ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 2 - 2026

Research on aero-engine fault diagnosis by counting and damage repair and optimizing LSSVM with IGWO algorithm

Manshan Li[1,*] *and Hongping Wei*[1]

[1] CSN Digital Intelligence Technology Co., Ltd., Guangzhou, Guangdong, 510000, China

SUMMARY: *In this paper, for the complex fault diagnosis problem of aero-engine with high dimensional nonlinearity, an intelligent fault diagnosis method of LSSVM based on damage repair and modified gray wolf optimization is proposed to characterize the performance deterioration trend with a damage accumulation prediction model under the interaction of multiple parameters, and to obtain the main health indicator parameters by damage repair method. The amount of model input information is increased. The improved Gray Wolf algorithm utilizes chaotic sequences to increase population diversity and uses adaptive inertia weights to achieve a good balance between global search capability and local exploitation capability to effectively optimize LSSVM parameters. The method proposed in this paper accurately diagnoses multiple types of engine faults, reduces the model learning time and accelerates the convergence process, providing an effective and feasible technical means for aircraft engine condition monitoring and fault diagnosis.*

KEYWORDS: *aero-engine; fault diagnosis; damage repair; gray wolf optimization algorithm; least squares support vector machine*

1 Introduction

Aircraft engine is an important embodiment of a country's scientific and technological, industrial and national defense strength. Engines operate in extremely harsh environments for a long time, such as high temperature and pressure, high corrosion, high density, etc., which is a failure-prone system, and once the engine failure occurs during the flight of an airplane, it will lead to serious civil aviation safety accidents, which can be extremely harmful [1, 2]. Therefore, it is necessary to carry out fault diagnosis of the engine to detect and locate the faults in time to ensure the reliability and safety of the aircraft [3]. Since the support vector machine (SVM) has a complete statistical learning theoretical foundation and excellent learning performance, according to the structural risk minimization rule, the generalization ability of the learning machine can be maximized, which is very suitable for the predictive analysis problem under small samples such as aviation engines [4-6]. However, the superior performance of SVM needs to select appropriate parameter values to be reflected, and the good or bad values of the parameters greatly affect the accuracy and efficiency of fault diagnosis. Least Squares Support Vector Machine (LSSVM), due to its superiority in small samples and high-dimensional nonlinear pattern recognition, has been applied in engine fault diagnosis to effectively solve the above problems [7, 8]. However, in practical problems, because it is difficult to know the range of the optimal parameters beforehand, this method is not only difficult to find the optimal parameters accurately, but also less efficient by using traversal

*BDHI1234@163.com
https://doi.org/10.65102/is2026721 Published: 30 April, 2026

selection, so it is necessary to optimize the LSSVM in order to improve the diagnostic accuracy.

Gray wolf optimization algorithm (GWO) is a new type of swarm intelligence optimization algorithm inspired by the predatory behavior of gray wolf groups, in order to improve the global search ability of GWO, avoid premature convergence of the algorithm, and improve the accuracy of the original algorithm, the improved version of the gray wolf optimization algorithm (IGWO) has a faster speed and higher accuracy of the search for the optimum in optimizing LSSVM [9-11]. The nature of the engine use process is a “damage-degradation-repair” dynamic cycle, the current diagnostic methods are mainly focused on the judgment between the health of the engine and the failure, but generally ignored the planned maintenance of the phased restoration of the effect of the Ming, which may result in misjudgment or performance degradation [12]. Therefore, constructing an IGWO-LSSVM intelligent diagnostic model that can take into account the dynamic process of damage repair is the key to realizing engine fault diagnosis.

For the application of SVM and its related methods in aero-engine fault diagnosis, literature [13] proposed an adaptive directed SVM model, and analyzed its application in aero-engine turbine blade fault diagnosis by combining adaptive sampling techniques and introducing penalty coefficients, pointing out that the method significantly improves the accuracy, efficiency and stability of reliability analysis. Literature [14] explored the application of SVM in turboshaft engine fault diagnosis, and for the problem of scarcity of fault data, proposed the OC-SVM-TL algorithm through the introduction of migration learning and single-class SVM, examined its effectiveness in dealing with the differences of data in different engine states and operating conditions, and pointed out that the algorithm can still maintain superior detection performance when there is insufficient data in the target domain. Literature [15] applies SVM technology to improve the accuracy and efficiency of aero-engine fault diagnosis, and analyzes its advantages in dealing with high-dimensional nonlinear data by comparing different kernel functions and optimizing the parameters, and points out that this method significantly improves the fault identification accuracy and diagnosis speed compared with the traditional way. Literature [16] combines the improved local discriminant basis (LDB) algorithm with SVM for aero-engine fault diagnosis, analyzes the effective identification of faults such as friction impacts by SVM classifiers by using the normalized energy difference and relative entropy to select the optimal feature subspaces, and emphasizes the advantages of this method in the application of state differentiation. Literature [17] proposed the WCFSE-FSVM model for aero-engine vibration fault diagnosis by combining the wavelet coefficient feature Shannon entropy (WCFSE) method with the fuzzy FSVM, analyzed its advantages in feature extraction and SVM classification, and pointed out that the model has a stronger ability of learning, generalization, and anti-noise than other methods. Literature [18] proposed an aero-engine control system fault diagnosis method by fusing probabilistic neural network and SVM, optimized the key parameters by using particle swarm algorithm, verified that the optimized SVM model reached a high diagnostic accuracy of 98.8%, and emphasized the effective role of the method in improving the diagnostic reliability. Literature [19] developed a fault diagnosis method for aero-engine fuel regulator by establishing the regulator model and the engine inverse model, and utilizing the deviation of its output from the sensor data to isolate the faults, and at the same time, introduced the machine learning technique of correlation vector machine to construct the accurate nonlinear inverse model, and finally verified the effectiveness and application prospect of the method through hardware-in-the-loop experiments. Literature [20] proposed a fault diagnosis method based on multitasking graph SVM reasoning, which integrates the personalized features of the engine by constructing a graph structure and analyzes its adaptability to individual differences, pointing out that the method can significantly improve the diagnostic accuracy and reduce the rate of misdiagnosis, and emphasizing the key role of incorporating the individual differences of the engine into the model to improve the diagnostic reliability.

As for LSSVM and related methods for aero-engine fault diagnosis has also received extensive attention from the academic community, literature [21] developed an online fault diagnosis system for turbofan engine turbine path sensors, and analyzed its advantages in improving the generalization performance and reducing the model redundancy by combining with recursive simplified LSSVM regression algorithm optimized by genetic algorithm, and verified the system's ability to single, dual sensor typical faults, and verified the effective diagnostic capability of the system for single and dual sensors. Literature [22] reduces the computational complexity by developing an extended LSSVM, investigates its pairwise and primitive space solution algorithms, analyzes its effectiveness in regression and classification, and points out that the model has a good potential for application in aero-engine fault diagnosis. Literature [23] proposed a LSSVM model using cross-validation optimization for chiller fault detection and diagnosis, analyzed its performance in dealing with nonlinear and system-level faults, pointed out that the model performs better than other methods in terms of overall correctness, detection efficiency, and emphasized its potential application in fault diagnosis of aero-engine similarly complex systems. Literature [24] proposed the LSSVM-CIL method for the category imbalance problem in aero-engine fault diagnosis, and analyzed its improvement in optimizing the classification boundary and enhancing the computational stability by introducing a differentiated regularization parameter and incorporating a recursive simplification strategy, and verified its effectiveness as a candidate technique for real-time fault detection. Literature [25] proposed a LSSVM model optimized based on KPCA and Sparrow algorithm for fault detection of flight control system sensors, which solved the blindness of traditional LSSVM parameter selection through intelligent parameter search, and verified the diagnostic accuracy of the model as high as 98.23%, which is significantly better than other comparative algorithms. Literature [26] for the nonlinear aero-engine distributed control system with multi-packet transmission, time delay and packet loss, proposed a sliding window multicore LSSVM based online packet loss prediction and compensation method, and combined with the neural network PID sliding mode controller for the design of the method, verified through simulation, the method in 30% and 60% of the packet loss rate to significantly improve the prediction accuracy and the performance of the control response. Literature [27] proposed a real-time fault diagnosis method based on genetic algorithm optimization of LSSVM regression for the health monitoring of aviation engines, analyzed the effectiveness of the algorithm in using actual experimental data through data standardization and parameter optimization, and emphasized its high detection accuracy and engineering application value. Literature [28] for the demand of aero-engine thrust high-precision estimation, put forward a gravitational search algorithm optimization based on the LSSVM model, through the comparison with the particle swarm algorithm, it is pointed out that the optimization scheme has a better parameter optimization ability and generalization performance, and can effectively meet the requirements of direct thrust control.

Based on the actual needs of aero-engine fault diagnosis, this paper proposes a LSSVM model that combines damage repair strategy and IGWO, and uses it in aero-engine fault diagnosis. The model utilizes the damage repair strategy to solve the previous problem of analyzing the damage process as a single direction. It also establishes a comprehensive judgment index including repair records, so that the failure judgment has the ability to consider the condition of parts in a comprehensive way.

2 Key technologies and algorithms

2.1 Aircraft Engine Fault Diagnosis Techniques

The gradual evolution of the traditional aero-engine fault diagnosis method based on the experience of maintenance engineers into the current new fault diagnosis method based on data analysis and artificial intelligence algorithms is an inevitable trend in the development of the aero-engine fault diagnosis method, and one of the manifestations of the continuous improvement of the aero-engine fault diagnosis level. The empirical-based method relies on a large amount of maintenance information and uses an IF-THEN rule base to construct a one-to-one correspondence from failure symptoms to failure modes; whereas the fault tree is a process of gradually unfolding the upper fault events into the lower fault events according to the logical relationship until a detectable base fault event occurs. CBM is to store the past failure cases, and when a new failure occurs, some rules are used to find out the most similar case with the failure from a large number of historical cases to analyze and judge, in order to determine the cause of this failure. However, with the emergence of more and more complex failure modes in modern aviation engines, the above methods have certain limitations, such as the difficulty of acquiring knowledge, the trouble of maintaining rules, and the inability to adapt to new types of failure modes, which makes it difficult for the empirical methods to satisfy the requirements of the modern aviation industry for fault diagnosis technology.

Machine learning-based methods can automatically mine and classify fault features, greatly improving the degree of artificial intelligence of the fault diagnosis system, which is represented by neural networks, which adaptively acquires fault features through layer-by-layer nonlinear mapping of fault sample inputs, and corrects the connection weight coefficients between nodes of the network by using error inverse extrapolation, and its weight iteration formula is:

$$w_{ij}(t+1) = w_{ij}(t) - \eta \frac{\partial E}{\partial w_{ij}} \tag{1}$$

where w_{ij} represents the connection weights between neurons, η denotes the learning step size, and E is the network output error function.

The radial basis function network, on the other hand, has excellent nonlinear approximation ability and good generalization properties through the hidden layer unit constructed by Gaussian kernel function.

Support vector machine (SVM) is based on Vapnik-Chervonenkis dimensional theory, which converts the nonlinear classification problem in the original feature space into a linear separation problem in the high-dimensional feature space by kernel trick, and its constrained optimization objective function is formulated as:

$$\min_{w,b,\xi} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\xi_i \tag{2}$$

The constraint is $y_i(w^T\phi(x_i)+b) \geq 1-\xi_i$. where $\phi(x_i)$ denotes the kernel mapping transformation and C is the regularization parameter. The decision tree algorithm recursively constructs classification decision rules through the information gain criterion, and the random forest effectively improves the stability and noise resistance of the diagnostic system by aggregating the prediction results of multiple decision trees through Bootstrap.

2.2 Least Squares Support Vector Machine Algorithm

Least Squares Support Vector Machine (LSSVM) is an improvement to the traditional SVM, which mainly transforms the inequality constraints in SVM into equality constraints and replaces the original loss function (hinge loss) with the least squares criterion, but it still has the good generalization ability that SVM has to deal with nonlinear classification.

What's more, the computational efficiency is significantly improved by avoiding the complex quadratic programming solution process, which makes it a very practical machine learning method in the field of fault diagnosis. The mathematical foundation of the method is built on the Lagrange multiplier theory, by constructing the objective function:

$$\min_{w,b,e} J(w,e) = \frac{1}{2} w^T w + \gamma \frac{1}{2} \sum_{k=1}^{N} e_k^2 \tag{3}$$

and impose the equation constraints:

$$y_k = w^T \varphi(x_k) + b + e_k,\ k = 1, 2, \cdots, N \tag{4}$$

The original nonlinear optimization problem is skillfully transformed into a linear system of equations, where the weight vector w, bias term b, error variable e_k and regularization parameter γ together constitute the core parameter system of the model, and the kernel function mapping $\varphi(x_k)$ plays a key role in mapping the original feature space to the high-dimensional kernel space.

By applying the Karush-Kuhn-Tucker optimality condition and constructing the corresponding Lagrangian function, the following system of linear equations can be obtained:

$$\begin{bmatrix} 0 & 1_v^T \\ 1_v & \Omega + \gamma^{-1} I \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix} \tag{5}$$

where the kernel matrix elements are:

$$\Omega_{jk} = \varphi(x_j)^T \varphi(x_k) = K(x_j, x_k) \tag{6}$$

reflects the similarity measure between the samples, and the Lagrange multiplier vector □α contains the weights of the training samples' contributions to the final decision function, unlike traditional SVMs that rely on only a small number of support vectors for prediction, the decision function of the LSSVM is:

$$f(x) = \sum_{k=1}^{N} \alpha_k K(x_k, x) + b \tag{7}$$

It utilizes the information of all the training samples, and this full-sample participation mechanism provides a richer information base for the accurate modeling of complex failure modes, although it sacrifices the sparsity of the model to a certain extent.

LSSVM shows significant technical advantages in dealing with nonlinear problems in aero-engine fault diagnosis, which is mainly due to its flexible and diverse kernel function mechanism that can effectively map the intricate fault modes in the original feature space to linearly differentiable forms in the high-dimensional kernel space, thus greatly simplifying the design complexity of the classifier. Radial basis function kernel:

$$K(x_i, x_j) = \exp\left(-\frac{\left\| x_i - x_j \right\|^2}{2\sigma^2} \right) \tag{8}$$

Effective control of the balanced relationship between model complexity and generalization ability is achieved by precise adjustment of the kernel width parameter σ. When the value of σ is small, the model tends to fit the training data finely but may face the risk of overfitting. A larger value of σ, on the other hand, helps to improve the model's generalization performance but may lead to underfitting phenomena.

The polynomial kernel is:

$$K(x_i, x_j) = (x_i^T x_j + c)^d \tag{9}$$

Through the reasonable selection of order parameter d and constant term c, it can adapt to different degrees of nonlinear mapping relations, which provides another effective way for the recognition of complex fault patterns by simulating the nonlinear transformation characteristics of the activation function of neural networks.

Although the sparsity of the LSSVM model is reduced compared with SVM, due to its faster computational speed, it is able to process a large amount of sensor information online during the actual aero-engine health assessment process, which is of some practical value for early fault detection. In addition, due to the existence of the regular term can limit the size of the weights to avoid overfitting phenomenon, so even a small number of fault samples can also obtain a better recognition effect. In addition, many parameter optimization methods such as cross-validation, grid search method, genetic algorithm, particle swarm optimization, etc. can be adopted to optimize the model performance of LSSVM under various fault detection tasks.

2.3 Gray Wolf Optimization Algorithm

The grey wolf optimization algorithm originates from the in-depth observation and mathematical abstraction of the hunting behavior of wolf packs in nature, and this bionic intelligence algorithm transforms the strict social hierarchy of wolf packs into the core mechanism of optimized search. The algorithm design cleverly utilizes the division of labor among the four ranks of wolves: α, β, δ, and ω. Among them, α wolves are responsible for decision making, β wolves are responsible for assisting coordination, δ wolves perform specific tasks to maintain order, and ω wolves follow instructions and participate in collective actions.

This hierarchical mapping mechanism enables the algorithm to designate the individual with the best fitness in the current population as the □α wolf, the second and third best individuals correspond to the β and δ wolves, respectively, and the rest of the individuals are categorized as the ω-wolf class, which achieves the asymptotic approximation of the global optimal solution through hierarchical management. The mathematical modeling of wolf hunting forms the basis of the computational framework of the algorithm, and the position update formula is used in the encircling prey phase:

$$D = \left| C \cdot X_p(t) - X(t) \right| \tag{10}$$

$$X(t+1) = X_p(t) - A \cdot D \tag{11}$$

describes the gradual approach of the individual to the position of the prey, where $X_p(t)$ denotes the current position vector of the prey, and A and C are the coefficient vectors controlling the convergence behavior, computed by the formulas, respectively:

$$A = 2a \cdot r_1 - a \tag{12}$$

$$C = 2 \cdot r_2 \tag{13}$$

where the convergence factor a decreases linearly from 2 to 0 during the iterative process, and the random vectors r_1 and r_2 both take values in the range of $[0,1]$. The attacking prey phase updates the gray wolf position by combining the guiding effects of the three optimal individuals, which is calculated as follows:

$$\vec{D}_\alpha = \left|\vec{C}_1 \cdot \vec{X}_\alpha - \vec{X}\right|, \quad \vec{D}_\beta = \left|\vec{C}_2 \cdot \vec{X}_\beta - \vec{X}\right|, \quad \vec{D}_\delta = \left|\vec{C}_3 \cdot \vec{X}_\delta - \vec{X}\right| \tag{14}$$

$$\vec{X}_1 = \vec{X}_\alpha - \vec{A}_1 \cdot \vec{D}_\alpha, \quad \vec{X}_2 = \vec{X}_\beta - \vec{A}_2 \cdot \vec{D}_\beta, \quad \vec{X}_3 = \vec{X}_\delta - \vec{A}_3 \cdot \vec{D}_\delta \tag{15}$$

$$\vec{X}(t+1) = \frac{\vec{X}_1 + \vec{X}_2 + \vec{X}_3}{3} \tag{16}$$

3 Research methodology

3.1 Introduction of damage repair mechanisms

The complexity of the operating environment of the aircraft engine determines the diversity and inevitability of its damage formation, mechanical stress cyclic loading, high-temperature thermal shock, vibration fatigue, and corrosive media erosion and other factors interacting, the engine core components gradually produce fatigue crack expansion, surface wear and tear, and geometric deformation accumulation of degradation phenomena. Damage evolution is beyond the traditional maintenance idea that damage is an irreversible process, from the beginning of the emergence of damage detection, and damage treatment, in the completion of its functional repair at the same time also for the monitoring system of the health state evaluation provides additional information on the damage repair aspects. The evolution of damage accumulation over time is modeled as in the Paris damage model:

$$\frac{da}{dN} = C(\Delta K)^m \tag{17}$$

where a denotes the crack length, N is the number of cycles, ΔK is the stress intensity factor amplitude, and C and m are material constants.

And the damage evolution model after the repair treatment follows the modified cumulative damage model:

$$D_{repair}(t) = D_{original}(t) \cdot \eta(t) + \int_{t_0}^{t} f_{repair}(\sigma, T, \phi) d\tau \tag{18}$$

where $\eta(t)$ is the repair efficiency function, f_{repair} is the new damage rate introduced by the repair process, and ϕ represents the repair process parameters. This theoretical description system lays a mathematical foundation for the effective integration of damage repair information into the fault diagnostic model, so that the diagnostic system obtains the ability to more accurately assess the health state of the engine and predict the performance evolution trend.

Currently, the commonly used repair techniques in aircraft powerplant maintenance include laser deposition, plasma spray welding, cold spraying, welding patching, cutting, etc. These repair techniques have different requirements for damage and follow different principles for repair. Comparison of the technical indicators of a variety of repair methods is shown in Table 1, among these methods, the laser cladding process is through the focused high energy density laser beam irradiation to the surface of the workpiece, and at the same time to the processed area to feed a certain amount of alloy powder material, so that the powder material and the substrate surface of the material fused together, cooling and solidification to form a new cladding layer integrated with the parts of the body; cladding layer thickness error of ± 0.1mm; The remaining life of the component after repair is generally up to 85% to 95% of the remaining life of the original component, which is mainly applied to engine fan blade front-end damage, high-pressure pressurized gas turbine blade tip damage, turbine blade thermal barrier coating spalling and oxidation damage.

Table 1: Comparison of Performance Parameters of Different Damage Repair Techniques

Repair technology	Repair accuracy (mm)	Strength recovery rate (%)	Repair efficiency	Cost level
Laser cladding	± 0.1	85-95	High	Medium
Plasma spraying	± 0.2	70-85	Medium	Lower
Cold spraying	± 0.15	80-90	Medium	Medium
Welding repair	± 0.3	90-98	Lower	Lower
Mechanical processing	± 0.05	95-100	High	Low
Composite repair	± 0.1	88-96	Medium	Higher

Plasma spraying uses plasma arc to heat the ceramic or metal powder to make it molten and then sprayed at high speed to the repaired surface, generating a coating with a bonding strength of 30 to 80 MPa, which has good application in repairing the thermal barrier coating spalling in the engine combustion chamber, the wear and tear on the surface of the pressurizer blades as well as the broken surface of the inside of the magazine. Cold spraying using supersonic airflow will be solid metal powder particles accelerated to 500 ~ 1200 m/s of high-speed movement, the use of kinetic energy to achieve the particles and the substrate plastic deformation combined with a low temperature of the substrate during the repair process to avoid the heat-affected zone, especially suitable for heat-sensitive materials repair treatment.

Welding technology has argon arc welding, electron beam welding, friction welding, etc., used to eliminate engine shaft parts cracks, magazine cracks and air leakage, pipe leakage and other structural defects, the strength of its weld can generally reach about 90% of the strength of the original substrate. Machining technology is the use of cutting, grinding, polishing and other ways to trim parts to achieve the requirements of dimensional accuracy. Although this method is simple and reliable, it changes the dimensional size of the part, and needs to be coupled with dimensional compensation methods to ensure that the repaired part can be assembled normally.

The contribution of damage repair to improve the robustness of the fault diagnosis system is manifested in the synergistic cooperation of various aspects, which can not only make up for the damaged part from the physical level, but also provide more information for the diagnosis from the information level in order to enhance the validity and reliability of the diagnostic results. The blade is a typical high-stress component, which is susceptible to fatigue damage under complex operating conditions, which eventually leads to crack generation and expansion, and its development rate can be described by the modified Parris equation:

$$\frac{da}{dN} = C(\Delta K_{eff})^m \cdot f(R, T, \phi) \tag{19}$$

where ΔK_{eff} is the effective stress intensity factor amplitude, R is the stress ratio, T is the temperature, ϕ is the environmental factor, the crack expansion resistance of the blade after laser cladding repair is significantly improved, and the threshold value of crack expansion in the repaired area ΔK_{th} is increased by 15%-25% compared with that of the original material, and this change in the material properties has a direct impact on the spectral distribution of vibration signals, and it provides an important basis for the fault diagnostic algorithm to differentiate between the repaired and unrepaired blades. Important basis.

Bearing raceway spalling fault repair involves complex physicochemical changes such as surface remelting, tissue refinement, residual stress regulation, etc. The contact fatigue life of the repaired bearing is calculated according to the theory of Lundberg-Palmgren:

$$L_{10} = \left(\frac{C}{P}\right)^p \cdot k_1 \cdot k_2 \cdot k_3 \tag{20}$$

where C is the dynamic load rating, P is the equivalent dynamic load, p is the life index, and k_1, k_2, and k_3 are the correction factors for repair process, material modification, and surface treatment, respectively.

The introduction of repair history information enables the diagnostic system to establish a more accurate component degradation model, and by recording detailed information such as time, location, process parameters and repair effect of each repair, the system constructs a cumulative damage equation containing repair factors:

$$D_{total}(t) = \sum_{i=1}^{n} w_i \cdot D_i(t) + \sum_{j=1}^{m} \beta_j \cdot R_j(t) \tag{21}$$

Where $D_i(t)$ represents the cumulative damage of the ind damage mechanism, w_i is the corresponding weight coefficient, $R_j(t)$ represents the influence function of the jth repair, and β_j is the repair effect coefficient, this mathematical model can more accurately predict the future health state evolution trend of the component, and significantly improve the fault prediction accuracy and reliability.

Repair program - fault type correlation matrix describes the degree of effectiveness of various types of repair programs for various types of faults, on the basis of which the repair program - fault type - the construction of the correlation matrix of detection information, so that the fault diagnosis based on the maintenance record is targeted. That is, in the fault diagnosis process, according to the maintenance of the component, and take different diagnostic programs and corresponding diagnostic parameters. Fault diagnosis based on the maintenance status is the maintenance results are fed back to the fault diagnosis module as an auxiliary means to determine whether the maintenance is successful. When the fault occurs after repair but the repair effect is not good, then the diagnostic threshold and alarm mode to make the corresponding correction, in order to prevent diagnostic errors caused by maintenance failure, to a certain extent, improve the diagnostic accuracy of the system.

3.2 IGWO algorithm to optimize LSSVM

In this paper, a new grey wolf optimization algorithm is proposed to improve the traditional grey wolf search, which has significant advantages in terms of parameter optimization for least squares support vector machines. Multifaceted improvement measures such as inertia weight factor, initial position chaos processing, and individual variation are proposed, and the problem that the original gray wolf algorithm tends to fall into local optimal solutions in the continuous function optimization process is effectively overcome. In order to solve the problem that LSSVM parameter optimization is easy to fall into local extreme value, Logistic chaotic mapping is used in the algorithm to improve the initialized population, and its chaotic sequence formula is:

$$x_{k+1} = \mu x_k (1 - x_k) \tag{22}$$

where $\mu = 4$ is the traversability and randomness characteristics of chaotic parameters, which ensures that the initial wolf pack position presents a more uniform distribution pattern in the parameter space, effectively eliminating the phenomenon of searching blind zones caused by the traditional random initialization method.

The dynamic inertia weighting mechanism abandons the limitations of the original linear convergence factor and adopts a nonlinear decreasing strategy for weight updating:

$$w(t) = w_{\max} - (w_{\max} - w_{\min}) \cdot \left(\frac{t}{T_{\max}} \right)^2 \tag{23}$$

where $w_{\max}$ and $w_{\min}$ denote the maximum and minimum inertia weights, respectively, t is the current number of iterations, and $T_{\max}$ is the maximum number of iterations, this design concept allows the algorithm to maintain a strong global exploration capability in the early stages of iteration while shifting the focus to local fine search in the later stages.

The adaptive mutation strategy dynamically adjusts the mutation probability according to the population diversity metric $diversity(t)$:

$$P_m(t) = P_{m_0} \cdot \exp\left(-\beta \cdot \frac{diversity(t)}{diversity_{\max}} \right) \tag{24}$$

The frequency of mutation operations is automatically increased when the population diversity measure decreases, thus effectively preventing the algorithm from converging to a suboptimal solution prematurely.

The mathematical modeling process of the LSSVM parameter optimization problem constitutes the theoretical basis for the practical application of the IGWO algorithm, and the design of the optimization objective function must establish a balancing mechanism between the model prediction accuracy, generalization ability, and computational complexity, and the joint optimization of the regularization parameter γ and the kernel function parameter σ can be expressed as follows:

$$\min f(\gamma,\sigma) = w_1 \cdot RMSE_{cv}(\gamma,\sigma) + w_2 \cdot \frac{1}{Accuracy_{cv}(\gamma,\sigma)} + w_3 \cdot Complexity(\gamma,\sigma) \tag{25}$$

where $RMSE_{cv}$ represents the cross-validation root mean square error, $Accuracy_{cv}$ represents the cross-validation accuracy, $Complexity$ is the model complexity measure, and the selection of the weighting coefficients w_1, w_2, and w_3 directly affects the quality of the optimization results.

The parameter search space was set in the range of $\gamma \in [10^{-3}, 10^{3}]$ and $\sigma \in [10^{-3}, 10^{3}]$, and a logarithmic scale division strategy was used to ensure that the search process could cover different orders of magnitude of parameter combinations. The position vector $X = [\log_{10}(\gamma), \log_{10}(\sigma)]$ of each individual wolf encodes a set of candidate parameters, and the mapping transformation from the search space to the actual parameter domain is realized by exponential transformation $\gamma = 10^{X_1}$ and $\sigma = 10^{X_2}$. The fitness function was evaluated using a k-fold cross-validation strategy, where the training dataset was randomly divided into k subsets, and the $k-1$ subsets were selected in turn to train the LSSVM model, while the remaining subsets were used for validation, and the fitness values were computed by the weighted average of the k validation results:

$$F(T) = \sum_{i=1}^{n} \alpha_i \cdot f_i(T) \tag{26}$$

where $f_i(T)$ denotes the performance index of the i nd fold validation, α_i is the corresponding weighting factor, and n is the number of cross-validation folds.

The position update mechanism of the IGWO algorithm is specially designed to adapt to the special requirements of continuous parameter space in LSSVM parameter optimization, and the algorithm maintains the hierarchical management structure of traditional gray wolf optimization. At the same time, an improved position update strategy is introduced to significantly improve the search efficiency. α wolf, β wolves and δ wolves correspond to the best, second best and third best individuals in the current population, respectively. Their position vectors contain the optimal parameter combination information, and the ω-wolf group updates its position by synthesizing the guidance information of the three leaders, and the improved position update formula is:

$$\vec{X}_1 = \vec{X}_\alpha - \vec{A}_1 \cdot \left| \vec{C}_1 \cdot \vec{X}_\alpha - \vec{X} \right| \cdot w(t) \tag{27}$$

$$\vec{X}_2 = \vec{X}_\beta - \vec{A}_2 \cdot \left| \vec{C}_2 \cdot \vec{X}_\beta - \vec{X} \right| \cdot w(t) \tag{28}$$

$$\vec{X}_3 = \vec{X}_\delta - \vec{A}_3 \cdot \left| \vec{C}_3 \cdot \vec{X}_\delta - \vec{X} \right| \cdot w(t) \tag{29}$$

$$\vec{X}(t+1)=\frac{\vec{X}_1+\vec{X}_2+\vec{X}_3}{3}+mutation(t) \tag{30}$$

Variation term $mutation(t)$ uses a Gaussian perturbation mechanism:

$$mutation(t)=N(0,\sigma_m^2)\cdot P_m(t) \tag{31}$$

where $N(0,\sigma_m^2)$ is a zero-mean Gaussian random number and σ_m is a variance strength parameter.

The boundary handling strategy ensures that the parameter search process always stays within the valid range, and a bounce-back mechanism is used when the position of an individual exceeds the search boundary:

$$X_{new}=X_{bound}+rand()\cdot(X_{bound}-X_{old}) \tag{32}$$

where X_{bound} is the boundary value and $rand()$ is a $[0,1]$-interval random number.

4 Experiments and analysis of results

4.1 Experimental design

In this study, a comprehensive experimental verification platform is constructed based on a certain turbofan engine, and a complete data set covering different states such as normal operation, typical faults, and damage repair is systematically acquired through a combination of multi-operating condition operation tests and artificial fault implantation. The experimental platform adopts distributed multi-sensor network architecture, and the sensor arrangement takes into account the constraints of fault sensitivity, signal reliability and engineering feasibility to realize the synchronous high-frequency acquisition of operating parameters in key parts of the engine. The configuration of the data acquisition system is shown in Table 2, and the vibration sensor adopts the PCB 353B04 model, with a range of ± 500g, a sampling frequency of 25.6kHz, and a measurement accuracy of ± 1%. Temperature monitoring using K-type thermocouples, temperature measurement range covers 0-1200 ° C, sampling frequency 100Hz, accuracy ± 0.5 ° C. Pressure sensor selection Kulite XCS-062, range 0-10MPa, sampling frequency 10kHz, accuracy ± 0.1% FS. speed sensor for Bentley 3300, speed range 0-20,000rpm, sampling frequency up to 25.6kHz, measurement accuracy 440.1% FS. Speed sensor is Bentley 3300, speed range 0- 20000rpm, sampling frequency 1kHz, accuracy ± 0.01%. The whole data acquisition process strictly follows the specification of aero-engine test to ensure the double guarantee of test safety and data quality.

Table 2: Parameters of Experimental Equipment

Device name	Model specification	Technical parameters	Sampling frequency	Measurement accuracy
Vibration sensor	PCB 353B04	Range: ±500g	25.6kHz	±1%
Temperature sensor	Kulite XCS-062	Range: 0-1200 °C	100Hz	±0.5 °C
Pressure sensor	Bentley 3300	Range: 0-10MPa	10kHz	±0.1%*FS*
Rotational speed sensor	NI PXIe-6358	Range: 0-20000 RPM	1kHz	±0.01%
Data acquisition card	PCB 482C05	16-bit resolution	1.25MS/s	±2.69mV
Signal conditioner	Dell Precision 7820	Gain 1 to 1000 times	DC-100kHz	±0.5%
Computer system	Kulite XCS-062	Intel Xeon E5-2687W	-	-

The descriptive statistics of the experimental data are shown in Table 3. The data acquisition adopts a multi-stage hierarchical acquisition strategy, and the baseline data acquisition is carried out in the normal engine running state, and the working conditions are set to cover the typical running points of slow, intermediate, and maximum continuous thrust, etc. Each working condition is run continuously for 30 minutes to ensure the thermal balance of the system. Each condition runs continuously for 30 minutes to ensure the thermal balance of the system. The acquisition frequency is set at 25.6kHz to meet the demand for high-frequency vibration signal analysis, while low-frequency parameters such as temperature and pressure are synchronously acquired at a frequency of 100Hz, resulting in 12,000 sets of sample data for the normal state. Failure state data are obtained by manually implanting typical failure modes, including four main types of bearing failures: inner ring defects, outer ring wear, blade cracks and carbon deposits in the combustion chamber, with three levels of damage for each type of failure: slight, moderate and severe. Bearing faults are manufactured by EDM on the raceway surface with artificial defects of 0.2-1.0mm in width and 0.1-0.5mm in depth, and blade cracks are simulated by wire-cutting process at the root of the blade with a length of 5-20mm and a depth of 1-5mm. Combustion chamber carbon deposits are formed by controlling fuel quality and combustion parameters to form carbon layers of different thicknesses on the inner wall. 9,000 sets of data are collected for each failure condition, and the total number of failure state samples is 36,000 sets. Repair state data collection is carried out after the completion of the damage repair process, the repair method covers laser cladding, plasma spraying, machining and other technologies, the quality of the repair is verified by non-destructive testing and qualified for data collection, and the post-repair operation test is set up under the same conditions as before the failure in order to quantitatively assess the repair effect, with 6,000 sets of sample data for the repair state, and the total number of experimental data reaches 54,000 sets.

Table 3: Descriptive statistics of experimental data

Data type	N	Feature Dimension	Mean range	SD range	Skewness range	Kurtosis range
Normal state	12000	42	0.15-2.87	0.23-1.45	-0.12-0.18	2.85-3.21
Bearing failure	9000	42	0.28-4.32	0.41-2.18	0.25-1.47	3.42-5.89
Blade failure	9000	42	0.19-3.65	0.35-1.89	-0.08-0.95	3.15-4.76
Combustion chamber failure	9000	42	0.22-3.98	0.38-2.05	0.18-1.23	3.28-5.12
Other faults	9000	42	0.26-3.54	0.33-1.97	0.12-1.08	3.07-4.85
Repair status	6000	42	0.17-3.12	0.27-1.62	-0.05-0.76	2.98-4.23
Total	54000	42	0.15-4.32	0.23-2.18	-0.12-1.47	2.85-5.89

Preliminary filtering of the original collected data and extraction of valid information, the vibration signal using Butterworth low-pass filter to do anti-aliasing processing, and set the cut-off frequency to 40% of the sampling frequency to prevent the phenomenon of spectral leakage; temperature and pressure signals are smoothed with a moving average filter, the size of the moving window depending on the specific circumstances. Using the overlap window to analyze the acquired segment of the signal is divided into a number of segments, each segment contains a data volume of N = 2048 sample points, the overlap rate between the two segments is 50%, in order to ensure that the time intervals represented by the two neighboring segments are continuous; the window function type is selected as the Hanning Hanning Window, which can reduce the impact of the energy leakage and the fence phenomenon on the frequency resolution; from the time, frequency, and the time-frequency domain The relevant characteristic parameters in the vibration signal are obtained from time, frequency and time-frequency domains. The time domain includes statistical parameters such as mean, variance, skewness coefficient, steepness coefficient, waveform factor, impulse factor, etc. The frequency domain generally utilizes the FFT transform to obtain the power spectral density, spectral center of gravity, and the ratio of energy in the frequency band. In the time-frequency domain, wavelet packet decomposition is used to obtain the energy and entropy features of each frequency band, which are composed of 42-dimensional feature vectors. In this paper, Z-score standard deviation is used to standardize the data, and different features with different measurement units are transformed into the same digital range value, and the standardization formula is:

$$x_{norm} = \frac{x - \mu}{\sigma} \tag{33}$$

Among them, μ and σ are the characteristic mean and standard deviation, respectively, and the outlier detection adopts the 3σ criterion to exclude the data points that obviously deviate from the normal distribution, and the datasets are randomly assigned as training, validation and test sets in the ratio of 7:2:1 to ensure that the various types of samples are uniformly distributed in different datasets.

4.2 Experimental results and analysis

In this paper, an aero-engine fault diagnosis methodology based on the constructed comprehensive experimental platform taking into account damage repair and the improved Gray Wolf optimization algorithm to optimize the least squares support vector machine is comprehensively validated. The experiment adopts a multilevel evaluation system, covering the dimensions of single fault type identification, concurrent diagnosis of multiple faults, repair state assessment, and real-time performance testing. The comparison benchmarks include classical methods such as traditional support vector machine, back propagation neural network, random forest, standard least squares support vector machine, and particle swarm optimization least squares support vector machine, which are fairly compared by unified data sets and evaluation criteria. The comparison results of the experimental results are shown in Table 4. The experimental results show that the improved Gray Wolf Optimization-Least Squares Support Vector Machine method proposed in this paper exhibits significant advantages in all four types of typical fault diagnosis tasks. The overall diagnostic accuracy reaches 97.8%, which is an average improvement of 6.5 percentage points compared with the traditional method, fully verifying the effectiveness and advancement of the method.

Table 4: Comparison of fault diagnosis accuracy rates by different methods

Diagnostic methods	Bearing failure (%)	Blade failure (%)	Combustion chamber failure (%)	Other faults (%)	Average accuracy rate (%)
Traditional support vector machine	89.3	87.6	85.2	88.1	87.6
BPNN	91.2	89.4	87.8	90.3	89.7
Random Forest	92.8	90.7	89.1	91.6	91.1
Standard least squares support vector machine	93.5	91.8	90.4	92.7	92.1
PSO-LSSVM	95.1	93.2	92.3	94.2	93.8
GA-LSSVM	94.7	97.6	96.8	93.8	93.4
IGWO-LSSVM	98.4	93.6	97.5	98.2	97.8

The identification correct rate of the method proposed in this paper in bearing fault diagnosis is 98.4%, which is 9.1% higher than that of the method based on the support vector machine, mainly because the improved gray wolf optimization algorithm can effectively optimize the parameters of the least-squares support vector machine and incorporate the damage repair information, while the bearing faults generally show certain periodic shock characteristics, the optimal kernel function parameters obtained by using the improved gray wolf optimization algorithm can The optimal kernel function parameters obtained by the improved gray wolf optimization algorithm can better distinguish this nonlinear fault characteristic.

Meanwhile the repair history information gives the algorithm more bases for state discrimination. The model test results show that the blade fault identification accuracy is 97.6%, which is 5.8% higher than the least squares support vector machine, indicating that the high frequency vibration signals caused by the blade crack growth are better described by the newly constructed radial basis kernel function, and the adaptive search capability based on the improved Gray Wolf algorithm also ensures the best pairwise relationship between the regularization factor and the parameters of radial basis kernel function, which can overcome the over-learning and under-learning phenomena well.

Among them, the accuracy of combustion chamber fault detection is 96.8%, combustion chamber faults are mainly based on temperature field and pressure fluctuations to determine the existence of faults, and then use the modified Gray Wolf algorithm under the collaborative decision-making of a variety of signals to assist in dealing with the mixed-type faults, and its accuracy is improved by 11.6% compared with the previous one; the rest of the various fault categories are able to achieve a high classification accuracy of 98.2%, respectively, and include the compressor instability, turbine blade over-temperature and pipe rupture faults, and other categories. The improved Gray Wolf Optimization-Least Squares Support Vector Machine algorithm is used to obtain a suitable classifier for multiple categories by performing a global search for the optimal parameters, which establishes an effective distinction between different fault types.

5 Conclusion

This paper establishes a least squares support vector machine fault diagnosis model based on the combination of damage repair theory and the improved gray wolf optimization algorithm.

The application of damage repair theory is a major feature of this paper, which breaks through the limitations of the original model, and utilizes the information containing the damage repair process as the evaluation index, so that the fault diagnosis results can reflect the overall condition of the part. The experimental results show that the diagnostic method containing repair information for all types of fault identification in the average accuracy of 0.6 percentage points, improve the gray wolf algorithm for parameter optimization accuracy of 97.8%, than the BPNN algorithm convergence speed increased by 8.1%. There are still some areas for further improvement and shortcomings in this study.

The fault mechanism model is not deeply mined, and the existing method is mainly based on statistical learning of data, which is more lacking in the physical analytical modeling of faults, and does not well grasp the law of fault development, which may lead to the fact that the method cannot be effectively applied to emerging fault forms and fault prognosis. Multi-sensor data fusion technology needs to be improved, although the introduction of vibration, temperature, pressure and other types of sensors. However, the integration method is relatively simple and does not take into account the complementarity and redundancy of information between the sensors, and how it affects the diagnostic effect in the case of sensor failure or poor signals remains to be further explored. These shortcomings point out the direction for the subsequent research, which needs to be gradually improved through the combination of theoretical deepening, algorithmic improvement and engineering practice.

About the Author

Manshan Li (1972.09-), male, Han ethnicity, from Hanyang, Hubei Province, holds a master's degree and works as a civil aviation engineer. His research focuses on the application of aviation safety big data and artificial intelligence technologies.

Hongping Wei (1991.07-), male, Han ethnicity, from Lanzhou, Gansu Province, holds a bachelor's degree and works as a civil aviation engineer. His research focuses on the application of aviation safety big data and artificial intelligence.

References

[1] Shen, J., Chen, M., & Chen, L. (2023, October). Effect of High Temperature Steam Inhalation on Aeroengine Performance. In Asia-Pacific International Symposium on Aerospace Technology (pp. 65-79). Singapore: Springer Nature Singapore.

[2] Guo, Y., Sun, Y., & Li, L. (2020). Research on probabilistic risk assessment of aeroengine rotor failure. Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering, 234(16), 2337-2347.

[3] Lee, D., Kwon, H. J., & Choi, K. (2024). Risk-based maintenance optimization of aircraft gas turbine engine component. Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability, 238(2), 429-445.

[4] Valkenborg, D., Rousseau, A. J., Geubbelmans, M., & Burzykowski, T. (2023). Support vector machines. American journal of orthodontics and dentofacial orthopedics, 164(5), 754-757.

[5] Xue, P., Yu, J., & Zhang, H. (2025). Fault diagnosis using ISMA to optimize SVM parameters for aircraft engine damage repair. Scientific Reports, 15(1), 42152.

[6] Wang, Y. S., Ma, Q. H., Zhu, Q., Liu, X. T., & Zhao, L. H. (2014). An intelligent approach for engine fault diagnosis based on Hilbert–Huang transform and support vector machine. Applied acoustics, 75, 1-9.

[7] Zhou, H., Huang, J., & Lu, F. (2018). Parsimonious kernel recursive least squares algorithm for aero-engine health diagnosis. IEEE Access, 6, 74687-74698.

[8] Deng, F., Guo, S., Zhou, R., & Chen, J. (2015). Sensor multifault diagnosis with improved support vector machines. IEEE transactions on automation science and engineering, 14(2), 1053-1063.

[9] Hashem, M. H., Abdullah, H. S., & Ghathwan, K. I. (2023). Grey wolf optimization algorithm: a survey. Iraqi Journal of Science, 5964-5984.

[10] Kumar, A., Pant, S., & Ram, M. (2017). System reliability optimization using gray wolf optimizer algorithm. Quality and Reliability Engineering International, 33(7), 1327-1335.

[11] Xia, Y., Zhao, J., Ding, Q., & Jiang, A. (2021). Incipient chiller fault diagnosis using an optimized least squares support vector machine with gravitational search algorithm. Frontiers in Energy Research, 9, 755649.

[12] Chen, M., Qu, R., & Fang, W. (2022). Case-based reasoning system for fault diagnosis of aero-engines. Expert Systems with Applications, 202, 117350.

[13] Li, C., Wen, J. R., Wan, J., Taylan, O., & Fei, C. W. (2024). Adaptive directed support vector machine method for the reliability evaluation of aeroengine structure. Reliability Engineering & System Safety, 246, 110064.

[14] Zhu, Y., Du, C., Liu, Z., Chen, Y. B., & Zhao, Y. P. (2022). A turboshaft aeroengine fault detection method based on one-class support vector machine and transfer learning. Journal of Aerospace Engineering, 35(6), 04022085.

[15] Wang, X. (2025). Aero Engine Fault Diagnosis Based on Support Vector Machine. Procedia Computer Science, 262, 1352-1358.

[16] Cui, J., Shan, M., Yan, R., & Wu, Y. (2014). Aero-Engine Fault Diagnosis Using Improved Local Discriminant Bases and Support Vector Machine. Mathematical Problems in Engineering, 2014(1), 283718.

[17] Fei, C. W., & Bai, G. C. (2013). Wavelet Correlation Feature Scale Entropy and Fuzzy Support Vector Machine Approach for Aeroengine Whole-Body Vibration Fault Diagnosis. Shock and Vibration, 20(2), 341-349.

[18] Wang, B., Ke, H., Ma, X., & Yu, B. (2019). Fault diagnosis method for engine control system based on probabilistic neural network and support vector machine. Applied Sciences, 9(19), 4122.

[19] Zhou, J., Liu, Y., & Zhang, T. (2019). Fault diagnosis based on relevance vector machine for fuel regulator of aircraft engine. International Journal of Machine Learning and Cybernetics, 10(7), 1779-1790.

[20] Chen, C., Zheng, Q. G., Jiang, F. J., & Zhang, H. B. (2025). Fault Diagnosis for Aircraft Engines Based on Multi-task Graph SVM Inference with Individual Differences. International Journal of Aeronautical and Space Sciences, 1-24.

[21] Hu, Y., Zhu, J., Sun, Z., & Gao, L. (2020). Sensor fault diagnosis of gas turbine engines using an integrated scheme based on improved least squares support vector regression. Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering, 234(3), 607-623.

[22] Zhao, Y. P., Wang, J. J., Li, X. Y., Peng, G. J., & Yang, Z. (2020). Extended least squares support vector machine with applications to fault diagnosis of aircraft engine. ISA transactions, 97, 189-201.

[23] Han, H., Cui, X., Fan, Y., & Qing, H. (2019). Least squares support vector machine (LS-SVM)-based chiller fault diagnosis using fault indicative features. Applied Thermal Engineering, 154, 540-547.

[24] Xi, P. P., Zhao, Y. P., Wang, P. X., Li, Z. Q., Pan, Y. T., & Song, F. Q. (2019). Least squares support vector machine for class imbalance learning and their applications to fault detection of aircraft engine. Aerospace Science and Technology, 84, 56-74.

[25] Xue, H., & Zeng, H. (2022, December). Fault detection of least squares support vector machine flight control system based on sparrow algorithm. In International Workshop on Automation, Control, and Communication Engineering (IWACCE 2022) (Vol. 12492, pp. 240-246). SPIE.

[26] Guangfu, L., Xu, W., & Jia, R. (2020). Multi-packet transmission aero-engine DCS neural network sliding mode control based on multi-kernel LS-SVM packet dropout online compensation. Plos one, 15(6), e0234356.

[27] Huang, P., Yu, H., & Wang, T. (2022). A study using optimized LSSVR for real-time fault detection of liquid rocket engine. Processes, 10(8), 1643.

[28] Sheng, H., & Zhang, T. (2017). Aircraft engine thrust estimator design based on GSA-LSSVM. International Journal of Turbo & Jet-Engines, 34(3), 279-285.