CIMAC CONGRESS 25
ZURICH, MAY 19 – 23, 2025

# 2025 | 374

# Research on construction of hybrid feature fault diagnosis model of turbocharger based on data

Turbochargers & Air/Exhaust Management


## Qirong Yang, Harbin Engineering University

Hechun Wang, Harbin Engineering University
Chuanlei Yang, Harbin Engineering University
Yinyan Wang, Harbin Engineering University


DOI: https://doi.org/10.5281/zenodo.15227770

## ABSTRACT

As one of the critical systems of marine diesel engine, turbocharger plays an important role in ensuring the power of ship's sailing provided sustainably and stably by diesel engine and reducing the impact of emissions on the environment. The fault diagnosis of diesel turbocharger is difficult to identify directly, has a large number of faults, and many faults coexist, so the traditional fault diagnosis is difficult to solve the problem of turbocharger fault diagnosis. In this paper, the common fault types and related thermal parameters of diesel turbochargers are studied, and the fault data are obtained by simulation model. The principal component analysis is used to analyze the features of the fault data, and the main influence features of different faults are determined according to the weight value. Combined with machine learning and optimization algorithm, a construction method of hybrid fault diagnosis model based on feature screening is proposed. The results show that the hybrid model can accurately classify faults with a small number of input features. The fault diagnosis model can be effectively applied to the fault diagnosis work of turbocharger.

# 1 INTRODUCTION

As one of the critical systems of Marine diesel engine, turbocharger plays an important role in ensuring the power of marine sailing provided sustainably and stably by diesel engine and reducing the impact of emissions on the environment [1]. Due to the complex structure and long time in the harsh working environment, the fault of turbocharger often occurs, and the fault diagnosis is needed. Regular overhaul and maintenance of turbochargers is of great significance to improve engine fuel economy, performance and reduce exhaust emissions [2].

For turbocharger fault diagnosis, Vlatko et al. [3] used fault tree analysis (FTA) to analyze the fault of a turbocharger, so as to estimate the reliability of the system and predict the cause of the fault. With the development of artificial intelligence algorithms, many studies have applied machine learning to fault diagnosis. Adamkiewicz et al. [4] determined the working parameters of the turbocharger system for monitoring the normal operation of the turbocharger and diagnosing the turbocharger. The relationship between engine shaft speed and boost pressure, as well as maintenance indexes are established by machine learning algorithm. Wei et al. [5] propose a construct method which is based on one-class support vector machine (OSVM), affinity propagation (AP) and Gaussian mixture model (GMM). The multi-fault identification accuracy of Marine turbocharger system is higher, the calculation speed is faster, and the generalization ability is stronger.

In the research of fault diagnosis, it is usually necessary to observe the data containing multiple variables and collect a large amount of data. Multivariable large data sets will undoubtedly provide rich information for research and application, but also increase the workload of data collection to a certain extent. There may be correlations between many variables, which increases the complexity of problem analysis. Therefore, it is necessary to find a reasonable method to minimize the loss of information contained in the original features while reducing the features to be analyzed, so as to achieve the purpose of comprehensive analysis of the collected data [6].

In diesel engine fault diagnosis, PCA method is often used to reduce the dimension of characteristic parameters. Hou et al. [7] studied the problem of unbalanced fault data due to the randomness of marine fault and fault duration. Principal component analysis (PCA) is used to convert high-dimensional fault samples to low-dimensional fault samples to reduce the computational complexity. Su et al. [8] studied the multi-parameter prediction and fault warning of MDEs. A combined neural network prediction model PCA-CNN-BILSTM is proposed based on PCA to reduce dimensionality of data.

In this paper, the main features of turbocharger fault samples are analyzed by PCA, and a hybrid fault diagnosis model based on feature screening is proposed. The contributions of this study are as follows:

(1) The fault samples of turbocharger were obtained by two-stroke diesel engine simulation model, and the fault diagnosis model was constructed based on thermal parameters. The accuracy of fault diagnosis models constructed by different machine learning methods is analyzed.

(2) PCA was used for feature selection, and main features were screened according to feature weights. Analyze the thermal parameters with the highest correlation between different faults of turbochargers.

(3) According to the main characteristics of different faults, a hybrid feature fault diagnosis model based on PCA is proposed. Combined with the optimization algorithm, the model parameters are optimized to further improve the accuracy of the model.

This paper is organized as follows: Section 2 shows the fundamental algorithm theories and the evaluation index. Section 3 is the construction and verification of simulation model, the acquisition of thermal parameters and the acquisition of fault samples. In Section 4, PCA is used for feature screening, and a hybrid fault diagnosis model was constructed with main features for analysis. Conclusions are given in Section 5.

# 2 THEORETICAL BASIS

## 2.1 Machine learning

In this paper, several common machine learning algorithms are selected to construct fault diagnosis models, such as SVM[9-10],BP[11-12],RF[13-14],ELM[15-16], to construct diesel engine fault diagnosis models, and to study the feature analysis of different models.

BP and ELM belong to neural networks. The structure of neural network is shown in Figure 1. BP is a kind of multi-layer feedforward neural network, which is characterized by forward transmission of signal and back propagation of error. ELM is a Singl-hidden Layer Feedforward Neural Network. This method has the advantages of fast learning speed and good generalization performance.

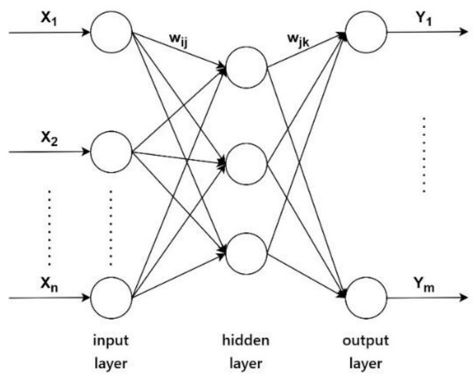


Figure 1. Neural network structure

Support Vector Machines (SVM) is a binary classification model. Its basic model is the linear classifier with the largest interval in the feature space, which distinguishes it from the perceptron most. SVM also includes kernel techniques, which make it essentially a nonlinear classifier. The learning strategy of SVM is to maximize the interval, which can be formalized as a problem of solving convex quadratic programming. It is also equivalent to the regularized problem of minimizing the hinge loss function.

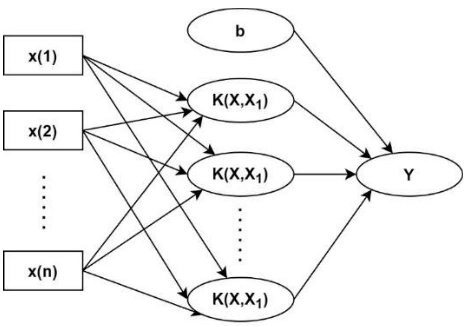


Figure 2. SVM structure

DT is an inductive learning algorithm based on example. The attribute values are compared in the inner nodes (non-leaf nodes) of the decision tree, and the branch down from the node is judged according to different attribute values, and the conclusion is obtained in the leaf nodes of the tree. RF is essentially a classifier containing multiple decision trees. The formation of these decision trees adopts random methods, so it is also called random decision trees. The trees in RF are not related to each other. Finally, the class with the most classification results in all decision trees is the final result.

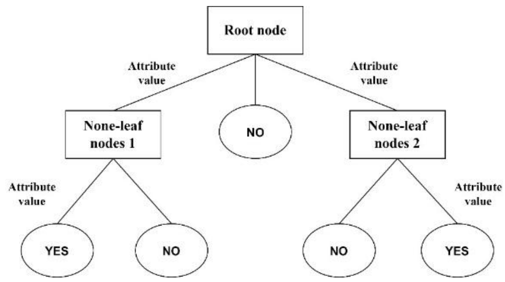


Figure 3. DT structure

## 2.2 Principal component analysis

Since there is a certain correlation between variables, it can be considered to change the closely related variables into as few new variables as possible, so that these new variables are pinions uncorrelated, so that fewer comprehensive indicators can be used to represent various types of information existing in each variable. Principal component analysis [17-18] is one of dimensionality reduction algorithms.

Principal Component Analysis (PCA) is the most commonly used linear dimension reduction method. Its goal is to map high-dimensional data to low-dimensional space through some linear projection, and it is expected that the information content (variance) of the data is maximum in the projected dimension. This uses fewer data dimensions while retaining more of the characteristics of the original data points. The process is shown in Figure 4.

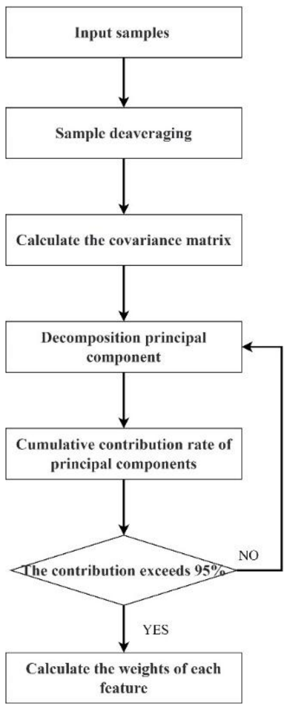


Figure 4. The structure of PCA

### 2.3 Evaluation index

As an evaluation of the real value and the predicted value, the accuracy can directly reflect the classification results in fault diagnosis, and the accuracy also reflects the accuracy of the fault diagnosis model.

F1-score is a statistical measure of the accuracy of a binary classification model. It takes into account both the precision and the recall. The F1-score can be seen as a kind of weighted average of model precision and recall. It has a maximum value of 1 and a minimum value of 0, and a larger value means that the model is more accurate.

In this paper, Micro-F1 is used as the evaluation index of the fault diagnosis model, the total precision and recall of all categories are calculated, and then F1-score is calculated. The calculation equation is shown in Equation 1.

$$F1 = 2 * \frac{precison * recall}{precison + recall} \tag{1}$$

## 3 SIMULATION MODEL CONSTRUCTION AND DATA ACQUISITION

The research object of this paper is MAN 6S35ME-B9.5 Marine low-speed intelligent diesel engine. Table 1 shows the main structural parameters of diesel engines.

Table 1. Main structural parameters of diesel engine

| Parameter | Unit | Value |
|---|---|---|
| Effective Power | kW | 3543 |
| Speed | r/min | 141 |
| Bore | mm | 350 |
| Stroke | mm | 1550 |
| Compression Ratio | — | 21.8 |
| Firing Order | — | 1-5-3-4-2-6 |
| Intake Mode | — | Charge Inter-cooling |

The 1D simulation model of diesel engine is shown in Figure 5. The left side is the diesel engine's intake system and intercooler, the middle part is the diesel engine's crankshaft and six cylinders and the corresponding injector for each cylinder, the right side is the exhaust system, and the bottom is the diesel engine's intake environment, turbocharging system and exhaust environment.

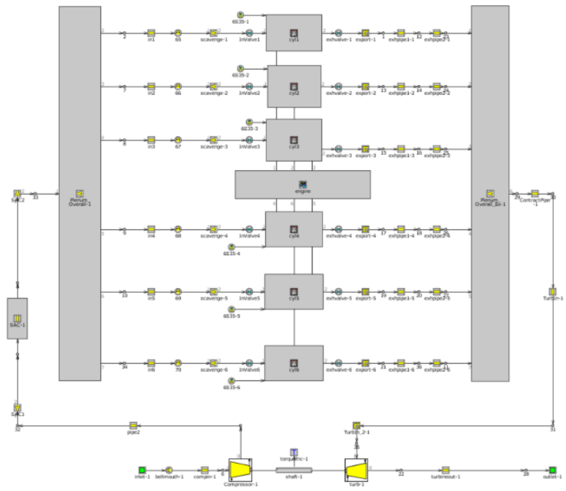

Figure 5. The 1D simulation model of diesel engine

In order to verify the accuracy of the diesel simulation model, parameters representing the performance of the diesel engine were selected for comparison. Table 2 shows the comparison between the simulated values and the test value of the bench test. By comparing the state parameters of the diesel engine under different loads, it can be seen that the error value is less than 3%, which meets the test error standard requirements.

Table 2. Comparison between test values and simulation values

| Parameter | Unit | Test | Simulation | Error |
|---|---|---|---|---|
| Indicated Mean Effective Pressure | bar | 16.52 | 16.29 | -1.39% |
| Indicated Power | kw | 3754.82 | 3824.06 | 1.84% |
| Effective Power | kw | 3382.77 | 3426.46 | 1.29% |
| Specific Fuel Consumption | /g·(kWh) −1 | 175.75 | 172.51 | -1.84% |
| Maximum explosion Pressure | Mpa | 18.98 | 18.87 | -0.58% |

| | | | | |
|---|---|---|---|---|
| Compression Ratio | | 3.61 | 3.67 | 1.66% |
| Compressor Efficiency | | 84.24 | 83.15 | -1.29% |
| Turbine Efficiency | | 86.24 | 87.06 | 0.95% |
| Exhaust Temperature Front Turbine | K | 685.57 | 672.65 | -1.88% |
| Exhaust Temperature After turbine | K | 520.24 | 510.63 | -1.85% |
| Exhaust Pressure Front Turbine | bar | 3.46 | 3.52 | 1.73% |

Since this paper mainly conducts fault simulation under 100% load of 1# cylinder, it is necessary to correct the pressure change in 1# cylinder, and the results are shown in Figure 6.

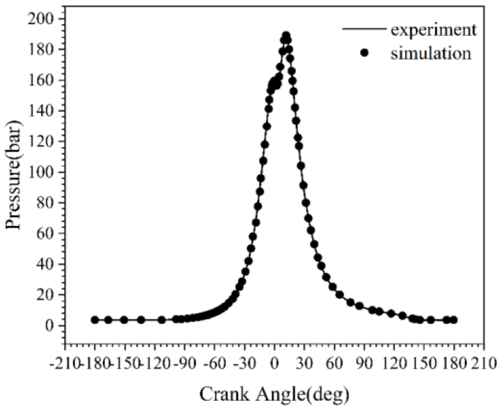


Figure 6. Pressure curve of 1# cylinder under 100% load

As can be seen from Figure 6, the change of test value and simulation value is consistent, the error is small, and the simulation accuracy is high. This model can be used to simulate the fault diagnosis of diesel engine.

The main object of fault research is the turbocharging system, and 6 states are selected for research, as shown in Table 3. The normal state is represented by Label F1. Select five faults of turbocharger and Label F2 to F6. Figure 4 shows the value range of faults in the simulation model.

Table 3. Turbocharger status

| Label | F1 | F2 | F3 | F4 | F5 | F6 |
|---|---|---|---|---|---|---|
| State | Normal | Compressor Fault | Turbine Nozzle Ring Block | Turbine Inefficiency | Turbine Exhaust Pipe Block | Bearing Wear |

Table 4. The value range of faults in the simulation model

| Fault | Parameter | Set Value | Fault Value |
|---|---|---|---|
| Compressor Fault | Compressor efficiency scaling factor | 1 | 0.8~0.95 |
| Turbine Nozzle Ring Block | Turbine flow coefficient | 1 | 0.8~0.95 |
| Turbine Inefficiency | Turbine efficiency scaling factor | 1.05 | 0.8~0.95 |
| Turbine Exhaust Pipe Block | Outlet pressure /bar | 1.05 | 1.05~1.20 |
| Bearing Wear | Turbocharger mechanical efficiency | 1 | 0.8~0.95 |

The working state of diesel engine is determined by various parameters of diesel engine. These parameters include the main engine speed, power, fuel consumption, etc. It has the advantages of large amount of information and stable source.

In this paper, 8 diesel engine thermal parameters related to turbochargers are selected as data features, and Label S1 ~ S8. Where, S1 is the boost pressure, S2 is the compressor outlet temperature, S3 is the turbine inlet temperature, S4 is the turbine exhaust temperature, S5 is the turbine inlet pressure, S6 is the turbine exhaust pressure, S7 is the turbocharger speed, and S8 is the air mass flow rate.

The analysis of thermal parameters mainly depends on the relative deviation of parameters, which can be expressed as:

$$\varepsilon = \frac{x_i - x_0}{x_0} \tag{2}$$

Where, $x_0$ is the parameter values under normal condition; $x_i$ is the parameter values under fault condition.

This paper selects Latin hypercube sampling (LHS) method to collect fault samples. LHS is a random sampling technique designed to reduce correlations between input variables, thereby improving the accuracy of Monte Carlo simulations. It divides the value range of each component into the same interval, and randomly draws a value in each interval to ensure the randomness and uniformity of the sample.

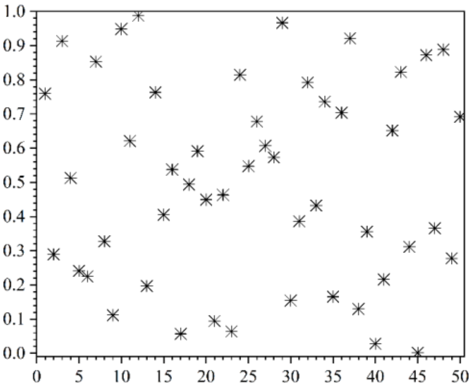

Figure 7. Latin hypercube sampling

For the fault of the turbocharger, 0~5% of the set value in Table 4 is selected as the value of the normal state, and the range of 5%-25% of the deviation from the normal value is selected as the value range of the fault for simulation. LHS is used to select samples without overlapping for the six faults in Table 3. Select 50 samples of each state and construct a dataset of 300 samples.

## 4 FAULT DIAGNOSIS MODEL

The above fault data set is processed, the fault diagnosis model is constructed by using machine learning, and the dimensionality reduction of data samples is processed by using PCA, and the parameters of the fault diagnosis model are optimized by combining the optimization algorithm to obtain a high-precision model.

### 4.1 Feature selection

Feature selection is an important problem in feature engineering, whose goal is to find the optimal feature subset. Feature selection can eliminate irrelevant or redundant features, so as to reduce the number of features, improve model accuracy, and reduce running time. And a simplified model of truly relevant features is selected to help understand the process of data generation.

PCA is used to extract the sample features, and the contribution rate of each component was calculated based on the maximum variance method. The component whose comprehensive contribution value is greater than 95% is selected as the main component. The composite score is calculated by the principal component, and the weight value of the index is obtained.

The fault sample contains 6 fault types and 8 feature parameters, and the weight value of each feature is analyzed according to PCA. Firstly, different fault samples are analyzed to screen out the main features. First, all the data were imported for principal component analysis to screen out the main features. Figure 8 shows the feature weights of the population sample.

As can be seen from Figure 8, the weight of feature S7 and feature S8 is large, indicating that features contain more fault information. The weight of feature S3 is 0, so the feature has little correlation with the fault.

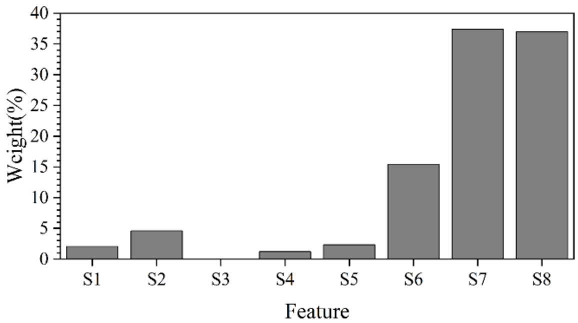


Figure 8. Feature weight of PCA

Next, the impact of 8 features on different faults is analyzed. Table 5 shows the feature weights of different faults. It can be seen from the table that the feature weights of different faults are different. In general, the weight distribution of each feature is similar to Figure 3. The weight of feature S7 and feature S8 is larger, and the weight of feature S3 is the smallest. For fault F3, the weight values of multiple features are the same, indicating that the information distribution is relatively uniform.

Table 5. Feature weights of different faults

| | S1 | S2 | S3 | S4 | S5 | S6 | S7 | S8 |
|---|---|---|---|---|---|---|---|---|

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| F1 | 2.439 | 5.442 | 0 | 1.203 | 2.770 | 26.837 | 30.660 | 30.649 |
| F2 | 0.006 | 26.679 | 0 | 0.009 | 0.006 | 18.833 | 27.233 | 27.233 |
| F3 | 15.852 | 15.846 | 0 | 15.831 | 15.851 | 4.922 | 15.852 | 15.847 |
| F4 | 0.006 | 0.004 | 0 | 0.010 | 0.005 | 13.455 | 43.251 | 43.269 |
| F5 | 0.001 | 0.016 | 0 | 0.007 | 0.000 | 33.309 | 33.319 | 33.347 |
| F6 | 0.004 | 0.004 | 0 | 0.010 | 0.006 | 13.456 | 43.261 | 43.259 |

## 4.2 Construction of fault diagnosis model based on feature selection

The 80% (240/300) samples of fault data were randomly selected as training samples and the rest as test samples. Table 6 shows the accuracy and F1-score of the four models.

Table 6. The accuracy and F1-score of fault diagnosis models

| | SVM | BP | RF | ELM |
|---|---|---|---|---|
| Accuracy | 96.67% | 96.67% | 91.67% | 100% |
| F1-score | 0.964 | 0.965 | 0.903 | 1 |

From Table 6, the accuracy of the model is similar to the value of F1-score. The accuracy of RF model is low, the accuracy is more than 91%. ELM has the highest accuracy and F1-score; In addition to the RF model, the accuracy of the other three models exceeded 96%, indicating that the sample features contained most of the information about the fault.

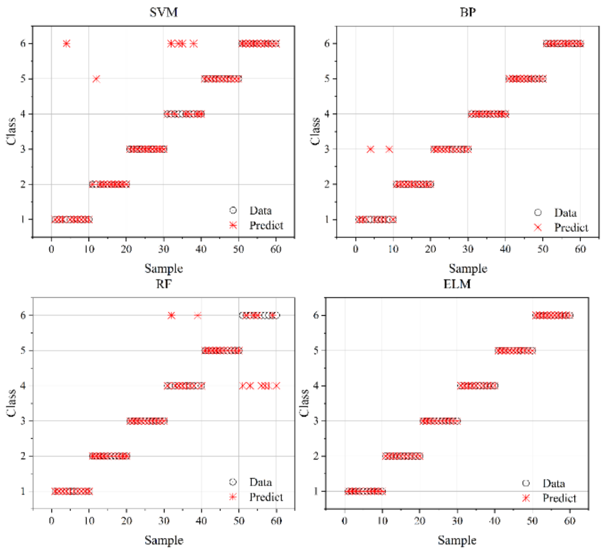


Figure 9. Classification results of the four models

As can be seen from Figure 9, there are differences in the fault class of misclassified samples. For the same sample, the BP model has more classification errors for class 1, the SVM model has more classification errors for class 4, and the RF model mainly concentrates on class 6.

According to PCA results in Figure 8, the importance of eight features of fault data is determined, that is, S8 = S7 > S6 > S2 > S5 > S1 = S4 > S3. Through analysis, the greater the feature weight, the higher the importance of the feature and the more sample information contained.

According to the weight order, different features are selected to construct the fault diagnosis model. Table 7 shows the results of fault diagnosis models constructed with different features.

Table 7. The results of fault diagnosis models constructed with different features.

| | | SVM | BP | RF | ELM |
|---|---|---|---|---|---|
| S7, S8 | Accuracy | 56.67% | 68.33% | 56.67% | 88.33% |
| | F1-score | 0.537 | 0.669 | 0.558 | 0.883 |
| S6, S7, S8 | Accuracy | 71.67% | 85.00% | 78.33% | 93.33% |
| | F1-score | 0.732 | 0.845 | 0.778 | 0.933 |
| S2, S6, S7, S8 | Accuracy | 91.67% | 96.67% | 90.00% | 100.00% |
| | F1-score | 0.916 | 0.966 | 0.899 | 1 |

| | | | | | |
|---|---|---|---|---|---|
| S2, S5, S6, S7, S8 | Accuracy | 98.33% | 96.67% | 93.33% | 100% |
| | F1-score | 0.981 | 0.959 | 0.969 | 1 |

From Table 7, the classification accuracy is low when there are fewer input features, indicating that a small number of features cannot effectively distinguish faults. With the increase of input features, the accuracy of classification and the F1 score of the model are gradually improved. When the input features are S2, S6, S7, S8, the accuracy of BP and ELM exceeds 95%, indicating that features S2, S6, S7, S8 contain most of the fault information. When the input features are S2, S5, S6, S7, S8, the classification results are close to those in the table. Features S1, S3, and S4 have little correlation with fault samples.

According to different faults, the features are selected to construct the mixed feature model of fault diagnosis. From Table 7, when the number of input features is 4, the accuracy of fault diagnosis model is improved, so the model with 3 input features is analyzed.

Figure 10 shows the confusion matrix of the four results, in which there are more classification error samples for Faults F1, F2, F4 and F6. Fault characteristics S6, S7, and S8 of F1, F2, F4, and F6 have overlapping information and cannot be distinguished from faults. Therefore, it is necessary to add feature input appropriately or further distinguish faults.

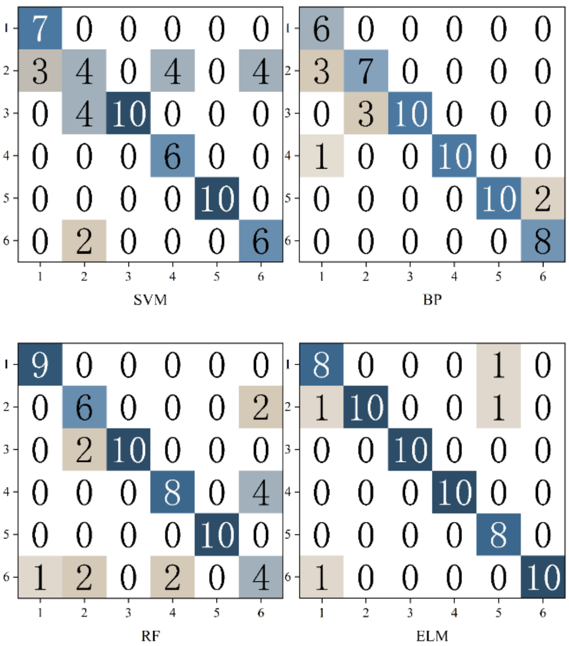


Figure 10. Confusion matrix of the four Few-feature models

According to Table 5, the features are distinguished. For faults F1, F4, F5 and F6, S6, S7 and S8 are used to construct the model. For faults F2 and F3, features S2, S7 and S8 are used for model construction. Fault diagnosis models with hybrid features are constructed respectively. Table 8 shows the classification results of the four hybrid features models, and Figure 11 shows the confusion matrix of the results.

Table 8. Accuracy and F1-score of hybrid features models

| | SVM | BP | RF | ELM |
|---|---|---|---|---|
| Accuracy | 68.33% | 76.67% | 91.67% | 98.33% |
| F1-score | 0.673 | 0.749 | 0.917 | 0.983 |

According to Table 7 and Table 8, by distinguishing the input characteristics of different faults, the classification accuracy of SVM model and BP model is reduced, while the classification accuracy of RF model and ELM model is improved.

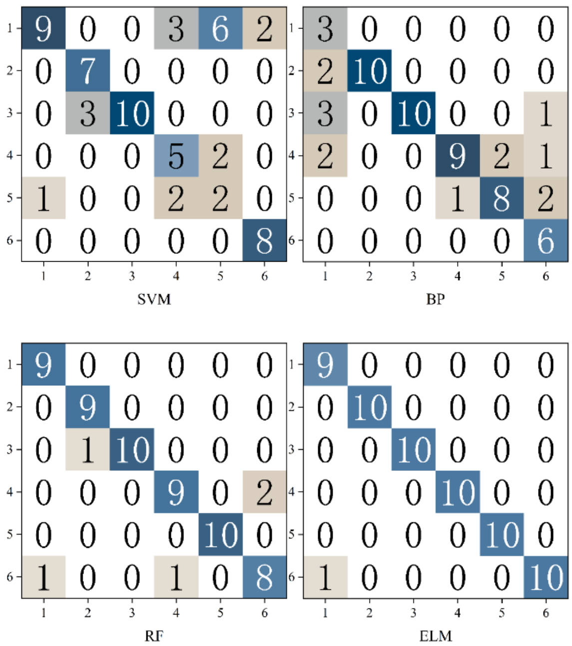


Figure 11. Confusion matrix of hybrid features models

Compared with Figure 10, by distinguishing input features, the error classification samples of faults F1, F2, F4 and F6 are reduced. The number of error classification samples of RF model and ELM model is reduced, and the accuracy is improved. However, for SVM model and BP model, the error classification samples of fault F5 increase, resulting in a decrease in the overall accuracy.

### 4.3 Construction of hybrid feature model

According to the feature importance of different problems, PCA can screen out the main features of different faults, eliminate the redundant features, and simplify the model. Distinguishing input features can reduce the number of misclassified samples and improve the accuracy of classification to a certain extent. In order to further improve the accuracy of the fault diagnosis model, an optimization algorithm is used to optimize the parameters of models. Combined with optimization algorithm, a new construction method of fault diagnosis hybrid feature model is proposed. The process is shown in Figure 12.

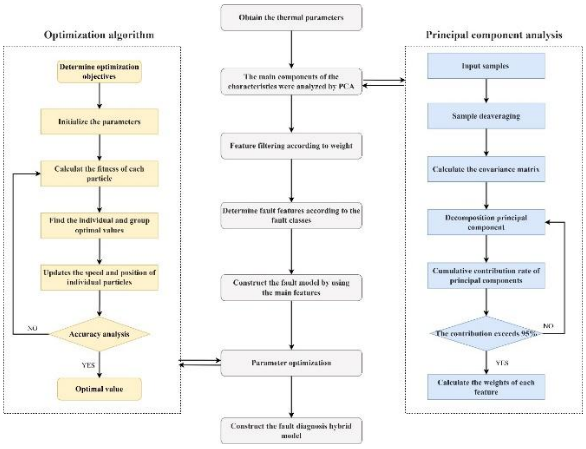


Figure 12. Flowchart of fault diagnosis hybrid model construction

The construction method of fault diagnosis hybrid feature model is as follows:

Step 1: Obtain the fault data of the diesel turbocharger, collect the corresponding thermal parameters, and build the fault data set.

Step 2: Use PCA to extract the main components of fault features, and the feature weights are calculated according to the total contribution rate.

Step 3: Filter the main features of different faults according to the feature weight. The fault diagnosis model is constructed and the results are analyzed.

Step 4: Optimize the parameters of the model using the optimization algorithm to further improve the classification accuracy of the model.

Step 5: Export the fault diagnosis hybrid feature model.

In machine learning, the optimizer is a crucial component. It is responsible for adjusting the parameters of the model during model training to minimize or maximize a certain loss function. The goal of the optimizer is to find a set of parameters such that the error between the model predictions on a given data and the actual results is minimal. In this paper, particle swarm optimization (PSO) was used to optimize the parameters [19-20].

Table 9 shows the optimized classification results of the four models with input feature 3. Compared with Table 8, it can be seen that the classification accuracy of SVM model and BP model has been effectively improved, while the parameters of RF model and ELM model are already optimal.

Table 9. Classification results of the optimized hybrid features models

| | SVM-OPT | BP-OPT | RF-OPT | ELM-OPT |
|---|---|---|---|---|
| Accuracy | 80.00% | 86.67% | 91.67% | 98.33% |
| F1-score | 0.793 | 0.749 | 0.917 | 0.983 |

Figure 13 shows the classification results of the Few-feature (FF) model, Hybrid features (HF) model and Hybrid feature model after optimization (HF-OPT). As can be seen from the figure, compared with FF model and HF model, HF-OPT model can effectively reduce the misclassified samples and improve the accuracy. Among the four machine learning algorithms, RF model and ELM model have stronger sample processing ability and higher model precision.

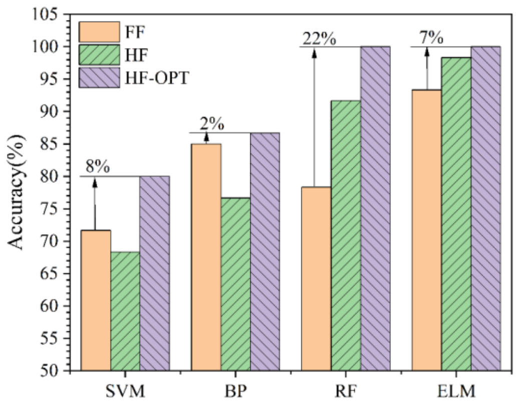


Figure 13. Comparison of accuracy of the three models

## 6 CONCLUSION

This paper analyzes the typical faults of turbochargers, and obtains the fault samples of turbochargers through simulation models. The fault diagnosis model of turbocharger is constructed by using machine learning. In this paper, a hybrid feature fault diagnosis model construction method is proposed by combining PCA and optimization algorithm.

(1) Feature selection based on PCA can effectively eliminate redundant features. In the case of ensuring the high accuracy of classification results, the fault diagnosis model is simplified. In the eight features of turbocharger in this paper, the accuracy of fault diagnosis model can reach 96% by inputting four main features.

(2) The main impact features of different faults are different. By distinguishing the input features of different fault samples, the number of misclassified samples can be effectively reduced. The accuracy of fault diagnosis model is further improved by reducing the impact of irrelevant features.

(3) With the reduction of input features, the model's ability to classify fault samples decreases. The hybrid feature fault diagnosis model can improve the classification ability of fault samples under a small number of input features.

# References

[1] Anantharaman, M., Islam, R., Khan, F., et al. 2019. Data Analysis to Evaluate Reliability of a Main Engine. *TransNav: International Journal on Marine Navigation and Safety of Sea Transportation*, 13(2):403-407.

[2] Nyongesa, J.A., Park, H.M., Lee, M.C., et al. 2024. Experimental evaluation of the significance of scheduled turbocharger reconditioning on marine diesel engine efficiency and exhaust gas emissions. *Ain Shams Engineering Journal*, 15(8):102845-102845.

[3] Vlatko, K., Josip, O., Ladislav, S. and Jelena, Č. 2020. Fault Tree Analysis and Failure Diagnosis of Marine Diesel Engine Turbocharger System. *Journal of Marine Science and Engineering*, 8(12):1004-1004.

[4] Adamkiewicz, A. and Nikończuk, P. 2022. An attempt at applying machine learning in diagnosing marine ship engine turbochargers. *Eksploatacja i Niezawodność – Maintenance and Reliability*, 24(4),

[5] Wei, Y., Liu, H., Chen G. and Ye, J. 2020. Fault Diagnosis of Marine Turbocharger System Based on an Unsupervised Algorithm. *Journal of Electrical Engineering & Technology*, 15(1):1331-1343.

[6] Pani, K.A. 2021. Non-linear process monitoring using kernel principal component analysis: A review of the basic and modified techniques with industrial applications. *Brazilian Journal of Chemical Engineering*, 39(2):1-18.

[7] Hou, L., Hou, L., Zhang J. and Du B. 2020. A Fault Diagnosis Model of Marine Diesel Engine Fuel Oil Supply System Using PCA and Optimized SVM. *Journal of Physics: Conference Series*, 1576(1):012045-.

[8] Su, Y., Gan H. and Ji, Z. 2024. Research on Multi-Parameter Fault Early Warning for Marine Diesel Engine Based on PCA-CNN-BiLSTM. *Journal of Marine Science and Engineering*, 12(6):965-965.

[9] Li S., Wang Y. and He, Z. 2024. Bushing fault diagnosis based on SVM and the improved sparrow search algorithm. *Journal of Physics: Conference Series*, 2815(1):012031-012031.

[10] Rajeev K. and S. R A. 2024. Bearing fault diagnosis using multiple feature selection algorithms with SVM. *Progress in Artificial Intelligence*, 13(2):119-133.

[11] Zenghua C., Lingjian Z., He L., et al. 2024. Research on bearing fault diagnosis based on improved genetic algorithm and BP neural network. *Scientific Reports*, 14(1):15527.

[12] Jiao R., Li S., Ding Z. et al. 2024. Fault diagnosis of rolling bearing based on BP neural network with fractional order gradient descent. *Journal of Vibration and Control*, 30(9-10):2139-2153.

[13] Jing T., Yuwei Z., Fengling Z., et al. 2023. A novel intelligent method for inter-shaft bearing-fault diagnosis based on hierarchical permutation entropy and LLE-RF. *Journal of Vibration and Control*, 29(23-24):5357-5372.

[14] Gao R.J., Liu K.Y., Duan J.C., et al. 2024. Research on fault diagnosis of electric gate valve in nuclear power plant based on the VMD-MDI-ISSA-RF model. *Annals of Nuclear Energy*, 207:110701-110701.

[15] Xiang W., Wu Y., Peng C., et al. 2024. A Fault Diagnosis Method for Electric Check Valve Based on ResNet-ELM with Adaptive Focal Loss. *Electronics*, 13(17):3426-3426.

[16] Ke, W. and Fengqing, Q. 2023. Fault diagnosis of gearbox based on Fourier Bessel EWT and manifold regularization ELM. *Scientific Reports*, 13(1):14486-14486.

[17] Díaz, M.J., Barboza, C.C., Portnoy, I. and Espinosa, T.C. 2024. A Novel Approach for Fault

Detection and Diagnosis in Multi-mode Processes Based on PCA, Random Forest, and K-means Clustering. *International Journal of Control, Automation and Systems*, 22(11):3499-3508.

[18] Khurshid, A. and Pani, K.A. 2024. An integrated approach combining randomized kernel PCA, Gaussian mixture modeling and ICA for fault detection in non-linear processes. *Measurement Science and Technology*, 35(7).

[19] Jiang, D. and Wang, Y. 2023. Fault Diagnosis of Power Electronic Circuits Based on Adaptive Simulated Annealing Particle Swarm Optimization. *Computers, Materials & Continua*, 76(1):295-309.

[20] Shaban, G.E., Ahmed, H.K. and Moustafa A.M.H. 2023. Fault diagnosis for PV system using a deep learning optimized via PSO heuristic combination technique. *Electrical Engineering*, 105(4):2287-2301.