

Machine Learning-Assisted Design of Yttria-Stabilized Zirconia Thermal Barrier Coatings with High Bonding Strength

Pengcheng Xu, Can Chen, Shuizhou Chen, Wencong Lu,* Quan Qian,* and Yi Zeng*

Cite This: *ACS Omega* 2022, 7, 21052–21061

Read Online

ACCESS |



Metrics & More

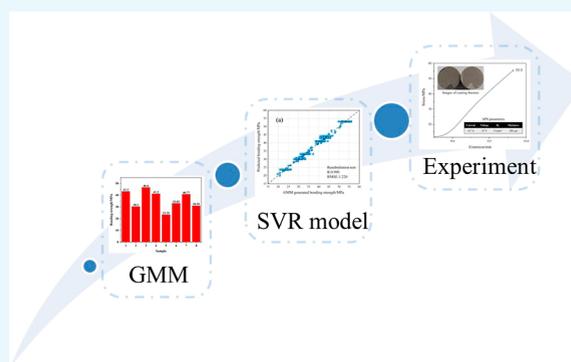


Article Recommendations



Supporting Information

ABSTRACT: As a high-quality thermal barrier coating material, yttria-stabilized zirconia (YSZ) can effectively reduce the temperature of the collective materials to be used on the surface of gas turbine hot-end components. The bonding strength between YSZ and the substrate is also one of the most important factors for the applications. Herein, the Gaussian mixture model (GMM) and support vector regression (SVR) were used to construct a machine learning model between YSZ coating bonding strength and atmospheric plasma spraying (APS) process parameters. First, GMM was used to expand the original 8 data points to 400 with the *R* value of leave-one-out cross-validation improved from 0.690 to 0.990. Then, the specific effects of APS process parameters were explored through Shapley additive explanations and sensitivity analysis. Principal component analysis was used to explain the constructed model and obtain the optimized area with a high bonding strength. After experimental validation, the results showed that under the APS process parameters of a current of 617 A, a voltage of 65 V, a H₂ flow of 3 L min⁻¹, and a thickness of 200 μm, the bonding strength increased by more than 19% to 55.5 MPa compared with the original maximum value of 46.6 MPa, indicating that the constructed GMM–SVR model can accurately predict the bonding strength of YSZ coating.



After experimental validation, the results showed that under the APS process parameters of a current of 617 A, a voltage of 65 V, a H₂ flow of 3 L min⁻¹, and a thickness of 200 μm, the bonding strength increased by more than 19% to 55.5 MPa compared with the original maximum value of 46.6 MPa, indicating that the constructed GMM–SVR model can accurately predict the bonding strength of YSZ coating.

INTRODUCTION

Gas turbines have been key heat-to-power conversion devices in the field of power generation since their inception, and improvement of their thermal efficiency has become an important research direction for scientific researchers.^{1,2} However, the improvement of thermal efficiency of a gas turbine requires a significant increase in the operating temperature of the combustion chamber, which brings new challenges to maximize the operating temperature of the gas turbine components.³ Thermal barrier coatings (TBCs) have been widely used for hot metal parts in advanced gas turbines and diesel engines to improve thermal protection to improve the thermal efficiency and performance.^{4–6} The TBC system refers to a complex coating system composed of a metal substrate, a bonding layer, and a surface ceramic coating.^{7,8} At present, yttria-stabilized zirconia (YSZ) has been one of the most widely used TBC materials.^{9,10} A stable or partially stable structure can be formed at high temperatures with Y₂O₃ added as a stabilizer to ZrO₂ to effectively alleviate the thermal mismatch problem of the ceramic layer and improve the bonding strength between the ceramic coating and the substrate.¹¹ The bonding strength is an important indicator to measure the bonding of the ceramic coating and the substrate.¹² Materials with higher bonding strength tend to have a better ability to withstand temperature changes; the higher the bonding quality, the less likely the cracking. In addition, the bonding strength of YSZ coating and substrate

materials is closely related to the choice and process parameters of the preparation process.¹³ The industrial preparation of TBCs mainly adopts two technologies, plasma spraying and electron beam-physical vapor deposition.¹⁴ Therefore, improving the bonding strength of the ceramic coating and the substrate by optimizing the process parameters of the YSZ coating spraying process is an effective and meaningful work.

Machine learning is the core of artificial intelligence with the ability to reorganize the existing knowledge structure and figure out implicit relationships, and it has been applied in many areas such as medical treatment, finance, materials, and chemistry with significant progress.^{15–19} In materials design, machine learning has occupied an important part in the development and design of alloys, polymers, perovskites, and other materials by virtue of the advantages of obtaining performance and trends from available data without knowing the underlying physical mechanism.^{20–24} Yang et al.²⁵ used machine learning combined with high-throughput screening

Received: March 26, 2022

Accepted: May 30, 2022

Published: June 9, 2022



and pattern recognition back-projection technology to break the upper limit of the hardness of the existing high-entropy alloys and designed the hardness of $\text{Co}_{18}\text{Cr}_7\text{Fe}_{35}\text{Ni}_5\text{V}_{35}$ to be 1148 HV, which is 24.8% higher than the hardness of the alloy with the highest hardness in the original data set. Chen et al.²⁶ used a step-by-step screening method of the packaging algorithm to screen out a subset of features for ridge regression, XGBoost, and support vector regression (SVR) models and integrated the three models to design low-melting-point alloys. Zhang et al.²⁷ proposed an interpretable strategy based on machine learning combined with Shapley additive explanations (SHAP) to accurately predict the formative nature of organic–inorganic hybrid perovskites (HOIPs) and screened out 198 non-toxic HOIP candidate materials with formative probability >0.99. Meanwhile, Lu et al.²⁸ collected an imbalanced formability data set of synthesized HOIPs to explore potential compositions, including 539 positive and 24 negative samples. The imbalanced machine learning was applied in predicting the experimental formability, and the classification model achieved a leaving-one-out cross-validation accuracy of 100.0% and a test accuracy of 96.1%. The important features, namely, A site atomic radius, A site ionic radius, and tolerance factor, were drawn out to reveal their relation to the formability. In the design of machine-learning-aided materials, the descriptors used for modeling are usually composed of structural and compositional information of materials, while the influence of experimental parameters of synthesis or characterization on the properties of materials would be often ignored. For YSZ ceramic coating materials, the bonding strength tends to vary under different spraying process conditions. Construction of a quantitative model to map the relationship between process parameters and bonding strength through machine learning can effectively avoid the workload of modifying materials, improving the bonding strength with the optimized process parameters.

The flowchart of this work is shown in Figure 1. We prepared and characterized the bonding strength of eight YSZ

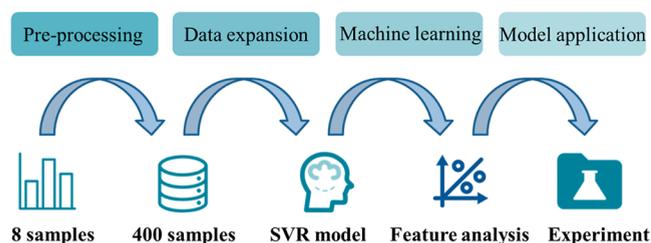


Figure 1. Flowchart of using machine learning to improve the bonding strength of the YSZ coating in this work, including preprocessing with feature selection, data expansion with a GMM, machine learning with algorithms, and model application with experimental validation.

coatings under different spraying process parameters through experiments. With the process parameters as descriptors, machine learning algorithms were used to construct a prediction model of bonding strength. In order to improve the accuracy of the model, a Gaussian mixture model (GMM) was applied to expand the original data set from 8 to 400. After model construction based on the data expansion, SHAP and sensitivity analysis were applied to figure out specific effects of atmospheric plasma spraying (APS) process parameters on the bonding strength of YSZ. The optimized process parameters

that theoretically exceeded the maximum bonding strength of the original data set were obtained after feature analysis, which was validated by experiment with the determined bonding strength in the corresponding process as high as 55.5 MPa.

MATERIALS AND METHODS

Experimental Methodology. The TBC system prepared in the experiments includes a metal substrate, a bonding layer, and a ceramic layer. A nickel-based superalloy was taken as the substrate with the size being a cylinder of Φ 25 mm \times 4 mm. The bonding layer was prepared by spraying NiCrCoAlY powder on the metal substrate with vacuum plasma spraying (VPS, Oerlikon Metco AG, Switzerland). Tables 1 and 2,

Table 1. Chemical Composition of NiCrCoAlY Powder in Bond Coat (wt %)

element	Co	Ni	Cr	Al	Y
content	bal.	29–35	29–35	5–11	0.1–0.8

respectively, list the chemical composition of the bonding layer powder and the VPS spraying parameters. The ceramic layer was prepared by spraying ZrO_2 -4 mol % Y_2O_3 on the bonding layer using APS technology (Oerlikon Metco AG, Switzerland). Table 3 lists the APS spraying parameter settings, including current, voltage, spraying power, the flow rate of Ar and H_2 , Ar/ H_2 , and spraying thickness. The quantitative relationship between process parameters and YSZ bonding strength was explored by using the process parameters of APS as descriptors. Based on the ASTM C733-13 standard, the bonding strength between the YSZ coating and the substrate with a bonding layer (Φ 25 mm \times 4 mm) was characterized by a universal testing machine (Instron-5592), while E7 epoxy resin with a bonding strength greater than 70 MPa was chosen to be the adhesive. In the process of characterization, the change of tensile force is recorded, and the bond strength is calculated according to the formula $P = F/S$, where P is the bonding strength (MPa), F is the maximum tensile force (kN), and S is the coating area ($\times 10^3 \text{ m}^2$).

Algorithm Detail. In this work, there are only eight data available for machine learning, which belong to a very typical small dataset. There are generally two processing methods to deal with the small data set. The first strategy is to choose a machine learning algorithm suitable for small sample modeling, such as support vector machines (SVMs). SVM is an algorithm based on the kernel functions. The existence of kernel functions enables SVM to determine the segmentation hyperplane with less support vectors, which brings the good performance of the constructed model even with small sample data.^{29,30} The general idea of SVM is that it maps the input vector into high dimensional space and finds a most optimal hyperplane as the criterion to classify the samples. In classification, SVM is named the support vector classifier (SVC) as well. The target of SVC is to get the classification line with the maximal margin hyperplane to make samples of different types furthest from each other. In regression, SVM is also called SVR. In SVR, the insensitive channels ε is used to handle the problem of weighing empirical and structural risks. Specifically, the error is ignored when the predicted value \hat{y}_i meets $|y_i - \hat{y}_i| \leq \varepsilon$, otherwise, the error is $|y_i - \hat{y}_i| - \varepsilon$. The deviation is concerned only when it is greater than ε in the empirical risk calculation. Similar to the constraint conditions

Table 2. VPS Parameters

	current/A	voltage/V	thickness/ μm	Ar/L min^{-1}	H ₂ /L min^{-1}	spay distance/mm
VPS-BC	700	65.5	100	50	9	120

Table 3. APS Parameters

sample	current/A	voltage/V	power/KW	Ar/L min^{-1}	H ₂ /L min^{-1}	Ar/H ₂	thickness/ μm
1	503	69.3	34.71	30.00	5.00	6	200
2	503	69.3	34.71	30.00	5.00	6	400
3	641	69.3	44.42	30.00	5.00	6	200
4	641	69.3	44.42	30.00	5.00	6	400
5	520	68	35.36	30.00	5.00	6	300
6	665	68	45.22	30.00	6.00	5	300
7	617	65	40.11	30.00	3.00	10	300
8	617	65	40.11	30.00	3.00	10	400

of SVC, SVR takes the value of margin as the standard to improve the prediction accuracy of the model.

The second strategy is virtual sample generation. From the perspective of the number of samples, the prediction accuracy of the model could be improved by increasing the number of samples. The GMM is a probabilistic model assuming that all data points are generated from a limited number of Gaussian mixtures.^{31,32} If n observations $X = \{X_1, \dots, X_n\}$ are generated by the mixed distribution P , each vector X_i is p -dimensional, and the distribution P is composed of G components, then the maximum mixed likelihood function of the distribution could be obtained using eq 1

$$L_M(\theta_1, \dots, \theta_G; \gamma_1, \dots, \gamma_n | x) = \prod_{i=1}^n \sum_{k=1}^G \pi_k f_k(x_i | \theta_k) \quad (1)$$

$$\left(\pi_k \geq 0; \sum_{k=1}^G \pi_k = 1 \right)$$

where $f_k(x_i | \theta_k)$ represents the density function of k -th category; θ_k is the corresponding parameter; and π_k is the weight parameter representing the probability that an observation belongs to the k -th category. If $f_k(x_i | \theta_k)$ is a multivariate normal distribution, then P is a Gaussian mixture distribution in which θ_k is composed of the mean value μ_k and the covariance matrix Σ_k . The density function $f_k(x_i | \theta_k)$ is shown in eq 2

$$f_k \left(x_i \left| \mu_k, \Sigma_k \right. \right) = \frac{\exp \left\{ -\frac{1}{2} (x_i - \mu_k)^T \Sigma_k^{-1} (x_i - \mu_k) \right\}}{2\pi^{\mu/2} |\Sigma_k|^{1/2}} \quad (2)$$

The Gaussian mixture distribution can be described by the probability density function represented by the weighted average of the Gaussian density functions, and the specific description is shown in the following eq 3

$$P(x | \theta) = \sum_{k=1}^G \pi_k f_k \left(x_i \left| \mu_k, \Sigma_k \right. \right) \quad (3)$$

GMM is essentially a density estimation algorithm. It can be seen from equation that by adjusting the weight π_k , the probability density function curve of the mixed model would be greatly affected to fit the non-linear function of any shape. The generation probability model describing the small sample data of YSZ bonding strength can be constructed through GMM. With the parameters solved by the expectation

maximization algorithm, the virtual samples meeting the expectation could be generated through the obtained generation model.

Computational Platform. The process of the machine learning model construction was conducted on the machine learning software package called ExpMiner and the online platform called OCPMDM, both of which were developed in our laboratory. The software of ExpMiner can be downloaded from the website of the Laboratory of Materials Data Mining in Shanghai University (<http://materials-data-mining.com/home#>). OCPMDM can be accessed at <http://materials-data-mining.com/ocpmdm/>.

RESULTS AND DISCUSSION

Data Generation. All data in this work were derived from experiments. The experimental bonding strength of YSZ coatings under different APS parameters is shown in Figure 2. Combining the figure with Table 3, it can be found that when other process parameters remain unchanged, the bonding strength is positively correlated with power and negatively correlated with thickness. However, the detailed influence of the process parameters on the bonding strength could not be observed only from Figure 2 and Table 3 nor the

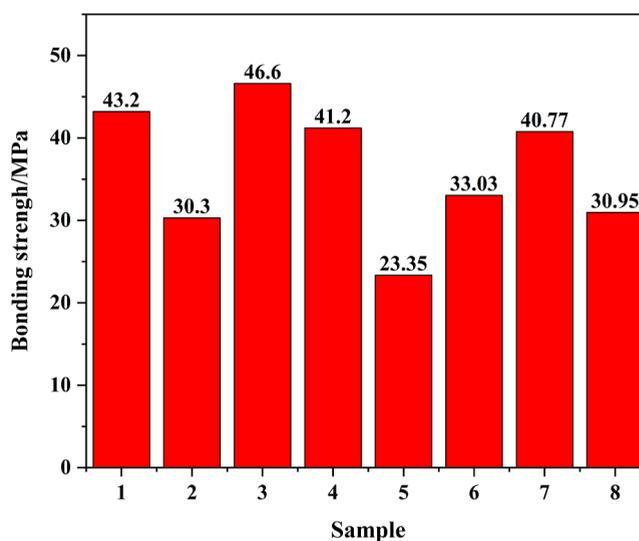


Figure 2. Experimental bonding strength of coating of eight samples with different APS parameters. Direct modeling for bonding strength prediction.

optimal parameters can be obtained for the improvement of bonding strength. Therefore, we have considered the bonding strength as the target variable and the APS process parameters as descriptors to construct a machine learning model to further explore the specific relationship between the process parameters and the bonding strength.

Using machine learning algorithms to construct a model is divided into data collection, feature selection, model selection, parameter optimization, and model evaluation. In data collection, the bonding strength of the samples in Figure 2 is set as the target variable, while the corresponding APS parameters in Table 3 are set as the descriptors. Feature selection aims to remove redundant variables and screen out the descriptors strongly related to the target variable to reduce the training time and improve prediction accuracy. Machine learning algorithms for feature selection such as maximum correlation and minimum redundancy, genetic algorithms, and recursive elimination methods are generally used to select the optimal descriptor subset for modeling. However, considering the experimental feasibility because the descriptors in this work are process parameters, it is more appropriate to use domain knowledge to select descriptors for modeling. The column Ar in Table 3 should be removed for the values are constant throughout the column. Accordingly, the column Ar/H₂ should also be removed for the values are completely linearly related to the values in the column H₂. The column power is the product of current and voltage, and it could be removed because the power could be controlled by adjusting the current and voltage. After removing redundant descriptors, the following descriptors are available for modeling: current, voltage, H₂, and thickness. Model selection refers to selecting the algorithm with the highest prediction accuracy from many modeling algorithms according to the evaluation functions. In this part, algorithms including ordinary least square (OLS) linear regression, random forest regression (RFR), decision trees regression (DTR), partial least squares (PLS), multiple linear regression (MLR), artificial neural network (ANN), and SVR are carried out for comparison. Because the SVR is a kernel-based algorithm, the choice of the kernel function will greatly affect the model prediction accuracy. Hence, the influence of different kernel functions on the SVR is also considered. The correlation coefficient (*R*) and root mean square error (RMSE) of leave-one-out cross-validation (LOOCV) are adopted as evaluation functions to evaluate the performance of the constructed model. The results are shown in Table 4. It can be found that compared with other algorithms, ANN performs the best with the highest *R* and lowest RMSE to be the optimal algorithm for modeling. In the

Table 4. *R* and RMSE of the Bonding Strength in LOOCV of Different Algorithms Based on Original Eight Samples

algorithms	<i>R</i>	RMSE
OLS	0.580	7.183
RFR	−0.005	8.362
DTR	−0.116	10.705
PLS	0.483	12.045
MLR	0.544	10.854
ANN	0.690	6.279
SVR-linear kernel	0.475	8.738
SVR-Gaussian kernel	−0.026	8.948
SVR-polynomial kernel	0.302	12.967

algorithm of ANN, the parameters of the number of input layer nodes (N_{input}), the number of hidden layer nodes (N_{hidden}), the number of output layer nodes (N_{output}), the learning rate from the input layer to hidden layer (rate 1), the learning rate from the hidden layer to output layer (rate 2), and the momentum term have a significant impact on the prediction accuracy of the model. To further improve the performance of the ANN model, grid search is used to optimize the ANN parameters with the RMSE of LOOCV as the evaluation index. Grid search refers to looping through all the candidate parameters, trying every possibility to get the best performing parameters. The starting value, ending value, step size of N_{hidden} , rate 1, rate 2, and momentum term in the grid search optimization of ANN are shown in Table S1 of the Supporting Information. The optimized parameters are shown in Table 5. After

Table 5. Optimized ANN Parameters

N_{input}	N_{hidden}	N_{output}	rate 1	rate 2	momentum term
4	4	1	0.32	0.12	0.69

parameter optimization, *R* of LOOCV has increased from 0.690 to 0.758, and the corresponding RMSE has also reduced from 6.279 to 6.009. In model evaluation, the resubstitution test and LOOCV are employed to further evaluate predictive ability of the model. The resubstitution test aims to test the self-consistency of the prediction method by predicting the modeling data. LOOCV is to take a data set containing *k* samples, of which *k*-1 is used as the training set, and the remaining one is used as the test set. Then, select the next one as the test set, and the remaining *k*-1 as the training set. The results are obtained until all samples are predicted as the test set. The evaluation result of LOOCV can be used to determine whether the model has the situation of overfitting. In addition to the resubstitution test and LOOCV, there should be data exclusive to the modeling data as an independent test set to test the predictive ability for external data. However, considering that there are only eight data for modeling in this work, if part of samples are taken as the test set, the data would have a greater negative impact on the prediction accuracy of the model. Moreover, LOOCV essentially performed eight independent tests with the test sample size of 1, which could also make up for the lack of independent test evaluation. The results of the resubstitution test and LOOCV are shown in Figure 3. Although *R* of the resubstitution test can reach 1.000, *R* of LOOCV being 0.758 still indicates that the ANN model constructed with eight samples is far from satisfactory.

To improve the prediction accuracy of the model, GMM was used to generate virtual samples. The Gaussian mixture distribution that best fits the original eight samples was calculated by GMM, while the corresponding descriptor values of the virtual samples were obtained by sampling in the obtained distribution, and the target variable of the virtual samples was obtained by the nearest-neighbor regression algorithm without weight. 50 virtual samples were generated for each sample of the original data by GMM. After virtual sample generation, a total of 400 samples were collected with the data size increased by 50 times compared with the original data set of only eight samples. The 400 virtual samples could be available in the file named “dataset.txt” of the Supporting Information to be directly used by ExpMiner, which were randomly divided into a training set of 320 samples and a test

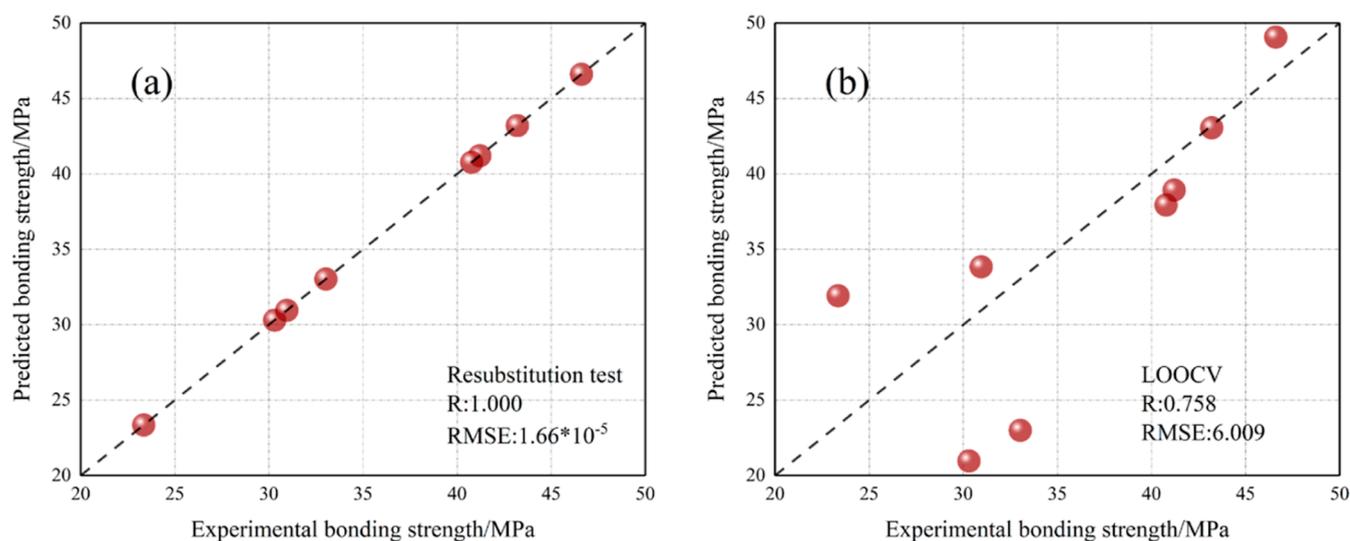


Figure 3. Experimental bonding strength vs predicted bonding strength with corresponding R and RMSE based on (a) resubstitution test and (b) LOOCV. GMM-based modeling for bonding strength prediction.

set of 80 samples according to a ratio of 4:1. Because the data set has changed, the model selection step should be repeated to ensure that the most suitable algorithm could be found for modeling. The R and RMSE values of different methods of LOOCV and 10-fold cross validation (10-fold CV) are shown in Table 6, from which it can be concluded that after data

Table 6. R and RMSE of the Bonding Strength in LOOCV and 10-Fold Cross Validation of Different Algorithms Based on 400 Samples

algorithms	R_{LOOCV}	$RMSE_{LOOCV}$	$R_{10\text{-fold CV}}$	$RMSE_{10\text{-fold CV}}$
OLS	0.868	4.409	0.869	4.397
RFR	0.983	1.619	0.982	1.687
DTR	0.975	1.977	0.971	2.137
PLS	0.868	4.391	0.869	4.376
MLR	0.868	4.391	0.868	4.386
ANN	0.988	1.391	0.989	1.336
SVR-linear kernel	0.868	4.525	0.868	4.521
SVR-Gaussian kernel	0.989	1.312	0.989	1.324
SVR-polynomial kernel	0.989	1.305	0.989	1.306

expansion, the SVR with the polynomial kernel function is the optimal algorithm with the highest R and lowest RMSE of both LOOCV and 10-fold CV. In the SVR algorithm with the polynomial kernel, the insensitive loss function ϵ and the capacity parameter C have a significant impact on the prediction accuracy of the model. After parameter optimization by grid search method with the RMSE of LOOCV as the evaluation functions, the optimal parameters were ϵ of 0.02 and C of 15. The starting value, ending value, and step size of ϵ and C in the grid search optimization of SVR are shown in Table S2. The trend of the RMSE of LOOCV with ϵ and C is shown in Figure S1. Under the optimal parameters, the results of the resubstitution test, LOOCV, 10-fold CV, and independent test are shown in Figure 4. After data expansion by GMM, the prediction accuracy has been greatly improved with R of LOOCV increasing from 0.690 to 0.990. In addition, the R value of the independent test set is as high as 0.986, also demonstrating the good generalization ability of the constructed model. The result of the independent test also made

up for the limitation of the lack of the independent test because the original data set of eight samples was too limited to be divided for the extra independent test set. The constructed SVR model with the optimal parameters named "SVR model.mod" file is available at <https://github.com/luktian/models>, which could be directly imported into the software of ExpMiner and used for the bonding strength prediction. In addition, the y -scrambling of the repeatability measure was adopted to further verify the stability of the model, avoiding random fluctuations caused by dataset division. The dataset of 400 samples was randomly divided 30 times into the training set and the test set after algorithm selection and parameter optimization with the R and RMSE of LOOCV, 10-fold CV, and independent test set as evaluation function to validate the stability of the model. R , RMSE with the corresponding average and standard deviation values (σ) of LOOCV, 10-fold CV, and independent test are shown in Table S3. It could be found that the constructed model has shown good predictability and stability according to the average R and RMSE of the independent test being higher than 0.988 and lower than 1.310 as well as the small σ being lower than 0.114.

Feature Analysis. Feature analysis refers to the statistical and physical analysis of the modeling descriptors to further explore the relationship between important descriptors and the target variable. It should be noted that all the feature analysis is only for the training set of the optimal model. For the constructed SVR model, SHAP and sensitivity analysis are used to explore the selected descriptors. SHAP belongs to a feature analysis method based on game theory to analyze the contribution of each feature to the predicted value of the model, which would assign a value to each feature of every sample to indicate the contribution of the feature to model predictions.^{33,34} The assigned value is also called the SHAP value of the feature, which is the weighted average of all possible differences. All features could be ranked according to the SHAP value to represent the quantitative contribution to the target variable. The main purpose of sensitivity analysis is to evaluate whether the results obtained under given conditions are sufficiently reliable when other conditions are not fully satisfied, which could be used to investigate the

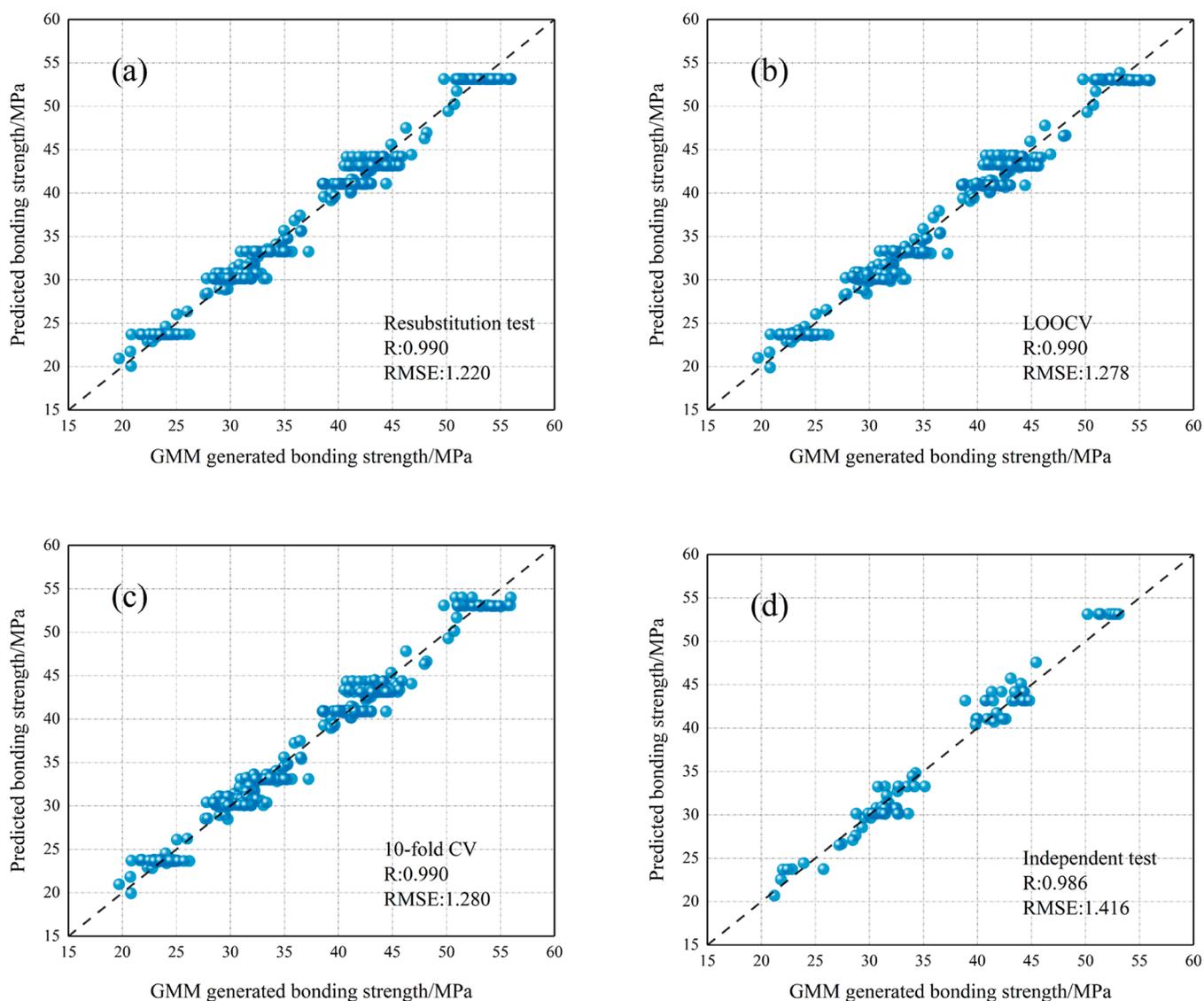


Figure 4. GMM-generated bonding strength vs predicted bonding strength with corresponding R and RMSE based on (a) resubstitution test, (b) LOOCV, (c) 10-fold cross validation, and (d) independent test.

change of the target variable with a certain descriptor under the condition of fixing other descriptors.^{35,36}

The SHAP analysis and sensitivity analysis of the descriptor are shown in Figure 5. In Figure 5a, the ranking of the descriptor contribution to the predicted value of the model could be obtained according to the SHAP values. It can be seen from Figure 5a that the descriptor contributing the most to the SVR model of bonding strength prediction is the thickness, followed by the flow rate of H_2 , current, and voltage. SHAP analysis can rank descriptors according to their contribution to the model, while the exploration of the sensitivity of bonding strength to changes in specific descriptors requires sensitivity analysis. Figure 5b–e illustrates the sensitivity analysis of the modeling descriptors. It can be seen that the bonding strength has a negative correlation with the flow rate of H_2 and a strong positive correlation with the voltage. However, for thickness and current, there is a negative correlation at first and a positive correlation after reaching the lowest point. However, in general, the bonding strength is negatively correlated with the thickness and the flow rate of H_2 as well as positively correlated with current and voltage. In the

process of coating deposition, the residual stresses of quenching stress and cooling stress would occur. The quenching stress is derived from the rapid formation of the layered structure during the deposition process, while the cooling stress comes from the mismatch of the thermal expansion coefficient of the coating and the substrate. The quenching stress and cooling stress increase with the coating thickness, which promotes the initiation and propagation of microcracks to lead to the decrease of the bonding strength after the tensile test.^{37–39} From the importance ranking of the descriptors by SHAP, it can be concluded that the quenching stress and cooling stress that increase with the coating thickness are the most important factors leading to the decrease of the bonding strength compared to the impact of other factors.

APS Parameters Optimization. The purpose of the machine learning model construction for bonding strength prediction is to achieve a breakthrough in the bonding strength by optimization of the APS process parameters. After obtaining the quantitative trend of the APS process parameters on the bonding strength through feature analysis, the process

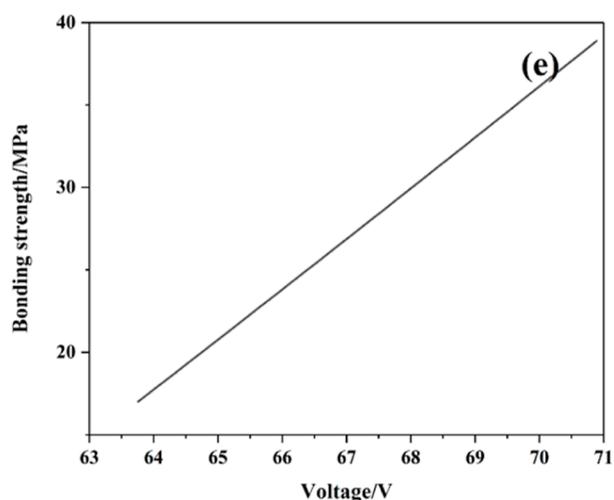
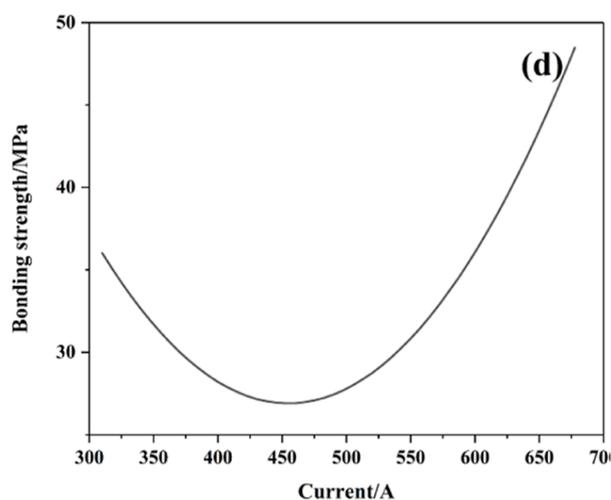
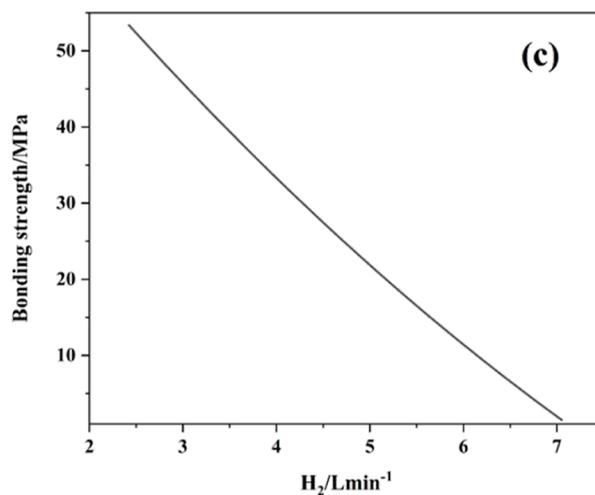
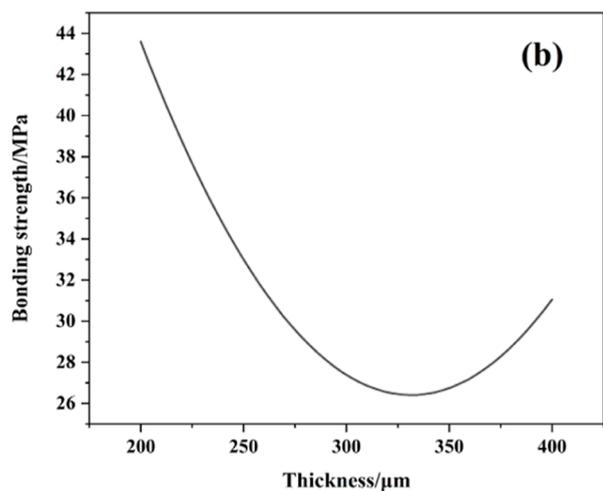
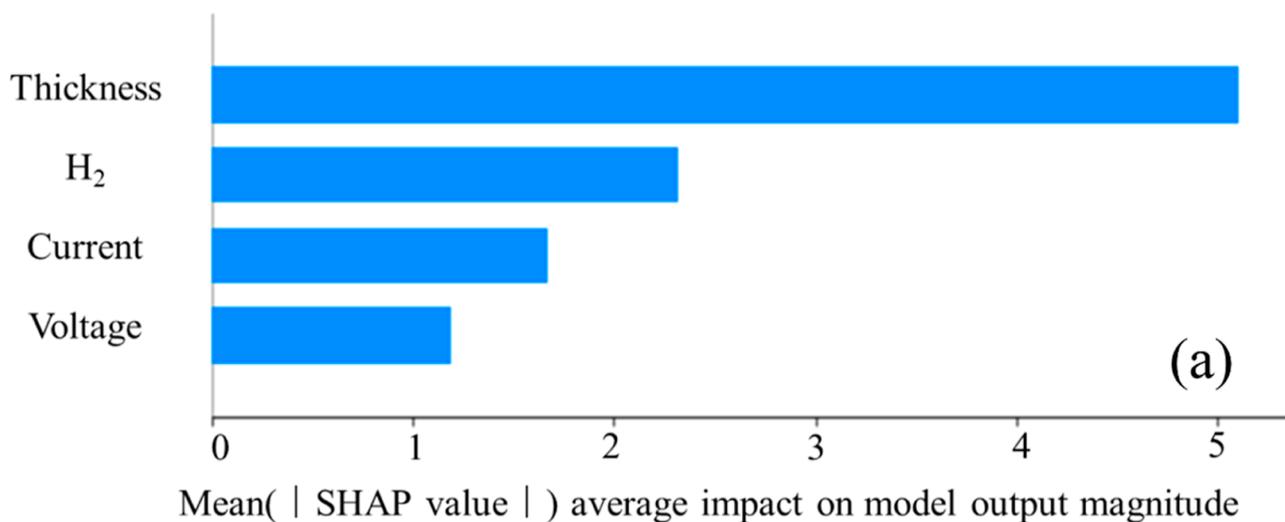


Figure 5. Feature analysis of (a) SHAP and sensitivity analysis of (b) thickness, (c) the flow rate of H₂, (d) current, and (e) voltage.

parameters could be optimized to improve the predicted value of the bonding strength. The optimized process parameters and the corresponding predicted bonding strength are shown in Table 7. Under this process, the maximum predicted

bonding strength by the model could reach as high as 68.771 MPa, while the highest value of the bonding strength in the original 8 data sets is 46.6 MPa. After APS parameter optimization by the machine learning model, the bonding

Table 7. Optimized APS Parameters and the Corresponding Predicted Bonding Strength

predicted bonding strength/MPa	current/A	voltage/V	H ₂ /L min ⁻¹	thickness/μm
67.577	700	71	2.5	200
53.137	617	65	3	200

strength is increased by 47.58%. However, in the origin data in Table 3, the distribution range of the current value is 503–665 A, indicating that the current value is difficult to reach 700 in the experiment. Fortunately, among the 400 virtual samples generated by GMM, there still exist some samples with predicted bonding strength higher than 46.6 MPa, among which the highest predicted value could reach up to 55.95 MPa. The specific process parameters are a current of 617 A, a voltage of 65 V, H₂ of 3 L min⁻¹, and a thickness of 200 μm. Besides, it can be observed that most of the process parameter values corresponding to the virtual samples with a bonding strength greater than 46.6 MPa are floating under the optimal process parameters, which can be regarded as an error in the experiment. The range of the optimized process parameter values does not deviate from the eight samples of the source data, indicating that the process parameters are equipped with the experimental feasibility.

Model Explanation. In this work, the results predicted using the SVR model can be explained by using material pattern recognition. Pattern recognition methods include statistical pattern recognition and syntactic pattern recognition. In this work, statistical pattern recognition is used to extend the descriptors to the multidimensional space of the sample projection. By applying appropriate computer pattern recognition technology to identify the distribution area of samples of various shapes, a mathematical model describing the distribution range of various samples in a multidimensional space can be obtained.⁴⁰ The pattern recognition method used in this work is principal component analysis (PCA), which can calculate two principal components PCA (1) and PCA (2) by a linear combination of descriptors to form an optimal discriminant plane.⁴¹

Taking the APS process parameters as the feature set; the bonding strength as the target variable; the training set as the data set; samples with the bonding strength greater than 46.6 MPa as positive samples; the rest as negative samples, the PCA projection diagram is shown in Figure 6. The rectangular area in the figure refers to the optimized area. In Figure 6, there are 46 samples in the optimized area, of which there are 43 positive samples and 3 negative samples. Positive samples can account for 92.73%, which is much higher than 57.54% in the training set. As long as the calculated PCA (1) and PCA (2) satisfy the boundary conditions of the optimized region, the probability of obtaining a positive sample can be improved. The boundary conditions of the optimized area are shown as eqs 4–7

$$-0.7651 \leq \text{PCA}(1) \leq 1.592 \quad (4)$$

$$-2.204 \leq \text{PCA}(2) \leq -1.004 \quad (5)$$

$$22.113 \leq 0.004316[\text{Current}] + 0.2877[\text{Voltage}] + 0.5133[\text{H}_2] - 0.003342[\text{Thickness}] \leq 24.470 \quad (6)$$

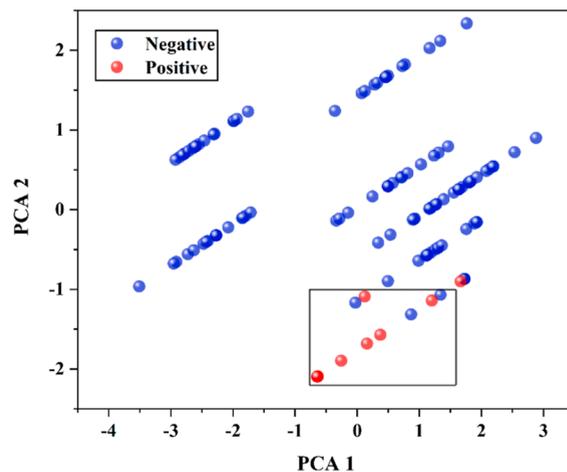


Figure 6. Pattern recognition of different samples by using the PCA method. The positive samples in the optimized area account for 92.73%.

$$12.482 \leq -0.002189[\text{Current}] + 0.1673[\text{Voltage}] + 0.2793[\text{H}_2] + 0.01116[\text{Thickness}] \leq 13.682 \quad (7)$$

Experimental Validation. To verify the model prediction of the optimized parameters, the bonding strength of YSZ coating was determined by experiments. The stress–extension curve and the corresponding APS parameters as well as the image of coating fracture are shown in Figure 7. Obviously, the

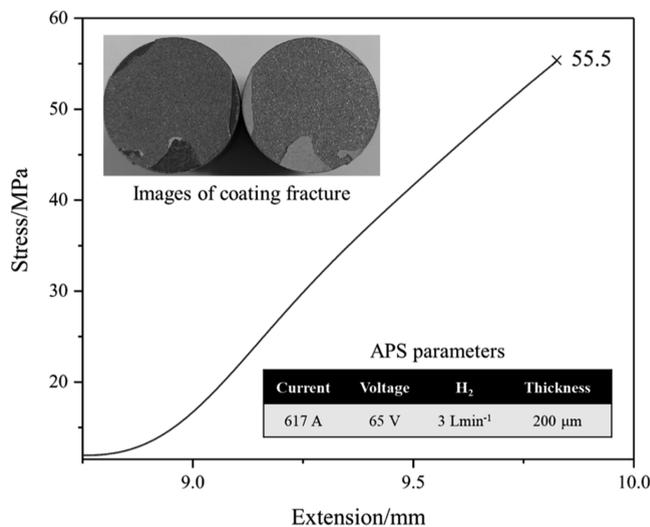


Figure 7. The stress–extension curve of the YSZ coating with optimized APS parameters and the corresponding images of coating fracture.

bonding strength with optimized APS parameters is better, which is 19.10% higher than the best value in the original data. In addition, the predicted bonding strength under the APS parameter is 53.137 MPa with the absolute error of 2.363 MPa, indicating the ideal prediction accuracy of the constructed GMM–SVR model. This prediction error can be reduced by constructing models by collecting more experimental data in future studies. The result of experimental verification demonstrates that the constructed GMM–SVR model could

assist the optimization of the APS process to promote the bond strength of YSZ coatings.

CONCLUSIONS

In this work, APS parameters were taken as descriptors to construct an SVR model for predicting the bonding strength of YSZ thermal barrier materials. GMM was used to expand the data set from the original 8 data points to 400 with the increase of R of LOOCV from 0.690 to 0.990, which has proved that GMM could solve the problem of low prediction accuracy and the lack of the independent testing due to the limited data. After model construction, SHAP and sensitivity analysis were adopted to analyze the relationship between descriptors and bonding strength. The results show that the thickness is a major factor in the bonding strength. The bonding strength is negatively correlated with the thickness and H_2 , but positively correlated with current and voltage. The parameters of APS were optimized through feature analysis and PCA. After optimization and experimental validation, the determined bonding strength with optimized APS parameters could reach 55.5 MPa, which is 19.109% higher than the maximum value of 46.6 MPa in the original eight data sets.

Although we have applied machine learning to achieve breakthrough of bonding strength by optimizing APS parameters, there still exist more improvements of this work needing to be realized in the future work. First, the flow rate of Ar in the APS process parameters is a constant column, which is unavailable to the model construction to explore the influence of the flow rate of Ar on the bonding strength. In subsequent work, the flow rate of Ar can be changed to further explore the influence of this parameter on the bonding strength. Second, although the method of GMM can improve the model accuracy of small data set modeling, the original eight sets of data are still too limited to better understand universality of the patterns found by feature analysis. This limitation can be solved through active learning. After the experimental validation with optimized APS parameters, the samples can be put back into the data set to reform a bigger data set for machine learning. Followed by active learning, the data size could be purposefully increased through iterative loops to achieve the two-way optimization of the model and APS process parameters simultaneously.

The algorithms of SVM and GMM were used to deal with machine learning of the small size of the dataset in this work, achieving the satisfactory results for experimental validation. SVM is an applicable machine learning method with a solid theoretical foundation, which has a wide range of applications in modeling of small size of datasets in materials science. The concept of "margin" in SVM could be used to obtain a structured description of data distribution, thereby reducing the requirements for data size and data distribution. Besides, the constructed model by SVM tends to have excellent generalization ability. However, SVM is sensitive to the selection of kernel function and its parameters. For different data, how to choose the optimal kernel function and parameters is still a challenge. GMM adopts the idea of a mixture model to find the distribution of data and obtains virtual samples by sampling based on the mixture Gaussian distribution to improve the performance of the model by expanding the data size. The results have shown that the GMM algorithm could improve the accuracy of small-data machine learning models with good applicability to be widely applied in other material fields. Li et al.³² have applied the GMM into the

Tennessee Eastman process and an industrial hydrocracking process to improve the performance of a machine learning model with a small size of dataset. Therefore, the combination of SVM and GMM to process small-data machine learning modeling can not only be applied to optimize the bonding strength of YSZ coatings but also can be used for small-data modeling in other material fields. However, there still exists room for the improvement of GMM such as anomaly point analysis. There may be anomaly points in the virtual samples generated by GMM. For the materials data, the characteristic values of materials tend to have a certain range, and the generated data are difficult to ensure if it conforms to the characteristics of actual materials. The analysis of anomaly points in GMM is still a direction for further research. Although the materials synthesis and characterization technology become more and more mature, most of the material data still belong to the small data due to the high cost of experiments or calculations. For small-data machine learning tasks, in addition to algorithm-based processing methods such as SVM and GMM, the amount of data can be expanded through high-throughput experiments and calculations, active learning, transfer learning, and the establishment of a complete material database and data processing platform.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acsomega.2c01839>.

Starting value, ending value, and step size information for ANN; SVR parameter optimization; and trend of the RMSE of LOOCV with ϵ and C (PDF)
Dataset of 400 virtual samples (TXT)

AUTHOR INFORMATION

Corresponding Authors

Wencong Lu – Materials Genome Institute, Shanghai University, Shanghai 200444, China; Department of Chemistry, College of Sciences, Shanghai University, Shanghai 200444, China; orcid.org/0000-0001-5361-6122; Email: wclu@shu.edu.cn

Quan Qian – School of Computer Engineering and Science, Shanghai University, Shanghai 200444, China; Email: qqian@shu.edu.cn

Yi Zeng – The State Key Lab of High Performance Ceramics and Superfine Micro-structure, Shanghai Institute of Ceramics, Chinese Academy of Sciences, Shanghai 200050, China; Email: zengyi@mail.sic.ac.cn

Authors

Pengcheng Xu – Materials Genome Institute, Shanghai University, Shanghai 200444, China

Can Chen – The State Key Lab of High Performance Ceramics and Superfine Micro-structure, Shanghai Institute of Ceramics, Chinese Academy of Sciences, Shanghai 200050, China

Shuizhou Chen – School of Computer Engineering and Science, Shanghai University, Shanghai 200444, China

Complete contact information is available at:

<https://pubs.acs.org/doi/10.1021/acsomega.2c01839>

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the National Key Research and Development Program of China (no.2018YFB0704400), the Key Program of Science and Technology of Yunnan Province (grant no. 202002AB080001/2), and the Key Research Project of Zhejiang Laboratory (no. 2021PE0AC02).

REFERENCES

- (1) Mohamed, O.; Khalil, A. Progress in Modeling and Control of Gas Turbine Power Generation Systems: A Survey. *Energies* **2020**, *13*, 2358.
- (2) Singh, K. Advanced Materials for Land Based Gas Turbines. *Trans. Indian Inst. Met.* **2014**, *67*, 601–615.
- (3) Krishna Anand, V. G.; Parammasivam, K. M. Thermal barrier coated surface modifications for gas turbine film cooling: a review. *J. Therm. Anal. Calorim.* **2020**, *146*, 545–580.
- (4) Thakare, J. G.; Pandey, C.; Mahapatra, M. M.; Mulik, R. S. Thermal Barrier Coatings—A State of the Art Review. *Met. Mater. Int.* **2020**, *27*, 1947–1968.
- (5) Yang, X.; Zhang, J.; Lu, Z.; Park, H.-Y.; Jung, Y.-G.; Park, H.; Koo, D. D.; Sinatra, R.; Zhang, J. Removal and repair techniques for thermal barrier coatings: a review. *Trans. IMF* **2020**, *98*, 121–128.
- (6) Mauer, G.; Vaben, R. Technology vision. *Surf. Eng.* **2013**, *27*, 477–479.
- (7) Dhonne, S.; Mahalle, A. M. Thermal barrier coating materials for SI engine. *J. Mater. Res. Technol.* **2019**, *8*, 1532–1537.
- (8) Liu, Q.; Huang, S.; He, A. Composite ceramics thermal barrier coatings of yttria stabilized zirconia for aero-engines. *J. Mater. Sci. Technol.* **2019**, *35*, 2814–2823.
- (9) Mauer, G.; Jarligo, M. O.; Mack, D. E.; Vaßen, R. Plasma-Sprayed Thermal Barrier Coatings: New Materials, Processing Issues, and Solutions. *J. Therm. Spray Technol.* **2013**, *22*, 646–658.
- (10) An, K. J. Assessment of the Thermal Conductivity of Yttria-Stabilized Zirconia Coating. *Mater. Trans.* **2014**, *55*, 188–193.
- (11) Song, M.; Guo, H.; Abbas, M.; Gong, S. Thermal deformation of Y2O3 partially stabilized ZrO2 coatings by digital image correlation method. *Surf. Coat. Tech.* **2013**, *216*, 1–7.
- (12) Ang, A. S. M.; Berndt, C. C. Investigating the anisotropic mechanical properties of plasma sprayed yttria-stabilised zirconia coatings. *Surf. Coat. Tech.* **2014**, *259*, 551–559.
- (13) Ramachandran, C. S.; Balasubramanian, V.; Ananthapadmanabhan, P. V. Multiobjective Optimization of Atmospheric Plasma Spray Process Parameters to Deposit Yttria-Stabilized Zirconia Coatings Using Response Surface Methodology. *J. Therm. Spray Technol.* **2010**, *20*, 590–607.
- (14) Mauer, G.; Vaßen, R. Coatings with Columnar Microstructures for Thermal Barrier Applications. *Adv. Eng. Mater.* **2019**, *22*, 1900988.
- (15) Arun Bhavsar, K.; Abugabah, A.; Singla, J.; Ali AlZubi, A.; Kashif Bashir, A.; Nikita. A Comprehensive Review on Medical Diagnosis Using Machine Learning. *Comput. Mater. Contin.* **2021**, *67*, 1997–2014.
- (16) Alanazi, H. O.; Abdullah, A. H.; Qureshi, K. N. A Critical Review for Developing Accurate and Dynamic Predictive Models Using Machine Learning Methods in Medicine and Health Care. *J. Med. Syst.* **2017**, *41*, 69.
- (17) Cai, J.; Chu, X.; Xu, K.; Li, H.; Wei, J. Machine learning-driven new material discovery. *Nanoscale Adv.* **2020**, *2*, 3115–3130.
- (18) Wei, J.; Chu, X.; Sun, X. Y.; Xu, K.; Deng, H. X.; Chen, J.; Wei, Z.; Lei, M. Machine learning in materials science. *InfoMat* **2019**, *1*, 338–358.
- (19) Haghghatlari, M.; Hachmann, J. Advances of machine learning in molecular modeling and simulation. *Curr. Opin. Chem. Eng.* **2019**, *23*, 51–57.
- (20) Qiu, Z.; Sugio, K.; Sasaki, G. Classification of Microstructures of AlSi Casting Alloy in Different Cooling Rates with Machine Learning Technique. *Mater. Trans.* **2021**, *62*, 719–725.
- (21) Wu, S.; Kondo, Y.; Kakimoto, M.-a.; Yang, B.; Yamada, H.; Kuwajima, I.; Lambard, G.; Hongo, K.; Xu, Y.; Shiomi, J.; Schick, C.; Morikawa, J.; Yoshida, R. Machine-learning-assisted discovery of polymers with high thermal conductivity using a molecular design algorithm. *npj Comput. Mater.* **2019**, *5*, 1–11.
- (22) Xu, P.; Chang, D.; Lu, T.; Li, L.; Li, M.; Lu, W. Search for ABO₃ Type Ferroelectric Perovskites with Targeted Multi-Properties by Machine Learning Strategies. *J. Chem. Inf. Model.* **2021**, DOI: 10.1021/acs.jcim.1c00566.
- (23) Tao, Q.; Xu, P.; Li, M.; Lu, W. Machine learning for perovskite materials design and discovery. *npj Comput. Mater.* **2021**, *7*, 1–18.
- (24) Li, L.; Tao, Q.; Xu, P.; Yang, X.; Lu, W.; Li, M. Studies on the regularity of perovskite formation via machine learning. *Comput. Mater. Sci.* **2021**, *199*, 110712.
- (25) Yang, C.; Ren, C.; Jia, Y.; Wang, G.; Li, M.; Lu, W. A machine learning-based alloy design system to facilitate the rational design of high entropy alloys with enhanced hardness. *Acta Mater.* **2022**, *222*, 117431.
- (26) Chen, H.; Shang, Z.; Lu, W.; Li, M.; Tan, F. A Property-Driven Stepwise Design Strategy for Multiple Low-Melting Alloys via Machine Learning. *Adv. Eng. Mater.* **2021**, *23*, 2100612.
- (27) Zhang, S.; Lu, T.; Xu, P.; Tao, Q.; Li, M.; Lu, W. Predicting the Formability of Hybrid Organic-Inorganic Perovskites via an Interpretable Machine Learning Strategy. *J. Phys. Chem. Lett.* **2021**, *12*, 7423–7430.
- (28) Lu, T.; Li, H.; Li, M.; Wang, S.; Lu, W. Predicting Experimental Formability of Hybrid Organic-Inorganic Perovskites via Imbalanced Learning. *J. Phys. Chem. Lett.* **2022**, *13*, 3032–3038.
- (29) Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297.
- (30) Shawe-Taylor, J.; Sun, S. A review of optimization methodologies in support vector machines. *Neurocomputing* **2011**, *74*, 3609–3618.
- (31) Reynolds, D. A.; Quatieri, T. F.; Dunn, R. B. Speaker Verification Using Adapted Gaussian Mixture Models. *Digit. Signal Process.* **2000**, *10*, 19–41.
- (32) Li, L.; Kumar Damarla, S.; Wang, Y.; Huang, B. A Gaussian mixture model based virtual sample generation approach for small datasets in industrial processes. *Inf. Sci.* **2021**, *581*, 262–277.
- (33) Lundberg, S. M.; Erion, G. G.; Lee, S.-I. Consistent Individualized Feature Attribution for Tree Ensembles. **2019**, arXiv:1802.03888. arXiv preprint.
- (34) Lundberg, S. M. L.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions. **2017**, arXiv:1705.07874. arXiv preprint.
- (35) Xu, P.; Lu, T.; Ju, L.; Tian, L.; Li, M.; Lu, W. Machine Learning Aided Design of Polymer with Targeted Band Gap Based on DFT Computation. *J. Phys. Chem. B* **2021**, *125*, 601–611.
- (36) Shi, L.; Chang, D.; Ji, X.; Lu, W. Using Data Mining To Search for Perovskite Materials with Higher Specific Surface Area. *J. Chem. Inf. Model.* **2018**, *58*, 2420–2427.
- (37) Chen, C.; Song, X.; Li, W.; Zheng, W.; Ji, H.; Zeng, Y.; Shi, Y. Relationship between microstructure and bonding strength of yttria-stabilized zirconia thermal barrier coatings. *Ceram. Int.* **2022**, *48*, 5626–5635.
- (38) Wang, J.; Sun, J.; Zhang, H.; Dong, S.; Jiang, J.; Deng, L.; Zhou, X.; Cao, X. Effect of spraying power on microstructure and property of nanostructured YSZ thermal barrier coatings. *J. Alloys Compd.* **2018**, *730*, 471–482.
- (39) Wang, Y.; Liu, Q.; Zheng, Q.; Li, T.; Chong, N.; Bai, Y. Bonding and Thermal-Mechanical Property of Gradient NiCoCrAlY/YSZ Thermal Barrier Coatings with Millimeter Level Thickness. *Coatings* **2021**, *11*, 600.
- (40) Kubat, M. Designing neural network architectures for pattern recognition. *Knowl. Eng. Rev.* **2000**, *15*, 151–170.
- (41) Xie, Y.; Sun, P. Terahertz data combined with principal component analysis applied for visual classification of materials. *Opt. Quantum Electron.* **2018**, *50*, 1–12.