y}$ predicts all output values of $\hat{y}_i$, the model is called the baseline model and has $R^2 = 0$. When an ML model performs worse than the baseline model, it will have a negative R^2.

3. Results and discussion

3.1. XGBoost model and feature importance

Taking logarithms on both sides of the dimensionless Arrhenius Eq. (1a) yields

$$\ln\left(\frac{\Delta w}{\Delta w_{00}}\right) = \ln\left(\frac{\Delta w_0}{\Delta w_{00}}\right) + m\ln\left(\frac{t}{t_0}\right) - \frac{Q}{RT} \tag{3a}$$

which exhibits the additive contributions of $\ln(\frac{\Delta w_0}{\Delta w_{00}})$, $m\ln(\frac{t}{t_0})$ and $-\frac{Q}{RT}$ to $\ln(\frac{\Delta w}{\Delta w_{00}})$. Obviously, the testing condition features of temperature T and exposure time t are explicitly shown in the Arrhenius equation suggested by the domain knowledge. The contributions from the eight chemical element features to the oxidation weight gain are via the three physical parameters of Δw_0, m and Q. Taking the additive behavior, we constructed two new testing condition features of $\ln(\frac{t}{t_0})$ and $-\frac{1}{RT}$.

An XGBoost model is first established with all the 10 original features including 8 alloying element features and two testing features of T and t, and the 10-CV prediction of Δw versus the true value on the 205 data is shown in Fig. S2 in Supplementary Information, showing a prediction accuracy with a determination coefficient of $R^2 = 0.955$. For comparison, another XGBoost model is developed with the new feature set including 8 alloying element features and two new testing condition features of $\ln(\frac{t}{t_0})$ and $-\frac{1}{RT}$ to predict the logarithm of weight gain, i.e., $\ln(\frac{\Delta w}{\Delta w_{00}})$. Fig. 2(a) shows the 10-CV prediction of $\ln(\frac{\Delta w}{\Delta w_{00}})$ versus the true value, indicating that the prediction accuracy is improved from $R^2 = 0.955$ to $R^2 = 0.987$ due to the use of $\ln(\frac{t}{t_0})$, $-\frac{1}{RT}$ and $\ln(\frac{\Delta w}{\Delta w_{00}})$. The improvement might be a clue that the suggested Arrhenius equation is valid to describe the oxidation behaviors, although such an ML model cannot provide an analytic formula.

To analyze the contribution of each feature, SHAP values are calculated from the XGBoost model. Based on the second established XGBoost model, all point-wise SHAP values ϕ_{ij} are calculated, where subscripts "i" and "j" denote sample i ($i = 1, \cdots, n$) and feature x_j ($j = 1, \cdots, m$), respectively. Then, the mean $\bar{\phi}_j$ and standard deviation $\sigma^2_{\phi_j}$ of SHAP values, and the absolute mean $|\bar{\phi}_j|$ of SHAP values for each feature x_j ($j = 1, \cdots, m$) are calculated from

$$\bar{\phi}_j = \frac{1}{n}\sum_{i=1}^{n}\phi_{ij}, \tag{3b}$$

$$\sigma^2_{\phi_j} = \frac{1}{n-1}\sum_{i=1}^{n}\left(\phi_{ij} - \bar{\phi}_j\right)^2, \tag{3c}$$

$$\left|\bar{\phi}_j\right| = \frac{1}{n}\sum_{i=1}^{n}\left|\phi_{ij}\right|, \tag{3d}$$

where $n = 205$ and $m = 10$ are the numbers of data and features, respectively. In the present work, we propose to rank the feature importance, F_{ϕ_j}, by the product of the absolute mean of SHAP values times the standard deviation of SHAP values for each feature, viz.,

$$F_{\phi_j} = \left|\bar{\phi}_j\right|\sigma^2_{\phi_j}. \tag{3e}$$

The proposed feature importance F_{ϕ_j} takes the classical feature importance $|\bar{\phi}_j|$ of SHAP values as one component. This means that the higher the absolute mean of SHAP values of a feature is, the greater the contribution of the feature to the response will be. In addition, the proposed feature importance F_{ϕ_j}, in analogy with Principal Component Analysis [53], takes the standard deviation of SHAP values into account. The above statistical indicator is extremely significant in the control and design of an appropriate response value by choosing a right value of the feature. Consider the extremal case that if all SHAP values of a feature are high and distributed in a very narrow range, this feature is important to the objective, but it is hard to control the objective by varying such feature value because its standard deviation of SHAP values is very small. On the other hand, if the standard deviation of SHAP values of a feature is large, varying the feature content will tune the objective and therefore the feature will play a more important role in the design, manufacture, or service of materials.

Fig. 2(b) shows the feature importance by F_{ϕ_j}, indicating that the two testing condition features of $-\frac{1}{RT}$ and $\ln(\frac{t}{t_0})$ rank at the top. The results imply that the testing temperature and exposure time play the most important roles in the high-temperature oxidation of HEAs in air atmosphere. In general, every alloying element feature importance is lower than the two testing condition features. Moreover, the feature importance of Fe and Ni is zero, indicating that the feature Fe and Ni have not been used during the generation of XGBoost model. This is due to that the atom percentages of Cr, Ni and Fe for each datum are equal in the collected oxidation HEAs dataset, that is, Cr(at.%):Ni(at.%):Fe(at.%)=1:1:1. This means that when feature Cr is used for splitting, the feature values of Ni and Fe are split accordingly so that features of Ni and Fe are not used for splitting. This phenomenon is called the masking effect [54].

As mentioned above, the two most important features are experimental conditions of $-\frac{1}{RT}$ and $\ln(\frac{t}{t_0})$. Fig. 2(c) illustrates much diversity in the plot of $\ln(\frac{\Delta w}{\Delta w_{00}})$ versus $-\frac{1}{RT}$, and Fig. 2(d) indicates even much great diversity in the plot of $\ln(\frac{\Delta w}{\Delta w_{00}})$ versus $\ln(\frac{t}{t_0})$. The great diversity is caused by the difference in material composition. As listed in Table 1, the 205 data are located in a ten-dimensional space of Al $\in$ [0, 30.23] at.%, Cr $\in$ [16.67, 23.26] at.%, Fe $\in$ [16.67, 23.26] at.%, Co $\in$ [0, 23.26] at.%, Ni $\in$ [16.67, 23.26] at.%, Cu $\in$ [0, 20.00] at.%, Ti $\in$ [0, 16.67] at.%, Si $\in$ [0, 6.12] at.%, $T \in$ [973, 1373] K, and $t \in$ [2, 401] h. The eight element features cause diversity in $\ln(\frac{\Delta w}{\Delta w_{00}})$ versus $-\frac{1}{RT}$ (or $\ln(\frac{t}{t_0})$) and it is significant for the design and development of anti-oxidation HEAs that how the material composition affects the oxidation performance of the HEAs. Under the guidance of domain knowledge, the TCLR method has the ability to explore the influence of material composition on oxidation performance.

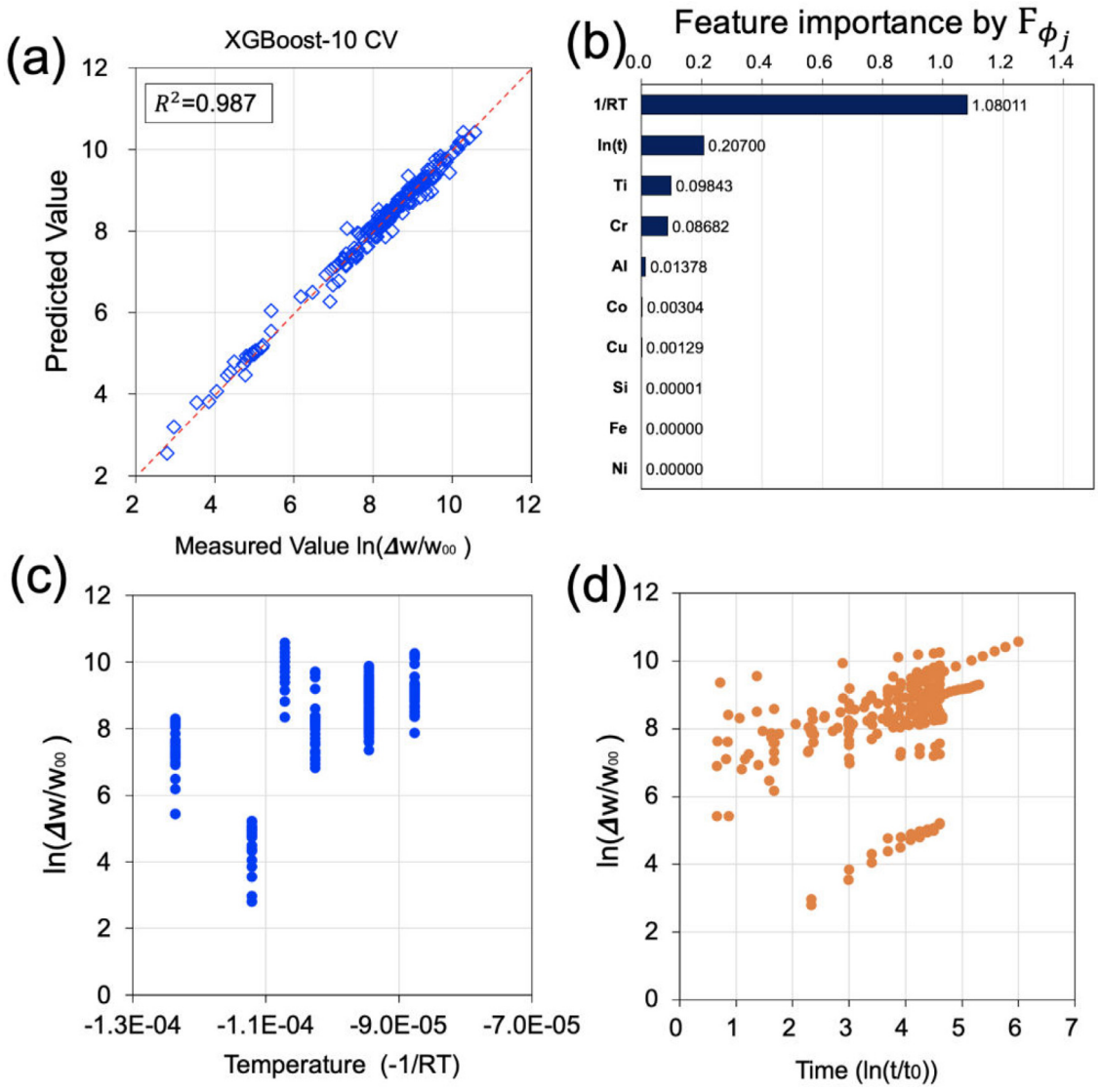

Fig. 2. (a) The 10-CV predicted values from XGBoost model versus measured values, (b) the feature importance ranking by F_{ϕ_j}, the plots of (c) $\ln(\frac{\Delta w}{\Delta w_{00}})$ versus $-\frac{1}{RT}$ and (d) $\ln(\frac{\Delta w}{\Delta w_{00}})$ versus $\ln(\frac{t}{t_0})$ on the collected HEAs oxidation dataset.

3.2. Extract the activation energy and time exponent from the dataset by TCLR

As discussed above, the activation energy Q and time exponent m determine the oxidation mechanism of HEAs in high-temperature air atmosphere, while the chemical composition and testing conditions determine the values of Q and m. This is because the chemical composition and microstructure are the intrinsic internal factors that dictate the oxidation mechanism and the external testing conditions may change the mechanism of oxidation, e.g., increasing temperature converts grain boundary predominated diffusion to lattice predominated diffusion. Experimental measurement of activation energy Q (or time exponent m) is conducted to test on a given material at various temperatures (or exposure time) with other testing conditions fixed [1] and then using Arrhenius equation (or power law) determines the value of Q (or m). The TCLR algorithm is designed for such purpose to identify, from a complex dataset, some points that approximately meet the ideal experiment and have a good linear correlation between objective and feature. In the present work, the TCLR algorithm is used to extract the activation energy Q and time exponent m from the high dimensional space dataset by dividing the whole feature space into many subdomains, called leaves, and the data in each leaf have high Linearity Goodness (LG) between $\ln(\frac{\Delta w}{\Delta w_{00}})$ and $-\frac{1}{RT}$ (or $\ln(\frac{t}{t_0})$), meaning that a leaf gives a value of Q (or m) so that a spectrum of Q (or m) is formed by the many leaves. Table S2 lists the optimized values of hyperparameters of TCLR and the two whole TCLR trees are given in Supplementary Information also. As an example, a small portion of the TCLR tree is shown for $\ln(\frac{\Delta w}{\Delta w_{00}})$ versus $-\frac{1}{RT}$ in Fig. 3(a) and $\ln\frac{\Delta w}{\Delta w_{00}}$ versus $\ln(\frac{t}{t_0})$ in Fig. 3(c). In the blue leaf of the TCLR in Fig. 3(a), there are eight data and their feature values are in the regions of Al $\in$ [18.18, 25.00] at.%, Cr $\in$ [16.67, 18.18] at.%, Fe $\in$ [16.67, 18.18] at.%, Co $\in$ [16.67, 18.18] at.%, Ni $\in$ [16.67, 18.18] at.%, Cu = 0 at.%, Ti $\in$ [8.33, 9.09] at.%, Si = 0 at.%, $T \in$ [973, 1373] K, and $t \in$ [2, 4] h. Fig. 3(b) shows the linear fitting of $\ln(\frac{\Delta w}{\Delta w_{00}})$ versus $-\frac{1}{RT}$ on the eight data and the slope yields the activation energy Q = 83.194 kJ/mol, indicating that at a given temperature there are still few data that have slightly different values of $\ln(\frac{\Delta w}{\Delta w_{00}})$. The result indicates that the slightly different values of $\ln(\frac{\Delta w}{\Delta w_{00}})$ at a given temperature are at-

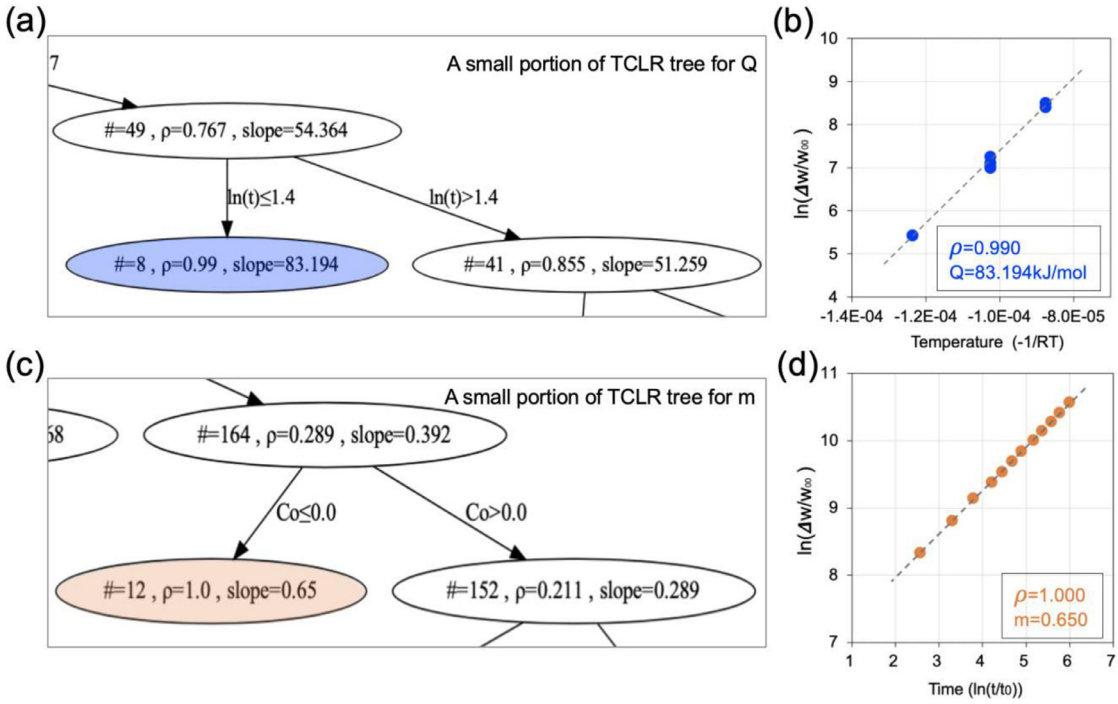

Fig. 3. A small portion of the TCLR tree for (a) activation energy and (c) time exponents, where typical leaves are highlighted; (b) the $\ln(\frac{\Delta w}{\Delta w_{00}})$ versus $-\frac{1}{RT}$ of the eight data in the highlighted leaf for activation energy, (d) the $\ln(\frac{\Delta w}{\Delta w_{00}})$ versus $\ln(\frac{t}{t_0})$ of the twelve data in the highlighted leaf for time exponents.

tributed to the other features. Nevertheless, within such a small subdomain in the feature space, the eight data share the same activation energy $Q = 83.194$ kJ/mol for oxidation at the temperature range during the exposition time period. There are twelve data in the light red leaf of Fig. 3(c) and the twelve data are within such a small subdomain in the feature space of = Cr = Fe = Ni = Cu = Ti = 16.67 at.%, Co = Si = 0 at.%, $T = 1123$ K and $t \in [13,\ 401]$ h, implying that the twelve data are generated on a certain HEA and the tests are conducted at a given temperature for various exposure time periods. As expected, the ideal tests lead to excellent linearity between $\ln(\frac{\Delta w}{\Delta w_{00}})$ and $\ln(\frac{t}{t_0})$ with the time exponent $m = 0.650$, as shown in Fig. 3(d), where only one datum exists at an exposure time. The value of $m = 0.650$ might indicate that the oxidation during the time period $t \in [13,\ 401]$ h is predominated by diffusion [43].

Fig. S3(a) shows the whole TCLR result of $\ln(\frac{\Delta w}{\Delta w_{00}})$ versus $-\frac{1}{RT}$, which is developed by setting the Pearson correlation coefficient ($\rho \geq 0.8$) and hence has 23 leaves and the data in each leaf range from 3 to 25. The TCLR yields the values of activation energy Q and Fig. 5(a) and Table S3 in Supplementary Information show the variation of activation energy and the subspace of features on each leaf with a certain activation energy, respectively, which forms a spectrum of the activation energy Q in the feature space. The obtained activation energy Q varies from 49.46 kJ/mol to 261.90 kJ/mol, which is all in good agreement with the previously reported values of HEAs experiment [32,41,47–49,]. The activation energy Q is the energy barrier for thermal energy to overcome in the oxidation process, which is vitally important to indicate how the temperature affects the oxidation behavior. At a given temperature, the larger activation energy indicates that the oxidation reaction is more difficult to occur, which corresponds to a stronger oxidation resistance [55]. The activation energy Q, found from the collected data with the same processing route, depends mainly on the alloy composition. The eight alloy elements vary over a wide range, as listed in Table 1, and cause the variation in activation energy Q.

Similarly, the whole TCLR result of $\ln(\frac{\Delta w}{\Delta w_{00}})$ versus $\ln(\frac{t}{t_0})$, developed with the same splitting criterion $\rho \geq 0.8$, is shown in Fig. S3(b) and has 18 leaves and the data in each leaf range from 7 to 20. The TCLR yields the values of time exponent m and Fig. 5(b) and Table S4 show the variation of time exponent m and the subspace of features on each leaf with a certain value of time exponent m, respectively, which forms a spectrum of the time exponent m in the feature space, indicating that the time exponent m forms a bimodal distribution that 175 data are in the range of [0.2, 0.7] and 30 data are in the range of [0.9, 1.1]. The time exponent m reflects how fast the oxidation weight gain with exposure time. The larger the time exponent is, the faster the oxidation weight gain will be. In the present study, all data pass the splitting criterion in the two TCLR trees. If some data cannot pass the criterion, these data do not follow the proposed formula. In this case, caution must be used to examine the reliability of these data and/or reconsider the proposed formula. With the determined two spectrums of Q and m, the spectrum of pre-factor or $\ln(\frac{\Delta w_0}{\Delta w_{00}}) = \ln(\frac{\Delta w}{\Delta w_{00}}) - m\ln(\frac{t}{t_0}) + \frac{Q}{RT}$ is obtained straightforwardly.

3.3. Analytical formula

If the oxidation mechanism does not vary during the experiments, the activation energy is independent of exposure time, although temperature may cause the change in the oxidation mechanism and then the change in activation energy [54,56]. The oxidation temperatures in all the HEA data range from 973 K to 1373 K and the activation energy is assumed to be independent of temperature. Thus, the activation energies of the 205 HEAs data depend only on material composition or the eight element features. A quadratic function feature pool is constructed with the eight elements, containing 16 features of Al, Cr, Fe, Co, Ni, Cu, Ti, Si, Al^2, Cr^2, Fe^2, Co^2, Ni^2, Cu^2, Ti^2 and Si^2. Then, LASSO algorithm with LOOCV is conducted to select features for the regression of activation energy Q. Fig. 4(a) shows the coefficient of determination

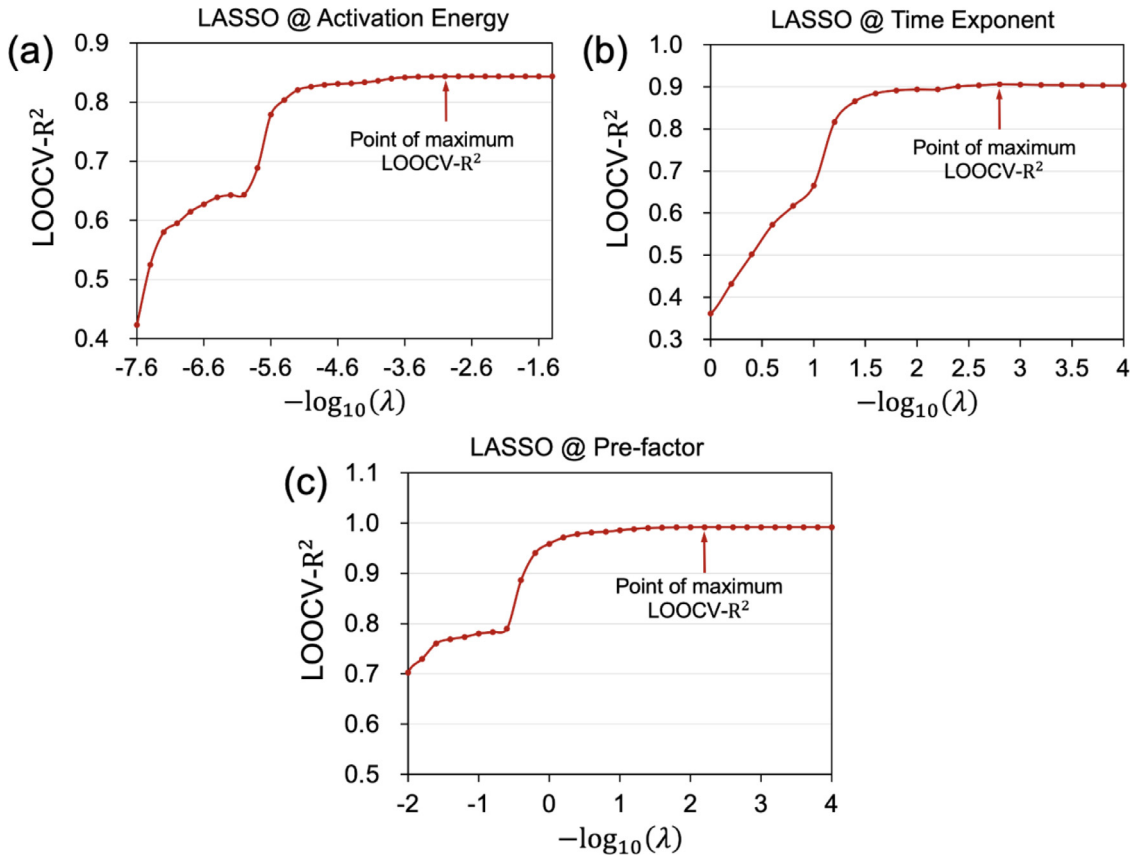

Fig. 4. The LOOCV coefficient of determination R^2 versus the penalty coefficient $-\log_{10}(\lambda)$ in LASSO for (a) activation energy Q, (b) time exponent m and (c) pre-factor.

R^2 of LOOCV reaches the maximum at the penalty coefficient of $\lambda = 10^3$, which selects 10 features of Cu, Ti, Al^2, Cr^2, Fe^2, Co^2, Ni^2, Cu^2, Ti^2 and Si^2 from the feature pool, and gives the analytic expression of activation energy Q as

$$Q^{\mathrm{LR}} = 123.015\mathrm{Al}^2 + 66.051\mathrm{Cr}^2 + 66.051\mathrm{Fe}^2 + 178.086\mathrm{Co}^2 + 66.051\mathrm{Ni}^2 + 8198.492\mathrm{Cu} - 229.703\mathrm{Cu}^2 - 20753.850\mathrm{Ti} + 967.148\mathrm{Ti}^2 + 929.394\mathrm{Si}^2 \ (\mathrm{kJ/mol}). \quad (4)$$

The linear analytical formula Eq. (4) has a fitting performance of $R^2 = 0.847$, and the formula prediction of Q versus the value from TCLR is plotted in Fig. S4 (a). It is interesting to note that the coefficients of elements Cr, Fe, and Ni in Eq. (4) are all the same due to the fixed ratio of Cr:Fe:Ni=1:1:1. The result indicates when few features fix their ratio and vary synchronously, their contribution to the task will be the same. The analytical formula shows that with the studied element ranges, the higher Al, Cr, Co and Si concentrations, the larger activation energy will be. The larger activation energy corresponds to stronger oxidation resistance. Al, Cr and Si atoms can form dense and protective oxide layers of Cr_2O_3, Al_2O_3 and SiO_2 that provide strong resistance to oxidation [36,37]. Ni has high corrosion resistance to acids and alkalis at high-temperature environment [57] and thus alloying higher Ni content may enhance the oxidation resistance of HEAs in high-temperature testing conditions. It deserves more systematic investigation to understand the roles of elements Cu, Co and Ti in the oxidation behaviors of HEAs in air.

Based on the domain knowledge, the time exponent is assumed to depend only on the material composition. The same approach for determining formula Eq. (4) was carried out to find an analytic formula between the time exponent m and element features. Fig. 4(b) shows the LOOCV R^2 reaches the maximum at the penalty coefficient of $\lambda = 10^{-2.8}$, which selects 15 features of Al, Cr, Fe, Co, Ni, Cu, Ti, Al^2, Cr^2, Fe^2, Co^2, Ni^2, Cu^2, Ti^2 and Si^2 from the feature pool and results in the analytic expression

$$m^{\mathrm{LR}} = 579.51\mathrm{Al} - 0.069\mathrm{Al}^2 + 608.80\mathrm{Cr} - 0.964\mathrm{Cr}^2 + 608.80\mathrm{Fe} - 0.964\mathrm{Fe}^2 + 542.86\mathrm{Co} + 1.455\mathrm{Co}^2 + 608.80\mathrm{Ni} - 0.964\mathrm{Ni}^2 + 579.50\mathrm{Cu} - 0.068\mathrm{Cu}^2 + 593.39\mathrm{Ti} - 1.596\mathrm{Ti}^2 + 94.316\mathrm{Si}^2 - 58,372.892. \quad (5)$$

As expected, the coefficients of elements Cr, Fe, and Ni in Eq. (5) are all the same due to the fixed ratio in the HEAs. The analytic formula Eq. (5) has a high fitting accuracy of $R^2 = 0.911$, and the formula prediction of m versus the value from TCLR is plotted in Fig. S4(b). In formula Eq. (5), the coefficients of the first-order and second-order terms are both positive for Co and the coefficient of Si is positive, indicating that increasing the content of these two elements definitely increases the time exponents. However, the coefficients of the first-order and second-order terms are not all positive or negative for the other six elements, meaning that the influence of these six elements on the time exponent is complex and non-monotonic.

Again, the pre-factor $\ln(\frac{\Delta w_0}{\Delta w_{00}})$ is assumed to depend on only eight element features. The same approach for determining formula Eq. (4) was carried out to find an analytic formula between the pre-factor and element features. Fig. 4(c) shows the LOOCV R^2 reaches the maximum at the penalty coefficient of $\lambda = 10^{-2.2}$, which selects 13 features of Al, Cr, Co, Cu, Ti, Al^2, Cr^2, Fe^2, Co^2, Ni^2, Cu^2, Ti^2 and Si^2 from the feature pool and yields the analytic

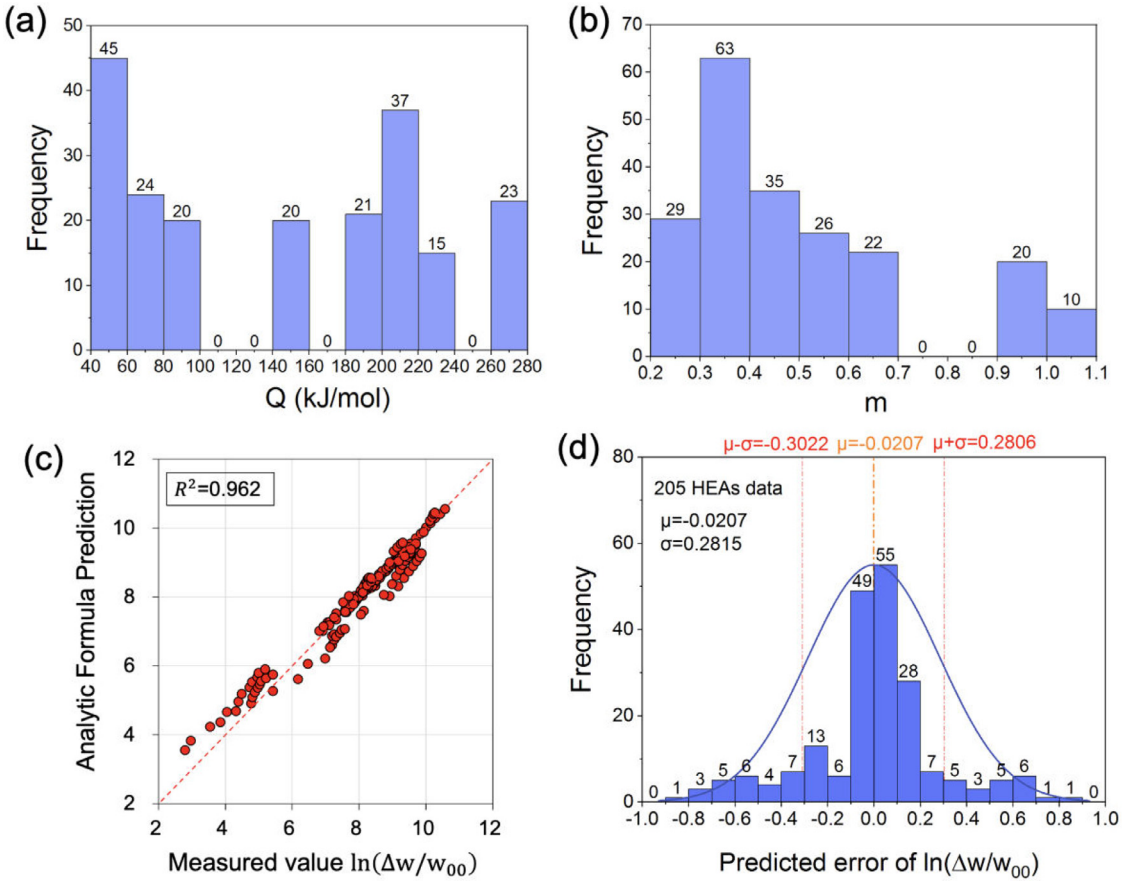

Fig. 5. The distribution of the (a) activation energy Q and (b) time exponent m obtained by TCLR; (c) the prediction from the high interpretative formula Eq. (7) versus the measured value and (d) the prediction error histogram on whole HEAs data.

Table 2
Summation of three ML models for the oxidation behavior of HEAs.

Data size	Features	ML algorithms	Providing analytic formula (Yes or No)	Refs.
300	Six elements: Al, Cr, Fe, Mn, Ni and W, oxidation temperature and exposure time	Artificial Neural Network (ANN)	No	[42]
214	Fourteen elements: Ni, Al, Cr, Ti, Si, Cu, Fe, Co, Mn, V, Zr, Hf, Ta and Nb	Tree-based Pipeline Optimization Tool (TPOT) and Elastic Net	No	[58]
205	Eight elements: Al, Cr, Fe, Co, Ni, Cu, Ti, and Si, oxidation temperature and exposure time	XGBoost, SHAP, TCLR, LASSO and LSLR	Yes	The present work

formula

$$\ln\left(\frac{\Delta w_0}{\Delta w_{00}}\right)^{LR} = -1353.273Al + 0.179Al^2 - 4268.779Cr + 2.320Cr^2 + 2.320Fe^2 - 1262.339Co - 3.604Co^2 + 2.320Ni^2 - 1352.441Cu + 0.143Cu^2 - 1389.475Ti + 4.030Ti^2 - 220.136Si^2 + 136,299.812 \quad (6)$$

with an excellent fitting performance of $R^2 = 0.998$, and the formula prediction of pre-factor versus the value from TCLR is plotted in Fig. S4(c).

Finally, putting the analytic expressions Eqs. (4), (5) and (6) together gives

$$\ln\left(\frac{\Delta w}{\Delta w_0}\right)^{LR} = \ln\left(\frac{\Delta w_0}{\Delta w_{00}}\right)^{LR} + m^{LR}\ln\left(\frac{t}{t_0}\right) - \frac{Q^{LR}}{RT}. \quad (7)$$

The formula Eq. (7) analytically describes the time and temperature-dependent oxidation behavior of HEAs in air. The contributions of testing temperature T and exposure time t to the oxidation weight gain $\ln(\frac{\Delta w}{\Delta w_{00}})$ are explicitly shown in the equation, and the contributions of HEA compositions to the oxidation weight gain $\ln(\frac{\Delta w}{\Delta w_{00}})$ are via the three parameters of $\ln(\frac{\Delta w_0}{\Delta w_{00}})$, m and Q. As expected, the high interpretative formula of Eq. (7) has a strong predictive power for 205 HEAs data, as shown in Fig. 5(c), with a fitting performance of $R^2 = 0.971$. The predicted error of 205 HEAs data is shown in Fig. 5(d), which follows a normal distribution with a mean value of $\mu = -0.0207$ and a standard deviation of $\sigma = 0.2851$. The errors vary from −0.82 to 0.86, and all errors are smaller than $\pm 3\sigma$, meaning that all data are fitted extremely well. It should be noted that the analytic expressions of Eqs. (4)–(6) contain only the eight chemical elements of Al, Cr, Fe, Co, Ni, Cu, Ti, and Si and therefore, it might not be appropriate to apply them to other HEAs systems with other elements. The comparison of three ML models for the oxidation behavior of high entropy al-

loys is shown in Table 2, indicating the present work is the only one that provides analytic formula.

4. Conclusions

The present work demonstrates the success of the domain knowledge-guided ML in the discovery of a high interpretive formula for the oxidation weight gain of HEAs in air. The established analytical formula for the oxidation behavior of the HEAs is extremely important for the fast evaluation of the high temperature oxidation behavior of the HEAs, and thus provides guidance to the inverse design of the HEAs against high temperature oxidation. More significantly, the strategy of domain knowledge guided ML paves the bright avenue for the development of materials informatics. The main outcomes of the present work are indexed as follows:

(1) The domain knowledge suggests the generalized dimensionless Arrhenius equation of $\frac{\Delta w}{\Delta w_{00}} = \frac{\Delta w_0}{\Delta w_{00}}(\frac{t}{t_0})^m \exp(-\frac{Q}{RT})$ based on the fact that the oxidation is a thermally active process and the collected oxidation data are conducted at various testing temperatures.
(2) A novel statistical measurement F_{ϕ_j} of feature importance is proposed based on SHAP values. The feature importance F_{ϕ_j} is an indicator that combined the mean of absolute SHAP values and the standard deviation of the SHAP values.
(3) The TCLR algorithm has the ability to efficiently split high dimensional HEAs data into many terminal nodes (leaves) and data in every leaf possess a good linear relationship between the objective and splitting feature.
(4) The spectrums of activation energy and time exponent are obtained from two TCLR trees, which also yield the spectrum of pre-factor naturally. The activation energy varies from 49.46 kJ/mol to 261.90 kJ/mol and the time exponent m varies from 0.23 to 1.10, both are in good agreement with experiment results of the HEAs oxidation.
(5) The three spectrums are regressed by using the chemical element features, resulting in a high interpretative formula. The analytical formula has extremely high prediction accuracy with a determination coefficient R^2=0.971 and provides guidance in the development of anti-oxidation HEAs and in the control and protection of HEAs oxidation in air at high temperatures.

Software

All ML approaches are performed on Python. The SHAP values are calculated by the SHAP library, the TCLR results are calculated by the open source TCLR package in Python, and the LASSO algorithm is available in scikit-learn libraries.

Availability of data and source codes

All experimental data collected in the study are contained in Supplementary Information. Source codes of the programmers are available at: https://github.com/qinghuawei/HEAs-oxidation. The source code of TCLR is available at: https://github.com/Bin-Cao/TCLRmodel.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Qinghua Wei: Conceptualization, Investigation, Methodology, Data curation, Writing – original draft. **Bin Cao:** Methodology. **Lucheng Deng:** Data curation. **Ankang Sun:** Data curation. **Ziqiang Dong:** Supervision. **Tong-Yi Zhang:** Supervision, Methodology, Writing – review & editing.

Acknowledgments

This work was financially supported by the National Key Research and Development Program of China (No. 2018YFB0704400), the Key Program of Science and Technology of Yunnan Province (No. 202002AB080001–2), the Key Research Project of Zhejiang Laboratory (No. 2021PE0AC02), and the Shanghai Pujiang Program (No. 20PJ1403700).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jmst.2022.11.040.

References

[1] B. Cao, S. Yang, A.K. Sun, Z.Q. Dong, T.Y. Zhang, J. Mater. Inf. 2 (2022) 4.
[2] Q.H. Wei, J. Xiong, S. Sun, T.Y. Zhang, Sci. Sin. Tech. 51 (2021) 722–736 (in Chinese).
[3] J. Xiong, T.Y. Zhang, S.Q. Shi, Sci. China Technol. Sci. 63 (2020) 1247–1255.
[4] A. Leitherer, A. Ziletti, L.M. Ghiringhelli, Nat. Commun. 12 (2021) 1–13.
[5] S. Sun, R. Ouyang, B. Zhang, T.Y. Zhang, MRS Bull. 44 (2019) 559–564.
[6] J. Xiong, S.Q. Shi, T.Y. Zhang, J. Mater. Sci. Technol. 87 (2021) 133–142.
[7] J. Xiong, T.Y. Zhang, J. Mater Sci. Technol. 121 (2022) 99–104.
[8] T.Y. Zhang, An Introduction to Materials Informatics (I): the Elements of Machine Learning, Science Press, Beijing, 2022.
[9] H. Wei, H. Bao, X. Ruan, Nano Energy 71 (2020) 104619.
[10] J. Kirman, A. Johnston, D.A. Kuntz, M. Askerka, Y. Gao, P. Todorović, E.H. Sargent, Matter 2 (2020) 938–947.
[11] D. Xue, P.V. Balachandran, J. Hogden, J. Theiler, D. Xue, T. Lookman, Nat. Commun. 7 (2016) 1–9.
[12] M. Zhong, K. Tran, Y. Min, C. Wang, Z. Wang, C.T. Dinh, E.H. Sargent, Nature 581 (2020) 178–183.
[13] S. Ramakrishna, T.Y. Zhang, W.C. Lu, Q. Qian, J.S.C. Low, J.H.R. Yune, S.R.J. Kalidindi, Intell. Manuf. 30 (2019) 2307–2326.
[14] T. Lookman, P.V. Balachandran, D. Xue, R. Yuan, npj Comput. Mater. 5 (2019) 1–17.
[15] C. Wen, Y. Zhang, C. Wang, D. Xue, Y. Bai, S. Antonov, Y. Su, Acta Mater. 170 (2019) 109–117.
[16] P.V. Balachandran, B. Kowalski, A. Sehirlioglu, T. Lookman, Nat. Commun. 9 (2018) 1–9.
[17] L. Yan, Y. Diao, Z. Lang, K. Gao, Sci. Technol. Adv. Mater. 21 (2020) 359–370.
[18] K.M. Jablonka, G.M. Jothiappan, S. Wang, B. Smit, B. Yoo, Nat. Commun. 12 (2021) 1–10.
[19] J.A. GarridoTorres, V. Gharakhanyan, N. Artrith, T.H. Eegholm, A. Urban, Nat. Commun. 12 (2021) 1–9.
[20] J.H. Wang, J.N. Jia, S. Sun, T.Y. Zhang, Eng. Fract. Mech. 259 (2022) 108160.
[21] Y. Chen, S. Wang, J. Xiong, G. Wu, J. Gao, Y. Wu, X. Mao, J. Mater. Sci. Technol. 132 (2023) 213–222.
[22] X.Y. Zhou, J.H. Zhu, Y. Wu, X.S. Yang, T. Lookman, H.H. Wu, Acta Mater. 224 (2022) 117535.
[23] B. Cantor, I.T.H. Chang, P. Knight, A.J.B. Vincent, Mater. Sci. Eng. A 375 (2004) 213–218.
[24] Y. Zhang, T.T. Zuo, Z. Tang, M.C. Gao, K.A. Dahmen, P.K. Liaw, Z.P. Lu, Prog. Mater. Sci. 61 (2014) 1–93.
[25] M. Vaidya, G.M. Muralikrishna, B.S.J. Murty, Mater. Res. 34 (2019) 664–686.
[26] Y.S. Huang, L. Chen, H.W. Lui, M.H. Cai, J.W. Yeh, Mater. Sci. Eng. A 457 (2007) 77–83.
[27] V. Dolique, A.L. Thomann, P. Brault, Y. Tessier, P. Gillon, Surf. Coat. Technol. 204 (2010) 1989–1992.
[28] Q.L. Xu, Y. Zhang, S.H. Liu, C.J. Li, C.X. Li, Surf. Coat. Technol. 398 (2020) 126093.
[29] M. Zhang, J.X. Hou, H.J. Yang, Y.Q. Tan, X.J. Wang, X.H. Shi, J.W. Qiao, Int. J. Miner. Metall. Mater. 27 (2020) 1341–1346.
[30] Y. Wei, Y. Fu, Z.M. Pan, Y.C. Ma, H.X. Cheng, Q.C. Zhao, X.G. Li, Int. J. Miner. Metall. Mater. 28 (2021) 915–930.
[31] D.B. Miracle, J.D. Miller, O.N. Senkov, C. Woodward, M.D. Uchic, J. Tiley, Entropy 16 (2014) 494–525.
[32] J. Dąbrowa, G. Cieślak, M. Stygar, K. Mroczka, K. Kulik, T. Berent, M. Danielewski, Intermetallics 84 (2017) 52–61 .

[33] K. Pan, Y. Yang, S. Wei, H. Wu, Z. Dong, Y. Wu, X.J. Mao, Mater. Sci. Technol. 60 (2021) 113–127.
[34] W. Kai, F.P. Cheng, C.Y. Liao, C.C. Li, R.T. Huang, J.J. Kai, Mater. Chem. Phys. 210 (2018) 362–369.
[35] W. Kai, F.P. Cheng, Y.R. Lin, C.W. Chuang, R.T. Huang, D. Chen, C.J.J. Wang, J. Alloy. Compd. 836 (2020) 155518.
[36] A.O. Moghaddam, N.A. Shaburova, M.V. Sudarikov, S.N. Veselkov, O.V. Samoilova, E.A. Trofimov, Vacuum 192 (2021) 110412.
[37] A.O. Moghaddam, M. Sudarikov, N. Shaburova, D. Zherebtsov, V. Zhivulin, I.A. Solizoda, E. Trofimov, J. Alloy. Compd. 897 (2022) 162733.
[38] C.M. Liu, H.M. Wang, S.Q. Zhang, H.B. Tang, A.L. Zhang, J. Alloy. Compd. 583 (2014) 162–169.
[39] W. Kai, C.C. Li, F.P. Cheng, K.P. Chu, R.T. Huang, L.W. Tsay, J.J. Kai, Corros. Sci. 108 (2016) 209–214.
[40] G. Laplanche, U.F. Volkert, G. Eggeler, E.P. George, Oxid. Met. 85 (2016) 629–645.
[41] X. Chen, Y. Sui, J. Qi, Y. He, F. Wei, Q. Meng, Z. Sun, J. Mater. Res. 32 (2017) 2109–2116.
[42] S.K. Dewangan, V. Kumar, Int. J. Refract. Hard Met. 103 (2022) 105777.
[43] P. Ampornrat, G.S. Was, J. Nucl. Mater. 371 (2007) 1–17.
[44] D.B. Miracle, O.N. Senkov, Acta Mater. 122 (2017) 448–511.
[45] R. Tibshirani, J. R. Stat. Soc. B 58 (1996) 267–288.
[46] T.Y. Zhang, B. Cao, S.Y. Zhang, S. Sun, Tree-classifier for linear regression software, No. 2021SR1951267, (2021).
[47] S. Wang, Z. Chen, P. Zhang, K. Zhang, C.L. Chen, B.L. Shen, Vacuum 163 (2019) 263–268.
[48] Y.Y. Liu, Z. Chen, Y.Z. Chen, J.C. Shi, Z.Y. Wang, S. Wang, F. Liu, Vacuum 169 (2019) 108837.
[49] A. Mohanty, J.K. Sampreeth, O. Bembalge, J.Y. Hascoet, S. Marya, R.J. Immanuel, S.K. Panigrahi, Surf. Coat. Technol. 380 (2019) 125028.
[50] T. Chen, C. Guestrin, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16), Association for Computing Machinery, New York, U.S, 2016, pp. 785–794.
[51] S.M. Lundberg, S.I. Lee, in: Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17), Curran Associates Inc., Red Hook, New York, USA, 2017, pp. 4768–4777.
[52] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, E. Duchesnay, J. Mach. Learn. Res. 12 (2011) 2825–2830.
[53] D. Whitlark, G.H. Dunteman, J. Mark. Res. 27 (1990) 243.
[54] G. Louppe, Understanding Random Forests: from Theory to Practice, University of Liege, 2014 Ph.D. Thesis.
[55] H. Liu, M.M. Xu, S. Li, Z.B. Bao, S.L. Zhu, F.H. Wang, J. Mater. Sci. Technol. 54 (2020) 132–143.
[56] P.E.R. Kofstad, Nature 179 (1957) 1362–1363.
[57] M. Uusitalo, P. Vuoristo, T. Mäntylä, Corros. Sci. 46 (2004) 1311–1331.
[58] J.A. Loli, A.R. Chovatiya, Y. He, Z.W. Ulissi, M.P. de Boer, B.A. Webler, Oxid. Met. 98 (2022) 429–450.