vehicles

Article

A Hybrid Data-Driven Method for Main Circuit Gound Faults Diagnosis in Electrical Traction Drive Systems

Xinyao Hou [1], Juntong Liu [2], Jinxin Zhang [2] and Qiang Ni [2,*]

[1] School of Locomotives and Rolling Stock, Guangzhou Railway Polytechnic, Guangzhou 511300, China; houxinyao@gtxy.edu.cn
[2] School of Automation, Guangdong University of Technology, Guangzhou 510006, China; 2112304001@mail2.gdut.edu.cn (J.L.); 2112404503@mail2.gdut.edu.cn (J.Z.)
* Correspondence: nq666@gdut.edu.cn

Abstract: A main circuit ground fault (MCGF) is a typical system fault in an electrical traction drive system (ETDS). When two or more MCGFs occur, it will cause serious accidents. Therefore, it is particularly important to detect and handle MCGFs in a timely manner. To improve the efficiency of train operation and ensure driving safety, this paper proposes a hybrid data-driven MCGF diagnosis method. First, the voltage signals related to the fault are selected according to the mechanism analysis of the MCGF, and then the initial feature variables are constructed according to these voltage signals. Secondly, the initial feature variables of different types of MCGF are analyzed in the time and frequency domains by wavelet transform, and four feature indicators are calculated. Finally, the fault feature indicators are trained by random forest to obtain a model for subsequent fault diagnosis. After comparative experiments using various machine learning methods, it was found that the RF used in the proposed method has a better diagnostic effect, and the correct isolation rate exceeds 99%.

Keywords: electrical traction drive system (ETDS); main circuit ground fault (MCGF); fault diagnosis; wavelet transform; random forest (RF)

Citation: Hou, X.; Liu, J.; Zhang, J.; Ni, Q. A Hybrid Data-Driven Method for Main Circuit Gound Faults Diagnosis in Electrical Traction Drive Systems. *Vehicles* **2024**, *6*, 1872–1885. https://doi.org/10.3390/vehicles6040091

Academic Editor: Loránd Szabó

Received: 27 September 2024
Revised: 28 October 2024
Accepted: 30 October 2024
Published: 5 November 2024

1. Introduction

Due to the rapid growth of the modern transportation field, high-speed trains have become an indispensable mode of transportation [1]. As the core of the train, the electrical traction drive system (ETDS) can provide power for the train through the AC-DC-AC circuit [2]. Since the train operates in various complex outdoor environments for a long time, various faults may potentially arise [3], especially the failure of the ETDS, which can significantly affect the train's normal operation and even cause safety accidents [4]. Hence, to guarantee the train's operating efficiency and safety, when a fault occurs in the ETDS, it is very important research content to diagnose the occurrence and type of the fault in time [5].

Among all the fault studies on ETDSs, component-level fault diagnosis results are the most prevalent, including open-circuit faults of the rectifier IGBT, sensor faults, open-circuit faults of the inverter, etc. Ref. [6] proposed a parallel direct fault diagnosis method for open-circuit faults of the IGBT, using an improved Kalman filter (KF) to evaluate the voltage of the capacitor by comparing the measured value and the estimated value. Ref. [7] proposed a comprehensive diagnosis technique for grid-side current sensor failures and DC bus voltage sensor failures based on reduced-order observers. Ref. [8] proposed a coupled inductor-based open-circuit fault diagnosis method for the power switch of the Aalborg inverter, which does not require additional sensors or diagnostic circuits.

However, the component-level fault diagnosis methods mentioned above all focus on a specific module fault. An ETDS includes modules such as the traction transformer, four-quadrant rectifier, inverter, traction motor, etc. [9], as shown in Figure 1, which is an electro-mechanical system with a complex coupling relationship between multiple modules.

Therefore, a fault diagnosis method that begins with a particular module does not fully consider the coupling interactions of an ETDS, which will lead to some faults, such as system-level faults, not being accurately diagnosed and protected. Throughout the train's actual operation, the traction control unit (TCU) will continue to collect various signals from the ETDS [10], but in the traditional diagnosis method, the TCU only uses one signal to see whether it exceeds the fault warning threshold and does not make full use of other signals [11], which leads to inaccurate diagnosis of the fault type. Therefore, it is an urgent problem to study the system-level fault diagnosis methods of ETDSs by combining the multiple signals collected by the TCU, which can accurately diagnose the fault type and take differentiated protection measures.

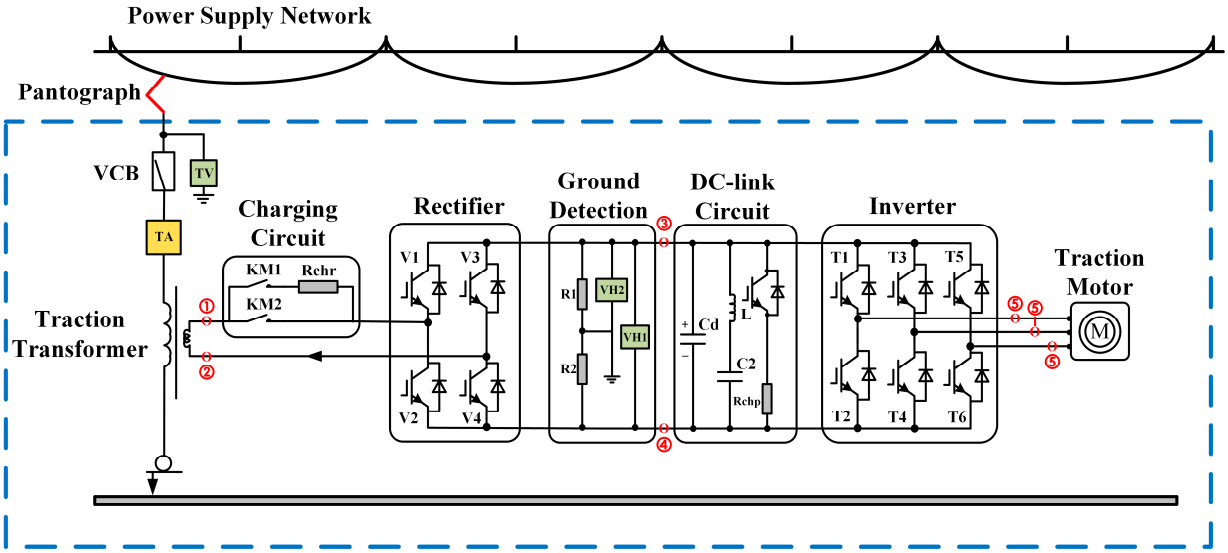

Figure 1. Schematic illustration of the main circuit of an ETDS.

A main circuit ground fault (MCGF) is a classic system-level fault in ETDSs, which can be divided into single-point ground faults and multi-point ground faults. Generally speaking, a single-point ground fault will not have a significant effect on the normal operation of the system, but once a multi-point ground fault arises in the system, it will cause a large overcurrent in the main circuit, which will have a destructive effect on the entire system [12]. To avoid this situation, it is essential to explore the ground fault diagnosis method, so that when a single-point ground fault arises, it can be diagnosed, and protection measures can be taken to prevent the occurrence of multi-point ground faults.

In the existing train main circuit ground fault diagnosis methods, the traditional method is to only detect the grounding voltage and use a simple upper- and lower-limit alarm method to achieve fault diagnosis, but this method cannot distinguish ground faults on the rectifier side from those of the inverter side well, and staff are required to check one by one. This not only increases the parking time but also affects the maintenance efficiency. Ref. [12] proposed a ground fault diagnosis method based on typical correlation analysis, but when the measured signal is mostly an AC signal containing high-frequency noise, its diagnostic performance needs to be improved. Therefore, the limitations of existing methods require us to study more comprehensive and effective diagnostic methods.

Current fault diagnosis research can generally be categorized into three methods: model-based, data-driven [13], and signal analysis-based [14]. Ref. [15] developed a model of the pulse rectifier grounded in the structural analysis of the traction system and proposed a method to diagnose the faults of the pulse rectifier sensor and IGBT. Ref. [16] established a predictive current model and a hybrid logic dynamic (MLD) model, which were hybridized to diagnose the open-circuit switch failure of a permanent magnet synchronous motor drive. The efficacy of the model-based diagnostic method depends largely on the accuracy of the

established mathematical model. However, as mentioned earlier, the ETDS is a complex coupled system. The complex coupling relationship and changes in operating conditions make it very difficult to accurately model ETDSs. For data-driven methods, there is no need to accurately model the system. Ref. [17] developed a new data-driven sensor fault detection and classification method founded on improved slow feature analysis (SFA). Ref. [18] proposed a data-driven high-speed train running-gear fault diagnosis method, which added a belief rule base in view of deep slow feature analysis. In contrast to the traditional method, the approach suggested by [18] is more effective in lowering the likelihood of fault alarms. It can be seen from [17,18] that data-driven diagnosis methods map relationships and acquire knowledge from extensive historical data, which are the effective solution for ETDS fault diagnosis. However, in engineering practice, ground faults do not occur often, and because the damage caused by a multi-point ground fault is very large, the TCU will also take protection measures very quickly, resulting in a very short time to collect fault signals when a fault occurs. Hence, it is very difficult to attain a large quantity of ground fault historical data.

In the fault detection and diagnosis (FDD) method grounded in signal analysis, the measured signal already contains rich fault information, which can be derived from time, frequency, and time–frequency domain joint analysis to identify different fault types [19]. Among the many time–frequency domain analysis methods, the traditional fast Fourier transform (FFT) can evaluate the frequency characteristics of the signal but cannot provide information about how these components change over time. The short-time Fourier transform (STFT) uses a fixed window, but the choice of window size affects the accuracy of frequency and time [20]. To resolve this challenge, the wavelet transform changes the window size according to the signal frequency, shifts the wavelet basis on the time axis, and performs convolution operations so as to better balance the accurate analysis of time and frequency. However, because of the complex and variable operating conditions of high-speed trains, the application of FDD methods based on signal analysis on trains is limited. Although the above three methods are not suitable for the diagnosis of MCGFs alone, there are still many things that can be learned from signal analysis-based and data-driven methods.

Driven by the lack of existing research, this article introduces a hybrid data-driven diagnostic method for MCGFs in ETDSs. This approach leverages the benefits of signal analysis, data-driven techniques, and the mechanisms of MCGFs. The key innovations and contributions of this article can be summarized as follows:

- Based on the mechanism analysis of MCGFs, initial feature variables are constructed by obtaining data of three voltage signals.
- Wavelet transform is used to analyze different MCGF types in the time and frequency domains, and the corresponding feature indicators are calculated.
- A diagnosis framework for MCGFs is proposed, incorporating fault detection and identification using random forest (RF), with its effectiveness validated through field experiments.

The structure of this article is organized as follows: Section 2 examines the mechanisms of MCGFs and establishes the initial feature variables. Section 3 introduces the diagnostic method and framework based on wavelet transform and RF. Section 4 outlines the experimental process and compares the results obtained. Finally, Section 5 concludes this article.

2. Analysis of MCGF Mechanism and Feature Variables

2.1. Analysis of MCGF Mechanism

There are five typical MCGF types in ETDSs, which are as follows: positive side of the rectifier to ground, negative side of the rectifier to ground, positive side of the DC-link circuit to ground, negative side of the DC-link circuit to ground, and inverter side to ground, as shown in Table 1.

Table 1. Typical MCGF types in ETDS.

Fault Type	Fault Location	Ground Description
C_1	①	Positive side of the rectifier
C_2	②	Negative side of the rectifier
C_3	③	Positive side of the DC-links
C_4	④	Negative side of the DC-links
C_5	⑤	Inverter side

The ground fault detection circuit shown in Figure 1 is composed of two large resistors R_1 and R_2 and two voltage sensors VH_1 and VH_2. The voltage sensor VH_1 measures the DC voltage U_D, and VH_2 is used to collect the voltage of ground detection circuit U_{GD}. The values of R_1 and R_2 are identical, so the connection between U_D and U_{GD} in normal operation condition can be described as

$$U_{GD} = \frac{1}{2} \cdot U_D \tag{1}$$

When an MCGF occurs, the values of U_D and U_{GD} will change, and the relationship between them will change accordingly. Therefore, the discussion of their relationship will be shown as follows.

- MCGF C_1.

When the positive side of the rectifier is grounded, the magnitude of U_{GD} changes from 0 to U_D, and the change frequency is the same as the frequency of the rectifier. The relationship between U_D and U_{GD} is as follows:

$$U_{GD} = (1 - V_A) \cdot U_D \tag{2}$$

where V_A indicates the switching condition of the rectifier. When V_1 and V_4 are turned on, $V_A = 1$; when V_2 and V_3 are turned on, $V_A = 0$, as shown in Figure 1.

- MCGF C_2.

When the negative side of the rectifier is grounded, the magnitude of U_{GD} changes from 0 to U_D, and the change frequency is the same as the frequency of the rectifier. The relationship between U_D and U_{GD} is as follows:

$$U_{GD} = (1 - V_B) \cdot U_D \tag{3}$$

where V_B indicates the switching condition of the rectifier. When V_2 and V_3 are turned on, $V_B = 1$; when V_1 and V_4 are turned on, $V_B = 0$.

- MCGF C_3 and C_4.

When the DC-link circuit is grounded, the connection between U_D and U_{GD} is determined by the type of MCGF, and it can be represented as

$$U_{GD} = \begin{cases} 0, & positive\ grounding \\ U_D, & negative\ grounding \end{cases} \tag{4}$$

- MCGF C_5.

When the inverter side is grounded, the magnitude of U_{GD} changes from 0 to U_D, and the change frequency is the same as the frequency of the inverter. The inverter output side is three-phase symmetrical, so we only discuss one phase. The relationship between U_D and U_{GD} is as follows:

$$U_{GD} = (1 - V_U) \cdot U_D \tag{5}$$

where V_U indicates the switching condition of the inverter. When T_1, T_3, or T_5 is turned on, $V_U = 1$; when T_2, T_4, or T_6 is turned on, $V_U = 0$.

2.2. MCGF Feature Variables Extraction

In this section, the original data are first normalized to be within the range of 0 to 1. According to the above discussion, the DC-link voltage U_D can become an important evaluation criterion for ground faults. Furthermore, the ground detection voltage U_{GD} has a clear mathematical relationship with U_D, so an initial characteristic variable z_1 can be constructed through these two voltages, which can be represented as follows:

$$z_1 = \overline{U}_{GD} - \frac{1}{2} \cdot \overline{U}_D \tag{6}$$

It is clear from the z_1 of different grounding faults in Figure 2 that under normal operating conditions, the value of z_1 is almost 0. From Table 2, the value of z_1 can well distinguish the situation of DC-link grounding, but there is no way to distinguish the other three types of MCGF because their z_1 all vary between $-0.5\ U_D$ and $0.5\ U_D$.

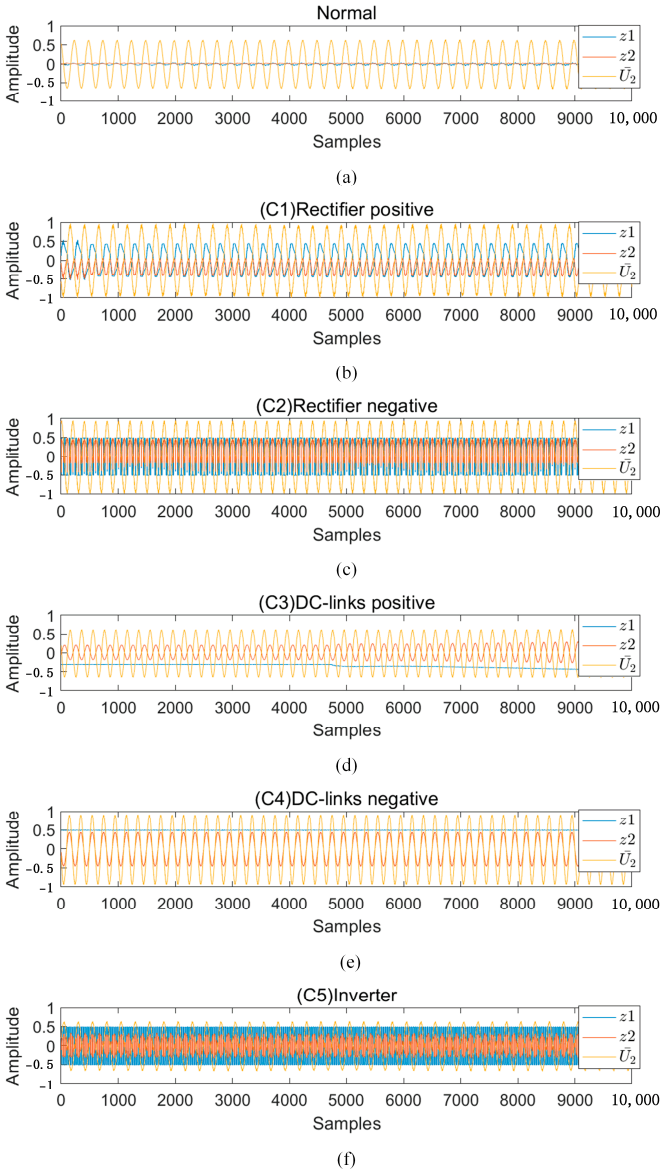

Figure 2. Features of each MCGF type. (**a**) Features of normal conditions. (**b**) Features of C_1. (**c**) Features of C_2. (**d**) Features of C_3. (**e**) Features of C_4. (**f**) Features of C_5.

Table 2. Change of z_1 under different ground faults.

Fault Type	Change of z_1
C_1	The value of z_1 will vary between $-0.5\ U_D$ and $0.5\ U_D$, and the frequency of change is consistent with the rectifier switching frequency.
C_2	
C_3	The value of z_1 will remain at $-0.5\ U_D$.
C_4	The value of z_1 will remain at $0.5\ U_D$.
C_5	The value of z_1 will vary between $-0.5\ U_D$ and $0.5\ U_D$, and the frequency of change is consistent with the inverter switching frequency.

It is also necessary to combine other signals to construct a new feature variable to better distinguish all ground faults. From the previous mechanism analysis, C_1, C_2, and C_5 are closely related to the switching change frequency of their respective converters. Therefore, it also essential for the signals related to the switching state to be combined. The traction transformer secondary voltage U_2 contains the rectifier switching information. Hence, a new feature variable z_2 is constructed to better characterize the ground fault and distinguish the other three types of MCGF. z_2 can be expressed as follows:

$$z_2 = z_1 \cdot \overline{U}_2 \tag{7}$$

In summary, z_2 can distinguish various MCGFs well. Next, it is necessary to analyze z_2 in the time–frequency domain to better represent its characteristics. Then, the corresponding feature indicators are established for subsequent model training.

3. Proposed Method

Due to the interconnected relationships in the ETDS, the entire system is relatively complicated, and the action changes of some modules are likely to cause some signal changes. The train will experience various operating states during operation, such as slowing down and stopping at the station, restarting and accelerating after picking up passengers, etc. The action changes of the corresponding system modules under different operating states are also different, resulting in the voltage signals related to the MCGFs likely being unstable. Therefore, to better obtain the characteristics of MCGFs, this paper uses wavelet transform to obtain time–frequency domain characteristics.

3.1. Wavelet Transform

In the analysis of the frequency domain, fast Fourier transform (FFT) is frequently used. Although it is widely used, this method only reflects the overall frequency characteristics of the signal and cannot be applied to the analysis of unstable signals. For non-stationary signals like grounding signals, there are short-time Fourier transform (STFT), wavelet transform, and other methods. STFT divides the non-stationary global signal into stable local signals; it has certain requirements for the setting of the window size, thus limiting its application. At present, wavelet transform is a better method for processing non-stationary signals. The wavelet transform conducts multi-scale refinement analysis on the signal by utilizing scaling and translation. It serves as a localized transformation in both time and frequency, enabling effective extraction of signal information.

Figure 3 shows a schematic diagram of multi-scale one-dimensional wavelet decomposition, where X is the input one-dimensional signal. The first level decomposition decomposes the X signal of length n into approximate coefficients cA_1 and detail coefficients cD_1 of length $\left(\frac{n-1}{2}\right) + M$. Here, M is half the length of the high-range and low-range filters used in wavelet transformation. cA_1 denotes the low-frequency component of the signal, while cD_1 signifies the high-frequency component of the signal. The multi-scale one-dimensional wavelet decomposition functions are as follows:

$$[C, L] = wavedec(X, Y, 'wname') \tag{8}$$

$$cA_3 = appcoef(C, L, 'wname', 3) \tag{9}$$

$$[cD_1, cD_2, cD_3] = \text{det}coef(C, L, [1, 2, 3]) \tag{10}$$

where X is the original signal, which is z_2 here; Y is the decomposition order, which is set to 3 here; and *wname* is the wavelet type, which is db_2 here.

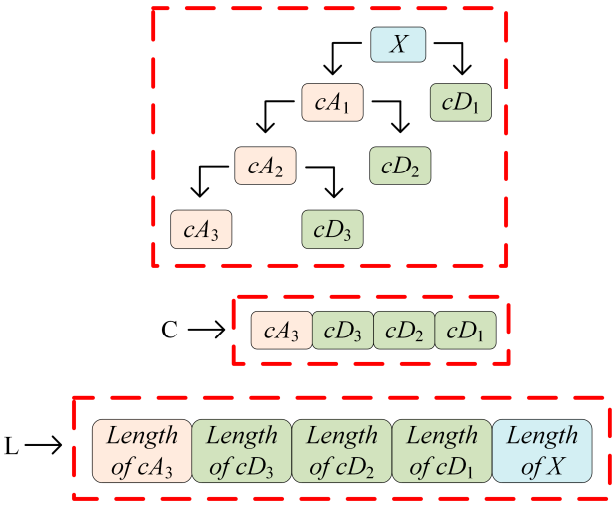

Figure 3. Schematic diagram of multi-scale one-dimensional wavelet decomposition.

After normalizing the original data, this paper uses db_2 as the wavelet basis to perform wavelet 3-level decomposition on the characteristics of different types of grounding faults. The wavelet transform results of z_2 for all types of ground faults are displayed in Figure 4.

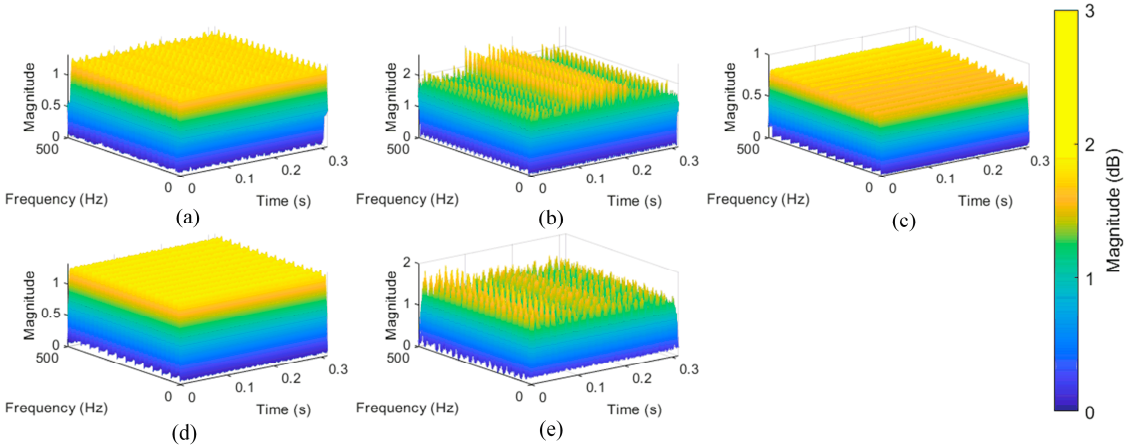

Figure 4. The wavelet transform results of z_2 for each type of MCGF. (**a**) Positive side of the rectifier C_1. (**b**) Negative side of the rectifier C_2. (**c**) Positive side of DC-links C_3. (**d**) Negative side of DC-links C_4. (**e**) Inverter side C_5.

3.2. Construction of Feature Indicators

After the time–frequency domain analysis of wavelet transform, the fault feature indicators can be constructed by the wavelet transform analysis results. As can be seen from Figures 2 and 4, generating feature indicators in both the time–frequency and time areas effectively captures various types of MCGF. Consequently, a second sliding window is employed to extract information from the time area of z_2 and the time–frequency area of the wavelet transform conclusions of z_2, resulting in a time–frequency area matrix (*TFM*).

The following definitions pertain to four feature indicators and the combined feature vector (*MFV*):

$$TFM = cA_3 + cD_1 + cD_2 + cD_3 \tag{11}$$

$$J_1(k) = \frac{1}{N} \sum_{i=k}^{N+k-1} \sum_{j=k}^{N+k-1} TFM_{ij} \tag{12}$$

$$J_2(k) = pk\_pk\{TFM\} \tag{13}$$

$$J_3(k) = \frac{1}{N} \sum_{t=k}^{N+k-1} z_2(t) \tag{14}$$

$$J_4(k) = \frac{1}{N-1} \sum_{t=k}^{N+k-1} [z_2(t) - J_3(k)]^2 \tag{15}$$

$$MFV(k) = [J_1(k)\ J_2(k)\ J_3(k)\ J_4(k)] \tag{16}$$

where N denotes the length of the sliding window, and the value of N in this article is set to 2500; *pk_pk* indicates the peak-to-peak value of the series; and k signifies each point within the feature indicators.

3.3. Random Forest

RF is an outstanding ensemble learning algorithm that combines multiple decision trees for classification [21], which has excellent performance in the presence of noisy and weakly discriminative data. In RF, each dataset is randomly replaced, while some features are randomly chosen as input. Multiple decision trees are then combined to build the entire model. Finally, in the classification problem, the majority of classification results are selected as the final result.

The specific procedures of the RF algorithm are outlined as follows:

- RF uses Bootstrap sampling to randomly extract N_d samples from the training dataset with replacement to generate multiple sub-datasets, where N_d refers to the number of samples in the training dataset. The size of each sub-dataset is the same as the original dataset. The process is expressed in the following formula:

$$D = \{(x_1, y_1), (x_2, y_2), \cdots, (x_{N_d}, y_{N_d})\} \tag{17}$$

$$D' = \{(x_i,\ y_i)\}, i = \{1, 2, \ldots, N_d\} \tag{18}$$

  where D is the training dataset and D' is the new dataset after sampling.
- For each sub-dataset, when building a decision tree, a feature subset is randomly selected from all features, and then a decision tree is trained using each sub-dataset and the corresponding feature subset.
- Multiple decision trees are combined into one model.
- The results of multiple decision trees are integrated for classification, and the final classification result is the category that receives the most votes.

Random forest has excellent capabilities, but it also has certain limitations. First, it performs poorly on unbalanced data and tends to be biased towards the majority class. Second, it takes a long time to train in the case of large datasets. However, in this article, the number of samples for each fault type is 10,000, and there will be no unbalanced data. In addition, as mentioned previously in this article, the probability of ground faults is relatively low, and it is very difficult to obtain a large amount of data. The overall data volume in this article is relatively moderate, and it will not lead to excessive training time while ensuring good results. Furthermore, the model complexity of random forest causes a lack of interpretability and internal operability, which makes it difficult to reflect the influence of a single feature or the interaction between multiple features, but these issues do not affect the final diagnostic results of this article. The topic of feature influence and interaction needs further research.

3.4. MCGF Diagnosis Framework

The entire MCGF diagnosis framework includes an offline part and an online part, as illustrated in Figure 5.

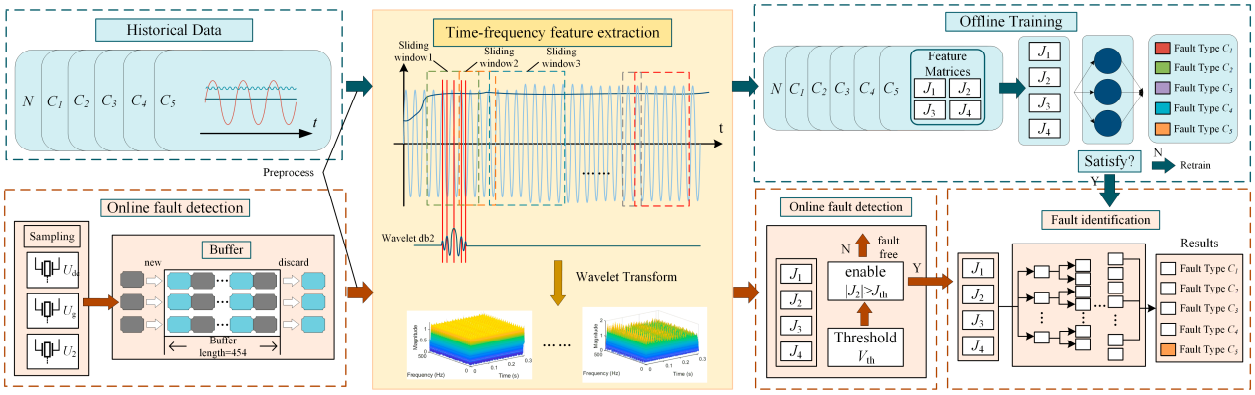

Figure 5. MCGF diagnosis framework of the proposed method.

In the offline phase, historical data undergo preprocessing to derive the initial feature variable z_2 for all MCGFs and normal conditions. Subsequently, the time–frequency amplitude matrices that specifically characterize the MCGFs are developed using wavelet transform. After that, four feature indicators are calculated from the results of the wavelet transform and combined into a feature matrix. Finally, a diagnostic model that can accurately map different types of MCGF is trained based on RF. Specifically, the preprocessing stage involves normalizing the initial data to an interval between 0 and 1 and calculating two initial feature variables z_1 and z_2. In the RF training process, the data in this article come from the actual train model. When a ground fault occurs during operation, the data before and after the failure are recorded through wave recording technology, allowing us to collect multiple fault samples of different fault types. There are 10,000 samples of normal condition and of each fault type, for a total of 60,000 samples. A total of 60% of them are used for training, 20% for verification, and 20% for testing.

In the online part, the live sampled data are stored in the buffer by the TCU. The TCU preprocesses and performs wavelet transform on the real-time sampled data and then calculates the feature indicators. When the feature indicator J_2 is greater than the threshold Jth (Jth = 0.01), the fault identification program will be enabled, and the four feature indicators will be synthesized into a feature matrix for use in the offline training model, thereby diagnosing the specific type of MCGF.

4. Experiment Verification

4.1. Experimental Data

Figure 6 shows the ETDS in an actual electric locomotive. To fully verify the feasibility of the diagnosis method proposed in this paper, a fault diagnosis experiment was carried out on-site. The experimental data include normal data and five types of MCGF data under different operating conditions.

In practice, as shown in Figure 5, the fault diagnosis board of the TCU needs to detect whether an abnormality occurs every 10 ms. After an abnormality occurs, the offline trained model is called to identify the fault type.

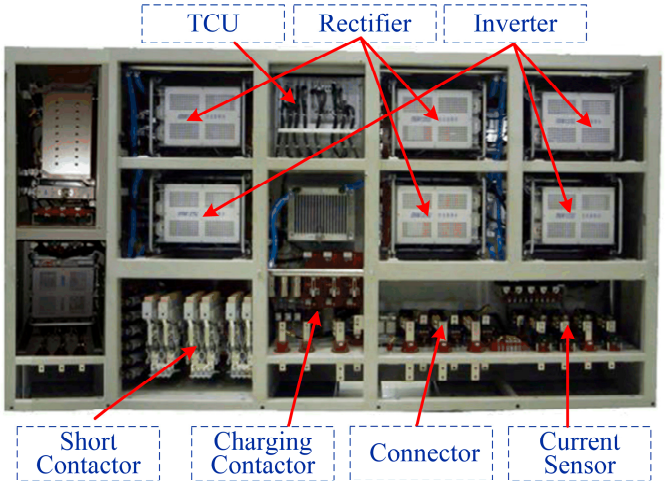

Figure 6. Electrical locomotive diagram of ETDS.

4.2. Result Analysis

In this subsection, the correct isolation rate (CIR) is utilized to intuitively illustrate the diagnostic performance of the proposed method, comparing it with two other machine learning methods to highlight its superiority. A high correct isolation rate (CIR) means that the cause of the fault can be found quicker, the maintenance efficiency can be improved, and the train stop time can be reduced. The expressions of the evaluation indicators are provided as follows:

$$CIR = \frac{The\ number\ of\ faults\ correctly\ located}{N_t} \tag{19}$$

where N_t represents the number of fault data.

Figure 7 shows the feature indicators of different types of MCGF. As shown in Figure 7b, when the ETDS is in normal operation, the value of J_2 is always 0, but when a ground fault occurs, J_2 will change differently depending on the type of ground fault and exceed the Jth threshold, thereby starting the fault identification function.

In the process of model training and fault identification, three machine learning methods, extreme learning machine (ELM), gradient boosting machine (GBM), and random forest (RF), were used. It can be seen from Table 3 that, on the one hand, the CIR of ELM is low when diagnosing C_3 and C_5 ground faults; on the other hand, the accuracy of GBM in diagnosing C_1 and C_5 ground faults is not high enough. Among the different types of grounding fault diagnosis, only the CIR of RF reaches more than 99%, which proves that RF has a better diagnostic effect.

Table 3. Results of different machine learning methods in CIR.

Fault Type	ELM	GBM	RF (Proposed)
C_1	98.68%	95.45%	100%
C_2	100%	100%	100%
C_3	92.23%	99.85%	99.92%
C_4	99.97%	99.98%	99.4%
C_5	93.17%	97.68%	99.77%

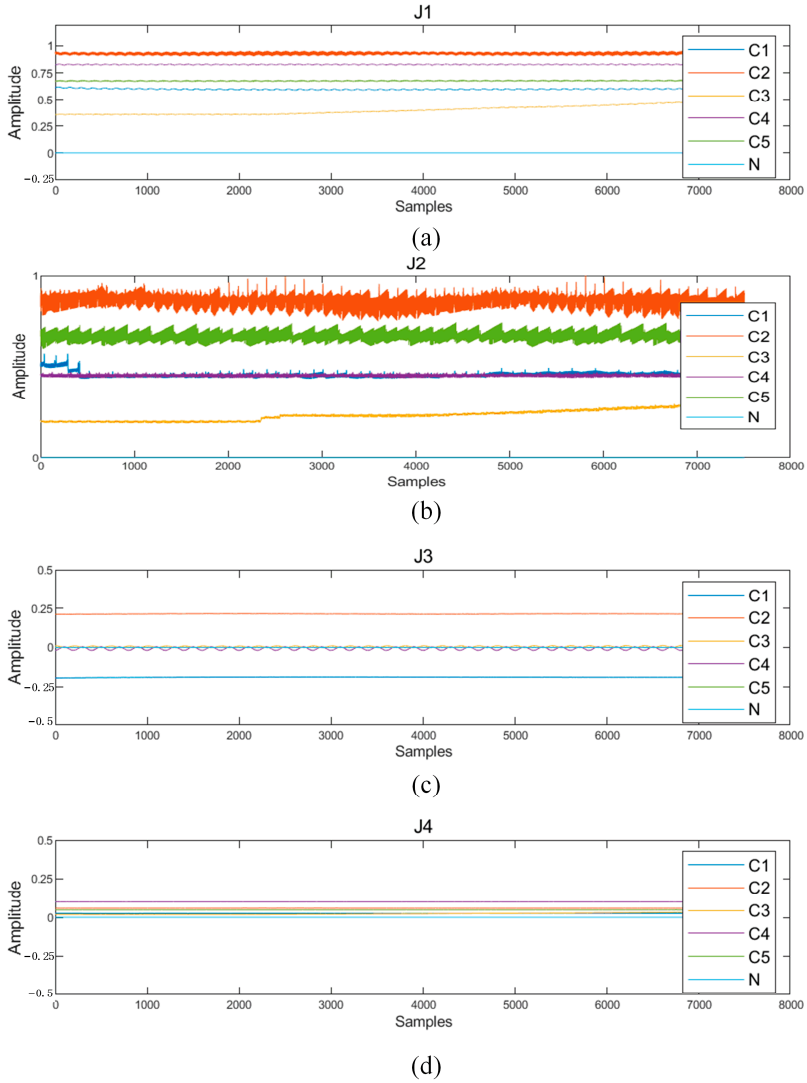

Figure 7. Feature indicators of each MCGF type and normal condition. (**a**) J1 of each MCGF type and normal condition. (**b**) J2 of each MCGF type and normal condition. (**c**) J3 of each MCGF type and normal condition. (**d**) J4 of each MCGF type and normal condition.

During the entire diagnostic process, the sampled data are first preprocessed. In the sampled data shown in Figure 8a, when the ground fault C_5 occurs, U_D and U_2 do not change, and only U_{GD} fluctuates, which fully demonstrates that a single-point ground fault will not disrupt the overall functioning of the system. Subsequently, the data are transformed by wavelet and then calculated to obtain four feature indicators. As presented in Figure 8b, following the occurrence of ground fault C_5, J_2 has obviously exceeded the threshold, and fault identification is enabled. Finally, the ground fault type is determined, as shown in Figure 8c, completing the entire MCGF diagnosis process.

Compared with the previous work performed in [22], first, after preprocessing the data, the time–frequency feature analysis method adopted in this paper is wavelet analysis. Compared with the STFT in [22], there is no need to consider the setting of the window size so as to better balance the accurate analysis of time and frequency. In terms of model prediction, we compared our model not only with the ELM algorithm in the literature but also with the GBM algorithm, and the RF algorithm with better diagnostic performance was obtained, thereby improving the entire diagnostic system, which has certain significance for the diagnosis of train main circuit grounding faults.

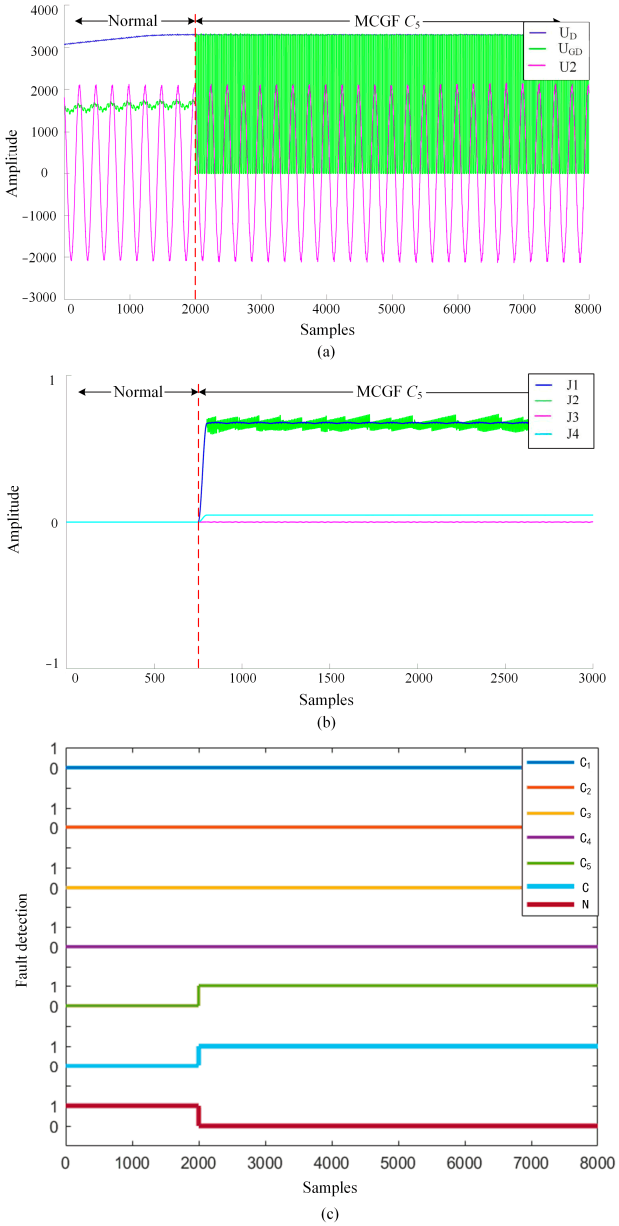

Figure 8. Diagnosis result of fault type C_5. (**a**) Field initial sampling data. (**b**) Feature indicators. (**c**) Diagnosis result of the proposed method.

5. Conclusions

In this article, a hybrid data-driven ground fault diagnosis method for an ETDS is proposed. Two initial feature variables are proposed by combining three original voltage signals, and then the fault features are analyzed in the time–frequency domain by wavelet transform. After that, four feature indicators are calculated and fused into a characteristic vector. Finally, machine learning methods are used for offline training and online fault identification, and the framework of the entire diagnosis process is proposed. In terms of machine learning methods, by comparing the results of field experiments with three different machine learning methods, we find that RF is the most effective method. The proposed method can well distinguish various types of ground faults and reduce the interference of high-frequency noise in the diagnosis results, which can further ensure the operation efficiency and safety of the train.

Author Contributions: Software, J.Z.; Writing—original draft preparation, J.L.; Writing—review and editing, X.H.; Project administration, Q.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Guangdong Province Youth Innovation Talent Project for Ordinary University (2022KQNCX220) and University Research Project of Guangzhou Education Bureau (202235280).

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

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

References

1. Zhang, X.; Niu, Y.; Mao, S.; Cai, Y.; He, R.; Ai, B.; Zhong, Z.; Liu, Y. Resource allocation for millimeter-wave train-ground communications in high-speed railway scenarios. *IEEE Trans. Veh. Technol.* **2021**, *70*, 4823–4838. [CrossRef]
2. Chen, H.; Jiang, B.; Ding, S.X.; Huang, B. Data-Driven Fault Diagnosis for Traction Systems in High-Speed Trains: A Survey, Challenges, and Perspectives. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 1700–1716. [CrossRef]
3. Zhang, K.; Gou, B.; Xiong, W.; Feng, X. An Online Diagnosis Method for Sensor Intermittent Fault Based on Data-Driven Model. *IEEE Trans. Power Electron.* **2023**, *38*, 2861–2865. [CrossRef]
4. Zhang, K.; Jiang, B.; Yan, X.-G.; Mao, Z. Incipient Fault Detection for Traction Motors of High-Speed Railways Using an Interval Sliding Mode Observer. *IEEE Trans. Intell. Transp. Syst.* **2019**, *20*, 2703–2714. [CrossRef]
5. Chen, H.; Jiang, B.; Chen, W.; Yi, H. Data-driven Detection and Diagnosis of Incipient Faults in Electrical Drives of High-Speed Trains. *IEEE Trans. Ind. Electron.* **2019**, *66*, 4716–4725. [CrossRef]
6. Jia, H.; Deng, Y.; Hu, X.; Deng, Z.; He, X. A Concurrent Diagnosis Method of IGBT Open-Circuit Faults in Modular Multilevel Converters. *IEEE J. Emerg. Sel. Top. Power Electron.* **2023**, *11*, 1021–1034. [CrossRef]
7. Xu, S.; Xu, X.; Du, H.; Wang, H.; Chai, Y.; Zheng, W.X.; Chen, H. Comprehensive Diagnosis Strategy for Power Switch, Grid-Side Current Sensor, DC-Link Voltage Sensor Faults in Single-Phase Three-Level Rectifiers. *IEEE Trans. Circuits Syst. I Regul. Pap.* **2024**, *71*, 3343–3356. [CrossRef]
8. Xiao, C.; Wu, W.; Chen, J.; Koutroulis, E.; Blaabjerg, F. Open-Circuit Fault Diagnosis and Fault-Tolerant Control for a Coupled-Inductor-Based Aalborg Inverter. *IEEE Trans. Ind. Electron.* **2024**, *71*, 14044–14053. [CrossRef]
9. Gong, Z.; Huang, D.; Jadoon, H.U.K.; Ma, L.; Song, W. Sensor-Fault-Estimation-Based Tolerant Control for Single-Phase Two-Level PWM Rectifier in Electric Traction System. *IEEE Trans. Power Electron.* **2020**, *35*, 12274–12284. [CrossRef]
10. Huang, J.; Jiang, B.; Xu, C.; Wang, N. Slipping Detection of Electric Locomotive Based on Empirical Wavelet Transform, Fuzzy Entropy Algorithm and Support Vector Machine. *IEEE Trans. Veh. Technol.* **2021**, *70*, 7558–7570. [CrossRef]
11. Dong, H.; Chen, F.; Wang, Z.; Jia, L.; Qin, Y.; Man, J. An Adaptive Multisensor Fault Diagnosis Method for High-Speed Train Traction Converters. *IEEE Trans. Power Electron.* **2021**, *36*, 6288–6302. [CrossRef]
12. Chen, Z.; Li, X.; Yang, C.; Peng, T.; Yang, C.; Karimi, H.R.; Gui, W. A data-driven ground fault detection and isolation method for main circuit in railway electrical traction system. *ISA Trans.* **2019**, *87*, 264–271. [CrossRef] [PubMed]
13. Mao, Z.; Xia, M.; Jiang, B.; Xu, D.; Shi, P. Incipient Fault Diagnosis for High-Speed Train Traction Systems via Stacked Generalization. *IEEE Trans. Cybern.* **2022**, *52*, 7624–7633. [CrossRef] [PubMed]
14. Hoang, D.T.; Kang, H.J. A Motor Current Signal-Based Bearing Fault Diagnosis Using Deep Learning and Information Fusion. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 3325–3333. [CrossRef]
15. Li, X.; Xu, J.; Chen, Z.; Xu, S.; Liu, K. Real-Time Fault Diagnosis of Pulse Rectifier in Traction System Based on Structural Model. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 2130–2143. [CrossRef]
16. Huang, W.; Du, J.; Hua, W.; Bi, K.; Fan, Q. A Hybrid Model-Based Diagnosis Approach for Open-Switch Faults in PMSM Drives. *IEEE Trans. Power Electron.* **2022**, *37*, 3728–3732. [CrossRef]
17. Ji, H. Data-Driven Sensor Fault Diagnosis Under Closed-Loop Control With Slow Feature Analysis. *IEEE Sens. J.* **2022**, *22*, 24299–24308. [CrossRef]
18. Cheng, C.; Qiao, X.; Luo, H.; Wang, G.; Teng, W.; Zhang, B. Data-Driven Incipient Fault Detection and Diagnosis for the Running Gear in High-Speed Trains. *IEEE Trans. Veh. Technol.* **2020**, *69*, 9566–9576. [CrossRef]
19. Chen, H.; Jiang, B. A Review of Fault Detection and Diagnosis for the Traction System in High-Speed Trains. *IEEE Trans. Intell. Transp. Syst.* **2020**, *21*, 450–465. [CrossRef]
20. Zhao, Z.; Li, G. Synchrosqueezing-Based Short-Time Fractional Fourier Transform. *IEEE Trans. Signal Process.* **2023**, *71*, 279–294. [CrossRef]

21. Sotnikov, D.; Lyly, M.; Salmi, T. Prediction of 2G HTS Tape Quench Behavior by Random Forest Model Trained on 2-D FEM Simulations. *IEEE Trans. Appl. Supercond.* **2023**, *33*, 6602005. [CrossRef]
22. Ni, Q.; Luo, H.; Liu, J.; Zhan, Z.; Li, X.; Zhao, Z.; Lai, L.L. A Feature Vector Learning-Based Method for Diagnosing Main Circuit Ground Faults in Electrical Traction Drive Systems. *IEEE Trans. Power Electron.* **2024**, *39*, 2537–2545. [CrossRef]