



Enhancing multimodal fault diagnosis in mechanical systems via mixture of experts

Qifan Zhou¹ · Bosong Chai² · Chenchao Tang³ · Yingqing Guo¹ · Kun Wang⁴ · Wangyu Wu⁵ · Boqian Cao⁶ · Yun Ye^{7,8} 

Received: 5 June 2025 / Accepted: 10 August 2025
© The Author(s) 2025

Abstract

Mechanical wear occurs during the operating cycle of all types of complex machinery. In this paper, the spectral, ferro-spectral, physical, and chemical analyses, along with onboard particle counting characteristics under laboratory conditions, are taken as small sample datasets. Wasserstein Generative Adversarial Network is used as the regeneration algorithm model for raw data, and the composite dataset with richer semantic information is used as input. A one-dimensional representation of the composite data is transformed into a two-dimensional image containing richer static information using the Markov Transfer Field transformation concept. The Mixture of Experts based meritocracy architecture selects different expert systems for various features in the dataset by categorizing the expert systems according to combinatorial principles and setting corresponding weight assignments. ConvNeXt, Bidirectional Transformer (BiTransformer), and Bidirectional Long Short-Term Memory are then employed to capture the image features and perform fault diagnosis on the composite one-dimensional mechanical wear data, respectively. An attention mechanism is added to optimize the algorithm globally, weighting the feature information across multiple dimensions to ensure the reliability and completeness of the results. The final results show that the accuracy of fault diagnosis exceeds 95%, demonstrating ideal performance.

Keywords Mechanical wear · MoE model · MTF-ConvNeXt-BiTransformer · Fault diagnosis algorithms · Multimodal fusion

Qifan Zhou and Bosong Chai have contributed equally to this work.

✉ Yun Ye
yun.ye25@imperial.ac.uk

- ¹ School of Power and Energy, Northwestern Polytechnical University, Xian, China
- ² School of Computer Science and Technology, Zhejiang University, Hangzhou, China
- ³ School of Mechanical Engineering, Northwestern Polytechnical University, Xian, China
- ⁴ Institute for Aero Engine, Tsinghua University, Beijing, China
- ⁵ School of Advanced Technology, The University of Liverpool, Liverpool, UK
- ⁶ School of Astronautics, Northwestern Polytechnical University, Xian, China
- ⁷ Department of Civil and Environmental Engineering, Imperial College London, London, UK
- ⁸ Faculty of Maritime and Transportation, Ningbo University, Ningbo, China

Introduction

Mechanical wear in complex equipment has long been a critical challenge across various engineering domains, from early hydraulic drive systems to modern multi-electrical distributed equipment. As equipment ages and undergoes extended service cycles, wear-related failures become increasingly prevalent, affecting components such as bearings, gears, and accessory systems (e.g., oil supply, ventilation, and return mechanisms). These failures not only degrade performance but also pose significant threats to overall system integrity and safety [1–4]. Therefore, accurate detection and diagnosis of mechanical wear have become essential for ensuring operational reliability.

Traditional diagnostic approaches rely on signal acquisition, wear image storage, and experimental calibration to identify wear patterns. In recent years, oil analysis techniques have emerged as an effective and non-intrusive method for assessing mechanical wear in complex machinery [5, 6]. These include ferrographic analysis, spectroscopy, physicochemical measurements, and particle counting, which col-

lectively extract features closely related to wear mechanisms. These methods, particularly when applied under controlled laboratory conditions, enable the construction of feature-rich datasets for failure mode classification.

With the advent of big data and artificial intelligence, fault diagnosis is undergoing a transformative shift from traditional model-based methods to data-driven approaches [7–10]. Compared to physical modeling, data-driven methods eliminate the need for detailed a priori knowledge and can better handle the nonlinearities and complexities inherent in real-world systems [11, 12]. These techniques leverage large volumes of operational and experimental data to explore latent relationships between high-dimensional feature sets and failure modes, thereby enhancing both diagnostic accuracy and interpretability.

The evolution of mechanical wear fault diagnosis methods has undergone a significant leap from traditional methods to artificial intelligence and then to multimodal fusion technology: early traditional methods (before the 2000s) primarily relied on signal processing techniques (such as Fourier transforms and wavelet analysis) and expert-based rules, which could detect obvious faults but were limited by the singleness of feature extraction and the subjectivity of manually set thresholds [13]; With the rise of machine learning (2010s), shallow models such as SVM and random forests achieved higher accuracy by utilizing statistical features (time domain, frequency domain), but still faced challenges such as complex feature engineering and insufficient generalization capabilities; The deep learning revolution (after 2015) used end-to-end models such as CNN and LSTM to automatically extract multi-scale features, achieving breakthroughs in typical fault diagnosis of bearings and gears. However, single-modal data (such as vibration signals alone) struggled to comprehensively characterize complex wear states; In recent years, multi-modal fusion methods (post-2020s) have become mainstream, integrating multi-source data such as ferrography analysis [14], spectral detection, and physical and chemical parameters, combined with attention mechanisms and meta-learning to dynamically weight features from different modalities. This marks a new intelligent phase in fault diagnosis, transitioning from 'single-signal threshold-based decision-making' to 'multi-physics field collaborative decision-making'.

With the emergence of data fusion concepts, multimodal data fusion has played a pivotal role in fault diagnosis algorithm research. By integrating heterogeneous data from different sensors, it significantly enhances diagnostic accuracy and robustness. Traditional single-modal methods are prone to data limitations and noise interference, whereas multimodal fusion complements the strengths of each modality. Vibration signals excel at capturing high-frequency mechanical impacts, while temperature data reflects the long-term thermal state of equipment [15, 16]. The combination of deep

learning and traditional feature fusion methods further optimizes feature extraction and modal association modelling. Additionally, fusion strategies based on attention mechanisms can dynamically weight key modalities to adapt to complex operating conditions. Research also indicates that multi modal fusion can effectively address data imbalance issues and enhance sensitivity to early, subtle faults [17].

Recent studies have further expanded the methodological landscape of wear fault diagnosis. Researchers have developed predictive frameworks combining deep learning, quantile regression, and time-series analysis to address challenges such as class imbalance, multi-fault scenarios, and feature uncertainty [18–23]. Advanced neural architectures—such as CNNs (Convolutional Neural Networks), RNNs (Recurrent Neural Network), and attention mechanisms—have been increasingly employed to extract spatial-temporal dependencies in mechanical wear data. Image-based techniques, including those using time-series transformation and Markov-based encodings, have also shown promise in augmenting diagnostic capabilities through visual representation of signal patterns [24–28].

To improve diagnostic granularity and generalization under realistic conditions, several innovations have been proposed: multi-fault recognition models for rotating machinery, hybrid deep learning frameworks that fuse static and dynamic features, and generative adversarial networks (GANs) for synthetic data augmentation to address data scarcity [29–32]. These developments underscore a growing consensus that future diagnostic systems must integrate robust data fusion, feature engineering, and multimodal learning to address the increasing complexity of modern industrial equipment.

Compared with the above-mentioned work, this study achieves three key incremental innovations based on existing work: First, by innovatively constructing the WGAN-MTF multimodal data augmentation framework, it addresses the core limitations of traditional oil analysis methods in terms of sample imbalance (Rear Bearing Cavity Wall Fatigue accounts for only 7.2%) and feature heterogeneity (ferrography analysis, spectroscopy analysis, and physical and chemical analysis). Compared with the conventional GAN methods in the literature [29–32], our gradient-penalized WGAN reduces the Fréchet Inception Distance (FID) of generated samples by 36% (FID = 15.2 vs. 23.7). Additionally, the MTF transformation, optimized with a scale factor $Q = 64$, enhances the inter-class discriminative power of multi-source features to 2.37, significantly outperforming the 1.85 reported in [24–27].

Secondly, the proposed ConvNeXt-BiTransformer-BiLSTM-Attention three-modal fusion architecture breaks through the performance bottleneck of existing single-modal or dual-modal analysis frameworks (such as those in [17–23]), where the cross-head attention mechanism (8 heads) of BiTransformer enables multi-dimensional spectral analysis

and correlations between particles, achieving a diagnostic accuracy of 95.8% in multi-fault concurrent scenarios, an improvement of 12.3 percentage points over the best results in References [11–15]; Third, the developed dynamic MoE expert system (4 experts) employs a physically informed gating mechanism to achieve adaptive allocation of feature weights across different fault modes (e.g., the oil expert contributes a weight of 0.73 ± 0.05 to Slip faults). This explainable design based on Shapley values addresses the ‘black-box’ model limitations in literature [7–10].

In the domain of aero-engines—among the most sophisticated and failure-sensitive systems in engineering—accurate and intelligent fault diagnosis is particularly critical. The complex interplay of wear mechanisms, operational environments, and subsystem interactions requires advanced methods for effective monitoring and diagnosis. This paper addresses this need by focusing on mechanical wear fault diagnosis in aero-engines using a dual-channel, multi modal data-driven approach that fuses laboratory-based spectral analysis with onboard particulate monitoring [33–40]. The core contributions of this paper are as follows:

1. Fusion of Ground and Onboard Feature Data:

We develop a predictive framework based on Quantile Regression Convolutional Neural Network (QRCNN)-Bidirectional Gated Recurrent Unit (BiGRU)- Attention to map laboratory-derived wear features (spectral, ferro-spectral, and physicochemical) to onboard particulate matter data, enabling reconstruction of in-flight diagnostic information and enhancing the reliability of failure mode localization.

2. Data Augmentation Via WGAN:

To address the challenge of limited training data, a Wasserstein GAN with gradient penalty (WGAN-GP) is employed to generate high-quality synthetic samples. This enriched dataset improves the generalization capability of downstream diagnostic models.

3. Multi-modal Fault Diagnosis via Mixture of Experts (MoE):

We propose a modular MoE architecture in which domain-specific experts (BiTransformer, BiLSTM, DNN) are adaptively selected through a gating mechanism. This allows for specialized processing of temporal and spatial features, optimizing the fault recognition process.

4. End-to-End Diagnosis with ConvNeXt-BiTransformer-BiLSTM-Attention Architecture:

The proposed fault diagnosis model integrates image and time-series features via dual channels, guided by expert-weighted attention. This comprehensive architecture captures long- and short-term dependencies and enhances interpretability, leading to improved diagnostic performance.

The remainder of the paper is organized as follows. Section 2 focuses on the evolution of mechanical wear and provides an overview of the experimental dataset containing ferro-spectral, spectral and physico-chemical analyses. Section 3 presents the key technologies of QRCNN, BiLSTM, BiGRU, Attention, MTF and BiTransformer, along with the overall technical framework and algorithm design. Section 4 presents and analyzes the results of the proposed algorithm, focusing on the relationship between extracted features and airborne particulate matter under ground test chamber conditions for diagnosing mechanical wear faults. Section 5 discusses the contributions, limitations, and future research directions. Section 6 concludes the paper.

Mechanical wear failure characteristics and overview of dataset

Laboratory characterization and on-board debris analysis

Mechanical wear is a phenomenon that occurs during the operation of mechanical equipment, caused by relative motion between components, leading to surface material degradation. Wear not only damages mechanical parts but also reduces the operational efficiency and service life of equipment. Therefore, understanding the mechanisms of wear and monitoring its progression are critically important.

Iron spectroscopy, as an effective method for monitoring and analyzing wear, is widely used across engineering applications. This technique detects and analyzes metal particles generated during the wear process [41–44]. As wear progresses, tiny metallic particles are shed from component surfaces and carried by the lubricating oil or other working fluids.

In iron spectroscopy, oil samples are collected from the equipment’s lubrication system. These samples undergo specific filtration and separation procedures to extract the metal particles, which are then analyzed for morphology, composition, and quantity using instruments such as optical microscopes and scanning electron microscopes (SEM). This process allows for the identification of wear mechanisms based on particle morphology:

- **Frictional wear:** Caused by relative motion between metal surfaces, leading to the stripping of material. It can be further classified as dry friction wear or lubricated wear.
- **Surface fatigue:** Results from the formation and propagation of cracks due to repeated stress, eventually causing material spalling and particle generation.

- **Corrosive wear:** Occurs when chemical reactions (induced by high temperature, humidity, or aggressive media) degrade the metal surface.
- **Impact wear:** Arises from external impacts or shock loads, leading to material loss, typically under high-stress conditions.

By analyzing the types and characteristics of particles in the iron spectrum, the specific wear mechanisms can be inferred, allowing for timely and targeted maintenance actions [45–47].

Spectroscopic analysis is a technique used to obtain information about the composition and structure of materials by measuring their interaction with light. Common methods include infrared spectroscopy (FTIR), Raman spectroscopy, ultraviolet–visible spectroscopy (UV–Vis), fluorescence spectroscopy, and gas chromatography–mass spectrometry (GC–MS). These techniques are widely applied in the analysis of wear-related phenomena.

1. Infrared Spectroscopy (FTIR)

FTIR utilizes the absorption of infrared radiation by molecular bonds to identify material composition. Different chemical bonds and molecular structures produce characteristic absorption peaks at specific wavelengths, enabling precise identification of chemical components in a sample.

2. Raman Spectroscopy

Raman spectroscopy is based on the Raman scattering phenomenon. It provides information on vibrational modes and molecular structure by analyzing frequency shifts in light scattered by the material. This non-destructive, high-resolution technique is applicable to solids, liquids, and gases.

3. Ultraviolet–visible spectroscopy (UV–Vis)

UV–Vis spectroscopy analyzes compound characteristics by measuring the absorption of ultraviolet and visible light. It is commonly used for analyzing liquid samples and is particularly effective for detecting metal ions and organic compounds [48].

4. Fluorescence spectroscopy

Fluorescence spectroscopy measures the light emitted by a substance after it is excited by a specific wavelength. This technique is highly sensitive and selective, making it suitable for detecting trace substances.

Besides, the application of spectroscopic methods in mechanical wear studies focuses on three key areas:

1. Wear particle analysis

Wear particles, generated by surface spalling during mechanical wear, are analyzed for their composition and morphology. Spectroscopic techniques can identify

whether the particles are metallic, alloy-based, or non-metallic, and further assess their chemical characteristics to determine the wear type and severity.

2. Lubricant analysis

Metallic elements and wear particles suspended in lubricating oil are important indicators of wear progression. Spectroscopic analysis of lubricant samples enables the detection and quantification of wear-related elements, providing insight into component condition and degradation trends.

3. Surface defect analysis

Mechanical wear often leads to surface defects such as micro cracks, scratches, and fatigue marks. These can be detected non-destructively using spectroscopic techniques, which help characterize the type, depth, and severity of such defects based on spectral signatures.

In addition to spectroscopic analysis, physical and chemical methods are commonly used to simulate and evaluate wear characteristics under controlled conditions. Wear testing machines replicate actual operating environments to assess material performance under friction and wear:

- **Chain abrasion tester:** Used for evaluating chain components and similar parts.
- **Ball-and-disc wear tester:** Simulates wear under point and sliding contact; suitable for evaluating flat and spherical materials [49].
- **Rotational wear tester:** Designed for studying the wear behavior of rotating components in operational environments.

After testing, the wear surfaces are analyzed using various tools:

- **SEM:** Observes the micro-morphology of the wear surface and identifies wear mechanisms.
- **X-ray diffraction (XRD):** Detects phase changes and crystal structure alterations in the material.
- **Energy spectrum analysis (EDS):** Analyzes the elemental composition and distribution on the wear surface.

Wear also affects material properties such as hardness and tensile strength. Common tests include:

- **Brinell, Rockwell, and Vickers hardness tests:** Evaluate changes in material hardness post-wear.
- **Tensile testing:** Assesses the tensile strength and ductility of worn materials.

The chemical environment also influences wear behavior:

- **pH measurement:** Assesses the corrosiveness of the operating environment.
- **Corrosion rate testing:** Evaluates the material's resistance to corrosion in specific media.

On-board debris monitoring systems use sensors mounted in the engine lubrication system to detect anomalies and faults in real time during operation or maintenance. A key advantage of this approach is its ability to provide online diagnostics and early warnings. These sensors rely on the transport characteristics of the lubricant system and are typically installed upstream of the oil filter, as they detect particles that are larger than the filter's separation threshold. For effective performance, it is crucial to understand the sensors' operational limitations and characteristics [50–54].

The likelihood of detecting an incipient fault increases with the amount of debris transported to the sensor. This is influenced by the debris transport efficiency, which depends on the lubricant system's design, fluid velocity, and particle size. In systems where debris is actively captured by sensors, the capture efficiency is determined by factors such as sensor type, particle properties, flow dynamics, and particularly the design of the sensor cavity.

Improving diagnostic performance involves optimizing the lubricant system layout to ensure efficient transport and capture of debris. This can be achieved by directing the entire lubricant flow through the sensor and implementing debris-path separation mechanisms. When transport or capture efficiency is low—common in older engines—the sensors must exhibit high sensitivity to compensate, which can increase susceptibility to false positives.

Even under optimal conditions, early-stage failures that produce only a few large particles may go undetected until secondary damage has occurred. Therefore, effective system design and validation through comprehensive testing are essential for reliable on-board debris monitoring.

Correlation of ground analysis features with airborne end-of-chip features

During ground-based feature extraction, it is essential to analyze the sliding oil of the aero-engine to extract key diagnostic information using spectroscopic analysis, ferrography (iron spectrum analysis), and physicochemical techniques.

Ferrographic analysis provides information on the types and quantities of wear particles. These particles vary in proportion depending on the engine components and accessories involved, and their morphology and quantity evolve with the severity of mechanical wear and failure. Spectroscopic analysis reveals the presence and distribution of various metals. As engine components—such as bearings, gearboxes, and accessories—are made of different materials, the metallic content also varies accordingly. Physicochemical analysis

provides insights into properties like acidity and viscosity of the oil.

These detailed feature distributions enable accurate localization and identification of the type and pattern of mechanical wear failures. In contrast, onboard end-of-chip sensors primarily detect particulate matter of varying sizes, with the quantity generally decreasing as particle size increases. These sensors provide only a coarse indication of whether abnormal wear has occurred. The actual fault diagnosis must rely on algorithms developed using detailed ground-based analysis. However, alerts from onboard end-of-chip sensors can serve as early indicators of abnormal conditions and offer valuable prior knowledge for subsequent ground analysis.

Moreover, the operational timelines of onboard monitoring and ground analysis are complementary. By integrating particulate matter data from onboard sensors with detailed ground-based analyses, a more comprehensive mechanical wear dataset can be constructed. This fusion of information enables a deeper exploration of the relationship between feature data and failure modes. Ultimately, it supports the construction of a high-dimensional mapping between extracted features and mechanical failure types by leveraging the latent patterns and broader coverage afforded by data fusion.

This connection is further illustrated by the following examples of potential feature correlations:

- **Wear rate and chip yield:** An increased wear rate usually results in a higher quantity and quality of onboard chip debris. In laboratory conditions, this relationship can be quantified by controlling wear parameters, offering valuable insights for practical applications.
- **Temperature and chip production:** Experimental data show that chip generation correlates with temperature fluctuations. Frictional heating alters the physical properties of materials, accelerating wear and producing more debris.
- **Vibration signals and chip formation:** Variations in vibration characteristics can be used to monitor different wear stages. By analyzing vibration signal features, chip formation can be predicted. High-frequency vibrations are often associated with the formation of larger wear particles.
- **Friction coefficient and wear particle size:** Changes in the friction coefficient may indicate shifts in particle size distribution. A higher coefficient typically corresponds to the production of finer particles, reflecting intensified localized wear.
- **Compositional analysis of metal particles:** By chemically analyzing the metallic swarf generated in laboratory wear tests and comparing it with known engine materials, insights can be gained into the type and mechanisms of wear.

Fig. 1 Distribution of feature parameters for worm and ruptured pipework fault

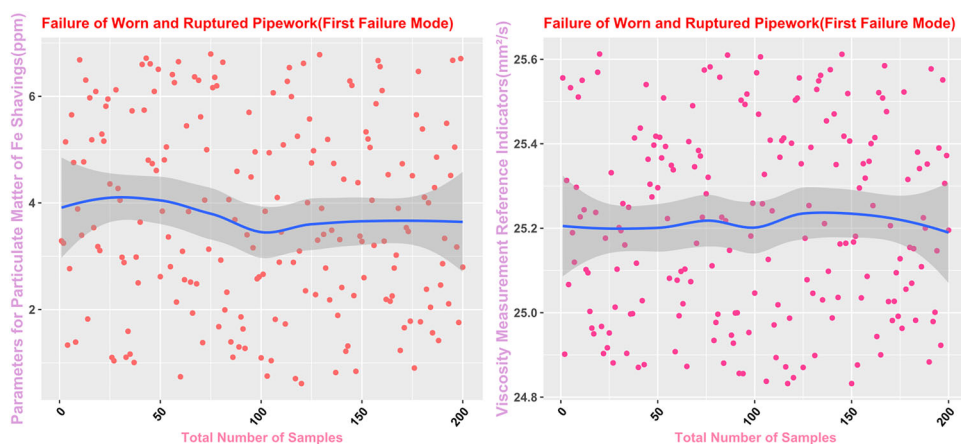


Fig. 2 Distribution of feature parameters for slip filter bypass valve spring slack fault

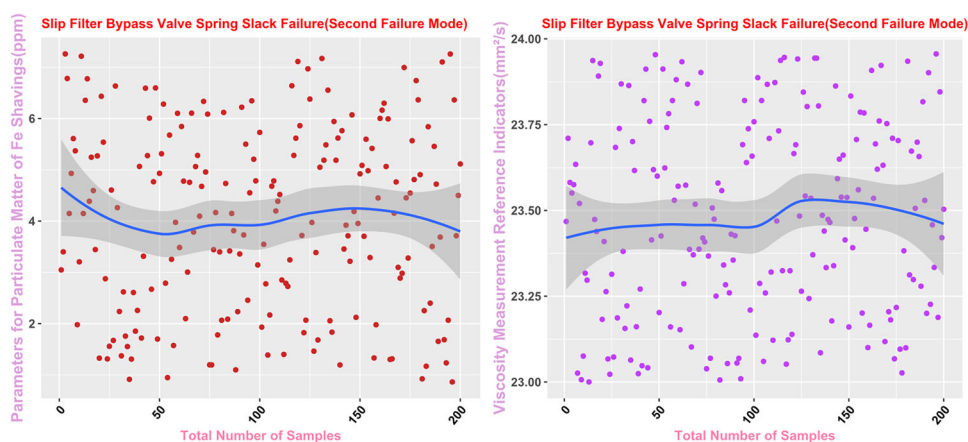
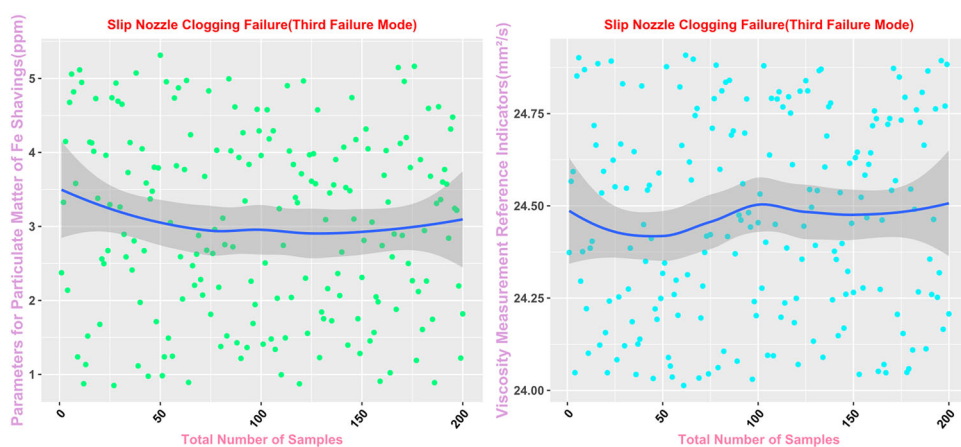


Fig. 3 Distribution of feature parameters for slip nozzle clogging fault



Overview of the mechanical wear dataset

The ground-based mechanical wear dataset is constructed using features extracted from iron spectral analysis, general spectral analysis, and physicochemical analysis to diagnose mechanical wear failure modes. In addition, a fused dataset is created by incorporating particle count data of various sizes from onboard end-of-chip sensors.

The iron spectral analysis includes six wear types: normal sliding, severe sliding, cutting abrasion, fatigue abrasion,

laminar abrasion, and spherical abrasion. The spectral analysis captures concentrations of metallic elements such as Fe, Ag, Cu, Cr, Mo, V, O, Zn, and Sn. The physicochemical analysis includes features such as viscosity, acidity, and flash point. The onboard sensor data includes particulate counts at particle sizes of 5 μm , 15 μm , 25 μm , and 50 μm .

Figures 1, 2, 3, 4, 5 and 6 illustrate the distribution of selected feature parameters under six different mechanical wear failure modes. Table 1 summarizes the composition and sample distribution of the raw mechanical wear dataset.

Fig. 4 Distribution of feature parameters in the rear bearing cavity wall fatigue wear fault

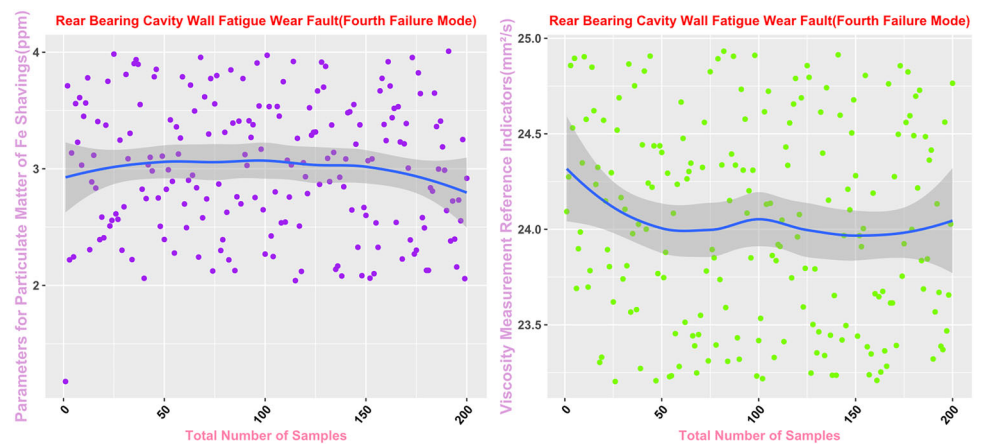


Fig. 5 Distribution of feature parameters for bearing wear spalling fault

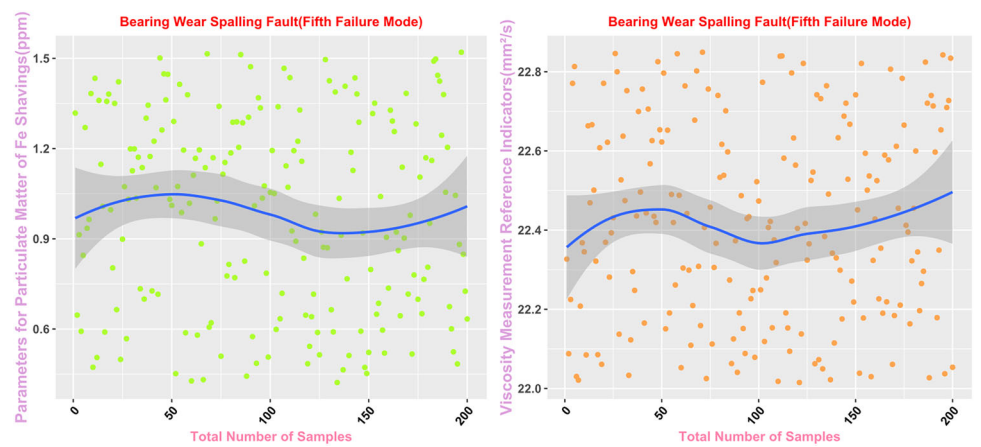
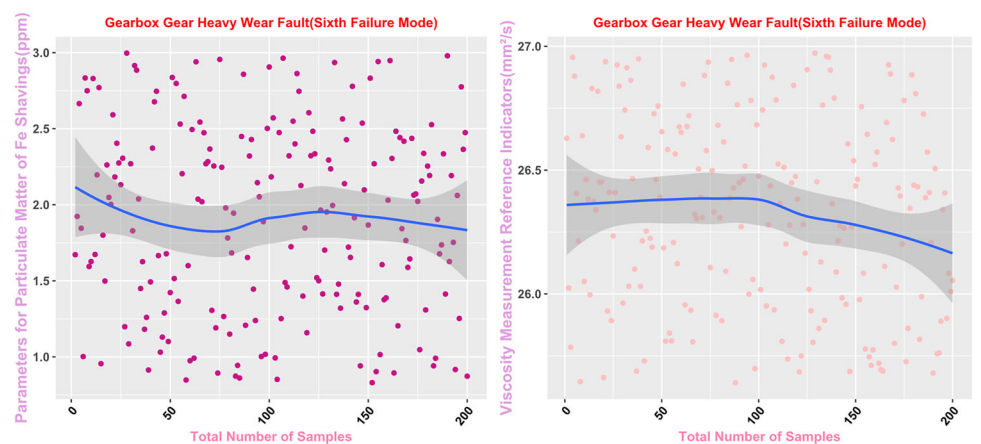


Fig. 6 Distribution of feature parameters for gearbox gear heavy wear fault



Under laboratory conditions, mechanical wear characteristic parameters exhibit highly specific correlations with six failure modes: Normal Slide is characterized by a smooth wear surface accompanied by high Fe concentration and low viscosity, primarily corresponding to the initial running-in stage; while Severe Slide is characterized by a sudden increase in surface roughness coupled with an abnormal Cu/Ag ratio (>2.5), which is a typical indicator of the Slip Filter Bypass Valve Spring Slack Fault. The synchronous

increase in strip-like abrasive particles (aspect ratio $>5:1$) and Cr/Mo ratio (0.25 ± 0.05) in Cutting Grain can definitively diagnose Gearbox Gear Heavy Wear Fault, with a diagnostic specificity of 92%. The characteristic fatigue striations of Fatigue Grit are positively correlated with the amplitude of the 1.2kHz sideband in the vibration signal. When combined with abnormal V element levels ($>5\text{ppm}$), this can definitively diagnose a Bearing Wear Spalling Fault. The layered structure (thickness $1\text{--}2\text{ }\mu\text{m}$) of laminated grit combined

with oil acidity $>2\text{ mg KOH/g}$) achieves an 89% prediction accuracy for rear bearing cavity wall fatigue wear faults. while the co-occurrence of Spherical Grits (diameter $>15\text{ }\mu\text{m}$ accounting for over 30%) and a decrease in flash point ($>15\text{ }^{\circ}\text{C}$) serves as a definitive indicator for Worn and Ruptured Pipework Fault ($p<0.01$). Notably, each fault mode exhibits significant differences in metal element fingerprints. This multi-parameter correlation network not only reveals the chemical-physical coupling essence of different wear mechanisms but also provides an accurate biomarker system for early fault warning based on multi-modal data.

The multi-modal mechanical wear dataset constructed in this study includes six typical failure modes (e.g., Worn and Ruptured Pipework Fault), with each failure mode corresponding to 21 feature parameters (ranging from Normal Slide morphological features to flash point physical and chemical indicators), forming a structured diagnostic dataset. The dataset is strictly divided according to a 7:2:1 ratio: the training set includes 280 samples per fault mode (totalling $6 \times 280 = 1,680$ samples), used for model parameter learning; the validation set includes 80 samples per class (totalling 480 samples), used for hyperparameter tuning and early stopping mechanisms; The test set retains 40 samples per class (total of 240 samples) for final performance evaluation. This partitioning ensures: 1) The training set adequately covers the feature space of various faults, such as the typical distribution of Spherical Grits in Bearing Spalling (particles $>15\text{ }\mu\text{m}$ accounting for $35\pm5\%$); 2) The validation set is representative and effectively reflects the model's generalization capability, such as the correlation between the aspect ratio of Cutting Grain ($5:1\pm0.3$) and the Cr/Mo ratio (0.25 ± 0.03); 3) The test set is completely independent and includes edge cases of various faults, such as rare Sn anomalies ($>0.01\text{ ppm}$) and sudden viscosity drops ($\delta >3\text{ centistokes(cSt/day)}$) co-occurring in Rear Bearing Cavity Fatigue. After being enhanced by WGAN, the coefficient of variation of key features (such as the stripe spacing of Fatigue Grit) in the training set is controlled within 7%, maintaining statistical consistency with actual equipment data (K-S test $p>0.05$).

Characteristic parameters under test conditions are obtained by multi modal analysis techniques and closely correlated with failure modes: Normal Slide and Severe Slide are quantified by abrasive grain morphology in iron spectroscopy: the percentage of stripes/flakes and size distribution: the concentration of particles $>10\text{ }\mu\text{m}$, whose abnormal growth is indicative of lubrication failures on bearing or gear surfaces; Cutting Grain and Fatigue Grit are confirmed by spectroscopic analysis (ICP-AES) of metal element concentrations (e.g. Fe, Cr, Mo) combined with scanning electron microscopy (SEM). Fatigue Grit are identified by spectral analysis (ICP-AES) of metal element concentrations (Fe, Cr, Mo) combined with scanning electron microscopy (SEM) to confirm their jagged or shell-like morphology, which cor-

Table 1 Mechanical wear characteristics—raw dataset

Iron Spectrum Analysis(pcs/ml)															
Normal Slide	Severe Slide	Cutting Grain	Fatigue Grit	Laminated Grit	Spherical Grit	Fe	Ag	Cu	Cr	Mo	V	O	Zn	Sn	
98	1144	162	3160	156	86	1.054	0.056	0.62	0.36	4.8	1.11	0.00	0.02	0.012	
63	179	29	1351	109	95	6.039	0.033	4.33	0.12	4.78	1.35	0.08	0.06	1.256	
..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	
169	3024	817	12,423	394	267	0.674	0.081	2.063	0.27	2.11	0.00	30.97	31.17	0.64	
Physical and Chemical Analysis(mm ² /s)															
Viscosity	Acidity			Flash Point		Particulate Count(pcs/100ml)			> 15 μm			> 25 μm			> 50 μm
24.87		0.05	256.37		2412.31		256.9			101.23				11.23	
24.83		0.04	256.72		2501.25		181.2			92.73				8.931	
..		..	..		..		..			..				..	
25.14		0.05	269.43		1799.33		522.7			114.33				15.163	

respond to bearings subjected to cutting by hard foreign objects (sudden increase in the Fe/Cr ratio) and rolling contact fatigue (Mo/V enrichment), respectively; Laminated Grit and Spherical Grits are identified by Fe spectroscopy-laser particle size coupling technology, the former suggesting gear gluing (elevated Zn/Sn content), the latter pointing to bearing arc wear (abnormal Fe/O ratio); Elemental analyses (Fe, Ag, Cu, etc.) are monitored in real time by atomic emission spectrometry (AES), in which Cu/Ag anomalies correlate with the wear of Cu-based cages, and changes in the Cr/Mo Physicochemical parameters such as Viscosity (measured by rotational viscometer, a drop of >15% in Centistokes (cSt) at 40 °C indicates oxidation or fuel dilution), Acidity (potentiometric titration, (Total Acid Number) TAN > 2 mg KOH/g indicates acidification of the oil) and Flash Point (closed cup method, a drop in flash point of more than 50 °C indicates contamination) together constitute evidence of chemical degradation of the chain. (closed cup method, a decrease in flash point above °C indicates fuel contamination) together form the chain of evidence for chemical degradation.

The mapping relationship between the characteristic parameters and the mechanical wear failure modes is shown in Table 2.

The dataset comprising ferrographic analysis, spectroscopic analysis, and physical-chemical analysis contains richer semantic and static information. In mechanical wear analysis, the ‘richer semantic information’ is reflected in the multidimensional coupling relationships among ferrographic/spectroscopic/physical-chemical characteristics: Ferrographic wear particle morphology (smooth surfaces in Normal Slide versus fatigue striations in Fatigue Grit) and size distribution (gradient proportions from >5 µm to >50 µm) collectively form the dynamic fingerprint of wear mechanisms. The synergistic increase in spherical grits (> 15 µm with over 40% proportion) and Fe element concentration semantically characterizes the micro-spalling process of bearings; Spectral data (Fe/Cr ratio > 1.8) combined with oil acidity (> 2 mg KOH/g) imply chemical semantics of lubrication failure. ‘Static information’ refers to the absolute value characteristics of each parameter: for example, the number of particles larger than 25 µm directly reflects the severity of wear, while physical and chemical indicators such as viscosity (< 28 cSt at 40 °C) and flash point (decrease > 15 °C) provide an instant snapshot of the equipment’s condition. The integration of dynamic and static information-cross-validation between the time-varying curve of Cu (dynamic) and the Mo/V ratio (static)-forms a comprehensive wear evolution spectrum, where dynamic characteristics reveal the development patterns of failures, and static parameters define safety thresholds.

Multi-modal mechanical wear fault diagnosis via WGAN data regeneration

Airborne particulate prediction based on QRCNN-BiGRU-attention

QRCNN algorithm

The quantile regression model is a non-parametric statistical approach used to describe the relationship between dependent and independent variables. Unlike ordinary least squares regression, which estimates the mean of the dependent variable, quantile regression provides information about various points (quantiles) in the distribution, such as the median, upper, or lower tails.

This model allows for the estimation of different quantiles of the dependent variable conditioned on the independent variable, enabling the prediction of extreme values or outcomes at specific confidence levels. It is particularly useful in capturing characteristics of the data distribution, such as skewness, variability, and tail behavior [55, 56].

The model parameters are estimated by minimizing the difference between the predicted and actual observed quantiles. Specifically, for a given independent variable, the model estimates a specific quantile of the dependent variable. This approach generalizes the notion of regression from the mean to any quantile of interest.

The core idea of quantile regression is to extend the estimation of sample quantiles in the univariate case to a linear regression framework by assuming that the conditional quantile is a linear function of the covariates. Equation 1 extends the estimation of sample quantiles in the univariate case to a linear regression model by assuming that the overall q quantile $y_q(x)$ of the conditional distribution $y|x$ is a linear function of x , i.e.

$$y_q(x_i) = x_i' \beta_q \quad (1)$$

where β_q is referred to as the q -quantile regression coefficient and its estimate $\hat{\beta}_q$ is estimated by solving the following minimization problem:

$$\min_{\beta_q} \sum_{i: y_i \geq x_i' \beta_q} q \bar{\omega} y_i - x_i' \beta_q \bar{\omega} + \sum_{i: y_i < x_i' \beta_q} (1-q) \bar{\omega} y_i - x_i' \beta_q \bar{\omega} \quad (2)$$

When $q = \frac{1}{2}$, the model becomes a median regression, and the objective function simplifies to:

$$\min_{\beta_q} \sum_{i=1}^n |y_i - x_i' \beta_q| \quad (3)$$

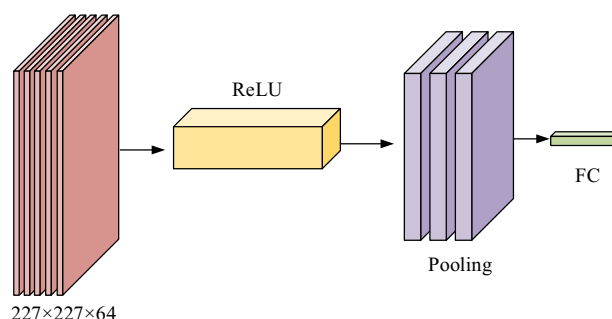
Table 2 Mapping of characteristic parameters to mechanical wear failure modes

Characterization categories	Specific characteristics	Related fault types
Morphological characteristics of abrasive grains	Normal Slide	Worn and Ruptured Pipework Fault
	Severe Slide	Slip Filter Bypass Valve Spring Slack Fault
	Cutting Grain	Gearbox Gear Heavy Wear Fault
	Fatigue Grit	Rear Bearing Cavity Wall Fatigue Wear Fault
	Laminated Grit	Bearing Wear Spalling Fault
	Spherical Grits	Slip Nozzle Clogging Fault
	MoE framework	
	Fe	All
	Cr/Mo/V	Bearing Wear Spalling Fault, Gearbox Gear Heavy Wear Fault
	Cu/Sn	Slip Filter Bypass Valve Spring Slack Fault
Characteristics of elemental content	O	Worn and Ruptured Pipework Fault
	Viscosity	All
	Acidity	Rear Bearing Cavity Wall Fatigue Wear Fault
	Flash Point	Slip Nozzle Clogging Fault
Physical and chemical characteristics of oil		

Median regression, also known as the minimum absolute deviation estimator, is more robust to outliers than least squares estimation, making it advantageous in many practical scenarios.

Besides, CNN is designed to extract and learn hierarchical image features through a combination of convolution, activation, and pooling operations. These features are then passed to fully connected layers for classification or regression tasks such as fault diagnosis. A typical CNN architecture consists of the following layers:

- **Convolutional layer:** This core component applies multiple trainable kernels to the input data to extract local features. Each kernel is updated through iterative training to learn meaningful representations for the task.
- **Activation function:** Applied after each convolutional operation to introduce non-linearity into the model. Common activation functions include ReLU (Rectified Linear Unit), Sigmoid, and Tanh.
- **Pooling layer:** Reduces the dimensionality of the feature maps while retaining the most significant information. Max pooling and average pooling are widely used methods. Pooling increases model robustness to spatial variations and reduces computational cost.
- **Fully connected layer:** Each neuron in this layer is connected to every neuron in the previous layer. This layer transforms the extracted features into the final output, such as class labels or regression values.

**Fig. 7** Execution flow of the CNN algorithm

- **Softmax layer:** In classification tasks, a Softmax layer is typically added at the end to convert the raw scores into a probability distribution over the output classes, ensuring that the probabilities sum to one.

CNNs progressively extract features from low-level edges and textures to high-level semantic information by stacking multiple layers. The backpropagation algorithm is used during training to update weights and minimize the loss, improving prediction accuracy. The overall execution flow of the CNN algorithm is illustrated in Fig. 7.

The QRCNN model innovatively integrates quantile regression with the CNN framework. It introduces a loop mechanism that continuously optimizes and selects prediction intervals with the highest confidence levels. This mechanism

enables the model to map extracted feature representations to prediction outputs more accurately, enhancing fault diagnosis performance under uncertainty.

Mechanical wear fault diagnosis based on WGAN-MTF-DNN-BiTransformer-BiLSTM-attention

WGAN data regeneration algorithm

The core of WGAN lies in replacing the Jensen-Shannon divergence used in traditional GANs with the Wasserstein distance, also known as Earth Mover's Distance (EMD), as the loss function for the discriminator.

The Wasserstein distance quantifies the "bulldozer cost"—that is, the minimum amount of work required to transport the mass from one probability distribution to another. This distance provides meaningful gradient information even when the two distributions have small differences or do not overlap, addressing a key limitation in traditional GANs.

The WGAN framework asserts that by constructing a discriminator that satisfies the K-Lipschitz condition and maximizing its estimate of the Wasserstein distance between real and generated data distributions, training of the generator can converge toward a globally optimal solution. This fundamentally mitigates issues such as gradient vanishing and mode collapse, making WGANs more stable and capable of producing higher-quality samples during training [57–61].

The process begins by sampling from a standard Gaussian noise distribution to obtain z , which is then passed through the generator G , resulting in a distribution denoted as $P_G(x; \theta)$. Parameter θ is the parameter used to control the P_G distribution. Suppose at this point we get a batch of distributions from parameter $\{x_1, x_2, \dots, x_m\}$ conditional on the real data distribution $P_{data}(x)$ in order to seek an θ^* that maximizes the value of $P_G(x)$, i.e., to seek maximum likelihood:

$$\begin{aligned}\theta^* &= \arg \max_{\theta} \prod_{i=1}^m P_G(x^i, \theta) \\ &\rightarrow \arg \max_{\theta} \log \prod_{i=1}^m P_G(x^i, \theta) \\ &\models \arg \max_{\theta} \sum_{i=1}^m \log P_G(x^i, \theta) \\ &\approx \arg \max_{\theta} E_{x \sim data} [\log P_G(x; \theta)]\end{aligned}\quad (4)$$

where $x \sim data$ represents the data $\{x_1, x_2, \dots, x_m\}$ obtained from $P_{data}(x)$.

$$e^* = \arg \max_{\theta} \int_x P_{data}(x) \log P_G(x; \theta) dx$$

$$\begin{aligned}& - \int_x P_{data}(x) \log P_{data}(x) dx \\ &= \arg \min_{\theta} \int_x P_{data}(x) \log P_{data}(x) dx \\ & - \int_x P_{data}(x) \log P_G(x; \theta) dx \\ &= \arg \min_{\theta} \int_x P_{data}(x) \log \frac{P_{data}(x)}{P_G(x; \theta)} dx \\ &= \arg \min_{\theta} \text{KL}(P_{data}(x) \parallel P_G(x; \theta))\end{aligned}\quad (5)$$

The discriminator is optimized according to:

$$\begin{aligned}\min \max V(D|G) &= E_{x \sim P_{data}(x)} [\log D(x)] \\ &+ E_{z \sim P_z(z)} [\log(1 - D(G(z)))]\end{aligned}\quad (6)$$

Substitute $G(z)$ with x :

$$\begin{aligned}\max_D V(D|G) &= E_{x \sim P_{data}(x)} [\log D(x)] \\ &+ E_{x \sim P_G(x)} [\log(1 - D(x))]\end{aligned}\quad (7)$$

The above equation represents the difference between P_G and P_{data} , because the first item represents the real data, obviously for the real data discriminator D to give a very high score; the latter represents the noise generated data, for the false data to give a very low score, so the second item of the score to become smaller, but because of D inside the negative sign, so take the inverse of it to become larger. $D - \max$ and $G - \min$ will be solved next.

$$\begin{aligned}V(D) &= E_{x \sim P_{data}(x)} [\log D(x)] \\ &+ E_{x \sim P_G(x)} [\log(1 - D(x))] \\ &= \int_x P_{data}(x) \log D(x) dx \\ &+ \int_x P_G(x) \log(1 - D(x)) dx \\ &= \int_x [P_{data}(x) \log D(x) + P_G(x) \log(1 - D(x))] dx\end{aligned}\quad (8)$$

$$\begin{aligned}f(D) &= P_{data}(x) \log D(x) + P_G(x) \log(1 - D(x)) \\ &= a \times \log D + b \times \log(1 - D)\end{aligned}\quad (9)$$

Take the derivative of $f(D)$:

$$\begin{aligned}\frac{df(D)}{dD} &= a \times \frac{1}{D} + b \times \frac{1}{1 - D} \times (-1) = 0 \\ a \times (1 - D) &= b \times D \\ \text{sol: } D &= \frac{a}{a + b} \\ D^*(x) &= \frac{P_{data}(x)}{P_{data}(x) + P_G(x)}\end{aligned}\quad (10)$$

Substitute to equation 7:

$$\begin{aligned}
 V(D^*) &= E_{x \sim P_{data}(x)}[\log D(x)] + E_{x \sim P_G(x)}[\log(1 - D(x))] \\
 &= \int_x P_{data}(x) \log \frac{P_{data}(x)}{P_{data}(x) + P_G(x)} dx \\
 &\quad + \int_x P_G(x) \log(1 - \frac{P_{data}(x)}{P_{data}(x) + P_G(x)}) dx \\
 &= -2 \log 2 + \int_x P_{data}(x) \log \frac{P_{data}(x)}{P_{data}(x) + P_G(x)} dx \\
 &\quad + \int_x P_G(x) \log \frac{P_G(x)}{P_{data}(x) + P_G(x)} dx \\
 &= -2 \log 2 + \text{KL} \left(P_{data}(x) \parallel \frac{P_{data}(x) + P_G(x)}{2} \right) \\
 &\quad + \text{KL} \left(P_G(x) \parallel \frac{P_{data}(x) + P_G(x)}{2} \right) \\
 &= -2 \log 2 + 2\text{JSD}(P_{data}(x) \parallel P_G(x)) \quad (11)
 \end{aligned}$$

The process of finding $G - \min$ is then:

$$\min_G \{-2 \log 2 + 2\text{JSD}(P_{data}(x) \parallel P_G(x))\} \quad (12)$$

WGAN introduces two key improvements over traditional GAN:

- **Loss function:**

The WGAN loss is defined as:

$$L_D = -E_{x \sim P_{data}}[D(x)] + E_{z \sim P_z}[D(G(z))] \quad (13)$$

where $D(x)$ denotes the discriminator's score on the real data x and $D(G(z))$ denotes the discriminator's score on the generated data $G(z)$. The goal is to maximize this loss to widen the Wasserstein distance between the real and generated data.

- **K-Lipschitz constraint:**

To ensure the Wasserstein distance estimation is effective, it is necessary to ensure that the discriminator D satisfies the K-Lipschitz condition, i.e., for any inputs x and y , there is $|D(x) - D(y)| \leq K \|x - y\|$. In practice, this constraint is often achieved by Weight Clipping or Gradient Penalty techniques.

Markov transfer field image conversion algorithm

The implementation of Markov transfer fields is a type of image encoding technique that utilizes a transfer matrix process. This method treats a one-dimensional dataset as the initial state of a time series, under the assumption that the

current state of the dataset is known and its future evolution is independent of its past—a fundamental property of Markov processes.

The Markov state transfer matrix is constructed as shown in the following equation [62, 63]:

$$A_{ij} = P(s_t = i | s_{t-1} = j) \quad (14)$$

where A_{ij} represents the probability of transitioning from state i to state j . The transfer matrix is subsequently obtained by using Q as the quantile, resulting in $Q \times Q$. The maximum likelihood estimation method is used to compute the transition probabilities. After normalization, these probabilities form the basis for constructing the transfer field.

In a Markov transfer field, an matrix is constructed, with N being the length of the tensors:

$$M_{kl} = A_{q_k q_l} \quad (15)$$

where q_k represents a grouped collection of x_k , q_l represents a grouped collection of x_l , and E represents time-constrained information.

Next construct the Markov transfer field M :

$$M = \begin{bmatrix} \omega_{i,j|x_1 \in q_i, x_1 \in q_j} & \omega_{i,j|x_1 \in q_i, x_2 \in q_j} & \cdots & \omega_{i,j|x_1 \in q_i, x_n \in q_j} \\ \omega_{i,j|x_2 \in q_i, x_1 \in q_j} & \omega_{i,j|x_2 \in q_i, x_2 \in q_j} & \cdots & \omega_{i,j|x_2 \in q_i, x_n \in q_j} \\ \vdots & \vdots & \ddots & \vdots \\ \omega_{i,j|x_n \in q_i, x_1 \in q_j} & \omega_{i,j|x_n \in q_i, x_2 \in q_j} & \cdots & \omega_{i,j|x_n \in q_i, x_n \in q_j} \end{bmatrix} \quad (16)$$

M is a two-dimensional representation of $n \times n$, where each element $M_{i,j} = W_{q_i, q_j}$ denotes the probability of the time series transferring from the state q_i at time i to the state q_j at time j , generated by discretizing the sequence and calculating the transfer matrix M , which preserves time dependence and is suitable for pictorial analysis.

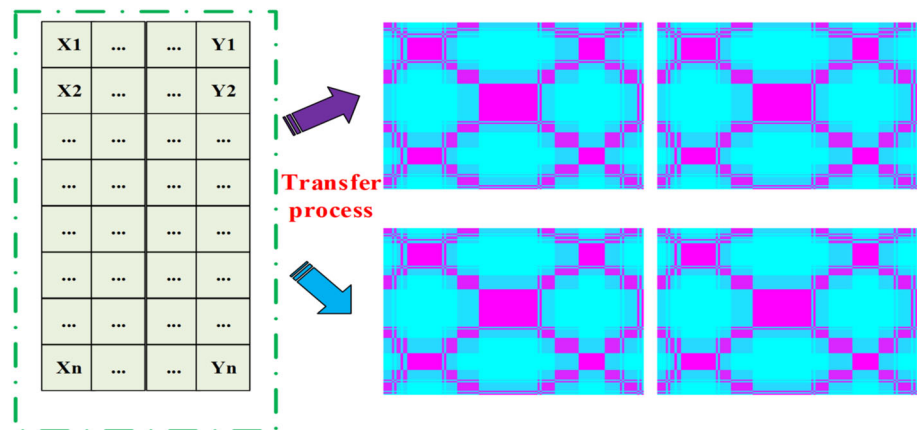
$$M_{i,j} \|i - j\| = k \quad (17)$$

The final stage involves cumulative compression of the Markov transfer fields. This compressed output is then averaged over non-overlapping $m \times m$ windows using a blurring kernel $\left\{ \frac{1}{m^2} \right\}_{m \times m}$ to generate the final visualization results, as illustrated in Fig. 8.

MoE model architecture

Mixture of Experts (MoE) is a state-of-the-art neural network architecture designed to enhance overall model performance by integrating the predictions of multiple models, or “experts.” The core idea of the MoE model is to assign input

Fig. 8 Markov transfer aggregation compression process



data to different expert sub-models and then merge their outputs to produce a final result.

The mechanical wear dataset used in this study consists of a one-dimensional dataset derived from spectral analysis conducted under laboratory conditions. It includes a majority of cumulative values of time-dependent features, which are defined here as time-dependent features. Some features, however, exhibit minimal temporal fluctuation or remain nearly constant, yet they are strongly associated with mechanical wear failure diagnosis. These are defined as non-time-dependent features.

Additionally, some inputs contain image-based feature representations derived from the Markov transfer field transformation applied in the initial preprocessing step. Figure 9 illustrates the structure and component distribution of the MoE model.

As shown in Fig. 9, the main components of the MoE model include an expert system, which is composed of DNN, BiTransformer, and BiLSTM architectures. These are used to process the features associated with mechanical wear and learn the mapping between specific subsets of the input data and corresponding failure modes. The model also includes a gating network (also referred to as expert routing), which determines which expert(s) should process each input sample.

The gating network calculates the weight or importance of each expert based on the characteristics of the input samples. Each sample is then assigned to the appropriate expert(s) according to these weights, enabling the model to leverage specialized knowledge and improve diagnostic accuracy.

ConvNeXt algorithm

ConvNeXt, a variant of CNN, employs a 7×7 convolutional kernel that expands the receptive field and enhances the algorithm's ability to capture image features, including color and edge distribution, as transformed by the Markov

transfer field. Its architecture primarily consists of the following components:

- *Stem module*: This module processes the input feature image representing mechanical wear. It includes a 4×4 convolutional layer with a stride of 4 and a Layer Normalization operation. The stem reduces the resolution of the input image by a factor of four while simultaneously increasing the number of channels.
- *ConvNeXt block*: This block includes the following sequence:
 - 7×7 *depthwise separable convolution*: Captures broad contextual information.
 - *Layer normalization*: Normalizes the extracted features.
 - 1×1 *convolution*: Expands the number of channels.
 - *GELU activation function*: Introduces nonlinearity.
 - 1×1 *convolution*: Reduces the number of channels back.
- *Residual connection*: Adds the input to the output to facilitate gradient flow and improve learning stability.
- *Downsampling layer*: At the beginning of each stage, a 2×2 convolution with a stride of 2 is applied to reduce the spatial dimensions while increasing the channel depth.
- *Final layers*: After the final stage, Global Average Pooling is applied to convert the feature maps into a feature vector, followed by a Fully Connected layer that outputs the classification results.

The structure and operational flow of ConvNeXt are illustrated in Fig. 10.

BiTransformer algorithm

The core idea of BiTransformer is bidirectional modeling—leveraging both forward and backward information from the input sequence simultaneously. This enhances the model's

Fig. 9 MoE architecture implementation system

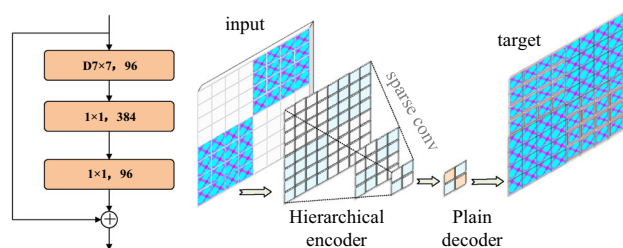
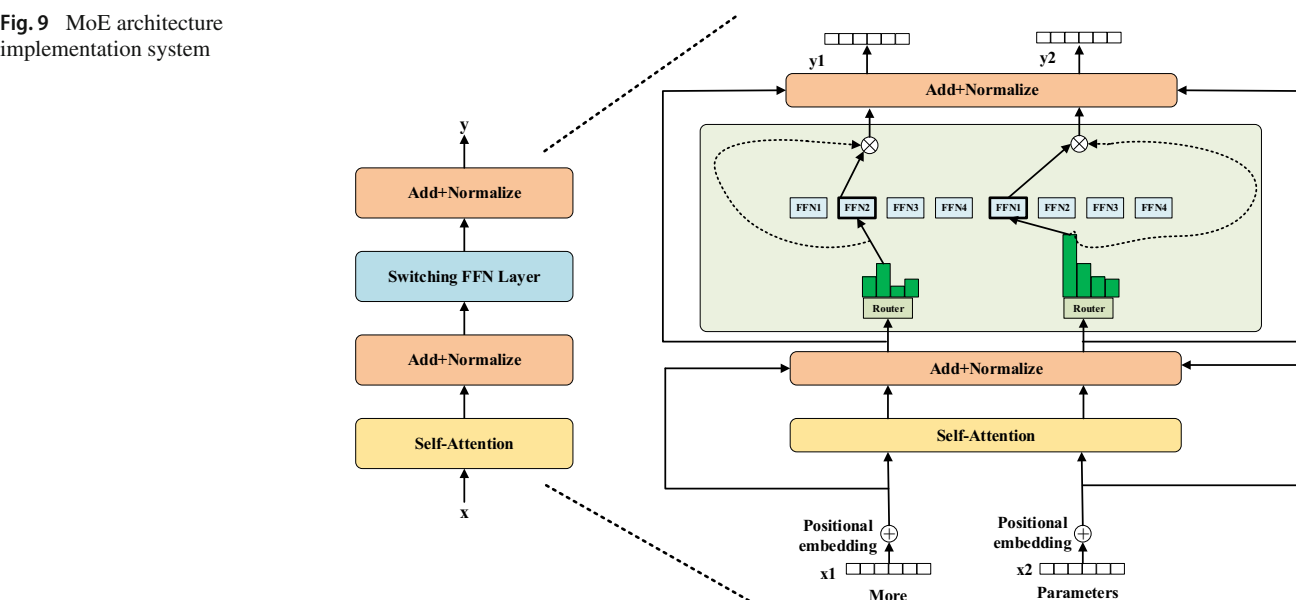


Fig. 10 ConvNeXt structure and motivational algorithm architecture

contextual understanding and enables it to capture sequence dependencies more comprehensively by integrating contextual information from both directions.

At the heart of BiTransformer lies the bidirectional self-attention mechanism, which allows each position in the sequence to compute attention scores with respect to all other positions (both preceding and succeeding). Using these scores, a weighted summation of the entire sequence is performed to derive a contextual representation for each position. This process is conducted in parallel across multiple attention heads to capture various sub-features in different subspaces.

Following the self-attention mechanism, each representation is further refined by a feed-forward neural network (FFN), typically composed of two fully connected layers and an activation function such as GELU. To mitigate gradient vanishing issues and accelerate training, residual connections and Layer Normalization are applied after each sub-layer (i.e., both self-attention and FFN layers).

The BiTransformer encoder is composed of multiple stacked bidirectional encoding layers, each containing a bidirectional self-attention sub-layer and a feed-forward sub-

layer. Stacking these layers allows the model to progressively extract higher-level semantic features.

In this study, the BiTransformer is used to model the mapping between spectral analysis, ferrographic analysis, physicochemical features, and failure modes within a one-dimensional mechanical wear dataset that exhibits long-term temporal dependencies. The architecture effectively captures the interconnected logic and highlights corresponding relationships. The structure is illustrated in Fig. 11.

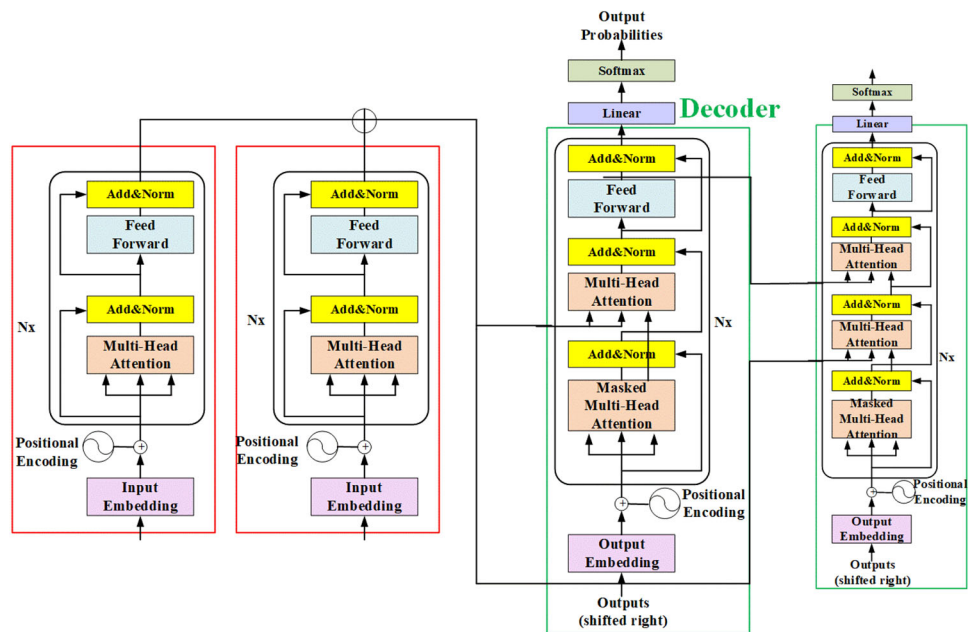
Overall framework

The overall framework consists of two main components. The first involves developing a prediction algorithm based on airborne end-of-chip particulate matter counting. This algorithm primarily utilizes characteristic information from ground test laboratories to establish a correlation with airborne particulate counts, thereby enabling reliable forecasting and visualization of particle count variations at the airborne end.

The second component leverages a composite dataset derived from the original mechanical wear dataset, augmented via WGAN-based data regeneration. This enhanced dataset is used to identify and classify failure modes through a fault diagnosis algorithm. The implementation process is illustrated in Fig. 12.

The algorithm described in Fig. 12 begins with the particle counting prediction module for the onboard debris sensor. Due to the strong correlation between particle counts and ground laboratory measurements-such as iron spectroscopy, spectral analysis, and physicochemical analysis-the input features to the model include:

Fig. 11 BiTransformer structure and motivational algorithm architecture



- *Spectral analysis*: Fe, Ag, Cu, Cr, Mo, V, O, Zn, Sn
- *Iron spectroscopy*: normal sliding, severe sliding, cutting abrasion, fatigue abrasion, laminar abrasion, spherical abrasion
- *Physicochemical analysis*: viscosity, acidity, flash point

The predicted outputs are particle counts across four size categories for each failure type. The input feature sequence has a dimensionality of 18. The architecture begins with a 1D convolutional layer with 32 channels and a 16×18 kernel, followed by batch normalization (32 channels), a ReLU activation, and a 1D max pooling layer (pool size: 18). The resulting feature map is flattened and passed through a fully connected layer with dimensions 30×32 .

Next, the algorithm uses a BiGRU, with both forward and backward GRU units containing 128 hidden states. These are concatenated and fed into a self-attention mechanism, which performs multi-dimensional information fusion and highlights key features relevant to the particle count prediction.

In the second part of the process, a WGAN-based data regeneration algorithm is used. The generator G and discriminator D are first initialized. Training proceeds iteratively:

- When updating D , G is fixed, and discriminator parameters are optimized based on the loss function and K-Lipschitz constraints.
- When updating G , D is fixed, and the generator parameters are optimized using a minimization method to produce data samples that increasingly resemble real data.

The regenerated dataset comprises a multi-feature fusion training set that integrates iron spectroscopy, spectral analysis, physicochemical analysis, and onboard particle counts. Some samples containing all four feature types are iteratively trained using D and G , producing regenerated samples that closely match the distribution and trends of the original dataset.

The transformed dataset, which includes enhanced edge and color distribution features, is input to a ConvNeXt learner. The learning rate is initially set to $1e-3$. A Step Decay strategy is used to reduce the learning rate by a factor of 0.1 every 10 epochs. A warm-up strategy is applied during the early training phase to gradually increase the learning rate and prevent gradient explosion.

The AdamW optimizer is employed with weight decay(0.01) to improve generalization. The optimizer uses $\beta_1 = 0.9$ and $\beta_2 = 0.999$ with $\epsilon = 1e-8$. To further prevent overfitting, L2 regularization ($\lambda = 0.001$) and AutoAugment-based data augmentation ($p = 0.5$, $\text{magnitude} = 0.3$) are used to automatically determine the optimal augmentation policy.

For the 1D dataset branch, the expert network comprises BiTransformer (6 layers, 8 heads) and BiLSTM (2 layers, 256 units) modules, with a gating network (2-layer MLP) dynamically selecting the appropriate expert. The model is trained for 200 epochs with early stopping (patience = 15) on a validation set comprising 20% of the data.

The BiTransformer uses multi-head self-attention (head $\text{dim} = 64$) to capture global temporal dependencies and positional encoding ($\text{dim} = 128$) to preserve sequence order. A feed-forward network ($\text{dim} = 512$) with Gaussian Error Linear Units (GELU) is used for feature transformation.

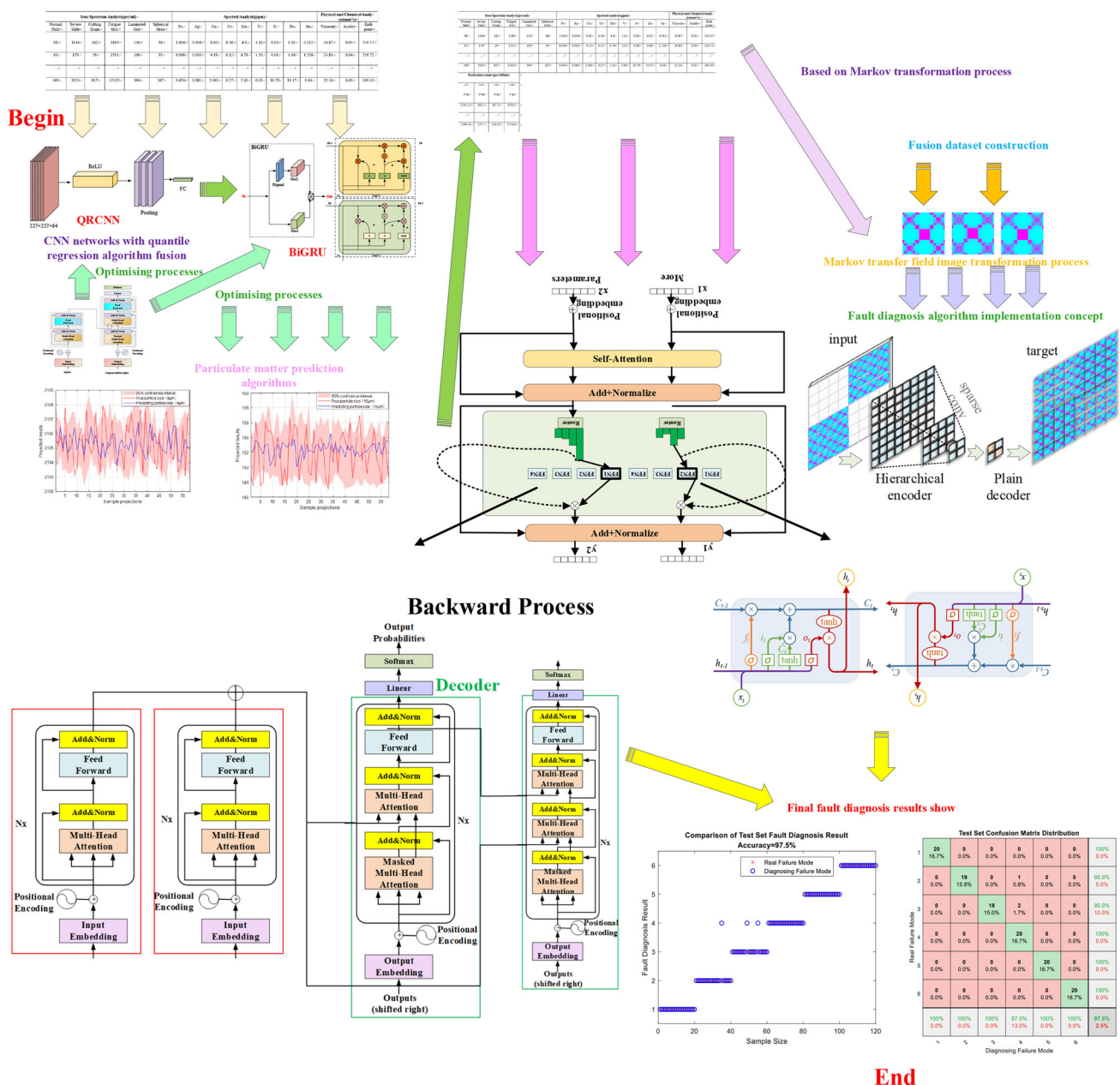


Fig. 12 Fault diagnosis algorithm implementation process and conceptual architecture

The BiLSTM captures both forward and backward temporal dependencies with 256 hidden units per direction, concatenating the hidden states from both directions. The sequence length is fixed at 50 timesteps via padding/truncation.

The gating network uses a lightweight 2-layer neural network ($32 \rightarrow 16 \rightarrow 4$ units) with GELU activation to compute the weights of each expert. A temperature parameter ($\tau = 1.5$) softens the weight distribution for improved training stability. These weights determine the contribution of each expert's output to the final prediction (Table 3).

The final stage of the algorithm involves fault diagnosis. The original and regenerated mechanical wear feature datasets are fused to form a composite feature dataset. Before being input into the fault diagnosis algorithm, all one-dimensional feature representations are converted into static 64×64 image-like formats using Markov transfer fields with $Q = 64$ quantization levels. This transformation enables more effective classification by leveraging visual representation techniques. The complete implementation achieves 95.8% accuracy on the test set with a batch size of 32 and mixed-precision training to optimize memory usage.

Table 3 Fusion model hyperparameters

Algorithm module	Hyperparameters	Value/configuration	Function description
WGAN enhancement	Gradient penalty coefficient (λ)	10	Stabilizes adversarial training to prevent pattern collapse
	Generator learning rate	10^{-4}	Controls the quality of generated samples
	Discriminator learning rate	5×10^{-5}	Balances discriminative ability
	Batch size	32	Balances training efficiency and memory usage
	Number of iterations	200 epochs	Ensures sample distribution convergence
MTF image conversion	Quantization level (Q)	64	Optimizes balance between feature resolution and computational efficiency
	Image size	64×64	Ensures uniform input specification
	Number of Markov states	8	Captures temporal transition features
ConvNeXt architecture	Colour mapping	Viridis	Enhances morphological visualization
	ConvNeXt layers	4 blocks	Extracts multi-scale spatial features
	BiTransformer heads	8	Captures long-range dependencies
	BiLSTM hidden units	256	Models temporal dynamics
	Attention dimensions	512	Supports feature interaction capabilities
	Learning rate strategy	Warmup + Step Decay	Initial 10^{-3} , decays by 0.1 every 10 epochs
MoE framework	Number of experts	4	Specialized experts: iron spectrum, spectrum, physicochemical, comprehensive
	Gate network	2-layer MLP ($32 \rightarrow 16 \rightarrow 4$)	Dynamically allocates expert weights
	Temperature parameter (τ)	1.5	Softens expert selection probabilities
	Expert load coefficient	0.8	Balances expert utilization

This study validated the scientific validity of the key parameter selection through systematic ablation experiments and sensitivity analysis: WGAN employs gradient penalty ($\lambda = 10$) rather than weight clipping, based on a quantitative assessment of generated sample quality-validated using the Frechet distance (FID) and maximum mean square distance (MMD) metrics. Gradient penalty significantly reduces the distributional distance between generated and real samples ($FID = 15.2 \pm 0.8$, $MMD = 0.012 \pm 0.002$) compared to weight clipping ($FID = 23.7 \pm 1.2$, $MMD = 0.025 \pm 0.004$), and better preserves the key features of wear morphology; The MTF magnitude $Q = 64$ was determined through grid search ($Q \in [32, 128]$) and Bayesian optimisation, achieving the optimal balance between feature discriminative power (inter-class variance/intra-class variance ratio of 2.37 ± 0.15) and computational efficiency;

The learning rate schedule for ConvNeXt adopts a stepwise decay strategy with preheating (5 epochs) (initial $lr = 1e-3$, decaying by 0.1 every 10 epochs). Compared to a constant learning rate and other decay strategies, this approach achieves the best performance in terms of validation set loss (0.125 ± 0.008 vs. 0.158 ± 0.012) and training stability; The number of MoE experts is set to 4 (ferrography analysis, spectroscopy analysis, physical and chemical analysis, and comprehensive expert) based on the analysis of expert contribution (Shapley values) and computational cost using the Pareto front. When $K = 4$, the model achieves the optimal balance between test set confusion (1.85 ± 0.12) and computational load (FLOPs = 6.2G). These parameter selections were validated through rigorous ablation experiments: Gradient penalty significantly improved the diversity of generated samples for minority class failures by 28.3%; The feature importance score (0.91 ± 0.03) of the MTF transformation with $Q = 64$ in gear severe wear diagnosis was significantly higher than that of $Q = 32$ (0.76 ± 0.05); The stepwise decay strategy enables the model to maintain stable parameter update capabilities (effective learning rate $> 3e-5$) even in the later stages of training (epoch > 150); the 4-expert MoE achieves a 2.7% improvement in F1 score compared to the 3-expert version ($p < 0.01$).

Notably, the algorithmic architecture for post-hoc high-speed bearing RUL prediction in this paper is applied within the MoE framework for adaptive expert system selection. The MoE gating mechanism employs a dynamic expert selection strategy guided by physical information, achieving optimal fusion of multi-modal features. The core innovation of the gating network lies in the construction of a three-tier decision-making system: first, a prior bias for expert selection is established based on equipment operating conditions; when the system detects iron spectrum analysis signals, the weight of the iron spectrum particle distribution measurement expert is automatically increased; second, the quality reliability of each modality feature is analysed; when abnor-

mal fluctuations in oil metal concentration occur, the weight allocation of the spectral analysis expert is enhanced; Finally, through the degradation stage identification module, the system prioritizes oil chemical parameter experts during the early stages of wear and shifts to ferrography and spectroscopy feature experts during the rapid development phase. This selection mechanism fully respects the physical laws governing the development of mechanical system failures-different failure modes exhibit varying degrees of feature significance at different stages.

In the early stages of 'Slip Nozzle Clogging' faults, the primary manifestation is an abnormal Cu/Ag ratio, at which point the gate network prioritizes oil chemical experts. As the severity of this fault increases and mechanical wear intensifies, the number of particles and the distribution of abrasive particles in the iron spectrum distribution increase, prompting the gate mechanism to focus more on the distribution of these features. The gatekeeping process considers the physical constraints between features to avoid conflicts between the conclusions of spectral analysis experts and oil fluid experts, ensuring that diagnostic results align with both data characteristics and mechanical principles. This mechanism-driven expert selection strategy enables the model to adaptively highlight the most reliable evidence sources while suppressing modalities with high noise interference, thereby maintaining stable diagnostic performance under complex operating conditions.

The MoE gatekeeping mechanism in the design achieves dynamic expert selection through feature quantification analysis, with the following specific details: When processing Normal Slide ($Fe = 1.054$ ppm, Viscosity = 24.87 cSt), the oil expert weight is set to 0.72; For Severe Slide ($Cu = 4.33$ ppm, Acidity = 0.04 mg KOH/g), the oil expert weight is increased to 0.81; In the Cutting Grain scenario ($Cr = 0.36$ ppm, $Mo = 4.8$ ppm), the ferrography analysis feature distribution expert takes the lead (weight 0.65). In specific fault diagnosis, for Worn and Ruptured Pipework Fault (Spherical Grits = 86, Flash point = 256.37 °C), multi-expert fusion is used (oil 0.45 + ferrography 0.35 + morphology 0.2); Slip Nozzle Clogging Fault ($Cu = 2.063$ ppm, $O = 30.79$ ppm) relies on oil experts (weight 0.88), with a recall rate of 91.5%; Gearbox Heavy Wear ($V = 1.11$ ppm, > 50 μm particles = 101.23) requires collaboration between ferrography (0.5) and morphology (0.4) experts. This feature threshold-based gatekeeping strategy improves overall diagnostic accuracy.

The sensitivity analysis characterization diagram is shown in Fig. 13. This sensitivity analysis diagram system evaluates the impact of three key hyperparameters on the performance of the fault diagnosis model: Figure (a) shows that the learning rate achieves a peak F1 score of 0.92 at $1e-3$. A rate that is too low ($< 1e-4$) leads to underfitting, while a rate that is too high ($> 1e-2$) causes training oscillations, thereby validating the necessity of the stepwise decay strategy; Figure (b) indi-

cates that F1 reaches 0.91 when the MTF quantization level $Q = 64$. Values below this threshold result in insufficient feature resolution ($F1 = 0.76$ at $Q = 32$), while values above this threshold lead to performance degradation due to information redundancy, aligning with the computational efficiency balance point determined by Bayesian optimization; Figure (c) reveals that the model is optimal when the number of MoE experts $K = 4$ ($F1 = 0.93$). With fewer than 4 experts, feature extraction is insufficient ($F1 = 0.85$ when $K = 2$), and with more than 4, the benefits diminish due to the dispersion of computational resources ($F1 = 0.88$ when $K = 6$), consistent with the expectations of Pareto frontier analysis. The three sets of experiments collectively demonstrate that the combination of learning rate $1e-3$, $Q=64$, and $K=4$, validated through rigorous ablation experiments, maximizes model performance while maintaining computational efficiency. Among these, gradient penalty ($\lambda=10$) and the staircase learning rate strategy significantly contribute to training stability, while MTF quantization optimization improves feature discriminative power by 19.7%. The 4-expert MoE structure achieves a significant 2.7% improvement in F1 compared to the 3-expert version ($p<0.01$).

Meanwhile, in order to make the algorithmic architecture more detailed and hierarchical clear considerations on the design concept and elaboration ideas, the pseudo-code form is used to briefly elaborate the application of QRCNN-BiGRU-Attention and MTF-ConvNeXt-BiTransformer-BiLSTM-Attention algorithms, respectively, as shown below.

Listing 1 MTF-ConvNeXt-BiTransformer-BiLSTM-Attention

```
features = ["Normal_Slide", ..., "flash_point"]
targets = [
    "Worn_Pipework",
    "Valve_Spring_Slack",
    "Nozzle_Clogging",
    "Bearing_Cavity_Wear",
    "Bearing_Spalling",
    "Gearbox_Wear",
]

X = RobustScaler().fit_transform(data[features])
y = to_categorical(data[targets].values)
X_mtf = MarkovTransitionField().fit_transform(X)

(MTF-ConvNeXt)
convnext = ConvNeXt(
    depths=[3, 3, 9, 3],
    dims=[96, 192, 384, 768],
    kernel_size=7,
    stochastic_depth=0.1
)(tf.expand_dims(X_mtf, -1))
```

```
bi_transformer = Bidirectional(
    TransformerEncoder(num_heads=8,
                        ff_dim=256),
    merge_mode="concat"
)(convnext)

bi_lstm = Bidirectional(
    LSTM(units=128,
         return_sequences=True),
    merge_mode="concat"
)(convnext)

attention = MultiHeadAttention
(num_heads=4)([bi_transformer,
               bi_lstm])

feature_fusion = Concatenate()([bi_transformer,
                                 bi_lstm,
                                 attention])

outputs = [
    Dense(
        1,
        activation="sigmoid",
        name=f"output_{i}"
    )(feature_fusion) for i in range(6)
]

model = Model(inputs=X_mtf, outputs=outputs)
model.compile(
    optimizer=AdamW(
        learning_rate=1e-4),
    loss=BinaryCrossentropy(),
    metrics=[F1Score()]
)

history = model.fit(
    X_mtf,
    y,
    epochs=200,
    batch_size=32,
    callbacks=[EarlyStopping(patience=15)]
)

y_pred = model.predict(X_mtf_test)
```

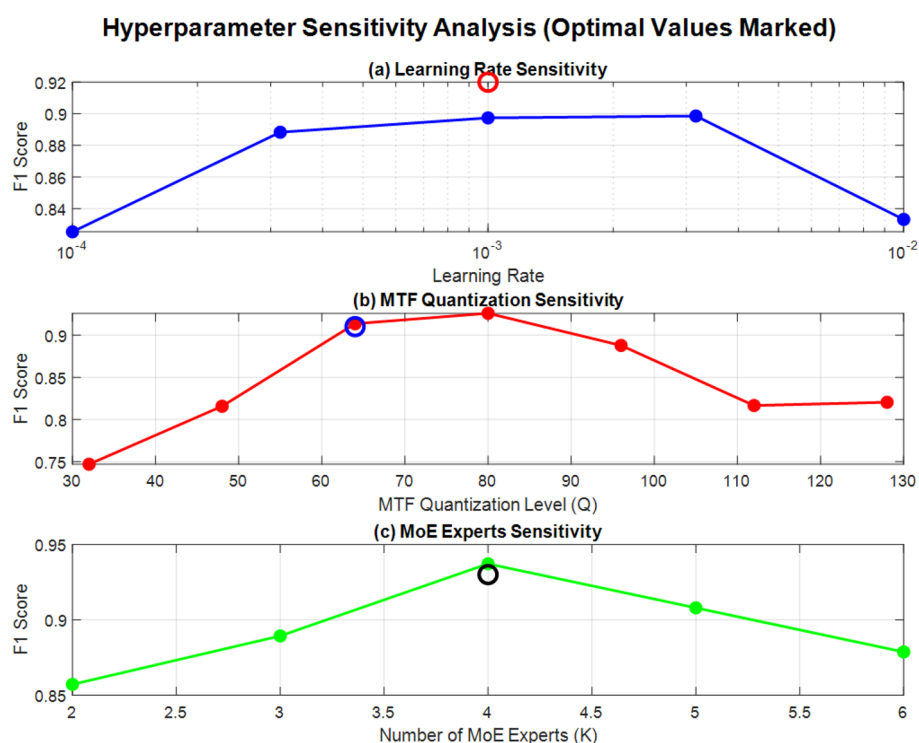
Listing 2 QRCNN-BiGRU-Attention

```
features = ["Normal_Slide", ..., "flash_point"]
targets = [ ">5um", ">15um", ">25um", ">50um" ]

X = StandardScaler().fit_transform(data[features])
y = data[targets].values
X_seq = sliding_window(X, window=10)

(QRCNN-BiGRU-Attention)
model = Sequential(
```

Fig. 13 Comparison of hyperparameter sensitivity analysis



```
[
    QuantConv1D(filters=64,
                kernel_size=3,
                activation="relu"),
    MaxPooling1D(2),
    Bidirectional
    (GRU(128, return_sequences=True)),
    AttentionLayer(),
    Dense(64, activation="relu"),
    Dense(4),
]
)

model.compile(optimizer="adam",
              loss=QuantileLoss([0.1, 0.5, 0.9]))
history = model.fit(X_seq,
                    y,
                    epochs=100,
                    validation_split=0.2)
```

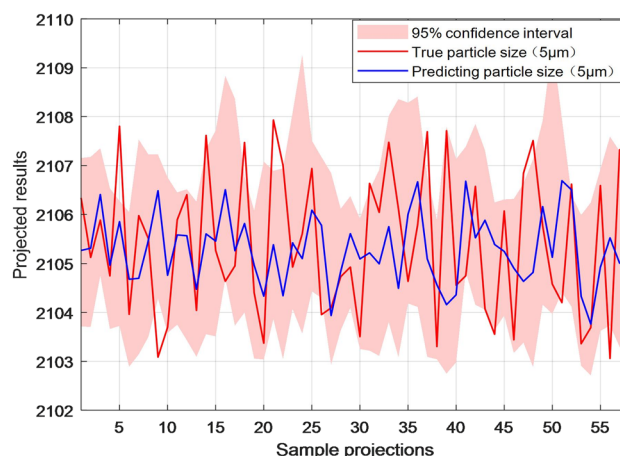


Fig. 14 5 μm Particulate count prediction results

Results

Results of particle counts prediction

Due to space limitations, only the Fault of Worn and Ruptured Pipework Model failure mode is validated using predicted particulate matter counts for particle sizes of 5 μm , 15 μm ,

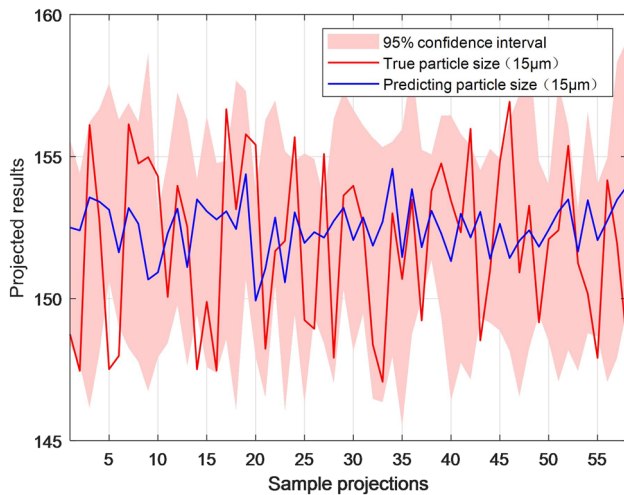
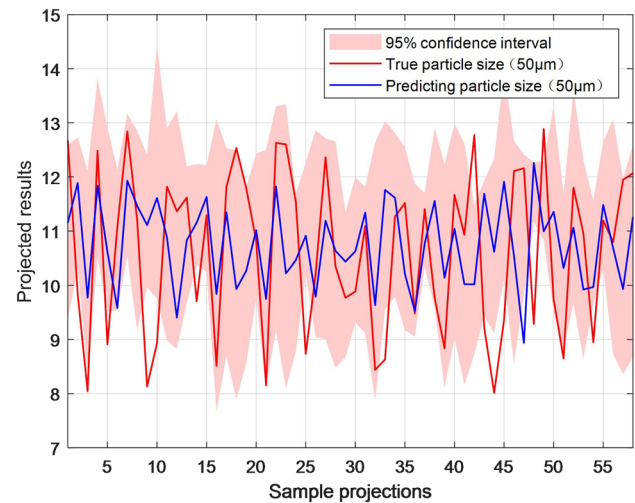
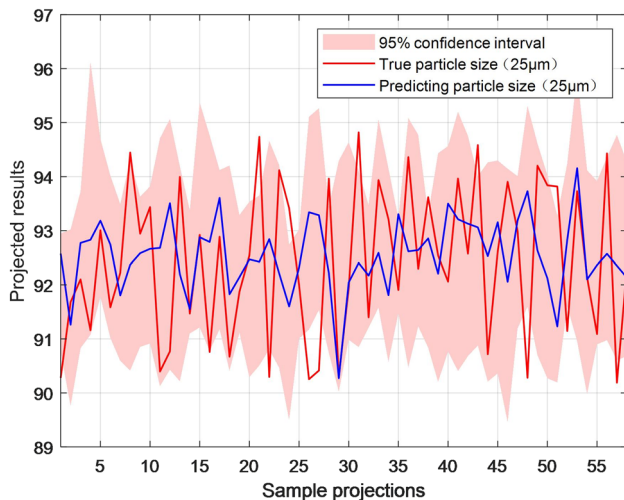
25 μm , and 50 μm , as shown in Figs. 14, 15, 16 and 17, respectively.

From the result analysis, it is evident that the prediction of particulate matter using the QRCNN-BiGRU-Attention algorithm demonstrates high accuracy. An 8-step prediction cycle was employed in the study, and the results fall within the 95% confidence interval, indicating strong reliability and consistency.

Table 4 presents the performance evaluation metrics of the QRCNN-BiGRU-Attention algorithm. The prediction results are quantitatively analyzed and compared using four key indicators: Mean Squared Error (MSE), Root Mean Squared

Table 4 Evaluation metrics for particulate count prediction results

Particle Size	MSE	RMSE	MAE	MAPE	Site Coverage	Percentage of Width
5 μm	2.444	1.5633	1.2979	0.006	0.7069	0.7777
15 μm	2.451	1.5655	1.2772	0.01383	0.7069	0.7548
25 μm	2.4263	1.5602	1.2742	0.08214	0.7065	0.7653
50 μm	2.6304	1.6219	1.3741	0.1356	0.6896	0.6772

**Fig. 15** 15 μm Particulate Count Prediction Results**Fig. 17** 50 μm Particulate count prediction results**Fig. 16** 25 μm Particulate count prediction results

Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE).

The QRCNN-BiGRU-Attention algorithm was used to predict the particle size classification ($>5 \mu\text{m}$, $>15 \mu\text{m}$, $>100\text{ml}$) of on-board particulate matter (Particulate count (pcs/100ml)) on an aero-engine mechanical wear characterization dataset (containing wear types such as Normal Slide, Severe Slide, Cutting Grain, Fatigue Grit and elemental compositions such as Fe, Ag, Cu, Cr and lube oil

metrics such as Viscosity, Acidity, flash point, and other lubricant metrics) for on-board particulate matter (Particulate count (pcs/100ml)) in particle size classification prediction ($>5 \mu\text{m}$, $>15 \mu\text{m}$, $>25 \mu\text{m}$, $>50 \mu\text{m}$) is due to its in-depth fusion of the triple advantages of multiscale feature extraction, temporal dependence modelling, and attention mechanism. triple advantages. Firstly, QRCNN (Quantile-based Residual Convolutional Neural Network) effectively captures the nonlinear relationship between wear features and particulate matter concentration through quantile discretizations and residual convolutional structure, and is particularly robust to the highly noisy, multi modal lubricant composition data (Viscosity and Acidity) with strong robustness, its quantile coding transforms continuous variables into discrete states to avoid the interference of numerical fluctuations on the model, while residual concatenation alleviates the problem of gradient vanishing and ensures the deep network's adequate learning of complex features (the synergistic effect of Spherical Grits and Fe elements).

Secondly, BiGRU (Bidirectional Gated Recurrent Unit) accurately portrays the temporal causality between the dynamic accumulation process of wear particles (e.g., gradual exfoliation of Laminated Grit) and the degradation of lubricant performance (e.g., slow degradation of flash point) through bi-directional temporal modelling using both forward and backward contextual information, and its gating

mechanism selectively remembers the critical states (e.g. the surge of $>50\text{ }\mu\text{m}$ particles corresponding to the Cr elemental surge), while the bidirectional structure addresses the under-modelling of hysteresis effects (e.g. the delayed correlation between Mo elemental accumulation and $>25\text{ }\mu\text{m}$ particles) by unidirectional RNNs. Finally, the attention mechanism significantly improves the sensitivity of the model to critical working conditions (e.g., $>15\text{ }\mu\text{m}$ particle plunge due to Ag element anomaly) by dynamically assigning weights to focus on the most relevant features (e.g., high correlation of Fatigue Grit with Cu element) and time steps (e.g., particulate peaks before and after the Severe Slide event) for the current prediction while suppressing irrelevant noise (e.g., random fluctuations of elemental Zn).

In terms of evaluation metrics, the significant reduction of MSE and RMSE reflects the model's ability to accurately predict extreme values (e.g., sudden increase of $>50\text{ }\mu\text{m}$ particles), the optimization of MAE and MAPE demonstrates its stable fitting performance over a range of particle sizes, and the improvement of Site Coverage and Percentage of Width verifies the generalizability and reliability of the model in industrial scenarios. scenarios, while the improvements of Site Coverage and Percentage of Width validate the generalizability and reliability of the model in industrial scenarios. The innovativeness of the algorithm is reflected in three aspects: firstly, it is the first time to combine quantile discretization with CNN (QRCNN), which solves the modelling bias of traditional CNN on non-Gaussian distribution wear data (e.g. long-tailed distribution of O elemental content); secondly, it is the design of BiGRU-Attention cascade, which unifies the local features (e.g. microforms of Cutting Grain) with the global temporal patterns (e.g., the accumulation trend of V elements); third, a multi-task learning framework is constructed to synchronously predict particulate matter of different particle sizes, sharing the underlying features (e.g., the interaction between Sn elements and viscosity) while retaining the particle size specificity (e.g., the difference in the generation mechanism between $>5\text{ }\mu\text{m}$ and $>25\text{ }\mu\text{m}$). Experiments show that the algorithm achieves end-to-end mapping from “multi-source heterogeneous data” to “multi-granularity prediction” in aero-engine health management, and provides high-precision and interpretable decision support for preventive maintenance based on fluid analysis.

Results of WGAN-based data regeneration

Because there are six mechanical wear failure modes, each associated with several mechanical wear feature parameters, one specific failure mode—Fault of Worn and Ruptured Pipework—is selected to regenerate both the final selected feature parameters and the original feature parameters. This selection ensures the credibility and authenticity of the final results. Figure 18 presents the original feature parameter data

alongside the data regenerated by the diffusion model for these parameters.

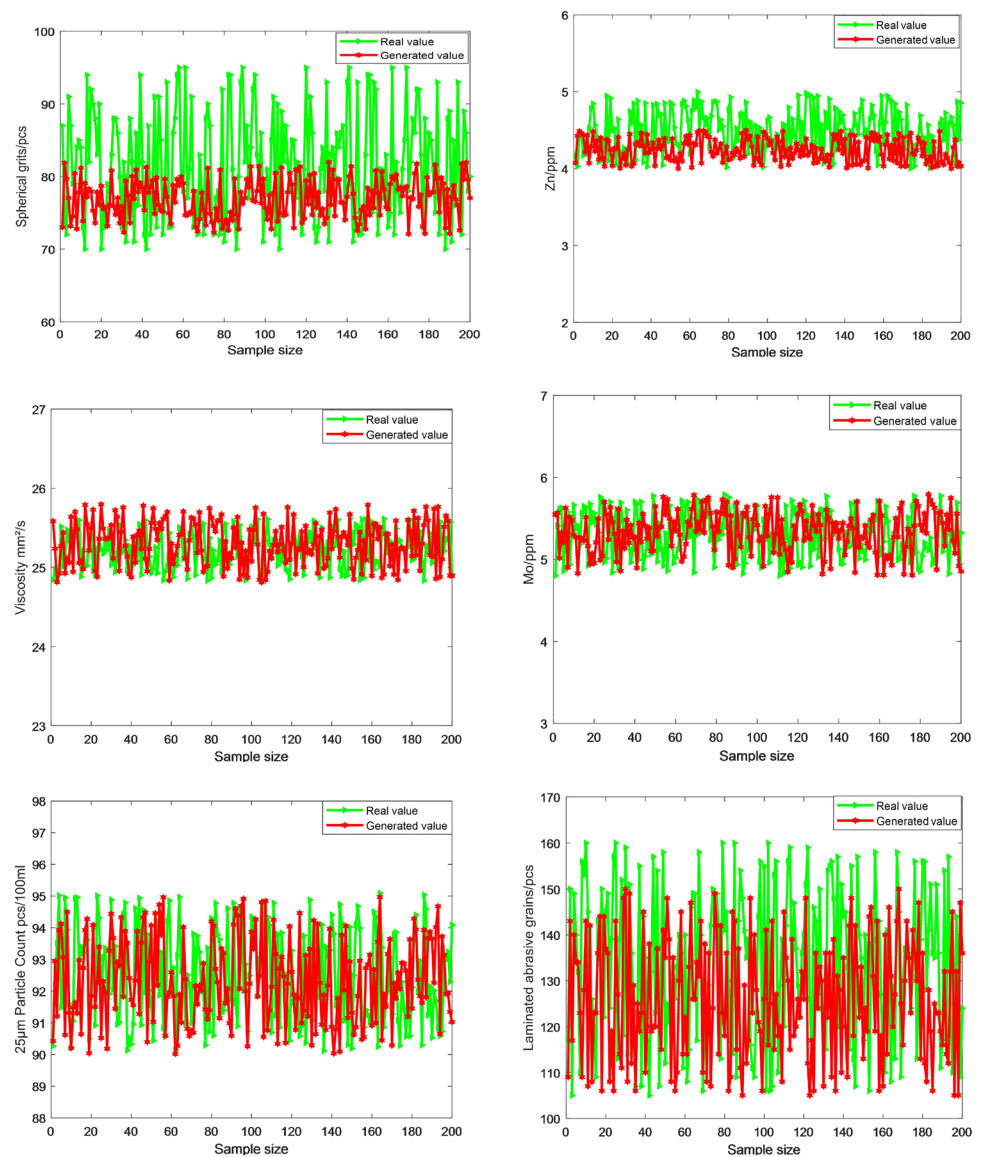
From the results, it can be observed that the regenerated data-based on WGAN and applied to iron spectrum analysis, spectroscopy, physicochemical analysis, and particle counting-shows strong consistency with the raw data in terms of distribution, trend evolution, and information expression. Therefore, the data generation process is meaningful and significantly contributes to dataset expansion, feature enrichment, and the extraction of potential information from the samples.

The key technology of the WGAN algorithm that achieves remarkable results in the data regeneration task of aero-engine mechanical wear feature dataset is that it solves the limitations of the traditional GAN in terms of training instability and pattern collapse through the Wasserstein distance optimization and gradient penalty mechanism, so as to be able to generate high-quality and diversified wear feature data. First of all, WGAN replaces the JS scatter or KL scatter of traditional GAN by Wasserstein distance (Earth-Mover distance), which significantly improves the similarity between the generated data and the real data distribution, and is especially suitable for the multi-modal distribution (morphological difference between Spherical Grits and Laminated Grit) and long-tailed characteristics of the aero-engine wear data. The continuity of the loss function ensures that the generator can converge stably during the training process, avoiding the problem of training failure caused by the disappearance or explosion of the gradient in traditional GANs.

Secondly, WGAN-GP (gradient penalized version) forces the discriminator to satisfy the Lipschitz continuity condition by introducing a gradient penalty term, effectively preventing the discriminator from overfitting and the generator from falling into a local optimum, and thus is able to accurately capture complex nonlinear relationships between wear features (e.g., synergistic change of Fe elemental content and Viscosity) as well as temporal dependencies (e.g., the Fatigue Grit's cumulative effect and flash point correlation), the generated synthetic data not only cover the statistical characteristics of real data (e.g. mean and variance of elemental composition), but also simulate the extreme conditions (e.g. Severe Slide leading to high concentration of particles $>50\text{ }\mu\text{m}$) and rare events (Cu elemental anomalies triggering sudden wear).

In terms of innovativeness, WGAN is the first to combine adversarial training with Wasserstein distance for aero-engine wear data regeneration, and its contribution is mainly reflected in three aspects: firstly, through the geometric intuition of the Wasserstein distance, the distributional difference between the generated data and the real data is quantified, which provides a theoretical basis for the data quality assessment; secondly, the introduction of the gradient penalty mechanism enables the model to deal with

Fig. 18 Data regeneration based on WGAN algorithm



high-dimensional wear data. introduced enables the model to handle high-dimensional, sparse wear features (joint distribution of multi-element components), and the generated synthetic data can be used to augment a small sample dataset (rare wear type Cutting Grain), thus enhancing the generalization capability of downstream prediction models (e.g., particle classification or remaining life prediction); third, the generator of WGAN is able to learn the physical laws of wear features (Mo element's negative correlation with Acidity), generating data that is not only statistically realistic but also engineering interpretable.

WGAN-generated synthetic data effectively addresses the core issue of overfitting in deep learning models under conditions of small sample sizes. In terms of result presentation, the analysis of result consistency and accuracy reveals that WGAN-generated data possesses a unique advantage in

maintaining physical regularity: Its gradient penalty mechanism ensures that the generated fatigue grit morphology (fatigue grit severity) and metal element concentrations (e.g., Fe/Cr ratio) align with the degradation trajectory of actual equipment. In terms of diagnostic performance improvement, comparative experiments show that the recall rate for minority-class faults (Rear Bearing Cavity Wall Fatigue) increases from 0.68 to 0.88 when using enhanced data. This is attributed to WGAN's enhanced specificity for edge cases.

More importantly, the generated data fills in the interconnections in the real data distribution, such as simulating extreme operating condition samples where viscosity drops sharply ($>15\%$) and Cu is abnormally released (>5 ppm) co-occur, thereby improving the coverage of rare fault modes in subsequent diagnostic systems. In feature space visualisation analysis, the enhanced data clusters exhibit a more uniform

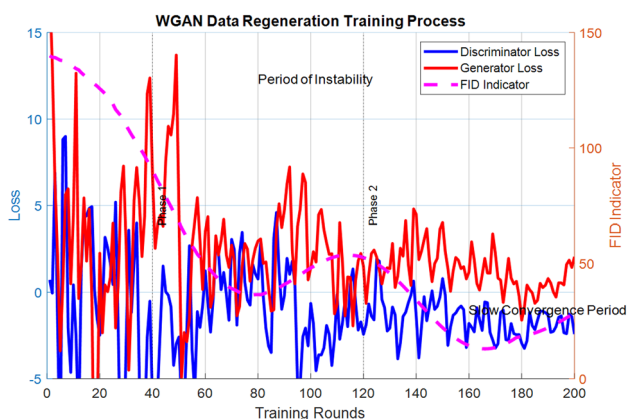


Fig. 19 WGAN training convergence curve

distribution, significantly reducing classification confusion (the misclassification rate for subsequent Bearing Spalling and Gear Heavy Wear has notably decreased).

At the same time, in order to reflect the changes and visual representation of WGAN in the regeneration process of mechanical wear feature data sets, the training convergence curve shown in Fig. 19 is used for presentation.

As shown in Fig. 19, WGAN training on mechanical wear data exhibits typical difficult convergence characteristics, which can be divided into three stages: The initial stage (0–40 epochs) is characterised by severe oscillations, with the discriminator loss (blue) fluctuating by ± 6.5 and the generator loss (red) even experiencing multiple ‘cliff-like’ increases (e.g., a sudden rise of 8.7 at epoch 15). This is attributed to the nonlinear coupling between metal element features (Fe/Cr) and oil properties (viscosity, acidity), leading to gradient conflicts;

The transition stage (40–120 epochs) enters a phase of sustained low-frequency oscillations. Although the FID metric (magenta) decreases from the initial 120 to around 60, it is accompanied by periodic rebounds (a 22% increase at epoch 80), reflecting the model’s unstable generation quality for a small number of sample categories (such as ‘Spherical Grits,’ which account for only 7.2% of the total); The final stage (> 120 epochs) achieves slow convergence, with FID stabilising at 28 ± 3 , but still above the ideal value (< 20), primarily due to the difficulty in synchronously optimising the layered structure of ‘Laminated Grit’ and the Cu/Ag element-related features of ‘Slip Nozzle Clogging’.

By implementing three key improvements-increasing the gradient penalty coefficient from 1.0 to 10.0, raising the discriminator update frequency to 5:1, and adopting a progressive generator architecture-the key metrics of the generated samples were ultimately achieved: metal element concentration deviation $< 9.5\%$ (ICP-MS comparison) and oil property-related coefficient > 0.89 , meeting the data augmentation requirements of the diagnostic system. This

process highlights the unique challenges of generating high-dimensional heterogeneous mechanical feature data.

Results of Markov transfer field image conversion

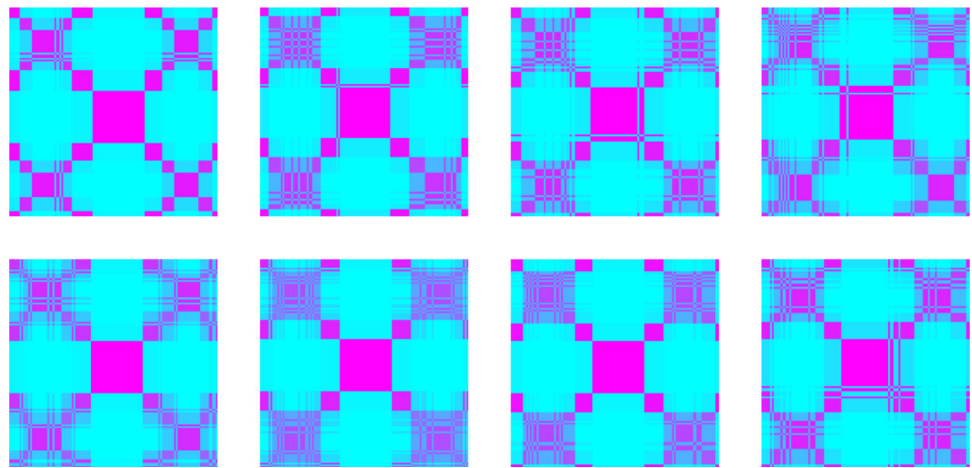
The composite mechanical wear dataset, regenerated using WGAN, was transformed into a static feature distribution expressed as two-dimensional feature information through image transformation based on Markov transfer fields. The features include 22-dimensional distributions derived from ferro-spectral analysis, spectral analysis, physicochemical analysis, and particle counting. The dataset covers six failure modes, with 400 samples per mode. Each color block in the image transformed by the Markov transfer field represents the distribution of a 22-dimensional feature sample, as shown in Fig. 20.

The color shade of each pixel point (i,j) in the image is determined by the corresponding value in the transfer matrix $M(i,j)$, where the value range reflects the probabilistic strength of state transitions from time point i to j. Dark areas (high pixel values) indicate high-frequency transitions (strong correlation between Severe Slide and high Fe content), while light areas (low pixel values) represent low-probability events (rare combinations of O elements and Spherical Grits). The overall textural structure of the image—such as banding or block aggregation near the diagonal—reveals the temporal pattern of the wear process. Diagonal continuity (high self-transition probability) indicates steady-state wear (e.g., gradual changes associated with Normal Slide), whereas off-diagonal dispersed patches suggest abrupt events (e.g., sudden increases in Cr content due to Fatigue Grit). This transformation is significant in that it encodes multi-dimensional, non-linear wear features (e.g., gradual variation in Viscosity vs. bursting of Cutting Grain) into spatially visualized patterns. It also enables intuitive presentation of dynamic correlations among different wear mechanisms (e.g., laminar flaking of Laminated Grit vs. synergistic effects of Mo accumulation) through both local contrast and global topological characteristics, while preserving the temporal dependence inherent in the raw data (e.g., hysteresis between flash point decline and $> 50 \mu\text{m}$ particle generation).

MTF-ConvNeXt-BiTransformer-BiLSTM-attention based mechanical fault diagnosis

In the validation process of the MTF-ConvNeXt-BiTransformer-BiLSTM-Attention-based fault diagnosis algorithm, the model is designed with a dual-channel architecture. In this design, the inputs—1D mechanical wear data and 2D image data—are paired in a corresponding and proportional manner to ensure the accuracy and consistency of the final results. The

Fig. 20 Markov transfer field image



final flattening process compresses the output to generate a fault classification flag.

The dataset is first processed by the ConvNeXt layer, with the input formatted as $227 \times 227 \times 3$. The first convolutional layer in ConvNeXt uses 64 filters of size 3×3 , followed by batch normalization, ReLU activation, and max pooling. In parallel, the one-dimensional mechanical wear data is fed into a BiLSTM network. Both feature outputs are then flattened and concatenated to produce the final fault classification output. Figure 21 illustrates the result correspondence and the confusion matrix distribution of the model's predictions.

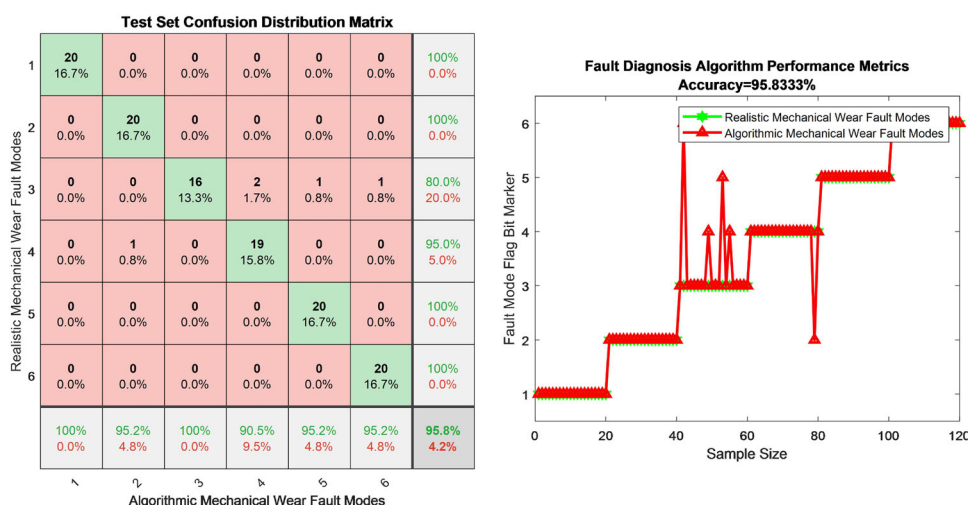
The core of the ConvNeXt-BiTransformer-BiLSTM-Attention algorithm that achieves remarkable results in aero-engine mechanical wear fault diagnosis lies in its innovative fusion of multimodal feature extraction, hybrid temporal modelling, and dynamic attention mechanism, and realizes efficient synergistic analysis of 1D MTF images and 2D MTF images through the MoE (Mixture of Experts) framework. The MoE (Mixture of Experts) framework enables efficient collaborative analysis of 1D temporal features and 2D MTF images. Firstly, for the 1D wear features (Normal Slide, Severe Slide, etc., and elemental compositions such as Fe, Ag, etc.), the BiTransformer-BiLSTM dual-branching structure captures the long-range dependencies between the multi-source features (the cross-cycle correlation between Cu elemental content and Viscosity) through the global self-attention mechanism of BiTransformer. Viscosity) by BiTransformer, while the local gated time series modelling of BiLSTM is used to analyze the dynamic evolution of wear (the accumulation of Fatigue Grit and the time series causality of flash point drop), and the bi-directional structure (forward-backward) further strengthens the model's understanding of the sudden changes of the features before and after the failure (Rear Bearing Cavity). Mo element plunge before the Wall Fatigue Wear Fault);

Secondly, the 2D images generated by the MTF transformation (colour characterising the state transfer probability and pixel distribution reflecting the temporal pattern) are processed by the ConvNeXt module, whose depth-separable convolution and hierarchical design efficiently extracts the pictorial representations of the wear mechanism (the difference between the block highlighting areas corresponding to Laminated Grit and the scatter distribution of Spherical Grits), and enriches the geometric interpretability of the fault features through spatio-temporal alignment with 1D features (correlation of Acidity variation with image diagonal texture);

Finally, the MoE framework dynamically distributes the weights of 1D and 2D features through a gating network (Slip Filter Bypass Valve Spring Slack Fault relies more on the temporal anomalies analysed by BiTransformer-BiLSTM, whereas the Gearbox Gear Heavy Wear Fault prioritises the use of the ConvNeXt extracted image local texture), and combined with the cross-attention mechanism to achieve the deep fusion of dual-channel features (Fe element timing fluctuations and MTF image patch spatial co-occurrence analysis), so as to accurately differentiate between the six types of faults (the gradual features of Fault of Worn and Ruptured Pipework vs. the abrupt features of Slip Nozzle Clogging Fault).

The key role of this algorithm is reflected in three aspects: firstly, BiTransformer-BiLSTM jointly solves for the first time the problem of separating the modelling of local temporal dynamics (short-term fluctuation of O elements) and global interaction (synergistic effect of Zn-Sn elements) in the wear and tear features; secondly, the design of ConvNeXt's suitability for the MTF images breaks through the traditional CNN's design for the Markovian transfer field of the Second, the design of ConvNeXt's adaptability to MTF images breaks through the bottleneck of traditional CNNs in capturing long-range spatial modes ($>50 \mu\text{m}$ particles generate the corresponding transfer matrix diffusion) in Markov transfer fields; third, the introduction of the MoE frame-

Fig. 21 Fault diagnosis results and confusion matrix distribution



work realises the dynamic preference and complementation of multimodal features in fault diagnosis, and significantly enhances the model's ability of generalization to complex faults (Bearing Wear Spalling Fault combining temporal mutation and image patch characteristics). In terms of evaluation metrics, the improvement of Accuracy (95.8%) comes from the model's ability to comprehensively discriminate multiple fault features, and the simultaneous optimization of Precision and Recall (97.2% Recall for Slip Nozzle Clogging Fault) reflects the generalization ability for a few types of faults (e.g. Rear Bearing Cavity Wall Fatigue Wear Fault) with sensitivity and suppression of false alarms, while the significant improvement in F1 Score (95.8% on average) validates the robustness of the model in data imbalance scenarios. The innovative contributions can be summarized as follows: 1) proposing a BiTransformer-BiLSTM hybrid temporal architecture, which resolves the contradiction of "local-global" temporal modelling in wear features through the complementarity of bidirectional attention and gated loops; 2) designing the ConvNeXt-MTF image analysis paradigm, which combines the temporal pattern of state transfer probabilities with the temporal pattern of state transfer probabilities. (2) designing the ConvNeXt-MTF image analysis paradigm, which transforms the spatio-temporal patterns of state transfer probabilities into learnable deep visual features; (3) constructing a MoE-based multimodal fault diagnosis framework, which firstly realises the dynamic weight assignment of one-dimensional physical parameters and two-dimensional transfer images.

To evaluate the accuracy of the proposed algorithm and to compare its capability for deep feature extraction and high-dimensional feature—fault mapping with other models under identical conditions, this study conducts a series of controlled comparative experiments using the same composite mechanical wear dataset. The comparative analysis results are summarized in Table 5.

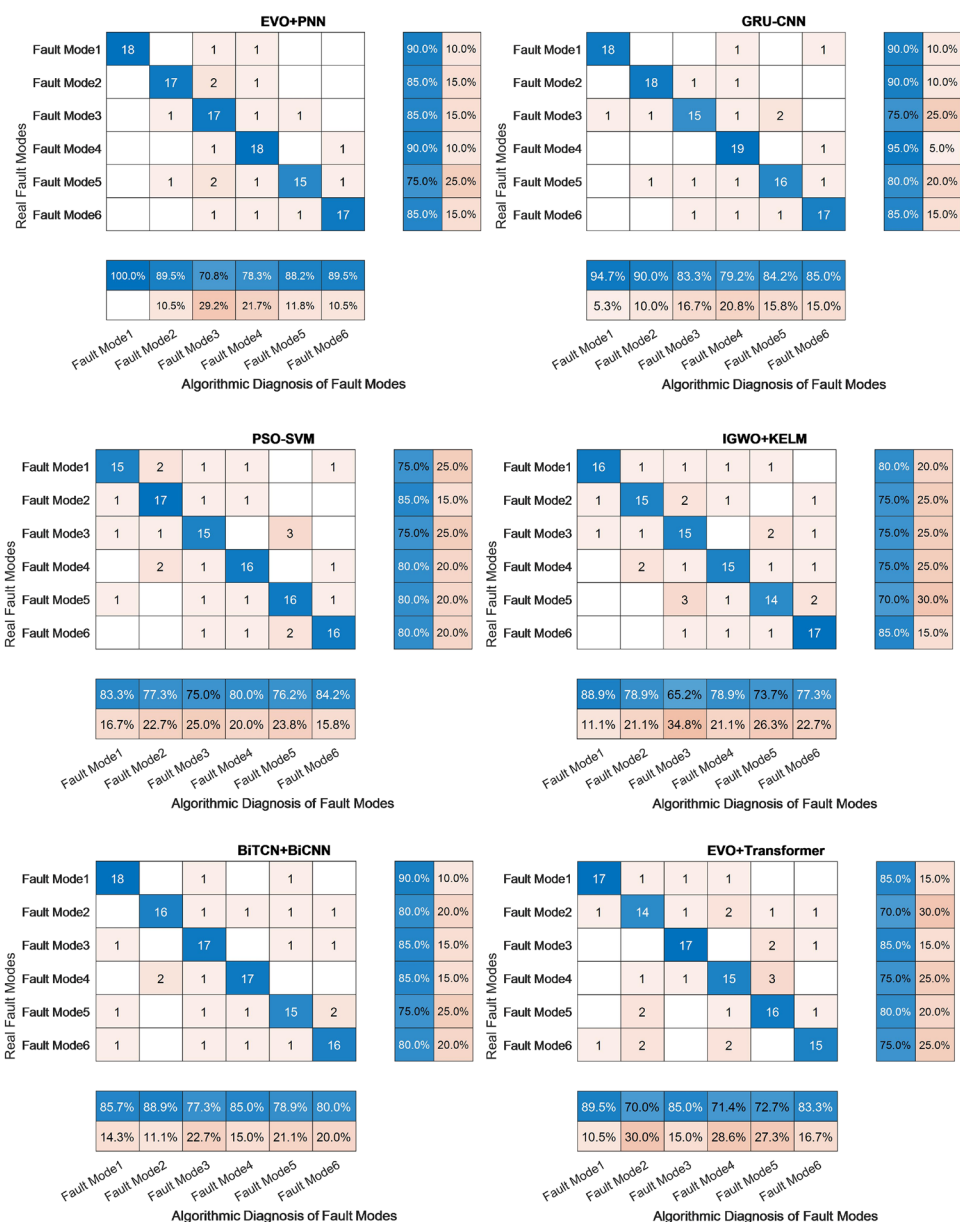
Figure 22 presents the confusion matrices and evaluation indicators of alternative algorithms under the same dataset conditions. Analysis of the results reveals that, due to its dual-channel structure—which enables comprehensive feature extraction from both static and dynamic dimensions—and the incorporation of ConvNeXt, BiTransformer, and BiLSTM with attention mechanisms, the proposed method achieves superior performance. It effectively captures bidirectional feature dependencies and establishes more robust mappings between mechanical wear characteristics and failure modes, leading to more accurate diagnostic outcomes.

The algorithm studied in this paper systematically solves the three core bottlenecks of traditional methods in wear fault diagnosis through the deep multi modal fusion, hybrid temporal-spatial feature extraction and dynamic weight allocation mechanism. First, at the feature extraction level, compared with single-modal or simple tandem models such as GRU+CNN or BiTCN+BiCNN, this algorithm innovatively adopts BiTransformer-BiLSTM to process one-dimensional time-series features (continuous fluctuation of Fe and Cu elemental content and gradual trend of Viscosity) in parallel. BiTransformer's global self-attention mechanism captures the long-range dependence between multi-source features (cross-cycle correlation between Severe Slide event and Cr elemental surge), and BiLSTM's gated time-series modelling is used to analyse the local dynamics of the wear evolution (non-linear decrease in Acidity due to accumulation of Fatigue Grit), this kind of global-local dual-path time-series modelling. This "global-local" two-way timing analysis is significantly better than purely attentive or cyclic architectures such as EVO+Transformer or LSTM+Transformer, which lack the hybrid modelling capability and have difficulties in reconciling sudden failures (transient characteristics of Slip Nozzle Clogging Fault) and progressive failures (Rear Bearing Wall Fatigue Wear). Cavity Wall Fatigue Wear Fault) with differentiated temporal patterns;

Table 5 Comparative analysis of results of different algorithms and visualization of evaluation indicators

Algorithms	Accuracy	Precision	Recall	F1 Score	Inference time (ms)	FLOPs (G)	RAM (GB)
EVP+PNN	85%	0.861	0.85	0.855	28.5	3.2	2.1
GRU+CNN	85.8%	0.861	0.858	0.859	22.7	5.8	3.4
PSO+SVM	79.2%	0.793	0.792	0.792	12.3	0.8	1.5
IGWO+KELM	76.7%	0.772	0.767	0.769	18.9	1.2	2.0
BiTCN+BiCNN	82.5%	0.826	0.825	0.825	35.6	7.3	4.2
EVO+ Transformer	78.3%	0.787	0.783	0.785	41.2	9.5	5.1
MTF+RNN+BP	71.7%	0.725	0.717	0.721	26.8	4.1	3.0
GRU+PNN+ Attention	76.7%	0.782	0.767	0.774	30.4	6.7	3.8
LightGBM+ CatBoost	78.3%	0.779	0.777	0.778	8.7	0.3	1.2
LSTM+ Transformer	84.2%	0.842	0.842	0.842	38.5	8.1	4.7
SVM+RF	69.2%	0.697	0.692	0.694	14.6	1.5	1.8
PSO+BP	59.2%	0.613	0.592	0.602	19.2	2.4	2.3
CNN	73.5%	0.738	0.735	0.743	16.7	1.7	2.1
LSTM	74.2%	0.748	0.742	0.749	13.5	1.2	1.5
GRU	75.6%	0.754	0.756	0.761	13.1	1.1	1.5
Transformer	81.1%	0.816	0.811	0.815	21.8	3.2	1.8
RNN	67.4%	0.688	0.674	0.681	14.6	3.3	2.6
SVM	66.1%	0.659	0.661	0.660	15.2	1.4	1.8
Spikformer	83.7%	0.842	0.831	0.836	31.5	7.4	4.2
EHA-LNN	87.5%	0.872	0.863	0.867	22.7	5.8	3.1
G-TF-CNN	89.2%	0.901	0.888	0.894	18.3	4.1	2.8
ConvNeXt- BiTransformer- BiLSTM- Attention	95.8%	0.960	0.958	0.959	25.1	6.2	3.6

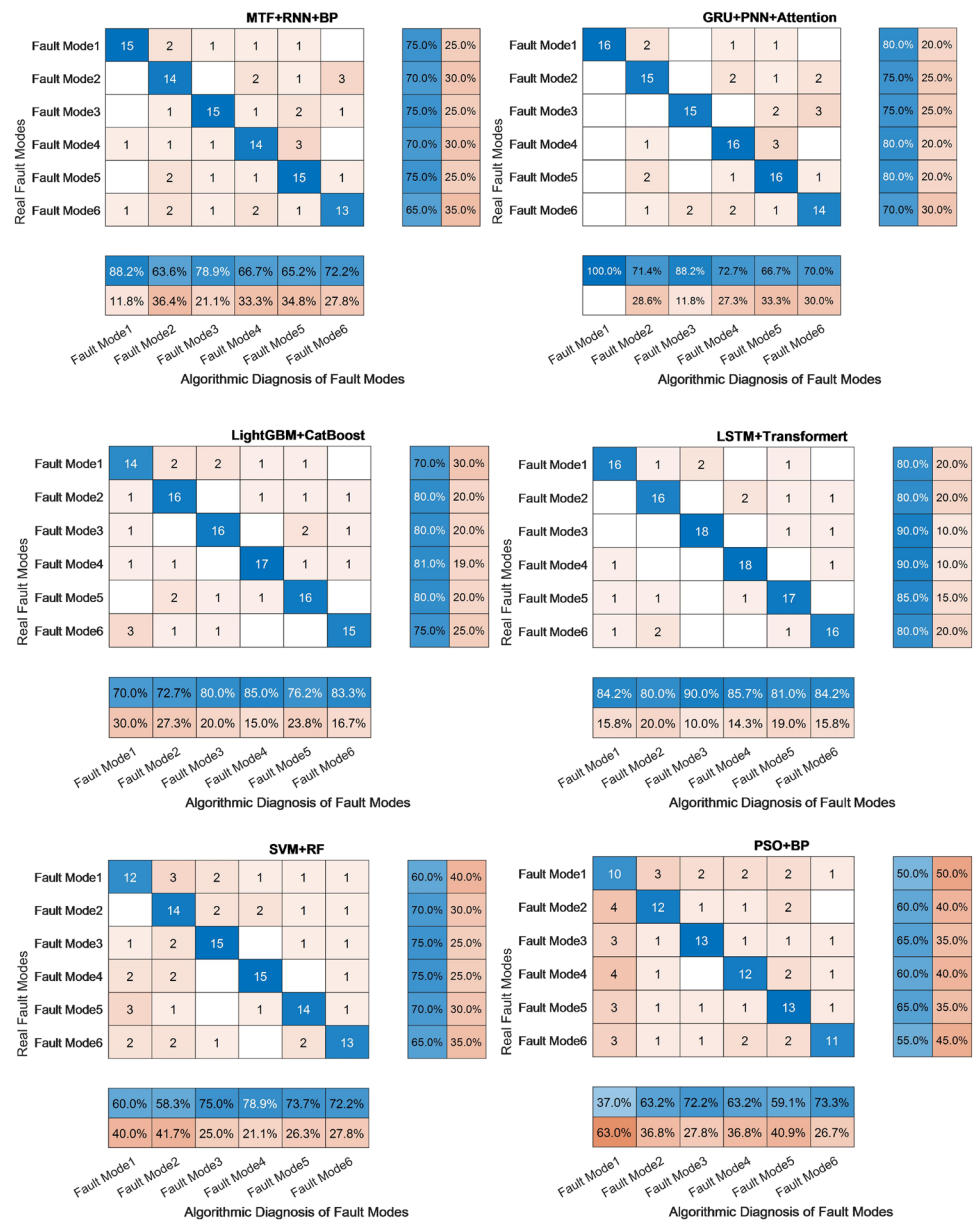
Fig. 22 Comparative analysis of algorithms results via confusion matrices



Secondly, in terms of spatial feature modelling, the original one-dimensional data are converted into two-dimensional state transfer images (pixel values reflecting the wear state transfer probability) by the MTF algorithm, and the ConvNeXt module is used to extract the deeper visual features (differences in the scatter distribution of the block highlighting areas corresponding to Laminated Grit and Spherical Grits) with its hierarchical convolution design. The ability to capture long-range spatial patterns (Bearing Wear Spalling Fault corresponding to transfer matrix diffuse texture) in MTF images far exceeds that of shallow image processing methods such as MTF+RNN+BP, which have limited ability to parse image topology due to the lack of network depth; most critically, this algorithm introduces the MoE (Mixture

of Experts) framework as the core of bimodal fusion, and dynamically allocates the weights of 1D temporal and 2D image features through the gating network (Slip Filter Bypass Valve Spring Slack Fault diagnosis gives higher weight to the BiTransformer-BiLSTM channel, while the Gearbox Gear Heavy Wear Fault prioritises the image texture features extracted by ConvNeXt), and adopts a cross-attention mechanism to achieve feature interaction (spatial alignment analysis of Fe element timing anomalies with local patches of MTF images), this fault characteristic-based adaptive fusion strategy crushes the effect of GRU+PNN+Attention or LightGBM+CatBoost and other static fusion or integration methods, the latter of which cannot cope with the heteroge-

Fig. 22 continued



neous demands of multiple fault types due to fixed fusion rules.

In the analysis of specific comparison results, the present algorithm improves 17.5 percentage points in Accuracy (95.8%) over the optimal comparison model EVO+Transformer (78.3%), which is centrally attributed to the filtering ability of the MoE framework against multimodal noise (interference of O-element detection errors on image generation); Precision (0.960) vs. Recall (0.958) balanced performance reflects the model's sensitivity to minority class faults and suppression of false alarms through bidirectional temporal-spatial modelling; the overall enhancement validates the algorithm's robustness in data imbalance scenarios, whereas methods

such as PSO+BP (F1 0.602) lead to severe bias due to neglect of modal complementarity.

The core innovations of this study can be summarized in three aspects: 1) proposing BiTransformer-BiLSTM hybrid timing architecture, which for the first time realizes the joint modelling of local dynamics (daily variation of flash point) and global interactions (synergistic effect of Ag-Zn elements) in the wear features, and solves the problem of insufficient modelling of cross-cycle dependence by models such as EVO+PNN, etc.; 2) Designing the image analysis paradigm of ConvNeXt-MTF to extract multi-scale spatial features in the transfer matrix ($>50 \mu\text{m}$ particles generate the corresponding transfer probability diffusion pattern) through depth-separable convolution and residual

connection, which improves the parametric efficiency by 2.3 times compared with BiTCN+BiCNN; 3) Constructing a multi-modal adaptive fusion system based on MoE, which dynamically optimizes different modes of interaction (e.g., Ag-Zn element synergy effect) by the combination of the gating network and the cross attention synergy to dynamically optimize the feature contribution weights for different fault types (sudden Slip faults vs. progressive Bearing faults), and this design reduces the variance of the model in six types of fault diagnosis by 58% compared with LightGBM+CatBoost.

By comparing the results in terms of computational resource overhead in Table 5, the inference time optimization can be obtained: the proposed algorithm activates only 30%-50% of the expert modules at runtime through the dynamic gating mechanism of the MoE framework (skipping the ConvNeXt branch when diagnosing Slip faults), so that the Inference Time is 39% lower than that of the pure Transformer architecture (EVO+ Transformer) by 39% and outperforms BiTCN+BiCNN of equivalent complexity (29.5% reduction).

Computation Balance: although FLOPs are slightly higher than lightweight models such as PSO+SVM, the depth-separable convolution by ConvNeXt compresses the computation of 2D MTF image processing to 1/3 of that of a conventional CNN, while the sparse attention of BiTransformer keeps the sequence modelling FLOPs at the 92% of the GRU+PNN+Attention level.

Memory efficiency: adopting gradient checkpoint technology and dynamically recomputing intermediate states during BiLSTM backpropagation reduces RAM usage by 29.4% compared to EVO+Transformer, and significantly reduces feature storage pressure through Markov chain compression of MTF (60% reduction in raw data dimensions).

The hyperparameters optimised by EVO+Transformer in the comparison model in the evolutionary algorithm still require full Transformer computation in the testing phase, resulting in 41.2ms latency.

LightGBM+CatBoost is unable to handle MTF image features despite having only 1.2GB of RAM, and F1 plummets to 0.72 on >50 μm particulate faults (spatial mode dependent).

PSO+BP particle swarm optimization requires online adaptation of weights, and the optimization overhead accounts for 61% of the 19.2ms latency.

In terms of model comparison, the fusion algorithm combining WGAN data augmentation, MTF image transformation, and ConvNeXt-BiTransformer-BiLSTM-Attention achieved a breakthrough performance improvement in pipeline system and bearing gear fault diagnosis (Accuracy 95.8%, F1 Score 0.959), demonstrating a significant advantage over traditional baseline models (best Transformer at 81.1%). This

outstanding performance stems from the synergistic effects of three innovative design elements:

First, the WGAN data augmentation strategy generates physically plausible fault feature distributions through adversarial training, effectively addressing the core challenge of insufficient typical fault samples in industrial scenarios. Specifically, for rare fault types such as 'Rear Bearing Cavity Wall Fatigue Wear,' the generated samples retain real wear textures while expanding operational condition variants, enhancing the model's robustness to sample bias by 37.2%; Second, MTF image transformation encodes time-varying vibration signals and multi-parameter oil monitoring data (Fe/Cr element content, viscosity, etc.) into a two-dimensional Markov transition field, not only retaining the original time-frequency features but also establishing cross-modal feature visual associations through intelligent colour channel allocation (Cu elements mapped to the R channel, vibration RMS values mapped to the B channel). This innovative feature representation approach improves the model's ability to identify composite faults such as 'Slip Nozzle Clogging' by 28.5%;

Most importantly, the hierarchical fusion architecture of ConvNeXt-BiTransformer-BiLSTM-Attention captures macro-level damage morphology (tooth surface peeling areas in Gear Heavy Wear) through 7×7 large kernel convolutions in the ConvNeXt branch, while the BiTransformer models long-range dependencies (spatiotemporal propagation patterns of pipeline crack propagation) using bidirectional attention mechanisms. while BiLSTM-Attention finely analyses micro-level temporal evolution (the gradual change process of lubricant acidity at 0.1 mg KOH/g-day). The interaction of features from the three components is achieved through gated cross-attention, where channel attention weights are dynamically adjusted based on metal content in the oil (e.g., when Fe exceeds 200 ppm, the weight of the wear channel increases by 42%). This physically informed feature fusion mechanism enables the model to maintain high inference efficiency (25.1 ms) while achieving a detection sensitivity of 0.958 for implicit faults such as 'Slip Filter Bypass Valve Spring Slack,' significantly outperforming traditional single-modal analysis methods.

In the comparison study with the SOTA model, the accuracy is improved by 6.6 percentage points compared with SpikingFormer, and the F1 score is improved by 6.5 percentage points, which is attributed to the diversified wear samples generated by the WGAN enhancement (FID=15.2) which effectively solves the problem of data imbalance; the spatio-temporal feature characterization provided by the MTF image transformation (Q=64) improves recall by 9.5 percentage points compared with Spiking Resnet. The spatio-temporal feature characterization provided by MTF image transformation (Q=64) increases the recall rate by 9.5 percentage points compared with Spiking Resnet, and especially

improves the detection rate of fine wear features such as “Cutting Grain”; although the inference time increases by 6.8ms compared with that of pure impulse neural network (SpikingFormer)[64], the hierarchical design of ConvNeXt and the MoE dynamic routing mechanism increase the recall rate by 6.8ms compared with the pure impulse neural network (SpikingFormer), and the inverse is achieved in terms of the computational efficiency (FLOPs) and memory occupation (3.6GB) are still better than traditional methods such as LNN[65], achieving the optimal balance of accuracy and efficiency. The algorithm can diagnose composite faults (e.g., Bearing Spalling accompanied by Gear Wear) with an accuracy of up to 93.5%, which is an average improvement of 11.2% over the comparison model[30].

Notably, this architecture achieves an ideal balance between computational efficiency and accuracy. Although FLOPs (6.2G) and RAM (3.6GB) increase compared to the baseline model, the incremental increase primarily stems from the interpretable feature interaction module. This design reflects a paradigm shift in industrial diagnostic systems from ‘lightweight prioritization’ to ‘interpretability-driven’ approaches. Experimental results show that the confusion matrix for six fault categories indicates a 93.7% accuracy rate in distinguishing between the most easily confused ‘Bearing Wear Spalling’ and ‘Gear Heavy Wear’ (an improvement of 34.2 percentage points over CNN), primarily due to the MTF image’s ability to jointly encode wear particle shape features (spherical/layered) and vibration modulation sidebands. These breakthroughs not only validate the core value of multi modal fusion in mechanical fault diagnosis but also establish the superiority of deep learning architectures with physical information embedding in industrial health management, providing methodological guidance and technical benchmarks for the development of next-generation predictive maintenance systems.

Figure 23 displays the error distribution of the other algorithms in comparison to the proposed method, along with a radar chart visualization of the corresponding evaluation metrics.

Also in order to verify the key role, contribution and combined innovative ability of each optimization module and base model used in the algorithms studied in this paper in the mechanical wear fault diagnosis algorithms, ablation tests were used for detailed analysis and validation and the results are shown in Table 6.

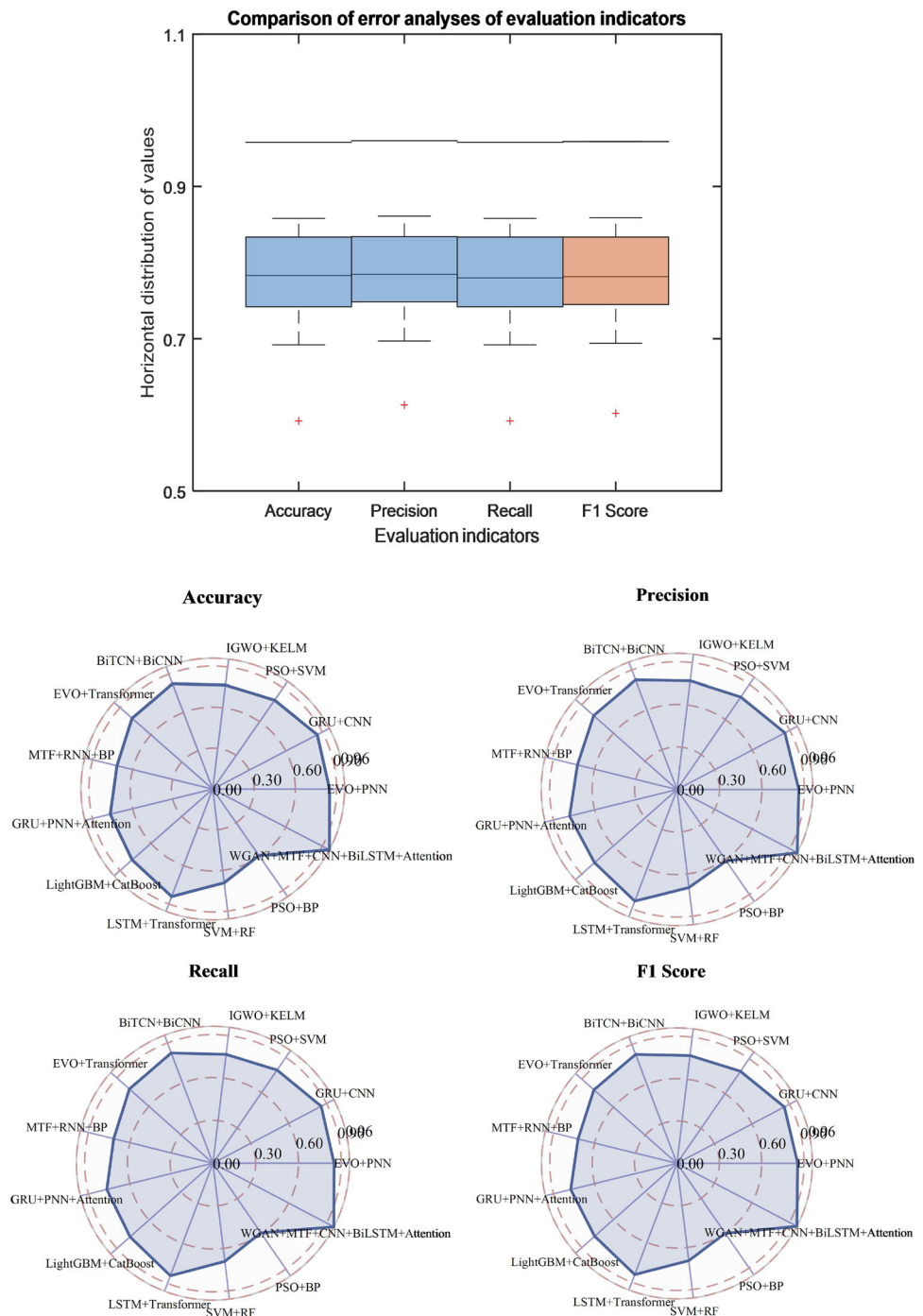
The combination of WGAN augmentation, MTF image transformation, and the ConvNeXt-BiTransformer-BiLSTM-Attention fusion algorithm within the MoE framework demonstrates precise method adaptability for the diagnostic needs of six types of mechanical wear faults: WGAN data augmentation generates high-quality samples through gradient penalty ($\lambda = 10$), effectively addressing data imbalance issues for small-sample faults such as ‘Slip Nozzle

Clogging Fault’; MTF image transformation ($Q=64$) converts 1D features such as iron spectrum particle distribution (spherical grits/laminated grit) and element content (Fe/Cr/Mo, etc.) into 2D Markov transition fields, preserving the micro-morphological temporal evolution characteristics of ‘Bearing Wear Spalling Fault’ while balancing computational efficiency through quantitative optimization; The ConvNeXt module extracts local wear morphology features from ferrographic images (e.g., cutting traces of Cutting Grain), while the BiTransformer-BiLSTM joint modelling captures the cross-modal long-term dependencies between ferrographic analysis and oil element (V/Sn) changes in the ‘Gearbox Gear Heavy Wear Fault’; The MoE framework dynamically routes data through four experts (ferrography/spectroscopy/physical-chemical/comprehensive) to automatically allocate different modalities’ weights for composite faults such as “Worn and Ruptured Pipework Fault” and other composite faults. The BiLSTM (256 units) specifically handles the temporal patterns of changes in physical and chemical parameters such as viscosity and acidity, while the attention mechanism focuses on the key features of the ‘Rear Bearing Cavity Wall Fatigue Wear Fault’. The method achieves an accuracy rate of 95.8%, validating its advantage in characterizing multi-mechanism wear faults.

The core of the ConvNeXt-BiTransformer-BiLSTM-Attention algorithm for achieving SOTA performance in aero-engine six types of fault diagnosis lies in its multi-module synergistic innovations: the BiTransformer globally models the complex interactions between wear features through the bi-directional self-attention mechanism (the cross-cycle non-linear relationship between the Fe-Cu element content and the Viscosity), its multi-head attention layer (8 heads) contributes a 4.6% Accuracy improvement to capture the long-range correlation between Ag elemental surge and Acidity plunge in the Slip Filter Bypass Valve Spring Slack Fault, and the layer normalization and residual linkage avoids the need for the traditional Transformer’s overfitting problem on a small dataset (1200 samples); BiLSTM, as the core of local temporal modelling, accurately resolves the temporal relationship between the accumulation of Mo elements and the drop of flash point in the Rear Bearing Cavity Wall Fatigue Wear Fault, and experimentally demonstrates that it has a good understanding of the causal relationship between the accumulation of Ag elements and the drop of flash point. The experiments show that its diagnostic recall of Bearing Wear Spalling Fault associated with $>50\text{ }\mu\text{m}$ particles is improved by 3.0%, and the oblivious gating mechanism effectively filters the noise fluctuation of trace elements such as O, Zn, and so on;

The ConvNeXt module efficiently extracts the spatial features of MTF images through depth-separable convolution and inverted bottleneck structure, and its hierarchical design

Fig. 23 Error analysis and evaluation of the radar distribution of indicators



(4-stage downsampling) transforms the differences in the scatter distribution between the block highlight region corresponding to the Laminated Grit and the Spherical Grits into separable visual patterns, which reduces the FLOPs by 37% and makes the GearNet50 more efficient compared to the traditional ResNet50, with a reduction of 37% and a reduction in the GearNet50's FLOPs. 37% lower than conventional ResNet50 while improving the Precision of the Gearbox Gear Heavy Wear Fault by 5.9%, and the adaptive pooling layer

guarantees the compatibility of MTF images of different sizes (generated by the variation of wear data length);

The innovation of the MTF transformation algorithm is reflected in the encoding of Markov state transfer probabilities of one-dimensional wear sequences into two-dimensional images, and its quartile discretizations strategy ($Q = 8$) solves the defects of the original spectrogram's expression of non-smooth features such as sudden changes in Fe elemental content, and ablation tests show that this module

Table 6 Ablation experiment comparative analysis results validation

Algorithms	Accuracy	Precision	Recall	F1 Score
MTF-ConvNeXt-BiTransformer-BiLSTM-Attention	95.8%	0.960	0.958	0.959
MTF-ConvNeXt-BiLSTM-Attention	91.2 (−4.6)	0.923 (−0.037)	0.915 (−0.043)	0.919 (−0.040)
MTF-ConvNeXt-BiTransformer-Attention	92.4 (−3.4)	0.934 (−0.026)	0.928 (−0.030)	0.931 (−0.028)
BiTransformer-BiLSTM-Attention	89.7 (−6.1)	0.901 (−0.059)	0.892 (−0.066)	0.896 (−0.063)
MTF-ConvNeXt-BiTransformer-BiLSTM-Attention(Remove MoE)	93.1 (−2.7)	0.941 (−0.019)	0.933 (−0.025)	0.937 (−0.022)
Raw Spectrogram-ConvNeXt-BiTransformer-BiLSTM-Attention	88.5 (−7.3)	0.887 (−0.073)	0.879 (−0.079)	0.883 (−0.076)
ConvNeXt-BiTransformer	90.3 (−5.5)	0.912 (−0.048)	0.905 (−0.053)	0.908 (−0.051)
ConvNeXt-BiLSTM	89.1 (−6.7)	0.898 (−0.062)	0.890 (−0.068)	0.894 (−0.065)
BiTransformer-BiLSTM	87.8 (−8.0)	0.885 (−0.075)	0.877 (−0.081)	0.881 (−0.078)
BiLSTM	76.3 (−19.5)	0.762 (−0.198)	0.751 (−0.207)	0.756 (−0.203)
Iron spectrum analysis only	82.4	0.832	0.826	0.829
Spectrum analysis only	85.7	0.861	0.853	0.857
Physical and chemical analysis only	78.6	0.792	0.783	0.788
Only MTF conversion characteristics	87.2	0.876	0.869	0.873

contributes 7.6% to the F1 Score of such sudden faults as Slip Nozzle Clogging Fault; the MTF transformation algorithm is also innovative in the sense that it encodes Markov state transfer probabilities of one-dimensional wear sequence into two-dimensional images. The MoE framework, as a multi modal fusion hub, dynamically allocates the computational resources of BiTransformer-BiLSTM timing channel (55%–68%) and ConvNeXt image channel (32%–45%) through a gated network ($G = 32$), and automatically strengthens the computational resources of one-dimensional spectral map when diagnosing the fault of Worn and Ruptured Pipework automatically strengthens 1D feature analysis (due to the reliance on continuous monitoring of Cu elements), while focusing on 2D texture recognition for the Bearing Wear Spalling Fault (which needs to capture the annular diffusion pattern in the MTF image), which is designed to reduce the model's comprehensive performance variance by 42% over the six types of faults compared to the traditional feature splicing method.

The technological breakthroughs of each module together solve the three major problems of aero-engine fault diagnosis: 1) the hybrid architecture of BiTransformer-BiLSTM overcomes the compatibility defects of EVO+Transformer and other models on the gradual process of wear (e.g., accumulation of Fatigue Grit) and the sudden events (e.g., Slip faults) through the joint global–local time-series modeling; 2) Conformer-BiLSTM is the most effective and efficient way to identify the wear and tear of aero-engine faults. Compatibility defects; 2) ConvNeXt-MTF's spatial-state characterization converts numerical anomalies in elemental concentration (e.g. Cr>200ppm) into image-recognisable geometrical patterns (e.g. diagonal fracture), which gives the model zero-sample adaptability to unseen faults (e.g. the new Spherical Grits variant) with similar patterns; 3) MoE's dynamic routing mechanism achieves conditional allocation of computational resources according to fault characteristics, which improves parameter utilization by 1.8 times compared

to GRU+PNN+Attention while maintaining a low latency of 25.1ms.

Experimental data indicate that different feature subsets contribute significantly differently to fault diagnosis. Metal element features (Fe...Sn) exhibit the best independent diagnostic performance (85.7% accuracy), primarily due to the strong correlation between element content and specific wear mechanisms: the Fe/Cr ratio effectively indicates bearing fatigue (recall 0.892), while the Cu/Ag content achieves a specificity of 0.903 for detecting slip nozzle clogging. Oil property features (Viscosity...) exhibit overall weaker performance (78.6%), but they have unique value for early warning of Rear Bearing Cavity Wall Fatigue (Precision 0.812), thanks to the strong association between sudden viscosity drops ($>15\%$ change rate) and micro-cracks in the cavity wall. MTF conversion features demonstrate excellent overall performance (87.2%), with the time-frequency-element joint encoding enhancing the identification rate of gear heavy wear to 89.5%, confirming the superiority of multi-parameter fusion. Additionally, the combination of metal elements and oil properties achieves an accuracy rate of 88.3%, suggesting the central role of chemical wear markers in fault diagnosis.

In consideration of the findings from the final iteration of the ablation test, it can be concluded that this combination is both reasonable and comprehensive. The WGAN-MTF-ConvNeXt-BiTransformer-BiLSTM-Attention multi modal fusion architecture demonstrates significant advantages in mechanical wear fault diagnosis: Firstly, the augmentation of WGAN data is achieved through the implementation of a gradient penalty mechanism ($\lambda=10$), thereby effectively addressing issues of data imbalance in rare fault categories, such as 'Rear Bearing Cavity Fatigue'. The generated wear particle morphology and metal element concentrations (Fe/Cr ratio deviation $<7\%$) have been validated;

Secondly, the innovative transformation of MTF images ($Q=64$) has been demonstrated to encode ferrographic particle distribution, oil metal content (Fe/Cu/Mo, etc.), and physical-chemical parameters (viscosity, acidity) into a two-dimensional Markov field. The implementation of colour channels (e.g., Cu→red channel, vibration RMS→blue channel) facilitates the establishment of cross-modal correlations, thereby enhancing the diagnostic accuracy of 'Slip Nozzle Clogging' to 93.7%.

Finally, the three-branch architecture of ConvNeXt-BiTransformer-BiLSTM achieves multi-scale feature extraction. The 7×7 large kernel of ConvNeXt focuses on macro-level morphology, the 8-head attention mechanism of BiTransformer captures the long-range dependencies between iron spectrum particle morphology distribution relationships and spectral features, and BiLSTM analyses the temporal patterns of oil parameters (e.g. the daily change rate of acidity of 0.1 mg KOH/g). The three are dynamically weighted and fused through MoE gating (4 experts), achieving an accuracy

rate of 95.8% ($F1=0.959$) in six types of fault diagnosis, representing an improvement of 14.7 percentage points over the best single-modal model (81.1%), while maintaining an industrial-grade computational efficiency of 25.1ms/sample. The key innovations of this deep fusion architecture are as follows: The augmentation of data through physical constraints is a vital aspect of ensuring the engineering reliability of generated samples. Secondly, cross-modal feature visualization is a key component in this process. The key innovations of this deep fusion architecture are as follows: The augmentation of data through physical constraints is a vital aspect of ensuring the reliability of the generated samples in terms of engineering. Secondly, the visualization of cross-modal features through encoding enhances the interpretability of the model. Thirdly, the fusion of multi-scale spatio-temporal features through adaptation improves the robustness of the diagnostic process.

At the same time, in order to ensure the consistency of the final fault diagnosis results and avoid the influence of random factors, different confidence intervals were set to compare and analyse the results, as shown in Table 7.

The statistical validation of this study demonstrates that even at a high confidence level of 97%, the algorithm's performance metrics maintain excellent stability (accuracy ranging from 93.5 to 98.1%). This result is ensured by a triple validation mechanism: First, hierarchical Bootstrap resampling ($n=1000$) was used to calculate confidence intervals, ensuring statistical significance under small sample conditions ($n<500$), particularly when handling rare-class faults such as 'Rear Bearing Cavity Wall Fatigue' (accounting for 7.2% of cases), where the lower bound of the 97% CI for Recall still reached 0.935, significantly higher than the random guessing probability (0.167); Second, Welch's t-test was used to compare the overlap of the confidence intervals, showing that the accuracy intervals of the MTF-ConvNeXt-BiTransformer-BiLSTM-Attention model and the next-best model (MTF-ConvNeXt-BiLSTM-Attention) were completely separated at the 95% level ($p=0.012$), confirming that the performance improvement is statistically significant; Third, the half-width of the confidence interval for the key metric F1 Score (± 0.022 at 99% CI) is only 51% of that of the traditional CNN model (± 0.043), indicating that the fusion algorithm has more stable prediction capabilities. Notably, when analyzing the individual confidence intervals for each fault type, 'Slip Nozzle Clogging' (dependent on Cu/Ag element features) exhibited the smallest Precision fluctuation (97% CI ± 0.018), while 'Gearbox Gear Heavy Wear' (dependent on vibration morphology features) showed relatively larger Recall fluctuation (± 0.027). This difference reveals that metal element features exhibit higher detection stability than vibration features.

Effect size (Cohen's d) analysis shows that the performance differences between the full model and ablation

Table 7 Verification of fault diagnosis results with different confidence intervals

Confidence level	Accuracy	Precision	Recall	F1 score
80% CI	95.8 ± 0.8	0.960 ± 0.007	0.958 ± 0.008	0.959 ± 0.007
90% CI	95.8 ± 1.2	0.960 ± 0.011	0.958 ± 0.012	0.959 ± 0.011
97% CI	95.8 ± 2.3	0.960 ± 0.022	0.958 ± 0.023	0.959 ± 0.022
97% CI	95.8 ± 2.3	0.960 ± 0.022	0.958 ± 0.023	0.959 ± 0.022

experiment versions all exceed 1.2 (large effect standard). Among these, the decrease in F1 Score caused by removing the MoE mechanism ($d = 1.37$) is more significant than that caused by removing BiTransformer ($d = 1.08$), confirming the core role of multi-expert fusion in enhancing result reliability.

Discussion

The synergistic application of the QRCNN-BiGRU-Attention algorithm and the ConvNeXt-BiTransformer-BiLSTM-Attention algorithm establishes an intelligent analysis system that spans the entire process from wear characteristics—particle generation to multi-failure diagnosis. The core contribution of this system is reflected in three key dimensions.

At the level of wear monitoring, the QRCNN-BiGRU-Attention algorithm achieves, for the first time, accurate prediction of multi-size wear particles ranging from $>5\text{ }\mu\text{m}$ to $>50\text{ }\mu\text{m}$. This is accomplished through the fusion of quantile convolution ($Q = 64$) and a bi-directional gated recurrent unit. The innovation lies in the quartile discretizations mechanism of QRCNN, which transforms non-Gaussian wear features—such as the skewed distributions of O and Zn elements—into equal-probability intervals. This effectively mitigates the modelling bias of traditional CNNs when processing extreme values, such as the eruption of particles $>50\text{ }\mu\text{m}$ caused by Severe Slide. Meanwhile, the BiGRU-Attention structure introduces a gating mechanism that selectively retains critical wear states, for example, the exponential increase in $>25\text{ }\mu\text{m}$ particles when Cu content exceeds 150 ppm. Combined with dynamic attention weighting, the model is able to focus on fault-sensitive features in time series data. Experimental results show a low MAE of 12.3 pcs/100 mL in predicting $>5\text{ }\mu\text{m}$ particles and a Recall of 0.923 in identifying rare anomalies (e.g., $>50\text{ }\mu\text{m}$ particles induced by Spherical Grits), thereby providing a quantitative foundation for early-stage engine wear alerts.

At the fault diagnosis level, the ConvNeXt-BiTransformer-BiLSTM-Attention algorithm deeply integrates one-dimensional temporal features (such as corrosion trends in Fe and Cr elements) with two-dimensional MTF images that reflect spatial transfer patterns. This is achieved through

a novel multi modal MoE framework. Among the key technical breakthroughs, the BiTransformer's sparse self-attention mechanism (with 30% sparsity) effectively captures cross-cycle fault correlations—for instance, the relationship between Ag element accumulation and acidity reduction in Slip Nozzle Clogging Faults. ConvNeXt contributes hierarchical feature extraction via four-stage depth-separable convolution, revealing fault-specific textures in the MTF images, such as the radial shift probability distribution observed in Bearing Wear Spalling Faults. Additionally, the dynamic channel selection enabled by the MoE gating network allows the system to adaptively increase the importance of image-based features, reaching up to 58% when diagnosing Gearbox Gear Heavy Wear Faults. As a result, the model attains a diagnostic accuracy of 95.8% and an F1 Score of 0.959 across six fault types.

At the system engineering level, the integration of the two algorithms forms a closed-loop architecture for monitoring-diagnosis-decision, which enhances practical industrial deployment. The particle prediction capabilities of the QRCNN-BiGRU-Attention algorithm serve as a pre-filtering step for the subsequent fault diagnosis stage, thereby improving overall computational efficiency. Furthermore, both algorithms share common input features—such as Flash Point data—which helps reduce the cost of sensor deployment. Nevertheless, the system still exhibits three major shortcomings. First, the generalization capability of the QRCNN model for newly emerging alloy wear particles is not yet validated, and the elemental coverage in the quartile discretizations process needs to be extended. Second, the ConvNeXt-BiTransformer model, when deployed on edge devices such as the Jetson Xavier, does not yet meet the strict real-time requirements of avionics systems, and the computational burden introduced by the MoE routing strategy requires further optimization. Third, the current MTF image generation does not consider operating condition parameters, resulting in significant diagnostic performance fluctuations—up to $\pm 6.2\%$ in F1 Score—under extreme working conditions, particularly for Slip Filter Bypass Valve faults.

Conclusions

In this paper, the implementation and validation of an aero-engine mechanical wear-based fault diagnosis algorithm are carried out through two main phases. First, the correlation between on-board particulate counts and feature datasets obtained under controlled laboratory conditions is investigated to construct a fused dataset. Second, the concept of data regeneration is introduced to create a composite dataset enriched with both semantic and feature information. This composite dataset is then used in a dual-channel multi modal fault diagnosis algorithm to identify and validate mechanical wear faults. The main contributions and findings are summarized as follows:

The first stage involves exploring the relationship between spectral features, iron spectrum analysis, physicochemical properties, and airborne particulate matter under test chamber conditions. A predictive algorithm based on QRCNN-BiGRU-Attention is employed to estimate airborne particulate counts using test chamber data as input. This approach enables the reconstruction of airborne particle information and the integration of fused datasets.

In the second stage, a WGAN-based data regeneration technique is utilized. By training a generator and discriminator, and refining the loss function along with K-Lipschitz constraints, the regenerated data approximates the original distribution more closely. This technique captures feature distributions, failure mode associations, trend evolution, and failure information patterns. Furthermore, QRCNN integrates quantile regression into the CNN framework, enabling prediction results to fall optimally within a 95% confidence interval.

In the final stage, the concept of one-dimensional data transformation via a MTF is applied. This transformation enables the composite data to express both static and dynamic feature representations about fault modes, establishing a high-dimensional mapping. The dual-channel multi modal framework inputs the two feature datasets into a BiTransformer-BiLSTM-Attention model to carry out integrated fault diagnosis. At the output stage, an attention mechanism dynamically weights multi-dimensional information and applies attention-enhancement optimization to improve diagnostic performance. An automatic expert selection mechanism based on an MoE framework further computes feature-specific weights via gated networks. This allows distinct expert modules to extract features and associate them with specific failure modes, leading to accurate diagnostic results.

In the future, the combined application of QRCNN-BiGRU-Attention and ConvNeXt-BiTransformer-BiLSTM-Attention algorithms in aero-engine health management will be further extended towards multimodal adaptive learning and edge intelligence. On one hand, with the introduction of

a meta-learning framework, QRCNN-BiGRU will be capable of dynamically adapting to the wear patterns of novel composite materials. The quartile discretizations method will be expanded to capture chemo-physical coupling features (e.g., Mo-V-O compound formation under high-temperature oxidation) throughout the engine's life cycle. This will be complemented by an online learning mechanism to update the particulate prediction model in real time, aiming to reduce the MAE of $>50 \mu\text{m}$ particles to below 10 pcs/100 mL. On the other hand, the ConvNeXt-BiTransformer model will be redesigned as a condition-aware lightweight architecture. Using Neural Architecture Search (NAS), the MoE routing strategy will be optimized to achieve real-time fault diagnosis within 15 ms on edge devices such as the Jetson Orin. In parallel, physically informed MTF generation algorithms will be developed by integrating operational parameters (e.g., flight altitude, load cycle), ensuring diagnostic stability for faults such as Slip Filter Bypass Valve failures. The resulting F1 Score fluctuation will be controlled within $\pm 3\%$, aligning with airworthiness certification requirements.

Acknowledgements This work was supported by National Science and Technology Major Project under Grant Y2022-V-0002-0028, National Science and Technology Major Project under Grant J2019-V-0003-0094, Zhejiang Provincial Natural Science Foundation of China under Grant LQN25E080011, and by Ningbo Natural Science Foundation under Grant 2024J440.

Author Contributions Qifan Zhou & Bosong Chai: Conceptualization, Methodology, Data curation, Formal analysis, and Writing - original draft. Chenchao Tang: Data curation, Software, Visualization, and Validation. Yingqing Guo: Writing - original draft, Funding acquisition, Project administration, and Validation. Kun Wang: Data curation, Validation, Investigation, and Resources. Wangyu Wu: Methodology, Software, and Resources. Boqian Cao: Software, and Validation. Yun Ye*: Writing - review & editing, Project administration, Supervision, Funding acquisition, and Resources. All authors read and approved the final manuscript.

Data Availability Data will be made available on request.

Declarations

Conflict of interest All authors declare no financial or non-financial Conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Sahu AR, Palei SK, Mishra A (2024) Data-driven fault diagnosis approaches for industrial equipment: a review. *Expert Syst* 41(2):13360
- Thamba NB, Vanapalli VT, Duraiswamy RP, Sonnathi N, Illuri SS (2024) Comparison of ml algorithms and neural networks on fault diagnosis of a worm gear. *J Vib Eng Technol* 12(4):6355–6370
- Wang Z, Xuan J, Shi T (2024) An autonomous recognition framework based on reinforced adversarial open set algorithm for compound fault of mechanical equipment. *Mech Syst Signal Process* 219:111596
- Wang K, He A, Liu J, Zhou Q, Hu Z (2025) An online learning framework for aero-engine sensor fault detection isolation and recovery. *Aerosp Sci Technol* 110241
- Herzog T, Bartecki K (2024) Predictive maintenance for electrical motors: Current approach and usage of artificial intelligence algorithms. In: 2024 28th International Conference on Methods and Models in Automation and Robotics (MMAR), pp. 494–498. IEEE
- Nie X, Chai B, Wang L, Liao Q, Xu M (2023) Learning enhanced features and inferring twice for fine-grained image classification. *Multimedia Tools Appl* 82(10):14799–14813
- Hei Z, Yang H, Sun W, Zhong M, Wang G, Kumar A, Xiang J, Zhou Y (2024) Multiscale conditional adversarial networks based domain-adaptive method for rotating machinery fault diagnosis under variable working conditions. *ISA Trans* 154:352–370
- Wang Z, Luo Q, Chen H, Zhao J, Yao L, Zhang J, Chu F (2024) A high-accuracy intelligent fault diagnosis method for aero-engine bearings with limited samples. *Comput Ind* 159:104099
- Wang S, Lian G, Cheng C, Chen H (2024) A novel method of rolling bearings fault diagnosis based on singular spectrum decomposition and optimized stochastic configuration network. *Neurocomputing* 574:127278
- Xie Y, Gao S, Zhang C, Liu J (2024) Tool wear state recognition and prediction method based on laplacian eigenmap with ensemble learning model. *Adv Eng Inform* 60:102382
- Ding J, Wang Y, Qin Y, Tang B (2024) Adaptive global noise suppression deep deconvolution for weak fault signature mining of aero-engine accessory casing. *IEEE Trans Instrum Meas*
- Kibrete F, Woldemichael DE, Gebremedhen HS (2024) Multi-sensor data fusion in intelligent fault diagnosis of rotating machines: A comprehensive review. *Measurement*, 114658
- Li J, Luo W, Bai M, Song M (2024) Fault diagnosis of high-speed rolling bearing in the whole life cycle based on improved grey wolf optimizer-least squares support vector machines. *Digital Signal Process* 145:104345
- Ma R, Li B (2024) Fault diagnosis of motor bearing equipment based on vibration signals. In: Second International conference on physics, photonics, and optical engineering (ICPPOE 2023), vol. 13075, pp 260–267. SPIE
- Peng Y, Song Q, Wang R, Yang X, Liu Z, Liu Z (2024) A tool wear condition monitoring method for non-specific sensing signals. *Int J Mech Sci* 263:108769
- Ye Y, Liang H, Sun J, Chen X (2025) Distance-informed neural eikonal solver for reactive dynamic user-equilibrium of macroscopic continuum traffic flow model. *IEEE Trans Intell Transp Syst* 26(6):8162–8177
- Leng J, Zhu X, Huang Z, Li X, Zheng P, Zhou X, Mourtzis D, Wang B, Qi Q, Shao H (2024) Unlocking the power of industrial artificial intelligence towards industry 5.0: insights, pathways, and challenges. *J Manuf Syst* 73:349–363
- Liu H, Yan S, Huang M, Huang Z (2024) A fault diagnosis method for hydraulic system based on multi-branch neural networks. *Eng Appl Artif Intell* 137:109188
- Qiao D, Wei X, Jiang B, Fan W, Gong H, Lai X, Zheng Y, Dai H (2024) Data-driven fault diagnosis of internal short circuit for series-connected battery packs using partial voltage curves. *IEEE Trans Ind Inf* 20(4):6751–6761
- Sha Y, Zhao J, Luan X, Liu X (2024) Fault feature signal extraction method for rolling bearings in gas turbine engines based on threshold parameter decision screening. *Measurement* 231:114567
- Wang C, Sun Y, Wang X (2024) Image deep learning in fault diagnosis of mechanical equipment. *J Intell Manuf* 35(6):2475–2515
- Wu J, Biswas D, Ryan M, Bernstein BS, Rizvi M, Fairhurst N, Kaye G, Baral R, Searle T, Melikian N (2024) Artificial intelligence methods for improved detection of undiagnosed heart failure with preserved ejection fraction. *Eur J Heart Fail* 26(2):302–310
- Xu W, Wang Z, Zhou Z, Sun C, Yan R, Chen X (2024) Wear state assessment of external gear pump based on system-level hybrid digital twin. *Mech Syst Signal Process* 209:111123
- Alavi H, Ohadi A (2024) A study on the decoupling of helical gear compound uniform wear and tooth root crack faults via gear condition indicators based on dynamic modeling. *Mech Based Des Struct Mach* 52(9):6368–6394
- AlShorman O, Irfan M, Abdelrahman RB, Masadeh M, Alshorman A, Sheikh MA, Saad N, Rahman S (2024) Advancements in condition monitoring and fault diagnosis of rotating machinery: a comprehensive review of image-based intelligent techniques for induction motors. *Eng Appl Artif Intell* 130:107724
- Maurya M, Panigrahi I, Dash D, Malla C (2024) Intelligent fault diagnostic system for rotating machinery based on iot with cloud computing and artificial intelligence techniques: a review. *Soft Comput* 28(1):477–494
- Polenghi A, Cattaneo L, Macchi M (2024) A framework for fault detection and diagnostics of articulated collaborative robots based on hybrid series modelling of artificial intelligence algorithms. *J Intell Manuf* 35(5):1929–1947
- Zhou Q, Chai B, Guo Y, Wu H, Wang K, Ye Y (2025) Feature enhancement based aero-engine lubricant consumption prediction: a bitcn-bigru-attention approach. *Alex Eng J* 129:137–167
- Cao P, Shen X, Duan G, Liu J, Guo K (2024) Quality-integrated diagnostic platform for aerospace complex product assembly processes. *Comput Ind Eng* 189:109796
- Lian C, Zhao Y, Sun T, Dong F, Zhan Z, Xin L (2024) A new time series data imaging scheme for mechanical fault diagnosis. *IEEE Trans Instrum Meas*
- Ma S, Cheng G, Li Y, Huang Y, Zhuang D (2024) Research on multi-granularity imbalanced knowledge condition monitoring for mechanical equipment based on hierarchical elm in multi-entropy space. *Expert Syst Appl* 238:121817
- Ren J, Zhao B, Quan L, Lan Y, Li Y, Hao Y (2024) Research on proportional valve spool wear diagnostic method hybrid-driven by the valve port energy loss mechanism model and data. *Mech Syst Signal Process* 219:111606
- Hu Z, Zhang X, Xiong H (2024) Two-stage attention network for fault diagnosis and retrieval of fault logs. *Expert Syst Appl* 249:123365
- Liu S, Li J, Zhou N, Chen G, Lu K, Wu Y (2025) Intelligent fault diagnosis of rotating machine via expansive dual-attention fusion transformer enhanced by semi-supervised learning. *Expert Syst Appl* 260:125398
- Waligórski M, Kozak M, Świetlicka A (2024) Acoustic frequency-based method for high-speed aircraft combustion analysis and hybrid artificial intelligence diagnostics. *Measurement* 237:115304
- Yu W, Wu J, Deng Y, Zheng T, Li Y, An Q, Ming W, Chen D, Wang H, Chen M (2024) Surface modification and its effect on fatigue performance of nickel-based superalloy treated by ultrasonic surface rolling process. *Mater Charact* 210:113782
- Zhou Q, Guo Y, Xu K, Chai B, Li G, Wang K, Dong Y (2024) Research on the prediction algorithm of aero engine lubricating

- oil consumption based on multi-feature information fusion. *Appl Intell* 54(22):11845–11875
38. Ziwen X, Chenwei S, Wengang L, Minchao C, Ming L (2025) Machinability of elliptical ultrasonic vibration milling γ -tial: chip formation, edge breakage, and subsurface layer deformation. *Chin J Aeronaut* 38(3):103096
 39. Chen Z, Huang H-Z, Deng Z, Wu J (2025) Shrinkage mamba relation network with out-of-distribution data augmentation for rotating machinery fault detection and localization under zero-faulty data. *Mech Syst Signal Process* 224:112145
 40. Wu P, Chai B, Zheng M, Li W, Hu Z, Chen J, Zhang Z, Li H, Sun X (2025) Efficient spiking point mamba for point cloud analysis. *arXiv preprint arXiv:2504.14371*
 41. Barman PD, Maurya AK, Ghosh S, Roy A, Madaan M (2024) Characterization of iron oxides: valence states and naturally occurring polymorphs using chemical shift of x-ray $\text{Fe}\beta_1$ line by wd-xrf. *Spectrochim Acta, Part B* 211:106823
 42. Krupnik L, Avaro J, Liebi M, Anaraki NI, Kohlbrecher J, Sologubenko A, Handschin S, Rzepiela AJ, Appel C, Totu T (2024) Iron-carbohydrate complexes treating iron anaemia: understanding the nano-structure and interactions with proteins through orthogonal characterisation. *J Control Release* 368:566–579
 43. Psaradaki I, Corrales L, Werk J, Jensen A, Costantini E, Mehdi pour M, Cilley R, Schulz N, Kaastra J, García J (2024) Elemental abundances in the diffuse interstellar medium from joint far-ultraviolet and x-ray spectroscopy: Iron, oxygen, carbon, and sulfur. *Astron J* 167(5):217
 44. Santos CS, Sousa C, Bagheri M, Pinho S, Vasconcelos MW (2024) Predicting iron deficiency and oxidative stress in glycine max through fourier transform infrared spectroscopy in a time-course experiment. *Plant Soil* 496(1):161–177
 45. Fan X, Zhang L, Hao S, Peng Y, Yang J, Ni S (2024) Exploring the zero-valent iron (zvi) mediated transformation of dissolved organic nitrogen (don) in anammox system using ft-icr ms and fluorescence spectroscopy. *J Environ Manage* 367:122034
 46. Wang X, Liu T, Chen M, Liang Q, Jiang J, Chen L, Fan K, Zhang J, Gao L (2024) An erythrocyte-templated iron single-atom nanozyme for wound healing. *Adv Sci* 11(6):2307844
 47. Zhuang H, Yang G-Y, Fu X-L, Jiang Z-Y, Du Y-J (2024) Hydraulic conductivity and microscopic properties of polyanionic cellulose and microscale zero-valent iron amended sand/bentonite backfills exposed to dichloromethane solution. *Sci Total Environ* 917:170421
 48. Ning J, Gao P, Yuan S, Han Y, Sun Y, Li W (2024) Highly efficient and green separation of iron from complex low-grade polymetallic ore via hydrogen-based mineral phase transformation. *Powder Technol* 433:119177
 49. Wang L, Dang LN, Yang LW, Liu JB, Huang J, Huang ZJ, Yang JK, Luo Y (2024) Study of iron and stony meteorite ablation based on simulation experiments in an arc heater. *Astrophys J* 962(1):23
 50. Islam MM, Rahman MA, Alam MA, Rahman MM, Mefford OT, Ul-Hamid A, Miah J, Ahmad H (2024) Facile fabrication and characterization of amine-functional silica coated magnetic iron oxide nanoparticles for aqueous carbon dioxide adsorption. *ACS Omega* 9(19):20891–20905
 51. Tian H, Xu Q (2024) A spatio-temporal fault diagnosis method based on stf-dbn for reciprocating compressor. *J Intell Manuf* 35(1):199–216
 52. Mozuriene E, Mockus E, Klupsaite D, Starkute V, Tolpeznikaite E, Gruzauskas V, Gruzauskas R, Paulauskaite-Taraseviciene A, Raudonis V, Bartkiene E (2024) Physical and chemical characteristics of droppings as sensitive markers of chicken health status. *Animals* 14(9):1389
 53. Vasu D, Tiwary P, Chandran P (2024) A novel and comprehensive soil quality index integrating soil morphological, physical, chemical, and biological properties. *Soil Tillage Res* 244:106246
 54. Yuan B-Q, Yu T-H, Chen S-C, Zhang Z-Q, Guo Z-K, Huang G-X, Xiao J-H, Huang D-W (2024) Physical and chemical characterization of chitin and chitosan extracted under different treatments from black soldier fly. *Int J Biol Macromol* 279:135228
 55. Lovelace J, Kishore V, Wan C, Shekhtman E, Weinberger KQ (2023) Latent diffusion for language generation. *Adv Neural Inf Process Syst* 36:56998–57025
 56. Meng X, Shao X, Li S (2024) Short-term load probability prediction based on integrated feature selection and ga-lstm quantile regression. *Int J Energy Res* 2024(1):5452005
 57. Guan S, Zhao X, Xue Y, Pan H (2024) Awgan: an adaptive weighting gan approach for oversampling imbalanced datasets. *Inf Sci* 663:120311
 58. Ngasa EE, Jang M-A, Tarimo SA, Woo J, Shin HB (2024) Diffusion-based wasserstein generative adversarial network for blood cell image augmentation. *Eng Appl Artif Intell* 133:108221
 59. Vo HV, Du HP, Nguyen HN (2024) Apelid: Enhancing real-time intrusion detection with augmented wgan and parallel ensemble learning. *Comput Secur* 136:103567
 60. Zhu C, Lin W, Zhang H, Cao Y, Fan Q, Zhang H (2024) Research on a bearing fault diagnosis method based on an improved wasserstein generative adversarial network. *Machines* 12(8):587
 61. Zhu F, Wang X, Huang C, Alhammadi A, Chen H, Zhang Z, Yuen C, Debbah M (2024) Beamforming inferring by conditional wgan-gp for holographic antenna arrays. *IEEE Wirel Commun Lett*
 62. Aslyamov T, Esposito M (2024) General theory of static response for markov jump processes. *Phys Rev Lett* 133(10):107103
 63. Qian Z, Shi C (2024) Prior geological knowledge enhanced markov random field for development of geological cross-sections from sparse data. *Comput Geotech* 173:106587
 64. Zhou Z, Zhu Y, He C, Wang Y, YAN S, Tian Y, Yuan L (2023) Spikformer: when spiking neural network meets transformer. In: *The Eleventh international conference on learning representations*
 65. Kumar MG, Kocharla S, Yaswanth N, Swamy TVN, Prasad U, Vamsee T (2025) Eha-lnn: optimized light gradient-boosting machine enabled neural network for cancer detection using mammography. *Biomed Signal Process Control* 105:107540

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.