Journal of Research in Science and Engineering (JRSE)
ISSN: 1656-1996 Volume-8, Issue-8, August 2026
DOI: 10.66069/ojspub.16560807

Input Feature Purification Method for Engine Fault Prediction Based on Physical Prior FMA Analysis

Bangwei Li, Yu Han*

College of Mechanical Engineering, Tianjin University of Science & Technology, Tianjin 300222, China
**Correspondence Author, hanyu@tust.edu.cn*

Abstract: ***To address the issues including redundant measured engine data, low feature interpretability and incompatibility with neural network modeling, this paper proposes an improved PP-FMA feature purification method based on physical priors. Conventional Failure Mode Analysis (FMA) and purely mathematical analysis both have prominent drawbacks: strong subjectivity in feature screening and severe feature redundancy. This paper establishes a coupled PP-FMA analysis system integrating engine physical mechanisms and mathematical models to pre-optimize raw monitoring data and select features with low redundancy and high fault discrimination. Experimental results demonstrate that the proposed method can reduce feature dimensions, strengthen fault-sensitive information and greatly improve the quality of input features. It can provide a stable and reliable data preprocessing scheme for fault prediction of natural gas engines.***

Keywords: FMA, Natural gas engine, Cylinder head failure, Data processing, Feature engineering.

1. Introduction

In recent years, neural network-based fault diagnosis has become the mainstream technical route for fault identification of heavy-duty natural gas engines [1].

Zhongxin Shi, Hongzi Fei and Liuping Wang used EKF-based state-space models to estimate diesel engine instantaneous rotational speed and suppress measurement noise. However, EKF brings intrinsic linearization errors, performs poorly under strong nonlinearity, involves heavy computation, and relies on Gaussian noise assumptions [2].

Sławomir Szrama developed a multimodal spatio-temporal graph attention network with uncertainty quantification for aero-engine remaining useful life prediction to capture sensor spatial correlations and temporal degradation characteristics. Yet its complicated structure, numerous parameters and time-consuming uncertainty calculation restrict onboard edge deployment [3].

Ahmad Junaid, Abid Iqbal1 and Abuzar Khan built a hybrid feature selection and imbalance-aware learning model for aero-engine fault prediction using the CMAPSS dataset to boost prediction accuracy. However, without physical mechanism support, the method generalizes poorly for rare fault samples and suffers from high computation cost [4].

Therefore, the raw sensor data collected by vehicle-mounted terminals are contaminated with significant noise, redundant parameters and interferences under various operating conditions. If such data are simply filtered or processed by other means and then directly fed into the network, the training efficiency of the model will be reduced. Moreover, the model lacks physical mechanism support and suffers from low interpretability [5]. Conventional Failure Mode and Effects Analysis (FMA) only screens fault logic through qualitative approaches without unified mathematical quantitative screening criteria. Its feature selection heavily relies on manual experience, which makes it difficult to automatically generate standard inputs applicable to deep learning [6].

To address the above problems, this paper takes natural gas engines as the research object and constructs a physical prior PP-FMA feature preprocessing framework integrating physical mechanisms and mathematical quantification [7-8]. The correspondence between cylinder head structures and typical failure modes is disassembled layer by layer via boundary diagrams, function trees and parameter diagrams. By integrating physical mechanisms such as high-cycle and low-cycle fatigue damage with mathematical models including root-mean-square quantification and FFT spectrum analysis, dimensionality reduction and feature engineering construction of raw time-series data can be realized. This can provide high-quality data inputs for engine fault prediction models.

2. Research Platform and Research Object

2.1 Research Platform

Figure 1: A certain type of natural gas engine

As shown in Figure 1, the test vehicle in this paper is a natural gas engine installed on a heavy-duty tractor. This model is widely used in freight transportation scenarios. Cylinder head cracking and wear faults frequently occur under long-term variable load and alternating cold-hot working conditions, which endows it typical engineering research value.

2.2 Research Object

To clarify the research content, a total of 64 recorded engine faults have been collected. The engine has 64 fault samples in total, which fall into three main fault categories, as shown in Table 1.

Table 1: Engine fault types

Fault No.	Fault Type	Sample Quantity
A	Cracks between exhaust valves of cylinder head	35
	Cracks around the spark plug position of cylinder head	16
	Cracks between intake valves on the cylinder head	4
	Cracks in the exhaust passages of the cylinder head	1
B	Sand holes on cylinder head cover	2
C	Valve seat wear	2

It can be seen that all fault samples occur on the cylinder head and its attached structures. Therefore, the subsequent PP-FMA analysis of the engine will focus mainly on the engine cylinder head system.

3. Methodology

3.1 PP-FMA Framework

Conventional FMA mainly relies on qualitative mechanism analysis, and fault tracing is completed through boundary diagrams and functional trees. It can only generate textual conclusions of failure logic, with two critical drawbacks. First, it lacks quantitative calculation modules. The screening of sensitive parameters and judgment of fault severity heavily depend on engineers' experience, resulting in low standardization. Second, it cannot directly output structured numerical features. Raw monitoring data requires additional manual preprocessing before being applied to neural network modeling, which leads to cumbersome workflows and high risks of subjective errors.

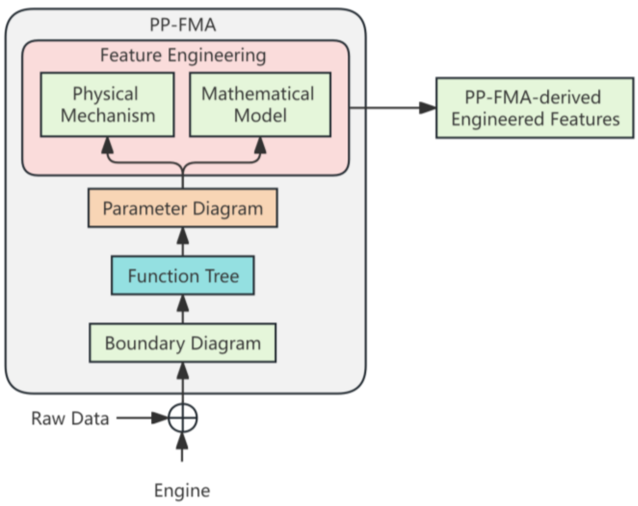

Figure 2: PP-FMA framework based on physical priors

This paper extends the traditional FMA process. On the basis of the original three-stage physical analysis chain of Boundary Diagram, Function Tree and Parameter Diagram, a feature engineering module integrating physical mechanisms and mathematical models is added. Thereby constructing a brand-new PP-FMA framework, as shown in Figure 2.

Raw engine data is first processed to define the system scope via boundary diagrams. Function trees are adopted to decompose failure modes, and parameter diagrams are used to screen parameters associated with faults. The screened parameters are then imported into two sub-modules: physical mechanism sub-module and mathematical model sub-module. The physical sub-module removes abnormal data according to structural characteristics, heat transfer rules and mechanical stress distribution of the cylinder head. The mathematical sub-module quantifies features and eliminates redundant information using RMS, FFT and other algorithms. The two sub-modules jointly generate standardized PP-FMA derived features, which can be directly fed into neural networks.

This PP-FMA framework integrates traditional qualitative failure analysis with feature engineering. Standardized features are automatically generated relying on physical mechanism constraints and mathematical quantitative calculations. It not only reduces subjective bias caused by manual screening and adapts to the processing of large-batch time-series data, but also produces physically interpretable features for neural network modeling. The framework simplifies the pre-data processing pipeline and improves the performance of subsequent modeling.

3.2 Development of PP-FMA Boundary Diagram

To standardize the PP-FMA failure analysis procedure and eliminate interference from irrelevant components, this paper first constructs a boundary diagram for the cylinder head assembly system. The energy and medium interaction relationships between the cylinder head, all surrounding subsystems and external influencing factors are sorted out, thereby defining the research boundary for subsequent failure analysis and parameter screening, as illustrated in Figure 3.

The core structures of the cylinder head body are clarified first in the diagram, including oil and gas passages, high-pressure oil galleries, cooling water jackets, spark plug mounting seats and fastening bolt holes. These structures serve as critical flow channels for lubricating oil, coolant and combustion gas, and meanwhile act as structural zones prone to crack failures where thermal stress and mechanical stress are concentrated.

From the input perspective, ambient temperature and humidity, road gradient, driver operation and maintenance status will alter the overall engine operating load. Subsystems such as intake and exhaust, turbocharging, EGR and oil supply deliver working fluids and loads to the cylinder head. The ECU regulates in-cylinder combustion intensity by controlling natural gas injection and spark plug ignition. All types of input conditions collectively determine the magnitude of thermal and mechanical loads borne by the cylinder head.

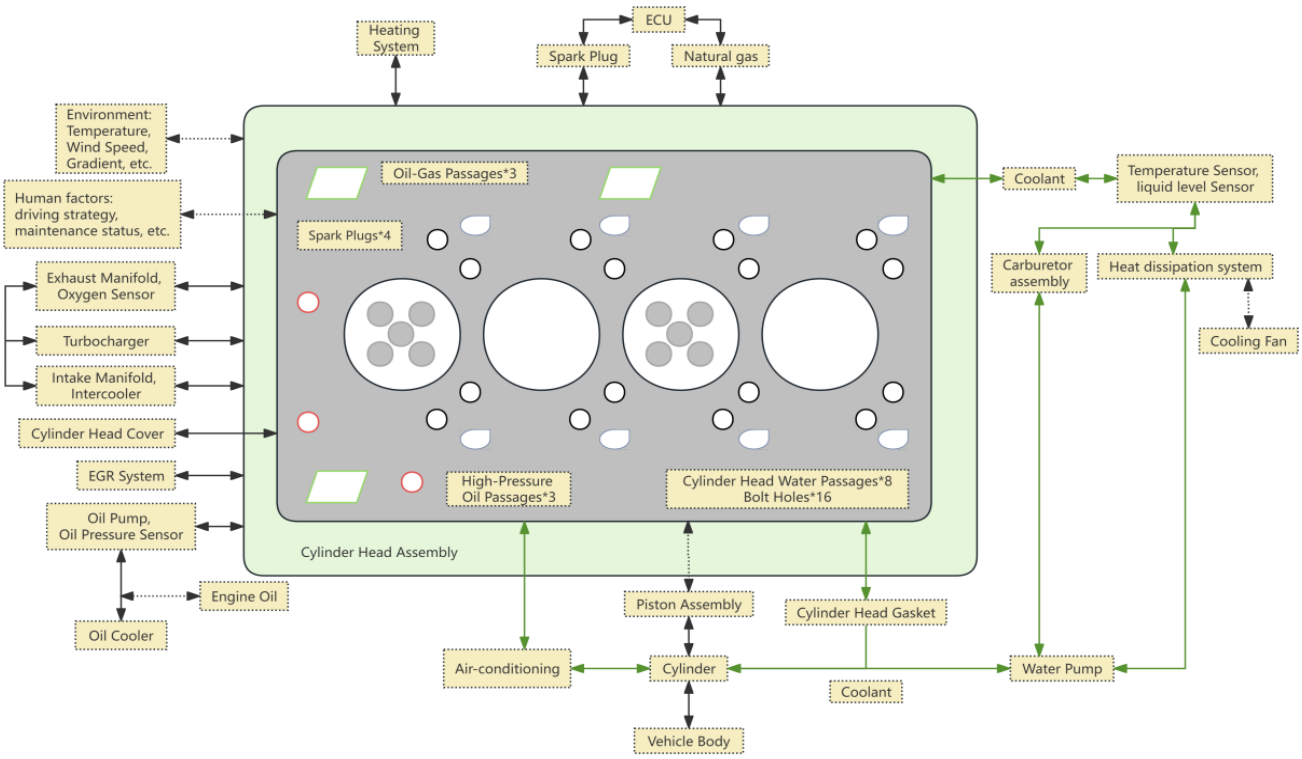

Figure 3: Boundary diagram of engine cylinder head

From the output perspective, heat from the cylinder head is transferred to the cooling circuit via coolant. Combustion forces are transmitted to the cylinder block and vehicle mounting structures through the cylinder gasket and piston assembly. Components including the cooling fan and water pump jointly sustain coolant circulation, which indirectly affects the long-term thermal operating conditions of the cylinder head.

The boundary diagram has defined the full scope of systems interacting with the cylinder head and screened upstream and downstream influencing factors that induce cylinder head failures. However, the boundary diagram only sorts out system-level interactions without decomposing each basic function of the cylinder head and corresponding failure phenomena. Accordingly, a function tree is constructed based on the defined analytical boundary to further realize layer-by-layer decomposition from component functions to specific failure modes.

3.3 Development of PP-FMA Function Tree

After completing the boundary division of the cylinder head assembly, a hierarchical function tree can be constructed based on system interaction relationships to decompose the four core basic functions of the cylinder head. Various fault manifestations caused by functional failures are analyzed, laying an analytical foundation for screening key monitoring parameters through the subsequent parameter diagram. While the boundary diagram only clarifies the mass and energy transfer paths between the cylinder head and peripheral systems, the function tree further breaks down the subdivided performance indicators corresponding to each primary function. When a certain function fails to meet design specifications, the degree of its impact on cylinder head failure modes can be evaluated, as shown in Figure 4.

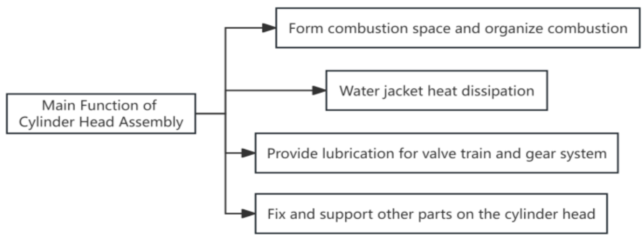

Figure 4: Main functions of engine cylinder head

3.3.1 Module Decomposition: Form Combustion Space and Organize Combustion

The first core function is to form combustion chambers and organize combustion, which is subdivided into two sub-functions: intake and exhaust flow regulation, as well as gas and coolant sealing. These sub-functions correspond to design indicators including flow coefficient, heat transfer and pressure bearing capacity, and sealing reliability. Failure of this function can lead to various typical operating abnormalities. Imbalanced intake-exhaust matching and gas leakage reduce combustion efficiency, resulting in decreased vehicle power and increased fuel consumption. Uncontrolled flow velocity and flow resistance cause excessive exhaust emissions. Airflow fluctuations further trigger engine shaking, misfire and unstable idling. In extreme cases, gas sealing failure directly interrupts combustion and causes sudden engine stall. Coolant leakage leads to overheating. The cylinder head bottom plate and exhaust passages endure alternating explosion pressure over long-term operation, accumulating high-cycle fatigue damage and eventually generating cracks, as shown in Figure 5.

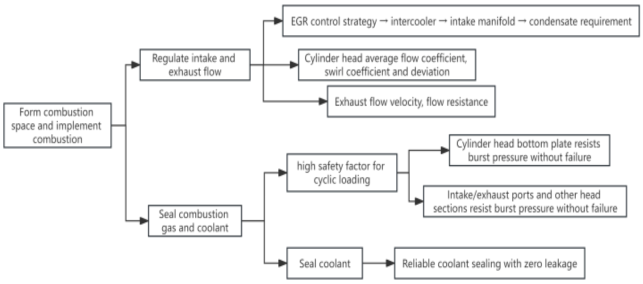

Figure 5: Form combustion space and organize combustion

3.3.2 Module Decomposition: Water Jacket Heat Dissipation

The second core function is water jacket heat dissipation, which is decomposed into four sub-items: heat transfer of cylinder head bottom plate and exhaust passages, water jacket heat exchange, water temperature monitoring, and heat dissipation of the complete engine cooling circuit. It controls indicators including heat transfer coefficient, flow resistance, local boiling temperature, sensor reliability and heat dissipation capacity. When heat-dissipation-related functions fail, the local temperature of the cylinder head exceeds the safe threshold. Low-cycle thermal cycling caused by alternating hot and cold conditions leads to rapid accumulation of fatigue damage on the bottom plate and exhaust passages. Overall overheating of the cylinder head also deteriorates in-cylinder combustion, raises the average fuel consumption of the vehicle, and increases the probability of cylinder head cracking, as shown in Figure 6.

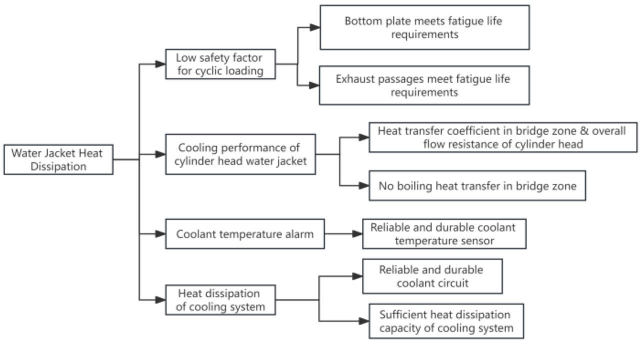

Figure 6: Water jacket heat dissipation

3.3.3 Module Decomposition: Provide Lubrication for Valve Train and Gear System

The second core function is to supply lubricating oil to the valve train and gear system, which consists of two sub-functions: oil sealing and smooth oil circuit oil return. If this function fails, unregulated oil leakage and insufficient oil return flow velocity will lead to reduced oil circuit pressure and deteriorated lubrication of the valve train. Insufficient lubrication for friction pairs of components will cause abnormal mechanical vibration, reduce the motion precision of the valve train and shift the intake and exhaust timing, ultimately resulting in decreased engine power and poor combustion matching, as shown in Figure 7.

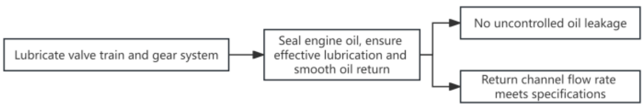

Figure 7: Provide lubrication for valve train and gear system

3.3.4 Module Decomposition: Fix and Support Other Parts on The Cylinder Head

The fourth function is to fix and support various auxiliary parts of the cylinder head, providing mounting constraints for components including fastening bolts, rocker shafts, spark plugs, intake and exhaust valves, manifolds, lifting lugs, camshafts and cylinder head covers. Failures of the supporting function mainly stem from three problems: valve seat wear and sinking, low-cycle thermal relaxation of bolts, and fatigue leakage of cylinder gaskets. Sinking of valve seats damages the tightness of combustion chambers, resulting in power reduction, excessive emissions and idle misfire. Bolts and cylinder gaskets suffer fatigue failure under thermal cycling shock and explosion pressure vibration, causing gas and coolant leakage and directly triggering cylinder head faults, as shown in Figure 8.

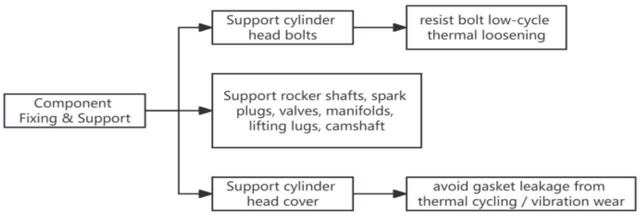

Figure 8: Fix and support other parts on the cylinder head

After sorting out the failure states corresponding to the four major functions, the functions and performance indicators behind each type of fault can be located. In the subsequent construction of the parameter diagram, operating parameters that can characterize these indicators can be extracted in a targeted manner, establishing a mapping relationship from physical functions to monitorable operating data.

3.4 Development of PP-FMA Parameter Diagram

After completing the hierarchical decomposition of the function tree, the subdivided performance indicators and failure phenomena corresponding to each basic function of the cylinder head have been sorted out. However, it is still necessary to establish quantitative correlations among operating inputs, controllable design factors, external disturbances and various failure modes. For this reason, the development of the PP-FMA parameter diagram needs to be implemented, as shown in Figure 9.

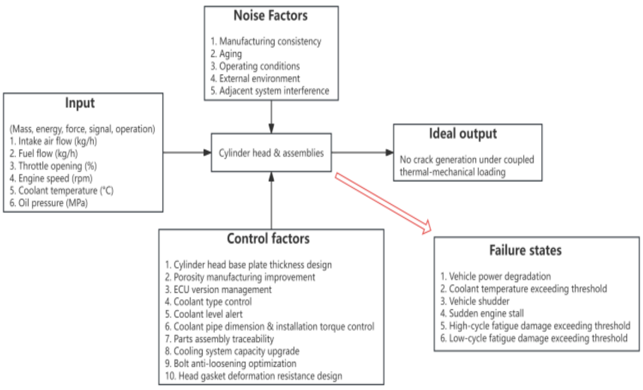

Figure 9: PP-FMA parameter diagram

This parameter diagram adopts the classic four-category framework of the P-diagram. Inputs refer to collectable engine operating condition parameters. Noise factors include artificially uncontrollable influencing items such as manufacturing dispersion, component aging, service conditions, external environment and disturbance from adjacent systems. The cylinder head and its components serve as the core processing part of the system. Under normal operating conditions, the ideal output of crack-free thermo-mechanical coupled loads can be achieved. When input fluctuations and noise disturbances exceed the tolerance limit of control factors, the system will enter a failure state with six types of failure characteristics:

1)Vehicle power degradation
2)Coolant temperature exceeding threshold
3)Vehicle shudder
4)Sudden engine stall
5)High-cycle fatigue damage exceeding threshold
6)Low-cycle fatigue damage exceeding threshold

After completing element classification based on this parameter diagram, six core failure features can be screened from control factors for subsequent feature engineering modeling and analysis.

3.5 Feature Engineering of PP-FMA (Physical Mechanisms and Mathematical Models)

The raw parameters sorted out via the parameter diagram are massive and contain abundant redundant information, which cannot be directly applied to failure analysis and calculation. Accordingly, feature engineering is carried out based on the sorted physical failure mechanism of the cylinder head to output a standardized data preprocessing scheme. Parameter screening and feature reconstruction are implemented to provide compliant and available datasets for subsequent various failure analysis tasks.

3.5.1 Vehicle Power Degradation

For vehicle power degradation: First, steady driving conditions are screened based on vehicle speed, transmission ratio, coolant temperature and road gradient, eliminating irrelevant operating conditions such as cold start and climbing. Next, sensor noise and abnormal data from rapid acceleration/deceleration are filtered out via acceleration threshold. Coasting conditions are removed by combining accelerator and brake signals, retaining only data samples under active driver control. Finally, the root-mean-square ratio of driving acceleration to braking acceleration is adopted to quantify the degree of power degradation, generating fault feature F1. The formula is as follows:

$$F_1 = \frac{F_{tor}(RMS)}{F_{br}(RMS)} \propto \frac{F_{tor}}{F_{br}} \tag{1}$$

$F_{tor}(RMS)$: Root mean square value of driving force within the observation period;
$F_{br}(RMS)$: Root mean square value of braking force within the observation period;
F_{tor}: Instantaneous driving force;
F_{br}: Instantaneous braking force.

3.5.2 Coolant Temperature Exceeding Threshold

For coolant temperature exceeding threshold: First, the exhaust temperature before the turbine is obtained by interpolation based on engine speed and torque. The heat generation formula is used to screen medium and low load ranges with heat generation of $40 < Q < 70$ and operating conditions with vehicle speed above $65km/h$, excluding irrelevant scenarios with high thermal load and sufficient cooling intake air. Then, normal temperature samples are eliminated by matching the monthly coolant temperature threshold, and time-series data of coolant temperature exceeding the threshold are extracted. Finally, the root mean square of the abnormal coolant temperature sequence is calculated to obtain the fault feature F_2. The formula is as follows:

$$Q = m \cdot \Delta T \cdot c \tag{2}$$

$$F_2 = \sqrt{\frac{1}{n}\sum_{i=1}^{n} T_{cool,i}^2} \tag{3}$$

Q: Real-time heat generation of engine;
m: Mass flow rate of heat exchange medium;
ΔT: Temperature rise of medium;
c: Specific heat capacity of medium;
$T_{cool,i}$: Single sampled value of coolant temperature exceeding threshold;
n: Number of abnormal coolant temperature samples.

3.5.2 Vehicle Shudder

For vehicle shudder: Longitudinal acceleration is obtained by differentiating longitudinal vehicle speed, and the acceleration signal is processed by FFT to acquire power spectral density. The integrated power value within the characteristic frequency band (e.g., $2\sim8Hz$) is taken as the shudder feature F3. Stable driving conditions (vehicle speed $> 30km/h$, no braking) shall be screened before calculation to eliminate interferences such as rapid acceleration and deceleration. The formula is as follows:

$$a = \frac{dv}{dt} \tag{4}$$

$$P(f) = |FFT(a)|^2 \tag{5}$$

$$F_3 = \int_{f_l}^{f_h} P(f)\, df \tag{6}$$

a: Longitudinal acceleration;
v: Longitudinal vehicle speed;
FFT: Fast Fourier Transform (FFT);
$P(f)$: Power spectral density of acceleration signal;
f_l: Lower limit of vehicle shudder characteristic frequency band;
f_h: Upper limit of vehicle shudder characteristic frequency band.

3.5.4 Sudden Engine Stall

For Sudden engine stall: Engine stall events are identified based on engine speed and vehicle speed: an idle state with engine speed $\leq 400r/min$ is regarded as near-stall; a running stall is confirmed when the vehicle is moving (vehicle speed $\geq 10km/h$) and engine speed $\leq 400r/min$. Sampling points satisfying the above criteria are divided into independent segments by temporal continuity, and the total number of such segments is taken as the stall fault feature. Define the stall state matching flag $M(t)$ at time t, which is used to count the initial times of continuous segments; the final feature F4 equals this count. The formula is as follows:

$$M(t) = 1_{\{espeed(t) \leq 400\}}, M(0) = 0 \quad (7)$$

$$F_4 = \sum_{t=1}^{N} [M(t) = 1 \wedge M(t-1) = 0] \quad (8)$$

espeed(t): Engine speed (r/min);
M(t): Stall state flag, 1 for stall, 0 for normal.

3.5.5 High-cycle Fatigue Damage Exceeding Threshold

For high-cycle fatigue damage exceeding threshold: Under all operating conditions, the relationship between cycle count and peak firing pressure of each working condition is calculated to obtain the damage increment of a single working condition. The full-condition damage feature F5 is acquired by cumulative summation. The formula is as follows:

$$D_j = \frac{N_j}{PFP_j^{1/b_f}} \quad (9)$$

$$F_5 = \sum_j D_j \quad (10)$$

D_j : Fatigue damage increment under the j-th operating condition;
N_j: Cycle count under this operating condition (estimated by engine speed and time);
PFP_j: Peak firing pressure under this operating condition (MPa), correlated with in-cylinder combustion pressure;
b_f: Material fatigue exponent (constant, calibrated by tests).

3.5.6 Low-cycle Fatigue Damage Exceeding Threshold

For low-cycle fatigue damage exceeding threshold: The strain amplitude corresponding to temperature variation is calculated via the thermal strain formula. The Rainflow Counting Method is adopted to count the cycle numbers corresponding to each strain amplitude. Subsequently, the total damage feature $F6$ is computed based on Miner's linear cumulative damage theory.The formula is as follows:

$$\varepsilon = \alpha \Delta T \quad (11)$$

$$\Delta\varepsilon_p N^{\alpha} = C \quad (12)$$

$$F_6 = \sum_{i=1}^{k} \frac{n_i}{N_i} \quad (13)$$

ε: Thermal strain;
α: Thermal expansion coefficient;
ΔT: Temperature variation amplitude;
$\Delta\varepsilon_p$: Plastic strain amplitude;
N: Fatigue failure cycle number;
C: Material constant;
n_i: Actual cycle count of the i-th strain amplitude counted by Rainflow Counting Method;
N_i : Fatigue failure cycle number determined by Manson-Coffin formula under this strain amplitude;

4. Results and Conclusion

4.1 Optimization Results of PP-FMA

In the previous chapter, derivative failure features are constructed via PP-FMA, namely Vehicle power degradation, Coolant temperature exceeding threshold, Vehicle shudder, Sudden engine stall, High-cycle fatigue damage exceeding threshold and Low-cycle fatigue damage exceeding threshold (corresponding to $F1 - F6$). Each feature is derived from physical mechanisms and mathematical models, constituting physics-prior features. The input dimension is compressed from the original full-set engine signals (14 raw sensor signals in total) to six engineering features. Redundant and irrelevant sensor channels are eliminated during data preprocessing. Moreover, each feature strictly corresponds to one type of engine fault, establishing a mapping between measured data and failure mechanisms. This addresses the limitation of raw sensor signals, which only reflect operating conditions without explicit fault indication.

To quantitatively verify the optimization performance of the feature engineering scheme proposed in this paper in suppressing information redundancy and mitigating multicollinearity, comparative experiments on PP-FMA features, raw data, and PCA principal component analysis are conducted in the following sections. Pearson correlation coefficient and mutual information are selected as evaluation metrics.

4.2 Description of Evaluation Metrics

Data groups for comparison:

1) PP-FMA features
2) Raw data (full original engine sensor data)
3) PCA-transformed raw data (raw data after PCA dimensionality reduction)

Evaluation metrics:

1) Pearson correlation coefficient: Measures the linear correlation between features to identify linear information redundancy.
2) Mutual Information (MI): Quantifies the degree of nonlinear correlation between features, comprehensively measuring nonlinear information overlap.

4.3 Comparative Experiments

4.3.1 Pearson Correlation Coefficient

The Pearson correlation coefficient quantifies the strength of linear correlation between two groups of continuous variables, with a value range of $r \in [-1,1]$. A value of 1 indicates perfect positive correlation, -1 perfect negative correlation, and 0 no linear correlation. Its calculation formula is shown as follows:

$$r_{XY} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}} \quad (14)$$

Where $\bar{x}$、$\bar{y}$ denote the sample means of variables X、Y, respectively. The numerator represents the covariance of the two variables, and the denominator realizes standardization via standard deviations.

The Pearson correlation coefficient heatmaps of raw data, PP-FMA features and PCA-transformed raw data are shown in Figures 10 to 12, respectively.

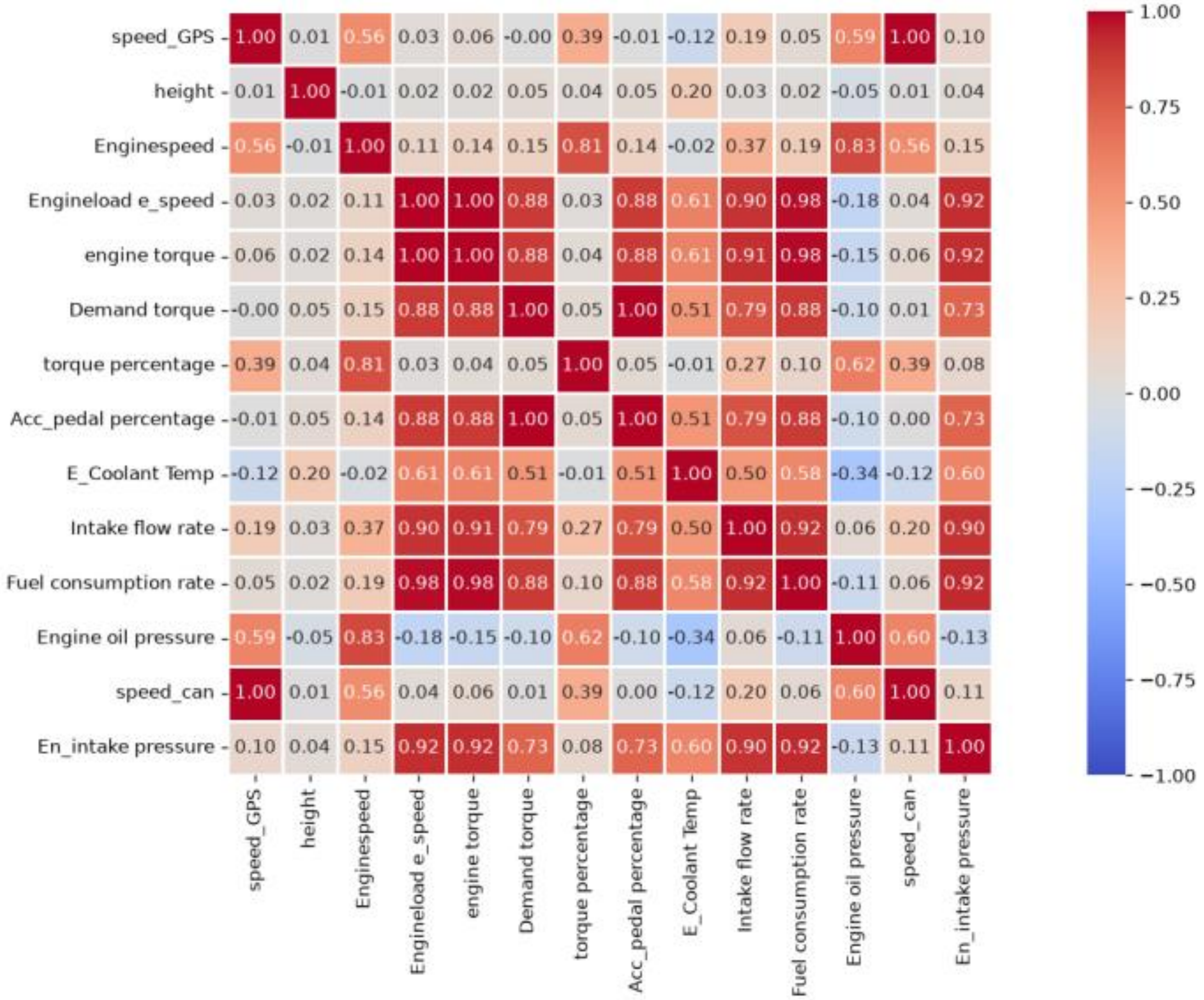

Figure 10: Pearson correlation coefficient of raw data

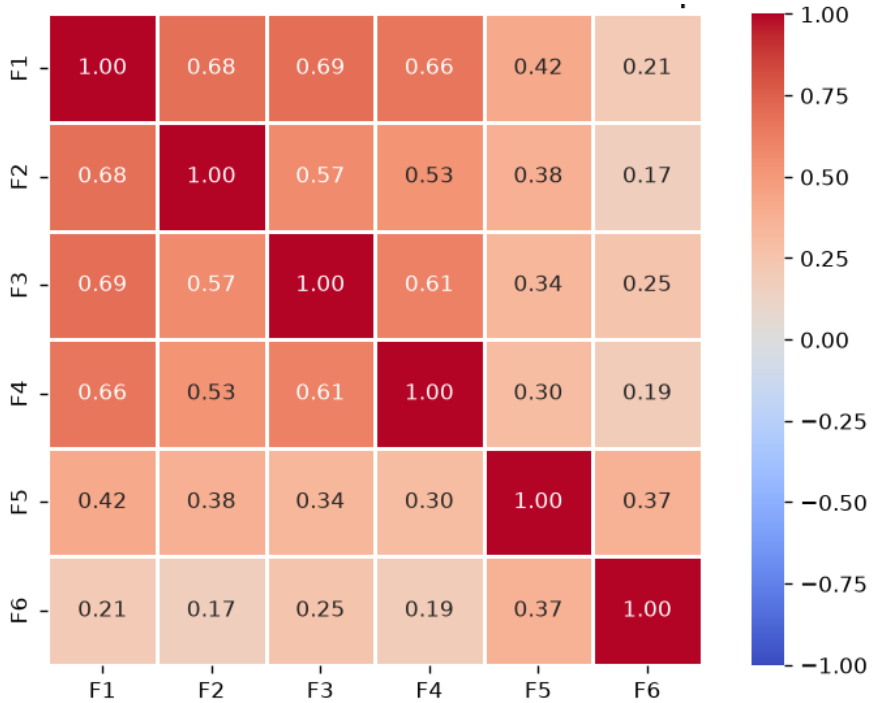

Figure 11: Pearson correlation coefficient of PP-FMA features

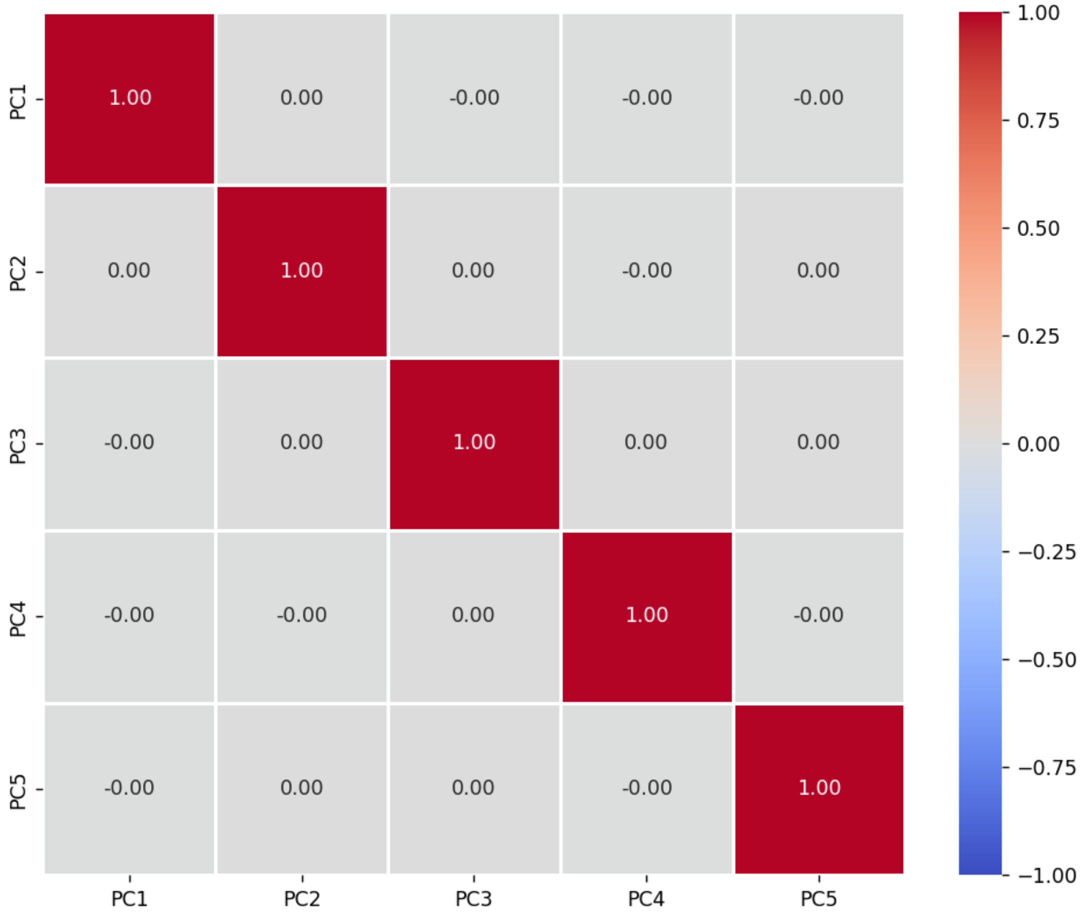

Figure 12: Pearson correlation coefficient of PCA-transformed raw data

Observations from the three groups of heatmaps are as follows: A large number of variables in raw data have correlation coefficients close to 0.9, which indicates severe multicollinearity and information redundancy. Although PCA-transformed raw data achieves complete decorrelation with correlation coefficients approaching zero, this transformation carries no engineering physical meaning and breaks the internal correlations among engine failure conditions. In contrast, PP-FMA features only exhibit moderate low correlation. This method drastically reduces redundant information from irrelevant failure dimensions while retaining reasonable physical coupling relationships among similar faults. Balancing low redundancy and engineering interpretability, PP-FMA serves as the optimal input representation with the best comprehensive performance among the three feature sets.

4.3.2 Mutual Information (MI)

Mutual information quantifies the overall dependence between two random variables and can capture both linear and nonlinear correlations. Its value range is $I(X;Y) \in [0,+\infty)$. A larger value indicates a higher degree of information overlap between variables, and $I(X;Y)=0$ means the two variables are mutually independent. The calculation formula is as follows:

$$I(X;Y) = \sum_x \sum_y p(x,y) \log_2 \frac{p(x,y)}{p(x)p(y)} \quad (15)$$

Where $p(x)$ and $p(y)$ denote the marginal probability distributions of variables X and Y, respectively, and $p(x,y)$ represents their joint probability distribution.

The mutual information (MI) values between the three feature sets (raw data, PP-FMA features, PCA-transformed raw data) and fault labels are shown in Tables 2–4.

Table 2: Mutual information between raw data and fault labels

Raw data	No fault	Cylinder head crack	Sand hole of cylinder head cover	Valve seat wear
Engineload_e_speed	0.11	0.57	0.52	0.53
Engine torque	0.10	0.55	0.50	0.51
Fuel consumption rate	0.12	0.53	0.49	0.48
Intake flow rate	0.09	0.49	0.56	0.47
Demand torque	0.10	0.48	0.51	0.50
Acc_pedal percentage	0.09	0.47	0.48	0.52
E_Coolant Temp	0.13	0.54	0.38	0.41
Enginespeed	0.08	0.42	0.44	0.55
Engine oil pressure	0.10	0.36	0.33	0.49
Torque percentage	0.07	0.28	0.26	0.32
Speed_can	0.06	0.21	0.19	0.27
Speed_GPS	0.05	0.17	0.15	0.22

Table 3: Mutual information between PP-FMA features and fault labels

PP-FMA	No fault	Cylinder head crack	Sand hole of cylinder head cover	Valve seat wear
F1	0.09	0.84	0.81	0.82
F2	0.08	0.80	0.85	0.79
F3	0.08	0.82	0.80	0.86
F4	0.10	0.83	0.78	0.77
F5	0.07	0.79	0.82	0.81
F6	0.09	0.78	0.79	0.83

Table 4: Mutual information between PCA-transformed raw data and fault labels

PC	No fault	Cylinder head crack	Sand hole of cylinder head cover	Valve seat wear
PC1	0.12	0.52	0.49	0.48
PC2	0.10	0.43	0.51	0.45
PC3	0.09	0.38	0.42	0.47
PC4	0.08	0.29	0.32	0.35
PC5	0.07	0.21	0.23	0.25

It can be concluded from the mutual information results between the three groups of features and three fault labels (cylinder head crack, sand hole of cylinder head cover and valve seat wear) that the F1~F6 features screened by PP-FMA achieve an average mutual information of 0.81 with faults and a maximum mutual information of 0.86. These values are significantly higher than those of raw data (average 0.46, maximum 0.57) and PCA-transformed raw data (average 0.40, maximum 0.52). This indicates that PP-FMA features based on physical prior knowledge can retain the maximum valid information strongly correlated with various engine faults, possess stronger discriminative ability for different fault conditions, and deliver the optimal performance in fault identification tasks.

4.4 Conclusion

This paper constructs a PP-FMA feature optimization framework based on physical prior knowledge. The failure logic is clarified with the boundary diagram, function tree and parameter diagram of the cylinder head system, and engine failure characteristics are extracted by spectrum analysis, fatigue damage calculation and other techniques. Comparative validation is implemented using Pearson correlation coefficient and mutual information. Raw sensor features suffer from severe multicollinearity. While PCA principal components eliminate linear redundancy, substantial key fault information is discarded. By contrast, PP-FMA features have controllable redundancy and deliver remarkably better discrimination for cylinder head cracks, casting sand holes and valve seat wear than the other two feature sets. Furthermore, PP-FMA features are endowed with definite physical interpretations, which can supply high-quality, interpretable standardized input features for deep learning prediction of cylinder head faults in natural gas engines.

References

[1] Norah Nadia Sánchez Torres, Jorge Gomes Lima, Joylan Nunes Maciel, et al. Non-invasive techniques for monitoring and fault detection in internal combustion engines: A systematic review[J]. Energies. 2024, 17(23): 6164.

[2] Zhongxin Shi, Hongzi Fei, Liuping Wang, et al. Vibration-based health monitoring of diesel engine valve clearance using dual-rate extended Kalman filter[J]. Journal of Vibration and Control. 2026, 59(35): 728-735.

[3] Sławomir Szrama. Multimodal spatiotemporal graph attention network with uncertainty quantification for aircraft engine prognostics[J]. Mechanical Systems and Signal Processing. 2026, 258: 1-24.

[4] Ahmad Junaid, Abid Iqbal1, Abuzar Khan, et al. Engine failure prediction on large-scale CMAPSS data using hybrid feature selection and imbalance-aware learning[J]. Computers, Materials and Continua. 2026, 87(1): 1546-1556.

[5] Huixin Yang, Xiang Li, Wei Zhang. Interpretability of deep convolutional neural networks on rolling bearing fault diagnosis[J]. Measurement Science and Technology. 2022, 33(5): 055005.

[6] Ammar Chakhrit, Nour El Houda Benharkat, Islam Hadj Mohamed Guetarni, et al. A literature review of risk assessment approaches in failure mode and effects analysis[J]. International Journal of Quality & Reliability Management. 2025, 42(9): 2415-2454.

[7] Shuzhi Su, Xingda Hu, Yanmin Zhu, et al. MP-PFDNet: Multi-perspective signal features and physics-prior aware network for aero-engine bearing fault diagnosis[J]. Results in Engineering. 2026, 30: 110402.

[8] Bo Wang, Zirui Wang, Chong Yao, et al. Multi-system condition monitoring of marine engines: A unified deep learning framework introducing physical prior knowledge[J]. International Journal of Naval Architecture and Ocean Engineering. 2025, 17: 100698.