Academic Journal of Applied Sciences
ISSN: 3105-0549 (Print), ISSN: 3105-0557 (Online) | Vol. 2, No. 3, 2026

Design and Implementation of a Data Analysis System for Heavy-Duty Engine Fault Diagnosis

Siyu Zhu *

School of Automobile and Traffic Engineering, Tianjin University of Technology and Education, Tianjin, 300222, China

* **Corresponding author**: (Email: siyu_zhu_tianjing@163.com)

Abstract: Aiming at the problems of lagging monitoring of heavy-duty engines under high load, long cycle and complex working conditions, difficulty in fault location and low maintenance efficiency, this paper designs and implements a lightweight, visual and integrated heavy-duty engine fault diagnosis data analysis system (V1.0). The system uses Python as the development platform, integrates five core capabilities of data access, health inspection, feature analysis, fault identification and report generation, and can perform centralized analysis and intelligent diagnosis of key operating parameters such as engine speed, load, coolant temperature, oil pressure, fuel pressure, intake pressure, exhaust temperature, and vibration value. Through data quality verification, multi-dimensional feature visualization, threshold rule discrimination and risk quantitative scoring, the system realizes rapid identification and accurate positioning of typical faults such as abnormal cooling, abnormal lubrication, abnormal fuel, abnormal vibration and abnormal intake. The test results show that the system can complete data import, operation situation analysis and fault diagnosis output without complex deployment conditions, effectively improve the efficiency of fault troubleshooting, reduce the experience dependence of maintenance personnel, and provide reliable technical support for the transformation of heavy-duty engines from passive maintenance to predictive maintenance.

Keywords: Heavy Engine, Health Assessment, Feature Extraction, Predictive Maintenance.

1. Introduction

In key fields such as transportation [1], construction machinery, mining [2], and marine equipment [3], heavy-duty engines serve as the core power units. They typically operate under long-term high-load, continuous conditions with multiple disturbances. Their operational status directly determines production continuity and safety. Common diagnostic components include combustion, lubrication, and cooling system faults [4]. Once abnormalities occur—such as abnormal coolant temperature, low oil pressure, or excessive mechanical vibration—minor faults can lead to increased fuel consumption, while severe ones may cause shutdown accidents. Furthermore, with the passage of time and processing cycles, mechanical wear inevitably leads to a loss of technical efficiency compared to the optimal state [5]. Therefore, efficient and accurate fault diagnosis is of great significance.

Currently, advanced automatic diagnosis technologies are evolving rapidly. Domestic research shows that the rail pressure signals of high-pressure common rail systems exhibit strong nonlinear characteristics, making traditional feature extraction methods difficult to apply [6]. To address overfitting and low accuracy in data-driven methods under limited samples, Tianjin University proposed a diagnosis model combining WACGAN and IRCNN [7], and also explored the dynamic characteristics of internal combustion engine rotor faults [8]. In addition, lightweight neural networks have been applied to accurately identify bearing faults under complex working conditions [9]. Internationally, end-to-end methods like MACNN [10] and enhanced digital twin-driven FDI methods combining sensor imaging mechanisms with CNNs [11] have significantly improved feature extraction and fault isolation capabilities.

Despite these advanced algorithms, most existing methods focus on specific components or require complex computing environments, making them difficult to quickly deploy for frontline maintenance personnel. Traditional diagnostic modes still suffer from lagging monitoring, difficulty in fault location, high dependence on personal experience, and low maintenance efficiency.

Against this background, this paper develops a heavy-duty engine fault diagnosis data analysis system (V1.0). Taking lightweight design, easy operation, and high practicality as the core objectives, it provides maintenance technicians with a one-stop tool. Without complex environmental configuration, the system supports rapid local import and visual analysis of CSV data. It can automatically complete health inspection, feature mining, fault identification, and report output, which significantly improves the efficiency and standardization of heavy-duty engine fault diagnosis, promoting the transformation from passive maintenance to predictive maintenance.

2. System Requirements Analysis and Overall Design

2.1. Demand Analysis

Data access requirements support local CSV format data import, compatible with running time, speed, load, temperature, pressure, vibration value, fault label and other fields; provide sample data fast loading function, easy to demonstrate and test.

Health inspection needs to automatically complete data integrity verification, missing value detection, sample size statistics, fault category coverage assessment, and generate intuitive data quality reports and health overviews.

Feature analysis requires statistical calculation and visual display of key parameters, including index mean, distribution characteristics, correlation relationship, trend changes, etc., to help users quickly grasp the operation rules of equipment.

Fault diagnosis needs to realize automatic identification of multi-type faults based on threshold rules such as temperature, pressure, vibration and load. Supports risk scoring, confidence calculation, high-risk sample ranking, and gives executable maintenance and disposal recommendations.

The ease of use requirement interface is simple, the operation process is clear, no login, no database configuration, one-click start, one-click import, one-click diagnosis, one-click export report, adapt to the scene of rapid diagnosis. The running example is shown in Figure 1.

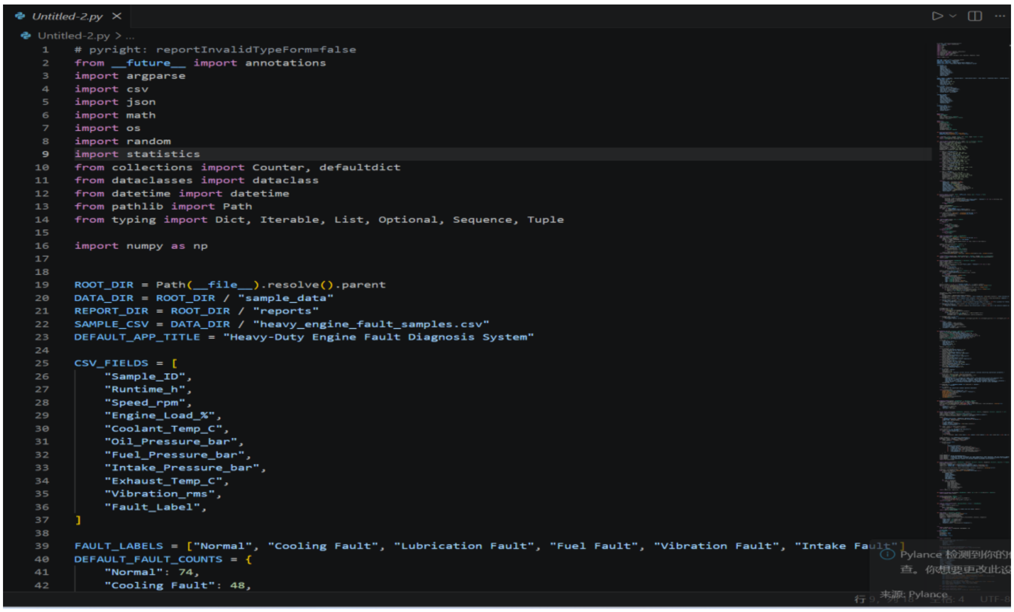


Figure 1. The running example

2.2. Overall Architecture of the System

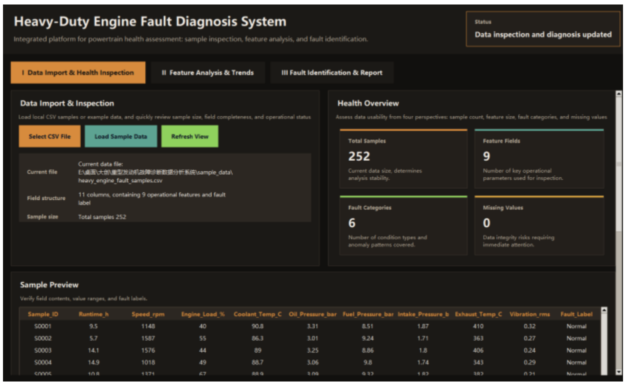
The system adopts modular design and is divided into three layers:

Data layer: responsible for data reading, format conversion, missing value processing, field verification and data object caching, providing a unified data source for the upper layer functions.

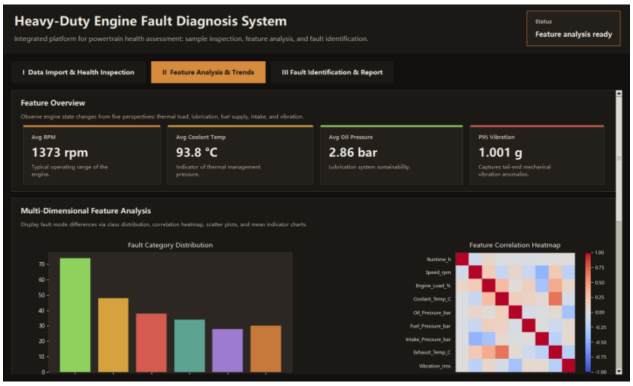
Function layer: It includes three core modules: data access and health inspection, feature analysis and operation situation, fault identification and diagnosis report. (The demonstration of the three core modules is shown in Figure.2.)

Presentation layer: It provides a graphical interactive interface to realize data display, chart visualization, operation button, result output and report export.

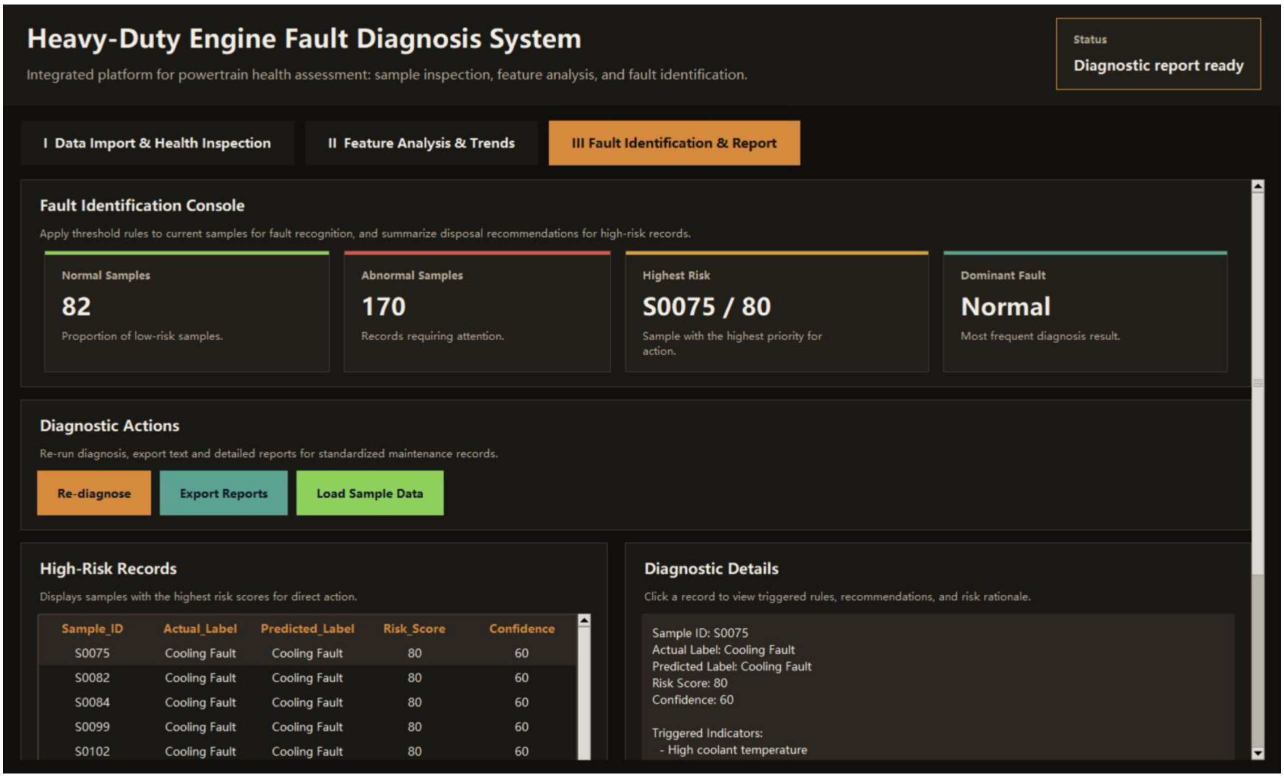
The overall process of the system is: data import → health inspection → feature analysis → fault identification → report export.



(a) data access and health inspection



(b) chart visualization



(c) fault identification and diagnosis report

Figure 2. The demonstration of the three core modules

3. Detailed Design of the System

3.1. Core Function Module Design

The following is the design idea and function realization of the core module

3.1.1. Data Access and Health Inspection Module

The module is the entrance of the system, and the main functions include:

File selection: support local CSV file import and sample data loading;

Field verification: automatic identification of sample number, running time, speed, load, temperature, pressure, vibration, fault category and other key fields;

Data statistics: calculate the total number of samples, the number of feature fields, the number of fault categories, the number of missing values;

(File selection, Field verification, Data statistics, several major functions are shown in Figure 3(a))

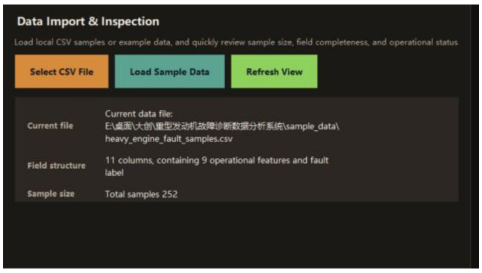
Sample preview: display the first 12 data, which is convenient to quickly check the field and numerical rationality;

(The sample preview function is shown in Figure 3(b))

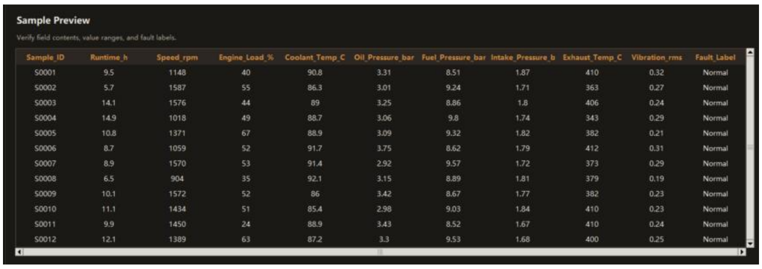
Inspection conclusion: automatic generation of data quality summary, including field integrity, category distribution, key index mean, high correlation feature pair prompt.

(Inspection conclusion function demonstration see Figure 3(c))

Through this module, the user can complete the data availability judgment within 10 seconds to determine whether to enter the subsequent in-depth analysis.



(a)



Sample_ID	Runtime_h	Speed_rpm	Engine_Load_%	Coolant_Temp_C	Oil_Pressure_bar	Fuel_Pressure_bar	Intake_Pressure_b	Exhaust_Temp_C	Vibration_rms	Fault_Label
S0001	9.5	1148	40	90.8	3.31	8.51	1.87	410	0.32	Normal
S0002	5.7	1587	55	86.3	3.01	9.24	1.71	363	0.27	Normal
S0003	14.1	1576	44	89	3.25	8.86	1.8	406	0.24	Normal
S0004	14.9	1018	49	88.7	3.06	9.8	1.74	343	0.29	Normal
S0005	10.8	1371	67	88.9	3.09	9.32	1.82	382	0.21	Normal
S0006	8.7	1059	52	91.7	3.75	8.62	1.79	412	0.31	Normal
S0007	8.9	1570	53	91.4	2.92	9.57	1.72	373	0.29	Normal
S0008	6.5	904	35	92.1	3.15	8.89	1.81	379	0.19	Normal
S0009	10.1	1572	52	86	3.42	8.67	1.77	382	0.23	Normal
S0010	11.1	1434	51	85.4	2.98	9.03	1.84	410	0.23	Normal
S0011	9.9	1450	24	88.9	3.43	8.52	1.67	410	0.24	Normal
S0012	12.1	1389	63	87.2	3.3	9.53	1.68	400	0.25	Normal

(b) The sample preview function

Figure 3. (c) Inspection conclusion

3.1.2. Feature Analysis and Operation Situation Module

Overview of key indicators: average speed, average coolant temperature, average oil pressure, high-risk vibration value;

Fault category diagram: display normal, abnormal cooling, abnormal lubrication, abnormal fuel, abnormal vibration, abnormal intake analysis;

Intelligent analysis and interpretation: automatically prompt the key focus directions such as thermal management risk, lubrication low pressure risk, and high vibration samples.

This module transforms abstract data into recognizable suggestions and feature analysis, reducing the threshold for technical personnel to analyze.

The function demonstration interface is shown in Figure 4.

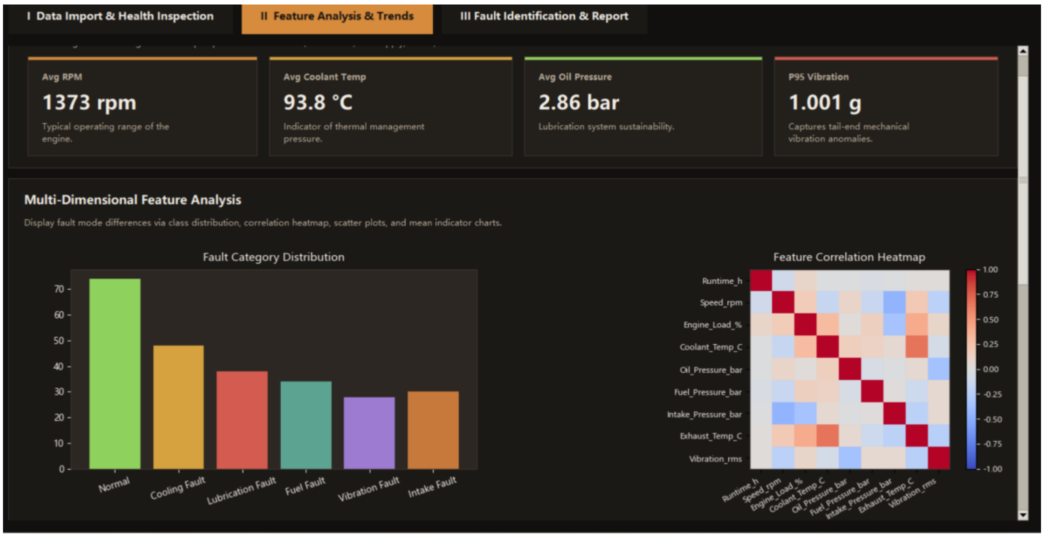


Figure 4. Feature analysis and operation situation

3.1.3. Fault Identification and Diagnosis Report Module

The core output module of the system realizes intelligent fault discrimination based on the rule engine:

Fault types: covering abnormal cooling, abnormal lubrication, abnormal fuel oil, abnormal vibration, abnormal air intake;

discriminant basis: too high / too low coolant temperature, low oil pressure, fuel pressure fluctuation, abnormal intake pressure, high exhaust temperature, excessive vibration value;

risk assessment: calculate the risk score and confidence for each sample, and sort them according to the risk level;

high-risk record: display the highest priority abnormal samples, support click to view details;

diagnostic description: display sample number, actual category, prediction category, hit rule, disposal suggestion;

report export: Generate text diagnostic reports and detailed forms, support archiving and standardized output.

(Fault identification and diagnosis report demo interface is shown in Figure 5)

(Export text reports and detailed table examples are shown in Figure 6.)

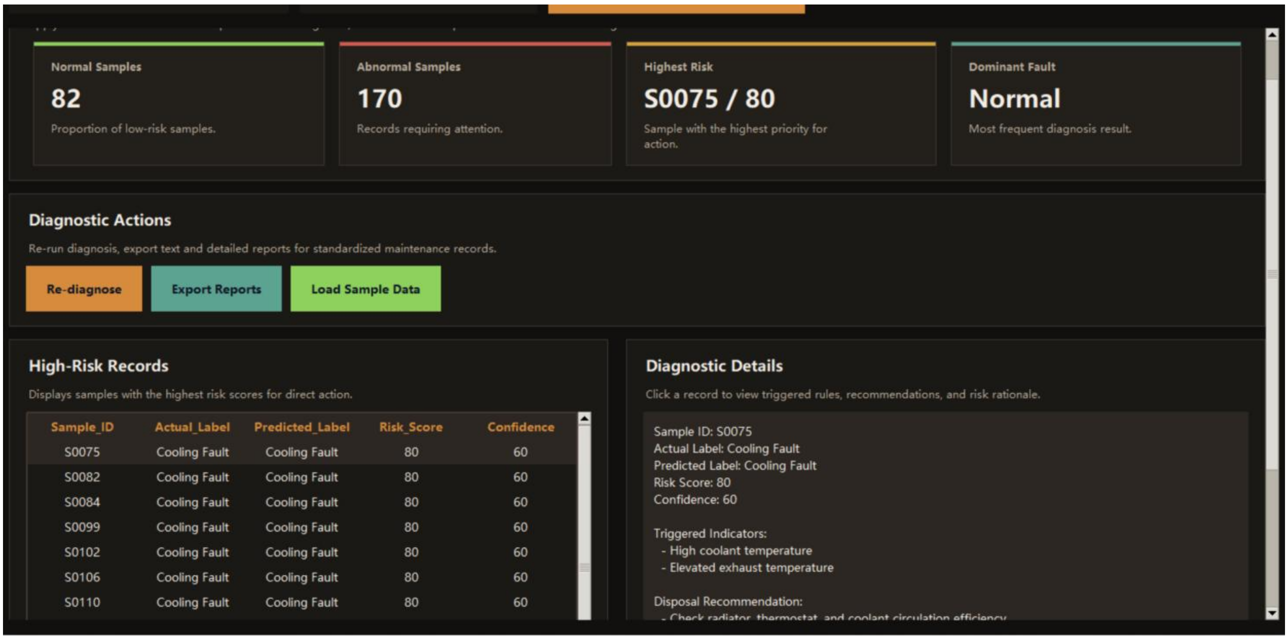
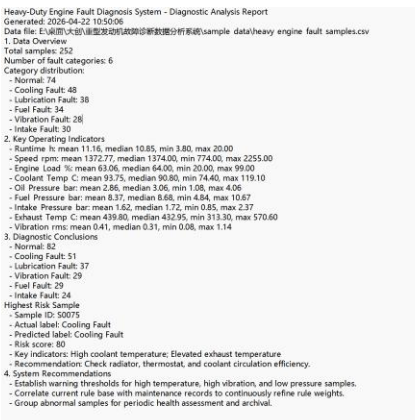
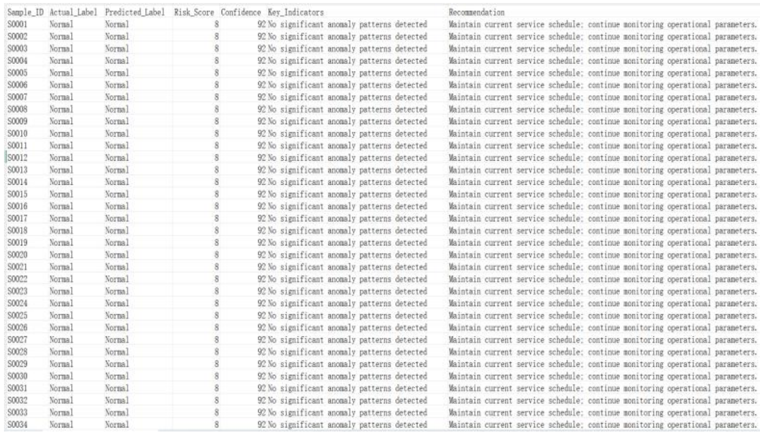


Figure 5. Fault identification and diagnosis report



Heavy-Duty Engine Fault Diagnosis System - Diagnostic Analysis Report
Generated: 2026-04-22 10:50:06
Data file: E:\桌面\大创\重型发动机故障诊断数据分析系统\sample data\heavy engine fault samples.csv
1. Data Overview
Total samples: 252
Number of fault categories: 6
Category distribution:
- Normal: 74
- Cooling Fault: 48
- Lubrication Fault: 38
- Fuel Fault: 34
- Vibration Fault: 28
- Intake Fault: 30
2. Key Operating Indicators
- Runtime h: mean 11.16, median 10.85, min 3.80, max 20.00
- Speed rpm: mean 1372.77, median 1374.00, min 774.00, max 2255.00
- Engine Load %: mean 63.06, median 64.00, min 20.00, max 99.00
- Coolant Temp C: mean 93.75, median 90.80, min 74.40, max 119.10
- Oil Pressure bar: mean 2.86, median 3.06, min 1.08, max 4.06
- Fuel Pressure bar: mean 8.37, median 8.68, min 4.84, max 10.67
- Intake Pressure bar: mean 1.62, median 1.72, min 0.85, max 2.37
- Exhaust Temp C: mean 439.80, median 432.95, min 313.30, max 570.60
- Vibration rms: mean 0.41, median 0.31, min 0.08, max 1.14
3. Diagnostic Conclusions
- Normal: 82
- Cooling Fault: 51
- Lubrication Fault: 37
- Vibration Fault: 29
- Fuel Fault: 29
- Intake Fault: 24
Highest Risk Sample
- Sample ID: S0075
- Actual label: Cooling Fault
- Predicted label: Cooling Fault
- Risk score: 80
- Key indicators: High coolant temperature; Elevated exhaust temperature
- Recommendation: Check radiator, thermostat, and coolant circulation efficiency.
4. System Recommendations
- Establish warning thresholds for high temperature, high vibration, and low pressure samples.
- Correlate current rule base with maintenance records to continuously refine rule weights.
- Group abnormal samples for periodic health assessment and archival.

(a) Text report



Sample_ID	Actual_Label	Predicted_Label	Risk_Score	Confidence	Key_Indicators	Recommendation
S0001	Normal	Normal	8	92	No significant anomaly patterns detected	Maintain current service schedule; continue monitoring operational parameters.
S0002	Normal	Normal	8	92	No significant anomaly patterns detected	Maintain current service schedule; continue monitoring operational parameters.
S0003	Normal	Normal	8	92	No significant anomaly patterns detected	Maintain current service schedule; continue monitoring operational parameters.
S0004	Normal	Normal	8	92	No significant anomaly patterns detected	Maintain current service schedule; continue monitoring operational parameters.
S0005	Normal	Normal	8	92	No significant anomaly patterns detected	Maintain current service schedule; continue monitoring operational parameters.
S0006	Normal	Normal	8	92	No significant anomaly patterns detected	Maintain current service schedule; continue monitoring operational parameters.
S0007	Normal	Normal	8	92	No significant anomaly patterns detected	Maintain current service schedule; continue monitoring operational parameters.
S0008	Normal	Normal	8	92	No significant anomaly patterns detected	Maintain current service schedule; continue monitoring operational parameters.
S0009	Normal	Normal	8	92	No significant anomaly patterns detected	Maintain current service schedule; continue monitoring operational parameters.
S0010	Normal	Normal	8	92	No significant anomaly patterns detected	Maintain current service schedule; continue monitoring operational parameters.
S0011	Normal	Normal	8	92	No significant anomaly patterns detected	Maintain current service schedule; continue monitoring operational parameters.
S0012	Normal	Normal	8	92	No significant anomaly patterns detected	Maintain current service schedule; continue monitoring operational parameters.
S0013	Normal	Normal	8	92	No significant anomaly patterns detected	Maintain current service schedule; continue monitoring operational parameters.
S0014	Normal	Normal	8	92	No significant anomaly patterns detected	Maintain current service schedule; continue monitoring operational parameters.
S0015	Normal	Normal	8	92	No significant anomaly patterns detected	Maintain current service schedule; continue monitoring operational parameters.
S0016	Normal	Normal	8	92	No significant anomaly patterns detected	Maintain current service schedule; continue monitoring operational parameters.
S0017	Normal	Normal	8	92	No significant anomaly patterns detected	Maintain current service schedule; continue monitoring operational parameters.
S0018	Normal	Normal	8	92	No significant anomaly patterns detected	Maintain current service schedule; continue monitoring operational parameters.
S0019	Normal	Normal	8	92	No significant anomaly patterns detected	Maintain current service schedule; continue monitoring operational parameters.
S0020	Normal	Normal	8	92	No significant anomaly patterns detected	Maintain current service schedule; continue monitoring operational parameters.
S0021	Normal	Normal	8	92	No significant anomaly patterns detected	Maintain current service schedule; continue monitoring operational parameters.
S0022	Normal	Normal	8	92	No significant anomaly patterns detected	Maintain current service schedule; continue monitoring operational parameters.
S0023	Normal	Normal	8	92	No significant anomaly patterns detected	Maintain current service schedule; continue monitoring operational parameters.
S0024	Normal	Normal	8	92	No significant anomaly patterns detected	Maintain current service schedule; continue monitoring operational parameters.
S0025	Normal	Normal	8	92	No significant anomaly patterns detected	Maintain current service schedule; continue monitoring operational parameters.
S0026	Normal	Normal	8	92	No significant anomaly patterns detected	Maintain current service schedule; continue monitoring operational parameters.
S0027	Normal	Normal	8	92	No significant anomaly patterns detected	Maintain current service schedule; continue monitoring operational parameters.
S0028	Normal	Normal	8	92	No significant anomaly patterns detected	Maintain current service schedule; continue monitoring operational parameters.
S0029	Normal	Normal	8	92	No significant anomaly patterns detected	Maintain current service schedule; continue monitoring operational parameters.
S0030	Normal	Normal	8	92	No significant anomaly patterns detected	Maintain current service schedule; continue monitoring operational parameters.
S0031	Normal	Normal	8	92	No significant anomaly patterns detected	Maintain current service schedule; continue monitoring operational parameters.
S0032	Normal	Normal	8	92	No significant anomaly patterns detected	Maintain current service schedule; continue monitoring operational parameters.
S0033	Normal	Normal	8	92	No significant anomaly patterns detected	Maintain current service schedule; continue monitoring operational parameters.
S0034	Normal	Normal	8	92	No significant anomaly patterns detected	Maintain current service schedule; continue monitoring operational parameters.

(b) Detailed table examples

Figure 6. Report

3.2. Data Structure Design

The key fields supported by the system include: sample number, running time (h), speed (rpm), engine load (%), coolant temperature (°C), oil pressure (bar), fuel pressure (bar), intake pressure (bar), exhaust temperature (°C), vibration (rms), fault category.

Fault category label definition: 0-normal, 1-cooling anomaly, 2-lubrication anomaly, 3-fuel anomaly, 4-vibration anomaly, 5-Intake anomaly.

3.3. Operating Environment Design

The system is based on Python development, no login, no deployment of Web services, operation mode:

(1) Open the system folder;

(2) Double-click to run heavy_engine_fault_system.py;

(3) Or use VS Code, PyCharm to execute the script directly.

After the interface is started, it can enter the operation, which has the characteristics of cross-platform, lightweight and easy migration.

4. System Implementation and Key Technologies

4.1. Development Tools and Dependency Libraries

The system is developed in Python language and mainly relies on libraries:

Pandas: data reading, cleaning, statistics and calculation;

matplotlib / Seaborn: data visualization, drawing histograms, heat maps, scatter plots;

tkinter: graphical interface construction;

pathlib: file path management;

cSV: Data format parsing.

4.2. Data Processing Flow Implementation

(1) Data import: Read the CSV file, check the integrity of the field, handle the abnormal format;

(2) Data validation: statistical missing values, repeated values, field legitimacy;

(3) Feature calculation: mean, median, correlation coefficient, distribution proportion;

(4) Fault identification: matching fault types according to

threshold rules;

(5) Result cache: Share data objects to three functional pages to achieve one-time data import and full-process use.

Acknowledgements

The authors gratefully acknowledge the financial support from xxx funds.

5. System Test and Result Analysis

5.1. Test Environment

Operating system: Windows 10/11

Python version: 3.8 and above

Test data set: 252 heavy-duty engine operation samples, including 6 states (normal+5 abnormal)

The number of fields: 11 columns, 9 running features+2 identification labels

5.2. Functional Test Results

Data access and inspection import 252 data, no missing values, complete fields, identify 6 fault categories, and the inspection conclusion is accurate. Key indicators: the average coolant temperature is 93.8 °C, the average oil pressure is 2.86 bar, the correlation coefficient between coolant temperature and exhaust temperature is 0.670, and the data availability is high.

The feature analysis and situation analysis system automatically generates distribution charts, correlation heat maps, scatter plots, and mean plots, which clearly show:

The number of normal samples is the largest;

Cooling anomaly is the main anomaly type;

There is a clear boundary between temperature and vibration.

The overall oil pressure is low and needs to be focused on.

The fault identification and report output system successfully identified 170 abnormal samples and 82 normal samples. The highest risk sample was S0075 (abnormal cooling), the risk score was 80, and the confidence level was 60. The diagnosis basis is clear, the disposal suggestion is targeted, and the report is exported completely.

6. Conclusion

The heavy-duty engine fault diagnosis data analysis system (V1.0) developed in this paper is oriented to the actual operation and maintenance requirements of heavy-duty power equipment, and realizes the integrated functions of data access, health inspection, feature analysis, fault identification, and report export. The system is implemented in a lightweight Python architecture. It does not require complex configuration. The interface is simple and convenient to operate. It can quickly process multi-source operating data of the engine, automatically identify typical faults, and give disposal suggestions. It effectively solves the problems of low efficiency, dependence on experience, and lag in anomaly detection of traditional diagnostic modes.

The test results show that the system meets the requirements of engineering application in data processing, visual analysis, fault identification accuracy, and operation convenience. It has high practical value and can provide reliable tool support for heavy-duty engine health management and predictive maintenance.

However, the current system (V1.0) still has certain limitations. First, the fault identification mechanism primarily relies on pre-defined threshold rules, which lacks adaptive adjustment capabilities for engines operating under significantly varying working conditions or extreme environments, potentially leading to false positives or omissions. Second, the system currently processes offline CSV data in batches, which falls short in achieving real-time online monitoring and dynamic early warning. Finally, the diagnostic dimension is limited to single-sample static feature analysis and has not yet deeply explored the temporal degradation trends and coupled fault mechanisms of the equipment.

Looking forward, future research will focus on the following directions: (1) Introducing machine learning and deep learning algorithms (e.g., Random Forest, LSTM) to replace or assist the rule engine, thereby enhancing the system's adaptive diagnostic capabilities and generalization in complex scenarios. (2) Developing a real-time data acquisition module to transform the system from offline analysis to online real-time monitoring, enabling dynamic health tracking. (3) Expanding the data structure to incorporate time-series data, facilitating the prediction of Remaining Useful Life (RUL) and further advancing the intelligent transformation of heavy-duty engine maintenance strategies.

Acknowledgements

This work was supported by the College Students' Innovation and Entrepreneurship Training Program of Tianjin University of Technology and Education (Grant No. 202510066116, Project Title: Heavy-duty engine fault expert system).

References

[1] State Administration for Market Regulation, & Standardization Administration of China. (2019). Railway rolling stock—Operational safety requirements—Part 1: General safety requirements for rolling stock (GB/T 25338.1-2019).

[2] National Mine Safety Administration. (2024). Safety technical specification for mining vehicles (2024). China University of Mining Press.

[3] State Administration for Market Regulation, & Standardization Administration of China. (2025). Technical requirements and test methods for marine diesel engine speed governing system (GB/T 3475-2025).

[4] Yu, Y., Ye, J., Li, B., et al. (2026). Fault diagnosis method of high-pressure common rail system based on MFE-MCFS-SVM. Transactions of Internal Combustion Engines, 44(2), 155-162.

[5] Tang, C., Bi, F., Huang, M., et al. (2025). Diesel engine fault diagnosis method based on WACGAN and IRCNN. Transactions of Internal Combustion Engines, 43(3), 270-278.

[6] Liu, Z. (2020). Theoretical and experimental study on dynamic characteristics of rotor faults of internal combustion generator set [Doctoral dissertation]. Tianjin University.

[7] Li, L. (2025). Bearing fault diagnosis method of diesel locomotive engine based on deep learning technology. Modern Manufacture & Upgrading, (9), 154-157.

[8] Yang, X., Bi, F., Cheng, J., et al. (2024). A multiple attention convolutional neural networks for diesel engine fault diagnosis. Sensors, 24(9).

[9] Jin, Z., Liu, J., Xu, M., et al. (2024). An enhanced digital twin-driven fault detection and isolation method based on sensor series imaging mechanism for gas turbine engine. Applied Thermal Engineering, 257(PB), 124308.

[10] Zhang, Y., Liu, P., & Wang, N. (2024). Fault types and maintenance measures of heavy-duty truck engines. Automotive Review, (3), 167-169.

[11] Giuseppe, C. (2022). Machine fault detection methods based on machine learning algorithms: A review. Mathematical Biosciences and Engineering, 19(11), 11453-11490.