Combined approach to capture the evolution of oxidation of Nickel based superalloys using data driven approaches

Nikhil Khatavkar and Abhishek Kumar Singh*
Materials Research Centre, Indian Institute of Science, Bangalore 560012, India

(Received 29 January 2024; revised 26 March 2024; accepted 9 April 2024; published 7 May 2024)

Nickel-based superalloys are an exceptional class of materials that are indispensable for high-temperature applications in the aerospace and power sector industries worldwide. The prolonged application of these materials in a demanding environment is hindered by the increased oxidation rates and deformation due to mass gain at high temperatures and the presence of corrosive agents. Calculating the oxidation properties using experimental techniques is laborious and highly cost/time intensive, which presents a considerable challenge to reducing the oxidation in these materials. In this work, we establish an extensive database consisting of the specific mass gain due to oxidation (Δm) and the parabolic oxidation rates (k_p) of nickel-based superalloys spanning all the superalloy generations. Highly accurate machine learning (ML) models are developed to predict (Δm) using artificial neural networks and tree-based XGBoost. The ML models are extended by unsupervised k means clustering to improve the accuracy of the models and generate insights on the composition-property linkages. Additionally, the ML model for k_p developed utilizing XGBoost yields unprecedented results with errors of 0.04. The ML model is analyzed using the SHapely Additive exPlanations parameters to determine the effect of individual features on the model. Further, we employ a genetic algorithm-based approach utilizing the developed ML models to minimize the k_p to improve the performance of the superalloys at high temperatures. The genetic algorithm-assisted optimization successfully yields several compositions for new Ni superalloys with up to 20% reduction in the k_p. This work presents essential advances for accelerating the targeted discovery of new materials for highly specialized and demanding applications.

DOI: 10.1103/PhysRevMaterials.8.053601

I. INTRODUCTION

Nickel-based superalloys are a special class of alloys used in the combustion turbines of power plants and aircraft [1,2]. The superalloys possess exceptional mechanical properties such as high yield strength, tensile strength, significant creep, and fatigue life [3]. Modern composition and microstructural optimization processes in the past decades have led to alloys that can sustain temperatures over 1100°C [4]. The presence of a continuous γ matrix and γ' precipitate phase is responsible for strengthening the superalloys via precipitate hardening [5,6]. The γ' phase exhibits a $L1_2$ crystal structure and forms coherent interfaces with the fcc-γ matrix phase. The γ/γ' interfaces block the dislocation movement to the γ' phase, thus limiting the dislocations to the narrow γ channels between the precipitates and leading to precipitation hardening [7]. Due to these properties, superalloys are employed in the fuel combustion sections of turbines, where the mechanical and temperature conditions are highly demanding.

Increasing the efficiency of the turbines is highly beneficial for the reduction of fuel usage, the weight of the engine, and emissions. One of the major route of enhancing the efficiency is through increasing the operation temperature of the engine [8,9]. While operating at high temperatures, significant deterioration of materials occurs due to corrosion, leading to deviations in the mechanical performance. The deterioration is experimentally quantified on the surface of superalloy using the specific mass-gain (Δm) (simply refered to as mass-gain hereafter) due to the surface oxide layer. The extent of the corrosion can also be measured via the parabolic rate of oxidation (k_p), in addition to the mass gain. The oxidation kinetics can be described using the Pilling and Bedworth's equation [10]:

$$(\Delta m/A)^2 = k_p t + C, \tag{1}$$

where, A is the area of the oxidised surface and C is a constant. Oxidation may also occur in the intergranular regions, along different grains, or below the external surface. The damage due to oxidation is especially significant in the vicinity of the new surface created by cracks, which are susceptible to the hot gases and particles. The continuous oxidation in these regions can accelerate the crack propagation leading to reduced service lifetimes of components [11].

The detrimental effects of oxidation in the superalloys can be reduced by adding elements such as aluminum and chromium. The addition of aluminum and chromium leads to the formation of a protective scale of Al_2O_3 or Cr_2O_3 [12]. Generally, the scales formed must have a slow growth rate and high adhesion to the superalloy in addition to being continuous on the surface. The slow growth of the oxide layer is essential for maintaining the ideal composition and properties of the superalloys for a longer duration. Alumina scale formation is possible for aluminum concentrations as low as five wt.%. However, formation of Cr_2O_3 requires higher

*abhishek@iisc.ac.in

concentrations of chromium [13]. The addition of silicon is also beneficial for combating hot corrosion attacks. Silicon up to 0.45 wt.% can improve oxidation resistance while having minimal impact on the mechanical properties [14]. Tantalum addition of up to 6 wt.% is found to be beneficial for oxidation resistance due to the formation of $NiTa_2O_6$ [15]. However, increasing tantalum concentration further leads to the degradation of oxidation resistance. In addition, the inclusion of highly reactive elements such as yttrium, hafnium, cerium, and lanthanum in limited quantities of 0.01–0.1 wt.% significantly improves the oxidation resistance. Significant improvement in alumina scale adhesion is also observed due to the presence of these trace elements [12]. In the fourth- and fifth-generation alloys, rhenium and ruthenium are added in significant quantities to enhance the superalloys' mechanical properties [16]. These elements form oxides such as Re_2O_7 and RuO_4, which have high vapor pressures [15,17]. Consequently, these elements are considered detrimental as they prevent the formation of a strong and cohesive alumina layer on the surface of the superalloy. The elements titanium, niobium, molybdenum, tungsten, and cobalt are also generally detrimental to nickel-based superalloys' oxidation properties. The oxidation properties of superalloys are also affected due to the synergistic effects of two or more elements. Tantalum increases the oxidation resistance for lower concentrations of aluminum, however, increasing the aluminum leads to discontinuous alumina scale and increased mass gain [18]. Similarly, the oxidation resistance due to aluminum is reduced due to the addition of tungsten due to the disruption of the oxide layers due to the formation of $NiWO_4$ [19].

The optimization of the oxidation properties is complicated due to the presence of large number of doping elements. The compositional space for the Ni superalloy swiftly becomes intractable experimentally due to the time and experimental costs. Moreover, often the mechanical and oxidation properties have contradictory requirements, which are difficult to optimize experimentally. Therefore, a new approach is needed to accelerate the process of discovery of new superalloy compositions for better oxidation resistance. In recent years, the introduction of artificial intelligence and machine learning to materials science has significantly accelerated new materials discovery and property improvement by learning and identifying essential patterns from past data. Supervised learning is a class of machine learning (ML), that uses alabeled data set to generate a function that maps the input features to the desired target/output value [20]. A good understanding of the past data through supervised machine learning then allows us to make highly accurate predictions of materials or properties for specified usage. Supervised machine learning algorithms have been successfully used to predict a wide range of materials properties [21–27]. Similarly, supervised ML has been successfully applied in superalloys to optimize parameters such as γ' phase fraction, solvus temperature, detrimental phase volume, and processing temperature. [28,29]. ML has also been used to predict the structural properties of superalloys, including the Vickers hardness [30,31], creep rupture lives [32,33]. In addition to this, unsupervised machine learning, which focuses to discover the relationships between the features, is being utilized to classify materials based image data [34], study the dislocation motion [35], and discovering thermoelectric, photovoltaic, and solid state conductor materials [36–38]. Since a large amount of nickel superalloy data is available, machine learning approach is best suited for optimizing the oxidation properties of the superalloys via unsupervised and supervised learning approaches.

In this work, utilizing data from experimental studies, we develop machine learning models for determining the oxidation resistance of superalloys. The ML models are trained to predict the mass gain (Δm) and the parabolic rate (k_p) of oxidation. The mass gain in a superalloy occurs due to the formation of various oxidation products. Mass gain and parabolic rate constant for oxidation are essential parameters that decide the service life of a superalloy component. Initially, ML models for mass gain are developed using artificial neural network and tree based methods. k-means clustering is employed in the data set to partition it into clusters with distinct characteristics. This is done to increase the efficiency and accuracy of the developed ML models. Further, the ML model for k_p developed and analyzed using SHapley Additive eXplanation (SHAP) to identify the essential features. A material design strategy is then discussed for reducing the detrimental effect of oxidation and extending the lifetime of superalloys.

II. METHODOLOGY

Initially, to select only the relevant features from the original dataset of 22 features, we apply least absolute shrinkage and selection operator (LASSO) [39]. LASSO solves the L1-penalty regression problem of finding the individual weights to minimize the error term [40]:

$$\min\left\{\frac{1}{N}||y-\mathbf{wx}||^2+\lambda||\mathbf{w}||_1\right\}, \tag{2}$$

where λ is the penalty term, which decides the degree of shrinkage of individual weights. A more significant value of λ leads to eliminating a larger number of features. After selecting essential features, artificial neural network (ANN) and eXtreme Gradient Boosting (XGBoost) algorithms are employed for training the ML models. Since the database generated consists of large data points, we prefer to employ the ANN and XGBoost as they scale better than the kernel-based algorithms [41].

Artificial neural networks are massively parallelized systems with the interconnection of many primitive units. ANN is modeled after biological neural networks of the brain to mimic the qualities of parallelization, distributed computation, generalization, learning ability, and fault tolerance [42]. ANN learns through detecting patterns and relationships in the data, fed sequentially to the network. The ANN consists of several hidden layers and a single output layer consisting of individual processing units known as neurons [43]. These neurons are the building blocks of ANN, which are interconnected to each other in a specific and rigid architecture [44]. Each layer in the ANN can accommodate several neurons. A schematic diagram for a single neuron is presented in Supplemental Material Fig. S1(a) [45]. The output layer may contain one neuron for regression or several neurons for classification problems. The architecture is feed-forward, which implies that

the data is transferred in a single direction only, as shown in Supplemental Material Fig. S1(b) [45].

Ensemble-based methods are best known for reducing bias components leading to low errors in prediction [46]. XGBoost is a decision tree-based ML model that uses gradient boosting algorithms for enhanced performance. Boosting technique can be used for both classification and regression problems. In boosting, weak learners are generated at each step, accumulating in a final model. When the learner for each step is decided by the direction of the gradient of the loss function, it is called gradient boosting [47]. For a data set with *m* features, a tree ensemble based ML model uses *k* different functions to predict the output given as [48]:

$$\hat{y_i} = \phi(x_i) = \sum_{k=1}^{K} f_k(x_i), \quad f_k \in F, \tag{3}$$

where $F = w_{q(x)}(q : \Re^m \rightarrow T)$ consists of all the regression trees and q is the unique structure of the tree. Each f_k is of a unique tree structure q with T independent leaves. The final prediction is calculated by summation of all the scores in the corresponding leaves. In a normal ensemble tree to learn a set of functions, the ML model minimizes the following objective function [48]:

$$\boldsymbol{L} = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k) \text{ where,}$$
$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \|w\|^2, \tag{4}$$

where, l is the convex loss function measuring the difference between the original value (y_i) and intermediate predicted value ($\hat{y}_i$). The second term in the equation (Ω) penalizes the model complexity. The λ term is a regularization term that smoothens the final weights and reduces overfitting. In gradient boosting, a greedy algorithm is employed that starts with a single leaf and adds complex branches iteratively to the tree [49]. The loss function in this case becomes

$$\boldsymbol{L}_{\text{split}} = \frac{1}{2}\left[\frac{\left(\sum_{i\in I_L} g_i\right)^2}{\sum_{i\in I_L} h_i + \lambda} + \frac{\left(\sum_{i\in I_R} g_i\right)^2}{\sum_{i\in I_R} h_i + \lambda} - \frac{\left(\sum_{i\in I} g_i\right)^2}{\sum_{i\in I} h_i + \lambda}\right] - \gamma, \tag{5}$$

where, g and h denote the first and second gradient statistics derivative of the convex loss function from Eq. (4) and I_L and I_R denote the instances to the left and right of the current split. XGBoost is a highly scalable end-to-end tree boosting system and can be used easily for large data set.

ML models that predict the target property can best explain the role of individual features. However, for complex models such as the ANN and XGBoost, original models cannot be used directly to determine the role of an individual feature. These ML models, however accurate, are essentially black boxes for extracting the target property. They provide no information in the prediction, which can be essential to enhancing understanding of the underlying mechanisms. Therefore, it is important to analyze the ML models to find relevant patterns for further improvement. To identify the relevant features, we employ SHAP analysis to the ML models developed using the data set [50]. SHAP assigns an importance value to individual features which are additive. These SHAP values are obtained by employing cooperative game theory, where the importance of individual features is assessed by the difference between the average prediction of an instance minus the average prediction using all instances. In this work, SHAP as implemented in python is employed on the ML models. We try to understand the complex dependence of the feature values on the oxidation characteristics of superalloys

In order to discover novel superalloys for better oxidation resistance, genetic algorithms (GA) are implemented to deliver novel compositions using the developed ML models. GAs are a subset of evolutionary algorithms, which search for a new set of solutions heuristically from the preexisting solutions [51,52]. The adeptness of the GA to search for a possible solution space is modeled after the Darwinian theory of natural evolution and genetic inheritance. The initial step is the recombination process, which is based on the application of two operators: crossover and mutation [53]. Crossover process is the equivalent of biological transfer of chromosomes. The features from the current best solutions are combined to be transferred to the new population. Further, the combined features are mutated by introducing random variations to enhance the uniqueness of the solutions. The best solutions in the population are tested using a fitness function. The fitness function controls the evolution process and can be used to implement constraints on the new population. Genetic algorithms do not require derivative information, which may not be available for all data sets and makes it immune to gradient-based methods' failure. Moreover, genetic algorithm has good parallelization capabilities and provides many possible solutions instead of a single solution. In this work, genetic algorithm was implemented to minimize the parabolic rate of oxidation using the ML models developed. The composition of individual elements was constrained to be in the range of the initial data set. Moreover, constraints imposed penalties on the entire composition if the sum was not equal to 100%.

III. RESULTS

A. Database validation and standardization

Development of novel superalloys for greater oxidation resistance is important to increase the safety and service life of the turbine components. In addition, increasing the oxidation resistance of the superalloys will reduce the operation costs. The schematic of the entire work is presented in Fig. 1. The availability of considerable volume of data on the oxidation characteristics enables us to perform ML modeling for oxidation resistance. Initially, an extensive superalloy materials database is established by manually collecting the data from literature, consisting of the mass gain and parabolic rate of oxidation [18,54–80]. Since, there is no standard method to report the compositions, and oxidation properties, all the entries of the database are verified individually. The selection of the superalloys is done such that only those compositions are included, which have nickel as the majority element. Further, to make the data unbiased towards any specific generation, we have included the superalloys of all the possible generations (1^{st} to 6^{th}). Our database consists of 2540 entries for mass gain and 875 entries for the parabolic rate constant. Only

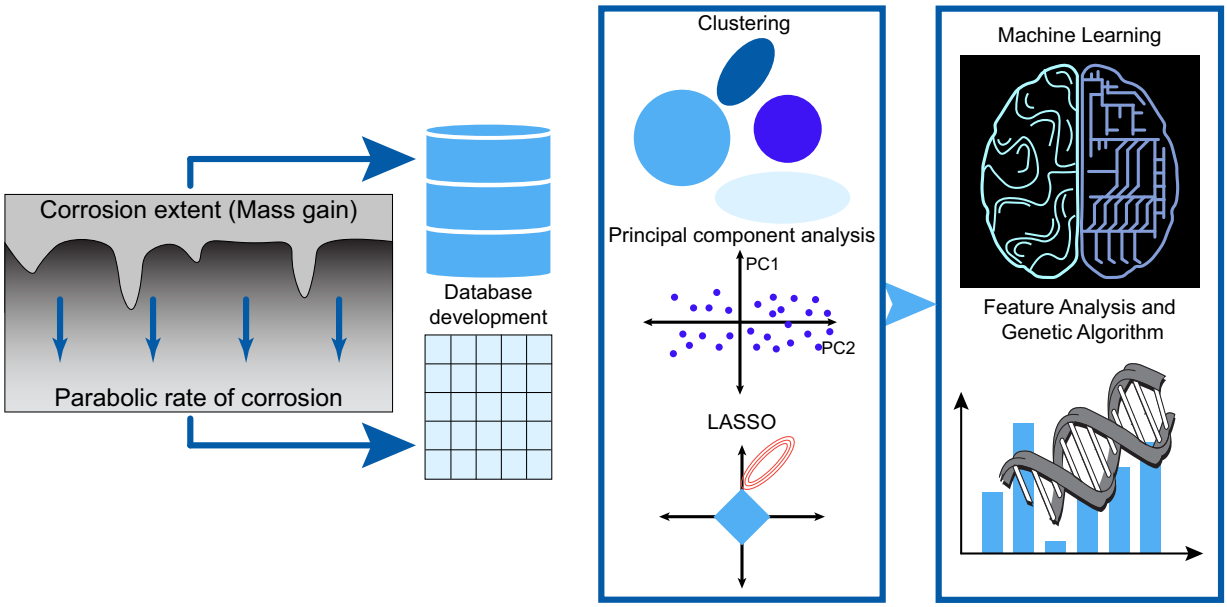

FIG. 1. Schematic of the present work. Initially, a database is established for corrosion properties, followed by feature engineering and machine learning.

basic descriptors were employed in this study to keep the ML models simpler to use, interpret and enhance user-friendliness. The descriptors consist of individual superalloy compositions (wt.%), time (hours), and temperature (°C) of exposure to the corrosive atmosphere.

In literature, the material is subjected to different types of environments for the study of oxidation. The data collected in this study only considers the experiments conducted in air. The database consists of 22 features, and two target properties, which are taken in log scale as the original values span several orders of magnitude. The details of all individual features and target property are presented in supplementary section table S1 [45]. Initially, validation steps were carried out to ascertain that the numerical values of all the database features lie within the acceptable limits. Any data set was considered only if the summation of individual elemental compositions added up to 100. Similarly, the values of time, temperatures, mass gain, and k_p were also checked for consistency in the units. The data set was then standardized using the Standard scalar in the Scikit-learn library [81]. Standard scalar is applied to the features in the data set, where the mean of all the feature distributions is brought to zero with a standard deviation of one. The standardized feature set for mass gain and rate of oxidation (k_p) is presented in Figs. S2(a) and S2(b). Standardization is crucial as it reduces the bias towards the features with larger variance since the features with a higher magnitude of variance may dominate over the other features.

B. ML models for mass gain

Nickel superalloys, subjected to temperatures in excess of 1100°C, experience corrosion and formation of surface oxide layer. The formation of the surface oxide layer leads to increase in the mass, which is measured by the weight gain curves. Since Δm measures the extent of corrosion, the prediction of Δm as a function of temperature is important. We initially employ the artificial neural network algorithm to predict Δm. The data was divided into training and testing data in a ratio of 90:10. To avoid overfitting, the architecture of ANN was fixed to 2 dense layers, with the output of each layer being fed to either the next dense layer or the output layer, as shown in Supplemental Material Fig. S1 [45]. Rectified linear unit, soft-max, or linear activation functions were applied to the individual neuron outputs. The number of neurons in different layers, the activation function of layers, learning rate, and epoch were considered hyperparameters, and hyperparameter optimization of ANN was carried out using the gridsearch algorithm. It was observed in the initial stage that the learning rate had a minor effect on the ANN training for the mass gain, and therefore we chose the default learning rate of 0.01. Next, we checked the effect of the number of epochs in training the ANN models. The ANN models performed poorly when the default/recommended value of epoch was employed. We searched for an optimal number of epochs for the ANN using gridsearch from one to five thousand. The results improved drastically when the number of epochs increased. After 1000 epochs, the results were seen to be converging to a constant value, and any further arbitrary increase does not lead to an increase in the accuracy of ANN models. Therefore for further ANN training, the number of the epoch was fixed to 1000. The best results for models using the ANN are presented in Table I. The results of ANN show good accuracy with acceptable error. The accuracy of the ML models is determined by calculating the root mean squared error (RMSE) and the coefficients of determination (R^2). The root mean squared error is represented as

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}\left(y_t^i - y_p^i\right)^2}{N}}, \tag{6}$$

where y_t and y_p are the true and predicted values of target variable and N is the total number of data set in the model.

TABLE I. The best results for the ANN models are presented along with the number of neurons in each layer. The learning rate and epochs are taken to be 0.01 and 1000, respectively. The R^2 and RMSE for the training and test set are reported.

Layer 1	Layer 2	R^2	RMSE
20	12	0.97/0.97	0.24/0.25
12	20	0.97/0.97	0.23/0.25
16	25	0.97/0.97	0.24/0.26
25	20	0.98/0.97	0.22/0.26
25	18	0.97/0.97	0.23/0.26

Since the error is squared while calculating the RMSE, it gives higher weight to the larger errors and thus is a better indicator of model accuracy than absolute error. The coefficients of determination are defined as the proportion of variance of the dependent variable, which is captured or explained by the combined independent variables. It is defined as

$$R^2 = 1 - \frac{\sum_i \left(y_t^i - y_p^i\right)^2}{\sum_i \left(y_t^i - y_{\text{mean}}\right)^2}. \tag{7}$$

The best result of $R^2 = 0.97/0.97$ and RMSE $= 0.24/0.25$ for the ANN model is obtained for 20 and 12 neurons in the first and second dense layer, respectively. The results for the ANN are similar to the work of Kim *et al.*, where ANN also provided a highly reliable prediction for a smaller data set of oxidation resistance over the response surface methodology [82]. The learning curve for the ANN model is plotted in Fig. 2(a). The trends for predicting the oxidation properties of various materials in literature are similar to our work, which confirms the effectiveness of the ANN models [83]. However, it is also observed that tree-based algorithms also predict the oxidation properties such as mass gain and k_p with equivalent accuracy, provided sufficient data is available [84,85]. Therefore, to explore different algorithms to improve the accuracy of the ML models developed, we trained XGBoost-based ML models for mass gain.

Initially, the XGBoost model was trained on the training data considering all the available compositional and experimental features. Hyperparameter tuning was carried out while training the ML models. For XGBoost, the eta, gamma, maximum depth, and maximum leaf nodes are the important hyperparameters to be tuned. The eta and gamma parameters

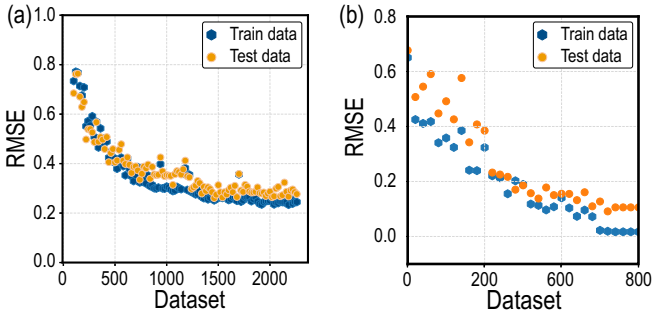

FIG. 2. Learning curve for (a) Δm, (b) k_p. The learning for Δm although converged shows considerable variation for the test data.

are the learning rate and regularization parameter, which control the rate of convergence and prevents overfitting. Max depth and maximum leaf nodes control the complexity of the tree models. The best result obtained for the XGBoost with all features has $R^2 = 0.97/0.97$ and RMSE $= 0.22/0.22$. It can be seen that the ML models provide only a marginal improvement in the prediction of properties. Therefore, to improve the XGBoost model, we partition the data into different clusters using k-means clustering and principal component analysis (PCA). The l LASSO is then applied to the individual clusters to select the most relevant features for ML. K-means clustering is the simplest and most widely used clustering algorithm. It is a form of partitional clustering algorithms where the data is segregated into distinct clusters (subsets) subject to specific clustering criteria [86]. The clustering criteria in K-means clustering is the sum of squared Euclidean distances between the cluster center (centroid) and the individual dataset. The algorithm starts with an initial random guess for partitioning the clusters. The clusters are updated by minimizing the squared distances between the individual dataset to its assigned cluster center. The error term (inertia) is represented as [87]:

$$E(m_1, m_2, ..., m_M) = \sum_{i=1}^{N} \sum_{k=1}^{M} I(x_i \in C_k) ||x_i - m_k||^2, \tag{8}$$

where M and N denote the number of clusters and features, respectively. x_i and m_k denote the coordinates of the i^{th} data set and k^{th} cluster in the feature space. The expression $I(x_i \in C_k)$ is equal to one if the data set x_i lies in the cluster C_k. The centroids of the M clusters are adjusted in each iteration to decrease the inertia term. The clusters converge when either the data reassignment to different clusters stops or the inertia stops decreasing further. The K-means clustering algorithm is straightforward and scales linearly with data [86]. We initially employ PCA on the entire data set to perform clustering efficiently. PCA is a multivariate technique that represents the entire data in a new set of orthogonal coordinate systems. The orthogonal axes in the new coordinate system are known as the principle components. The objective of PCA is to reduce the dimensionality of the data set by first transforming the entire data set into a new orthogonal coordinate system and retaining the axes (principal components) which exhibit maximum data variance. The data can be thus represented as

$$F^j = \sum_{i=1}^{\min(J-1,R)} \alpha_i^j \boldsymbol{I}_i + \bar{F}_i, \tag{9}$$

where, $\boldsymbol{I}_i$ are the principal components obtained and α_i^j are the corresponding coefficients or coordinates of the j^{th} data set. The maximum number of components is limited by either the dimensionality of the original data set J or by an objectively selected truncation level, R. The variance of data captured by successive principal components is presented in Fig. 3(a), where the variance of 80% is deemed sufficient to capture the essential characteristics of the database. Since the first ten principal components can capture 80% variance, therefore, we proceed with them to perform the clustering of the database using the k-means clustering. The default number of clusters generated by the algorithm is eight. However, we perform optimization to identify the ideal

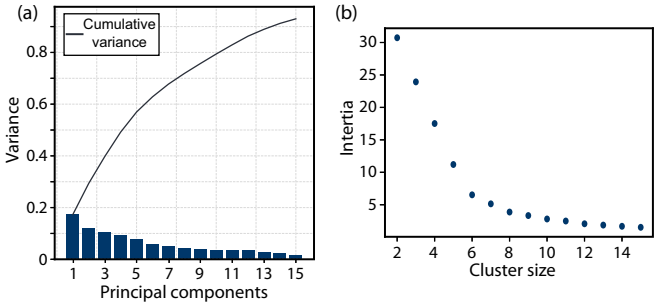

FIG. 3. (a) Variance of data captured by successive principal components. The first ten principal components capture 80% variance, hence they are selected for describing the database, (b) Identification of ideal clustering. The inertia (error) reduces considerably until the 6th cluster. The error reduction after the 6th cluster is minimal, giving rise to the elbow plot.

clustering method since proceeding with the default number of clusters in the algorithm may not give the most accurate result and have any physical significance. The ideal clustering is identified in Fig. 3(b). It can be seen from Fig. 3(b) that the error reduces considerably when the number of clusters is increased from one to six. However, after the six clusters, the error reduction is not significant, indicating that any further increase in clustering is ineffective in partitioning the data. Therefore, we partition the data into six clusters in this work. The clustering of data is presented in Fig. S2. In Fig. S2, the data is plotted in the different principal component directions, and the data is distinctly segregated into six clusters.

After successfully performing clustering on the data, we build ML models on these clusters individually. To develop accurate ML models, selecting highly correlated features to the target property is essential. We employ LASSO feature reduction to select the most relevant features for the individual cluster. LASSO minimizes the sum of square residual (error) subject to the L1 penalty term. Due to the nature of L1 constraint, some of the coefficients of features are zero leading to the elimination of these features during the training of the ML model. More features are eliminated for larger values of λ. We employ gridsearch for the λ parameter of LASSO, and the XGBoost algorithm is employed to train the ML models for all the clusters. The results for the XGBoost models are presented in Table II. The six clusters vary significantly in size, ranging from the largest cluster with 1858 data to the smallest cluster with 20 data. There is also a considerable variation in the individual compositions for different clusters. The ML models developed for clusters one, three, and six are highly accurate with high R^2 and low RMSE for both train and test data. However, the cluster size is insufficient for clusters two, four, and five to develop any meaningful ML models. We train new XGBoost models by combining the data from the three clusters into a new cluster seven. The data for these clusters is presented in Table II. Combining clusters two, four, and five into the new cluster results in a highly accurate ML model with R^2 of 0.99/0.99 and RMSE of 0.01/0.01. The ML models are then analyzed using the Pearson correlation coefficient to check the relevance of each feature in deciding the Δm. The Pearson correlation for all the clusters along with the features is shown in Fig. 4. Time of exposure and temperature are the common features for all the clusters in the present work. Both time and temperature show a significant positive correlation with the mass gain, indicating that an increase in the two features independently leads to a larger extent of oxidation. This trend is similar to the experimental work on Inconel alloys, where an increase in temperature or time of exposure leads to a larger mass gain in the superalloy [54]. For cluster 1, it can be observed that the elements tungsten, molybdenum, cobalt, and titanium show a positive correlation with the mass gain, indicating that the presence of these elements deteriorates the oxidation resistance of the superalloys. Chromium, zirconium, ruthenium, and aluminum negatively correlate with mass gain, leading to increased resistance of superalloy to oxidation. The results are consistent with experimental observations where it is seen that aluminum, chromium, and zirconium are highly beneficial for reducing oxidation and mass gain while molybdenum, tungsten, cobalt, and titanium are detrimental [12,18,88]. For cluster 3, titanium, nickel, niobium, and molybdenum show a positive correlation, while aluminum shows a negative correlation to the mass gain. The composition of chromium in cluster 3 shows a positive correlation, which has been attributed to the minimal variety of its composition in the data set of cluster 3. Therefore, the correlation of chromium is statistically insignificant and can be ignored. In cluster 6, the composition of tungsten, vanadium, and nickel shows a positive correlation while aluminum shows a negative correlation to the mass gain, which resembles the experimental results found in literature [12,89]. For cluster 7, the composition of nickel and silicon shows a positive correlation to mass gain, implying deterioration of oxidation resistance properties with increasing concentrations of the elements. A common trend is visible in all the ML models of the four clusters. The compositions of cerium, lanthanum, and gadolinium are either absent or have a very poor correlation with the oxidation property. This is due to the fact that these rare earth elements are added to the superalloys in trace amounts, which the ML models are unable to delineate properly.

TABLE II. The best results for the XGBoost models for various clusters. The R^2 and RMSE for the training and test set are reported.

Cluster	Size	R^2	RMSE	Features
1	1858	0.99/0.99	0.11/0.11	19
2	74	0.76/0.82	0.11/0.11	3
3	229	0.93/0.93	0.22/0.23	10
4	141	0.96/0.96	0.14/0.14	3
5	20	0.81/0.84	0.21/0.16	6
6	205	0.96/0.91	0.23/0.23	9
7 (2+4+5)	235	0.99/0.99	0.01/0.04	7

C. ML models for parabolic rate constant

After successfully developing accurate ML models for mass gain, which depict the extent of the oxidation, it is essential to develop ML models that describe the rate of oxidation in superalloys. To make the ML models for k_p consistent with the ML model for mass gain, we use the same feature set for

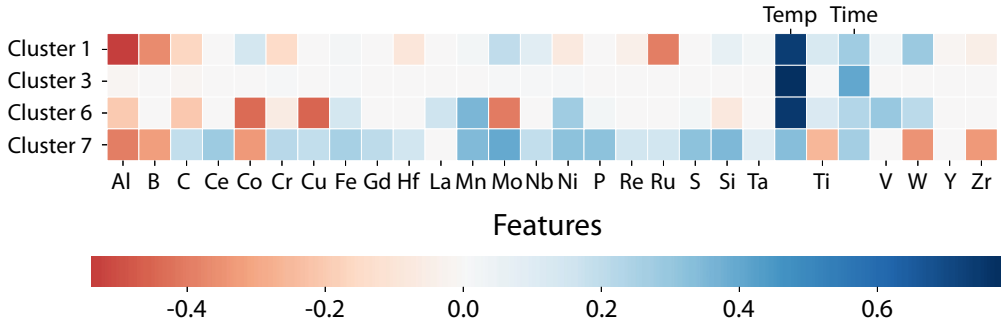

FIG. 4. Correlation of target property with the features for all the clusters. Blue (red) represent positive (negative) correlation of the feature of the cluster with the target property. Time and temperature are important parameters as the correlation is largley positive in all the clusters.

training the k_p ML model. After the validation of the collected data, 787 entries for k_p are chosen. Initially, LASSO is employed for selecting the essential features for predicting k_p. Further, the XGBoost algorithm is applied to develop the ML models. The hyperparameters of LASSO and XGBoost are optimized using the gridsearch algorithm. The best ML model presents high accuracy with R^2 of 0.99/0.99 and RMSE of 0.03/0.04. Learning curves for the XGBoost model presented in Fig. 2(b) confirm that no overfitting occurs for the model.

To better understand the XGBoost model and analyze the role of individual features in the prediction of k_p, analysis is performed for the XGBoost ML model using SHAP. SHAP is useful to understand and gather inferences directly from the ML model as opposed to the Pearson correlation coefficients, which present the correlation between the features and the target property. The effect of individual feature on the k_p as predicted by the XGBoost model is presented as summary plot in Fig. 5. In Fig. 5, the scaled value of individual feature is plotted against the SHAP value. The SHAP value indicates the magnitude of positive or negative impact a feature has on the output/ target property. Similar to the ML models for mass gain, temperature shows a positive correlation with k_p indicating that temperature is one of the dominating features for prediction of k_p. Positive correlation indicates that an increase in temperature leads to the increase in k_p. The distribution of temperature is much significant when compared to the other features, which implies that temperature has the most profound effect on increase of k_p. The SHAP value for temperature increase as the temperature increases (presented in red), indicating a positive correlation with k_p, which is detrimental for oxidation. The distribution of time reveals that the SHAP value is larger for short duration and smaller for longer duration, implying. This strongly agrees with experimental studies where the k_p decreases due to the formation of protective oxide coatings [90–92]. The elements aluminium, chromium, tungsten, tantalum, yttrium, hafnium show a negative correlation with k_p, implying that addition of these elements are beneficial to decrease the k_p. Iron, niobium, molybdenum, cobalt show a positive correlation which is detrimental for oxidation resistance. The correlation for the elements is in agreement with the experimental observations [12]. An important aspect of the correlations for k_p, similar to the mass gain ML models, is the correlation for the rare earth elements such as lanthanum and cerium. These materials show a negative correlation with the oxidation rate. It is experimentally established that lanthanum and cerium are highly beneficial for the oxidation resistance while ruthenium is detrimental [93,94]. These rare earth elements are added in trace quantities, which leads to difficulties in establishing accurate correlations. The effects of time and temperature on

TABLE III. The best parameters for genetic algorithm.

Parameter	Value
Max iteration	100
Size	100
Mutation probability	0.1
Elite ratio	0.05
Crossover probability	0.5
Parents portion	0.3

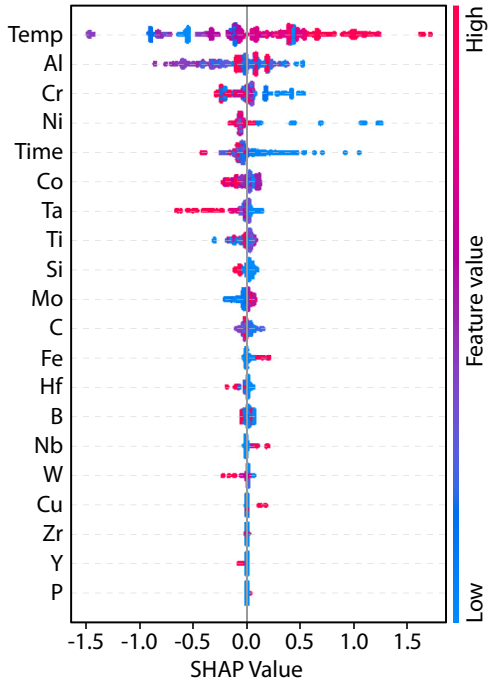

FIG. 5. Summary plot for k_p derived from SHAP. It represents the relation of the feature with k_p as described by the best ML model for k_p.

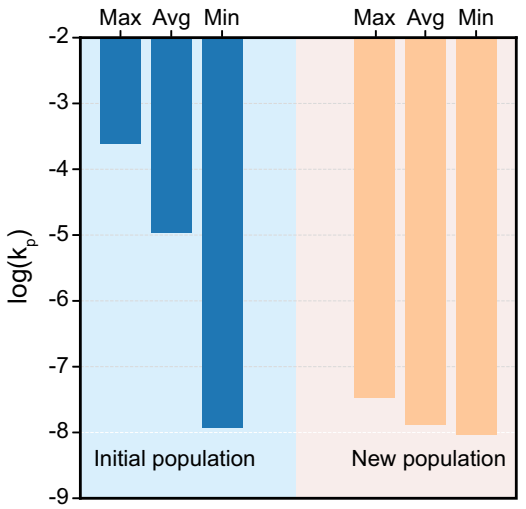

FIG. 6. Optimization of k_p using genetic algorithm. The maximum, average and minimum values of k_p for the initial and new population are presented. The average k_p for the new population is considerably low as compared to the initial population. Further, the minima obtained for the new population is also an improvement over the initial population.

k_p are well known. However, using SHAP values, we have determined the effect of each element on k_p, which is highly cost and labour intensive if done experimentally. The trends obtained using the ML model can now be employed to search for superalloy compositions with reduced k_p.

D. Oxidation rate optimization

After developing accurate ML models for oxidation and analyzing the trends for individual feature, we utilize genetic algorithm to search for new compositions of superalloys to minimize the oxidation in superalloys. Since the optimization of the property is based on the prediction of ML model, it is important that we utilize the best ML model. Since, there is a huge variation in the accuracy of the ML models for different clusters for mass gain, the ML model developed for the k_p is selected to optimize the oxidation properties. To optimize the composition of all elements present, we retrain the XGBoost ML model for k_p by including all the elements and using the same hyperparameters. The new ML model developed has R^2 of 0.99/0.99 and RMSE of 0.03/0.04. The overview of the entire dataset for parabolic oxidation rate is presented in Supplemental Table 1. The input range of the features are taken as the bounds for the optimization process. The fitness function, used to evaluate the performance of the new population, is defined such that any offspring with composition not equal to 100 is penalized. Similarly, penalty is imposed for less value of time and temperature. The optimization of hyperparameters is carried out iteratively and values giving the best results are provided in Table III. The new population obtained by GA shows improvement in the rate of oxidation results.

The compositions and the corresponding values of k_p are presented in the Supplemental Table 2 [45], while the gain in the results is presented in the Fig. 6. The optimization leads to the reduction of the k_p by 20%, which is unprecedented and can help in extending the life cycle of the superalloy components significantly. The new population obtained contains several new superalloys with k_p less than the current minimum in the database hence providing new options for exploring superalloys with improved oxidation resistance.

These new superalloy compositions predicted by the GA reveal interesting trends. It is widely known that addition of rhenium is important for enhancing the creep resistance properties. The successive generations of Ni-superalloys are based on the addition of rhenium [95]. However, the oxidation resistance of the superalloys suffer greatly due to rhenium [17,96]. In the new composition predicted by GA, the concentration of rhenium decreases significantly to enhance the oxidation resistance. Further, the concentration of the element tungsten is increased which can help to retain the creep resistance by inducing a negative lattice misfit [97,98]. The increase of ruthenium in the new superalloys can be beneficial to increase the microstructure stability and suppress the Topologically close-packed phase formation due to rhenium, tungsten and, other refractory elements. The concentrations of aluminium and chromium, which are highly beneficial for the oxidation resistance, are increased further, while the concentrations of Ta, Co, and Mo are reduced. The values of the k_p for the different compositions are for higher temperatures range. Therefore, the superalloy compositions in this work represent an advancement for developing novel superalloys with better oxidation resistance.

IV. CONCLUSIONS

We have established an extensive superalloy materials database with oxidation properties. The database consists of the alloy compositions, testing conditions, and the individual alloy's mass gain and parabolic rate constants. We utilize this database to develop ML models to predict oxidation mass gain and parabolic rate constant. The data for mass gain is first partitioned into six clusters to segregate the data with similar behavior and obtain better prediction accuracy. The best XGBoost ML model shows high accuracy with a minimum RMSE of 0.04. The ML model developed for k_p also shows great prediction accuracy with RMSE of 0.04. SHAP analysis of the k_p models elucidates the effect of different elements, including the minority alloying elements and the experiment conditions on the oxidation characteristics of the superalloy. The trends obtained for the elements are in agreement with the experimental observations. Further, utilizing genetic algorithm, we successfully reduce the parabolic rate of oxidation by 20% which is unprecedented. In this work, we have successfully demonstrated the utilization of supervised/unsupervised learning and genetic algorithms to accelerate new materials discovery and design new superalloy compositions for better oxidation resistance and extended service life.

The data and machine learning scripts of this study are available from the corresponding author, upon reasonable request.

ACKNOWLEDGMENTS

The authors acknowledge the financial support from Aeronautics Research Development Board

(ARDB), Ministry of Defence, India under Grant No. ARDB/GTMAP/01/2031993/M/I. The authors acknowledge the support from Materials Informatics Initiative of IISc (MI3). The authors acknowledge the support from the Institute of Eminence (IoE) scheme of The Ministry of Human Resource Development (MHRD), Government of India. The authors thank Materials Research Centre (MRC) and Supercomputer Education and Research Centre (SERC) at Indian Institute of Science, Bangalore for providing the required computational facilities. N.K. acknowledges S. Swetlana and M. Dey of Materials Research Centre, IISC, for their valuable insights and suggestions.

[1] M. Heilmaier, M. Krüger, H. Saage, J. Rösler, D. Mukherji, U. Glatzel, R. Völkl, R. Hüttner, G. Eggeler, C. Somsen *et al.*, Metallic materials for structural applications beyond nickel-based superalloys, JOM **61**, 61 (2009).

[2] G. R. Thellaputta, P. S. Chandra, and C. Rao, Machinability of nickel based superalloys: a review, Materials Today: Proceedings **4**, 3712 (2017).

[3] R. C. Reed, *The Superalloys: Fundamentals and Applications* (Cambridge University Press, Cambridge, 2008).

[4] E. Akca and A. Gürsel, A review on superalloys and IN_{718} nickel-based INCONEL superalloy, Periodicals of Engineering and Natural Sciences **3**, 15 (2015).

[5] S. Zhao, X. Xie, G. D. Smith, and S. J. Patel, Gamma prime coarsening and age-hardening behaviors in a new nickel base superalloy, Mater. Lett. **58**, 1784 (2004).

[6] S. Semiatin, S. Kim, F. Zhang, and J. Tiley, An investigation of high-temperature precipitation in powder-metallurgy, gamma/gamma-prime nickel-base superalloys, Metall. Mater. Trans. A **46**, 1715 (2015).

[7] W. Xia, X. Zhao, L. Yue, and Z. Zhang, Microstructural evolution and creep mechanisms in ni-based single crystal superalloys: A review, J. Alloys Compd. **819**, 152954 (2020).

[8] J. J. Lee, Can we accelerate the improvement of energy efficiency in aircraft systems? Energy Convers. Manage. **51**, 189 (2010).

[9] U. Schumann, Influence of propulsion efficiency on contrail formation, Aerosp. Sci. Technol. **4**, 391 (2000).

[10] S. Mrowec and A. Stokłosa, Calculations of parabolic rate constants for metal oxidation, Oxid. Met. **8**, 379 (1974).

[11] A. Stankowski, *Turbo Expo: Power for Land, Sea, and Air* (ASME, New York, 2002), Vol. 36088, pp. 1181–1195.

[12] R. Darolia, Development of strong, oxidation and corrosion resistant nickel-based superalloys: Critical review of challenges, progress and prospects, Int. Mater. Rev. **64**, 355 (2019).

[13] J. Stringer, High-temperature corrosion of superalloys, Mater. Sci. Technol. **3**, 482 (1987).

[14] A. Yeh, K. Kawagishi, H. Harada, T. Yokokawa, Y. Koizumi, T. Kobayashi, D. Ping, J. Fujioka, and T. Suzuki, Development of Si-bearing 4th generation Ni-base single crystal superalloys, Carbon NY **20**, 13 (2008).

[15] K. Kawagishi, A. Sato, A. Sato, T. Kobayashi, and H. Harada, Oxidation behavior of Ru-containing Ni-base single-crystal superalloys, *High-Temperature Oxidation and Corrosion*, Materials Science Forum, Vol. 522-523 (Trans Tech Publications Ltd., 2006), pp. 317–322.

[16] W. Xia, X. Zhao, L. Yue, and Z. Zhang, A review of composition evolution in ni-based single crystal superalloys, J. Mater. Sci. Technol. **44**, 76 (2020).

[17] K. Kawagishi, H. Harada, A. Sato, A. Sato, and T. Kobayashi, The oxidation properties of fourth generation single-crystal nickel-based superalloys, JOM **58**, 43 (2006).

[18] S.-J. Park, S.-M. Seo, Y.-S. Yoo, H.-W. Jeong, and H. Jang, Effects of Al and Ta on the high temperature oxidation of Ni-based superalloys, Corrosion Sci. **90**, 305 (2015).

[19] Y. Jin, J. Pei, C. Li, H. Li, Z. Wang, and Y. Liu, Effect of Cr and W addition on the oxidation behavior of Ni-8% Al alloy at 1000 °C, Vacuum **200**, 111044 (2022).

[20] R. Ramprasad, R. Batra, G. Pilania, A. Mannodi-Kanakkithodi, and C. Kim, Machine learning in materials informatics: Recent applications and prospects, npj Comput. Mater. **3**, 54 (2017).

[21] A. C. Rajan, A. Mishra, S. Satsangi, R. Vaish, H. Mizuseki, K.-R. Lee, and A. K. Singh, Machine-learning-assisted accurate band gap predictions of functionalized mxene, Chem. Mater. **30**, 4031 (2018).

[22] A. Mishra, S. Satsangi, A. C. Rajan, H. Mizuseki, K.-R. Lee, and A. K. Singh, Accelerated data-driven accurate positioning of the band edges of mxenes, J. Phys. Chem. Lett. **10**, 780 (2019).

[23] R. Juneja, G. Yumnam, S. Satsangi, and A. K. Singh, Coupling the high-throughput property map to machine learning for predicting lattice thermal conductivity, Chem. Mater. **31**, 5145 (2019).

[24] R. Juneja and A. K. Singh, Guided patchwork kriging to develop highly transferable thermal conductivity prediction models, J. Phys.: Mater. **3**, 024006 (2020).

[25] R. Kumar and A. K. Singh, Chemical hardness-driven interpretable machine learning approach for rapid search of photocatalysts, npj Comput. Mater. **7**, 197 (2021).

[26] R. K. Barik and A. K. Singh, Accelerated discovery of the valley-polarized quantum anomalous Hall effect in mxenes, Chem. Mater. **33**, 6311 (2021).

[27] M. Mukherjee, S. Satsangi, and A. K. Singh, A statistical approach for the rapid prediction of electron relaxation time using elemental representatives, Chem. Mater. **32**, 6507 (2020).

[28] P. Liu, H. Huang, S. Antonov, C. Wen, D. Xue, H. Chen, L. Li, Q. Feng, T. Omori, and Y. Su, Machine learning assisted design of γ' - strengthened cobase superalloys with multiperformance optimization, npj Comput. Mater. **6**, 1 (2020).

[29] Z. Qin, Z. Wang, Y. Wang, L. Zhang, W. Li, J. Liu, Z. Wang, Z. Li, J. Pan, L. Zhao *et al.*, Phase prediction of Ni-base superalloys via high-throughput experiments and machine learning, Mater. Res. Lett. **9**, 32 (2021).

[30] S. Swetlana, N. Khatavkar, and A. K. Singh, Development of Vickers hardness prediction models via microstructural analysis and machine learning, J. Mater. Sci. **55**, 15845 (2020).

[31] N. Khatavkar, S. Swetlana, and A. K. Singh, Accelerated prediction of Vickers hardness of Co-and Ni-based superalloys from microstructure and composition using advanced image processing techniques and machine learning, Acta Mater. **196**, 295 (2020).

[32] Y. Liu, J. Wu, Z. Wang, X.-G. Lu, M. Avdeev, S. Shi, C. Wang, and T. Yu, Predicting creep rupture life of ni-based single crystal superalloys using divide-and-conquer approach based machine learning, Acta Mater. **195**, 454 (2020).

[33] X. Hu, J. Wang, Y. Wang, J. Li, Z. Wang, Y. Dang, and Y. Gu, Two-way design of alloys for advanced ultra supercritical plants based on machine learning, Comput. Mater. Sci. **155**, 331 (2018).

[34] R. Cohn and E. Holm, Unsupervised machine learning via transfer learning and *k* means clustering to classify materials image data, Integr. Mater. Manuf. Innov. **10**, 231 (2021).

[35] D. C. Pagan, T. Q. Phan, J. S. Weaver, A. R. Benson, and A. J. Beaudoin, Unsupervised learning of dislocation motion, Acta Mater. **181**, 510 (2019).

[36] X. Jia, Y. Deng, X. Bao, H. Yao, S. Li, Z. Li, C. Chen, X. Wang, J. Mao, F. Cao *et al.*, Unsupervised machine learning for discovery of promising half-Heusler thermoelectric materials, npj Comput. Mater. **8**, 34 (2022).

[37] Z. Wang, J. Cai, Q. Wang, S. Wu, and J. Li, Unsupervised discovery of thin-film photovoltaic materials from unlabeled data, npj Comput. Mater. **7**, 1 (2021).

[38] Y. Zhang, X. He, Z. Chen, Q. Bai, A. M. Nolan, C. A. Roberts, D. Banerjee, T. Matsunaga, Y. Mo, and C. Ling, Unsupervised discovery of solid-state lithium ion conductors, Nat. Commun. **10**, 1 (2019).

[39] R. Tibshirani, Regression shrinkage and selection via the lasso, J. Royal Stat. Soc. B **58**, 267 (1996).

[40] R. Tibshirani, Regression shrinkage and selection via the lasso: a retrospective, J. R. Stat. Soc. B **73**, 273 (2011).

[41] E. Schulz, M. Speekenbrink, and A. Krause, A tutorial on Gaussian process regression: Modelling, exploring, and exploiting functions, J. Math. Psychol. **85**, 1 (2018).

[42] A. K. Jain, J. Mao, and K. M. Mohiuddin, Artificial neural networks: A tutorial, Computer **29**, 31 (1996).

[43] U. M. R. Paturi, S. Cheruku, and S. R. Geereddy, Process modeling and parameter optimization of surface coatings using artificial neural networks (ANNS): State-of-the-art review, Mater. Today: Proceedings **38**, 2764 (2021).

[44] S. Agatonovic-Kustrin and R. Beresford, Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research, J. Pharm. Biomed. Anal. **22**, 717 (2000).

[45] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevMaterials.8.053601 for details of the data and Supplemental figures.

[46] R. Mitchell and E. Frank, Accelerating the XGBoost algorithm using GPU computing, PeerJ Computer Science **3**, e127 (2017).

[47] J. H. Friedman, Stochastic gradient boosting, Computational Statistics & Data Analysis **38**, 367 (2002).

[48] T. Chen and C. Guestrin, XGBoost: A scalable tree boosting system, in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (ACM, 2016), pp. 785–794.

[49] B. Pan, Application of XGBoost algorithm in hourly PM2.5 concentration prediction, in *IOP Conference Series: Earth and Environmental Science* (IOP, London, 2018), Vol. 113, p. 012127.

[50] S. M. Lundberg and S.-I. Lee, A unified approach to interpreting model predictions, *Advances in Neural Information Processing Systems*, Vol. 30 (NIPS) (Curran Associates, Inc., 2017), pp. 1–10.

[51] T. Bhoskar, O. K. Kulkarni, N. K. Kulkarni, S. L. Patekar, G. Kakandikar, and V. Nandedkar, Genetic algorithm and its applications to mechanical engineering: A review, Mater. Today: Proceedings **2**, 2624 (2015).

[52] W. Paszkowicz, Genetic algorithms, a nature-inspired tool: A survey of applications in materials science and related fields: Part II, Mater. Manuf. Process. **28**, 708 (2013).

[53] C. Z. Janikow, A knowledge-intensive genetic algorithm for supervised learning, *Genetic Algorithms for Machine Learning* (Springer, Berlin, 1993), pp. 33–72.

[54] D. Saber, I. S. Emam, and R. Abdel-Karim, High temperature cyclic oxidation of Ni based superalloys at different temperatures in air, J. Alloys Compd. **719**, 133 (2017).

[55] C. Jang, K. Daejong, K. Donghoon, S. Injin, R. Woo-Seog, and Y.-s. Yoo, Oxidation behaviors of wrought nickel-based superalloys in various high temperature environments, TNMSC **21**, 1524 (2011).

[56] S.-J. Park, S.-M. Seo, Y.-S. Yoo, H.-W. Jeong, and H. Jang, Effects of Cr, W, and Mo on the high temperature oxidation of Ni-based superalloys, Materials **12**, 2934 (2019).

[57] S.-J. Park, S.-M. Seo, Y.-S. Yoo, H.-W. Jeong, and H. Jang, Statistical study of the effects of the composition on the oxidation resistance of Ni-based superalloys, J. Nanomater. **2015**, 929546 (2015).

[58] C. Juillet, A. Oudriss, J. Balmain, X. Feaugas, and F. Pedraza, Characterization and oxidation resistance of additive manufactured and forged IN718 Ni-based superalloys, Corrosion Sci. **142**, 266 (2018).

[59] G. Yan, W. Yu, and S. Shengping, Oxidation protection of enamel coated Ni based superalloys: Microstructure and interfacial reaction, Corrosion Sci. **173**, 108760 (2020).

[60] W. J. Nowak, Characterization of oxidized Ni-based superalloys by gd-oes, J. Analytic. Atomic Spectrom. **32**, 1730 (2017).

[61] F. Weng, H. Yu, C. Chen, and K. Wan, High-temperature oxidation behavior of Ni-based superalloys with Nb and Y and the interface characteristics of oxidation scales, Surf. Interface Anal. **47**, 362 (2015).

[62] J. Cao, J. Zhang, R. Chen, Y. Ye, and Y. Hua, High temperature oxidation behavior of Ni-based superalloy GH202, Mater. Charact. **118**, 122 (2016).

[63] L. Liu, S. Wu, Y. Chen and S. Lu, Oxidation behavior of Re-modified nickel-based superalloy between 950° C and 1150° C in air, TNMSC **26**, 1163 (2016).

[64] J. Wang, M. Chen, S. Zhu, and F. Wang, Ta effect on oxidation of a nickel-based single-crystal superalloy and its sputtered nanocrystalline coating at 900 − 1100° C, Appl. Surf. Sci. **345**, 194 (2015).

[65] O. Gupta, D. Mudgal, D. Puri, and S. Prakash, High temperature cyclic oxidation of Ni based superalloys at 900° C in air, International Journal of Advanced Scientific Research and Technology **2**, 486 (2012).

[66] H. Mallikarjuna, N. Richards, and W. Caley, Isothermal oxidation comparison of three Ni-based superalloys, J. Mater. Eng. Perform. **26**, 2014 (2017).

[67] F. Weng, H. Yu, C. Chen, and K. Wan, Fabrication of Ni-based superalloys containing Nb and their high tempera-

ture oxidation behaviors, Mater. Manufact. Process. **30**, 1364 (2015).

[68] J.-D. Cao, J.-S. Zhang, Y.-Q. Hua, Z. Rong, R.-F. Chen, and Y.-X. Ye, High temperature oxidation behavior of Ni-based superalloy GH586 in air, Rare Metals **36**, 878 (2017).

[69] J. Jiang, G. Xiao, Y. Wang, and Y. Liu, High temperature oxidation behavior of the wrought Ni-based superalloy GH4037 in the solid and semi-solid state, J. Alloys Compd. **784**, 394 (2019).

[70] D. Mudgal, S. Singh, S. Prakash *et al.*, High temperature cyclic oxidation behavior of Ni and Co based superalloys, JMMCE **11**, 211 (2012).

[71] D. W. Yun, S. M. Seo, H. W. Jeong, and Y. S. Yoo, The effect of Gd addition on the cyclic oxidation behavior and creep life of alumina-forming Ni-based superalloy, Corrosion Sci. **170**, 108694 (2020).

[72] S.-J. Park, S.-M. Seo, Y.-S. Yoo, H.-W. Jeong, and H. Jang, Effects of Ti on high temperature oxidation of Ni-based superalloys, Corrosion Sci. Technol. **15**, 129 (2016).

[73] B. Gao, L. Wang, Y. Liu, X. Song, S. Yang, and A. Chiba, High temperature oxidation behaviour of γ'-strengthened Co-based superalloys with different Ni addition, Corrosion Sci. **157**, 109 (2019).

[74] S. M. Das, M. P. Singh, and K. Chattopadhyay, Evolution of oxides and their microstructures at 800° C in a γ-γ'stabilised Co-Ni-Al-Mo-Ta superalloy, Corrosion Sci. **155**, 46 (2019).

[75] A. Evangelou, K. Soady, S. Lockyer, N. Gao, and P. Reed, Oxidation behaviour of single crystal nickel-based superalloys: Intermediate temperature effects at 450 − 550° C, Mater. Sci. Technol. **34**, 1679 (2018).

[76] Y.-x. Zhu, C. Li, Y.-c. Liu, Z.-q. Ma, and H.-y. Yu, Effect of Ti addition on high-temperature oxidation behavior of Co–Ni-based superalloy, J. Iron SteelRes. Int. **27**, 1179 (2020).

[77] Y. Hu, T. Cao, C. Cheng, L. Zhang, and J. Zhao, Oxidation behavior of a single-crystal Ni-based superalloy over the temperature range of 850° C − 950° C in air, Appl. Surf. Sci. **484**, 209 (2019).

[78] M. Taylor, H. Evans, S. Stekovic, and M. Hardy, The oxidation characteristics of the nickel-based superalloy, RR1000, at temperatures of 700–900° C, Mater. High Temp. **29**, 145 (2012).

[79] X. Song, W. Lei, L. Yang *et al.*, Effects of temperature and rare earth content on oxidation resistance of Ni-based superalloy, Prof. Nat. Sci.: Mater. Int. **21**, 227 (2011).

[80] H. Pei, Z. Wen, and Z. Yue, Long-term oxidation behavior and mechanism of DD6 ni-based single crystal superalloy at 1050° C and 1100° C in air, J. Alloys Compd. **704**, 218 (2017).

[81] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, Scikit-learn: Machine learning in Python, J. Machine Learn. Res. **12**, 2825 (2011).

[82] H.-S. Kim, S.-J. Park, S.-M. Seo, Y.-S. Yoo, H.-W. Jeong, and H. Jang, Regression analysis of high-temperature oxidation of Ni-based superalloys using artificial neural network, Corrosion Sci. **180**, 109207 (2021).

[83] L. B. Coelho, D. Zhang, Y. Van Ingelgem, D. Steckelmacher, A. Nowé, and H. Terryn, Reviewing machine learning of corrosion prediction in a data-oriented perspective, npj Mater. Degrad. **6**, 1 (2022).

[84] L. Yan, Y. Diao, Z. Lang, and K. Gao, Corrosion rate prediction and influencing factors evaluation of low-alloy steels in marine atmosphere using machine learning approach, Sci. Technol. Adv. Mater. **21**, 359 (2020).

[85] Y. Zhi, T. Yang, and D. Fu, An improved deep forest model for forecast the outdoor atmospheric corrosion rate of low-alloy steels, J. Mater. Sci. Technol. **49**, 202 (2020).

[86] A. K. Jain, M. N. Murty, and P. J. Flynn, Data clustering: a review, ACM Comput. Surv. **31**, 264 (1999).

[87] A. Likas, N. Vlassis, and J. J. Verbeek, The global k-means clustering algorithm, Pattern Recog. **36**, 451 (2003).

[88] R. A. MacKay, T. P. Gabb, J. L. Smialek, and M. V. Nathal, A new approach of designing superalloys for low density, JOM **62**, 48 (2010).

[89] C. A. Barrett, *A Statistical Analysis of Elevated Temperature Gravimetric Cyclic Oxidation Data of 36 Ni-and Co-base Superalloys Based on an Oxidation Attack Parameter*, Tech. Rep. (NASA, 1992), pp. 1–50.

[90] F. Pettit, G. Meier, M. Gell, C. Kartovich, R. Bricknel, W. Kent, and J. Radovich, Oxidation and hot corrosion of superalloys, Superalloys **85**, 651 (1984).

[91] N. N. Aung and X. Liu, Effect of temperature on coal ash hot corrosion resistance of inconel 740 superalloy, Corrosion Science **82**, 227 (2014).

[92] S. Lee, W. Young, and G. Vermes, Evaluation of additives for prevention of high temperature corrosion of superalloys in gas turbines, J. Eng. Power. **95**, 333 (1973).

[93] W. Jian, Y. Liang, H. Rui, Z. Tiebang, and L. Jinshan, Effect of lanthanum on oxidation behavior of Ni-20Cr-18W-1Mo alloys at 1373 K for 100 h in air, Rare Metal Mater. Eng. **43**, 2060 (2014).

[94] Z. Shi, L. Shizhong, H. Mei, and L. Jiarong, Influence of yttrium addition on high temperature oxidation resistance of single crystal superalloy, J. Rare Earth. **31**, 795 (2013).

[95] M. Huang and J. Zhu, An overview of rhenium effect in single-crystal superalloys, Rare Metals **35**, 127 (2016).

[96] A. Akhtar, S. Hegde, and R. Reed, The oxidation of single-crystal nickel-based superalloys, JOM **58**, 37 (2006).

[97] A. Sato, A.-C. Yeh, T. Kobayashi, T. Yokokawa, H. Harada, T. Murakumo, and J. Zhang, Fifth generation Ni based single crystal superalloy with superior elevated temperature properties, Energy Mater. **2**, 19 (2007).

[98] D. Jithin S., S. K. Selvaraj, K. Vijay Vishaal, and G. Sundaramali, Comparative analysis between 5th and 6th generation superalloys and previous generation superalloys, Adv. Mater. Sci. Eng. **2022**, 3530689 (2022).