

Article

Classification of Partial Discharge in Vehicle-Mounted Cable Termination of High-Speed Electric Multiple Unit: A Machine Learning-Based Approach

Yanhua Yang ¹, Jiali Li ², Zhenbao Chen ¹, Yong-Chao Liu ^{3,*}, Kui Chen ², Kai Liu ^{2,*}, Dong-Li Xin ², Guoqiang Gao ² and Guangning Wu ²

- ¹ CRRC Zhuzhou Electric Locomotive Co., Ltd., Zhuzhou 412001, China; 010200003547@crrecg.com (Y.Y.); 010200057433@crrecg.com (Z.C.)
- ² School of Electrical Engineering, Southwest Jiaotong University, Chengdu 611756, China; ljli@my.swjtu.edu.cn (J.L.); kchen@swjtu.edu.cn (K.C.); xindonglihero@my.swjtu.edu.cn (D.-L.X.); xnjdggq@swjtu.edu.cn (G.G.); gnwu@home.swjtu.edu.cn (G.W.)
- ³ Energy Department, UTBM, Université Bourgogne Franche-Comté, 90010 Belfort, France
- * Correspondence: yongchao.liu@utbm.fr (Y.-C.L.); liukai@swjtu.edu.cn (K.L.)

Abstract: This paper presents a machine learning-based approach to identify and separate partial discharge (PD) and two typical pulse interference (PI) signals in the vehicle-mounted cable terminations of high-speed electric multiple units (EMUs). First, a test platform was established to capture PD and two typical PI signals in these terminations. The acquired signals were then processed using the square envelope method to extract feature parameters, such as the rise time proportion, the left–right symmetry, and the upper–lower symmetry. PD signal classification was carried out on these signals, utilizing waveform parameters derived from a hierarchical clustering algorithm. The results validate that the extracted feature components effectively classify and separate PD and two typical PI signals in the vehicle-mounted cable terminations of high-speed EMUs.



Citation: Yang, Y.; Li, J.; Chen, Z.; Liu, Y.-C.; Chen, K.; Liu, K.; Xin, D.-L.; Gao, G.; Wu, G. Classification of Partial Discharge in Vehicle-Mounted Cable Termination of High-Speed Electric Multiple Unit: A Machine Learning-Based Approach. *Electronics* **2024**, *13*, 495. <https://doi.org/10.3390/electronics13030495>

Academic Editors: Anton Rassõlkin, David Pánek, Miklos Kuczmann and Tamás Orosz

Received: 14 December 2023
Revised: 11 January 2024
Accepted: 19 January 2024
Published: 24 January 2024



Copyright: © 2024 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/>).

Keywords: high-speed electric multiple units; vehicle-mounted cable termination; partial discharge; machine learning

1. Introduction

With the rapid development of high-speed railroads, the number of electric multiple units (EMUs) operating on high-speed railway lines, along with their operational density, is increasing [1,2]. Ensuring the safe operation of EMUs, each of which utilizes a single-phase AC power supply and the AC/DC/AC traction drive system, is crucial [3–10]. The vehicle-mounted high-voltage cable, a key component for power transmission, directly impacts the safe operation of EMUs [11,12]. The cable termination, being the core part of the vehicle-mounted cable, is a weak link due to its installation position, complex insulation structure, and harsh operating conditions, making it prone to partial discharge (PD) [13]. The extent and severity of PD tend to increase over time. Detecting the magnitude of cable termination discharge allows for a preliminary assessment of the severity of PD at the cable termination and the extent of insulation material damage [14,15]. The initial stage of PD emits a weak electromagnetic signal, and the testing site often contains various interference signals, including the periodic narrowband, the white noise, and the pulse interference (PI) signal [16,17]. The PI signal, in particular, resembles the shape of signals generated by the PD, making it challenging to distinguish in both time- and frequency-domain features, thereby seriously affecting the detection and classification of discharge signals. Consequently, effectively separating PD signals from mixed signals containing PI signals remains a significant research challenge.

At present, classical algorithms employed against pulse noise include the deviation compensation algorithm [18], the symbolic algorithm [19], the logarithmic cost function

algorithm [20], a series of algorithms based on the generalized maximum correntropy criterion [21], and the affine projection algorithm [22], among others. In [23], the affine projection generalized maximum correntropy filtering algorithm was proposed. This algorithm combines affine projection with generalized maximum correntropy ℓ_2 for system identification in impulsive noise environments, offering improved filtering accuracy and faster convergence without computing the inverse of the input data matrix. A stochastic PI suppression algorithm using particle swarm optimization was proposed in [24] to enhance neural networks for PD detection, boosting training speed and network generalization despite requiring substantial data. In [25], a combined step size method was developed to address the poor tracking ability of traditional variable step size by optimizing the L_1 paradigm of systematic error, thereby effectively identifying PI signals. Further, a method utilizing a multi-sample maximum correlation wavelet high-energy scale was proposed in [26] to suppress random PI signals and white noise. In [27], the PD signal detection and denoising under random PI signals using maximum overlap discrete wavelet transform was explored, demonstrating its superiority over empirical Bayesian wavelet transform. In [28], a PD separation method based on density peak fast search discovery clustering for inverter power supplies was proposed, allowing separation of PD signals and disturbances under inverter test conditions. Despite these advancements, challenges persist in improving convergence performance and managing computational increases.

Artificial intelligence-based approaches have been increasingly utilized in electric apparatuses for control, diagnosis, prognosis, and signal classification and separation [29–35]. Among these, the machine learning-based approach has gained significant attention due to its advantages in accuracy, adaptability, and efficiency [29–32]. In this paper, a machine learning-based approach is proposed for classifying and separating PD and two typical PI signals. This method is based on the waveform parameter analysis combined with hierarchical clustering, specifically addressing two typical PI signals in the PD signal detection in the vehicle-mounted cable termination of high-speed EMUs. The PD and two typical PI signals were obtained from a cable termination test platform built in the laboratory. The waveform characteristics were analyzed, a single pulse envelope was processed, and feature parameters such as the rise time ratio, the left–right symmetry, and the upper–lower symmetry of the waveform were extracted. These parameters were then combined with the hierarchical clustering algorithm for signal classification. The obtained results verified the effectiveness of the extracted feature components, demonstrating good separability. This approach successfully classifies and separates PD and two typical PI signals in the vehicle-mounted cable termination of high-speed EMUs, with the advantages of requiring a small data volume and providing fast operational speed.

The rest of this paper is organized as follows: Section 2 details the collection of typical PD and PI signals from the vehicle-mounted cable terminations of high-speed EMUs. Section 3 conducts feature analysis and introduces the methods used. Section 4 adopts a hierarchical clustering method to classify and separate the PD and PI signals, also demonstrating the effective differentiation of the extracted features on the target signals. Section 5 presents the conclusions.

2. Data Collection

The vehicle-mounted cable termination is structured with multiple layers, including the external umbrella skirt, the heat-shrinkable tube, the insulation tube, the stress tube, the main insulation, and the cable core. This complexity has led researchers conducting field analysis to identify that PD is most likely to occur between the stress tube and the main insulation [36]. To facilitate the collection of PD signals, manufacturers have intentionally created air gap defects in this area. The structure of this artificially defective cable termination is depicted in Figure 1. Additionally, high-speed EMU operations, involving frequent starting and stopping of high-power electrical appliances and space limitations in high-voltage systems, are prone to triggering the corona PI signal and the power supply PI signal. These two PI signals, with high-frequency characteristics overlapping with those of PD signals, are the focus of this study.

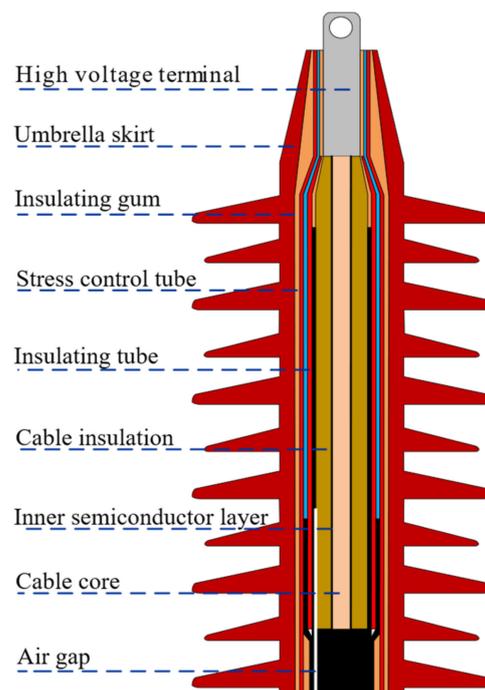


Figure 1. Cable termination structure.

The high-frequency current transformer (HFCT) is a commonly used tool for on-line detection of PD signals, known for its high sensitivity, easy installation, and strong anti-interference capabilities. In this study, the HFCT test loop was utilized, and the constant-voltage method was employed to capture both PD and two typical PI signals. The experimental setup involved a voltage of 27.5 kV and a sampling frequency of 100 MHz using a high-frequency oscilloscope. The collection methods for the PD and two types of PI signals are detailed as follows:

1. A brand-new cable termination without prefabricated defects was used, and a pin was inserted into the anti-corona ball to capture the single corona PI signal.
2. Anti-corona balls were installed on the new cable termination, and electric motors co-grounded with the cable termination were repeatedly started and stopped while boosting and adopting constant-voltage processes to obtain the single power supply PI signal.
3. Anti-corona balls were installed at the end of the prefabricated defective cable termination to obtain the single PD signal from the cable termination.
4. Without corona-proofing, the electric motors grounded jointly with the defective cable termination were repeatedly started and stopped during constant voltage application to obtain mixed PD signals containing both corona and power supply PI signals.

The above-mentioned signals were recorded in their time-domain waveforms, with each signal sample comprising no less than 50 groups, ensuring sufficient data for subsequent PI characterization. The structural schematic of the test platform is illustrated in Figure 2, while Figures 3–5 display the PD, the corona PI, and the power supply PI signals obtained from these experiments. In Figures 3–5 the blue line represents the applied voltage level, and the red line represents the discharge signal collected by the high-frequency oscilloscope.

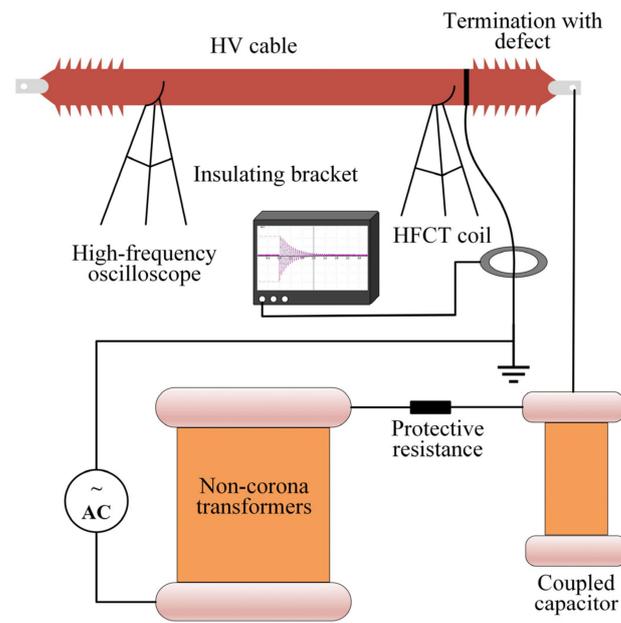


Figure 2. Test circuit wiring diagram.

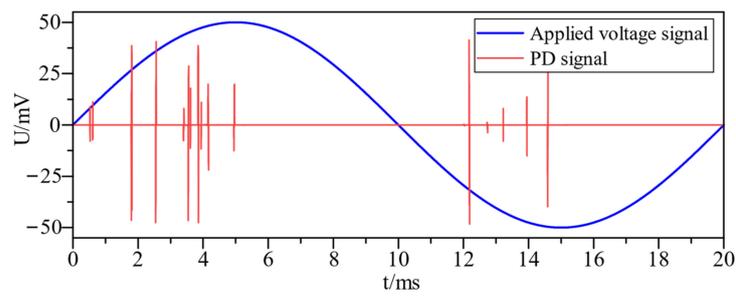


Figure 3. PD signal.

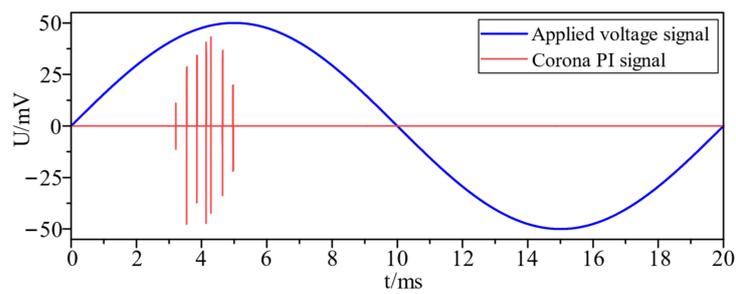


Figure 4. Corona PI signal.

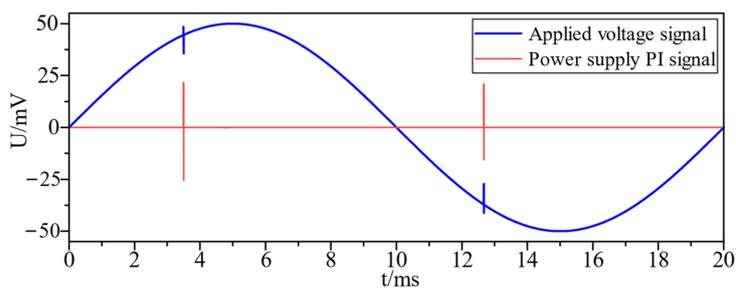


Figure 5. Power supply PI signal.

3. Signal Feature Extraction

3.1. Signal Feature Analysis

In observing the characteristics of single frequency cycle signals, it is evident that the PD signal at the cable termination occurs in both the positive and negative half-waves of the power frequency test voltage, predominantly in the first and third quadrants, and demonstrates a wide phase distribution. In contrast, the external corona PI signal mainly occurs in a narrower phase range within the first quadrant. The external power supply PI signal, triggered by switch actions, can appear in any half-wave and typically presents as isolated events, unlike the groupings seen in the PD and the external corona PI signals. All three pulse signals' amplitudes vary with the externally applied voltage, making phase and amplitude inadequate for separating PD signals from PI signals.

A further analysis of waveform parameters of individual time-domain pulse signals revealed distinct characteristics. As shown in Figure 6, the PD signals typically have a single-wave oscillation period of about 1500–3000 ns and a width of around 200 ns, exhibiting good attenuation and upper–lower symmetry. The external corona PI signal shows an oscillation period of approximately 600–1000 ns, a single oscillation width of about 50 ns, and poorer waveform symmetry. The power supply PI signal displays an oscillation period of about 300–600 ns, along with a single oscillation width of around 20 ns, and is concentrated in the positive amplitude region.

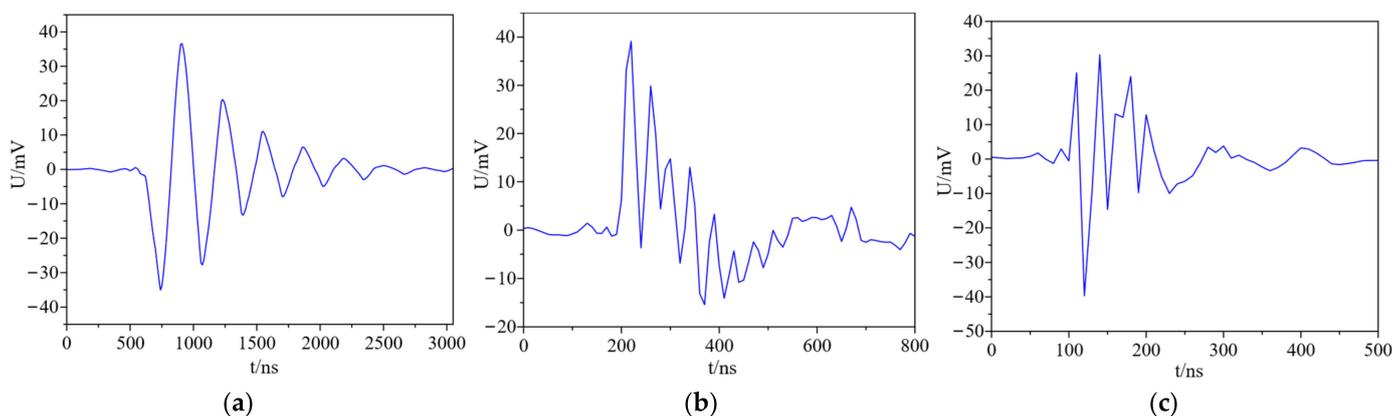


Figure 6. Single-wave time-domain waveform plot: (a) PD signal. (b) Corona PI signal. (c) Power supply PI signal.

In summary, there are clear differences in waveform characteristics between PD and PI signals, suggesting that signal waveform parameters are viable for distinguishing PD signals from PI signals.

3.2. Signal Envelope Processing

The extracted signal undergoes envelope processing to derive waveform parameter features. This study employed the square envelope method based on extremum points, which involves interpolating and fitting the signal envelope line using the pulse waveform's maximum points. The process squares the signal voltage value to yield a unipolar waveform, aiding in the asymmetry analysis of PD and PI signals. The methodology comprises the following steps:

1. Waveform extraction. First, the voltage value of the sampling point of the pulse signal is square and normalized to obtain the unipolar waveform with the following equation:

$$y(t) = x(t)^2 \quad (1)$$

where $y(t)$ is the unipolar waveform and $x(t)$ is the original HFCT signal.

$$\bar{y}(t) = \frac{y(t)}{\max(y(t))} \quad (2)$$

where $\bar{y}(t)$ denotes the normalized waveform, and $\max(y(t))$ denotes $y(t)$ in the maximum value.

2. Determine the maximum value. Obtain the maximum value points in the waveform to form a data sequence.
3. End extending. The endpoint mirroring method in the empirical modal decomposition algorithm is borrowed to extend the endpoints of the maximal value sequence [37].
4. Interpolation fitting. The expanded sequence of maximum values is fitted using cubic spline interpolation to obtain the square envelope signal.

The flow of the square envelope method based on the extreme points is shown in Figure 7, and the square envelope signal obtained is shown in Figure 8.

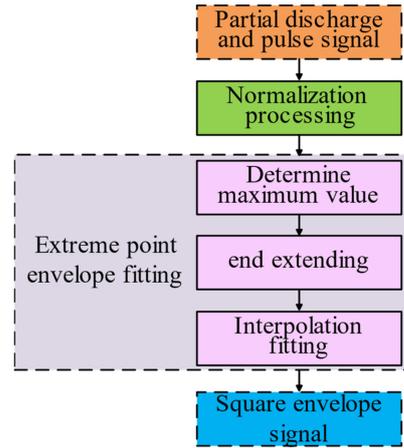


Figure 7. Flowchart of square envelope method.

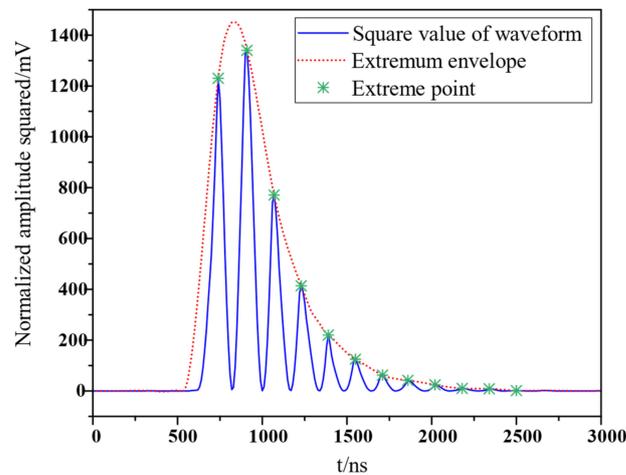


Figure 8. Signal square envelope.

3.3. Signal Feature Extraction

Although PD and PI signals are similar in the time–frequency domain, differences are identified upon analyzing their pulse waveforms. It was found that the envelope characteristics of these signals from different sources are distinct. Thus, waveform parameter characteristics such as the rise time percentage, the left–right symmetry, and the upper–lower symmetry were extracted for each signal type. The three characteristic parameters are defined as:

1. Envelope signal rise time percentage:

$$\alpha = t_{rise} / T_{total} \tag{3}$$

where t_{rise} is the rise time of the envelope signal from 10% amplitude to peak value, and T_{total} is the total duration of the envelope signal.

- The degree of symmetry between the left and right sides of the envelope signal:

$$\beta = A_p/A_t \tag{4}$$

where A_p and A_t correspond to the area of the envelope from the beginning to the peak and from the peak to the end, respectively.

- The degree of symmetry between the top and bottom of the pulse waveform:

$$\gamma = E_2/E_3 \tag{5}$$

where E_2 and E_3 are the second and third pole points of the unipolar waveform, respectively.

The number of individual waveforms was extracted by 200 for the PD signal, the corona PI signal, and the power supply PI signal, and the feature parameters were extracted based on the waveform parameter analysis method, as shown in Figure 9.

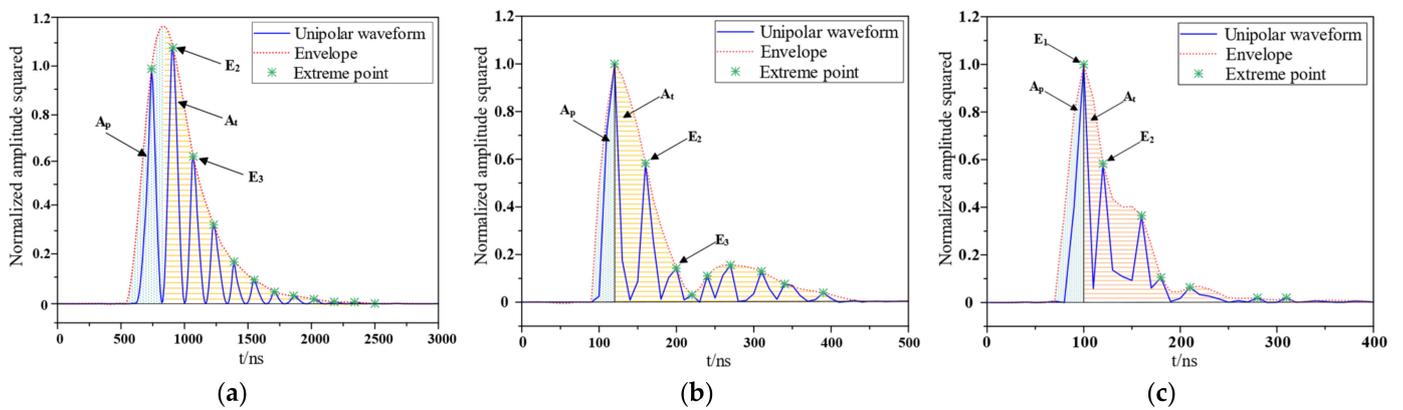


Figure 9. Signal square envelope characteristics: (a) PD envelope characteristics. (b) Corona PI envelope characteristics. (c) Power supply PI envelope characteristics.

Using these characteristic parameters, the distributions of $\alpha - \gamma$, $\beta - \alpha$, and $\beta - \gamma$ can be obtained, as shown in Figure 10. In these distribution plots, it becomes evident that the two-dimensional distributions of the eigenvalues self-aggregate into three distinct categories. This observation suggests that the extracted characteristic parameters effectively classify the three pulse signals. Despite some overlap in the distributions of these pulses, the differences between them are more pronounced. Consequently, the envelope signal rise time percentage α , left–right symmetry β , and pulse waveform upper–lower symmetry γ were identified as the feature parameters for fuzzy clustering. Selecting these parameters aided in effectively separating the PD signal from the two PI signals.

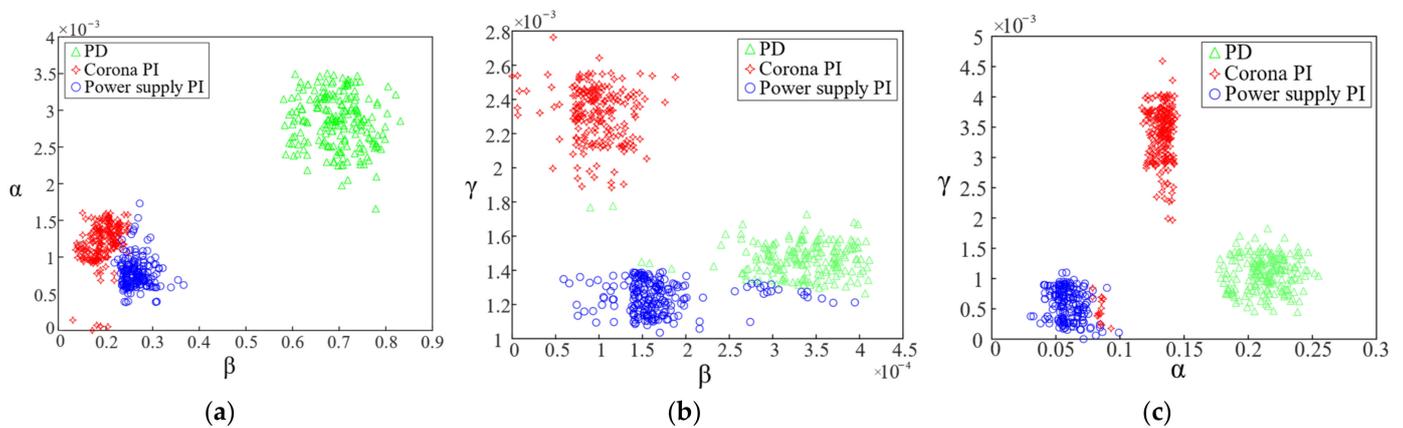


Figure 10. Distribution plot of two-dimensional features: (a) $\alpha - \gamma$. (b) $\beta - \alpha$. (c) $\beta - \gamma$.

3.4. Hierarchical Clustering

Hierarchical clustering is a method that employs a bottom-up approach, starting with each sample point and gradually merging those with high similarity. This process creates a hierarchical nested clustering tree by calculating the similarity between data points of different categories. In such a tree, the original data points of various categories form the lowest level, while the top level is the root node of a cluster. One key advantage of hierarchical clustering is that it does not necessitate a predefined number of clusters; the similarity of distances and rules is easily definable, and the hierarchical relationship of classes can be discerned. The methods for calculating class-to-class distance in hierarchical clustering include: the shortest-distance method (single linkage), the longest-distance method (complete linkage), the intermediate-distance method (average linkage), and the class-average method (ward linkage) [38].

The specific steps of the hierarchical clustering algorithm are as follows, with the flowchart depicted in Figure 11:

1. Calculating the distance or similarity between data points.
2. Initializing each data point as an individual cluster, and forming n_c small clusters denoted as $C = \{C_1, C_2, \dots, C_{n_c}\}$.
3. Based on the chosen distance calculation method, selecting the cluster pairs for merging, typically choosing the two clusters with the closest or most similar distances C_a and C_b :

$$\{C_a, C_b\} = \underset{C_i, C_j \in C, i \neq j}{\operatorname{argmin}} D(C_i, C_j) \tag{6}$$

where $D(,)$ is a function that measures the distance between two clusters.

4. Merging selected cluster pairs and updating the cluster structure.
5. Repeating the merging process until either a preset number of clusters is reached or all data points are consolidated into one cluster.
6. The final cluster center might be a mean vector, median vector, or other form of representative vector for all data points in each cluster.

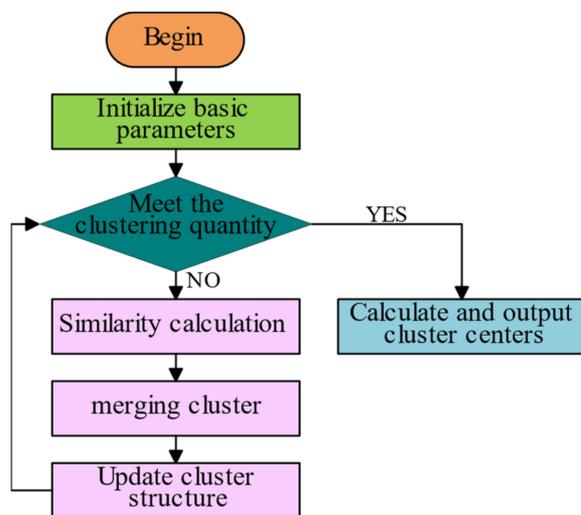


Figure 11. Hierarchical clustering flowchart.

4. Results and Discussion

Five accomplished tasks will be shown in this section. First, setting the clustering parameters and establishing performance evaluation criteria for the clustering methods. Next, investigating the impact of different methods for calculating category distance on the quality of the resulting clusters. This is followed by a comparison of the clustering methods used in this study with other clustering approaches, particularly focusing on mixed signal separation. Finally, the section concludes by discussing potential directions for future

research. Based on the waveform parameter analysis method, 200 feature parameters were extracted from the PD signal, the corona PI signal, and the power supply PI signal for the clustering algorithm classification.

4.1. Parameter Setting and Clustering Quality Evaluation Indexes

Three waveform parameters were identified as feature quantities for hierarchical clustering: the rise time share of the PD and PI envelope signals, left–right symmetry, and upper–lower symmetry. The number of clusters was set to $C = 3$, and Euclidean distance was used as the measure between samples. For the evaluation of clustering quality, the adjusted mutual information (AMI), Fowlkes–Mallows score (FMI), and Davies–Bouldin Index (DBI) were utilized.

The AMI is expressed in Equation (7). It is adjusted for the probability of clustering groups to match the real scenario, with a value range of $[-1, 1]$. A larger AMI value indicates that the clustering results are more consistent with the actual situation.

$$AMI = \frac{MI - E[MI]}{\text{mean}(H(U), H(V)) - E[MI]} \quad (7)$$

where U and V are the assignments to the N sample labels, the $H(U)$ and $H(V)$ are their entropies, MI is the maximum variance, and $E[MI]$ is the expected value of the maximum variance.

The FMI, representing the geometric mean of Precision and Recall, has a value range of $[0, 1]$; the closer to 1, the better the performance. The FMI is calculated as follows:

$$FMI = \frac{TP}{\sqrt{(TP + FP)(TP + FN)}} \quad (8)$$

where TP denotes that the sample pair is a nest in the true value and also a nest in the predicted value, FP denotes that the sample pair is a nest in the predicted value but not a nest in the true value, and FN denotes that the sample pair is a nest in the true value but not a nest in the predicted value.

DBI is calculated by dividing the sum of average intraclass distances of any two classes by the distance between their cluster centers, aiming to find the minimum value. A smaller DBI value indicates smaller intraclass distances and larger interclass distances. The DBI is expressed as

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} \frac{s_i + s_j}{d_{ij}} \quad (9)$$

where s_i denotes the average distance between each point of a cluster and the center of mass of that cluster, also known as the cluster diameter; d_{ij} denotes the distance between the centers of mass of clusters i and j ; and k is the current class.

4.2. Effect of Different Methods of Calculating Category Distances on the Quality of the Clusters Clustered

Different methods of calculating distances, also known as linkage methods, are suitable for various types of data and clustering structures. The choice of an appropriate linkage method significantly impacts the quality and interpretability of the clustering results. In this study, the effectiveness of four hierarchical clustering linkage methods was investigated. The clustering effect diagrams for these methods are illustrated in Figure 12, while Table 1 presents a clustering quality analysis for different distance calculation methods.

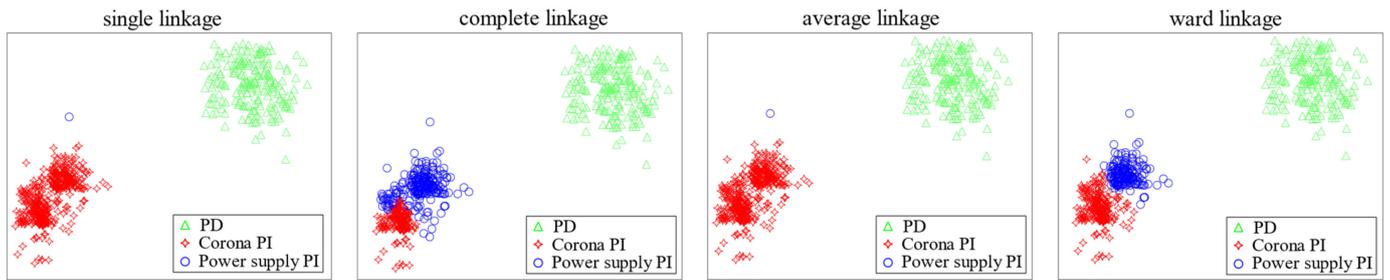


Figure 12. Clustering results of different linking methods.

Table 1. Clustering evaluation of different linking methods.

	Single Linkage	Complete Linkage	Average Linkage	Ward Linkage
AMI	−0.729	0.813	−0.729	0.897
FMI	0.773	0.868	0.773	0.940
DBI	0.702	0.461	0.702	0.286
Time	2.432 s	4.536 s	3.432 s	2.034 s

From Figure 12, it is observed that both single and average linkages tend to categorize the power supply PI signal almost entirely as the corona PI signal. A similar trend is noted with complete linkage, whereas ward linkage accurately identifies the three pulse signals. Table 1 further reveals that the AMI and FMI for the ward linkage are approximately 0.9, significantly higher than the other three methods. Additionally, the DBI is minimized, and the computation time is relatively short. Consequently, ward linkage was selected for constructing the hierarchical clustering. The results, depicted in Figure 13, demonstrate that various types of pulse signals cluster closely around their respective centers, with clear distinctions between different categories. This outcome suggests that the hierarchical clustering algorithm effectively identifies and separates PD and PI signals based on the extracted feature parameters (Figure 14). The clustering centers for the calculated PD and PI signals are detailed in Table 2.

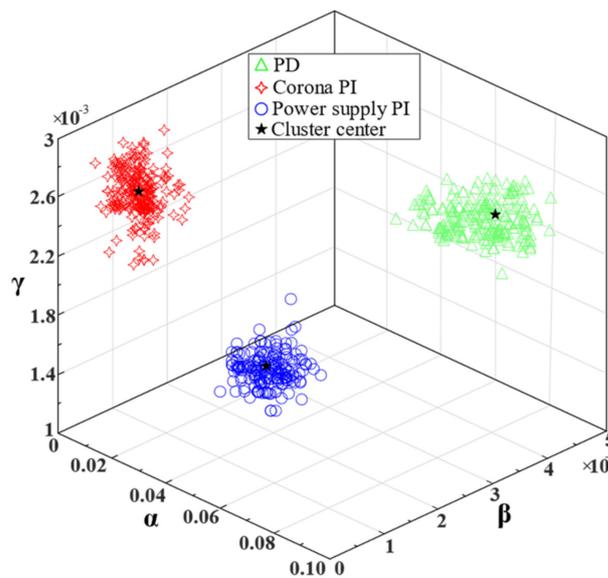


Figure 13. Hierarchical clustering results.

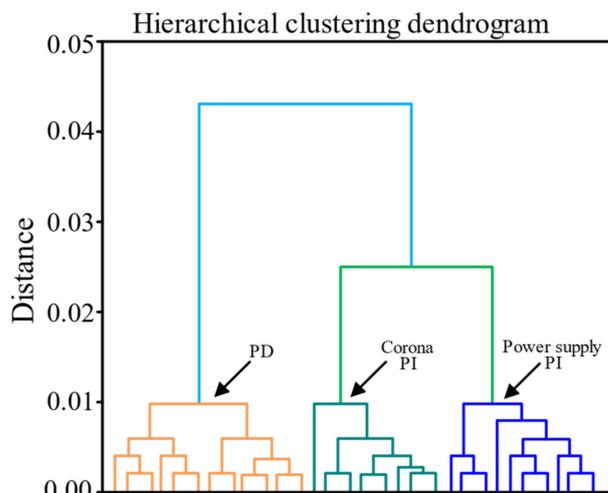


Figure 14. Clustering tree.

Table 2. Clustering centers.

Type	α	β	γ
PD signal	0.06313	4.361×10^{-4}	1.9451×10^{-3}
Corona PI signal	0.02102	1.320×10^{-4}	2.4743×10^{-3}
Power supply PI signal	0.01483	3.1828×10^{-4}	1.2312×10^{-3}

4.3. Comparison of Different Clustering Algorithms

In order to select the most suitable clustering algorithm and compare the effectiveness of different clustering algorithms in studying pulse disturbances in partial discharges, this study included classical clustering algorithms from traditional approaches. These were K-means clustering, fuzzy c-means clustering (FCM), and density-based spatial clustering of applications with noise (DBSCAN). The confusion matrices of four methods, including hierarchical clustering, are illustrated in Figure 15, with the clustering quality analysis presented in Table 3.

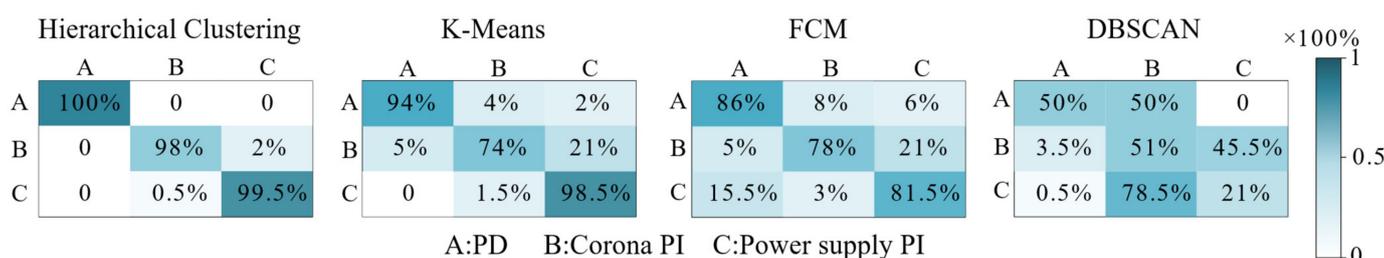


Figure 15. Confusion matrix for different clustering methods.

Table 3. Quality analysis of different clustering methods.

	Hierarchical Clustering	K-Means	FCM	DBSCAN
AMI	0.907	0.829	0.813	0.229
FMI	0.940	0.868	0.773	0.373
DBI	0.226	0.368	0.461	0.802
Time	2.432 s	5.036 s	5.032 s	8.034 s

From an examination of Figure 15 and Table 3, the overall performance ranking is observed as hierarchical clustering > K-means clustering > FCM > DBSCAN. Hierarchical clustering demonstrates an almost complete and correct identification of PD and power

supply PI signals, although it occasionally misclassifies the corona PI signal as the power supply PI signal. Its AMI and FMI are both greater than 0.9, and it has the smallest DBI, indicating its accuracy in recognizing the three types of signals. Among the four methods, DBSCAN shows the least favorable performance, proving unsuitable for PI recognition. Considering the running times of these methods, hierarchical clustering emerges as the most effective.

4.4. Mixed Signal Separation

On the PD testing platform established in the laboratory, a mixed signal comprising PD, external corona, and power supply PI generated by multiple vehicle-mounted cable terminations within one power frequency cycle was captured, as shown in Figure 16. We extracted multiple α , β , and γ parameters from mixed pulse signals. We calculated the Euclidean distance between the pulse feature vector to be recognized and the standard cluster center, in order to effectively distinguish different types of pulse signals. This method uses standard clustering centers to cluster and separate mixed signals. For instance, selecting a mixed signal example as shown in Figure 16, the classification results include 10 corona PI signals and 5 power supply PI signals, while the remaining are PD signals. The clustering results are shown in Figure 17. Figure 18 shows the PD signal after PI signal separation. The separation process reduces the total amplitude and discharge phase of the PD signal in the first and third quadrants. This reduction effectively reduces the impact of PI signals on PD detection accuracy, thereby improving the reliability of PD methods in evaluating the insulation status of vehicle-mounted cable terminations.

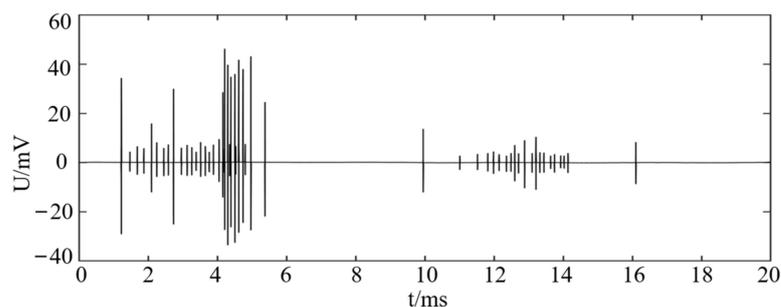


Figure 16. Mixed Signals.

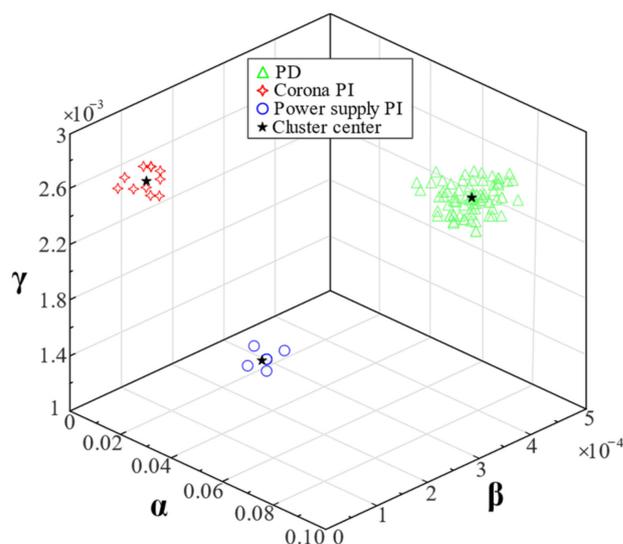


Figure 17. Mixed signal separation results.

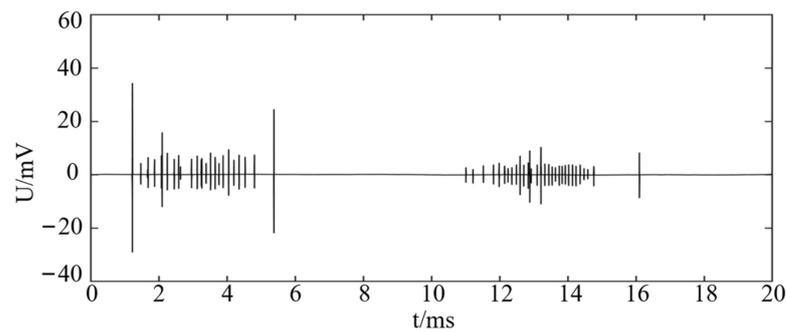


Figure 18. Partial discharge signal after separation of PI signals.

4.5. Discussion

This study focused on the high-frequency PD and PI signals at vehicle-mounted cable terminations of high-speed EMUs, yielding significant results that aid in separating PD signals from common PI signals. However, due to time constraints, this study faced certain limitations that warrant further investigation:

1. While this study simulated the generation of two common PI signals during the operation of high-speed EMUs and separated them from PD signals, the actual PI experienced in high-speed EMUs is more diverse than the two types discussed. Thus, additional research is needed to address the separation of these more varied signals.
2. The methodology employed here is effective for recognition and separation in scenarios with small sample sizes, such as samples ranging from tens to hundreds of data. Moreover, for larger sample conditions, such as when the data volume reaches thousands, integrating traditional methods with artificial intelligence approaches could offer more comprehensive and efficient solutions.

5. Conclusions

This paper presents a machine learning-based approach for identifying and separating PD and two typical PI signals from vehicle-mounted cable terminations of high-speed EMUs using waveform parameter analysis and a hierarchical clustering algorithm. The conclusions are as follows:

1. Envelope processing of the extracted PD and two typical PI signals yields characteristic parameters like the rise time percentage, the left–right symmetry, and the upper–lower symmetry of the enveloped signal waveforms. These feature parameters have proven effective in distinguishing the target signal.
2. The study proposes an innovative method, amalgamating waveform parameter analysis with a hierarchical clustering algorithm. Impressively, with AMI and FMI metrics surpassing 0.9, and the smallest DBI at 0.226 achieved in just 2.423 s, the approach demonstrates exceptional performance. These findings affirm the approach’s success in effectively distinguishing and isolating PD and two typical PI signals from vehicle-mounted cable terminations in high-speed EMUs.
3. The proposed method successfully isolates PD signals under mixed PI conditions, demonstrating the effectiveness and accuracy of the scheme. This advancement not only mitigates the impact of PI signals on PD detection but also achieves almost one hundred percent accuracy in identifying PD signals from mixed signals. Consequently, it enhances the accuracy of using PD measures to assess the insulation status of vehicle-mounted cable terminations in high-speed EMUs.

Author Contributions: Conceptualization, K.C.; methodology, J.L.; software, Z.C.; validation, Y.Y., G.G. and G.W.; formal analysis, Y.-C.L.; investigation, Y.Y.; resources, K.L.; data curation, Z.C. and D.-L.X.; writing—original draft preparation, J.L.; writing—review and editing, Y.-C.L. and K.L.; visualization, G.G.; supervision, G.W.; project administration, Z.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Natural Science Foundation of China (NSFC, U1966602, 52377161, 52007158), Excellent Young Scientists Fund of China (51922090), Southwest Jiaotong University new interdisciplinary cultivation project (YH1500112432273, YH15001124322105), and Fundamental Research Funds for the Central Universities (A0920502052301-170).

Data Availability Statement: The data presented in this study are available on request from the corresponding authors. The data are not publicly available due to ongoing further analysis and the anticipation of data sharing upon completion.

Conflicts of Interest: Authors Zhenbao Chen and Yanhua Yang are employed by CRRC Zhuzhou Electric Locomotive Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Cheng, Z.-P.; Kong, H.-W.; Ma, J.; Jia, L.-M. Overview of resilient traction power supply systems in railways with interconnected microgrid. *CSEE J. Power Energy Syst.* **2020**, *7*, 1122–1132.
2. Guo, L.; Cao, W.; Bai, L.; Zhang, J.; Xing, L.; Xiang, E.; Zhou, L. Fault diagnosis based on multiscale texture features of cable terminal on EMU of high-speed railway. *IEEE Trans. Instrum. Meas.* **2020**, *70*, 3502612. [[CrossRef](#)]
3. Ge, X.-L.; Pu, J.-K.; Liu, Y.-C. Online open-switch fault diagnosis method in single-phase PWM rectifier. *Electron. Lett.* **2015**, *51*, 1920–1922. [[CrossRef](#)]
4. Chen, H.; Jiang, B. A review of fault detection and diagnosis for the traction system in high-speed trains. *IEEE Trans. Intell. Transp. Syst.* **2019**, *21*, 450–465. [[CrossRef](#)]
5. Zuo, J.; Ding, J.; Feng, F. Latent leakage fault identification and diagnosis based on multi-source information fusion method for key pneumatic units in Chinese standard electric multiple units (EMU) braking system. *Appl. Sci.* **2019**, *9*, 300. [[CrossRef](#)]
6. Cheng, C.; Wang, J.; Chen, H.; Chen, Z.; Luo, H.; Xie, P. A review of intelligent fault diagnosis for high-speed trains: Qualitative approaches. *Entropy* **2021**, *23*, 1. [[CrossRef](#)]
7. Huang, W.; Kou, X.; Zhang, Y.; Mi, R.; Yin, D.; Xiao, W.; Liu, Z. Operational failure analysis of high-speed electric multiple units: A Bayesian network-K2 algorithm-expectation maximization approach. *Reliab. Eng. Syst. Saf.* **2021**, *205*, 107250. [[CrossRef](#)]
8. 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](#)]
9. Tang, Z.; Chen, Z.; Sun, J.; Lu, M.; Liu, H. Noise prediction study of traction arc tooth cylindrical gears for new generation high-speed electric multiple units. *Lubricants* **2023**, *11*, 357. [[CrossRef](#)]
10. Wu, Y.; Liu, X.; Zhou, Y. Deep PCA-based incipient fault diagnosis and diagnosability analysis of high-speed railway traction system via FNR enhancement. *Machines* **2023**, *11*, 475. [[CrossRef](#)]
11. Bai, L.; Fan, D.; Li, T.; Li, B.; Su, M.; Fan, S.; Zhang, L. Influence of surface discharge on the deterioration characteristics of ethylene-propylene rubber cable insulation under alternating current high voltage. *IET Sci. Meas. Technol.* **2022**, *16*, 293–304. [[CrossRef](#)]
12. Gao, G.; Zhou, S.; Yang, S.; Chen, K.; Xin, D.; Tang, Y.; Liu, K.; Wu, G. Accurate Identification Partial Discharge of Cable Termination for High-speed Trains Based on Wavelet Transform and Convolutional Neural Network. *Electric Power Syst. Res.* **2023**, *225*, 109838. [[CrossRef](#)]
13. Eigner, A.; Rethmeier, K. An overview on the current status of partial discharge measurements on AC high voltage cable accessories. *IEEE Electr. Insul. Mag.* **2016**, *32*, 48–55. [[CrossRef](#)]
14. Alvarez, F.; Ortego, J.; Garnacho, F.; Sanchez-Uran, M.A. A clustering technique for partial discharge and noise sources identification in power cables by means of waveform parameters. *IEEE Trans. Dielectr. Electr. Insul.* **2016**, *23*, 469–481. [[CrossRef](#)]
15. Pan, J.; Wang, M.; Hu, Q.; Li, C. A detection method of partial discharge signal based on wavelet. In Proceedings of the 7th International Conference on Integrated Circuits and Microsystems, Xi'an, China, 28–31 October 2022.
16. Zhong, J.; Bi, X.; Shu, Q.; Chen, M.; Zhou, D.; Zhang, D. Partial discharge signal denoising based on singular value decomposition and empirical wavelet transform. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 8866–8873. [[CrossRef](#)]
17. Hussein, R.; Shaban, K.-B.; El-Hag, A.-H. Denoising different types of acoustic partial discharge signals using power spectral subtraction. *High Voltage* **2018**, *3*, 44–50. [[CrossRef](#)]
18. Wang, W.Y.; Zhao, H.Q.; Lu, L. Bias-compensated constrained least mean square adaptive filter algorithm for noisy input and its performance analysis. *Digital Signal Process.* **2019**, *84*, 26–37. [[CrossRef](#)]
19. Sheng, Z.; Zhang, J.-S.; Han, H.-Y. Robust shrinkage normalized sign algorithm in an impulsive noise environment. *IEEE Trans. Circuits Syst. II Express Briefs* **2017**, *64*, 91–95.
20. Sayin, M.-O.; Vanli, N.-D.; Kozat, S.-S. A novel family of adaptive filtering algorithms based on the logarithmic cost. *IEEE Trans. Signal Process.* **2014**, *62*, 4411–4424. [[CrossRef](#)]
21. Chen, B.D.; Xing, L.; Zhao, H.-Q. Generalized correntropy for robust adaptive filtering. *IEEE Trans. Signal Process.* **2016**, *64*, 3376–3387. [[CrossRef](#)]

22. Ozeki, K.; Umeda, T. An adaptive filtering algorithm using an orthogonal projection to an affine subspace and its properties. *Electron. Commun. Jpn. Part 1 Commun.* **1984**, *67*, 19–27. [[CrossRef](#)]
23. Zhan, J.; Guo, L.-X.; Li, Y.-S. The bias-compensated proportionate nlms algorithm with sparse penalty constraint. *IEEE Access* **2020**, *8*, 4954–4962.
24. Shao, Z.; Huang, C.; Xiao, Y.; Zhao, Y.; Jiang, X. Application of PSO based neural network in suppression of stochastic pulse interference for partial discharge monitoring in large generators. *Autom. Electr. Power Syst.* **2005**, *29*, 49–52.
25. Huang, F.-Y.; Zhang, J.-S.; Zhang, S. Combined-step-size affine projection sign algorithm for robust adaptive filtering in impulsive interference environments. *IEEE Trans. Circuits Syst. II Express Briefs* **2016**, *63*, 493–497. [[CrossRef](#)]
26. Zhang, P.; Zhou, X.; Tian, T.; Wang, Y.; Li, X.; He, N.; Zhang, G.; Zhang, X.; Sun, J. Method of multi-sample maximum correlation wavelet high energy scale on location time difference calculation of partial discharge source. In Proceedings of the IEEE 4th Conference on Energy Internet and Energy System Integration, Wuhan, China, 30 October–1 November 2020.
27. Shams, M.-A.; El-Shahat, M.; Anis, H.-I. Detection and de-noising of pd signal contaminated with stochastic pulse interference using maximal overlap discrete wavelet transform. In Proceedings of the IEEE 3rd International Conference on Dielectrics, Valencia, Spain, 5–31 July 2020.
28. Li, Z.; Zhou, K.; Huang, Y.; Zhou, G.; Ye, B. A novel partial discharge pulse separation method for variable frequency resonant test. In Proceedings of the IEEE International Conference on High Voltage Engineering and Application, Beijing, China, 6–10 September 2020.
29. Kotsiantis, S.B.; Zaharakis, I.D.; Pintelas, P.E. Machine learning: A review of classification and combining techniques. *Artif. Intell. Rev.* **2006**, *26*, 159–190. [[CrossRef](#)]
30. Zhao, S.; Blaabjerg, F.; Wang, H. An overview of artificial intelligence applications for power electronics. *IEEE Trans. Power Electron.* **2021**, *36*, 4633–4658. [[CrossRef](#)]
31. Zhao, Z.; Wu, J.; Li, T.; Sun, C.; Yan, R.; Chen, X. Challenges and opportunities of AI-enabled monitoring, diagnosis & prognosis: A review. *Chin. J. Mech. Eng.* **2021**, *34*, 56.
32. Samanta, A.; Chowdhuri, S.; Williamson, S.S. Machine learning-based data-driven fault detection/diagnosis of lithium-ion battery: A critical review. *Electronics* **2021**, *10*, 1309. [[CrossRef](#)]
33. Hakim, M.; Omran, A.A.B.; Ahmed, A.N.; Al-Waily, M.; Abdellatif, A. A systematic review of rolling bearing fault diagnosis based on deep learning and transfer learning: Taxonomy, overview, application, open challenges, weaknesses and recommendations. *Ain Shams Eng. J.* **2023**, *14*, 101945. [[CrossRef](#)]
34. Liu, Y.-C.; Laghrouche, S.; N'Diaye, A.; Cirrincione, M. Hermite neural network-based second-order sliding-mode control of synchronous reluctance motor drive systems. *J. Frankl. Inst.* **2021**, *385*, 400–427. [[CrossRef](#)]
35. Chen, K.; Liao, Q.; Liu, K.; Yang, Y.; Gao, G.; Wu, G. Capacity degradation prediction of lithium-ion battery based on artificial bee colony and multi-kernel support vector regression. *J. Energy Storage* **2023**, *72*, 108160. [[CrossRef](#)]
36. Xin, D.; Wu, G.; Chen, K.; Liu, K.; Xie, Y.; Gao, G.; Xiao, S.; Tang, Y.; Sun, C.; Lin, M. Research on the evolution characteristics of interfacial defect inside the vehicle-mounted high-voltage cable termination for high-speed trains. *CSEE J. Power Energy Syst.* **2023**. [[CrossRef](#)]
37. Wu, Y.; Xin, H.; Wang, J.; Wang, X. Rolling bearing fault diagnosis based on the variational mode decomposition filtering and extreme point envelope order. *J. Vib. Shock* **2018**, *37*, 102–107.
38. Yang, J.; Parikh, D.; Batra, D. Joint unsupervised learning of deep representations and image clusters. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016.

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.