

Article

A New Incipient Faults Diagnosis Method Combining SAE and AdaBoost Algorithm for Vehicle Power Supply with Imbalanced Datasets

Yinlong Han ^{1,2,3}, Aimin An ², Wei Li ² and Haiying Dong ^{4,*}

¹ School of New Energy Engineering, Jiuquan Vocational Technical University, Jiuquan 735000, China; hanyinlong0718@163.com

² School of Automation and Electrical Engineering, Lanzhou University of Technology, Lanzhou 730030, China; anaiminll@163.com (A.A.); liwei@lut.edu.cn (W.L.)

³ Gansu Higher Research Institute of the Ministry of Education, Lanzhou 730030, China

⁴ School of New Energy and Power Engineering, Lanzhou Jiaotong University, Lanzhou 730030, China

* Correspondence: hdyong@ljztu.edu.cn

Abstract

For the incipient faults of vehicle power supply under imbalanced datasets, the traditional shallow network has the problems of limited feature extraction ability and the insufficient generalization ability of a single network model. In this paper, an AdaBoost-SAE deep ensemble diagnosis method, which combines the Stacked Auto-Encoder (SAE) deep network and Adaptive Boosting (AdaBoost) algorithm, is proposed. First, SAE is used as a weak classifier to learn and extract incipient fault features from the monitoring data of vehicle power supply. Secondly, in the iterative training process of the model, the classification performance of a single SAE is improved step-by-step by constantly adjusting the weights of the misclassified samples in the training set. Finally, the multiple weak classifiers are combined into strong classifiers by linear weighting to achieve accurate identification of incipient faults under imbalanced datasets. The test results demonstrate that the proposed method can mine deeper features of incipient faults and effectively improve the adverse effects of sample imbalance. Compared with traditional fault diagnosis models and a single SAE, the accuracy of the incipient fault diagnosis can reach 96.6%. Furthermore, the F1-scores of the various working conditions also increased significantly.

Keywords: vehicle power supply; incipient fault diagnosis; imbalanced datasets; stacked auto-encoder; adaptive boosting algorithm



Academic Editor: Jie Zhang

Received: 4 September 2025

Revised: 9 October 2025

Accepted: 15 October 2025

Published: 18 October 2025

Citation: Han, Y.; An, A.; Li, W.; Dong, H. A New Incipient Faults Diagnosis Method Combining SAE and AdaBoost Algorithm for Vehicle Power Supply with Imbalanced Datasets. *Processes* **2025**, *13*, 3343. <https://doi.org/10.3390/pr13103343>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The vehicle power supply has the characteristics of fast movement, lower noise operation, simple operation, all-weather work, etc. It has been widely used in army camping lighting, command, control, communication, artillery system, missile system, air force airport, etc. [1]. However, the working environment of the vehicle power supply is complex and changeable, resulting in many kinds of faults and high randomness. During operation, when incipient faults evolve into serious faults, the entire weapon system may be paralyzed and affect the combat effectiveness of the military. In severe cases, it may even cause serious accidents that endanger the safety of life and property. Therefore, if the early signs of common system faults can be found in time, the types of incipient faults can be accurately identified, and the defensive maintenance can be carried out in advance,

the occurrence of serious faults can be reduced or avoided, and the safety level of vehicle power supply can be improved.

Incipient faults are characterized by subtle manifestations, susceptibility to being masked by noise and disturbance, and difficulty in accurate differentiation between one another. Therefore, more powerful fault diagnosis methods are urgently needed. Vehicle power supply is itself a complex electromechanical system. Because it is difficult to establish a precise mathematical model, and prior knowledge is limited, traditional diagnosis methods based on analytical models and expert systems are not suitable for the incipient fault diagnostics of vehicle power supply [2]. With the rapid development of sensors and information technology, complex systems can obtain massive, numerous sources, and high-dimensional monitoring data [3]. Therefore, data-driven intelligent fault diagnosis methods have become an important way to solve the incipient fault diagnosis of complex systems. As a typical representative of data-driven models, deep learning (DL) can effectively extract high-level features of data from monitoring data and express original data features more completely due to its multi-hidden layer structure [4], opening up a new way for the research of incipient fault diagnosis. Li [5] proposed a fault diagnosis method that integrates feature selection with a deep learning SAE network, which achieved satisfactory results in diagnosing incipient faults in vehicle power supply. Husari and Seshadrinath [6] used the HCNN deep network to extract fault features, combined with the two-level SVM to classify the interturn faults of electric drive system, and evaluated the severity of faults. Peng [7] proposed an Adversarial Domain Adaptation Network with MixMatch (ADANM) for diagnosing incipient faults in permanent-magnet synchronous motors under multiple working conditions. Article [8] proposes a novel deep learning framework that leverages multi-sensor data to diagnose seven types of incipient interturn short-circuit faults in wind turbine generators.

The literature mentioned above applies the deep network to the incipient fault diagnosis of the balanced datasets, which has achieved a good diagnostic effect. However, a single deep network still has the disadvantage of poor generalization performance in an imbalanced dataset, resulting in a low fault recognition accuracy rate [9]. At present, the problem with the processing of imbalanced samples mainly starts from two aspects: the data sampling method and improved model algorithm [10], among which under-sampling, oversampling [11–13], and ensemble learning methods [14,15] are widely used. Under-sampling eliminates sample imbalances by reducing the number of samples in most classes, which may cause data information loss and affect the classification effect. Oversampling expands the number of minority samples, which tends to strengthen some sample features, causing the classifier to overfit. In 1995, Freund and Schapire [16] pioneered the AdaBoost algorithm based on the classic boosting ensemble algorithm, which forced weak classifiers to focus on difficult samples, and weighted and combined multiple weak classifiers into a strong classifier. AdaBoost algorithm can significantly improve the model learning accuracy and generalization ability, so that it has certain advantages in the classification of imbalanced sample datasets.

Considering the characteristics of ensemble algorithms and deep learning, this paper proposes an ensemble learning method based on AdaBoost-enhanced SAE network. This approach integrates the sequential weighting mechanism of AdaBoost with the powerful nonlinear feature extraction capacity of SAE, leveraging their complementary advantages to enhance diagnostic performance under imbalanced data conditions. Through multiple rounds of iterative weighting, it actively focuses on a few difficult-to-identify fault samples, effectively alleviating classification bias caused by data imbalance. Building upon the retained advantage of the original deep network in extracting subtle fault features, the proposed model significantly enhances its capability to handle imbalanced data distributions.

It is expected to achieve superior performance in diagnosing incipient faults of vehicle power supplies under imbalanced sample conditions.

2. Analysis of Incipient Faults in Vehicle Power Supply

Vehicle power supply is a typical complex electromechanical system. An analysis of its primary structure (see the schematic diagram in Figure 1) indicates that the key component determining its performance metrics and power quality is its diesel generator set. This core unit is primarily composed of a diesel engine with its electronic governor system and a synchronous generator with its excitation control system.

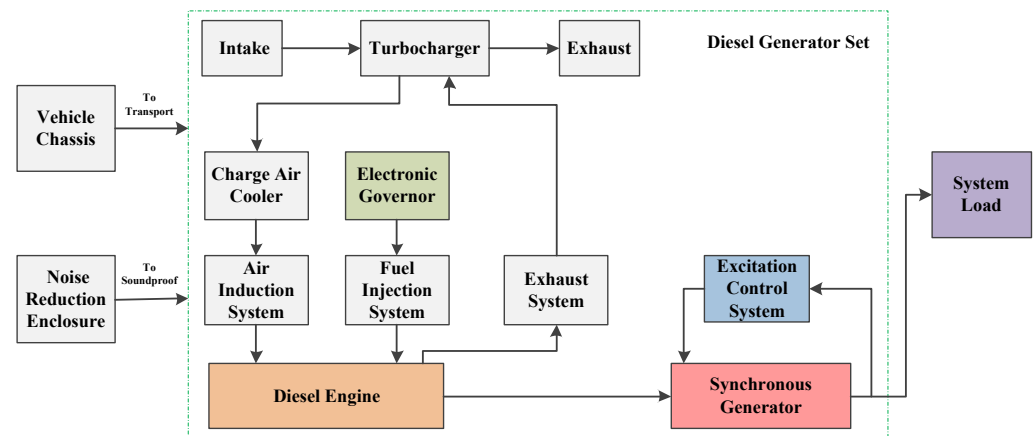


Figure 1. Schematic diagram of the structure of the vehicle power supply.

Due to the strong coupling between each module, the fault mechanism is relatively complex, and the fault types include both electrical and mechanical faults. In the long-term operation process of vehicle power supply, owing to factors such as load changes, component degradation, wear, and harsh environments, the system will gradually experience some failures, such as loss excitation of synchronous generators, unbalanced three-phase voltages of generators, interturn short-circuits, blockage of diesel fuel injection nozzles, electronic governor failure, etc. Once the above faults occur, the vehicle power supply will not work normally. However, the incipient fault of the vehicle power supply has the characteristics of non-obvious measurement signal symptom and loud noise. When an incipient fault occurs, the vehicle power supply can continue to operate in a particular state. But over time, the incipient fault will gradually evolve into a serious fault, which is essentially the process of system parameters from “quantitative change” to “qualitative change”, the mechanism of the failure is the same. Therefore, this study focuses on the early stages of the above common failures, which can be defined according to the four types of power station indicators in “Military Alternating Current Mobile Electric Power Plant, General Specification For” [17].

Failure mechanism and characteristic data are the basis for developing incipient fault diagnosis research. Considering the lack of previous historical monitoring data for vehicle power supply and the high cost of destructive tests, the virtual simulation of complex systems is not only feasible but also a future development trend. Therefore, relying on the team’s development of “Vehicle Power Supply Simulation System” for military users, based on the analysis of the vehicle power supply failure mechanism, by connecting auxiliary components or adjusting related module parameters on this Simulation System, different fault conditions were simulated and monitoring data were collected, and then, according to the main indicators of electrical performance of power stations in GJB 235A-97, combined with the experience of industry experts and historical data, four categories of incipient faults of vehicle power supply were finally determined. (This simulation system strictly follows

the “Technical Agreement for Vehicle Power Supply Model Simulation Software” and is designed specifically for high-fidelity modeling and simulation of 75 kW/120 kW low-noise vehicle power supply. As a core tool, the system has been widely used and tested by the military for many years, and its reliability, accuracy, and effectiveness have been highly recognized and fully verified by users. Figure 2 presents a comparative analysis of steady-state performance indicators between simulation experiments and physical experiments for a 120 kW vehicle power supply. Table 1 displays four categories of incipient faults in the vehicle power supply system, describing their characteristic manifestations and highlighting the associated risk consequences.

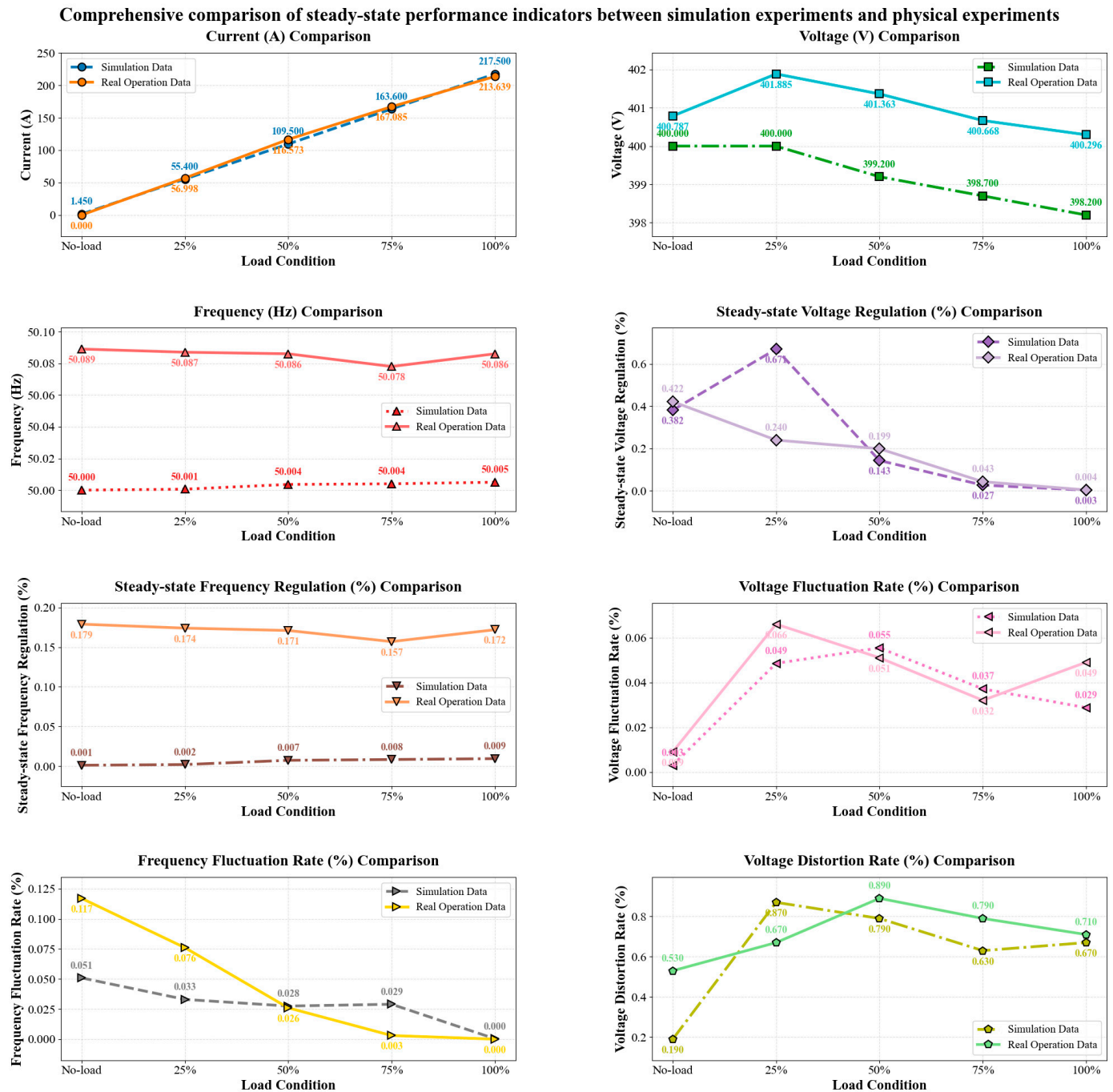


Figure 2. Comparison of steady-state performance indicators between simulation experiments and physical experiments.

Table 1. Description of incipient fault types and characteristics in vehicle power supply.

No.	Incipient Fault State	Risk of Fault
1	Partial loss of excitation in synchronous generator	The performance of the synchronous generator decreases and the load capacity decreases. Sudden downtime in severe cases [18–21].
2	Unbalance degree 1%~3% of three-phase voltage in synchronous generator	The insulation of the synchronous generator is overheated, which damages the insulation life. In severe cases, the insulation will be damaged, which will cause serious short-circuit faults [22–24].
3	Slightly blockage of diesel fuel injector	The power of the diesel motor decreases and the speed decreases. In severe cases, the diesel motor is weak in acceleration and difficult to start [25–28].
4	Electronic Governor degeneration	The speed of the diesel motor is higher than the standard speed and the speed is unstable. In severe cases, the speed of the diesel motor is out of control [29–31].

3. Incipient Faults Diagnosis Model of Vehicle Power Supply

3.1. Proposed Model: AdaBoost-SAE

Through the deep network structure built by deep learning, high-order information such as more abstract and detailed information can be mined from the input data, so it has the ability to automatically learn the essential features of data from the original samples, and these high-order features are undoubtedly the key factors to improve the performance of incipient fault diagnosis [32]. As a classic network model in the field of deep learning, SAE extracts high-order features layer-by-layer by “copying” the input data and realizing data dimensionality reduction in the process, converting complex input data into more representative features. Therefore, it is feasible to select SAE to extract incipient fault features of vehicle power supply and add a Softmax classifier to diagnose common incipient faults of vehicle power supply.

However, the actual operation of the vehicle power supply is mostly under normal working conditions, and the fault occurs only occasionally as a special working condition. Therefore, the number of samples under normal working conditions is much higher than that under fault conditions, resulting in a severe imbalance in the sample datasets. This imbalance will cause the information of the majority samples to cover up the information of the minority samples and then make the network parameter update more inclined to the majority samples when it is used for deep network training. Especially for the incipient faults whose symptoms are not obvious and lack data, the ability of deep network to extract its features is restricted, and the fault diagnosis effect is further affected.

AdaBoost algorithm, as a classic representative of ensemble algorithm boosting, combines multiple homogeneous weak classifiers into strong classifiers. Meanwhile, in the turn-by-turn learning of the weak classifiers, according to the misclassification of the training samples. In the subsequent learning and training of the weak classifier, it can automatically increase the weight of misclassified samples. The ability of the weak classifier to identify misclassified samples is gradually enhanced. It is precisely this mechanism that enables AdaBoost ensemble algorithm to have the ability to adaptively adjust the error rate of weak classifiers. It not only focuses the focus of weak classifiers on difficult samples, but also improves the generalization ability of weak classifiers, and the ensemble combination of multiple weak classifiers will show better classification performance than a single classifier.

To sum up, in order to obtain the deep, intrinsic characteristics of incipient faults of vehicle power supply and improve the identification ability of incipient faults in deep networks under imbalanced samples, this paper intends to combine multiple weak classifiers SAE into a strong classifier through the AdaBoost ensemble algorithm. A fault diagnosis

method for vehicle power supply based on the AdaBoost-SAE ensemble deep learning model is proposed to address the challenges of difficult extraction of incipient fault features and imbalanced samples distribution.

3.2. Model Description

For incipient fault diagnosis of vehicle power supply under imbalanced data conditions, the overall framework of the AdaBoost-SAE-based model is illustrated in Figure 3. The construction process of the diagnostic model includes three major phases: data acquisition and preprocessing, construction of the AdaBoost-SAE diagnosis model, and model validation.

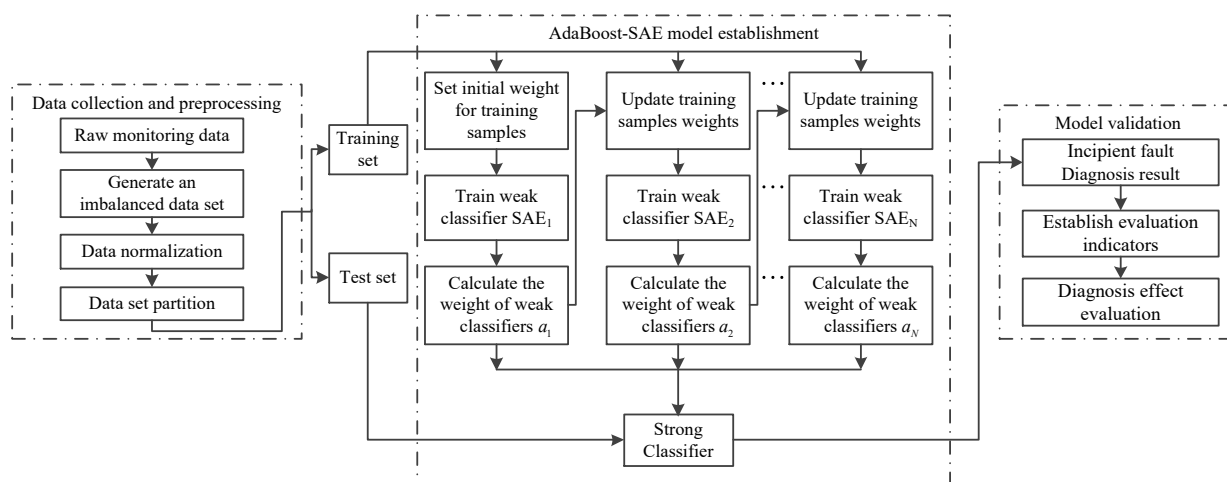


Figure 3. Incipient fault diagnosis model of vehicle power supply based on AdaBoost-SAE.

3.2.1. Data Collection and Preprocessing

The corresponding incipient faults in Table 1 were simulated through the vehicle power simulation platform developed by the team in advance, and the corresponding monitoring data was collected, and established a feature vector set of incipient faults of the vehicle power supply, which is composed of 15 kinds of relevant variables in the vehicle power supply monitoring data, specifically including active power (P), reactive power (Q), power factor ($\cos \varphi$), frequency (f), rotor speed (n), electromagnetic torque (T_e), stator voltage (U_s), stator current (I_s), excitation voltage (E_f), three-phase voltage (U_a , U_b , U_c), three-phase current (I_a , I_b , I_c). The feature vector set of the incipient fault of the vehicle power supply is expressed as $X \in \mathbf{R}^{N \times J}$, incipient fault category labels use One-Hot Encoding, and the dataset is represented as $y \in \mathbf{R}$.

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1J} \\ x_{21} & x_{22} & \dots & x_{2J} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \dots & x_{NJ} \end{bmatrix} \quad (1)$$

$$y = \{y_l = \text{sign}(k), k = 1, 2, \dots, 5\}_{l=1}^N \quad (2)$$

where N indicates the number of samples in the feature set and J indicates the number of variables in the feature set, namely $J = 15$; k is the number of fault categories, total from Table 1 common incipient fault types plus normal status $k = 5$.

According to the ratio of 10:1 between the normal working conditions and fault conditions, imbalanced datasets were generated from the feature vector set. The datasets were normalized and preprocessed to eliminate the dimensional influence of different features.

3.2.2. The Principle of the Proposed AdaBoost-SAE Algorithm

The AdaBoost-SAE model obtains a better and more comprehensive supervised learning model by combining multiple weak classifiers (SAE). The specific steps are as follows:

Step 1. Initialize the weight vector of the training samples in the imbalanced datasets of the vehicle power supply:

$$D_1 = (w_{11}, w_{12}, w_{13} \dots w_{1n}), w_{1i} = \frac{1}{n}, i = 1, 2, 3 \dots n \quad (3)$$

n denotes the number of training samples, w_{1i} is the weight of the i^{th} sample in the first iteration, and D_1 denotes the initial weight distribution of the samples.

Step 2. Training weak classifier SAE: one by one for each iteration update training $m = 1: M$ (M says the number of weak classifier), algorithm loop performs (a)–(d).

- (a) Assign the corresponding weights w_{mi} to all training samples, train the m^{th} weak classifier under this sample distribution and obtain the error rate of the weak classifier during this round of iterative training. Its value is the sum of the weights of misclassified samples.

$$e_m = \sum_{i=1}^N w_{mi} I(\text{SAE}_m(x_i) \neq y_i) \quad (4)$$

where w_{mi} is the weight of the i^{th} training sample in the SAE_m ; $\text{SAE}_m(x_i)$ is the prediction result of the m^{th} weak classifier, $x_i \in X$, and y_i is the true label of the incipient faults, $y_i \in Y$. $I(case)$ denotes when the $case$ is true, $I(case) = 1$, otherwise, its value is 0.

- (b) Calculate the weight of the m^{th} SAE in this round:

$$a_m = \frac{1}{2} \log \frac{1 - e_m}{e_m} + \log k \quad (5)$$

Among them, k is described in Formula (2).

- (c) Update the weight distribution of the training dataset samples:

$$D_{m+1} = (w_{m+1,1}, w_{m+1,2}, w_{m+1,3}, \dots w_{m+1,n}) \quad (6)$$

$$w_{m+1,i} = \frac{w_{mi}}{Z_m} \exp(-a_m y_i \text{SAE}_m(x_i)), i = 1, 2, \dots, n \quad (7)$$

In Formula (7), Z_m is a normalization factor, and its function is to make the sum of the training sample weights 1, which is defined as follows:

$$Z_m = \sum_{i=1}^n w_{mi} \exp(-a_m y_i \text{SAE}_m(x_i)) \quad (8)$$

- (d) Perform linear weighted fusion of all SAE according to Formula (9) to obtain the final strong classifier:

$$Y_M(x) = \sum_{m=1}^M a_m \text{SAE}_m(x) \quad (9)$$

Step 3. Model validation:

The test samples from imbalanced datasets were input into the strong classifier, and finally the class of incipient faults of the vehicle power supply is identified. The diagnostic effect of AdaBoost-SAE model was verified by different evaluation indexes.

3.3. Training of the Weak Classifier (SAE)

The weak classifier SAE is stacked with multiple Auto-Encoders (AE), which is used to extract high-order features of vehicle power supply incipient fault, and input the extracted feature vectors into the Softmax layer to predict the fault categories. In each iteration training process of AdaBoost-SAE model, SAE deep network performs parameter training according to updated sample weight distribution, which mainly includes unsupervised training and supervised fine-tuning process.

Perform layer-by-layer greedy training on the SAE, randomly initialize the initial weights and biases of the first layer of AE, train the network through the gradient descent algorithm to minimize the reconstruction error, and optimize the network parameters. Only the coding part was retained after the first AE network was trained. Subsequently, the high-order features extracted by the first AE are used as the input of the second AE, and the second network is trained. Analogously, after completing the training of the entire network, the output of the high-order features by the last layer of the network is the input of the Softmax classification layer.

After the forward unsupervised training, the weight and bias of each layer are taken as the initial weight and bias of the SAE, the error between the actual value of the sample label and the predicted value output by the Softmax layer is calculated by the backpropagation algorithm, and the weight and bias of each layer are fine-tuned using the Gradient Descent Method. Thus far, after fine-tuning, all parameters of SAE network are in a relatively optimal position, and the network training is completed. In the deep learning ensemble model, the structure of the weak classifier SAE network is shown in Figure 4. It employs a combination of unsupervised pre-training (forward) and supervised fine-tuning (backward) for feature learning.

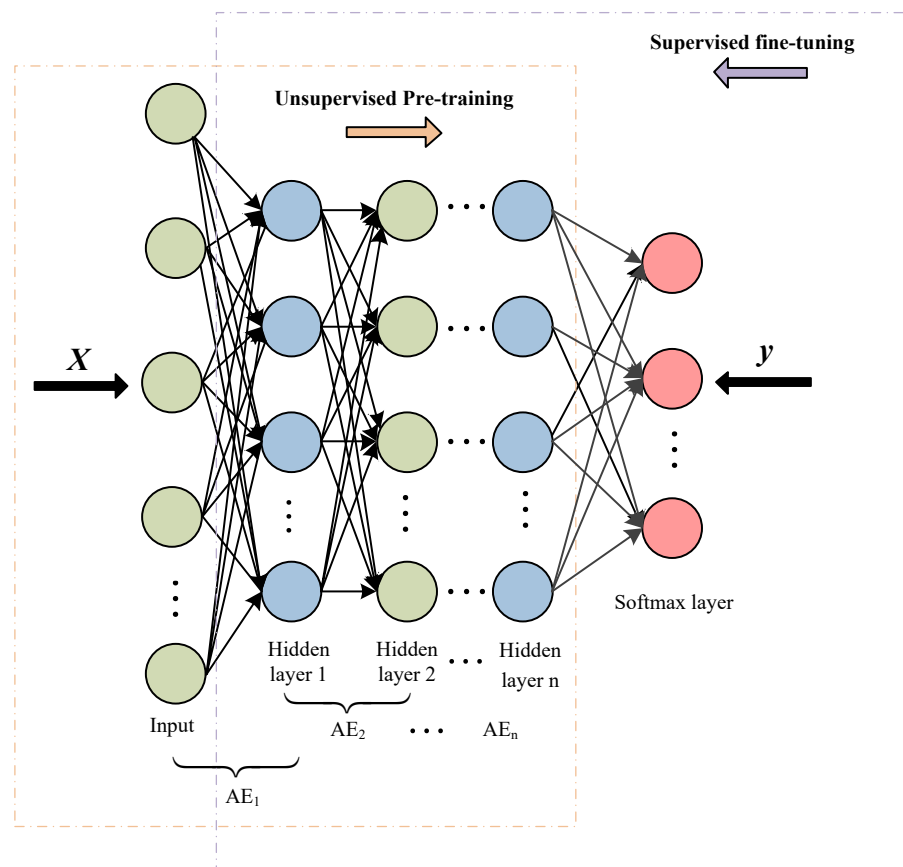


Figure 4. Structure and training process of the SAE model.

4. Experimental Results and Analysis

4.1. Imbalanced Datasets Description and Evaluation Indicators

As described in Table 1, five working conditions were simulated on the vehicle power supply simulation platform. Namely, normal working conditions, partial loss of excitation of the synchronous generator, unbalanced degree 1~3% of three-phase voltage in the synchronous generator, slight blockage of diesel fuel injector, and electronic governor degeneration. Operating data are collected and an imbalanced dataset is generated. In the dataset, the number of samples under normal working conditions was 5000, and the other four types of incipient fault conditions had 500 samples for each type. That is, the majority samples and minority samples had a 10:1:1:1:1 distribution. Table 2 describes the sample quantities and proportions under five working conditions of the vehicle power supply, along with their corresponding labels. In the imbalanced datasets, 60% of the samples were randomly selected as the train set (4200 samples), 20% of the samples were used as the validation set (1400 samples), and 20% of the samples were used as the test set. The simulation test hardware environment was a notebook computer with 16G memory and an i7-8750H processor.

Table 2. Description of imbalanced datasets of incipient fault of vehicle power supply.

Vehicle Power Supply Working Conditions	Number of Samples	Proportion/%	Working Condition Label
normal working condition	5000	71.43	1
partial loss of excitation of the synchronous generator	500	7.143	2
three-phase voltage imbalance of the synchronous generator 1~3%	500	7.143	3
slightly blocked fuel injector	500	7.143	4
electronic governor degradation	500	7.143	5

To objectively and quantitatively evaluate the effectiveness of the deep learning ensemble model in the diagnosis of common incipient faults in vehicle power supply, considering that its nature is still a multiclassification problem, the accuracy and F1-score in the classification task evaluation index are used to evaluate the diagnostic performance of the AdaBoost-SAE model. The F1-score is the weighted harmonic average of precision and recall rates. The higher the score, the stronger the fault diagnosis ability of the model. When calculating the evaluation index of the i -th class sample, the i -th class is regarded as the positive class and the other classes are regarded as the negative class. The calculation formula for the evaluation index is as follows:

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (10)$$

$$precision = \frac{TP}{TP + FP} \quad (11)$$

$$recall = \frac{TP}{TP + FN} \quad (12)$$

$$F1 = 2 \frac{precision \cdot recall}{precision + recall} \quad (13)$$

where TP represents a positive sample classified as the positive class by the model, TN represents a negative sample classified as the negative class by the model, and FP represents a positive sample classified as the negative class by the model. FN represents a negative sample classified as the positive class by the model.

4.2. Optimization of the Main Parameters of the AdaBoost-SAE Model

The network structure and parameters of the weak classifier SAE have an essential influence on the feature extraction and classification of incipient faults [33]; and the number of weak classifier SAE integrations (the number of iterations of the deep integration model) also affects the final diagnosis effect of the AdaBoost-SAE deep integration model. Therefore, it is necessary to first analyze the main parameters of the AdaBoost-SAE model and then select the optimal parameter settings to make the AdaBoost-SAE model have the best performance in diagnosing incipient faults under imbalanced datasets.

(1) Determining the network structure of the weak classifier (SAE)

For the analysis of the parameters in the SAE deep network, there is still no practical and effective method; therefore, traversal optimization is adopted. Based on a large amount of simulation experience in the early stage, after careful consideration, a test range was set for the number of SAE hidden layers and the number of hidden layer neurons. The number of hidden layers increased from 2 to 4, and the number of hidden layer neurons ranged from 5 to 12. Sigmoid function is selected as the activation function of the hidden layer, and MSE (mean-square error) function is used as the loss function. The network is trained using Gradient Descent algorithm to minimize the reconstruction error. The number of iterations for each AE was 2000, and the number of iterations for the overall fine-tuning of the network was 5000. The AdaBoost-SAE model cyclic iterative training number was uniformly set to 10, that is, the number of weak classifiers was 10. Meanwhile, to prevent model overfitting, dropout layers and L2 weight decay were introduced in each weak classifier (SAE). At the same time, when the model performance no longer improves in consecutive iterations, training is immediately stopped to limit the number of weak classifiers and prevent performance saturation and computational redundancy. Taking the incipient fault diagnosis accuracy of the vehicle power supply with imbalanced datasets as the evaluation index of the optimization test, the impact of the number of hidden layers of the weak classifier and the number of hidden layer neurons on the fault diagnosis accuracy of the deep integration model is shown in Table 3. At the same time, the accuracy of incipient fault diagnosis of the single SAE under different network structures is given in Table 3.

Table 3. The optimization result of the network structure of the weak classifier.

Number of Hidden Layers	Number of Hidden Layer Neurons	SAE	AdaBoost-SAE(10) ¹
2	9-5	91.0%	91.4%
2	9-7	90.9%	91.6%
2	10-8	91.9%	92.6%
2	10-7	92.0%	94.9%
2	12-5	92.2%	92.4%
3	6-10-7	91.3%	92.1%
3	8-12-7	91.4%	92.5%
3	10-8-7	91.6%	93.6%
3	9-8-5	92.3%	92.9%
3	12-7-6	91.1%	92.4%
4	10-8-6-5	91.6%	92.2%
4	8-10-12-7	90.3%	91.5%
4	12-9-7-5	91.9%	92.9%
4	10-9-7-5	90.4%	92.0%
4	9-7-6-5	90.8%	92.3%

¹ The AdaBoost-SAE(10) model was trained over 10 iterative rounds, meaning the model consists of 10 weak classifiers.

As shown in Table 3, under the premise of limiting the number of iterative trainings, the accuracy of the combination of the two-layer hidden layer (10-7) in the AdaBoost-SAE model is the highest, at 94.9%. In contrast, the fault diagnosis accuracy of several combinations of three and four hidden layers is relatively low. This shows that the SAE hidden layer structure of the weak classifier is a combination of (10-7), which is conducive to the AdaBoost-SAE model to learn the data characteristics of the incipient faults of the vehicle power supply under imbalanced datasets, and the diagnosis effect is better than other structural combinations. Comparing the fault diagnosis results of a single SAE with different network structures, it shows that the diagnosis accuracy can be improved by integrating SAE networks. However, the optimal network structure of the two models is not the same, which also indicates that the network structure of the simple SAE model and the AdaBoost-SAE model is not related.

(2) Determining the number of iterations of the AdaBoost-SAE model

To select a more appropriate number of weak classifiers (SAE) to form the AdaBoost-SAE model, an optimization experiment was performed on the number of weak classifiers. In the experiment, the hidden layer structure of SAE selected the combination of (10-7) in Table 3. The initial number of weak classifiers was 1. Each round of iterative training adds a weak classifier and calculates the accuracy of model diagnosis. The model stopped iterating when the accuracy was stable and no longer increased. Figure 5 illustrates the variation in fault diagnosis accuracy of the AdaBoost-SAE model as a function of the number of weak classifiers.

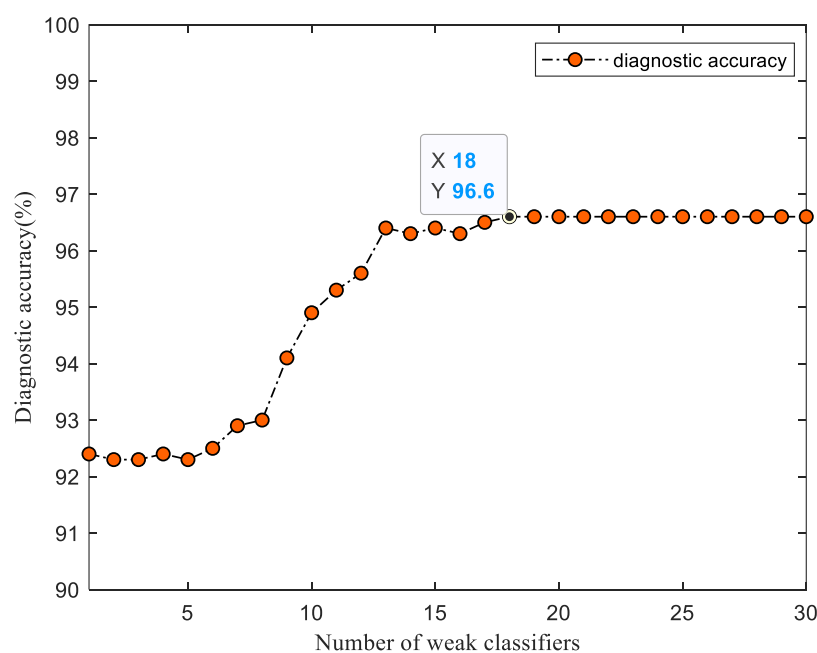


Figure 5. The variation in fault diagnosis accuracy with the number of weak classifiers.

It can be seen intuitively from Figure 5 that the number of weak classifiers (SAE) increases one by one from the initial one, and the overall fault diagnosis accuracy also shows an increasing trend. Finally, when the number of weak classifiers reaches 18, the model fault diagnosis accuracy tends to be stable, reaching 96.6%. After that, when the number of base classifiers is increased, the fault diagnosis accuracy remains unchanged, indicating that AdaBoost-SAE model can achieve better fault diagnosis results when the number of weak classifiers is 18.

Based on the above test results, in the subsequent process of fault diagnosis using the AdaBoost-SAE model, the key parameters of the model are set as follows: the number of weak classifiers is 18, the SAE network structure has four layers, including two AE in the hidden layer, and the number of neurons in the hidden layer is (10-7).

4.3. Analysis of Incipient Fault Diagnosis Results

The confusion matrix can directly and comprehensively reflect the diagnostic accuracy and recall rate of different incipient faults of vehicle power supply, as well as the number of samples misdiagnosed and missed diagnosis of various incipient faults. Therefore, the incipient fault diagnosis results of vehicle power supply based on AdaBoost-SAE model under imbalanced samples were visualized using the confusion matrix, as shown in Figure 6 (including train set, validation set, and test set). Here, only the confusion matrix of the diagnostic results of the test set is analyzed. The green cells along the main diagonal represent the correctly diagnosed samples and their proportions, while the gray cell in the bottom-right corner indicates an overall fault diagnosis accuracy of 96.6%. The first row of the matrix indicates that for the first working condition, 991 samples were correctly classified, while 37 samples were missed, resulting in a recall rate of 96.4% and a miss rate of 3.6%. The first column shows that 11 samples were misclassified as the first working condition, yielding a diagnostic precision of 98.9%.

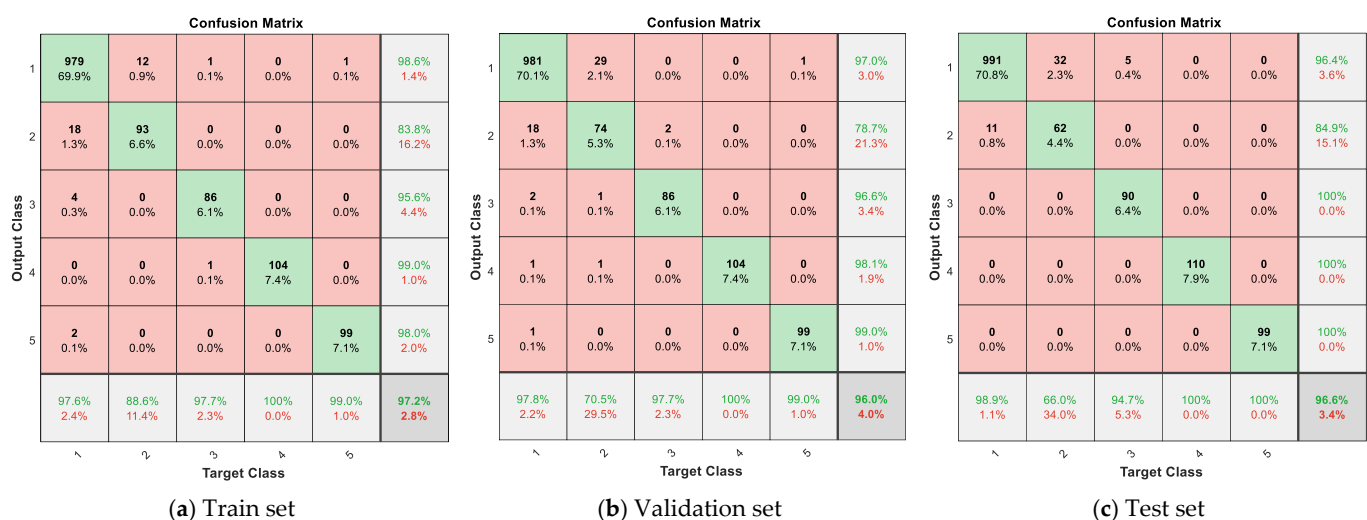


Figure 6. Incipient fault diagnosis results of AdaBoost-SAE model under imbalanced datasets.

To validate the performance advantages of the AdaBoost-SAE model in diagnosing incipient faults of vehicle power supply under imbalanced sample conditions, the following four models were established as reference groups in the simulation experiments: a BP model based on shallow neural network, a traditional machine learning algorithm SVM, an AdaBoost-BP ensemble model (which has achieved optimal accuracy after 20 rounds of iterative training), and a SAE network. Figure 7 comparatively presents the diagnostic result differences between the reference groups and the proposed model through confusion matrices.

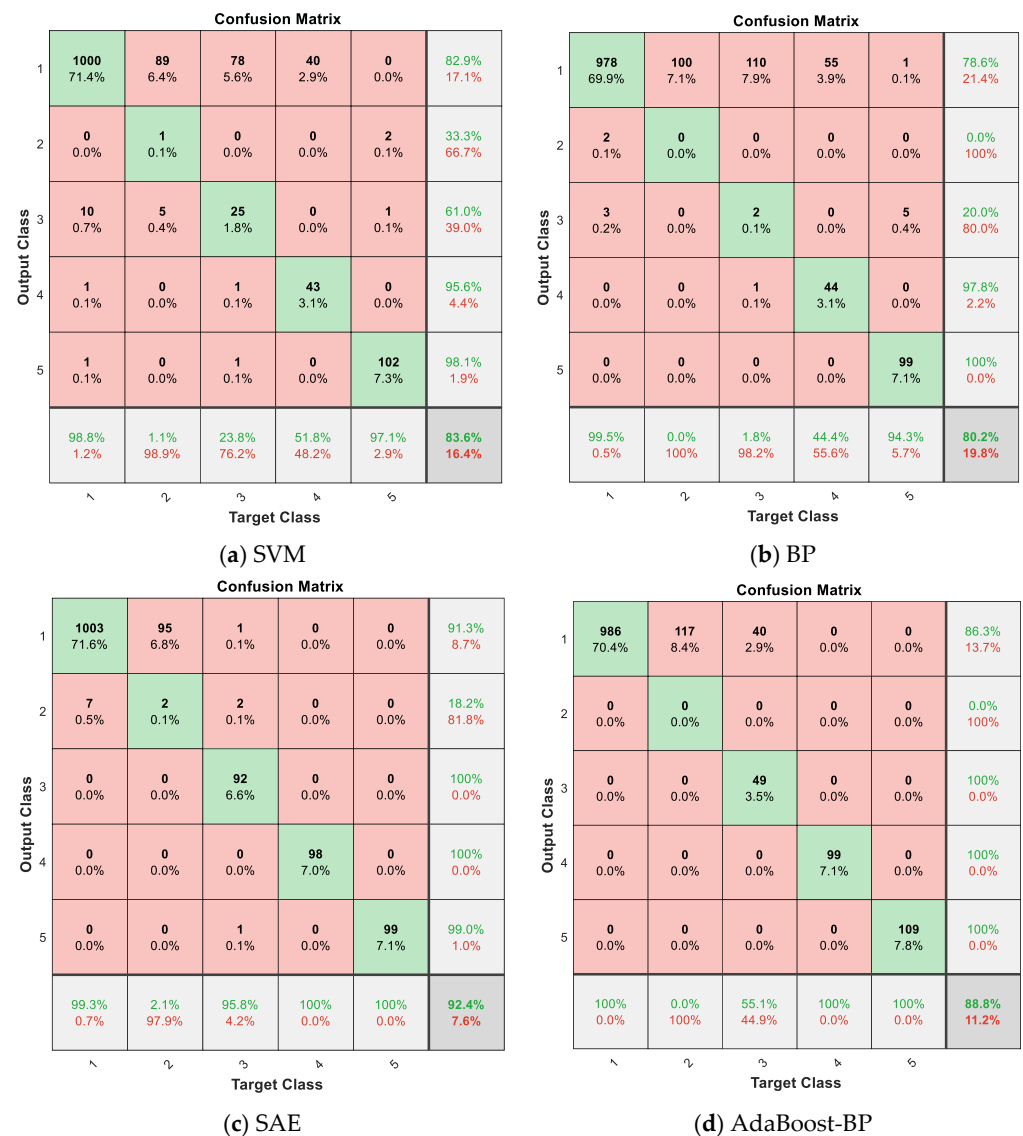


Figure 7. Incipient fault diagnosis results of different models under imbalanced datasets.

Figure 7a–d show the incipient fault diagnosis results of the SVM, BP, SAE, and AdaBoost-BP, the fault diagnosis accuracy under imbalanced datasets is 83.6%, 80.2%, 92.4%, and 88.8%, lower than the 96.6% accuracy of the AdaBoost-SAE model in Figure 6. For the diagnosis results of the four models mentioned above, owing to sample imbalance, even a large number of samples among the minority classes are misclassified (e.g., the 2nd and 3rd working conditions in BP, SVM and AdaBoost-BP and the 2nd working condition in the SAE model), due to the low proportion of the minority class samples in the total samples, the overall diagnostic accuracy of the model still reached a high level. Therefore, using only accuracy to evaluate the effect of fault diagnosis is neither objective nor scientific. Simultaneously, the influence of contingency and random factors in the training process of each model on the diagnosis results was considered to evaluate the fault diagnosis results more convincingly. We performed 10 fault diagnosis tests on each of the models and performed statistics on the diagnosis results. To obtain the average of the accuracy rate, training time, testing time, and F1-score of different models to evaluate the diagnostic performance of the model comprehensively. Table 4 summarizes the fault diagnosis accuracy, model training time, and testing time of the four reference group models, along with the deep learning ensemble models AdaBoost-SAE(5), AdaBoost-SAE(10), and

AdaBoost-SAE. Table 5 presents the F1-score and Macro-F1 of the four reference group models compared with the AdaBoost-SAE model.

Table 4. Comparative analysis of diagnostic accuracy rate, training time, and testing time across different models.

Model	Accuracy/%	Training Time/s	Testing Time/s
SVM	82.5	8.67	0.0079
BP	79.8	7.38	0.0073
SAE	92.3	68.54	0.0081
AdaBoost-BP	88.4	47.75	0.0218
AdaBoost-SAE(5) ¹	92.7	136.79	0.0245
AdaBoost-SAE(10) ²	94.9	375.86	0.0242
AdaBoost-SAE ³	96.6	485.74	0.0239

¹ The AdaBoost-SAE(5) model was trained over five iterative rounds, meaning the model consists of five weak classifiers. ² The AdaBoost-SAE(10) model was trained over 10 iterative rounds, meaning the model consists of 10 weak classifiers. ³ The AdaBoost-SAE was identified as the optimal model in the experiments. The model was trained over 18 iterative rounds, meaning the model consists of 18 weak classifiers.

Table 5. Comparison of F1-score and macro-F1 among different fault diagnosis models.

Model	F1-Score/%					Macro-F1/%
	Working Condition 1	Working Condition 2	Working Condition 3	Working Condition 4	Working Condition 5	
SVM	87.74	0.42	31.68	57.28	80.98	51.62
BP	86.59	0.46	6.93	55.39	79.43	45.76
SAE	95.13	3.77	96.85	99.86	99.45	79.01
AdaBoost-BP	92.45	0	69.57	100	100	72.40
AdaBoost-SAE	97.63	74.27	97.28	100	100	93.84

By comparing the diagnostic effects of the different models in Tables 4 and 5, it can be seen that

- (1) The average diagnostic accuracy of the AdaBoost-SAE model under imbalanced datasets was 96.6%, which is an increase of 4.3% compared with the single SAE and an increase of 7.8% compared with the integration model AdaBoost-BP. The SVM and BP models had poor fault diagnosis effects, with the average diagnosis accuracies of 82.5% and 79.8%.
- (2) Among the deep learning ensemble models, AdaBoost-SAE(5), which incorporates the smallest number of weak classifiers (five), required a training time of 136.79 s. In contrast, the AdaBoost-SAE model, integrated with 18 weak classifiers, exhibited the longest training time, reaching 485.74 s. These results indicate substantial variation in training times across the deep learning ensemble models, with a clear trend toward longer durations as the number of weak classifiers increases. Compared to the control models (SVM, BP, SAE, and AdaBoost-BP), the deep learning ensemble models demand considerably more training time due to their higher structural complexity. Nevertheless, the testing time of the AdaBoost-SAE model was 0.0239 s—only marginally longer than those of the four control models, with differences remaining within the millisecond range.
- (3) SVM and BP showed poor classification results in 2nd, 3rd, and 4th working conditions. The SAE and AdaBoost-BP models are also difficult to classify into the 2nd working condition. The AdaBoost-SAE model significantly improves the F1-score of each working condition on the premise of improving the overall fault diagnosis accuracy. For example, compared with the four models of SVM, BP, SAE, and AdaBoost-BP, the F1-score of the most difficult to diagnose 2nd working condition increased by 73.85%, 73.81%, 70.5%, and 74.27%, respectively.

- (4) Compared to the reference group models, the proposed AdaBoost-SAE model achieved a Macro-F1 score of 93.84%, significantly outperforming all others and highlighting its superior performance in fault diagnosis on imbalanced datasets.

The above results show that, compared with shallow neural networks and traditional machine learning, deep neural networks have obvious advantages in extracting incipient fault characteristics of vehicle power supply. AdaBoost algorithm can significantly reduce the impact of imbalanced training samples, thereby improving the diagnostic performance of the model. The AdaBoost-SAE model combines the advantages of SAE networks and AdaBoost algorithm. When the symptoms of incipient faults in the vehicle power supply are not obvious, and the data samples are imbalanced, it can effectively extract the features of incipient faults, better recognize some difficult-to-classify samples, and obtain stable diagnosis results for the same datasets, showing strong generalization ability. However, as an ensemble model integrating multiple deep neural networks, AdaBoost-SAE required 18 iterative training rounds to converge, resulting in a time-consuming training process. Nevertheless, given its application mode of “offline training and online testing”, the millisecond-level difference in testing time compared to the reference group models does not constitute a significant impact in practical engineering applications.

5. Conclusions

To enhance the diagnostic performance for incipient faults in vehicle power supplies under imbalanced datasets, this paper proposes a fault diagnosis model based on AdaBoost-SAE deep ensemble learning. The main contribution of this study lies in constructing an ensemble learning framework capable of effectively capturing subtle fault features and addressing class imbalance issues. Experimental results demonstrate that the proposed model significantly improves the overall accuracy of incipient fault diagnosis across multiple working conditions, particularly showing superior performance in identifying challenging samples. This model provides a technically practical solution for the reliable diagnosis of incipient faults in vehicle power systems.

However, this study has several limitations. First, although the AdaBoost-SAE model demonstrates excellent diagnostic performance, its iterative training process is more time-consuming compared to traditional fault diagnosis models. Second, the number of weak classifiers and the network’s hyperparameters still heavily rely on extensive manual experimentation rather than automatic configuration. Finally, due to the lack of validation by real data, the model’s practical applicability remains constrained. To address these limitations, future work will focus on the following aspects: first, optimizing the model architecture and training strategy, for instance, by introducing model pruning or distributed training to significantly reduce training time. Second, investigating adaptive parameter optimization algorithms to minimize the reliance on manual experimentation. Third, exploring transfer learning techniques to enhance the model’s generalization capability.

Author Contributions: Conceptualization, Y.H. and W.L.; funding acquisition, H.D.; methodology, Y.H.; resources, W.L. and H.D.; software, Y.H. and A.A.; validation, Y.H.; formal analysis, Y.H.; data curation, Y.H. and W.L.; writing—original draft preparation, Y.H. and W.L.; writing—review and editing, Y.H. and A.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Gansu Province Major Science and Technology Projects under grant number 25ZDGA001.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data is sourced from the “Vehicle Power Supply Simulation System”. Due to its involvement in the defense industry, the data is subject to strict confidentiality regulations and is not publicly available.

Acknowledgments: We thank the School of Automation and Electrical Engineering, Lanzhou University of Technology, for providing their support during the research work.

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

References

- Ren, X.J. Military Power Station of Three Research Analyses. *Movable Power Stn. Veh.* **2020**, *1*, 42–47.
- Li, J.; Zhou, D.H.; Si, X.S.; Chen, M.Y.; Xu, C.H. Review of incipient fault diagnosis methods. *Control Theory Appl.* **2012**, *29*, 1517–1529. Available online: https://kns.cnki.net/kcms2/article/abstract?v=2t0iREynv6ks2WIpAqmCplfE6awgUlqgAxOJi0SEeFAYdskorl8C-tndRQzFFbA1xfIU7CM85x4dykY1Ej_9C2m6aJVe0qMDi9VMilw7Eq8zmJgzyRhqq8UGZqabCAuSkcNw1R67rfxnB9M7P3RkINlzVSrEDZBiQLEQHs8YZSZp12vvtXLWCg==&uniplatform=NZKPT&language=CHS (accessed on 10 October 2024).
- Liu, B.; Li, C.; Yan, Y.; Zhang, G.Z.; Geng, Q.; Shi, T.N.; Xia, C.L. Review of Fault Diagnosis Techniques for Motor Drive Systems. *Proc. CSEE* **2023**, *43*, 5619–5634. [\[CrossRef\]](#)
- Wen, C.L.; Lv, F.Y.; Bao, Z.J.; Liu, M.Q. A Review of Data Driven-based Incipient Fault Diagnosis. *Acta Autom. Sin.* **2016**, *42*, 1285–1299. [\[CrossRef\]](#)
- Li, W.; Han, Y.L.; Sun, X.J. Incipient Fault Diagnosis Method of Vehicle Power Supply Based on Feature Optimization and Deep Learning. *Acta Armamentarii* **2022**, *43*, 2935–2944.
- Husari, F.; Seshadrinath, J. Incipient Interturn Fault Detection and Severity Evaluation in Electric Drive System Using Hybrid HCNN-SVM Based Model. *IEEE Trans. Ind. Inform.* **2021**, *18*, 1823–1832. [\[CrossRef\]](#)
- Peng, X.; Peng, T.; Yang, C.; Ye, C.; Chen, Z.; Yang, C. Adversarial domain adaptation network with MixMatch for incipient fault diagnosis of PMSM under multiple working conditions. *Knowl.-Based Syst.* **2024**, *284*, 12. [\[CrossRef\]](#)
- Wang, Q.; Cui, S.; Li, E.; Du, J.; Li, N.; Sun, J. Deep Learning-Based Fault Diagnosis via Multisensor-Aware Data for Incipient Inter-Turn Short Circuits (ITSC) in Wind Turbine Generators. *Sensors* **2025**, *25*, 2599. [\[CrossRef\]](#)
- Buda, M.; Atsuto, M.; Mazurowski, M.A. A Systematic Study of the Class Imbalance Problem in Convolutional Neural Networks. *Neural Netw.* **2018**, *106*, 249–259. [\[CrossRef\]](#) [\[PubMed\]](#)
- Xiang, H.X.; Yang, Y. Survey on imbalanced data mining methods. *Comput. Eng. Appl.* **2019**, *55*, 1–16. [\[CrossRef\]](#)
- Gu, Z.J.; Yang, X.Y.; Sui, H.; Zhang, Y. Semi-supervised under-sampling method for anomaly detection of industrial control data with class imbalance and overlap. *Appl. Res. Comput.* **2025**, *42*, 156–164. [\[CrossRef\]](#)
- Zhao, X.; Liya, Y.; Shaobo, L.; Li, C.; Feng, Y. An optimized oversampling-based federated transfer learning approach for rotating machinery cluster fault diagnosis. *J. Comput. Des. Eng.* **2025**, *12*, 154–172. [\[CrossRef\]](#)
- Prasoj, R.A.; Putra, M.A.A.; Apriyani, M.E.; Rahmanto, A.N.; Ghoneim, S.S.; Mahmoud, K.; Lehtonen, M.; Darwish, M.M. Precise transformer fault diagnosis via random forest model enhanced by synthetic minority over-sampling technique. *Electr. Power Syst. Res.* **2023**, *220*, 109361. [\[CrossRef\]](#)
- Usman, M.; Chen, H. EMRIL: Ensemble Method based on Reinforcement Learning for binary classification in imbalanced drifting data streams. *Neurocomputing* **2024**, *605*, 22. [\[CrossRef\]](#)
- Zhong, X.; Wang, N. Ensemble learning method based on CNN for class imbalanced data. *J. Supercomput.* **2024**, *80*, 10090–10121. [\[CrossRef\]](#)
- Freund, Y.; Schapire, R.E. A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. Syst. Sci.* **1997**, *55*, 119–139. [\[CrossRef\]](#)
- GJB 235A-97; Military Alternating Current Mobile Electric Power Plant, General Specification For. The Commission of Science, Technology and Industry for National Defence of the People's Republic of China: Beijing, China, 1997.
- Dai, D.R.; He, Y.L.; Zhang, W.; Wu, K.; Xu, M.X.; Zhang, Y.; Wang, X.L. Research on stator core and winding vibration of permanent magnet synchronous generator with demagnetization. *J. Vib. Eng.* **2025**, *1–10*. [\[CrossRef\]](#)
- Su, G.X. Simulation Analysis on Loss of Excitation Phenomenon for Synchronous Generator. *Electr. Mach. Technol.* **2023**, *03*, 28–30.
- Cui, G.; Xiong, B.; Huang, K.J.; Li, Z.G.; Ruan, L. Spatial Distribution Characteristics and Influencing Factors of Demagnetization of Permanent Magnet Motor for Electric Vehicle. *Trans. China Electrotech. Soc.* **2023**, *38*, 5959–5974. [\[CrossRef\]](#)
- Rasoulpour, M.; Amraee, T.; Sedigh, A.K. A Relay Logic for Total and Partial Loss of Excitation Protection in Synchronous Generators. *IEEE Trans. Power Deliv.* **2020**, *35*, 11. [\[CrossRef\]](#)
- Jin, S.; Liang, S.; Shi, L. Control of BDFG Wind Turbine Based on Virtual Synchronous Generator Under Unbalanced Grid. *Power Syst. Technol.* **2024**, *48*, 2426–2435. [\[CrossRef\]](#)
- Cao, Z.; Cheng, M.; Yan, X.M. Multi-objective Control of Dual-cage-rotor Brushless Doubly Fed Induction Generator Under Unbalanced Grid Voltage. *Proc. CSEE* **2024**, *44*, 293–304. [\[CrossRef\]](#)
- Zhu, W.; Huang, N.X.; Shu, H.Y.; Hu, J.; Tang, M. Dominant feature identification and index construction of three-phase voltage imbalance in substation area. *J. Electr. Power Sci. Technol.* **2025**, *40*, 141–149. [\[CrossRef\]](#)

25. Wang, J.; Huang, Y.; Gao, X.Y.; Wang, T.; Wang, X.; Hui, J. Blockage Location Algorithm of Multi-cylinder Fuel Injectors Based on Stacked Sparse Autoencoder. *Acta Armamentarii* **2024**, *45*, 3706–3717. Available online: <https://link.cnki.net/urlid/11.2176.TJ.20231213.1545.002> (accessed on 18 September 2025).
26. Yu, Y.H.; Jia, H.C.; Hu, J.; Hu, L.; Yang, J. Fault Diagnosis Technology of Diesel Engine Fuel Injector Clogging Based on Acoustic Emission. *Trans. CSICE* **2023**, *41*, 466–472. [[CrossRef](#)]
27. Işıklı, F.; Şentürk, G.; Sürmen, A. Effects of Valve, Armature, and Armature Pin Guidance on Diesel Injector Performance. *Appl. Sci.* **2024**, *14*, 5737. [[CrossRef](#)]
28. Yang, J.T.; Zhang, H.; Zhao, J.H. Influence of injector structure parameters on fuel injection quantity fluctuation under multiple injections. *J. Harbin Eng. Univ.* **2025**, *46*, 276–282. [[CrossRef](#)]
29. Zhang, X.C.; Huang, D.W. Typical Fault Analysis of Electronic Governor for a Certain Type of Ship Power Generation Diesel Engine. *Mech. Electr. Equip.* **2025**, *42*, 7–11. [[CrossRef](#)]
30. Zhang, J.; Wu, S.M.; Yu, Y.H. Analysis of Influence of Key Parameters of Typical Diesel Engine Speed Control System on Speed Control Performance. *Energy Energy Conserv.* **2020**, *09*, 144–145+181. [[CrossRef](#)]
31. Xun, X.H.; Xu, T.; Wang, H.; Song, Z.H. A Fault Case of A V-type Diesel Engine Caused by the Deterioration of the Components of the Governor. *Intern. Combust. Engine Parts* **2021**, *11*, 86–88. [[CrossRef](#)]
32. Wen, C.L.; Lv, F.Y. Review on Deep Learning Based Fault Diagnosis. *J. Electron. Inf. Technol.* **2020**, *42*, 234–248. [[CrossRef](#)]
33. Patel, H.R.; Shah, V.A. Shadowed Type-2 Fuzzy Sets in Dynamic Parameter Adaption in Cuckoo Search and Flower Pollination Algorithms for Optimal Design of Fuzzy Fault-Tolerant Controllers. *Math. Comput. Appl.* **2022**, *27*, 89. [[CrossRef](#)]

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