The Intelligent Fault Diagnosis of Diesel Engine Based on the Ensemble Learning

Mingqi Shao¹,a, Jin Wang²,* and Sibo Wang¹,b

¹School of Energy and Power Engineering, Wuhan University of Technology, Wuhan, PRC
²Rizhao Polytechnic, Rizhao, China

*Corresponding author e-mail: 568857508@qq.com, a1014232997@qq.com,
b565380793@qq.com

Abstract. As the source of power, internal combustion engine is widely used. Because the structure of internal combustion engine is complex and the working condition is bad, the possibility of failure is great. The condition monitoring and fault diagnosis of internal combustion engine can discover and eliminate the faults in time, and ensure the safety and reliability of the operation process. With the rapid progress of machine learning technology, fault diagnosis system is gradually moving towards intelligent development. The ensemble learning uses different methods to change the distribution of the original training samples, so as to build multiple different classifiers, and combine these classifiers to get a stronger classifier to make the final decision. In this paper, through the simulation model of 7k98mc diesel engine to simulate the normal and fault conditions of diesel engine, the ensemble learning algorithm is verified. The results show that the ensemble learning algorithm has higher accuracy and better stability than shallow classifiers.

1. Introduction
As the core equipment of marine power plant, marine diesel engine is widely used in the field of ships. However, due to its poor working conditions, it is prone to failure, which will affect the operation of the ship and may cause huge economic losses and even cause damage to key equipment and endanger personal safety [1]. If the fault can be detected and repaired according to the relevant operating data of the diesel engine, the accident can be avoided and the economic loss can be reduced. In recent years, the rapid progress of computer artificial intelligence and machine learning technology makes the fault diagnosis system gradually develop towards intelligent direction. Fuzzy set theory, artificial neural network (ANN), support vector machine (SVM), decision tree, hidden Markov chain, Bayesian network and other technologies have been widely used. Intelligent fault diagnosis is to simulate the process of human thinking, through the effective acquisition, transmission and processing of diagnostic information, simulation of human experts, with flexible strategies to make accurate judgments and optimal decisions on the running status and faults of monitoring objects. Compared with the fault identification method based on signal processing and feature extraction, intelligent fault diagnosis has the ability of learning function and automatic acquisition of diagnosis information for real-time fault
diagnosis, which greatly reduces the dependence on expert experience and human factors, so it becomes
the key application technology to realize mechanical fault diagnosis.

Tao x et al. [2] decomposes the vibration signals of normal and fault states by wavelet packet,
descrives the clustering of characteristic parameters by clustering analysis, and calculates the
Mahalanobis distance between unknown and known samples to diagnose faults. Liu et al. [3] used back
propagation neural network (BPNN) to classify diesel engine faults, and used hybrid particle swarm
optimization algorithm (psd-dv) with differential operator to adjust the weight and threshold of BP
neural network, so as to improve the convergence of BP neural network. Togun et al. [4] used BP neural
network to predict torque and brake fuel consumption rate of internal combustion engine, took
parameters such as speed and throttle position as input eigenvector, and torque and brake fuel
consumption rate as output. The test results showed that neural network can effectively predict torque
and brake fuel consumption rate. Liu et al. [5] collected the top dead center signal and cylinder head
vibration signal of the internal combustion engine, intercepted the vibration signal of each cylinder
combustion section according to the firing sequence and extracted the characteristic value, and input the
BP neural network to diagnose the misfire fault.

Among them, artificial neural network (ANN) and support vector machine (SVM) are commonly
used shallow classifier. In the neural network fault diagnosis method, the fault feature corresponds to
the input point of the neural network, and the diagnosis result corresponds to the output node of the
neural network. If there are too many training samples, the learning process will converge slowly and
easily fall into the local minimum, otherwise, if there are too few samples, there will be over fitting; it
is not easy to find the appropriate method to select the network node weight coefficient. For the fault
diagnosis of complex system, if the non-significant feature information is not fully mined, it is difficult
to achieve the purpose of accurate fault location and identification. Different from Ann's idea of relying
on large samples to train network, SVM is mainly suitable for the case of less fault samples. Its idea is
to use kernel function to map nonlinear problems in input space to high-dimensional feature space, and
construct linear function discriminator in high-dimensional space. However, SVM can only be used as
a binary classifier, although it can achieve a combination classifier with other methods, it usually
encounters problems such as computational difficulties. In the face of the non-significance of the fault
characteristics of complex systems and the variability of the relationship between the characteristics and
the fault, SVM seems to be inadequate. All in all, the above two methods are shallow machine learning
methods. The simplicity of their structure makes it difficult to accurately grasp the non-significance and
uncertainty changes of fault features, as well as the complex mapping relationship between features and
faults, and to mine the deep information in features. Therefore, further research and learning of the
ensemble learning model is the only way for fault feature extraction and fault classification.

2. The principle of different classifiers

2.1. Support Vector Machine
The function of SVM is to solve convex quadratic programming problems. For a given training set
\( (x_i, y_i), i = 1, 2, 3, \ldots, n \), \( x_i \) represent data point features, \( y_i \in \{-1, 1\} \) represent labels. The function of the
classifier is to find an optimal plane in the space \( \{ x \mid \beta^T x = 0, \beta \in R^n \} \) of the input variable \( x_i \). The
problem is turned into a dual problem by using Lagrange optimization method. The solution flow of
linear separable support vector as follows:

1. The original question is:

\[
\max \beta
\]

\[
y_i (\omega^T x_i + b) = \hat{y}_i \leq \hat{y}, \quad i = 1, 2, \ldots, n
\]

(2) It is specified that minimum function is 1, and original problem deformation is:
\[
\Phi(w) = \frac{1}{2} ||w||^2 = \frac{1}{2}(w \cdot w)
\]  
(3)

\[
y_i \left( (w \cdot x_i) + b \right) \geq 1, i = 1, \ldots, l
\]  
(4)

(3) Constructing Lagrange function:

\[
L(w, b, \alpha) = \frac{1}{2} ||w||^2 - \sum_{i=1}^{l} \alpha_i \left( y_i \left( (x_i \cdot w) + b \right) - 1 \right)
\]  
(5)

(4) According to the duality of Lagrangian function and some solutions, the original problem is transformed into:

\[
\min_{\alpha} \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \langle x_i, x_j \rangle - \sum_{i=1}^{n} \alpha_i
\]  
\text{s.t.} \alpha_i \geq 0, i = 1, \ldots, n
\]  
\[
\sum_{i=1}^{n} \alpha_i y_i = 0
\]  
(6, 7)

(5) The \( a^* \) is solved by the SMO algorithm.

(6) The \( w \) and \( b \) is solved by:

\[
w^* = \sum_{i=1}^{N} a_i^* y_i x_i
\]  
(8)

\[
b^* = y_j - \sum_{i=1}^{N} a_i^* y_i (x_i \cdot x_j)
\]  
(9)

(7) The classification decision function is get by:

\[
f(x) = \text{sign} \left( \sum_{i=1}^{N} a_i^* y_i (x \cdot x_i) + b^* \right)
\]  
(10)

Please follow these instructions as carefully as possible so all articles within a conference have the same style to the title page. This paragraph follows a section title so it should not be indented [2-3].

2.2. KNN classifier

The principle of KNN classifier is to classify by measuring the distance between different eigenvalues: if K samples in the feature space and most of the nearest samples in the feature space are of the same class, the similarity between the samples to be tested and the K samples will be taken as the class weight of the class of k nearest neighbour text, and the samples to be tested will be classified into the class with the largest similarity. Then the samples will also belong to this class Farewell. In the decision-making of classification, the method only determines the category of the samples to be tested according to the category of the nearest one or several samples.

The principle algorithm of KNN is as follows:

(1) Nearest neighbor method
It is known that \( S = \{(x_1, \theta_1), (x_2, \theta_2), \ldots, (x_n, \theta_n)\} \), \( x_i \) is feature vector of sample \( I \), and \( \theta_i \) is corresponding label. It is assumed that there are \( C \) categories, and \( \theta_i \in \{1, 2, \ldots, C\} \). Defining the distance \( d(x_i, x_j) \) of two samples, usually using Euclidean distance or Manhattan distance, the problem can be avoid of matching between objects. The Euclidean distance is:

\[
d(x, y) = \sqrt{\sum_{k=1}^{n} (x_k - y_k)^2}
\]

The Manhattan distance:

\[
d(x, y) = \sum_{k=1}^{n} |x_k - y_k|
\]

To the unknown samples \( x \), the nearest sample in \( S \) is \( x^* \).

\[
d(x, x^*) = \min_{j=1, \ldots, N} d(x_i, x_j)
\]

If it is written in the form of discriminant function, the discriminant function of category \( \omega_i \) can be written as follows:

\[
g_i(x) = \min_{j=1, \ldots, N} d(x_i, x_j), i = 1, \ldots, C
\]

The decision rule is the comparison size of various functions, namely:

\[
g_k(x) = \min_{i=1, \ldots, c} g_i(x), x \in \omega_k
\]

(2) K-nearest method

There is a certain risk in the decision-making based on the nearest sample. Select the first several known samples closest to the new sample, and use their class voting to determine the class of the new sample. This method is called k-nearest neighbor method. If \( N \) samples which is known belong to \( c \) classes \( w_i \) respectively, \( \omega_i, i = 1, \ldots, c \), the first \( k \) neighbors of new sample \( x \) in these samples, the discriminant function of class \( \omega_i \) is as follows:

\[
g_i(x) = ki, i = 1, \ldots, C
\]

The decision rules are:

\[
g_k(x) = \max_{i=1, \ldots, c} g_i(x), x \in \omega_k
\]

2.3. Random Forest classifier

Random forest is an algorithm that integrates multiple trees through the idea of integrated learning. Each tree is a classifier. For a sample, each tree has a classifier. The random forest classifier will take the category with the most votes as the output result. Random forest algorithm is divided into three parts: the construction of training set, the generation of decision tree and the generation of algorithm.

Decision tree is the tree structure of data structure with special meaning. Each non final node of decision tree represents the characteristics of data. According to different eigenvalues, the data is divided into several subsets. These subsets are all branches of the root node, and then each branch is divided.
The final node of decision tree is the final category label of data. For a sample feature vector, from the top of the decision tree to the final node, the category label is the category of the sample vector. The structure of decision tree is shown in Fig. 1

**Figure 1.** The structure of decision tree

Algorithm process of decision tree:
1. Information entropy: it is used to measure the uncertainty of random variables. If the things to be classified may be classified into \( n \) categories, the \( x_1, x_2, \ldots, x_n \) can get each probability \( p_1, p_2, \ldots, p_n \), then the entropy of \( x \) is defined as:

\[
H(X) = -\sum_{i=1}^{n} p_i \log p_i
\]  

(18)

2. Conditional entropy: Assuming that there are random variables \( (x, y) \), its joint probability distribution is:

\[
P(X = x_i, Y = y_j) = p_{ij}, i = 1, 2, \ldots, n, j = 1, 2, \ldots, m
\]

3. Information gain: Indicates that degree to which the uncertainty of \( Y \) is reduced when the information of characteristic \( X \) is known. It is defined as:

\[
g(D, A) = H(D) - H(D|A)
\]  

(19)

The calculation process:
1. Assuming that there is a set of data with \( n \) at node \( k \), the splitting function at this node can be defined as:

\[
f_{h}(x, \omega_n) = I(\Phi(x)^{T} \omega > 0): \mathbb{R}^{d+1} \times \mathbb{R}^{d+1} \rightarrow \{0,1\}
\]  

(20)

Where \( \omega_n \) is the required parameter in the training process, and \( I(\square) \) represent the matrix \( x \) of indicative function \( \Phi( x ) \).

2. In the form of \( f_{h}(x) \), the decision tree is recorded as disjunctive normal form:

\[
h(x) = \bigvee_{i=1}^{n} (\bigwedge_{j \in R_i} f_j(x) \bigwedge_{j \in L_i} \neg f_j(x))
\]  

(21)

Where \( n \) is the number of positive leaf nodes, representing the right parent node in the path from the \( i_{th} \) positive node to the root node, and the left parent node in the path from the \( i_{th} \) positive node to the root node.
(3) The S-function of logistic is used to express the determinant as follows:

\[
\bar{f}_n(x, \omega_k) = \frac{1}{1 + e^{-\sum_{j=1}^{n} \omega_{kj} \phi_j}}
\] (22)

Bring the differentiable disjunctive normal form into equation to get the differentiable normal form of decision tree:

\[
\tilde{h}(x) = 1 - \prod_{i=1}^{n}(1 - g_i(x))
\] (23)

(4) Defining the error function:

\[
E(\tilde{h}, S) = \sum_{m=1}^{M}(y_m - \tilde{h}(x_m))^2
\] (24)

(5) The gradient descent method is used to solve the minimum value:

\[
\frac{\partial E}{\partial \omega_{kj}} = -2\chi_j(y - \tilde{h}(x)) \times \left\{ \sum_{i \in R_L} (\prod_{r \in I}(1 - g_r(x))(1 - \tilde{f}_r(x))) - \sum_{i \in R_L} (\prod_{r \in I}(1 - g_r(x))g_l(x)\tilde{f}_k(x)) \right\}
\] (25)

Through the principle of random forest classifier, we can know that the construction of random forest is random, which can greatly reduce the phenomenon of random forest over fitting.

3. Simulation results

3.1. Simulation model

![Figure 2. The structure of diesel engine model](image)

According to the mathematical model of diesel engine, the Simulink module of each system of diesel engine is built. In order to facilitate the difference between multiple devices, suffix 1, 2, 3 are added after the model. The specific module structure is shown in Fig.2

The prototype of the simulation is 7k98mc diesel engine, which is a large-scale low-speed two-stroke diesel engine and widely used as the power unit of large ships. The main technical parameters are as follows in the Table.1.
Table 1. 7K98MC diesel engine parameters

| parameters              | value             |
|-------------------------|-------------------|
| Bore                    | 2600 mm           |
| Number of cylinders     | 7                 |
| Brake power at MCR      | 40055 kW          |
| Engine speed at MCR     | 94 r/min          |
| BMEP at MCR             | 18.2 bar          |
| Compression ratio       | 17.5              |
| Connecting and length   | 3220 mm           |
| EVO                     | 74°CA before BDC  |
| EVC                     | 87°CA after BDC   |
| SPO                     | 42°CA before BDC  |
| SPC                     | 42°CA after BDC   |
| Turbocharger units      | 3                 |

3.2. Model simulation calculation and verification

In order to verify the accuracy of the model, this paper selects the diesel engine to simulate under the load of 50% (74.6 R/min), 75% (85.4 R/min) and 100% (94 R/min). The governor will automatically adjust the diesel engine model to the set speed, and record the simulation value of the model under the corresponding steady state condition. In this paper, the solver of the model adopts the constant step size continuous algorithm, and the integration algorithm adopts the oldel1 method. The sampling step size is 0.001 and the sampling time is 150s.

In the verification, the test data of diesel engine is important. By comparing the test data with the simulation data, the accuracy and rationality of the model can be verified. The test data of diesel engine is shown in Table 2.

Table 2. Steady state simulation results and comparison with shop trials data

| Engine load (% MCR) | Power (kW) | BSFC g/(kWh) | $P_{\text{max}}$ bar | $P_{\text{com}}$ bar | $N_{\text{tc}}$ r/min | $P_{\text{sca}}$ bar | $T_{\text{ex}}$ K |
|---------------------|------------|--------------|---------------------|---------------------|----------------------|-------------------|------------------|
| 50                  | Calculated | 20226        | 179.60              | 99.43               | 72.62                | 7896              | 2.07             | 593.68          |
|                     | Measured   | 20028        | 179.46              | 98.00               | 72.00                | 7782              | 2.05             | 600.17          |
|                     | Error (%)  | 0.99         | 0.08                | 1.46                | 0.86                 | 1.47              | 0.89             | -1.08           |
| 75                  | Calculated | 30345        | 175.03              | 127.91              | 100.13               | 9710              | 2.86             | 611.63          |
|                     | Measured   | 30041        | 176.01              | 128.00              | 99.90                | 9670              | 2.87             | 614.57          |
|                     | Error (%)  | 1.01         | -0.55               | -0.07               | 0.23                 | 0.41              | -0.40            | -0.48           |
| 100                 | Calculated | 40462        | 177.69              | 138.40              | 125.67               | 10943             | 3.58             | 660.83          |
|                     | Measured   | 40055        | 177.98              | 139.40              | 126.70               | 10946             | 3.63             | 663.90          |
|                     | Error (%)  | 1.02         | -0.16               | -0.71               | -0.81                | -0.03             | -1.39            | -0.46           |

It can be seen from table 2 that the comparison data of steady-state simulation results and workshop test data are consistent, which verifies the rationality and correctness of the model. Therefore, the model can be used to simulate the thermal failure of diesel engine.

3.3. Fault simulation

In this experiment, the data measured under five common working conditions (normal, compressor fault, compressor fault, injection timing retry) are 100 sets of data in each working condition. There 15 features of each dataset, which are effective power, effective oil consumption rate, air-fuel ratio, maximum explosion pressure of cylinder, maximum combustion temperature of cylinder, outlet temperature of compressor, outlet temperature of compressor, outlet temperature of intercooler, pressure of scavenging
box, temperature of scavenging box, pressure of exhaust pipe, temperature of exhaust pipe, outlet pressure of turbine, outlet temperature of turbine, speed of turbine.

If in cylinder parameters are involved in fault simulation data acquisition, parameter selection and setting of 1# cylinder are shown in Table 3.

### Table 3. Parameter selection for diesel engine fault simulation

| Number | Fault                        | Parameters                     | Normal value |
|--------|------------------------------|--------------------------------|--------------|
| 1      | Compressor fault             | Compression efficiency         | 0.873        |
| 2      | Air cooler fault             | Cooler efficiency              | 0.90         |
| 3      | Injection timing advance     | Starting time of combustion    | 3.5°CA       |

Using the form of one-hot to preprocess the data to simplify the original data [1,0,0,0] means normal condition, [0,1,0,0] means compressor fault, [0,0,1,0] means air cooler fault, [0,0,0,1] means injection timing advance.

3.4. The comparison of different methods

For this diagnosis system, 70% of the training data, 30% of the test data and 30 times of training are used. The test results of different classifiers are as follows in the Fig.

![Accuracy Chart](image)

**Figure 3.** Comparison of accuracy of different algorithms

Fig.3 shows the comparison chart of the accuracy of three algorithms. It can be seen that the accuracy of the three algorithms to identify the fault state fluctuates significantly. The accuracy of SVM to identify the fault is basically maintained around 89%, the accuracy of DT to identify the fault is basically maintained around 91%, and the accuracy of RT to identify the fault is basically maintained around 91%. The result of rate recognition is the best, which is kept near 95%, and has a high accuracy, which shows that the fault model has a certain accuracy.
Fig. 4 shows the comparison of the average value of the three algorithms for the accuracy rate. It can be seen that the average value of SVM for the accuracy rate of fault recognition is 81.6%, and the average value of DT for the accuracy rate of fault recognition is 84.7%, while RT for the accuracy rate of fault recognition is still the highest, 95.1%, which is 13.5% higher than SVM and DT for the accuracy rate of fault diagnosis. And 10.4%. According to the simulation test results of fault model, it is concluded that RT fault analysis algorithm has better diagnosis effect than SVM and DT in fault accuracy analysis.

In the Fig. 5, the red line represents the median, the top line represents the maximum, the bottom line represents the minimum, the upper edge of the box represents the upper quartile, and the lower edge of the box represents the lower quartile. The higher the red line, the narrower the box shape, the smaller the fluctuation, the higher the stability, and the higher the position, the higher the recognition degree of the fault is, the more accurate it is. It can be seen that the recall rate of RF in fault analysis is better than that of SVM and DT.
Figure 6. Comparison of F1-score of different algorithms

Fig.6 shows the box chart of the three algorithms for the comprehensive evaluation index. First, from the box shape, the comprehensive evaluation index gradually narrows from SVM to RF shape, and then from the median, the comprehensive evaluation index gradually increases from SVM to RF value. Finally, through the comprehensive evaluation index of PEMFC in theory, the reliability of RF to the accuracy of fault model analysis is improved.

4. Conclusion
The integrated classifier is an important classification method. Compared with the simple classifier, it has higher accuracy, can process the input samples with high latitude features, does not need to reduce the dimension, and has the ability to evaluate the importance of each feature in the classification problem. In addition, the integrated classifier can obtain the unbiased estimation of internal generation error in the generation process besides, the random forest classifier can also deal with the default value problem well. As a classifier widely used in various fields, random forest is worth studying in terms of running speed and accuracy.

Acknowledgments
This work supported by the Students innovation and entrepreneurship training program of Wuhan University of Technology (20191049705016).

References
[1] Cai C, Zhang C, Gang L. A novel fault diagnosis approach combining SVM with association rule mining for ship diesel engine. IEEE International Conference on Information & Automation, 2017.
[2] Tao X, Lu C, Lu C, Wang Z. An approach to performance assessment and fault diagnosis for rotating machinery equipment. EURASIP J ADV SIG PR 2013.
[3] Bo L, Pan H, Li X. An Expert System for Fault Diagnosis in Diesel Engine Based on Wavelet Packet Analysis and Hybrid PSO-DV Based Neural Network., 2010.
[4] Togun NK, Baysec S. Prediction of torque and specific fuel consumption of a gasoline engine by using artificial neural networks. APPL ENERG 2010; 87: 349.
[5] Liu X, Zhang M, Yao F. Adaptive fault tolerant control and thruster fault reconstruction for autonomous underwater vehicle. OCEAN ENG 2018; 155: 10.