aerospace

MDPI

Article

A Dual-Branch Feature Construction for Hot Jet Remote Sensing of a Certain Aero-Engine Under Diverse Operating Conditions

Zhenping Kang, Yuntao Li, Yurong Liao, Xinyan Yang and Zhaoming Li *

Department of Electronic and Optical Engineering, Space Engineering University, Beijing 101416, China; zhenpingkang@hgd.edu.cn (Z.K.); lyt3007@hdg.edu.cn (Y.L.); liaoyr@hgd.edu.cn (Y.L.); xinyan_yang@hgd.edu.cn (X.Y.)

* Correspondence: lizm@hgd.edu.cn

Abstract

Aiming to address the problem of extracting the remote sensing FTIR spectral characteristics of the hot jet of a certain type of aero-engine under different working conditions, this paper proposes a feature construction algorithm for the remote sensing FTIR spectral characteristics of the aero-engine hot jet based on the fusion of the original spectral features and the deep spectral features. The infrared spectrum was collected at a distance of 280 m, covering the spectral range of 2.5–15 μm with a resolution of 1 cm^{-1}. The Neighborhood–Autoencoder Integration Dual-Branch Network (NAIDN) feature construction algorithm is proposed. This algorithm contains a neighborhood integration branch and an autoencoder branch. The neighborhood integration branch converts the radiation intensity values of discrete wavenumber points into local energy aggregation features through a sliding window, accurately extracting the key physical information in the original spectrum. The autoencoder branch uses a three-layer fully connected neural network architecture to mine the deep spectral features of the spectral data. The algorithms of the two branches not only retain the physical interpretability of spectral analysis but also capture the multi-parameter coupling information hidden in the hot jet spectrum through the representation learning ability of the autoencoder, achieving feature fusion across spatial dimensions. Compared with traditional feature construction algorithms, the dual-branch feature construction algorithm proposed in this paper has stronger comprehensive representation capabilities. The content of carbon dioxide (CO_2) and cyanide groups (-C≡N) in the hot jet under different operating conditions varies significantly. In the experiment, an unsupervised clustering algorithm, the Agglomerative Clustering classifier, is selected, and the classification accuracy of the features extracted by the algorithm in this paper reaches 92.97% on this classifier, thereby verifying the effectiveness of the algorithm in this paper.

Keywords: FTIR; hot jet of aero-engine; feature construction

Check for updates

Academic Editors: Konstantinos Kontis and Gang Sun

Received: 24 January 2026
Revised: 13 March 2026
Accepted: 8 April 2026
Published: 9 April 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Aero-engines [1–4] are the core components of aircraft, and their operating conditions directly affect flight safety and efficiency. During various operations such as takeoff, cruise, and landing, the engines encounter different aerodynamic loads, temperatures, and pressure environments, and their operating conditions are complex and variable. In-depth research on their operating characteristics under various conditions is of great significance for ensuring aviation safety and optimizing engine performance. To address the challenges of abnormal monitoring for aero-engines under different flight conditions, Sun et al. [5]

innovatively proposed a brand-new evaluation method based on state monitoring information. This method used kernel principal component analysis (KPCA) to construct the condition subspace of the aero-engine, thereby accurately depicting the performance evolution process of the engine. To solve the problem of abnormal monitoring for aero-engines under different flight conditions, Wang et al. [6] developed an adaptive anomaly monitoring framework. Firstly, they used the mean and standard deviation method to preprocess the flight data, achieving the automatic division of complex flight scenarios; then, they constructed a monitoring model combining the Sparrow Search Algorithm (SSA) and Kmedoids algorithm to ensure adaptive monitoring in all flight conditions; finally, they designed specific indicators to precisely evaluate the abnormal risk via state monitoring. Through actual flight data verification, this method demonstrated higher detection accuracy and a lower false alarm rate (FAR) in the detection of sudden and progressive anomalies in aero-engines. Liu et al. [7] proposed the Fourier graph network (OCFGNet) based on the representation of working condition characteristics, which maps the input flight sequence to a high-dimensional graph space. This space can integrate temporal and spatial dynamic features in a unified manner. The Fourier graph operator adopts a shared weight parameter mechanism, which can effectively balance information from adjacent nodes in different propagation sequences, fully capturing the spatio-temporal dependencies in the graph structure. Moreover, this operator also had high sensitivity to complex working conditions, achieving efficient attention while reducing computational complexity. Ding et al. [8] constructed a refined time–frequency neural network (RTFNN)-interpretable model by designing a refined time–frequency convolution kernel (RTFCK) that embeds high-order phase operators, achieving efficient aggregation of fault feature information energy, significantly improving the interpretability and fault feature extraction ability of the model. Li et al. [9] proposed the Spatio-Temporal Physical Field Intelligent Perception Method (SGM) based on diffusion generative models, providing a new approach for aero-engine combustion diagnosis and regulation, and helping to deeply understand the dynamic evolution of the complex internal physical processes of the engine under different operating conditions.

When using the Fourier infrared spectrometer to detect the operating conditions of an aero-engine, based on the absorption characteristics of substances to infrared light, it analyzes the composition by emitting infrared light, interacting with the sample, and processing the signals through Fourier transformation. Its core application is in the analysis of exhaust gas components, for judging combustion efficiency and faults, monitoring the state of lubricating oil to warn of wear and contamination, and also assisting in analyzing the composition of wear particles to locate the fault. This technology has advantages such as simultaneous detection of multiple components, rapid response, and non-destructiveness. Wolak et al. [10] compared the analytical results obtained from two devices in order to quickly assess the quality of lubricating oil using infrared spectroscopy. This assessment was based on the changes in selected physical and chemical properties of engine oil that occurred during actual operation. The changes in physical and chemical properties such as oxidation degree, nitration degree, sulfonation degree, carbon content, basic value (TBN), and additive percentage content were analyzed in terms of direction and intensity. Based on the obtained results, the statistical relationship between the two alternative devices was thoroughly described. To achieve the rapid determination of the gasoline induction period, Liu et al. [11] innovatively proposed a new analytical method based on Fourier transform–attenuated total reflection infrared spectroscopy (ATR-FTIR), and constructed a dedicated analytical system integrating spectral measurement, data processing, display, and storage. This system deeply integrated the Fourier transform infrared spectrometer module with metrological software. The sample display accessory was particularly crucial, using a zinc

selenide (ZnSe) 9-fold reflection ATR crystal coated with diamond film combined with a stainless steel cover with a sealing device. This not only ensured a constant optical path but also significantly improved the convenience of sample injection and cleaning. Yin et al. [12], based on data from Fourier transform infrared (FTIR) spectroscopy, Spectral Oil Analysis (SOA), and other conventional methods, discussed the oil monitoring experiment for the propeller steering. The experiment showed that the FTIR spectroscopy method could obtain results quickly and easily through laboratory analysis, and combined with the oil analysis of the spectrometer, the complementary information was most effective for the condition monitoring of marine machinery. Mike et al. [13] used infrared spectroscopy to analyze the antioxidant content and total acid value of synthetic turbine oil for aero-engines. Two-dimensional infrared correlation analysis was used to study and interpret the trends observed in the spectra, because acids form in the oil and antioxidant substances are depleted, which is a function of aging and engine wear. Principal component and partial least squares algorithms were used and compared to develop calibration and prediction models.

An autoencoder is an unsupervised learning model based on neural networks. Its core objective is to automatically learn the latent feature representation of data through the compression–reconstruction process. Its structure is symmetrical, consisting of an encoder (Encoder) and a decoder (Decoder), and is commonly used for data dimensionality reduction, denoising, feature extraction, and generation tasks [14–16]. In the field of anomaly detection, the nonlinear dimensionality reduction autoencoder (Autoencoder) demonstrates unique advantages. Sakurada et al. [17] selected artificial data generated by the Lorenz system and real data from spacecraft telemetry as samples, processed them through the autoencoder for dimensionality reduction, and compared it with traditional linear principal component analysis (PCA) and kernel principal component analysis (kernel PCA) to deeply explore its performance characteristics. The autoencoder can sensitively capture subtle anomalies and successfully detect abnormal data points that linear methods cannot identify, significantly improving the sensitivity of anomaly detection. Additionally, by extending the autoencoder to a denoising autoencoder (Denoising Autoencoder), the model performance is further optimized, and the detection accuracy and robustness are improved. Compared with kernel principal component analysis, the autoencoder achieves nonlinear dimensionality reduction without the need for complex kernel function calculations, reducing computational complexity and improving algorithm efficiency. Gonzalez et al. [18] employed an unsupervised variational autoencoder (VAE) to analyze a set of FTIR spectral data from multiple iron ore deposit reverse circulation (RC) drill core samples in the Pilbara region of Western Australia, in order to identify any potential anomalies. Yang et al. [19] proposed a data-driven method based on FTIR data to predict the characteristics of crude oil. The autoencoder was used to learn a new representation form for the dimensionality reduction of FTIR data. The learned low-dimensional representation was input into SVR to predict the characteristics of crude oil.

The main contents of this work are as follows:

(1) A Fourier infrared spectrometer was utilized to conduct precise field measurements on the hot jet of a certain type of aero-engine, and the spectral data of the hot jet generated by the engine under different operating conditions was obtained.

(2) A dual-channel feature construction algorithm for spectral analysis is proposed. This algorithm consists of two branches: neighborhood integration and an autoencoder, which respectively extract the original spectral features and deep spectral features of the aero-engine hot jet, breaking through the limitations of traditional single-modal feature representation.

(3) A feature selection and fusion strategy based on physical significance is proposed. By designing a feature selection algorithm based on peak area and a cross-space feature fu-

sion optimization mechanism, the accuracy and interpretability of clustering analysis were improved. This lays a solid data foundation for subsequent studies on hot jet characteristics comparison and fault diagnosis.

The structure of this thesis is mainly composed of five parts. Section 2 reviews the current status of the motion characteristics of different operating conditions of aero-engines and the methods for feature extraction. It briefly introduces the methods, contributions, and framework of this paper. Section 3 introduces the field experiment design for the measurement of hot jet spectra of aero-engines and the structure and details of the dual-channel feature spectral analysis method. Section 4 elaborates on our experimental content and results and conducts a detailed analysis of the experimental results. Section 5 presents our discussion on the experimental results, analyzes the advantages and limitations of the research methods, and explores potential improvement directions. Section 6 provides a systematic summary of the entire paper.

2. Related Work

Spectral feature extraction [20,21] is a crucial step in extracting useful information from spectral data, reducing data dimensionality, and improving the efficiency of subsequent analysis. The methods for spectral feature extraction include two major categories: traditional and deep learning. Traditional methods directly utilize the original spectra or extract features based on physical meanings, such as the original spectral features with the original data appearance preserved, giving them low computational cost and strong interpretability. Statistical and physical features focus on key information to achieve effective dimensionality reduction, but both are limited by high-dimensional redundancy and complex sample processing capabilities [22,23]. In the field of biomedical mass spectrometry analysis, the Morris research team proposed a new method for extracting mass spectrometry data features by integrating translation-invariant wavelet transform with average spectral peak detection [24]. This method achieved feature extraction and quantitative analysis through average spectra, enhanced signal processing capabilities using translation-invariant wavelet transform, and completed high-precision peak detection based on average spectra. The study systematically verified the effectiveness of this method through case analysis and simulation experiments, fully demonstrating the technical advantages of average spectra in peak detection. Additionally, the team innovatively constructed a computer mass spectrometry model based on physical mechanisms, providing a more theoretically supported algorithm framework for mass spectrometry analysis. He et al. [21] used pulse eddy current (PEC) technology to achieve non-contact and non-destructive welding quality monitoring. Using PEC technology, they quantitatively detected laser-welded aluminum alloy structures with porosity and crack defects. They constructed a detection system to obtain the PEC signals of different defects in laser weld seams. They calculated characteristic parameters such as the peak, peak time, fundamental amplitude, peak and rising curvature ratio, fundamental and third harmonic amplitude ratio, and marginal spectral peak of the PEC signals, quantitatively representing the type and size of laser welding defects. They established a defect identification model based on a support vector machine (SVM), using input characteristic parameters to identify the type and depth of laser weld seam defects. Deep learning methods, with the help of models such as CNN [25,26], LSTM [27], and Transformer [28], can automatically learn spectral features, adaptively extract nonlinear patterns, efficiently denoise, and handle complex data. However, they have limitations such as high computational cost and strong data dependence.

Traditional spectral feature extraction methods still have several key deficiencies. On the one hand, conventional methods mainly rely on peak positions and peak intensities, making it difficult to capture the overall changes in spectral absorption and making them

insensitive to subtle concentration differences under different operating conditions. On the other hand, some feature extraction methods based on deep learning can achieve high classification accuracy. Moreover, most existing algorithms cannot simultaneously retain local spectral information and mine deep nonlinear relationships, making it difficult to obtain robust and discriminative features from precious and small sample spectral data.

To overcome these shortcomings, this paper proposes a dual-branch feature extraction framework that integrates peak area calculation and a deep learning autoencoder. The peak area branch enhances the discrimination of concentration differences by integrating local spectral information and strengthens the quantitative expression of spectral changes. The autoencoder branch, on the other hand, adaptively extracts deep nonlinear features without the need for manual design. The combination of the two not only retains the clear physical interpretability brought by the peak area but also possesses the powerful feature expression ability of the autoencoder, thereby enabling more effective and reliable feature extraction for engine exhaust spectra under different operating conditions.

3. Method

3.1. Data Collection

We conducted an outdoor field experiment to collect the hot jet data of the aero-engine. The data collection process is shown in Figure 1. The measurement distance range of the spectrometer from the aero-engine was 127–280 m. During the measurement, the outdoor temperature was 18–20 °C, and the humidity was 19–73% Rh. The measurement instrument used was a Fourier infrared spectrometer, and the measurement mode was passive mode, requiring no external light source. The spectral resolution was 1 cm^{-1}, capable of resolving the fine rotational–vibrational spectra of molecules such as CO_2 and H_2O; the spectral measurement range was 2.5–15 μm, fully covering the characteristic absorption bands of the combustion products. The full view angle could reach 1.5°. The FTIR spectrometer is made in Beijing, China. During data acquisition, 30 sets of spectral data were acquired under each operating condition, measurement repeatability was assessed by computing the standard deviation of the spectral data, and radiometric calibration was conducted using a standard blackbody to ensure the accuracy and comparability of the spectral intensity. The software used for data processing is based on Python 3.9.



(**a**) (**b**)

Figure 1. Data acquisition (**a**) Schematic diagram of the field experiment scene. (**b**) Photograph of the experimental site.

Radiometric calibration of the Fourier transform infrared spectrometer is essential to ensure spectral data accuracy and reliability. This process involves precise calibration of

instrument wavelength, radiometric intensity, and other critical parameters through blackbody radiation reference measurements. The brightness temperature calculation follows Planck's radiation law [29], establishing traceability to fundamental thermodynamic principles.

$$\mathrm{T}(v) = \frac{hcv}{kln\{[L(v) + 2hc^2v^3]/L(v)\}} \quad (1)$$

where h represents the Planck constant, $h = 6.62607015 \times 10^{-34}$ J·s; c represents the speed of light, $c = 2.998 \times 10^8$ m/s; v represents the wave number, with the unit of cm^{-1}; k represents the Boltzmann constant, $k = 1.380649 \times 10^{-23}$ J/k; and $L(v)$ represents the radiant flux of a unit beam.

In the field experiment, we collected data during both the acceleration and non-acceleration states of the aero-engine. In this paper, the acceleration state is referred to as State 1, and the non-acceleration state as State 2. The total number of collected sample data points is 128. Among them, there are 56 samples in State 1 (acceleration) and 72 samples in State 2 (non-acceleration). We selected the first samples from the original data collected in each state and the background for presentation, as shown in Figure 2. To address the interference from the background environment, such as the atmosphere, we simultaneously collected the background spectra at the corresponding moments while collecting the spectral data of the hot jet. In the experiment, by differentially processing the original hot jet spectra and the background spectra, we effectively eliminated the environmental interference and thus obtained the pure hot jet spectra that could truly reflect the engine's operating conditions.

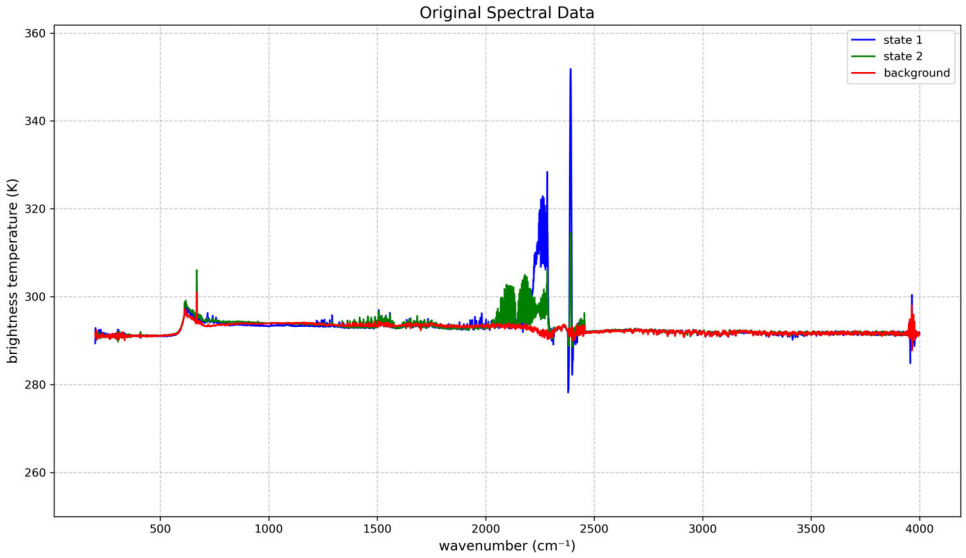


Figure 2. Original data collected.

3.2. Design of NAIDN Feature Construction Algorithm

The NAIDN feature construction algorithm adopts a dual-branch parallel architecture, including a neighborhood integral branch and an autoencoder branch. For the hot jet spectral data of aero-engines, it simultaneously realizes the parallel extraction of the original spectral features and the deep spectral features, effectively breaking the limitations of traditional single-modal feature representation. It not only retains the physical interpretability of spectral analysis, but also captures the implicit multi-parameter coupling information in the hot jet spectral data through the representation learning ability of the autoencoder, providing a more comprehensive feature basis for engine state recognition under complex operating conditions.

3.2.1. Neighborhood Integral Branch

The neighborhood integral branch, as the original spectral feature extraction module in the dual-channel feature construction framework, is designed closely around the physical essence of spectral analysis. Through the sliding window integral algorithm, it achieves efficient capture of key physical features in spectral data and retains their interpretability. Since the absorption of substances by light does not occur at a single precise wave number point but stems from the broadening effect of molecular vibration or rotational energy levels (such as doppler broadening caused by thermal motion, collision broadening, etc.), the actual absorption features manifest as a continuous increase in intensity within a certain wave number range, rather than the infinitely narrow single peak in the ideal model. As shown in Figure 3, the neighborhood integral converts the intensity information of discrete wave number points into local energy aggregation features with physical interpretability by simulating the actual shape of spectral absorption peaks.

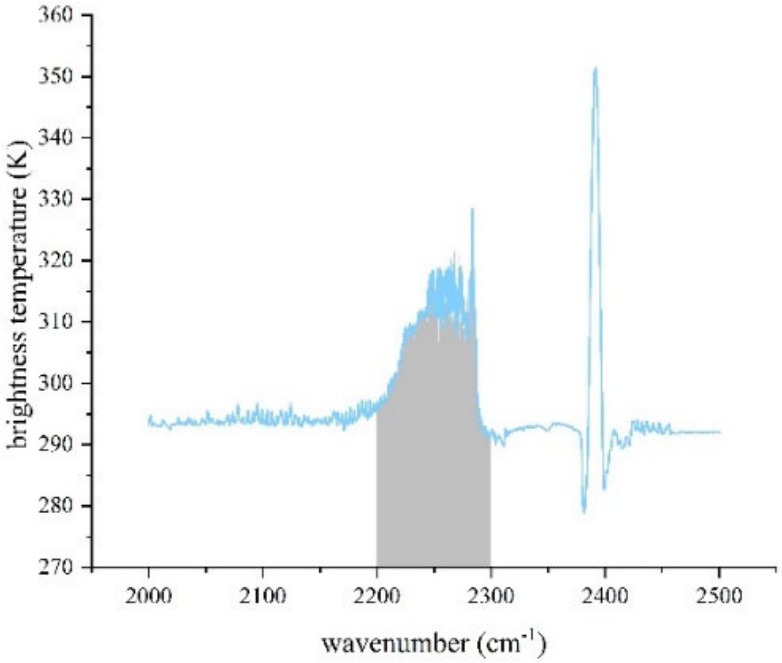


Figure 3. Neighborhood integral diagram.

The neighborhood integration is computed via a sliding summation window, which approximates localized energy distribution around each spectral wavenumber position. Physically, this metric quantifies the absorption band area corresponding to specific molecular functional groups or chemical bonds. Consistent with the Lambert–Beer law [30], these integrated areas demonstrate linear proportionality to analyte concentrations.

$$A = k \cdot l \cdot c \tag{2}$$

where A represents the absorbance, corresponding to the adjacent integral value; k represents the absorption coefficient; l represents the optical path length; and c represents the concentration of the absorbing substance.

Our spectral data is represented by each row, and for each sample's spectral signal, the following operation needs to be performed to extract the original spectral features of each sample.

Step 1: Calculate the neighborhood integral of a single sample: For the spectral signal of the sample $S_m = [S_m(1), S_m(2), \ldots, S_m(N)]$, N represents the total number of wave numbers. By summing through a local sliding window, the approximate integral area near

each wave number point is calculated, reflecting the total energy of that region. For the i-th wave number point in the spectrum S, the neighborhood integral A_i is defined as:

$$A_m(i) = \sum_{k=max(1,i-k)}^{min(N,i+k)} S_m(k) \tag{3}$$

where k represents the wavenumber on the left side of the window.

Step 2: For M samples, calculate the neighborhood integral matrix of all samples $A \in R^{M \times N}$:

$$A = \begin{bmatrix} A_1(1) & A_1(2) & \dots & A_1(N) \\ A_2(1) & A_2(2) & \dots & A_2(N) \\ \vdots & \vdots & \ddots & \vdots \\ A_M(1) & A_M(2) & \dots & A_M(N) \end{bmatrix} \tag{4}$$

where $A_M(i)$ represents the integral value of the m-th sample at the i-th wave number point within its neighborhood.

Step 3: Select key absorption regions: By calculating the total peak area across multiple samples, identify the wave numbers that contribute the most to all samples. Sort $T(i)$ in descending order and select the key wavenumber points.

$$T(i) = \sum_{m=1}^{M} A_M(i) \tag{5}$$

where $T(i)$ represents the total peak area of all samples at the i-th wave number point.

Step 4: Based on the indexed wave number points, extract the corresponding brightness temperature from the original hot jet data of the aero-engine as the key feature for the neighborhood integration branch extraction F_1.

3.2.2. Autoencoder Branch

The autoencoder [31,32], an unsupervised neural network architecture, operates through an encoder–decoder framework to achieve nonlinear data compression and reconstruction. This model's principal advantage resides in its capability to autonomously learn latent feature representations, eliminating dependency on manual feature engineering while significantly enhancing extraction efficiency. As illustrated in Figure 4, the symmetrical architecture comprises complementary encoder and decoder networks that collaboratively extract hierarchical features through unsupervised parameter optimization.

The encoder nonlinearly embeds the input data into a low-dimensional latent space through manifold learning (Equation (6)), while the decoder reconstructs the original input from this compressed representation via inverse mapping (Equation (7)). This symmetrical architecture enables bidirectional translation between high-dimensional observations and their latent embeddings.

$$h = \sigma(Wx + b) \tag{6}$$

where x represents the high-dimensional input, W represents the weight matrix, b represents the offset, $\sigma(\cdot)$ represents the nonlinear activation function, and h represents the output of the encoder.

$$\hat{x} = \sigma\left(\widetilde{W}h + c\right) \tag{7}$$

where $\widetilde{W}$ is the weight matrix related to W, c represents the offset, h represents the input of the decoder, and $\hat{x}$ represents the output of the decoder.

As illustrated in Figure 5, our framework implements a denoising autoencoder as its hierarchical feature-learning component within the dual-channel architecture. The encoder nonlinearly maps high-dimensional spectral inputs into compressed latent embeddings through manifold learning, achieving unsupervised representation learning. This archi-

tecture learns complex nonlinear dependencies via multilayer perceptrons with ReLU activations, overcoming the limitations of conventional linear decomposition methods. Specifically, the denoising mechanism mitigates instrumentation noise through corrupted input reconstruction, while the hierarchical decomposition mechanism reveals multi-scale spectral patterns: shallow layers resolve localized absorption features and fine-grained details, whereas deeper layers synthesize global spectral characteristics through progressive abstraction.

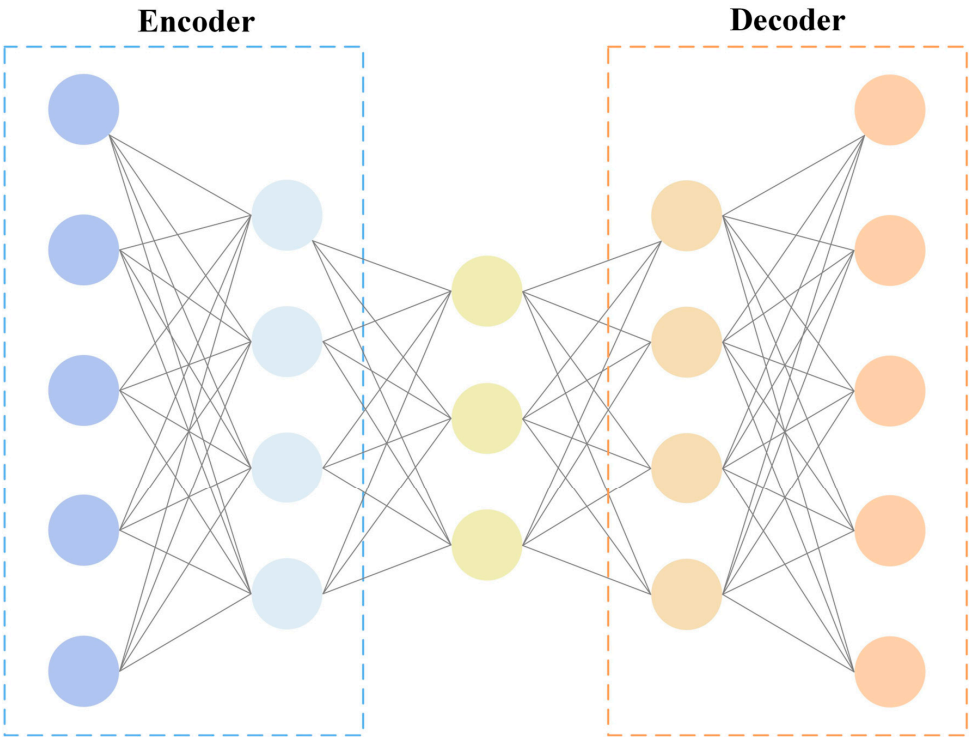


Figure 4. Structure of the autoencoder.

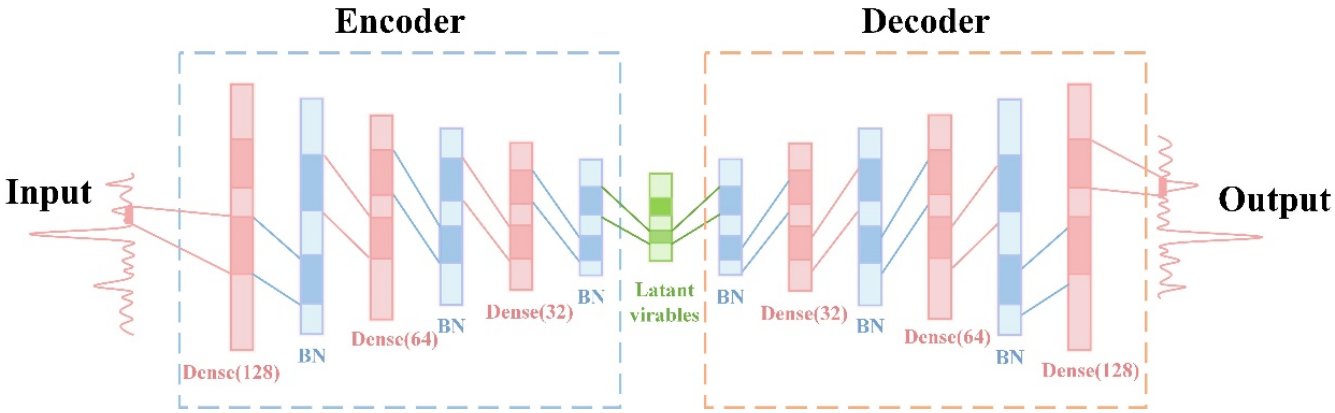


Figure 5. Network structure diagram of the autoencoder branch.

The input spectrum is $x \in R^n$, where n represents the total number of wave numbers. The encoder gradually achieves dimensionality reduction through three fully connected layers.

$$h_1 = ELU(W_1 x + b_1) \tag{8}$$

where $W_1 \in R^{128 \times n}$, $b_1 \in R^{128}$, and the output is $h_1 \in R^{128}$.

$$h_2 = ELU(W_2 h_1 + b_2) \tag{9}$$

where $W_2 \in R^{64\times128}$, $b_2 \in R^{64}$, and the output is $h_2 \in R^{64}$.

$$h_3 = ELU(W_3 h_2 + b_3) \tag{10}$$

where $W_3 \in R^{32\times64}$, $b_3 \in R^{32}$, and the output is $h_3 \in R^{32}$.

$$f = tanh(W h_3 + b) \tag{11}$$

where $W \in R^{d\times32}$, $b \in R^{d}$, and the output $f \in R^{d}$ refers to the extracted deep spectrum.

Batch normalization layers are implemented after every encoder fully connected layer to mitigate internal covariate shift induced by parameter updates during training. This technique standardizes layer inputs to zero-mean and unit-variance distributions through scale-variance transformation (Equation (8)), significantly improving the model convergence rate and generalization performance. For M aero-engine hot jet spectral samples, the resulting deep feature representation matrix F_2 encodes nonlinear manifold structures through hierarchical abstraction, where d denotes latent space dimensionality.

The training process of the encoder model is constrained by the mean square error loss function, which can be expressed as:

$$L = \frac{1}{N}\sum_{i=1}^{N} \parallel x_i - \hat{x_i} \parallel^2 \tag{12}$$

where N represents the number of samples, and x_i represents the i-th spectral data.

3.2.3. Feature Fusion

In the data processing procedure of the hot jet of an aero-engine, in order to achieve the effective integration of multi-dimensional information, this paper integrates the features extracted by the autoencoder branch and the neighborhood integration branch. The autoencoder branch, with its unique encoding–decoding structure, can capture the global low-dimensional representation of the data and extract the core features of the data, while the neighborhood integration branch focuses on local physical features and accurately extracts key information such as the original brightness temperature values of the aero-engine hot jet. After the integration of the two, a comprehensive feature matrix F containing the global low-dimensional representation and local physical features is formed, achieving a comprehensive coverage of the data features and laying a solid foundation for subsequent analysis.

$$F = \begin{bmatrix} F_1 \\ F_2 \end{bmatrix} \tag{13}$$

During the process of extracting features through the autoencoder branch, after the data is processed by the tanh activation function, its value range is limited within the range of $[-1, 1]$. In contrast, the original brightness temperature values of the hot jet of the aero-engine extracted by the neighborhood integration branch have a significantly different numerical range. Since the subsequent clustering algorithm is relatively sensitive to the consistency of the feature value range, features with different value ranges may lead to deviations in the clustering results, affecting the accuracy and reliability of the analysis. Therefore, to avoid the negative impact of the difference in feature value ranges on the clustering algorithm, this paper normalizes the integrated feature matrix F after fusion, uniformly mapping each dimension of the features to the same numerical interval, eliminating the dimensional differences and value range deviations among the features, ensuring that the clustering algorithm can perform accurate calculations based on standardized feature data, and improving the accuracy and effectiveness of data analysis.

4. Experiments and Results

The overall experimental process of this article is shown in Figure 6. The experiment is mainly divided into three parts: acquisition of hot jet spectral data of the aero-engine, feature extraction, and clustering.

4.1. Feature Construction Experiment

When processing the spectral data of an aero-engine's hot jet acquired from outdoor field experiments, the model first conducts preprocessing on the raw spectral data. Specifically, normalization is applied to ensure the consistency and stability of the data, thereby laying a solid foundation for accurate feature extraction in subsequent steps. In the experiment, since the resolution of the hot jet of the aero-engine is 1 cm^{-1}, the sliding window width of the neighborhood integral branch is set to 2, and the total length of the sliding window is 5. This is sufficient to capture the complete shape of the local absorption peak and does not cause feature blurring due to an overly large window. The step size is set to 1, and point-by-point sliding can retain the local information of all wave number points. The essence of the sliding window integration is the accumulation of local energy. The neighborhood integral value of a certain wave number point in the hot jet spectrum corresponds to the total radiative energy of molecular vibration or rotational energy level transitions near that wave number. The higher the integral value, the higher the molecular concentration and temperature in that area, which is directly related to the physical state of the hot jet. The neighborhood integration branch module works by integrating the brightness temperature information of the raw spectral data, from which it extracts 15 dimensional key original spectral features. Meanwhile, the autoencoder branch is dedicated to extracting the deep spectral features of the aero-engine's hot jet. Through three layers of dimensionality reduction operations, this branch compresses the high-dimensional spectral data into a 15-dimensional feature space, enabling more-efficient subsequent analysis.

Under two distinct operating conditions, the spectral curves of the aero-engine hot jet exhibit a highly consistent variation trend, with significant consistency in their constituent substances. The primary difference between the two lies in the content ratio of each substance, a characteristic that provides a key basis for subsequent operating condition differentiation. Notably, the peak intensity of CO_2 within the wavenumber range of 2387–2394 cm^{-1} is extremely high. This intense peak stands out prominently in the spectral graph and produces a significant masking effect: other characteristic peaks preceding the CO_2 peak have relatively weak signal intensities, which are easily obscured when contrasted with the strong CO_2 peak. This obscuration makes it challenging to accurately identify and analyze the substance information and associated characteristics represented by these weaker peaks. The neighborhood integration method demonstrates unique advantages in addressing this issue. Its implementation involves setting a local sliding window around each wavenumber point in the spectrum and summing the spectral intensity values within the window. Unlike methods that focus solely on the intensity of individual wavenumber points, this approach comprehensively considers the spectral information in the vicinity of each point. Specifically, it can integrate and amplify the information of the previously weak characteristic peaks that were masked by the strong CO_2 peak. Through this integration process, even those characteristic peaks that were submerged by the strong peak (and previously difficult to distinguish) have their embedded information preserved and presented in the form of neighborhood integration values. Ultimately, this method enables the effective extraction of the weaker characteristic peaks that appear before the strong CO_2 peak, providing a more robust means for the comprehensive and accurate analysis of the aero-engine hot jet's spectral characteristics, and for identifying the various material components it contains.

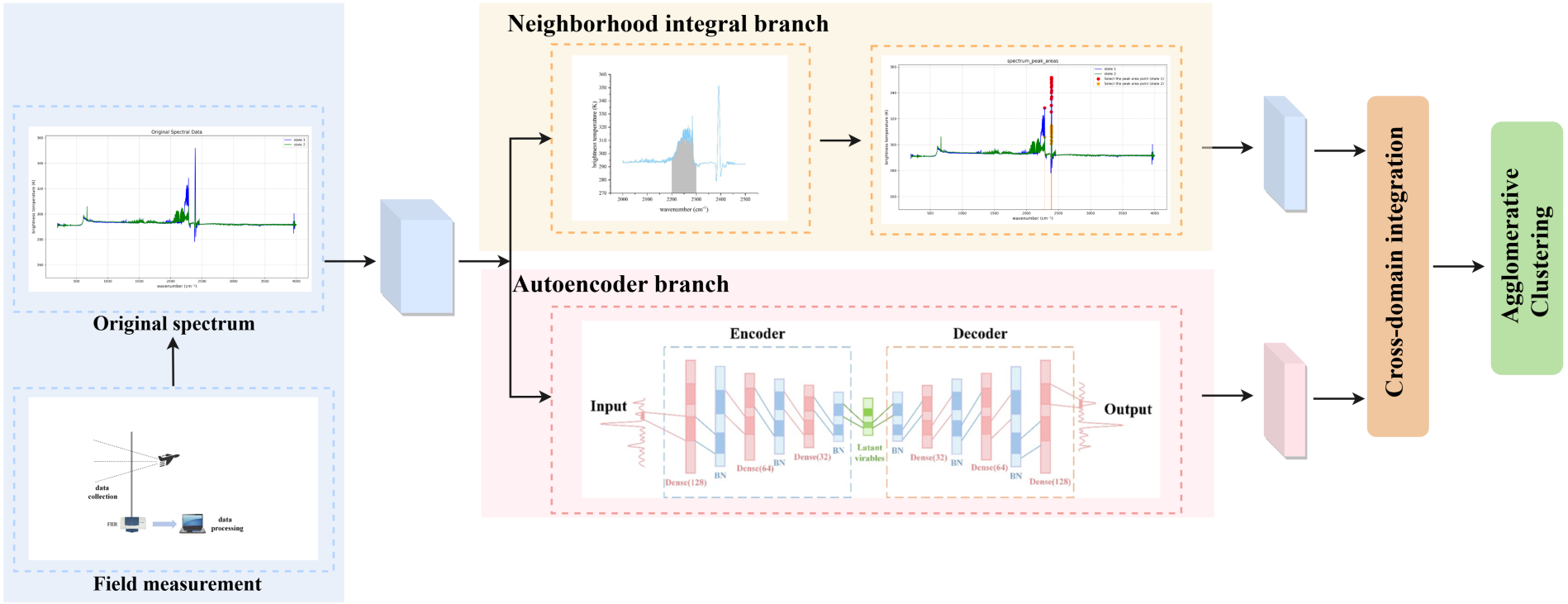


Figure 6. Experimental flowchart.

Based on the characteristic wavenumber point indices selected by the neighborhood branch module, the results of the extracted original spectral features are presented in Figure 7. Under the two distinct operating conditions, the hot jet spectra exhibit significant characteristics at specific wavenumber positions. Specifically, a clear absorption peak is observed at the wavenumber of 2283 cm^{-1}, a feature corresponding to compounds containing -C≡N. The presence of such cyanide-containing compounds confirms the existence of specific chemical components in the engine's hot jet. Cyanides are relatively common in organic synthesis chemistry; their origin in the hot jet can be attributed to two main pathways: either as reaction products of certain nitrogen-containing organic components in the fuel during high-temperature combustion, or as products of further conversion of some combustion intermediates. Notably, the detection of cyanide-containing compounds holds important indicative value for evaluating three key aspects of the engine combustion process: the chemical reaction pathways involved, combustion efficiency, and pollutant generation. This information thus provides critical insights for in-depth analysis of the engine's combustion performance and environmental impact.

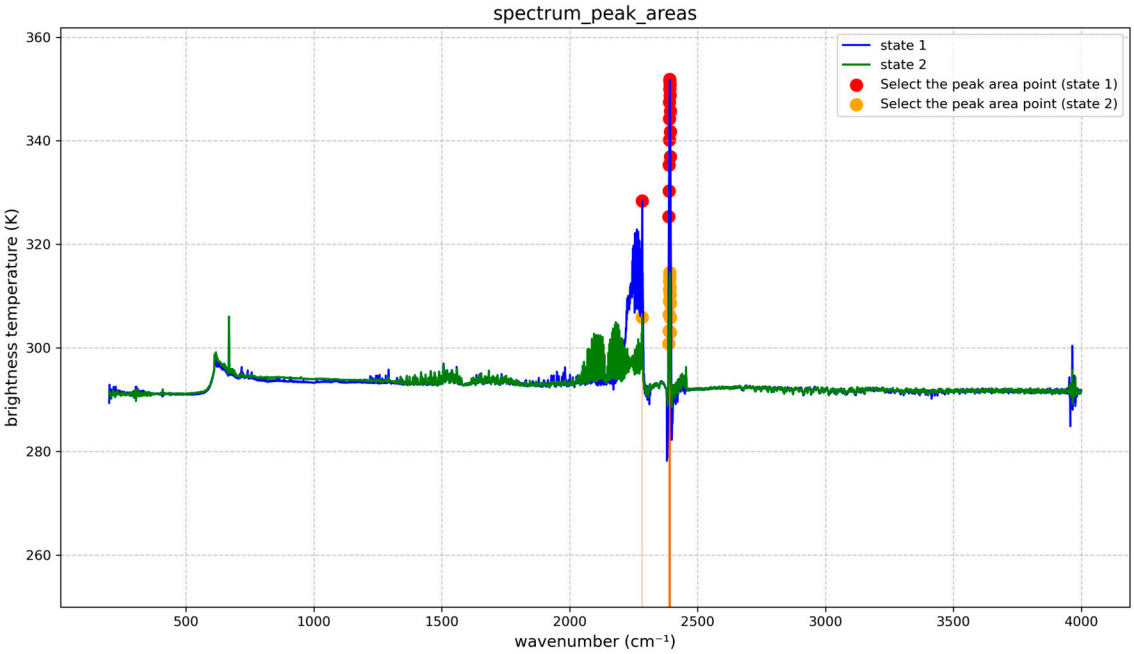


Figure 7. Graph showing results after neighborhood integration.

A similar prominent spectral feature is also observed within the wavenumber range of 2387–2394 cm^{-1}, which corresponds to CO_2—one of the most common byproducts of combustion. During the combustion of an aero-engine's fuel, the carbon–hydrogen components in the fuel react with oxygen in the air, generating large quantities of CO_2. The absorption peaks appearing in this wavenumber range directly reflect the concentration and existing state of CO_2 in the hot jet. By analyzing parameters such as the intensity of these absorption peaks, we can further assess the completeness of the engine's combustion process. For instance, abnormal absorption peak intensity may indicate an oxygen-rich or oxygen-deficient condition during combustion—both of which can affect the engine's performance and emission characteristics. Notably, combining the spectral characteristics of cyanide-containing compounds (at 2283 cm^{-1}) with those of CO_2 (at 2387–2394 cm^{-1}) provides critical spectroscopic evidence for in-depth research into three core aspects of aero-engines under different operating conditions: combustion mechanisms, pollutant generation pathways, and emission patterns.

Figure 8 presents the t-SNE visualization of the feature extraction results obtained by the model. From this visualization, we can preliminarily conclude that the fused features are capable of distinguishing the hot jet spectral data of the aero-engine under different operating conditions to a certain extent. This observation confirms the effectiveness of the method that combines autoencoder-derived features with neighborhood integration-based features—specifically, it demonstrates that the method has successfully captured the discriminative information in the spectral data, which is critical for distinguishing between different operating conditions.

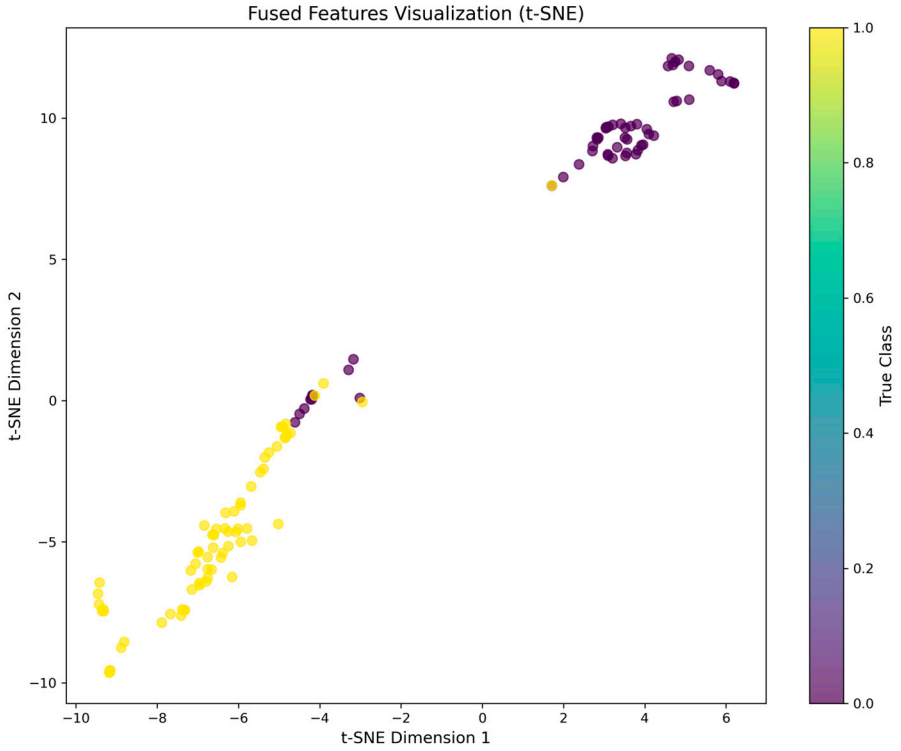


Figure 8. Visualization of feature extraction.

Figure 9 shows the 95% confidence interval comparison chart of the two types of samples at the extracted characteristic wave number points, indicating the average intensity differences and statistical stability of these two types of samples at these characteristic points. The intensity change trends of the two types of samples are highly synchronous: both reach a peak around 2391.2 cm^{-1} and then gradually decrease. This indicates that the acceleration treatment did not change the overall peak shape of the spectrum, but instead enhanced the overall absorption intensity.

Figure 10 shows the SHAP plots for the neighborhood branch and the autoencoder branch. The important features are mainly distributed within the range of 2283–2395 cm^{-1}. The neighborhood branch pays more attention to the feature at the position of 2392.1 cm^{-1}, while the autoencoder branch focuses more on the feature at 2389.2 cm^{-1}. This difference in feature focus points not only verifies the reliability of the 2283–2395 cm^{-1} range as a discriminative basis, but also reflects the complementarity of the two feature extraction strategies: the neighborhood branch focuses more on the local integral features of the spectrum, while the autoencoder branch focuses more on the abstract features encoded by a single point, providing interpretability support for the effectiveness of the subsequent fusion of features.

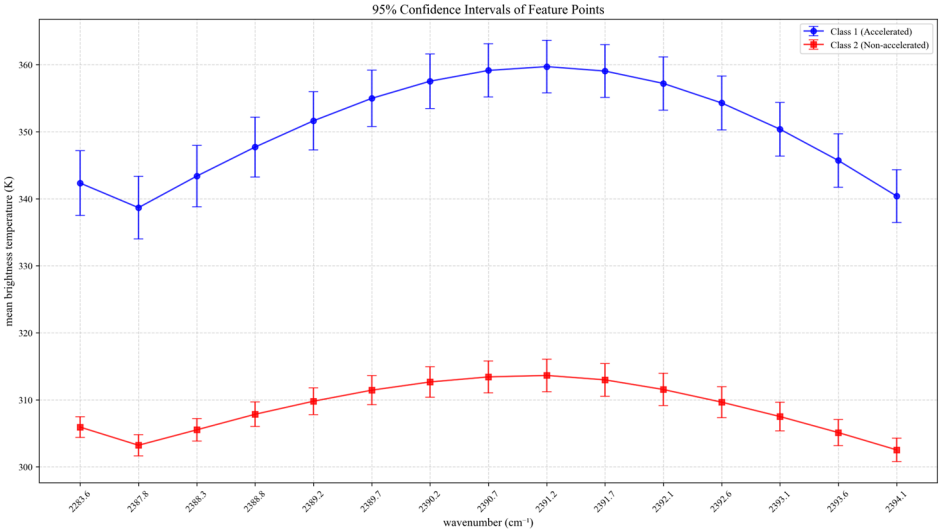
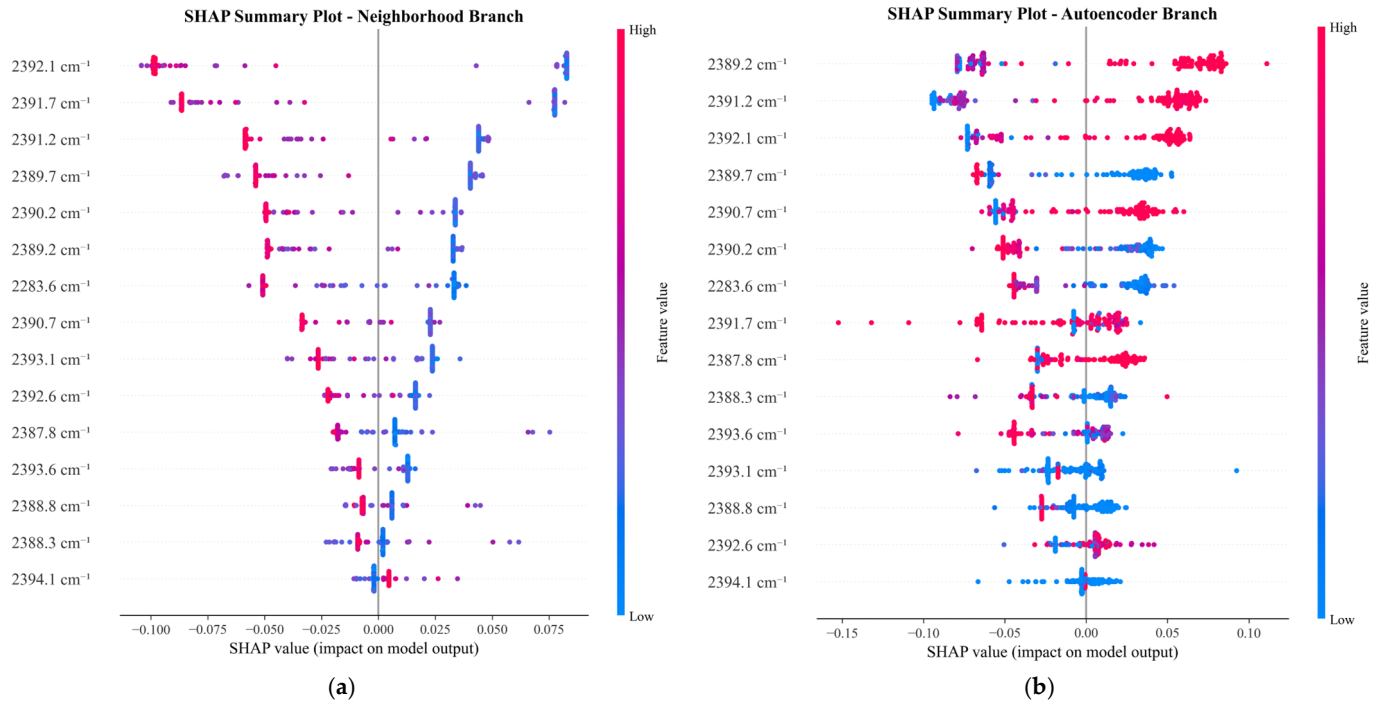


Figure 9. Confidence intervals of the two feature extraction branches.



(**a**) (**b**)

Figure 10. SHAP plots of the two feature extraction branches. (**a**) The results of the neighborhood branches; (**b**) The result of the autoencoder branch.

4.2. Validation Experiments for the Construction of Features

After extracting key original spectral features and key deep spectral features from the FTIR spectral data of the aero-engine wake flow via the proposed model, a multi-dimensional feature vector is constructed. To validate the effectiveness of this feature vector, an unsupervised clustering algorithm was employed for verification. These features of each sample are combined into a vector, thereby constructing a multi-dimensional feature matrix $F = \{f_1, f_2, \cdots, f_N\}$, where N represents the number of samples, and f_N is the multi−dimensional feature vector of the i−th sample. The extracted multi−dimensional feature matrix $F = \{f_1, f_2, \cdots, f_N\}$ is input into the Agglomerative Clustering model for clustering.

Agglomerative Clustering is a clustering algorithm based on the agglomerative hierarchical clustering method [33]. The core idea of this algorithm is to construct a clustering

hierarchy structure through a bottom-up strategy [34]. The algorithm starts with the finest granularity, treating each data point in the dataset as an independent initial cluster, and then continuously merges the clusters with the highest similarity (the closest distance) based on the pre-defined similarity measurement criterion until the preset stopping conditions are met, such as a specified number of clusters, a maximum distance threshold between clusters, or meeting a certain homogeneity standard. In practical applications, the methods for measuring the similarity between clusters are diverse. Common ones include Euclidean distance (measuring spatial geometric distance) and cosine similarity (evaluating the consistency of vector directions). In this study, we adopted the variance minimization-based method to measure inter-cluster distance. This method works by minimizing the dispersion of data points within each cluster, thereby enhancing the compactness (i.e., tightness of data distribution within clusters) and homogeneity (i.e., consistency of data attributes within clusters) of the final clustering results.

Step 1: Firstly, each sample is independently classified into a cluster. The merging cost is calculated based on the cluster centroid and the number of samples. Through sample quantity weighting, the influence during the merging of large clusters is balanced. At the same time, the difference between clusters is characterized by the square of the centroid distance. The smaller the value, the closer the data distribution after the clusters are merged, indicating a better clustering effect.

$$d(C_i, C_j) = |C_i| \cdot |C_j| \cdot \frac{1}{|C_i| + |C_j|} d^2(\mu_i, \mu_j) \tag{14}$$

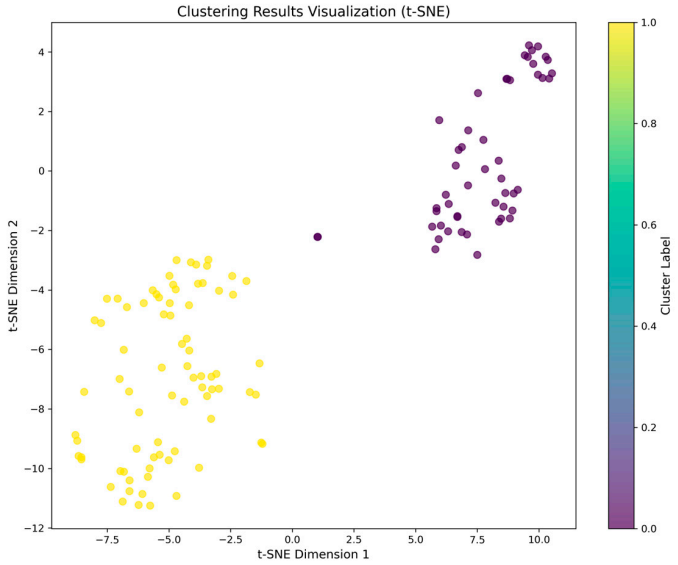
where $|C_i|, |C_j|$ represents the number of cluster samples, μ_i, μ_j represents cluster centroid, and $d(\mu_i, \mu_j)$ represents Euclidean distance.

Step 2: After calculating the merging cost for all $\frac{N \times (N-1)}{2}$ cluster pairs, select the pair with the lowest cost (C_a, C_b) for merging and generate a new cluster $C = C_a \cup C_b$. Assign weights based on the original cluster sample size to ensure that the new centroid position reasonably reflects the distribution characteristics of the merged data. The centroid of the new cluster is

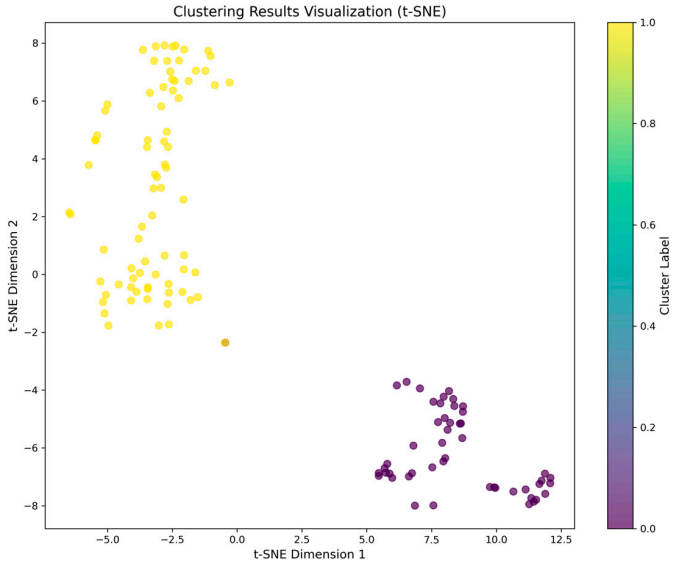
$$\mu = \frac{|C_a|\mu_a + |C_b|\mu_b}{|C_a| + |C_b|} \tag{15}$$

Step 3: Repeat Step 2 to achieve the Nth iteration, calculate the distances of the remaining $N - t + 1$ clusters. After each iteration, the number of clusters decreases by 1. As the merging process progresses, the algorithm continuously calculates the distances between the remaining clusters and dynamically adjusts the clustering structure. When the number of clusters drops to the preset value, the algorithm terminates.

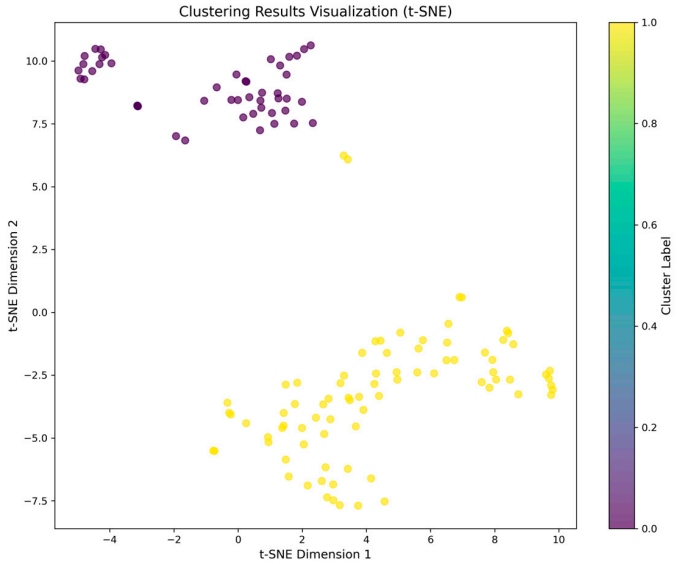
GMM [35] and kmeans [36] are also commonly used clustering algorithms. During the experiment, the feature vectors extracted by the proposed model were input into three distinct clustering models: Agglomerative Clustering, Gaussian Mixture Model (GMM), and kmeans. For visualization of the clustering results, t-SNE (t-distributed Stochastic Neighbor Embedding) was employed. Its core function is to convert the sample distance relationships in the high-dimensional feature space into a probability distribution in a low-dimensional space, while maximizing the preservation of the local neighborhood structure of the original data—this ensures that the relative proximity of samples in the low-dimensional visualization remains consistent with their relationships in the high-dimensional space. The visualization results of the clustering outcomes using t-SNE are presented in Figure 11.



(**a**)



(**b**)



(**c**)

Figure 11. Clustering results. (**a**) GMM; (**b**) kmeans; (**c**) Agglomerative Clustering.

We adopted the evaluation metrics of Accuracy [37,38], Precision [39,40], Recall [41,42], and F1 score [43,44]. Accuracy is defined as the ratio of correctly predicted samples (both true positives and true negatives) to the total number of samples. Precision refers to the proportion of actually positive samples among all samples predicted as positive (based on the model's output). Recall, by contrast, represents the proportion of correctly predicted positive samples relative to the total number of actual positive samples in the dataset. F1 Score is a core indicator in statistics for evaluating the prediction accuracy of binary classification models. Its main advantage lies in balancing Precision and Recall to comprehensively measure the overall performance of the model in identifying positive and negative samples. Specifically, it is represented by the harmonic mean of the two. Through the weighted harmonic mechanism, it avoids the model masking its true performance due to bias towards a certain type of prediction. The higher the value, the better the model achieves a balance between accurately identifying positive examples and reducing the omission of positive examples, and it is an important and practical standard for evaluating machine learning models. Table 1 shows the results of each clustering algorithm. It can be seen that Agglomerative Clustering performs outstandingly in the performance indicators of Accuracy, Precision, and F1 score, while kmeans and GMM perform relatively weakly. It can be seen from Figure 12 that the kmeans algorithm performs better in the Recall indicator.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{16}$$

$$Precision = \frac{TP}{TP + FP} \tag{17}$$

$$Recall = \frac{TP}{TP + FN} \tag{18}$$

$$F1\text{-}score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{19}$$

where TP represents true positive cases (where the prediction is correct and the actual outcome is positive); TN represents true negative cases (where the prediction is correct and the actual outcome is negative); FP represents false positive cases (where the prediction is incorrect and the actual outcome is negative); and FN represents false negative cases (where the prediction is incorrect and the actual outcome is positive).

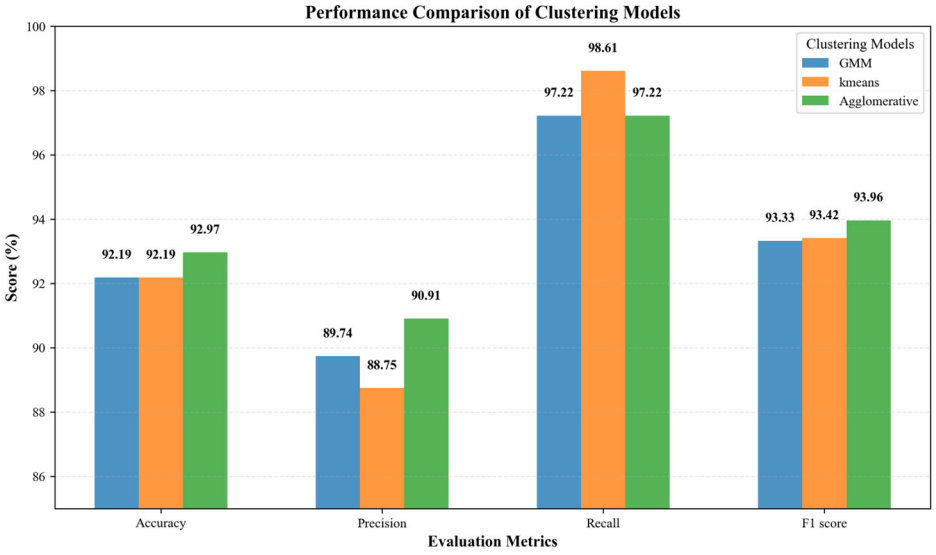


Figure 12. Clustering results.

Table 1. Results after clustering of each model.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
GMM	92.19	89.74	97.22	93.33
kmeans	92.19	88.75	98.61	93.42
Agglomerative Clustering	92.97	90.91	97.22	93.96

5. Discussion

This study focused on extracting the hot jet characteristics of an aero-engine under different operating conditions. We conducted the research using an on-site measurement method. To ensure data comparability, the experimental conditions strictly controlled the measurement environmental parameters (temperature, humidity, etc.) and geometric conditions (measurement distance, angle, etc.). A Fourier Transform Infrared spectrometer was used to collect spectral data from the central area of the engine's tail nozzle. Notably, the temperature of the engine's tail nozzle can reach 1500–2000 K, posing significant safety risks for direct measurement. In the on-site experimental environment, to safeguard the safety of measurement personnel and ensure the stable operation of the spectrometer, the instrument was deployed on the side of the engine. A measurement distance of 127–280 m was maintained between the hot jet source and the spectrometer [45]. While this setup of measurement angle and distance ensures operational safety, it inevitably leads to attenuation of the hot jet radiation signal. Specifically, during the transmission of the spectral signal through the atmospheric medium in the measurement path, the atmosphere exerts absorption and attenuation effects on the signal. However, in the described on-site data collection, the experimental distance was relatively short, and the temperature difference between the hot jet mixed gas and the background environment was approximately 300–400 °C. Given these conditions, the influences of the atmosphere and ambient factors are temporarily neglected in the analysis.

In the data analysis stage of the hot jet spectrum of the aero-engine, this paper focuses on the radiation brightness temperature spectrum in the 400–4000 cm^{-1} characteristic band. This band covers the characteristic absorption peaks of various key combustion products and contains abundant information about the combustion reaction. It is an important data window for revealing the essence of the combustion process. To deeply explore the characteristic information contained in the spectral data, the NAIDN feature extraction algorithm was designed in this paper. This algorithm innovatively adopts a dual-branch architecture, capturing the global features and local details of the spectral data through different branches, achieving hierarchical extraction of spectral features. One branch uses the local perception characteristics of the convolutional neural network to accurately locate the characteristic absorption peaks in the spectrum; the other branch uses the autoencoder architecture to learn the low-dimensional embedding representation of the spectral data and extract the intrinsic structural features of the data. The collaborative work of the two branches ensures the comprehensive extraction of complex features in the spectral data. As shown in Figure 10, the SHAP value distribution of the neighborhood branch is wider and has a larger absolute value (up to ± 0.075), indicating a stronger influence on the model's decision-making. Traditional spectral feature extraction methods mostly rely on peak intensity for analysis, while other neural network architectures, although achieving high classification accuracy, have features with weak interpretability. This study focuses on the spectral feature extraction of the same model under different operating conditions. Since the tail gas components of the same model are consistent, the differences mainly lie in the concentration of substances. To more accurately depict these concentration differences, this study uses peak area as the core feature and integrates the wavenumber neighborhood

to amplify the concentration differences between different operating conditions. In the experiment, we selected an appropriate half-width and step size of the sliding window, which not only retains the local information of all wavenumber points but also effectively enhances the sensitivity of the features to concentration changes.

After feature extraction is completed, the Agglomerative Clustering algorithm is used to verify the extracted features. This algorithm performs bottom-up hierarchical clustering to gradually aggregate similar spectral features, thereby achieving effective classification of spectral data under different operating conditions. The clustering results clearly show that in different operating conditions, the characteristic manifestations of cyanide compounds and carbon dioxide in the hot jet spectrum of the aero-engine are particularly prominent. Further analysis reveals that this characteristic difference may be the result of the combined effect of combustion thermodynamic conditions and chemical reaction kinetics paths under different operating conditions. Specifically, the thermodynamic parameters such as air–fuel ratio, temperature, and pressure in the combustion process directly affect the oxidation process of the fuel, and the significant characteristic of cyanide compounds reflects the degree of fuel pyrolysis and incomplete oxidation, and its concentration change is closely related to the combustion efficiency. While carbon dioxide is the main combustion product, its spectral signal enhancement is not only related to the generation amount but is also affected by subsequent secondary reactions and physical transmission processes. For example, the reverse water–gas shift reaction under high-temperature conditions leads to dynamic changes in the concentration of carbon dioxide, and the mixing process of the hot jet with the surrounding environment also changes its spatial distribution. Based on the above research findings, this provides new ideas and directions for engine combustion optimization. By adjusting the fuel injection strategy, such as adopting stratified combustion technology, a gradient distribution of fuel can be achieved, promoting the fine control of the combustion process; optimizing the airflow organization of the combustion chamber to enhance the mixing efficiency of fuel and air can effectively reduce the emission of incomplete combustion products and reduce the intensity of characteristic spectral signals, thereby achieving efficient and clean combustion of the aero-engine.

This study extracted multiple key features from the FTIR spectral data of an aero-engine under different operating conditions, including both the original spectral features and the deep spectral features. It also employed a feature selection and fusion strategy based on physical significance. This cross-space feature fusion can not only retain the physical interpretability of spectral analysis but also explore the inherent patterns of the data through deep learning. It is particularly suitable for complex hot jet scenarios influenced by multiple parameters, such as those of aero-engines, providing a more reliable feature basis for combustion state assessment and analysis of pollutant generation mechanisms. The current study is limited by the sample size. The current study is also limited by the model and operating condition dimensions covered by the samples. If we can obtain the hot jet multispectral data of all operating conditions of multiple aero-engines (covering small and large fan ratios, models, typical operating conditions such as takeoff at sea level, cruise, and high-altitude climb), we will use the existing feature extraction methods to explore more data features and thereby achieve the optimization design and intelligent diagnosis of aero-engines.

6. Conclusions

Aiming to address the problem of constructing the remote sensing FTIR spectral characteristics of the hot jet of a certain type of aero-engine under different working conditions, this paper proposes a remote sensing FTIR spectral characteristic construction algorithm for the fusion of original spectral features and deep spectral features of the aero-engine

hot jet. Firstly, the hot jet data of a certain type of aero-engine under different operating conditions is collected using the remote sensing Fourier transform infrared spectrometer, and the measured spectral data set is established; then, a construction algorithm integrating dual branches of energy features and deep learning features is proposed. This algorithm consists of two branches: the energy feature extraction branch and the deep learning feature extraction branch. The energy branch converts the radiation intensity values of discrete wavenumber points through the sliding window into local energy aggregation features, thereby accurately extracting the key physical information in the original spectrum. The deep learning branch adopts a three-layer fully connected neural network architecture of an autoencoder to explore the deep spectral features of the spectral data. The algorithms of the two branches not only retain the physical interpretability of spectral analysis, but also capture the multi-parameter coupling information hidden in the hot jet spectrum through the representation learning ability of the autoencoder, achieving feature fusion across spatial dimensions. Compared with traditional feature construction algorithms, the proposed dual-branch feature construction algorithm has stronger comprehensive representation ability. This experiment selects the unsupervised Agglomerative Clustering algorithm for research. The results show that the feature extraction strategy of this algorithm achieved a classification accuracy of 92.97% on the target classifier. This result successfully verifies the good effectiveness of the algorithm.

Author Contributions: Formal analysis, Y.L. (Yuntao Li).; investigation, Z.K. and Z.L.; software, Y.L. (Yurong Liao) and X.Y.; validation, Z.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Natural Science Foundation of China, grant number 62005320.

Data Availability Statement: The data presented in this study are available upon request from the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Vassilakis, E.; Besseris, G. The use of SPC tools for a preliminary assessment of an aero engines' maintenance process and prioritisation of aero engines' faults. *J. Qual. Maint. Eng.* **2010**, *16*, 5–22. [CrossRef]
2. Nan, G. Modeling and Dynamic Analysis of Shrouded Turbine Blades in Aero-Engines. *J. Aerosp. Eng.* **2016**, *29*, 04015021. [CrossRef]
3. Li, Y.F.; Lv, Z.; Cai, W.; Zhu, S.-P.; Huang, H.-Z. Fatigue Life Analysis of Turbine Disks Based on Load Spectra of Aero-engines. *Int. J. Turbo Jet-Engines* **2016**, *33*, 27–33. [CrossRef]
4. Cai, K.L.; Xie, S.S.; Wu, Y. Identification of aero-engines model based on T-S fuzzy model. *Tuijin Jishu/J. Propuls. Technol.* **2007**, *28*, 194–198. [CrossRef]
5. Sun, C. Operating Reliability Assessment for Aero-engine Based on Condition Monitoring Information. *J. Mech. Eng.* **2013**, *49*, 30.
6. Wang, J.; Song, Y.; He, T. A novel adaptive monitoring framework for detecting the abnormal states of aero-engines with maneuvering flight data. *Reliab. Eng. Syst. Saf.* **2025**, *258*, 110910. [CrossRef]
7. Liu, H.; Sun, Y.; Wang, X.; Wu, H.; Guo, Y.; Wang, H. Operating condition feature representation-based Fourier graph network for civil aircraft state estimation. *Reliab. Eng. Syst. Saf.* **2025**, *261*, 111085. [CrossRef]
8. Ding, J.; Wang, Y.; Qin, Y.; Tang, B. RTFNN: A refined time–frequency neural network for interpretable intelligent diagnosis of aero-engine. *Adv. Eng. Inform.* **2024**, *64*, 103048.
9. Li, Z.; Han, W.; Zhang, Y.; Fu, Q.; Li, J.; Qin, L.; Dong, R.; Sun, H.; Deng, Y.; Yang, L. Learning spatiotemporal dynamics with a pretrained generative model. *Nat. Mach. Intell.* **2024**, *6*, 1566–1579. [CrossRef]
10. Wolak, A.; Zaj, G. Changes in the operating characteristics of engine oils: A comparison of the results obtained with the use of two automatic devices. *Measurement* **2017**, *113*, 53–61. [CrossRef]
11. Liu, Y.F.; Yuan, H.F.; Song, C.F.; Xie, J.-C.; Li, X.-Y.; Yan, D.-L. Fast Determination of Induction Period of Motor Gasoline Using Fourier Transform Attenuated Total Reflection Infrared Spectroscopy. *Spectrosc. Spectr. Anal.* **2014**, *34*, 2929–2933.

12. Yin, Y.; Gao, H.; Yan, X.; Xiao, H. Application of the quantitative oil monitoring to analysing the operating condition of marine machinery. *Sci. China Ser. A* **2001**, *44*, 449–453.
13. Adams, M.J.; Romeo, M.J.; Rawson, P. FTIR analysis and monitoring of synthetic aviation engine oils. *Talanta* **2007**, *73*, 629–634. [CrossRef] [PubMed]
14. Xu, J.; Xiang, L.; Liu, Q.; Gilmore, H.; Wu, J.; Tang, J.; Madabhushi, A. Stacked Sparse Autoencoder (SSAE) for Nuclei Detection on Breast Cancer Histopathology Images. *IEEE Trans. Med. Imaging* **2016**, *35*, 119–130. [CrossRef]
15. Li, J.; Xia, C.; Chen, X. A Benchmark Dataset and Saliency-Guided Stacked Autoencoders for Video-Based Salient Object Detection. *IEEE Trans. Image Process.* **2017**, *27*, 349–364. [CrossRef]
16. Chhibra, S.S.; Chernyavskaya, N.; Maier, M.H.S. Autoencoders for real-time SUEP detection. *Eur. Phys. J. Plus* **2024**, *139*, 281. [CrossRef]
17. Sakurada, M.; Yairi, T. Anomaly Detection Using Autoencoders with Nonlinear Dimensionality Reduction. In *Proceedings of the MLSDA 2014 2nd Workshop on Machine Learning for Sensory Data Analysis*; ACM: New York, NY, USA, 2014.
18. Gonzalez, C.M.; Horrocks, T.; Wedge, D.; Holden, E.; Hackman, N.; Green, T. Anomaly detection in Fourier transform infrared spectroscopy of geological specimens using variational autoencoders. *Ore Geol. Rev.* **2023**, *158*, 14. [CrossRef]
19. Yang, S.B.; Moreira, J.; Li, Z. Predicting crude oil properties using fourier-transform infrared spectroscopy (FTIR) and data-driven methods. *Digit. Chem. Eng.* **2022**, *3*, 100031. [CrossRef]
20. Ta-Peng, T.; Rong-Ching, W. Application of artificial neural network on sound-signal recognition for induction motor. In *Proceedings of the National Science Council, Republic of China. Part A. Physical Science and Engineering*; National Science Council: Taipei, Taiwan, 1999; Volume 23.
21. He, K.F.; Liang, J.H.; Yong, J.F.; Shi, W.Q. Quantitative Detection of Laser Welding Defective Structure Based on Feature Exaction of the Pulsed Eddy Current Signal. *J. Mater. Eng. Perform.* **2023**, *32*, 6412–6422.
22. Sun, Z.; Li, Y.; Li, M.; Wang, N.; Liu, J.; Guo, H.; Li, B. Steel pickling rinse wastewater treatment by two-stage MABR system: Reactor performance, extracellular polymeric substances (EPS) and microbial community. *Chemosphere* **2022**, *299*, 134402. [CrossRef]
23. Ayalew, A.A.; Wodag, A.F. Extraction and Chromatographic Analysis of Ethiopian Oak Bark Plant for Leather Tanning Applications. *Chem. Afr.* **2023**, *6*, 1551–1560. [CrossRef]
24. Morris, J.S.; Coombes, K.R.; Koomen, J.; Baggerly, K.A.; Kobayashi, R. Feature extraction and quantification for mass spectrometry in biomedical applications using the mean spectrum. *Bioinformatics* **2005**, *21*, 1764–1775. [CrossRef] [PubMed]
25. Liu, Z.; Shen, C.; Fan, X.; Zeng, G.; Zhao, X. Scale-aware limited deformable convolutional neural networks for traffic sign detection and classification. *IET Intell. Transp. Syst.* **2020**, *14*, 1712–1722. [CrossRef]
26. Gan, C.; Yang, Y.; Zhu, Q.; Jain, D.K.; Struc, V. DHF-Net: A hierarchical feature interactive fusion network for dialogue emotion recognition. *Expert Syst. Appl.* **2022**, *210*, 118525. [CrossRef]
27. Sadeghian, E.; Dragomirescu, E.; Inkpen, D. Damage Detection for a Cantilevered Steel I-Beam through Deep-Learning Methods: LSTM, Multivariate Time-Series Transformer, and LSTM-Based Autoencoder. *J. Comput. Civ. Eng.* **2025**, *39*, e05324. [CrossRef]
28. Gu, Y.; Jin, F.; Zhao, J.; Wang, W. A hybrid lightweight transformer architecture based on fuzzy attention prototypes for multivariate time series classification. *Inf. Sci.* **2025**, *703*, 121942. [CrossRef]
29. Ade, P.A.R.; Aghanim, N.; Arnaud, M.; Arroja, F.; Ashdown, M.; Aumont, J.; Baccigalupi, C.; Ballardini, M.; Banday, A.J.; Barreiro, R.J. Planck 2015 results: XX. Constraints on inflation. *Astron. Astrophys.* **2016**, *594*, A20. [CrossRef]
30. Sassaroli, A.; Fantini, S. Comment on the modified Beer–Lambert law for scattering media. *Phys. Med. Biol.* **2004**, *49*, N255–N257. [CrossRef]
31. Vincent, P.; Larochelle, H.; Lajoie, I.; Bengio, Y.; Manzagol, P.A.; Bottou, L. Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion. *J. Mach. Learn. Res.* **2010**, *11*, 3371–3408.
32. Socher, R.; Huang, E.H.; Pennington, J.; Manning, C.D.; Ng, A. *Dynamic Pooling and Unfolding Recursive Autoencoders for Paraphrase Detection*; Curran Associates Inc.: Red Hook, NY, USA, 2011; pp. 801–809.
33. Michalakopoulos, V.; Sarmas, E.; Papias, I.; Skaloumpakas, P.; Marinakis, V.; Doukas, H. A machine learning-based framework for clustering residential electricity load profiles to enhance demand response programs. *Appl. Energy* **2024**, *361*, 122943. [CrossRef]
34. Balbi, E.; Cianfarra, P.; Crispini, L.; Tosi, S.; Ferretti, G. Hierarchical-agglomerative clustering analysis of geomorphic features applied to tectonic investigation of terrestrial planets: An example from Claritas Fossae, Mars. *Icarus* **2024**, *420*, 116197. [CrossRef]
35. Aldao, E.; Veiga-Lopez, F.; Gonzalez-Jorge, G.D. Enhancing UAV Classification with Synthetic Data: GMM LiDAR Simulator for Aerial Surveillance Applications. *IEEE Sens. J.* **2024**, *24*, 26960–26970. [CrossRef]
36. Chen, L.; Wang, K.; Li, M.; Wu, M.; Pedrycz, W.; Hirota, K. K -Means Clustering-Based Kernel Canonical Correlation Analysis for Multimodal Emotion Recognition in Human–Robot Interaction. *IEEE Trans. Ind. Electron.* **2023**, *70*, 1016–1024. [CrossRef]
37. Foody, G. Assessing the Accuracy of Remotely Sensed Data: Principles and Practices. *Photogramm. Rec.* **2010**, *25*, 204–205. [CrossRef]

38. Hashem, A.A.A.; Khalil, A.S.; Dabour, S.A.; El-Haig, W.M. Diagnostic accuracy of clinical examination, ultrasonography, and computed tomography in detecting and localizing posterior segment intraocular foreign bodies: A surgical correlation study. *J. Egypt. Ophthalmol. Soc.* **2025**, *118*, 296–301. [CrossRef]
39. Deffner, S. Towards enhanced precision in thermometry with nonlinear qubits. *Quantum Sci. Technol.* **2025**, *10*, 025009. [CrossRef]
40. Hu, Z.; Fan, X.; Zhao, Y.; Wu, W.; Liu, J. MHOE-DETR: A Ship Detection Method for Small and Fuzzy Targets Based on Satellite Remote Sensing Image Data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2025**, *18*, 20452–20468. [CrossRef]
41. Battut, A.; Ratovo, K.; Beaudouin-Lafon, M. OneTrace: Improving Event Recall and Coordination with Cross-Application Interaction Histories. *Int. J. Hum.-Comput. Interact.* **2024**, *41*, 3241–3258. [CrossRef]
42. Yan, X.; Li, Z.; Zhai, Y.; Liu, K.; Zhang, K.; Zhao, Z. LSKAFF-YOLO:Large Separable Kernel Attentional Feature Fusion Network for Transmission Tower Detection in High-Resolution Satellite Remote Sensing Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2025**, *18*, 19208–19222. [CrossRef]
43. Fujino, A.; Isozaki, H.; Suzuki, J. Multi-label Text Categorization with Model Combination based on F1-score Maximization. In Proceedings of the Third International Joint Conference on Natural Language Processing: Volume-II, Hyderabad, India, 7–12 January 2008.
44. Diallo, R.; Edalo, C.; Awe, O.O. *Machine Learning Evaluation of Imbalanced Health Data: A Comparative Analysis of Balanced Accuracy, MCC, and F1 Score*; Springer: Cham, Switzerland, 2025.
45. Golyak, I.; Glushkov, V.; Gylka, R.; Vintaykin, I.; Morozov, A.; Fufurin, I. Quantitative Remote Sensing of Sulfur Dioxide Emissions from Industrial Plants Using Passive Fourier Transform Infrared (FTIR) Spectroscopy. *Environments* **2026**, *13*, 61. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.