Article citation info:
Peng Y, Hai W, Xue P. Fault diagnosis analysis of gas turbine engine based on fuzzy algorithm. Diagnostyka. 2025;26(1):2025109. https://doi.org/10.29354/diag/200870.

DIAGNOSTYKA, 2025, Vol. 26, No. 1

e-ISSN 2449-5220
DOI: 10.29354/diag/200870

FAULT DIAGNOSIS ANALYSIS OF GAS TURBINE ENGINE BASED ON FUZZY ALGORITHM

Yana PENG [1,*], Wenjiong HAI [1], Peng XUE [2]

[1] School of Aviation Maintenance Engineering, Chengdu Aeronautic Polytechnic, Chengdu 610100, China
[2] Engineering Technology Training Center, Civil Aviation University of China, Tianjin 300300, China
[*] Corresponding author, e-mail: yanapeng@126.com

Abstract

To solve the problems of low efficiency and difficult feature extraction in traditional fault diagnosis methods, this study proposes an optimized Fuzzy C-Means clustering algorithm for diagnosing and analyzing gas turbine engine faults. This algorithm mainly introduces subtraction clustering, penalty factors, and data weights on the basis of the original fuzzy C-means clustering algorithm, thereby improving the generalization ability of the algorithm model and the credibility of the results. The optimized fuzzy C-means clustering algorithm had the highest level of accuracy value, with a value of 95.67%, which was 11.79% higher than the average accuracy of other algorithms. Meanwhile the optimized Fuzzy C-Means clustering algorithm improved the accuracy values of KNN, BP, SVM, and fuzzy C-means clustering algorithms by 19.65%, 12.26%, 3.55%, and 11.70%. The training set accuracy of the optimized fuzzy C-means clustering algorithm under four engine states was at the highest level, with an average improvement of 15.5%, 25%, 24%, and 16% in accuracy. The optimized fuzzy C-means clustering algorithm achieved an accuracy of 90.39% in the test set, with an average improvement of 16.13% in accuracy. The membership classification results indicated that the optimized fuzzy C-means clustering algorithm had a membership degree of 1.

Keywords: fuzzy c-means clustering algorithm, laplacian eigenmaps, fault diagnosis, turbine engine, value of affiliation

List of Symbols/Acronyms

Symbol	Explanation
i	Data in X
j	Data in X
W	Matrix of weighting factors
L	Laplace matrix
D	Diagonal matrix
t	Thermonuclear constant
J	Value function
N	Capacity
u	Degree of affiliation
c	Number of centroids
p	Number of categories
d	Distance metric
v	Center of clustering
r_a	Neighborhood radius
r_b	Corrected Neighborhood Scope
m	Distance between category center store and single point
a	Penalty factor

1. INTRODUCTION

With the development of China's industrial level, the performance of gas turbine engines is gradually improving. The increasingly complex working environment increases the probability of damage to key components such as turbine disks and blades of gas turbine engines, which in turn leads to increased operating and maintenance costs after being put into use [1-2]. Gas turbine engine is a power plant widely used in aviation, power and industrial fields. A gas turbine engine produces high temperature and high pressure gas by burning fuel, which pushes the turbine to rotate, thus outputting power. The performance of the engine is affected by a variety of factors, including intake conditions, mechanical condition, and combustion efficiency. Troubleshooting is a vital component of ensuring the reliability and safety of engines, preventing potential failures and performance degradation, and reducing downtime and repair costs. Gas turbine engine fault diagnosis refers to the reasonable judgment of the performance status and lifespan of each component based on the monitoring results after monitoring the

Received 2024-09-20; Accepted 2025-02-05; Available online 2025-02-10

relevant working parameters of each component of the engine [3-4]. The types of faults in gas turbine engines can be roughly divided into vibration faults, gas path component faults, and control system faults, among which gas path faults are more common [5-6]. The vibration faults mainly come from rotor unbalance, air flow instability, and gear meshing instability of transmission box. Gas path failure mainly refers to the type of failure caused by supercharger, turbine blade, intake pipe, and cooling system. Control system faults mainly refer to the types of faults caused by sensors, actuators, control system software, and electrical systems. As a widely used iterative clustering algorithm, Fuzzy C-Means clustering (FCM) algorithm has been applied by scholars in rapid diagnosis of diabetes and tumor classification at this stage, and has shown good fast classification ability and high accuracy [7-8]. In recent years, scholars have researched the problem of turbine engine failures. Sarwar et al. proposed a hybrid model of convolutional neural network and dimensionality reduction algorithm to improve the accuracy of fault information in gas turbine engines, and applied it to the operating conditions of gas turbine engines. The proposed hybrid model had high accuracy and an error of only 0.00173 [9]. Kordestani et al. proposed a parallel structure of Laguerre filter combined with fuzzy neural network based on heavy-duty gas turbine engines to predict the remaining service life of gas turbine engines, and applied it to estimate compressor fouling accumulation and filter defects. The established model had higher accuracy and estimation efficiency compared to other fault diagnosis and prediction models [10]. Xiong et al. used simulated annealing and genetic algorithm to optimize the simulated c-means clustering algorithm to overcome the sensitivity of the FCM to local minimum problems, and applied the improved FCM to the diagnosis of bearing faults. The optimized FCM improved accuracy by 9.22%, and compared to the introduction of the model, the accuracy of fault diagnosis increased by 15.56% [11]. Cheng et al. proposed an intelligent online sensor fault diagnosis algorithm for gas engine fault diagnosis to obtain high-precision data on the working state of gas engines. On the relevant dataset, the research algorithm significantly reduced the diagnosis time compared to other algorithms, with a diagnosis time of only 0.233s and an accuracy rate of over 95% [12]. Feng K et al. proposed a method for calculating the blade frequency of gas turbine blade faults based on sparse harmonic product spectrum, which combines Vold Kalman filtering with adaptive parameter optimization process. This method can separate blade related vibration from magazine vibration even under strong noise conditions, and the results show that the model can discover potential faults of blades earlier and more accurately [13]. Fahmi A T W K et al. applied a long short-term memory model to turbine fault detection and used a threshold method to determine the likelihood of turbine faults. The results showed that the accuracy of the model for turbine vibration anomalies reached more than 86% [14]. However, these methods mostly focus on studying individual fault types and have not validated the credibility of relevant evaluation indicators. Therefore, to solve the problem of gas turbine engine fault diagnosis, this study proposes an improved FCM. The innovation lies in introducing subtraction clustering, data weight, and penalty factors for optimization on the basis of the original FCM, which can improve the generalization ability and credibility of the algorithm. Meanwhile, the improved FCM introduces subtraction clustering and other optimization modules to improve the algorithm's efficacy and reliability in fault data classification.

2. METHODS AND MATERIALS

2.1. Extraction of operating data based on gas turbine engine

During the operation of a turbine engine, the measurement parameters recorded by sensors are the main data source for fault diagnosis. The relevant measurement parameters of a gas turbine engine can be divided into direct measurement and indirect measurement parameters. Direct measurement parameters are one of the components of state monitoring parameters, including engine inlet and outlet pressure, temperature, and fan material speed [15-16]. Indirect measurement parameters are obtained through calculations based on direct measurement parameters, mainly including engine flow rate, boost ratio, and fuel consumption rate. In light of the alterations observed in measurement-related parameters, it is imperative to ascertain the associated changes in component performance, delineate the pertinent fault types, and formulate suitable troubleshooting plans and methods. The principle of gas turbine engine failure is shown in Fig. 1.

Fig. 1 illustrates the logical relationship between influencing factors, changes in component performance, and the parameters to be tested. In the diagnosis of gas turbine engine faults, the strong coupling of data information and high noise content make it difficult to use relevant data fault information in engine fault diagnosis. Prior to conducting engine fault diagnosis, this study introduces Locality Preserving Projection (LPP) algorithm to reduce the dimensionality of turbine engine data and extract fault information. The core concept of LPP algorithm is based on a relatively rare Laplacian Eigenmaps (LE) algorithm, mainly for the purpose of dimensionality reduction [17-18]. In the initial space, the parameter dimension of dataset $X=[x_1, x_2, x_3, \cdots, x_n]$ is defined as m, and correlation transformation is performed based on matrix A to obtain the reduced spatial dataset

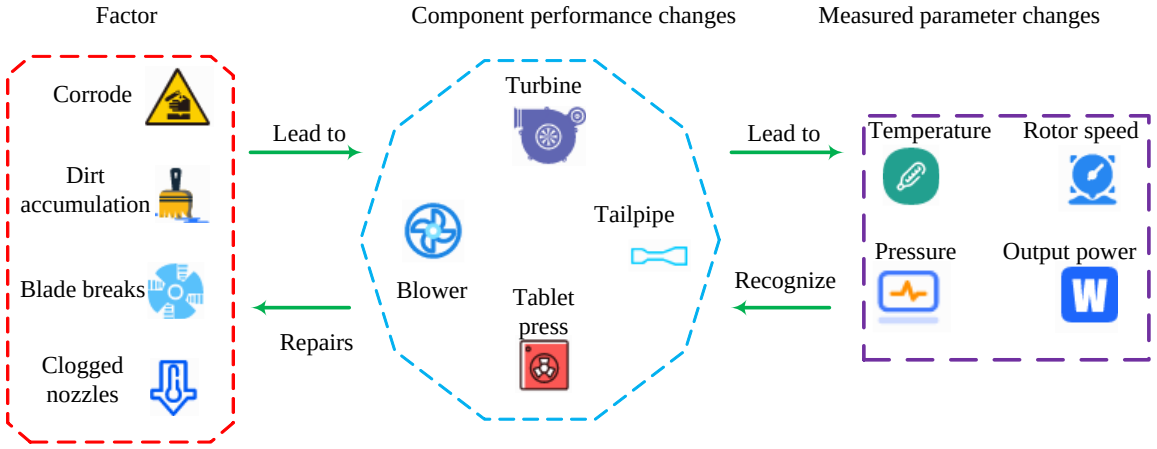


Fig. 1. Schematic diagram of gas turbine engine gas circuit troubleshooting

$Y=[y_1,y_2,y_3,\cdots,y_n]$. The final result is that the dimension m' of dataset Y is smaller than the dimension m of the original dataset. The calculation formula for improving the original algorithm and considering the global structure is shown in equation (1).

$$\sum_{ij}(y_i-y_j)^2W_{ij}-\frac{(y_i-\bar{y})^2}{n} \tag{1}$$

In equation (1), i and j respectively represent different data in X. W_{ij} represents a matrix composed of distance weight coefficients between i and j. To avoid eliminating scaling factors, the process of solving eigenvalues after adding relevant constraints is shown in equation (2).

$$\begin{cases} A^T XDX^T A=1 \\ (B-XLX^T)A=\lambda XDX^T A \\ B=\sum_{i=1}^{N}(x_i-\bar{x})(x_i-\bar{x})^T/N \\ L=D-W \\ D_{ii}=\sum_j W_{ji} \end{cases} \tag{2}$$

In equation (2), λ represents the eigenvalue. L represents the Laplacian matrix. D represents the diagonal matrix. The maximum value of D, D_{ii}, is taken. First, a data adjacency graph is constructed based on the K-Nearest Neighbor (KNN) algorithm. Due to the KNN algorithm's ability to adjust the value of k, it provides flexibility for the algorithm to adapt to different datasets and problem requirements. Moreover, the KNN model's ability to enhance with the increase of training samples makes it highly adaptable when dealing with dynamically changing datasets. The relevant schematic of the KNN algorithm is shown in Fig. 2.

In Fig. 2, the green triangle at the center point can be classified as a pink circle or a blue square. After describing the nearest neighbor relationship based on the above algorithm, if either data x_i or x_j are within the closest region of another data k, then a relationship line is used to connect x_i and x_j. Next, the weights of the corresponding relationship lines are selected and a thermonuclear function is introduced to calculate the weighted values of the connection lines between nodes, as shown in equation (3).

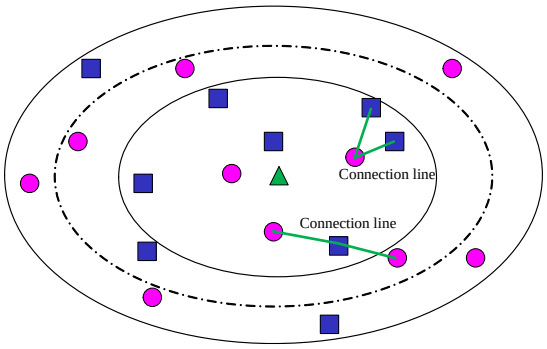


Fig. 2. Correlation diagram of KNN algorithm

$$W_{ij}=e^{\frac{\|x_i-x_j\|^2}{t}} \tag{3}$$

In equation (3), W_{ij} represents the weighted value, which is inversely proportional to the distance between x_i and x_j, where t represents the thermal constant. Finally, based on equation (2), the correlation matrix is calculated and information about the characteristic solution is obtained. Finally, the column vectors of the above matrices are combined to generate the transformation matrix A.

2.2. Gas engine fault diagnosis based on FCM

After extracting the fault characteristics of gas turbine engines through the LPP algorithm, the FCM is introduced to objectively diagnose the fault types of gas turbine engines. The likelihood of various fault types occurring is represented by membership indicators, and the fault diagnosis process is shown in Fig. 3.

In Fig. 3, after improving the FCM, credibility judgment and fault type identification can be performed based on the clustering results. The algorithm is based on fuzzy theory, and each target

obtains membership degrees of different numerical values on this basis. This result can reflect the interpretability of a single target for a certain category. The essence of minimizing the value function lies in the sum of squared errors between a single data point and the classification. The expression of the value function is shown in equation (4).

$$J_m = \sum_{i=1}^{N}\sum_{j=1}^{c} u_{ij}^m d_{ij}^2 \tag{4}$$

In equation (4), J_m represents the value function. N represents the number of capacities. u_{ij} represents the degree of membership. c represents the number of cluster center points. m represents the number of categories. d represents the distance measure between data or between cluster centers and data. Based on the value function, the Lagrange function is constructed to determine the membership degree and cluster center point, and its function expression is shown in equation (5).

$$J = \sum_{i=1}^{N}\sum_{j=1}^{c} u_{ij}^m d d_{ij}^2 + \sum_{i=1}^{N} \lambda_i (\sum_{j=1}^{c} u_{ij} - 1) \tag{5}$$

Based on equation (5), by taking the derivative of the membership function u_{ij} and substituting it into $\sum_{k=1}^{c} u_{ki} = 1$, the relevant expression for membership is shown in equation (6).

$$u_{ij} = 1 / \sum_{k=1}^{c} (\frac{d_{ij}}{d_{kj}})^{\frac{1}{m-1}} \tag{6}$$

In equation (6), after calculating the partial derivative based on the Lagrange function and removing non-zero coefficients, the equation based on the cluster center is shown in equation (7).

$$v_k = \frac{\sum_{i=1}^{n} u_{ki}^m x_i}{\sum_{i=1}^{n} u_{ki}^m} \tag{7}$$

In equation (7), v represents the cluster center. The type and quantity of gas turbine failures are unknown, and the location of the number of data points is uncertain. In view of this, this study introduces subtraction clustering combined with FCM to calculate data point density and determine the number of cluster centers, to alleviate the problem of inaccurate diagnosis due to parameter reasons. The core concept of subtractive clustering is that each sample in the dataset can be defined as a cluster, and the minimum cluster group is merged until a certain stopping condition is reached. Meanwhile, the density of the remaining data is corrected based on the point with the highest density, to avoid clustering in the same cluster of data [19-20]. Based on the number and location of clusters in the dataset, the distribution of data can be characterized, and the indicator of correlation density is shown in equation (8).

$$D_i = \sum_{j=1}^{n} \exp \| (-\frac{\| x_i - x_j \|^2}{(0.5 r_a)^2}) \tag{8}$$

In equation (8), r_a represents the neighborhood radius of the cluster center point, and the density contribution of data points outside the radius range of this point based on this point is 0, as expressed in equation (9).

$$r_a = \frac{1}{2} \min_k \left\{ \max_i \{ \| x_i - x_k \| \} \right\} \tag{9}$$

After determining the density value according to equation (9), the size of the density values of each data point is compared to determine whether the maximum value satisfies the iteration condition. If the iteration conditions are met, the initial cluster centers and number are obtained. If it is incorrect, corrections are made. The iteration conditions and correction equation are shown in equation (10).

$$\begin{cases} \frac{D_{ci+1}}{D_{ci}} \le \delta \\ D_i = D_i - D_{ck} \exp(-\frac{\| x_i - x_{ck} \|^2}{(0.5 r_b)^2}) \end{cases} \tag{10}$$

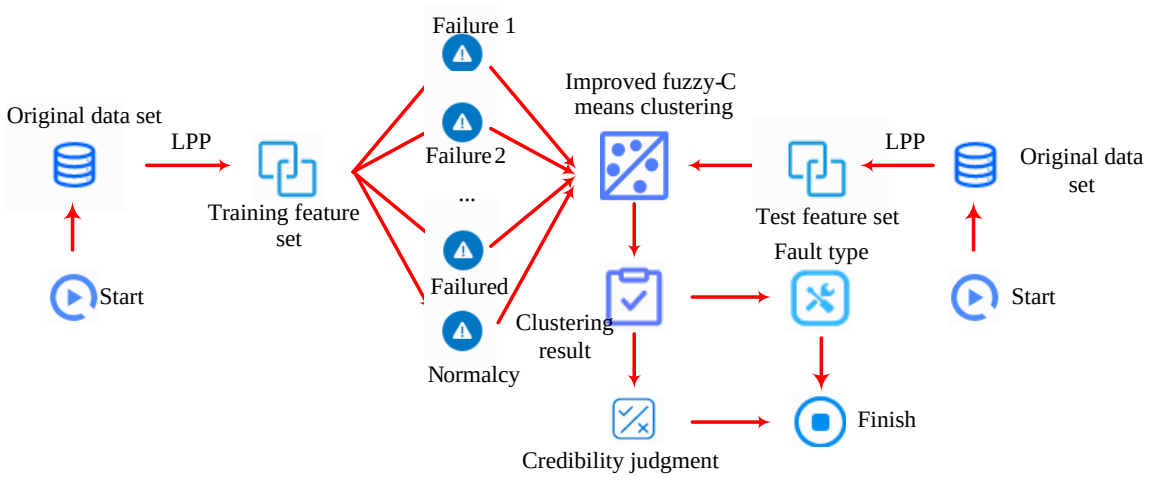


Fig. 3. Framework diagram of troubleshooting process

In equation (10), r_b represents the neighborhood range that must be corrected. After the center point correction is completed, other uncorrected data points are continued to be used as objects, and based on the iterative effect, the density values corresponding to the data points are ultimately made more stable. In response to the high degree of dispersion of data points within the gas turbine engine class and the strong distribution of data points between classes, this study introduces the contribution degree to divide different sample data based on the same class. Considering the problem of poor classification accuracy due to membership degree, there may be misclassification that affects the center position. Therefore, this study introduces data weight to evaluate the contribution of data of the same type [21]. After introducing data rights, the study praises or punishes the membership degree of the point by introducing a penalty factor to avoid the possibility of misclassification to other clusters. To amplify the contribution of individual data points to the classification results, a data weighted form is proposed, as shown in equation (11).

$$\begin{cases} w_i = \dfrac{\sum_{i=1}^{n} m_i}{m_i} \\ m_i = \dfrac{1}{n}\sum_{i=1}^{n} d_{ij} \end{cases} \tag{11}$$

In equation (11), m_i represents the distance between the class center point and a single point. The objective function and corresponding cluster center points are shown in equation (12).

$$\begin{cases} J_m = \sum_{i=1}^{N}\sum_{j=1}^{c} w_i u_{ij}^m d_{ij}^2 \\ v_i = \dfrac{\sum_{i=1}^{N} w(i) u_{ij}^m x_i}{\sum_{i=1}^{N} w(i) u_{ij}^m} \end{cases} \tag{12}$$

In equation (12), during the clustering process, noise points and irrelevant points are generally in a discrete state from the clustering center point. After introducing data weights, the magnitude of the change is small, and extreme changes may occur during the corresponding correction process, manifested as an increase in membership degree for those belonging to this category and a decrease in membership degree for those not belonging to this category. Therefore, in this study, a penalty factor is introduced and set to avoid the above situation, as shown in equation (13).

$$| u_{ij} - \frac{1}{c}\sum_{j=1}^{c} u_{ij} | \geq \sigma \quad or \quad u_{ip} \geq 0.5 \tag{13}$$

In equation (13), u_{ip} represents the maximum membership value. When it is found through comparison that the numerical significance of a certain data point is higher than the membership values of other data points, or the numerical value of that data point is greater than 0.5, corresponding operations are performed according to the penalty factor, as shown in equation (14).

$$a = u_{ip}(1 - d_{ip} / \sum_{j=1}^{c} d_{ij}) \tag{14}$$

In equation (14), a represents the penalty factor, and equation (15) is used to make relevant corrections to the membership degree.

$$\begin{cases} u_{ip} = 1 - a \sum_{j=1, j\neq p}^{c} u_{ij} = 1 - a - a u_{ip} \\ u_{ij} = a u_{ij}, i \neq p \end{cases} \tag{14}$$

In equation (14), a represents the penalty factor, and equation (15) is used to make relevant corrections to the membership degree.

$$\begin{cases} u_{ip} = 1 - a \sum_{j=1, j\neq p}^{c} u_{ij} = 1 - a - a u_{ip} \\ u_{ij} = a u_{ij}, i \neq p \end{cases} \tag{15}$$

After introducing a penalty factor into equation (15), the optimized algorithm mentioned above is applied to the fault diagnosis of gas turbine engines. After collecting the raw data, pre-processing operations are performed using the LPP algorithm to map the high-dimensional data set to a low dimensional space through dimensionality reduction, making it easier to extract the relevant features of each data point. Due to the possibility of human factors affecting the research results, subtraction clustering is used to represent the original center point position, and then the data point features extracted by the LPP algorithm are segmented based on an improved fuzzy algorithm. Finally, in this study, the penalty factor transformation matrix is used to calculate the corresponding data weights. Data belonging to the same class can be iteratively controlled within a more appropriate range, and corresponding gas turbine engine failure models can be established based on all data points. The algorithm flow is shown in Fig. 4.

In Fig. 4, first, based on equation (9), the r_a of each data point is obtained to determine the interval based on density data. Then, the density value corresponding to each data point is calculated, and whether the data point meets the corresponding stopping condition is determined. If it meets the condition, the corresponding initial cluster center point and number are obtained. If it does not meet the condition, the new neighborhood radius is adjusted and the new density index is recalculated. By combining the cluster center points with equation (6), the corresponding membership matrix is calculated, and then the data weight is calculated using equation (11). The objective function value is calculated using equation (12) and compared with the objective function value generated in the previous iteration. If it is less than σ, stop. If it is greater than σ, the penalty factor a is obtained

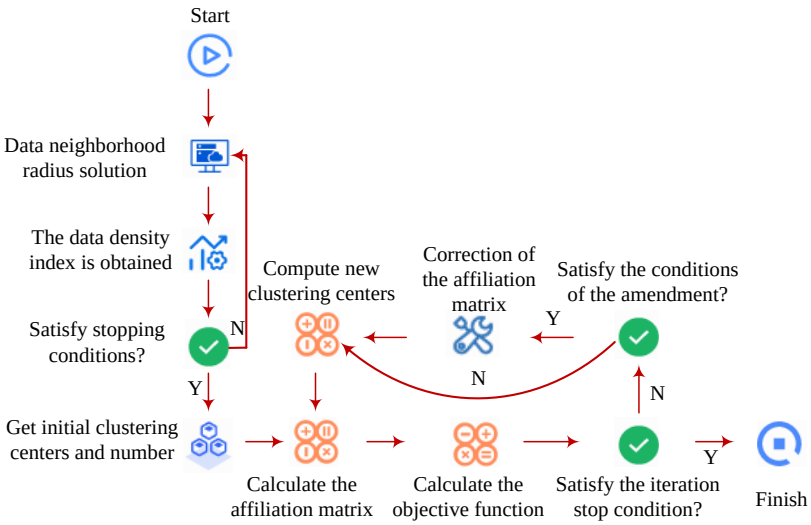


Fig. 4. Flowchart of fault diagnosis algorithm

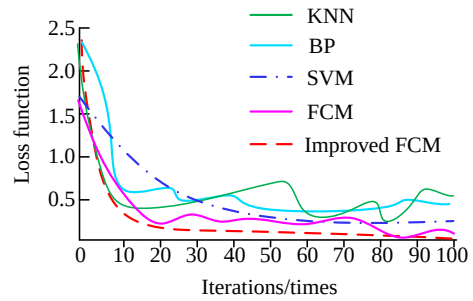
based on equation (14). Whether the correction conditions are met is compared. If the correction conditions are met, the cluster center points will be recalculated and iterated according to equation (15) after correction; If not satisfied, the cluster center will be recalculated directly. After the above operations, an improved FCM based gas turbine fault diagnosis model based on clustering centers can be obtained. The fault diagnosis process can be completed by inputting the corresponding raw data values during the detection process.

3. RESULTS

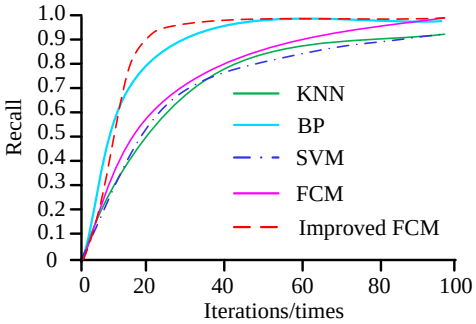
3.1. Performance analysis based on improved FCM

This study utilized a degradation simulation dataset of turbofan engines for relevant experimental verification. All experimental environments were as follows: the operating system was window10, the processor was Intel (R) Core (TM) i5-9300H, and the GPU was NVIDIA Quadro M5000. Meanwhile, four other classic algorithms were introduced for comparison, including KNN algorithm (with 4 neighbors), Backpropagation neural network (BP), Support Vector Machine (SVM), and FCM. The loss function curves and recall rates of the five algorithm models were shown in Fig. 5.

In Fig. 5 (a), the improved FCM loss function showed a rapid decline rate and had been reduced to an extremely low level. Compared to other algorithms, although they also showed a downward trend, the loss function value did not converge to a very small value. The KNN, BP, and FCM models showed significant oscillations, while the SVM algorithm model did not exhibit oscillations, but the final convergence value of the improved FCM loss function was larger. In addition, the loss on the improved FCM dataset was significantly lower than the other four algorithms, indicating that the improved FCM can fit the training data well and effectively suppress overfitting. In Fig. 5 (b), the improved FCM has converged at 25 iterations, and the recall rate was close to 1. The final recall rates of the two FCM and BP algorithm models remained basically the same, while the recall rates of the other two algorithms converged to 0.9. The accuracy of the 5 algorithm models was shown in Fig. 6.



(a) Curves of the loss functions of the five algorithm models



(b) Recall Comparison Chart

Fig. 5. Comparison of loss function and recall rate of 5 algorithms

In Fig. 6, the improved FCM had smaller intra-group errors, while the other groups, especially BP and FCM, had larger intra-group errors. The improved FCM had the highest level of accuracy value, with a value of 95.67%. The average accuracy values of the other four algorithms were 77.85%, 82.15%, 91.26%, and 84.26%, respectively, with an

average accuracy improvement of 11.79% compared to the previous methods.

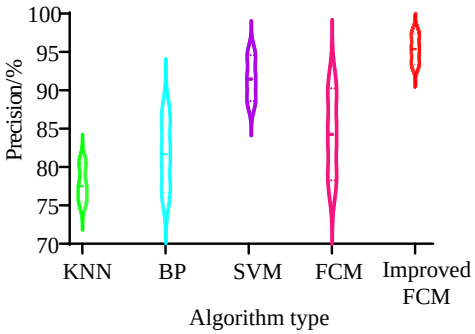


Fig. 6. Violin plot for comparison of accuracy of five algorithms

3.2. Fault diagnosis analysis of gas turbine engine based on improved FCM

Based on the above dataset, four types of engine states were selected, including Normal State (NS), Adjustable Blended Valve (ABV), Import Total Temperature Indication (ITI), and Exhaust Gas Temperature Indication (EGTI). The training set corresponded to 100, 25, 25, and 25 states, respectively. The confusion matrices of the 5 algorithm models for the training set are shown in Fig. 7.

In Fig. 7, based on the five algorithms under normal conditions, the accuracy rates were 75%, 85%, 80%, 94%, and 99%, respectively. The improved FCM had an average improvement of 15.5% in engine accuracy under normal conditions. Meanwhile, the improved FCM had the highest level of accuracy for three types of faults, with an average improvement of 25%, 24%, and 16%. The experimental results confirmed that the improved FCM achieved the highest accuracy in all categories. 100 sets of data were selected for the normal state and three types of fault states in the test set. Based on clustering theory and five algorithms, the dimensionality was reduced to two dimensions after normalization. The results of KNN, BP, and SVM dimensionality reduction were shown in Fig. 8.

In Fig. 8 (a), the dimensionality reduction effect was the worst, and the distribution of the four engine states was relatively discrete. In Fig. 8 (b), the SVM algorithm performed well in clustering ITI fault types, but performed poorly on other engine fault types. In Fig. 8 (c), the BP algorithm generally performed well in reducing the dimensionality and clustering of EGTI fault states, but its clustering performance for other engine states was not ideal. Overall, the clustering performance of the three algorithms for the four engine states was poor. FCM was further compared with improved FCM, and the clustering effect is shown in Fig. 9.

In Fig. 9 (a), FCM only had good clustering performance for EGTI and normal state types, and the clustering performance for the other two types was not ideal. In Fig. 9 (b), the improved FCM performed well in clustering all engine states. The results indicated that introducing subtractive clustering, penalty factors, and data weights based on FCM could improve the clustering ability of the original algorithm. The cross validation experimental results of 5 algorithms are shown in Fig. 10.

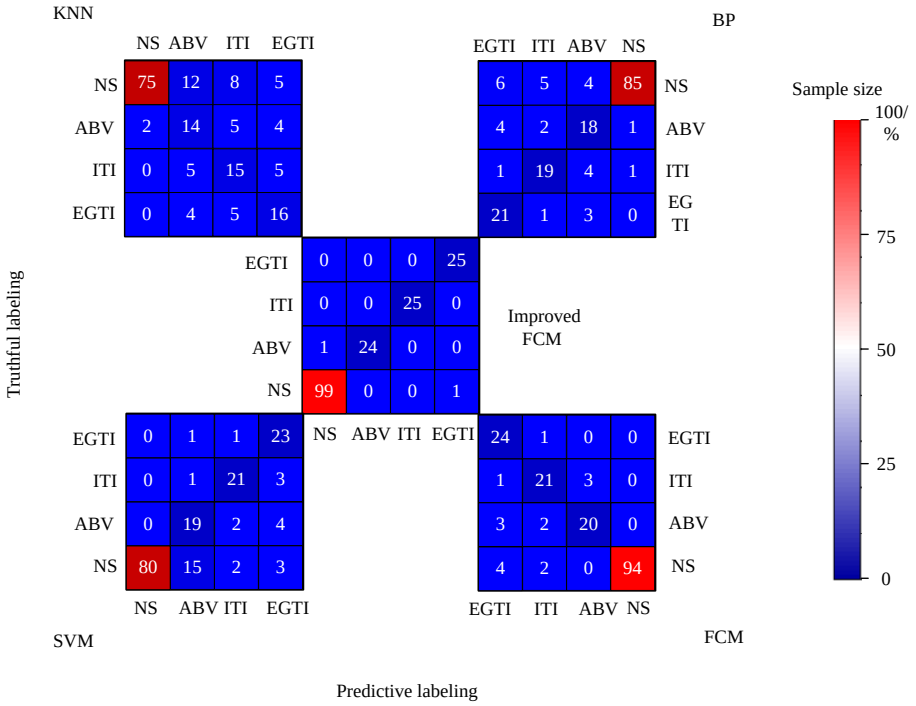
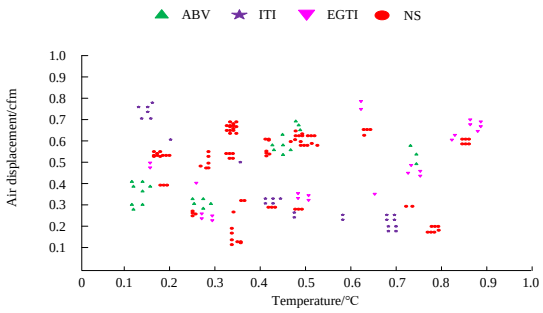
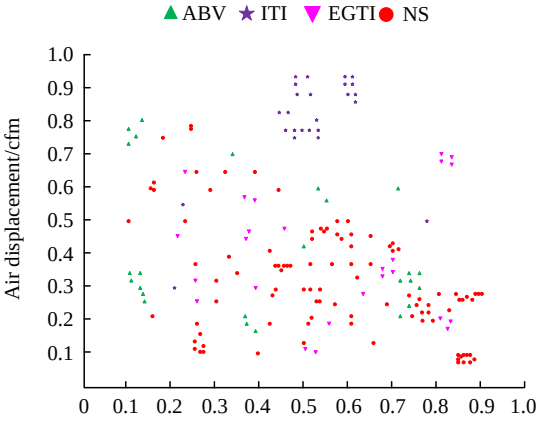


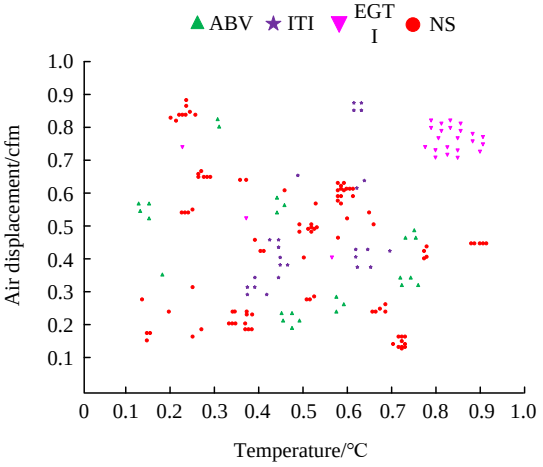
Fig. 7. Schematic diagram of the test centralized mixing matrix



(a) KNN clustering effect diagram



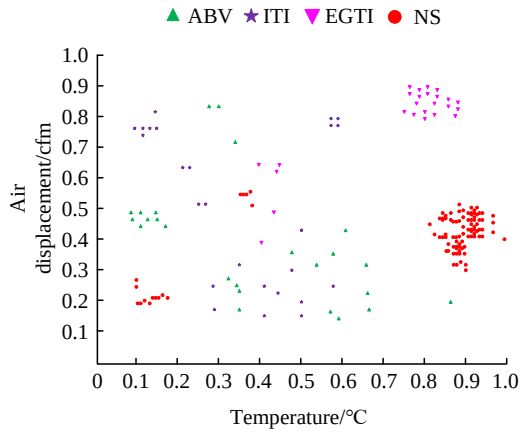
(b) SVM clustering effect diagram



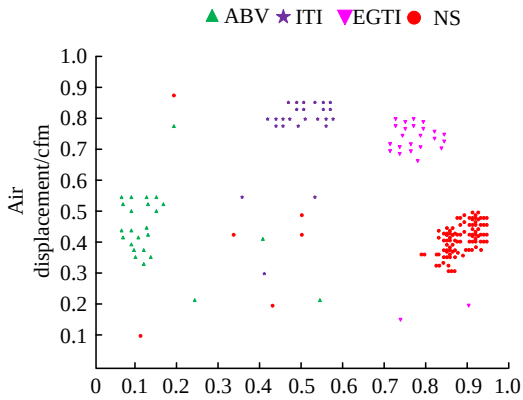
(c) BP clustering effect diagram

Fig. 8. Schematic diagram of dimensionality reduction by KNN, BP and SVM algorithms

In Fig. 10 (a), in the cross validation comparison chart, the improved FCM maintained the highest level of accuracy in each validation experiment, with average accuracies of 66.52%, 71.68%, 84.58%, 74.26%, and 90.39% for the five algorithms, respectively. The improved FCM had an average accuracy improvement of 16.13% compared to the other four algorithms. In Fig. 10 (b), the improved FCM had a smaller standard deviation compared to the other groups, indicating a lower degree of traditional FCM, this study introduced subtraction dispersion within the group. Compared with the



(a) FCM clustering effect diagram



(b) Improved FCM clustering effect diagram

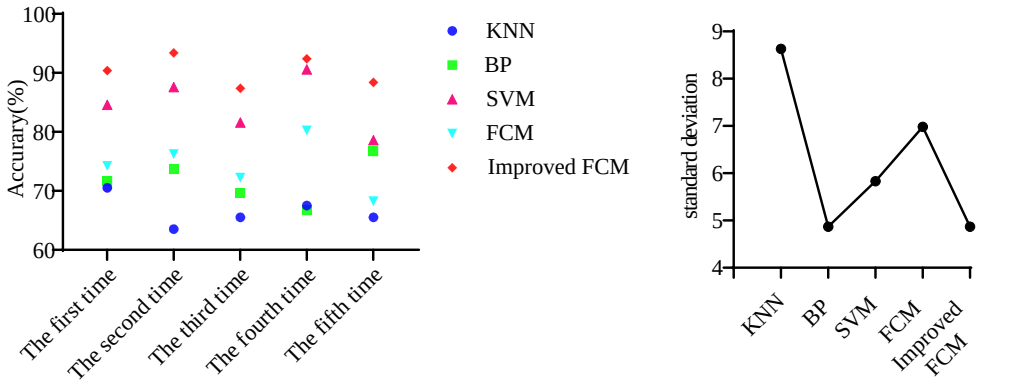
Fig. 9. Schematic diagram of dimension reduction of FCM and improved FCM

clustering, data weighting, and penalty factor optimization FCM to improve accuracy and reduce intra-group bias. The credibility of classification based on membership degree is shown in Fig. 11.

In Fig. 11 (a), in the ABV state, the membership degree of the KNN algorithm for classifying fault data was only 0.26, indicating that the credibility of the KNN algorithm for ABV fault data was not high. The membership values of BP, SVM, and improved FCM were all 1, indicating that the three algorithms had extremely high reliability in classifying ABV fault data. In Fig. 11 (b), the FCM optimized under ITI state had the highest membership degree, with a value of 1. In contrast, the membership value of the improved FCM increased, which meant the reliability value increased. Further research on EGTI and the membership results under normal conditions is shown in Fig. 12.

In Fig. 12 (a), under the EGTI state, the SVM algorithm had a certain regularity in classifying the membership degree of fault data. When the number of data sets was less than 50, the membership degree of SVM algorithm classification data fluctuated greatly, indicating that the credibility was high and low. When the quantity was greater than 50, the membership degree of SVM algorithm classification

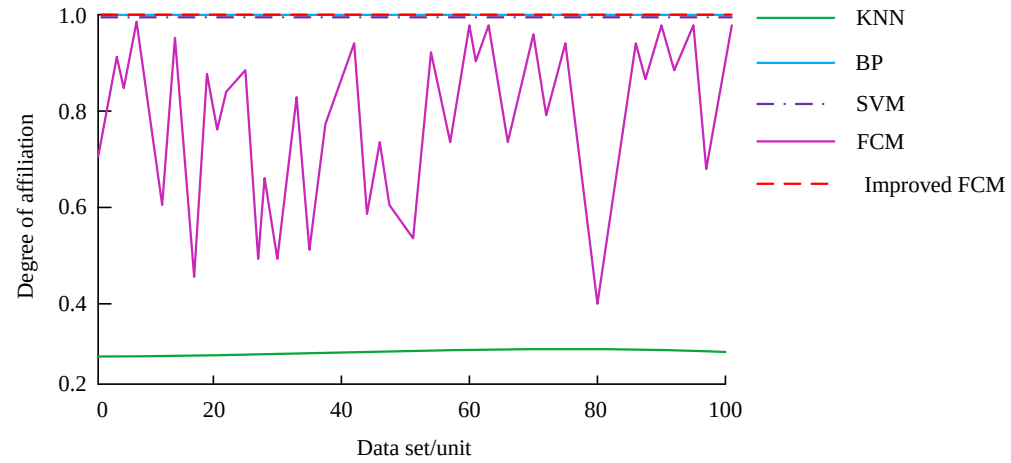
data was 1, indicating that the credibility was at a high level. Meanwhile, the membership degree values of BP and improved FCM were both 1. In Fig. 12 (b), for the data under normal conditions, the membership values of SVM and improved FCM reached 1, which was significantly higher than other algorithms. In summary, the improved FCM had the highest membership values in both normal engine conditions and three fault states, indicating a high level of reliability in classifying fault data. Finally, the study selected a typical gas turbine engine fault type and compared the reconstruction accuracy as well as error comparison of the above five algorithms for gas turbine engine gas path faults in practical applications. The results are shown in Table 1.



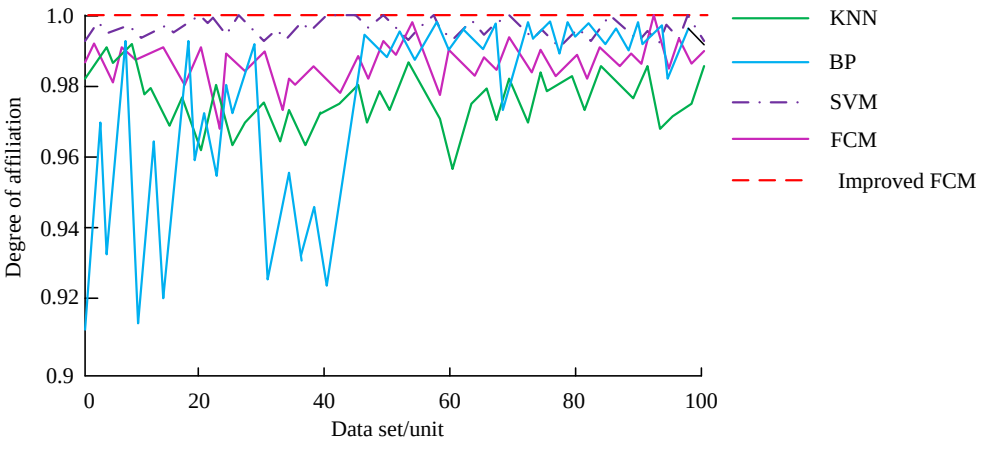
(a) Plot of the results of the cross-validation experiments of the five algorithms

(b) Comparison of the standard deviation of the five algorithms

Fig. 10. Cross-experimental validation comparison of 5 algorithms

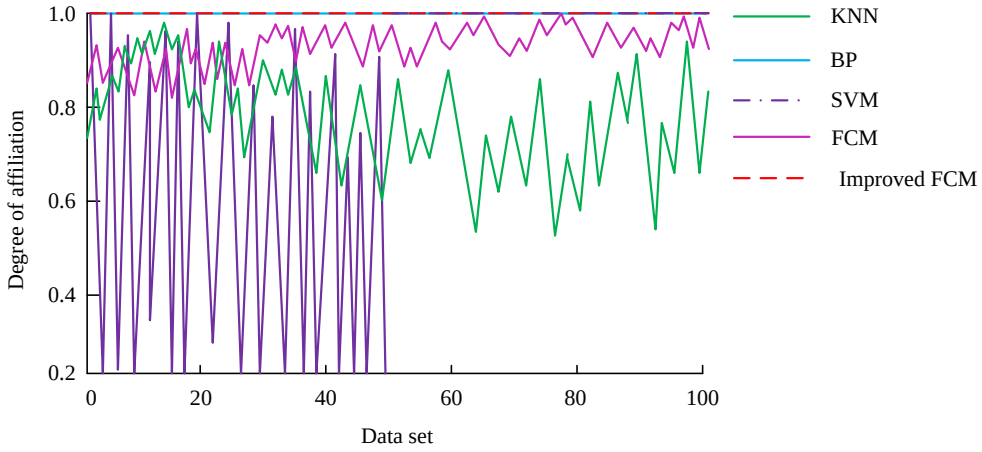


(a)Categorical affiliation results plot for ABV data

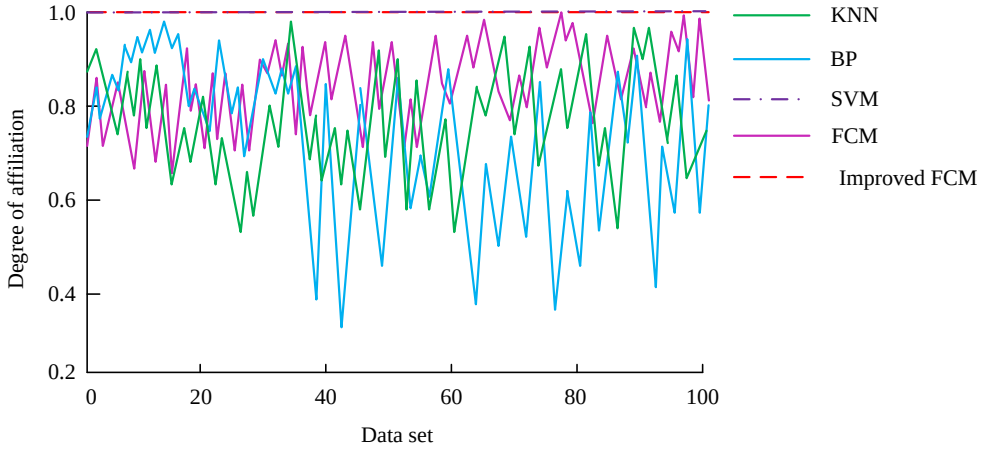


(b)Categorical affiliation results plot for ITI data

Fig. 11. Classification results of affiliation for ABV and ITI classification



(a) Categorical affiliation results plot for EGTI data



(b) Categorical affiliation results plot for NS data

Fig. 12. Affiliation classification results for EGTI and NS classification

Table 1. Comparison of accuracy and error of gas turbine engine gas circuit fault reconstruction

Algorithm type	Reconstruction accuracy (%)	MAE
KNN	80.65	4.99
BP	78.45	5.96
SVM	86.54	5.28
FCM	91.54	2.68
Improved FCM	98.47	1.09

Table 1 shows the comparison of the reconstruction accuracy as well as the error value of each algorithm for gas turbine engine gas circuit faults. The reconstruction accuracy value of the improved FCM algorithm was the largest, with a value of 98.47% and an average improvement of 14.18% over the reconstruction accuracy of the remaining four algorithms. At the same time, the improved FCM algorithm had the smallest Mean Absolute Error (MAE) value. The results showed superior performance in the diagnosis of typical fault types in gas turbines under real operating conditions. The study further compared the diagnostic accuracy of the above algorithms for air circuit faults, bearing faults, and starting faults. The results are shown in Table 2.

Table 2. Comparison of accuracy for three fault types

Algorithm type	Air line failure	Bearing failure	Starting fault
KNN	70.69%	71.54%	69.87%
BP	75.78%	78.58%	74.58%
SVM	84.58%	81.45%	83.59%
FCM	89.55%	88.59%	87.69%
Improved FCM	98.47%	94.58%	96.58%

Table 2 compares the diagnostic accuracy of the algorithms for the three fault types. The results showed that the improved FCM had the highest diagnostic accuracy for the three types of faults, at 98.47%, 94.58%, and 96.58%, respectively. Compared with the other four algorithms, the accuracy has increased by an average of 18.32%, 14.54%, and 17.64%. This indicated that in practical applications, the improved FCM had a higher diagnostic accuracy for typical gas turbine faults.

4. DISCUSSION

In the experimental results, the improved FCM achieved an accuracy of 90.39%, an average improvement of 16.13% compared to the other four algorithms. The recall rate results indicated that the improved FCM converged at 25 iterations, with a recall rate close to 1. The final recall rates of the two FCM and BP algorithm models remained basically the same, while the recall rates of the other two algorithms converged to 0.9. The accuracy value results showed that the improved FCM had the highest level of accuracy value, with a value of 95.67%, which was 11.79% higher than the average accuracy of other algorithms. Meanwhile, the improved FCM improved the accuracy values of KNN, BP, SVM, and FCM algorithms by 19.65%, 12.26%, 3.55% and 11.70% respectively. The improved FCM accuracy under four engine states was at the highest level, with an average improvement of 15.5%, 25%, 24%, and 16% in accuracy. Among them, the improved FCM improved the correctness of KNN, BP, SVM, and FCM algorithms by 24%, 14%, 19%, and 5% under normal type. The membership classification results indicated that the improved FCM had a membership value of 1 in all four engine states. Xiong et al. proposed a novel simulated c-means clustering algorithm based on simulated annealing and genetic algorithm for the diagnosis of bearing faults. The new algorithm solved other problems associated with fuzzy clustering algorithms while including initial cluster centroid sensitivity and convergence to local minima. In addition, the simulation results could be used as a classification criterion for identifying several bearing fault types. The results showed that the improved FCM improved the accuracy by 9.22%, and compared with the introduced model, the accuracy of fault diagnosis increased by 15.56%, which was consistent with the results of this study [11]. Cao et al. used an improved FCM for software fault location, and introduced four methods including Tarantula to evaluate the effectiveness of the proposed FCM. The results showed that the improved FCM had a higher level of accuracy compared to other algorithms. In terms of accuracy test case recognition, the improved FCM had lower false positive and false negative rates, which was consistent with the credibility results of fault data classification based on membership evaluation in this study [22]. Arneja T's research summarized the types and factors of engine failure, and reduced the risk of gas turbine engine through effective fault analysis methods [23]. In terms of engine fault diagnosis types, the research focused on the overall fault diagnosis of gas turbine engines, including vibration fault, gas path fault, and starting fault, which had a wider application range. Meanwhile, the diversity of fault types involves mechanical and thermal aspects, etc., while relevant literature are limited to bearing fault and inverter fault types, and their application range has certain limitations. The possible reason is that the improved FCM algorithm introduces optimization measures such as subtraction clustering, which leads to differences in fault type diagnosis of gas turbine engines. In summary, the improved FCM algorithm proposed in the research shows significant advantages in the fault diagnosis of gas turbine engines.

5. CONCLUSION

In this study, the FCM is first used as the basis for fault diagnosis of gas turbine engines. Subtractive clustering, data weighting, and penalty factors are introduced to optimize it. After comparing the performance with classic KNN, BP, SVM, and basic FCM, it is shown that the proposed improved FCM algorithm exhibits significant superiority in accuracy, precision, recall, and other aspects. Secondly, the results based on the confusion matrix indicate that the improved FCM exhibits the highest level of accuracy among the four types of engine states. Combining the results of membership degree correlation, it indicates that the improved FCM has a high level of accuracy in determining engine fault types, and it has a high level of credibility. The research offers the potential for more precise fault diagnosis in engine reliability, as well as reduced maintenance costs and an extended engine lifespan. Furthermore, the enhanced FCM algorithms have the potential for cross-domain application in fields such as power systems. Although the improved FCM proposed in the study has shown excellent superiority in gas turbine engine fault diagnosis, the faults that occur during actual turbine engine operation may be of multiple types coexisting. The simultaneous occurrence of multiple faults may make feature extraction of relevant data points difficult, and the dimensionality reduction effect may also be affected, leading to deviations in the subsequent diagnostic process. Future research can further improve the fault diagnosis depth of the model and incorporate some uncommon fault operation data into the research scope, thereby deepening the generalization ability of the model. At the same time, more attention can be paid to the diagnosis of multiple failure modes, that is, diagnostic strategies in which there are multiple faults in the same engine. In addition, cooperation with experts in thermodynamics as well as mechanical engineering on the failure mechanisms of gas turbine engines could be aimed at further optimizing the fault diagnosis algorithms.

Source of funding: *This research was conducted without any external funding.*

Author contributions: *research concept and design, Y.P.; Collection and/or assembly of data, W.H.; Data analysis and interpretation, W.H.; Writing the article, Y.P.; Critical revision of the article, P.X.; Final approval of the article, P.X.*

Declaration of competing interest: *The authors of this paper declares no conflict of interest.*

REFERENCES

1. Lee D, Kwon HJ, Choi K. Risk-based maintenance optimization of aircraft gas turbine engine component. Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability. 2024; 238(2):429-445. https://doi.org/10.1177/1748006X221135907.
2. Kennedy IR, Hodzic M, Crossan AN, Crossan N, Acharige N, Runcie JW. Estimating maximum power from wind turbines with a simple newtonian approach. Archives of Advanced Engineering Science. 2023; 1(1):38-54. https://doi.org/10.47852/bonviewAAES32021330.
3. Ghufron S, Prayogi S. Cooling system in machine operation at gas engine power plant at PT multidaya prima elektrindo. RIGGS: Journal of Artificial Intelligence and Digital Busines.s 2023; 1(2): 25-29. https://doi.org/10.31004/riggs.v1i2.21.
4. He A, Zeng Q, Zhang Y, Xie P, Li J, Gao M. A fault diagnosis analysis of afterburner failure of aeroengine based on fault tree. Processes. 2023; 11(7): 2086-2086. https://doi.org/10.3390/pr11072086.
5. Lee D, Kwon HJ, Choi K. Risk-based maintenance optimization of aircraft gas turbine engine component. Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability 2024; 238(2):429-445. https://doi.org/10.1177/1748006X221135907.
6. George B, Muthuveerappan N. Life assessment of a high temperature probe designed for performance evaluation and health monitoring of an aero gas turbine engine. International Journal of Turbo & Jet-Engines. 2023;40(2):139-146. https://doi.org/10.1515/tjj-2020-0037.
7. Anggrawan A, Mayadi M. Application of KNN machine learning and fuzzy C-means to diagnose diabetes. MATRIK: Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer 2023; 22(2): 405-418. https://doi.org/10.30812/matrik.v22i2.2777.
8. Kodipalli A, Fernandes SL, Dasar SK, Ismail T. Computational framework of inverted fuzzy C-means and quantum convolutional neural network towards accurate detection of ovarian tumors. International Journal of E-Health and Medical Communications (IJEHMC). 2023;14(1):1-16. https://doi.org/10.4018/IJEHMC.321149.
9. Sarwar U, Muhammad M, Mokhtar AA, Khan R. Hybrid intelligence for enhanced fault detection and diagnosis for industrial gas turbine engine. Results in Engineering. 2024; 21(2): 101841-101841. https://doi.org/10.1016/j.rineng.2024.101841.
10. Kordestani M, Mousavi M, Chaibakhsh A. A new compressor failure prognostic method using nonlinear observers and a Bayesian algorithm for heavy-duty gas turbines. IEEE Sensors Journal. 2023; 23(4): 3889-3900. https://doi.org/10.1109/JSEN.2022.3233585.
11. Xiong J, Liu X, Zhu X, Zhu H, Li H, Zhang Q. Semi-supervised fuzzy c-means clustering optimized by simulated annealing and genetic algorithm for fault diagnosis of bearings. IEEE Access. 2020; 8(1): 181976-181987. https://doi.org/10.1109/ACCESS.2020.3021720.
12. Cheng K, Wang Y, Yang X, Zhang K, Liu F. An intelligent online fault diagnosis system for gas turbine sensors based on unsupervised learning method LOF and KELM. Sensors and Actuators A: Physical. 2024; 365(2):114872-114872. https://doi.org/10.1016/j.sna.2023.114872.
13. Feng K, Xiao Y, Li Z. Gas turbine blade fracturing fault diagnosis based on broadband casing vibration. Measurement. 2023; 214(5): 112718-112729. https://doi.org/10.1016/j.measurement.2023.112718.
14. Fahmi ATWK, Kashyzadeh KR, Ghorbani S. Fault detection in the gas turbine of the Kirkuk power plant: An anomaly detection approach using DLSTM-Autoencoder. Engineering Failure Analysis. 2024; 160(5):108213-108229. https://doi.org/10.1016/j.engfailanal.2024.108213.
15. Eskandari MA, Karimi H, Sarvari A, Naderi M. Turbine inlet temperature effects on the start process of an expansion cycle liquid propellant rocket engine. Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering. 2023;237(1):42-61. https://doi.org/10.1177/09544100221090797.
16. Dev S, Lafrance S, Liko B, Guo HS. A study on effect of engine operating parameters on NOx emissions and exhaust temperatures of a heavy-duty diesel engine during idling. International Journal of Engine Research. 2023;24(3):982-998. https://doi.org/10.1177/14680874221076087.
17. Yi L, Zhu J, Wang Y, Liu J, Wang S. Short-term power load forecasting based on orthogonal PCA-LPP dimension reduction and IGWO-BiLSTM. Recent Patents on Mechanical Engineering. 2023; 16(1): 72-86. https://doi.org/10.2174/2212797615666221012091902.
18. Su S, Zhu G, Zhu Y, Ge B, Liang X. Coupled locality discriminant analysis with globality preserving for dimensionality reduction. Applied Intelligence. 2023; 53(6): 7118-7131. https://doi.org/10.1007/s10489-022-03409-3.
19. Zhang M, Parnell A. Review of clustering methods for functional data. ACM Transactions on Knowledge Discovery from Data. 2023; 17(7): 1-34. https://doi.org/10.1145/3581789.
20. Kusumadewi S, Rosita L, Wahyuni EG. Performance of fuzzy C-means (FCM) and fuzzy subtractive clustering (FSC) on medical data imputation. ComTech: Computer, Mathematics and Engineering Applications. 2024;15(1):29-40. https://doi.org/10.21512/comtech.v15i1.11002.
21. Shahrabi Farahani A, Mohammadi E, Alizadeh M. Utilizing artificial intelligence to develop an advanced compressor airfoil family for industrial, aero-derivative, and heavy-duty gas turbines. Proceedings of the Institution of Mechanical Engineers, Part A: Journal of Power and Energy. 2023; 237(6): 1170-1187. https://doi.org/10.1177/09576509231163350.
22. Cao H, Li L, Chu Y, Deng M, Wang P, Zhao C. A coincidental correctness test case identification framework with fuzzy C-means clustering. Multimedia Systems. 2023; 29(3): 1089-1101. https://doi.org/10.1007/s00530-022-01039-w.
23. Arneja T. Mechanical and thermal factors contributing to turbine engine failures. International Journal of I.C. Engines and Gas Turbines. 2024; 10(01): 7-12. https://doi.org/10.37591/IJICEGT.

Yana PENG obtained her Master's degree in Mechatronic Engineering from the University of Electronic Science and Technology of Chinain2004. Presently, she is working as s an associate professor at the Aviation Maintenance Engineering School of Chengdu Aeronautic Polytechnic. Her area of interest are information processing and automatic control research.
e-mail: yanapeng@126.com

Wenjiong HAI obtained her Master's degree in Mechanical Design and Theory from Nanjing University of Aeronautics and Astronauticsin2003. Presently, she is working as s an associate professor at the Aviation Maintenance Engineering School of Chengdu Aeronautic Polytechnic. Her area of interest are aircraft engine assembly and maintenance techniques research.
e-mail: hai_skylark@163.com

Peng XUE obtained his BE in computer science and technology from Civil Aviation University of China in 2006. He obtained his ME in aerospace promotion theory and engineering from Civil Aviation University of China in 2012. AT presently, he is working as a lecturer at the Engineering Technology Training Center of Civil Aviation University of China. His areas of interest is the fault diagnosis of aircraft engines.
e-mail: pxue@cauc.edu.cn