ORIGINAL RESEARCH ARTICLE

Machine Learning-Based Diagnosis of Thermal Barrier Coating Process Quality

Dongjie Sun[1,3] · Qing He[1,2] · Zhi Huang[1,3]

Abstract Based on machine learning algorithms, a method is proposed for quality diagnosis of atmospheric plasma spraying (APS) processes used in thermal barrier coatings with determined coating materials and processes, aiming to swiftly evaluate the quality of APS coatings. First, the three-dimensional morphology of the coating is reconstructed through surface interpolation fitting, employing one-dimensional morphology standards and abnormal training set samples of the plasma-sprayed thermal barrier coating. This algorithm enables the extraction of cross section data of the coating at any angle. The mapping relationship between the characteristic parameters of the Gaussian peak and the process and coating characteristics is thoroughly analyzed, and the 12-dimensional characteristic parameters are utilized to effectively represent the one-dimensional morphology samples. Subsequently, principal component analysis (PCA) and K-nearest neighbor (KNN) algorithms are employed for accurate prediction and classification of the process quality of coating samples. Additionally, an exploratory factor analysis (EFA) model is established to comprehensively depict the relationship between plasma spraying process parameters, the process, and the three-dimensional morphology of the coating. The experimental results show that the machine learning algorithm has high accuracy in quality diagnosis, and its robustness is further verified by K-fold cross-validation. When combined with the EFA model, the proposed method facilitates rapid feedback on process quality, enabling real-time evaluation. Overall, this innovative approach presents a novel solution for the quality diagnosis of atmospheric plasma spraying processes. The incorporation of machine learning techniques and the establishment of the EFA model contribute to enhanced efficiency and accuracy in the evaluation process, paving the way for advancements in thermal barrier coating applications.

Keywords APS · Gaussian peak · machine learning · quality diagnosis · thermal barrier coatings

Introduction

There are higher demands on aero engines' service life and thermal efficiency in the new aircraft specifications (Ref 1, 2). The high-pressure turbine blade is one of the core components of an aero engine (Ref 3), and it works in environments of high temperature and high pressure, high stress load, and strong airflow erosion. Through practice, it has been proved that thermal barrier coating technology is currently the most effective way to significantly increase the service temperature and life of high-temperature alloys (Ref 4-7), and the quality of coatings is one of the critical factors influencing performance of high-pressure turbine blade. Atmospheric plasma spraying (APS) is used widely to produce thermal barrier coatings, there are more than 50 process parameters that influence the coupling between plasma and powder by influencing particle temperature, velocity, molten state and particle distribution, which

✉ Qing He
    heqing@caams.org.cn

    Dongjie Sun
    sdj1127600951@163.com

1   Beijing Jinlunkuntian Special Machine Co., Ltd,
    Beijing 101113, China

2   Surface Engineering Research Institute, Chinese Academy of
    Agricultural Mechanization Sciences, Beijing 100083, China

3   College of Engineering, China Agricultural University,
    Beijing 100083, China

finally determine a coating's microstructure and performance (Ref 8, 9).

In order to guide the regulation of the APS process, some researchers established online monitoring means, such as online particle temperature, velocity monitoring and online performance (thickness, residual stress and mechanical properties) evaluation (Ref 10-13). Zhang et al. (Ref 14) explored the temperature and velocity trends of particles in the APS spraying process, as well as their connection with the coating morphology, and designed a comprehensive computational scheme for the integral solution of the control equations of the whole process of spraying, which achieved the characterization of the energy changes of their spraying process by a mathematical model. Ye et al. (Ref 15) used Spray Watch 2i (Oseir Ltd., Tampere, Finland) to monitor the process of supersonic plasma spraying, and investigated the relationship between the temperature of the particles in the spraying process and the variation of the spraying distance, which provided some theoretical references for the control of the spraying process parameters.

The protective ability of the thermal barrier coating to the sprayed substrate is determined by the microstructure of the coating, the macroscopic coating distribution and the interface morphology of the coating coverage area. In reference to Lorenzo-Bañuelos et al. (Ref 16), it is suggested that in order to optimize the three-dimensional morphological characteristics of coatings and their basic performance, it is important to establish a correlation between the standard process parameters of spraying and the morphological characteristics of coatings in practical spraying operations. In the realm of microscale morphology in coatings, the Gaussian distribution model has been widely applied for simulating coating profiles and predicting thickness (Ref 17-19). Furthermore, Wu et al. (Ref 20) have underscored the relationship between Gaussian function parameters and factors such as the actual feedstock flow rate, spray angle, and spraying distance. Trifa et al. (Ref 21) investigated the thickness distribution and interfacial morphology of APS coating profiles on a substrate. They made measurements of the coating profiles using two inspection devices: a coordinate measuring machine and a laser profiler. Additionally, the coating profile was fitted using a Gaussian function, and key parameters such as the maximum height, width at half-maximum height, area of deposit, and horizontal offset of the deposit peak relative to the reference line were extracted. These parameters were employed to define the geometric shape of the deposition model. A mathematical model was then established to elucidate the relationship between process parameters and the impact on the coating. Subsequently, Trifa et al. (Ref 22) analyzed the coating profile topography using this model and suggested the use of a set of parameters to comprehensively describe the essential features of the coating topography. This approach offers valuable insights for the process quality diagnosis of APS coatings. Although Trifa et al.'s research on coating profiles is primarily based on one-dimensional morphology (fixed-point deposition), it is essential to note that parameters such as the maximum height, full width at half maximum, and area of the one-dimensional Gaussian fitting curve directly impact the thickness of the final thermal barrier coating. The horizontal offset of the peak reflects the actual plasma jet center's deviation and the tilt degree of the molten particles, both of which directly influence the porosity of the final coating. Therefore, the study of coating profiles plays a crucial role in controlling the thermal insulation performance of the coating.

Machine learning lies at the core of the fourth industrial revolution, commonly referred to as Industry 4.0. In the realm of spraying research, Mauer et al. (Ref 23) shed light on the substantial advantages and heightened reliability stemming from the integration of machine learning tools in the current diagnosis and control of thermal spraying processes. Bobzin et al. (Ref 24) utilized Residual Neural Network (ResNet) and Support Vector Machine (SVM) for the efficient estimation of in-flight particle properties in APS. Zhu et al. (Ref 25) applied Convolutional Neural Network (CNN) to optimize APS process parameters, particularly beneficial for newcomers to the field. Liu et al. (Ref 26, 27) employed Artificial Neural Network (ANN) to predict High Velocity Oxygen Fuel (HVOF) sprayed coatings and developed two ANN models for improved process control. Kolluru et al. (Ref 28) utilized ANN to predict shear strength in spark plasma sintering (SPS), while Zhao et al. (Ref 29) used statistical tools to analyze the impact of process parameters on the porosity of YSZ coatings. In cold spray research, Wang et al. (Ref 30) employed feature selection and ANN to predict critical velocity. Canales et al. (Ref 31) utilized machine learning to describe the deposition window and predict efficiency in cold spraying, suggesting optimization of material properties. In Electrostatic Spray Deposition (ESD) research, Paturi et al. (Ref 32) used ANN and SVM to predict coating thickness, and a genetic algorithm (GA) to optimize ESD deposition parameters. The aforementioned studies offer valuable methods for controlling the spraying process and establishing a mapping model between process parameters and coating performance. However, these studies were unable to simultaneously establish a connection between these three components. Furthermore, model training requires a larger sample dataset to ensure its accuracy.

This paper employs laser three-dimensional scanning technology to achieve rapid detection of the one-dimensional morphology of plasma-sprayed thermal barrier

coatings. It conducts a digital analysis of the one-dimensional morphology, extracting Gaussian peak features from the coating cross section. Based on PCA-KNN and EFA models, the paper accomplishes swift diagnosis of coating quality and establishes the intrinsic correlation among coating process parameters, spraying processes, and coating quality. This provides a novel approach for diagnosing the process quality of thermal barrier coatings.

Experimental Procedures

Sample Preparation

Compared with the whole spraying process of industrial production, one-dimensional morphology coating is formed by fixed-point spraying, and two-dimensional morphology coating is formed by reciprocating linear motion spraying in the horizontal direction of the spray gun. At this time, the vertical feed motion is superimposed to form three-dimensional morphology coating. Macroscopically, it is similar to the process of point moving into line, line moving into surface and surface moving into body. The two-dimensional and three-dimensional morphology cannot directly reflect the coupling and migration of plasma jet and particles in all directions due to the stacking of molten particles. The one-dimensional morphology formed by fixed-point spraying provides a new perspective to directly study the influence of process parameters on quality, which lays a foundation for the subsequent study of two-dimensional morphology and three-dimensional morphology.

The one-dimensional morphology coating was prepared by using nano-agglomerated 8YSZ powder (38-61 μm, Beijing Jinlunkuntian Special Machine Co., Ltd., Beijing, China), was sprayed by 9 MB torch (Beijing Zhenbang Aerospace Precision Machinery Co., Ltd., Beijing, China), and the spraying power was 45 kW. The physical center of the anode outlet of the spray gun is perpendicular to the physical center of the sample. The spraying time of each sample was 10 s. Since in the plasma spraying process, the deviation variable generally consists of plasma gun power (affected by current, voltage and H_2 flow), powder feeding parameters, main gas flow (A_r flow) and nozzle usage time, the experimental parameters of the anomalous samples under the deviation process were selected based on these four perspectives. Concerning powder injection parameters, although the powder injection angle affects particle temperature and velocity, and the injection distance influences particle divergence width, research indicates that their impact on coating performance is significantly less than that of the carrier gas flow rate (Ref 33). In industrial batch production, once process parameters are established,

they are less susceptible to influences from the working environment and nozzle manufacturing errors. Therefore, this paper does not consider the injection angle and injection distance as deviation variables but focuses solely on the carrier gas flow rate as the variable for deviation.

Table 1 presents the detailed parameters of the APS (atmospheric plasma spraying) process in the experiment. Fig. 1 displays samples from the one-dimensional morphology training set. As the spraying state may undergo slight changes with an increase in nozzle working time during spraying, standard process parameters are consistently employed in the creation of standard samples. A new nozzle is utilized to consecutively spray S-1 to S-20, capturing not only the subtle variations in the spraying state but also enhancing the diversity of standard samples. In the preparation of abnormal samples, the impact of nozzle working time growth is considerably less significant than the influence of nozzle health status and process parameters on the coating. Therefore, four nozzles with varying degrees of burnout are used for the preparation of UB-1 to UB-4, and four nozzles exceeding their service life are employed in the creation of UO-1 to UO-4. For other types of samples that deviate from the standard process parameters, because the range of process parameters can be set is wider than the health status of the nozzle, new nozzles are used in the preparation, and 3 parameters are set up above and below the standard value, and 6 samples are prepared for each type. Finally, a total of 46 training set samples were obtained.

The coating morphology of the one-dimensional standard samples of fixed-point spraying under the standard process is almond-shaped, and the microscopic profile of the coating morphology under other deviation processes is greatly different from the profile under the standard process, so it is feasible to diagnose the spraying process by the microscopic profile characteristics of the coating.

Digital Representation

Due to the small quantity of the training set samples, it was difficult to extract image features in order to conduct process diagnosis by using image processing models commonly used in deep learning (such as convolutional neural networks, recurrent neural networks, etc.), so this design attempted to extract the features from the three-dimensional morphology of the coating, and to analyze the dataset by using machine learning algorithms applicable to small sample sizes.

There is a coupling between plasma and powder during the forming process of APS thermal barrier coatings, which leads to a Gaussian distribution on the bonding surface of the coating. Therefore, the analysis of Gaussian curves on

Table 1 APS process parameters of training set samples

| Training samples	Powder feeding rate, g/min	Powder carrier gas flow, L/min	Main gas (Ar) flow rate, L/min	Arc current, A	Voltage, V	Spray distance, mm	Spray angle, °
S-1-S-20	22.5	7	28	600	75	85	90
UB-1-UB-4	22.5	7	28	600	75	85	90
UO-1-UO-4	22.5	7	28	600	75	85	90
UF-1	22.5	2	28	600	75	85	90
UF-2	22.5	3	28	600	75	85	90
UF-3	22.5	4	28	600	75	85	90
UF-4	22.5	10	28	600	75	85	90
UF-5	22.5	11	28	600	75	85	90
UF-6	22.5	12	28	600	75	85	90
UP-1	22.5	7	28	460	75	85	90
UP-2	22.5	7	28	480	75	85	90
UP-3	22.5	7	28	500	75	85	90
UP-4	22.5	7	28	700	75	85	90
UP-5	22.5	7	28	720	75	85	90
UP-6	22.5	7	28	740	75	85	90
UA-1	22.5	7	18	600	75	85	90
UA-2	22.5	7	20	600	75	85	90
UA-3	22.5	7	22	600	75	85	90
UA-4	22.5	7	32	600	75	85	90
UA-5	22.5	7	34	600	75	85	90
UA-6	22.5	7	36	600	75	85	90
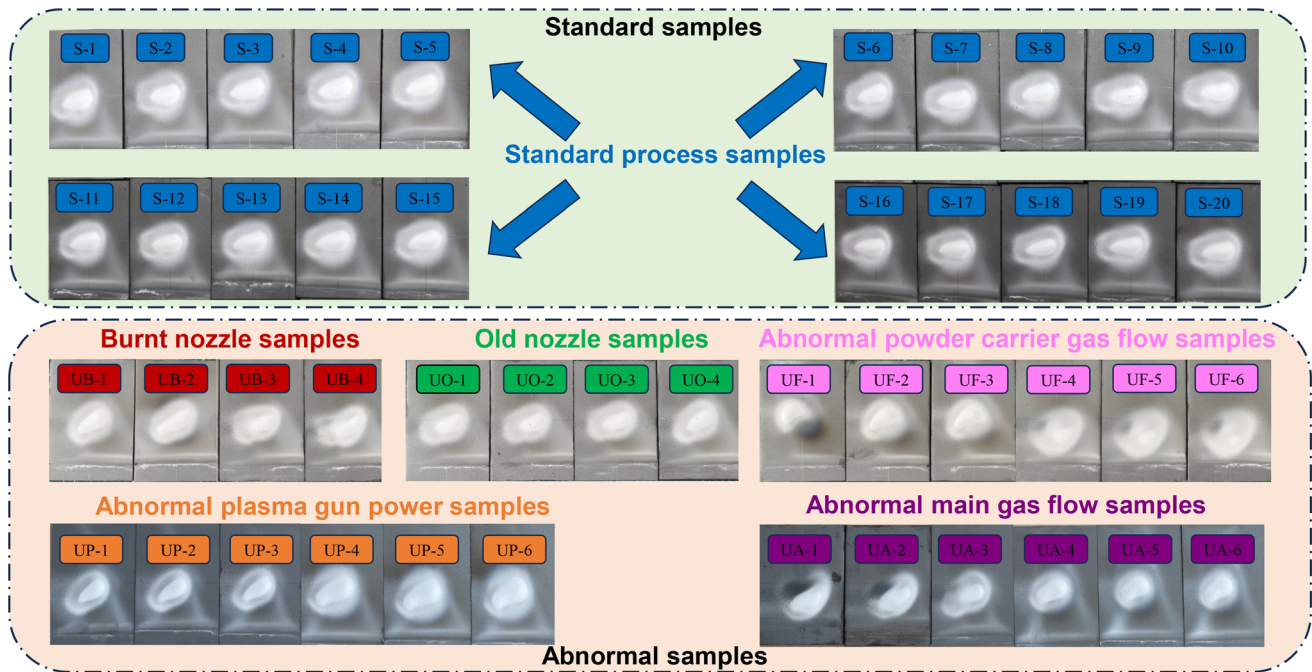
Fig. 1 One-dimensional morphology samples of the training set

the coating is essential for research on coating performance. The Gaussian function is as follows:

$$f(x) = Ae^{-\frac{(x-X_c)^2}{2\omega^2}} \tag{Eq 1}$$

A, X_c and ω are all real numbers, and, $A > 0$. where A is the peak value of Gaussian curve, X_c is the corresponding horizontal coordinates, and ω is the standard deviation. Fig. 2 is a Gaussian curve parametric plot. In the plot, full width at half maximum FWHM numerically equal to $2.354 * \omega$, the Gaussian peak area is the area bounded by the curve and the horizontal line.

In order to show the differences of the plasma spray process under different process parameters through one-dimensional morphology of the coatings, the Gaussian peak characteristic parameters were associated with the plasma spraying process to construct a characterization relationship, as shown in Table 2.

The laser 3D profilometer DS1050 (Cognex Corporation, Natick, Massachusetts, America) was used to scan the one-dimensional morphology samples to obtain (X, Y, Z) 3D point cloud data, with each sample consisting of approximately 4500 coordinate points (Fig. 3a). In order to extract more precise 3D morphological data of the coating, surface interpolation was used to reconstruct the original surface in 3D. Based on the initial point cloud space of the samples, a meshed plane of dimensions 150×150 was regenerated, and the interpolated meshed surface data were obtained using the cubic interpolation technique in the 3D space (Fig. 3).

In order to obtain the main feature information of the coating, a new algorithm was designed, which takes the vertical line from the highest point in the coating profile to the sample base plane as the axis, and extracts the data of the cr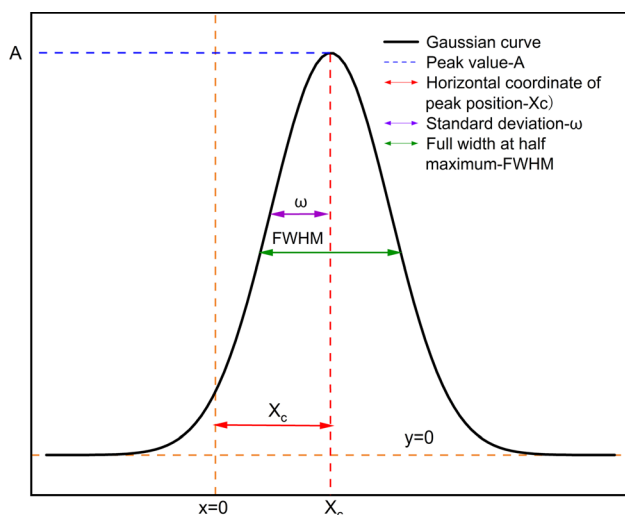oss section curve of the profile at any angle based on this axis. Fig. 4(a) and (b) show the schematic diagram of extracting the cross section data at any angle, and the flowchart of the algorithm, respectively. The principle of the algorithm is shown in Fig. 4. The main idea is to screen the actual three-dimensional shape data points whose distance from the section is less than the given error, and project them onto the section to obtain the two-dimensional data of the section.

Figure 5 shows the parameter information of peak height (A), offset (X_c), full width at half maximum (FWHM) and Gaussian peak area (Area) after Gaussian fitting of standard samples and each type of abnormal samples at 0°-180° cross section. It is evident that samples produced by the burnout nozzle and old nozzle exhibit lower A values compared to the standard sample. Furthermore, other parameters for both nozzles show non-smooth transitions across different cross sections, displaying some irregularities. Samples manufactured with deviations in the powder gas flow rate parameter have a significant impact on both A values and FWHM values, thereby influencing the melting state of particles. Deviations in the main gas flow rate parameter and spray power parameter have a pronounced effect on the X_c value, indicating a substantial displacement between the morphological center and the spray center. In contrast, the standard sample maintains consistently high levels for A, FWHM, and Area, with a smaller X_c. The transitions across different cross sections are smoother in the standard sample. Consequently, the coating profile of the standard sample is notably superior to that of the abnormal samples.

Figure 6 shows the Gaussian fitting curve of the coating cross section of sample number "S-7" and the original curve. The 0°, 45° and 90° cross section curves indirectly reflect the coupling effect of plasma and powder in the horizontal, vertical and inclined directions, showing the combination of molten particles and the surface, which is highly representative in the coating forming. In order to simplify the characterization parameters of the coating, the A, X_c, FWHM and Area of the three typical sections are used as the 12-dimensional features of the coating after digitization.

Method of Analysis

Principal Component Analysis (PCA) and K-Fold

Due to the limited size of the training dataset and the high dimensionality of the data, employing certain machine learning classification algorithms may lead to overfitting, potentially encountering the "curse of dimensionality" (Ref 34). In such cases, the application of techniques such as Principal Component Analysis (PCA) and K-fold cross-

Fig. 2 Gaussian curve characteristic parametric plot

Table 2 Characterization relationship between Gaussian peak characteristic parameters and the plasma spraying process

| Gaussian peak characteristic parameters	Plasma spraying process
A (peak value of Gaussian curve)	The degree of aggregation and particle flux in the high-temperature zone of the plasma jet (near the center of flame flow). The higher the value of A, the higher the degree of coupling of the powder particles to the high-temperature plasma.
X_c (horizontal coordinate of peak value)	Represents the offset of the physical center X_0 relative to the position of the nozzle outlet center where the gun is vertically aligned, and X_c represents the offset of the high-temperature and high-particle flux deposition position relative to the physical center when $X_c = 0$.
FWHM (Full width at half maximum)	Represents the degree of aggregation of powder in the high-temperature zone of plasma. The larger the value of FWHM, the flatter the coating profile, the lower the degree of aggregation of powder particles in the plasma high-temperature zone, and the larger the degree of dispersion of the delivered powder.
Area (Gaussian peak area)	To some extent, it can characterize the deposition efficiency of powder particles under the same spraying process and powder feeding conditions
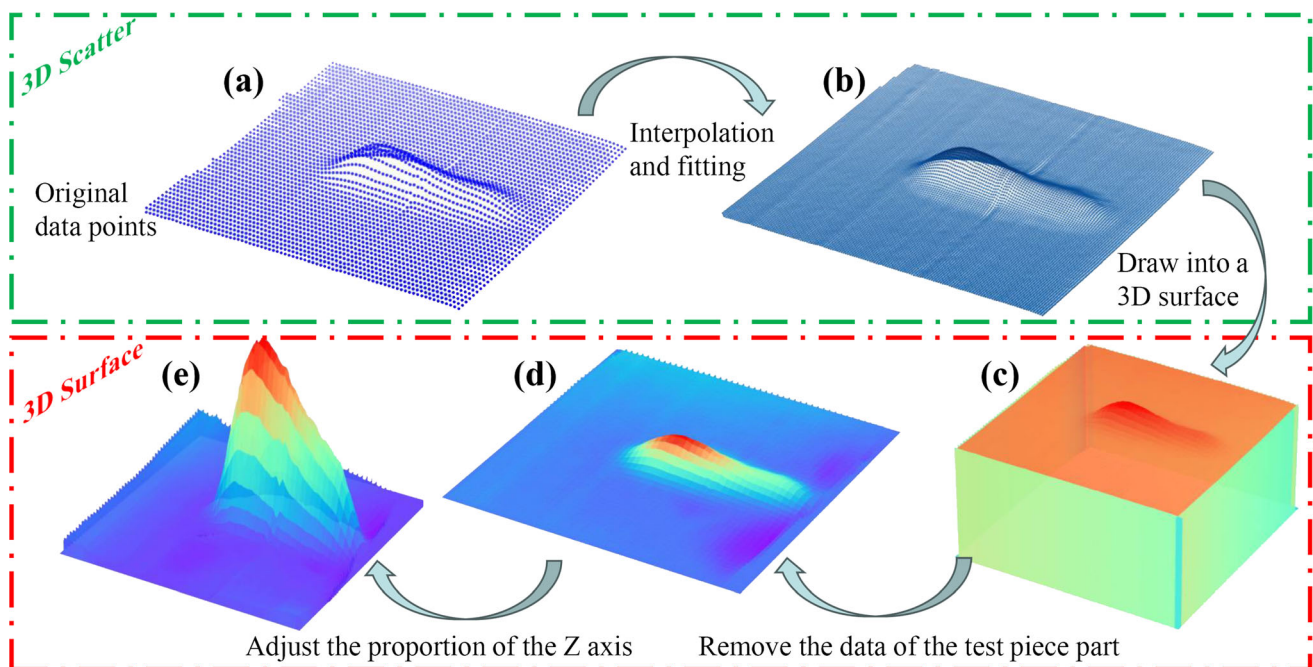
Fig. 3 The process of coating data extraction

validation becomes crucial for mitigating overfitting and ensuring the model generalizes well across different datasets (Ref 35, 36). PCA is a method that transforms the original data into a set of principal components, representing the most significant structures and variations within the data. This transformation effectively reduces the dimensionality of the data by eliminating redundant or uncorrelated features, thereby enhancing the model's performance (Ref 37-39). Assuming the input data matrix is represented by X,

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix} \tag{Eq 2}$$

Normalize the matrix X to obtain the normalized matrix X_s:

$$X_s = \left[X - (1, 1, \ldots, 1)^T x_{av} \right] \cdot \text{diag}\left(\frac{1}{s_1}, \frac{1}{s_2}, \ldots, \frac{1}{s_m} \right) \tag{Eq 3}$$

In (3), $x_{av} = (x_{av,1}, x_{av,2}, \ldots, x_{av,m})$, and s_j is the corresponding standard deviation.
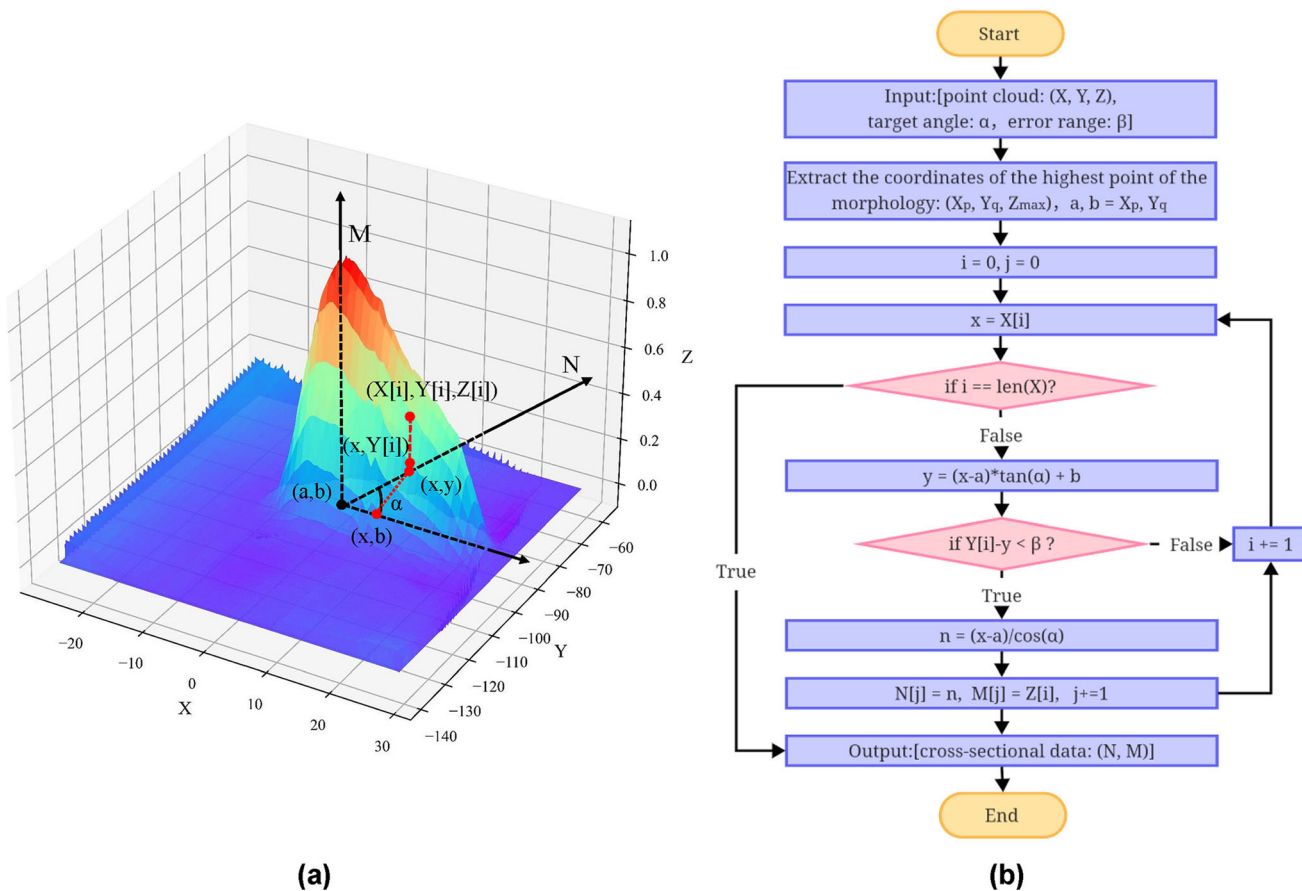
Fig. 4 Principle of the algorithm for extracting cross section data at any angle ((a) Schematic diagram for extracting cross section data at any angle, (b) the algorithm flowchart for extracting cross section data at any angle)

The coefficient matrix R is calculated as:

$$R = \frac{1}{m} X_s^T X_s \qquad (\text{Eq } 4)$$

Find the m eigenvalues of X and arrange the eigenvalues from largest to smallest as $\lambda_1, \lambda_2, \ldots, \lambda_n$ to obtain the arranged eigenvalue M_σ matrix as:

$$M_\sigma = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n) \qquad (\text{Eq } 5)$$

The associated eigenvectors can be determined from the eigenvalues. In practical scenarios, the selection of principal component scores is guided by the cumulative contribution to the variance.

The decomposition of X_s yields:

$$X_s = T \times P^T + E_s, T = X \times P \qquad (\text{Eq } 6)$$

In Eq 6, P is the principal molecular space loading matrix, T is the score matrix, E_s is the residual matrix.

K-fold cross-validation is another essential strategy to address overfitting. It involves dividing the dataset into K subsets, training the model on K-1 subsets, and validating it on the remaining subset. This process is repeated K times, with each subset serving as the validation set exactly once

(Ref 36). The overall performance is then averaged, providing a more robust evaluation of the model's ability to generalize to diverse data.

K-Nearest Neighbor (KNN)

Due to an imbalance in the number of standard and anomalous samples within the training dataset, the application of classification algorithms such as Support Vector Machine (SVM), Decision Tree (DT), and K-means may encounter challenges. However, the K-Nearest Neighbor (KNN) algorithm, being proximity-based, exhibits relative robustness in the context of imbalanced datasets. This resilience stems from its reliance on the distances between samples for classification (Ref 40, 41).

In the classification process, KNN computes the Euclidean distance between a new instance and instances in the training dataset. Given two n-dimensional samples: $x_i = \left[x_i^1, x_i^2, \ldots, x_i^n\right]$ and $x_j = \left[x_j^1, x_j^2, \ldots, x_j^n\right]$, the Euclidean distance d_{ij} is defined as follows:

Fig. 5 The subgraphs (a)-(h) are the peak height (A), offset (X_c), full width at half maximum (FWHM) and Gaussian peak area (Area) of the 0°-180 ° section of the coating profile of the standard sample and various abnormal samples, respectively. Among them, the process parameter deviations of abnormal samples in subgraphs (a), (c), (e) and (g) are set below the standard range, and subgraphs (b), (d), (f) and (h) are set above the standard range.
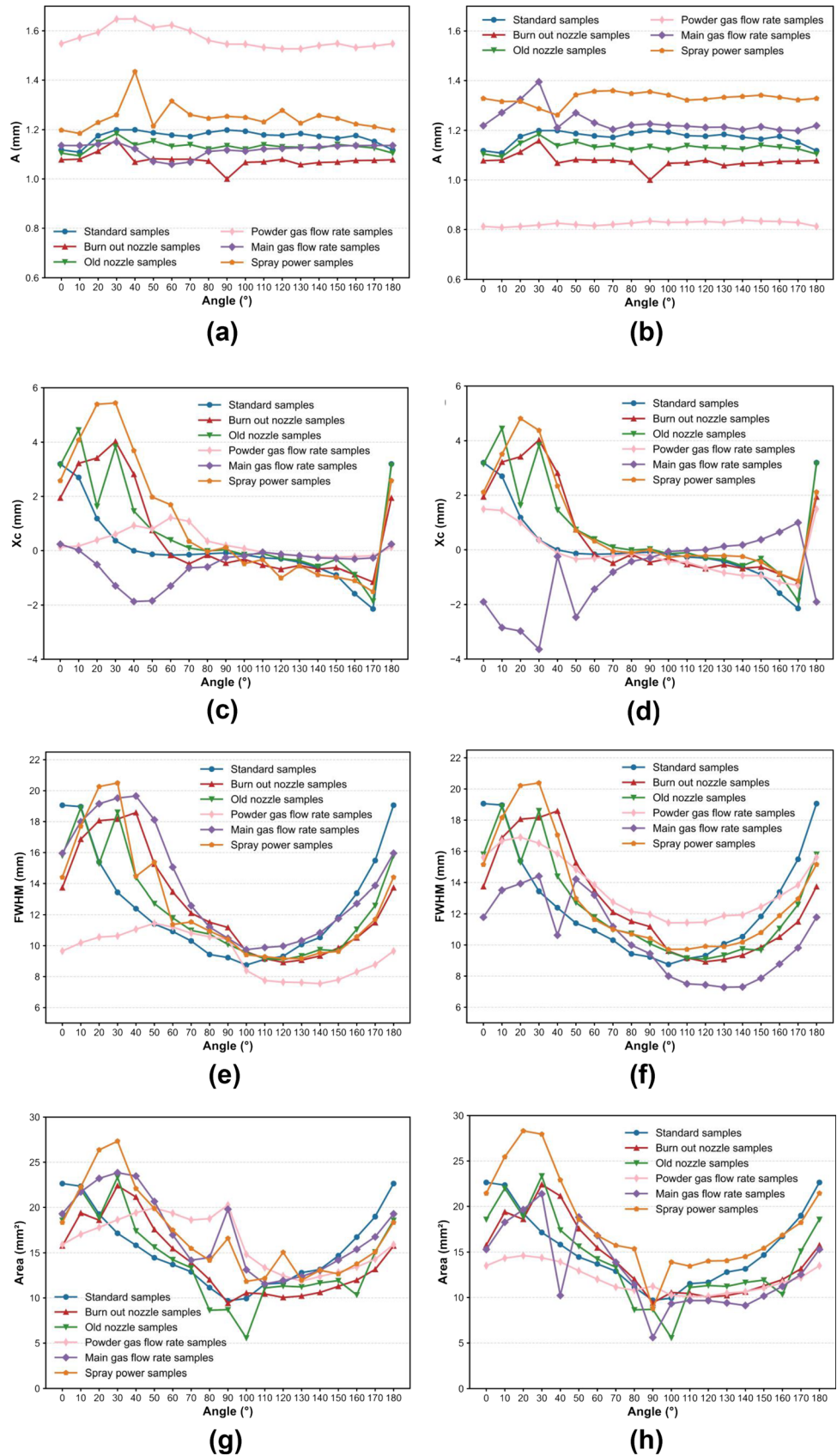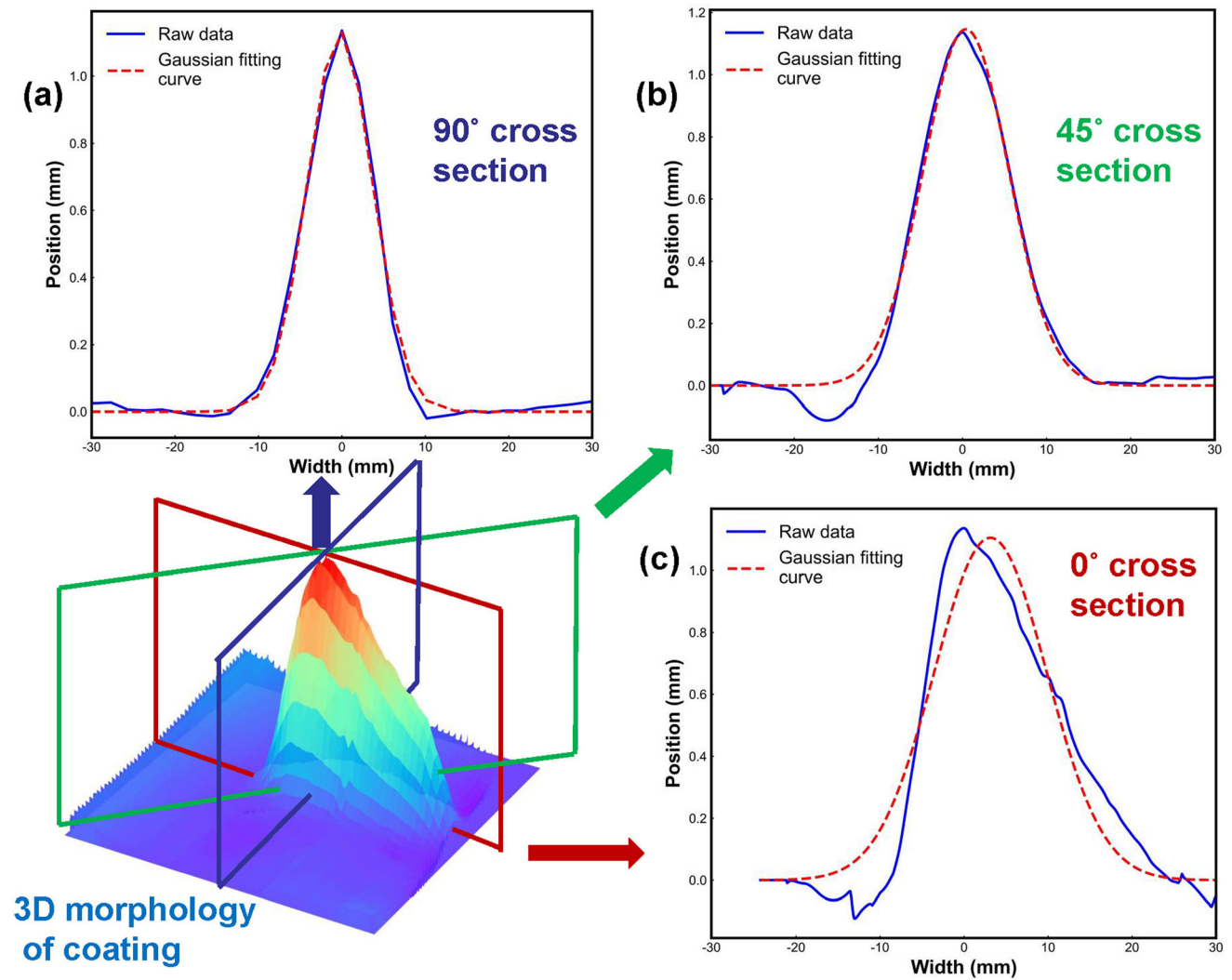
Fig. 6 No. "S-7" sample coating cross section Gaussian fitting curve and the original curve

$$d_{ij} = \sqrt{\sum_{k=1}^{n}(x_i^k - x_j^k)^2} \qquad (Eq\ 7)$$

Subsequently, it selects the closest k instances as neighbors and predicts the class based on their classifications (Ref 42). Given that KNN employs Euclidean distance as the distance metric, the data used for classification can either be the original 12-dimensional dataset or the 2-dimensional dataset resulting from PCA dimensionality reduction.

Exploratory Factor Analysis (EFA)

In order to analyze characteristic dimensions that deviate significantly between the anomalous samples and the standard samples on the coating morphology, Exploratory Factor Analysis (EFA) was applied. EFA is a multivariate statistical method aimed at expressing the complex

relationships between combinations of underlying factors by decomposing the covariance matrix between variables (Ref 43, 44). The matrix form is given by:

$$X = AF + \varepsilon \qquad (Eq\ 8)$$

In (Eq 8),

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix}, A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pm} \end{bmatrix}, F = \begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_m \end{bmatrix}, \varepsilon$$

$$= \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_p \end{bmatrix}$$

$$(Eq\ 9)$$

We denote F as the common factors or latent factors of X, A as the factor loading matrix, and ε as the unique

factors of X. $A = a_{ij}$, where a_{ij} represents the factor loading. Mathematically, it can be proven that the factor loading a_{ij} is the correlation coefficient between the i-th variable and the j-th factor, reflecting the importance of the i-th variable on the j-th factor.

In the algorithmic principle of the method, the eigenvalues of the correlation matrix correspond to the ideally chosen coordinate axes, while the eigenvectors represent the orientation of the data in the rotated coordinate system. The eigenvalues reflect the variance of the data in the corresponding dimension, while larger eigenvalues imply that the corresponding eigenvector direction contains more information. By analyzing the factor loadings of factors with large eigenvalues, it is possible to calculate the extent to which the original data contributes to the underlying factors, thus better describing the variability in the data (Ref 45, 46). By combining the Gaussian feature covariate characterization relationship with the factor loadings generated by the EFA model, the factor analysis can be performed by combining data from different sample sets, e.g., standard samples and anomalous samples under a certain deviation process, to find out the feature dimensions of the deviation process that have the greatest potential impact on the original data, and then to derive the dimension that has the greatest potential impact of each deviation process on the physical changes of the process, and thus to build a model of the relationship among the parameters of the plasma spraying process, the process, and the coating's three-dimensional topographic characteristics (shown in Fig. 7) in order to predict and optimize the coating's morphological characteristics, as well as to achieve better process control and quality improvement.

Results and Discussion

The Preparation of the Testing Sample Set

In order to use fewer test samples to reflect the morphological characteristics of the standard and anomalous classes of coatings, and then verify the effectiveness of the diagnostic model, Fig. 8 shows the set of test samples, including 2 samples of standard class (TS-1-TS-2), 1 sample of burnt nozzle (TB-1), 1 sample of old nozzle exceeding the service life (TO-1), 2 samples of abnormal powder carrier gas flow (TF-1-TF-2), 2 samples of abnormal main gas flow (TA-1-TA-2), and 2 samples of abnormal plasma gun power (TP-1-TP-2), making 10 cases in total (For the specific process parameter settings of the test set, refer to the training set parameters in Table 1, taking the median of the upper and lower deviation ranges for each type of anomaly sample).

Classification Prediction Algorithm Validation

Firstly, to mitigate the impact of different scales within the dataset, all sample data is standardized using the z-score. In order to select appropriate principal components that effectively eliminate redundant information in multidimensional data and prevent model overfitting, the covariance matrix of the standardized dataset is computed. The eigenvalues are then arranged in descending order, and the cumulative variance contribution corresponding to each eigenvalue (principal component) is illustrated in Fig. 9. Choosing 2 to 3 principal components captures a substantial portion of the information (variance contribution). For a more concise extraction of meaningful information and ease of visualization, 2 principal components are selected.

Figure 10 shows the test set of the PCA and KNN, through the PCA method, with sample data from 12-dimensional features down to two-dimensional features. Through the visualization, as shown in Fig. 10(a), which reflects the distribution of the training samples and test samples of each type after the reduction to 2 dimensions, there is a certain clustering tendency in the two-dimensional data distribution of the standard class and anomalous class samples after the reduction of dimensionality. Therefore, there is not only a big difference in the microscopic morphology of coatings, but there is also a large amount of information even after the digitization and data dimensionality reduction, so it is feasible to digitalize the samples, and the PCA algorithm is well validated.

Using the data after dimensionality reduction to validate the KNN algorithm, Fig. 10(b) shows the data visualization of the PCA-KNN test sample set. The species of the three training set samples closest to the test samples can be determined by calculating the Euclidean distance between the test samples and the training set samples, then prediction of the species of the test samples is available. After calculating, the accuracy of the KNN algorithm in predicting the test samples using the data after dimensionality reduction is 100%, and the accuracy of the prediction using the 12-dimensional original data before dimensionality reduction is 90%. Table 3 displays the actual categories, predicted results, and accuracy of the classification algorithm for each test sample.

Due to the relatively small sample size, potential overfitting may exist in this training set. The PCA-KNN model is tested using K-fold cross-validation, as illustrated in Fig. 11. For each fold, the confusion matrix is obtained, and performance metrics, including accuracy, precision, recall, and F1 score, are calculated. Specifically:

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}} \quad \text{(Eq 10)}$$
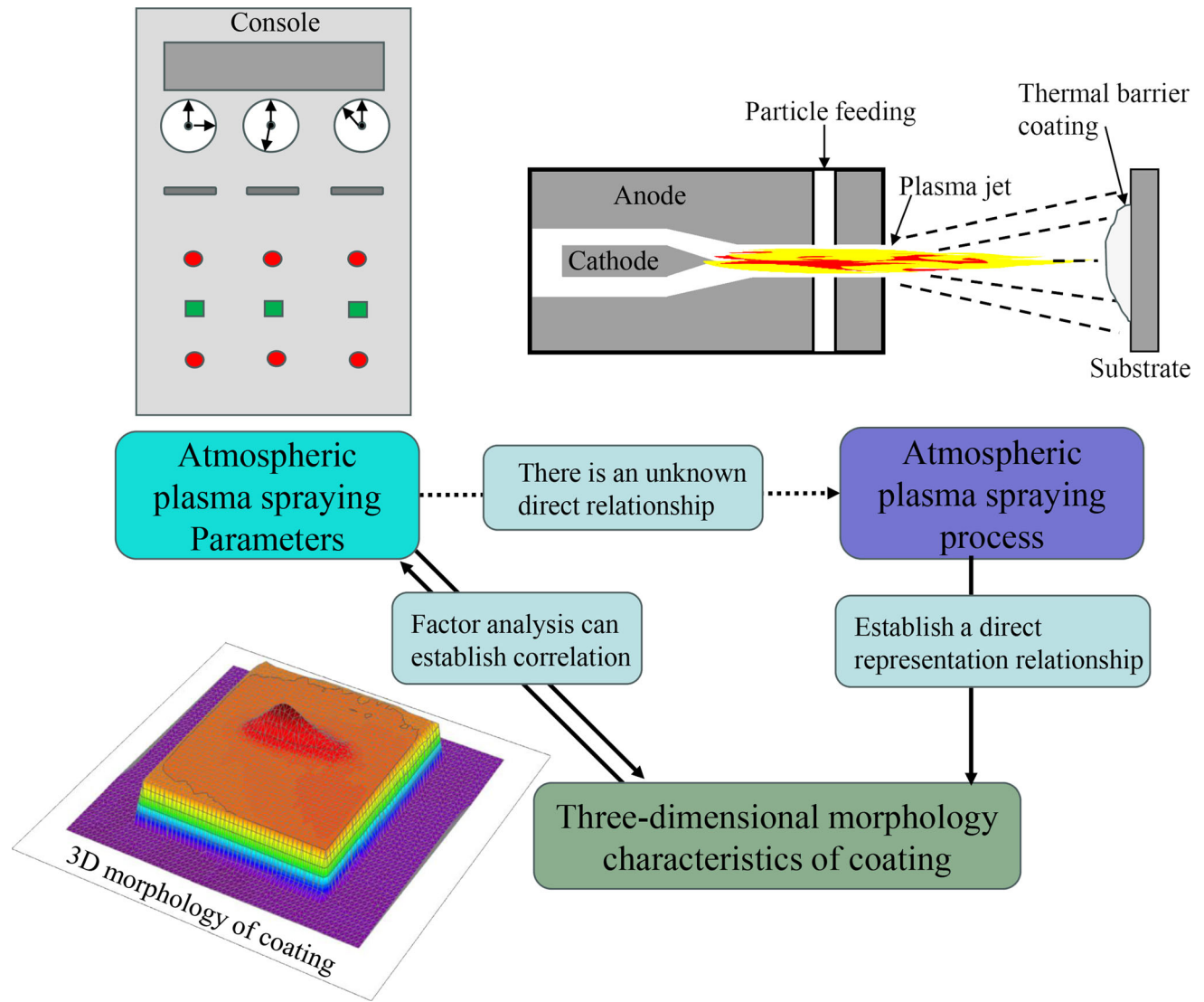
Fig. 7 Relationships between plasma spray process parameters and plasma spray process and the 3D morphological characteristics of coatings
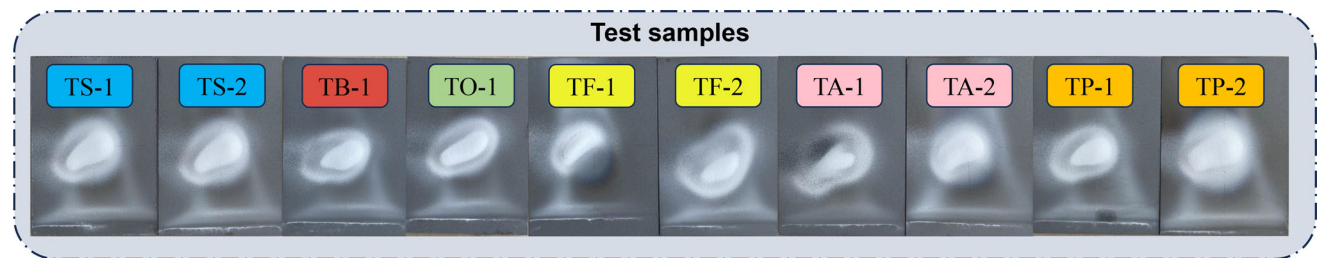

Fig. 8 Test sample set

$$\text{Precision} = \frac{\text{True positives}}{\text{True positives} + \text{False positives}} \qquad \text{(Eq 11)}$$

$$\text{Recall} = \frac{\text{True positives}}{\text{True positives} + \text{False negatives}} \qquad \text{(Eq 12)}$$

$$\text{F1 - score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \qquad \text{(Eq 13)}$$

After conducting a 5-fold cross-validation, the PCA-KNN model demonstrated an average accuracy of 94.7%. Additionally, the other three relevant metrics exhibited high values, suggesting its capability to perform well
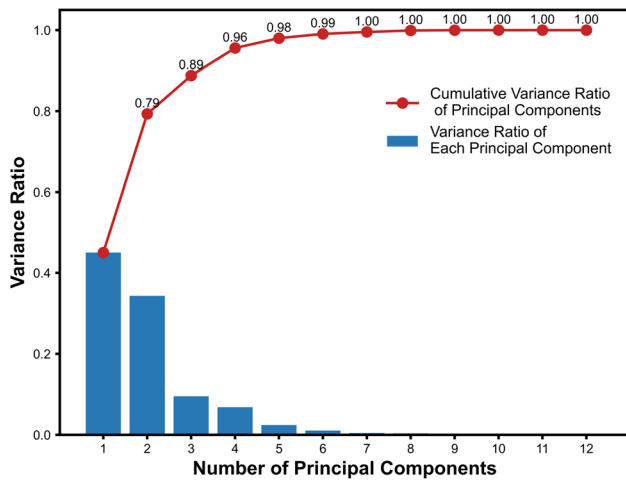
Fig. 9 Cumulative Variance Contribution of Principal Components (Eigenvalues of the Covariance Matrix)

across various data subsets. These results highlight the model's robust generalization ability and feasibility for practical applications.

Table 3 Actual categories of test samples and predicted results from the classification algorithm

Algorithm	Actual	KNN	PCA-KNN
Test samples	Classification (Standard 1, Abnormal 0)		
TS-1	1	1	1
TS-2	1	1	1
TB-1	0	1	0
TO-1	0	0	0
TF-1	0	0	0
TF-2	0	0	0
TA-1	0	0	0
TA-2	0	0	0
TP-1	0	0	0
TP-2	0	0	0
Accuracy (%)	…	90	100

Process Feedback Method

It can be known from the principle of the EAF model in the "Method of analysis" section of this article, the eigenvalues of the covariance matrix represent a measure of the variance of the data in the corresponding dimension on the
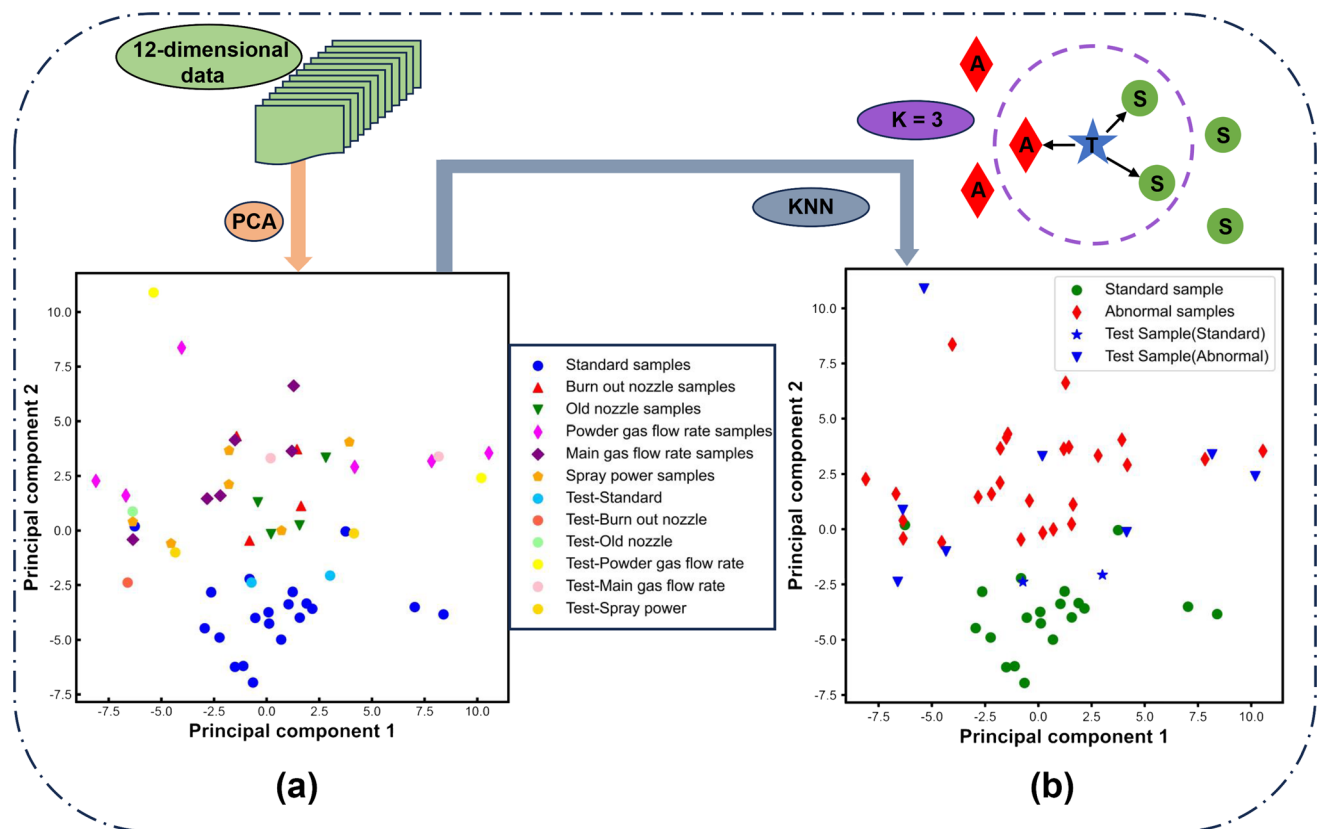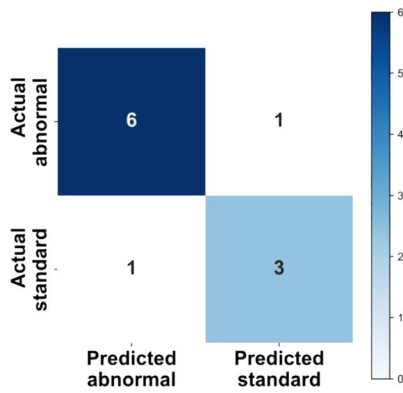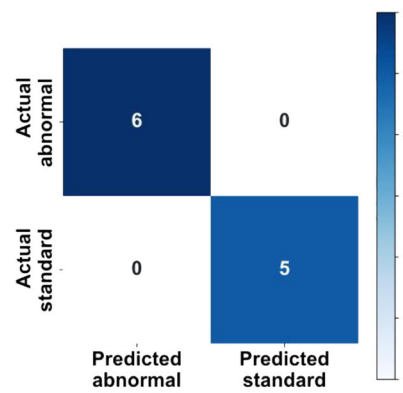

Fig. 10 Test set PCA and KNN

(a)

(b)

(c)

(d)

(e)

(f)

rotated coordinate system. The samples' factor analysis of standard and deviation process (abnormal powder feed gas flow) is shown in Fig. 12, and in Fig. 12(a), it can be seen that larger eigenvalues correspond to eigenvectors in the direction of data with greater variability, and four factors can cover the majority of the amount of information of the original data. Combining the magnitude of the eigenvalues of the selected factors and the factor loadings generated by the EFA model (shown in Fig. 12b), it is possible to get the raw characteristics of the main effects of this deviant process.

The ranking of the influence of feature parameters under different process parameters based on the EFA model is shown in Fig. 13, and as shown, it analyzes all the factors of anomalous samples and standard samples under deviated processes to get the relationships between deviated process parameters and Gaussian peak feature parameters of the coating morphology, which can reveal the influence of different types of deviated process samples on the coating morphology and establish a correlation model between
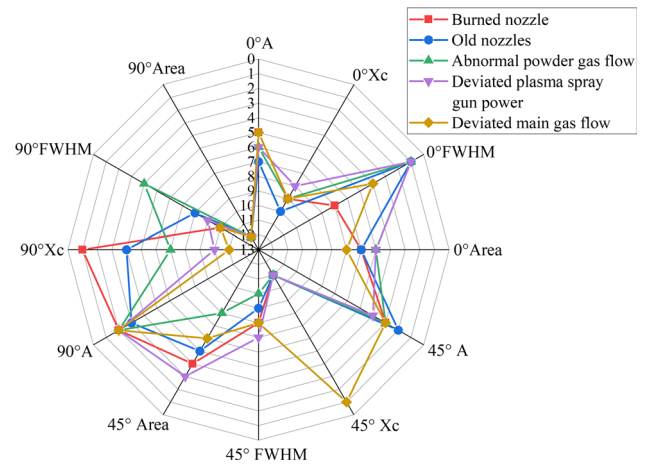

Fig. 13 Ranking of influence of characteristic parameters under different process parameters based on EFA model

them (the radial data in Fig. 13 indicates the ranking of the contribution of the feature parameters).

In practical engineering applications, with the comprehensive prediction classification algorithm and the EFA model, before workpiece spraying, the fixed-point APS process can be used to produce test specimens, according to the PCA-KNN algorithm to determine whether the process meets the requirements of the standard process. If the requirements are met, the workpiece can be coated.
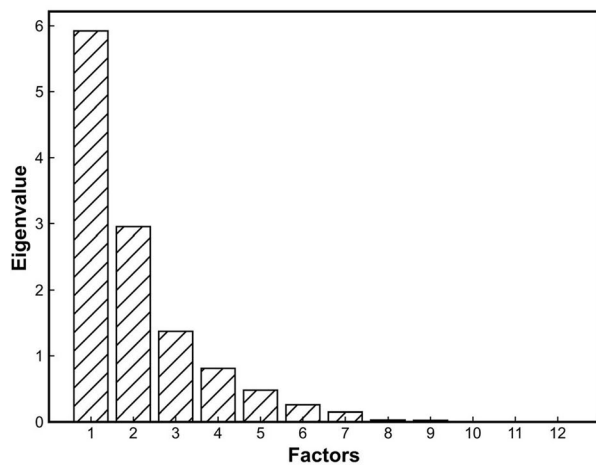
(a)
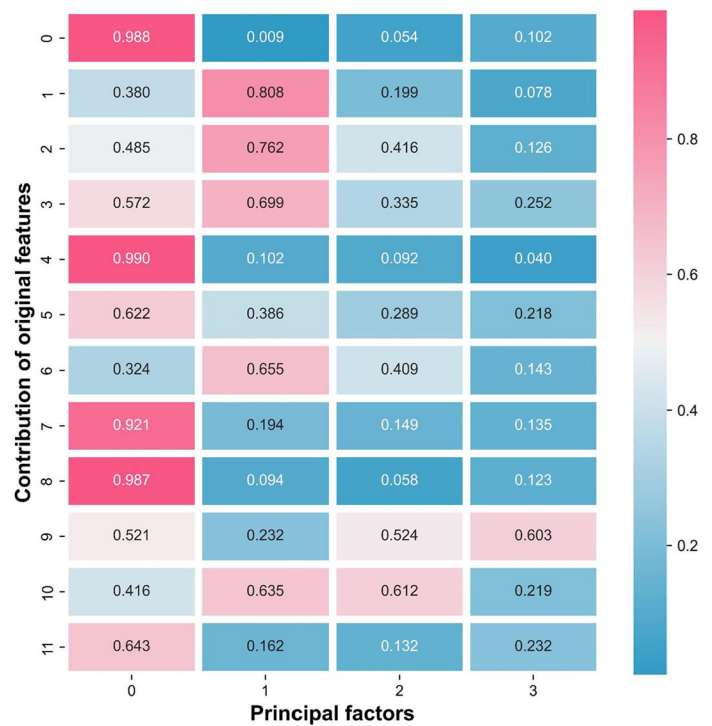
(b)

Fig. 12 Factor analysis of samples from standard and deviant processes (abnormal powder feed gas flow), ((a) Plot of factor number and eigenvalue relationship. (b) Heat map of EFA factor loadings)

Otherwise, it is necessary to obtain the original features with the largest difference between the test piece and the sample under the standard process, and combine the relationship model of process parameters-process-coating three-dimensional morphology characteristics established by EFA to adjust the relevant spraying process parameters until the new test piece is judged to meet the standard process.

Conclusions

Based on machine learning algorithms, this study proposes a process quality diagnosis method for thermal barrier coatings using macroscopic three-dimensional surface features obtained through atmospheric plasma spraying. The main conclusions are as follows:

1. After reconstructing the three-dimensional morphology of one-dimensional coating samples, the cross sectional morphology exhibits a Gaussian peak distribution. The study establishes a correlation between Gaussian peak feature parameters and process characteristics, enabling a digital analysis of feature parameters for standard and anomalous process samples, with extraction of typical feature data for analysis.
2. The effectiveness of the classification algorithm was assessed, with the PCA-KNN algorithm demonstrating high accuracy in the classification and diagnosis of process quality for coated samples. Following K-fold cross-validation, it achieved an average accuracy of 94.7%.
3. By standardizing the coating material, equipment, and developing a standard process, this study utilizes exploratory factor analysis to establish a correlation model between process parameters and coating characteristic morphology, enabling the diagnosis of process abnormalities. Furthermore, through additional data accumulation and secondary development, the model can be swiftly applied to different plasma spraying coatings, facilitating the rapid diagnosis of coating abnormalities in the mass production process.

References

1. J. Ni, K. Shi, S. Xue, J. Zhao, S. Liu, H. Liu, and Z. Li, Development and Outlook of Ceramic Materials for Thermal Barrier Coatings in Aircraft Engines, *Mater. Rep.*, 2021, **35**(S1), p 163-168.
2. P.G. Lashmi, P.V. Ananthapadmanabhan, G. Unnikrishnan et al., Present Status and Future Prospects of Plasma Sprayed Multilayered Thermal Barrier Coating Systems, *J. Eur. Ceram. Soc.*, 2020, **40**(8), p 2731-2745.
3. M. García-Martínez, J.C. del Hoyo Gordillo, M.P. Valles González, A. Pastor Muro, and B. González Caballero, Failure Study of an Aircraft Engine High Pressure Turbine (HPT) First Stage Blade, *Eng. Fail. Anal.*, 2023, **149**, p 107251.
4. C. Zhu, A. Javed, P. Li et al., A Study of the Microstructure and Oxidation Behavior of Alumina/Yttria-Stabilized Zirconia (Al2O3/YSZ) Thermal Barrier Coatings, *Surf. Coat. Technol.*, 2012, **212**, p 214-222.
5. N.P. Padture, M. Gell, and E.H. Jordan, Thermal Barrier Coatings for Gas-Turbine Engine Applications, *Science*, 2002, **296**(5566), p 280-284.
6. A.G. Evans, D.R. Mumm, J.W. Hutchinson, G.H. Meier, and F.S. Pettit, Mechanisms Controlling the Durability of Thermal Barrier Coatings, *Prog. Mater. Sci.*, 2001, **46**(5), p 505-553.
7. D.R. Clarke and, C.G. Levi, Materials Design for the Next Generation Thermal Barrier Coatings, *Annu. Rev. Mater. Res.*, 2003, **33**(1), p 383-417.
8. J. Sun et al., *Thermal Spray Science and Technology*, Metallurgical Industry Press, Beijing, 2013.
9. E. Bakan and R. Vaßen, Ceramic Top Coats of Plasma-Sprayed Thermal Barrier Coatings: Materials, Processes, and Properties, *J. Therm. Spray Technol.*, 2017, **26**, p 992-1010.
10. L.M. Chen and Q. Li, Current Status and Development of Plasma Spraying Technology, *Heat Treat. Technol. Equip.*, 2006, **27**(1), p 1-5.
11. P. Fauchais, G. Montavon, M. Vardelle, and J. Cedelle, Developments in Direct Current Plasma Spraying, *Surf. Coat. Technol.*, 2006, **201**(5), p 1908-1921.
12. P. Fauchais, Understanding Plasma Spraying, *J. Phys. D Appl. Phys.*, 2004, **37**(9), p R86.
13. Z. Wang, S. Huang, J. Xue et al., Current Status and Progress of Plasma Spraying Equipment, *China Surf. Eng.*, 2000, **4**, p 5-7.
14. D. Zhang, L. Zhang, X. Hu, and H. Zhang, Numerical Studies of Arc Plasma Generation in Single Cathode and Three-Cathode Plasma Torch and Its Impact on Plasma Spraying, *Int. J. Heat Mass Transf.*, 2016, **98**, p 508-522.
15. X. Ye, Z. Li, and Y. Zhu, Research on Plasma Spraying Particle Temperature Based on CCD Image Processing Technology, *New Technol. New Process*, 2005, **9**, p 59-60.
16. M. Lorenzo-Bañuelos, A. Díaz, D. Rodríguez, et al., Influence of Atmospheric Plasma Spray Parameters (APS) on the Mechanical Properties of Ni-Al Coatings on Aluminum Alloy Substrate, *Metals*, 2021, **11**, p 612.
17. J. Liu, R. Bolot, and S. Costil, Residual Stresses and Final Deformation of an Alumina Coating: Modeling and Measurement, *Surf. Coat. Technol.*, 2015, **268**, p 241-246.
18. M. Liu, H. Wu, Z. Yu et al., Description and Prediction of Multilayer Profile in Cold Spray Using Artificial Neural Networks, *J. Therm. Spray Technol.*, 2021, **30**, p 1453-1463.
19. T. Van Hoof, G. Fradet, F. Pichot et al., Simulation of Thermal Spray Coating on 3D Objects: Numerical Scheme and Aeronautic Test Case, *J. Therm. Spray Technol.*, 2019, **28**, p 1867-1880.
20. W. Hongjian, X. Xie, M. Liu, C. Chen, H. Liao, Y. Zhang, and S. Deng, A New Approach to Simulate Coating Thickness in Cold Spray, *Surf. Coat. Technol.*, 2020, **382**, p 125151.
21. F.I. Trifa, G. Montavon, C. Coddet, P. Nardin, and M. Abrudeanu, Geometrical Features of Plasma-Sprayed Deposits and Their Characterization Methods, *Mater. Charact.*, 2005, **54**(2), p 157-175.
22. F.I. Trifa, G. Montavon, and C. Coddet, On the Relationships Between the Geometric Processing Parameters of APS and the Al2O3–TiO2 Deposit Shapes, *Surf. Coat. Technol.*, 2005, **195**(1), p 54-69.

23. G. Mauer and C. Moreau, Process Diagnostics and Control in Thermal Spray, *J. Therm. Spray Tech.*, 2022, **31**, p 818-828.

24. K. Bobzin, W. Wietheger, H. Heinemann et al., Prediction of Particle Properties in Plasma Spraying Based on Machine Learning, *J Therm. Spray Technol.*, 2021, **30**, p 1751-1764.

25. J. Zhu, X. Wang, L. Kou, L. Zheng, and H. Zhang, Prediction of Control Parameters Corresponding to In-Flight Particles in Atmospheric Plasma Spray Employing Convolutional Neural Networks, *Surf. Coat. Technol.*, 2020, **394**, p 125862.

26. M. Liu, Y. Zexin, Y. Zhang, W. Hongjian, H. Liao, and S. Deng, Prediction and Analysis of High Velocity Oxy Fuel (HVOF) Sprayed Coating Using Artificial Neural Network, *Surf. Coat. Technol.*, 2019, **378**, p 124988.

27. M. Liu, Z. Yu, H. Wu et al., Implementation of Artificial Neural Networks for Forecasting the HVOF Spray Process and HVOF Sprayed Coatings, *J. Therm. Spray Technol.*, 2021, **30**, p 1329-1343.

28. S.K.P. Kolluru, S.A. Kalvakolanu, S.G. Chinthapanti, and S.T. Palakurthy, Machine Learning Approach to Model and Predict the Shear Strength of TZM-Graphite Joint Bonded by Spark Plasma Sintering, *Mater. Today Proc.*, 2023, **92**, p 1482–1492.

29. Y. Zhao, F. Peyraut, M.P. Planche et al., Experiments, Statistical Analysis, and Modeling to Evaluate the Porosity Influence in SPS Coatings, *J. Therm. Spray Technol.*, 2019, **28**, p 76-86.

30. Z. Wang, S. Cai, W. Chen et al., Analysis of Critical Velocity of Cold Spray Based on Machine Learning Method with Feature Selection, *J. Therm. Spray Technol.*, 2021, **30**, p 1213-1225.

31. H. Canales, I.G. Cano, and S. Dosta, Window of Deposition Description and Prediction of Deposition Efficiency Via Machine Learning Techniques in Cold Spraying, *Surf. Coat. Technol.*, 2020, **401**, p 126143.

32. U.M.R. Paturi, N.S. Reddy, S. Cheruku, S.K.R. Narala, K.K. Cho, and M.M. Reddy, Estimation of Coating Thickness in Electrostatic Spray Deposition by Machine Learning and Response Surface Methodology, *Surf. Coat. Technol.*, 2021, **422**, p 127559.

33. G. Mauer, R. Vaßen, D. Stöver et al., Improving Powder Injection in Plasma Spraying by Optical Diagnostics of the Plasma and Particle Characterization, *J. Therm. Spray Technol.*, 2011, **20**, p 3-11.

34. I. Ulaq, D.M. Khan, M. Hamraz, N. Iqbal, A. Ali, and Z. Khan, Optimal K-NEAREST Neighbours Based Ensemble for Classification and Feature Selection in Chemometrics Data, *Chemometr. Intell. Lab. Syst.*, 2023, **240**, p 104882.

35. A. Liang, H. Yunpeng, and G. Li, The Impact of Improved PCA Method Based on Anomaly Detection on Chiller Sensor Fault Detection, *Int. J. Refrig.*, 2023, **155**, p 184-194.

36. J. Li, F. Gao, S. Lin, M. Guo, Y. Li, H. Liu, S. Qin, and Q.Y. Wen, Quantum k-Fold Cross-Validation for Nearest Neighbor Classification Algorithm, *Phys. A: Stat. Mech. Appl.*, 2023, **611**, p 128435.

37. H. Li, *Statistical Learning Methods*, Tsinghua University Press, Beijing, 2012.

38. B. Ghojogh, F. Karray, and M. Crowley, Eigenvalue and Generalized Eigenvalue Problems: Tutorial. arXiv preprint arXiv:1903.11240.

39. F.L. Gewers, G.R. Ferreira, H.F.D. Arruda, F.N. Silva, C.H. Comin, D.R. Amancio, and L.D.F. Costa, Principal Component Analysis: A Natural Approach to Data Exploration, *ACM Comput. Surv.*, 2021, **54**(4), p 1-34.

40. S. Zhang, X. Li, M. Zong, X. Zhu, and D. Cheng, Learning k for knn Classification, *ACM Trans. Intell. Syst. Technol.*, 2017, **8**(3), p 1-19.

41. M.L. Zhang and Z.H. Zhou, ML-KNN: A Lazy Learning Approach to Multi-Label Learning, *Pattern Recogn.*, 2007, **40**(7), p 2038-2048.

42. M. Bansal, A. Goyal, and A. Choudhary, A Comparative Analysis of K-Nearest Neighbor Genetic, Support Vector Machine, Decision Tree, and Long Short Term Memory algorithms in machine learning, *Decis. Anal. J.*, 2022, **3**, p 100071.

43. M.C. Howard, A Review of Exploratory Factor Analysis Decisions and Overview of Current Practices: What We Are Doing and How Can We Improve, *Int. J. Hum. Comput. Interact.*, 2016, **32**(1), p 51-62.

44. H. Treiblmaier and P. Filzmoser, Exploratory Factor Analysis Revisited: How Robust Methods Support the Detection of Hidden Multivariate Data Structures in IS Research, *Inf. Manag.*, 2010, **47**(4), p 197-207.

45. S.E. Kim, I.W. Seo, and S.Y. Choi, Assessment of Water Quality Variation of a Monitoring Network Using Exploratory Factor Analysis and Empirical Orthogonal Function, *Environ. Modell. Softw.*, 2017, **94**, p 21-35.

46. L.B. Iantovics, C. Rotar, and F. Morar, Survey on Establishing the Optimal Number of Factors in Exploratory Factor Analysis Applied to Data Mining, *Wiley Interdiscip. Rev. Data Mining Knowl. Discov.*, 2019, **9**(2), p e1294.