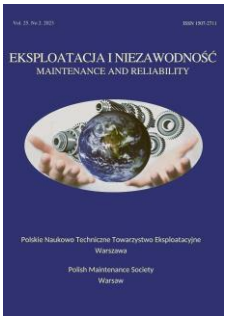


EksploatacjaiNiezawodnosc – Maintenance and Reliability
Volume 28 (2026), Issue 4
journal homepage: http://www.ein.org.pl

Article citation info:
Wu W, Ye W, Zhao M, Zhang Y, Ma Z, Reliability enhancement in substation partial discharge real-time monitoring system based on web data flow-EEMD singular value entropy and distributed computing,EksploatacjaiNiezawodnosc – Maintenance and Reliability 2026: 28(4)http://doi.org/10.17531/ein/220093

Reliability enhancement in substation partial discharge real-time monitoring system based on web data flow-EEMD singular value entropy and distributed computing

Indexed by: Web of Science Group

Wentao Wu[a,*], Wei Ye[a], Minghui Zhao[a], Yunfeng Zhang[a], Zhiming Ma[a]

[a] Qinghai Dehong Electric Power Technology Co., Ltd, China

Highlights

- EEMD and singular value entropy combined for accurate PD signal characterization.
- ISSA-SVM optimizes PD type recognition accuracy in GIS equipment.
- Hadoop-based distributed architecture enables real-time PDM data processing.
- PCA-ISSA-SVM integration reduces feature dimensionality.
- Hybrid algorithm outperforms traditional methods in real-time performance.

Abstract

In power systems, reliable substation operation is critical for ensuring power supply stability and continuity. The reliability of substations directly affects the performance of the entire power system. However, existing substation monitoring systems have significant deficiencies in signal feature extraction accuracy and partial discharge (PD) type detection, which substantially limits the reliability and timeliness of fault diagnosis. To address these issues, this study proposes a hybrid approach based on ensemble empirical mode decomposition singular value entropy (EEMD-SVE). Moreover, a real-time PD monitoring system is developed using Hadoop and the hybrid algorithm. The suggested hybrid algorithm's performance is contrasted with that of alternative algorithms. Experimental outcomes revealed that the algorithm's F1 score was 97.88%, the average MSE value was 1.537, the average RMSE value was 0.462, and the fitting coefficient value was 0.968, all of which were better than the comparison algorithm. Subsequently, the application effect of the proposed monitoring system was analyzed. According to the findings, the system outperformed the comparison system in identifying four different types of partial discharges, with accuracy rates of 98.6%, 97.3%, 96.9%, and 97.8%, respectively. Furthermore, the study's suggested monitoring system outperformed the comparative system with an error rate of 0.14% and a throughput of 1132 requests per second. In conclusion, the algorithm and monitoring system proposed in this study are effective in enhancing the reliability of substation operation and fault diagnosis capabilities, providing a theoretical basis for partial discharge research.

Keywords
web data flow, EEMD, singular value entropy, distributed computing, substation partial discharge

1. Introduction

With the rapid development of the social economy, the growth of electricity demand and the continuous expansion of the power grid scale, the reliable operation of substations is of vital importance to the stability and safety of the power system. As an important part of the power grid, the substation's internal gas insulation switchgear (GIS) is in long-term operation. Due to the existence of defects such as air gaps and impurities, partial discharge (PD) signals are often triggered [1]. It can serve as

(*) Corresponding author.
E-mail addresses: W. Wu (ORCID: 0009-0009-4225-1485) johnnynelson789@163.com, W. Ye (ORCID: 0009-0009-5589-2775) 13997232476@139.com, M. Zhao (ORCID: 0009-0002-1878-4248) zmh2025001@163.com, Y. Zhang (ORCID: 0009-0008-8164-8226) 16608602288@163.com, Z. Ma (ORCID: 0009-0000-5958-5638)m15597031932_1@163.com

a foundation for assessing the insulation condition of equipment by detecting and identifying PD signals, guaranteeing the stable and secure functioning of substation equipment [2]. These partial discharge signals not only affect the insulation performance of the equipment, but may also cause faults, thereby affecting the reliability of the substation. Although many methods currently monitor GIS PDs in real time through the Web data stream of the monitoring system, there are still problems such as low accuracy in signal feature extraction and low accuracy in PD type identification. This limits the reliability and timeliness of substation fault diagnosis. Many experts have done relevant research. In an effort to address the issue of low PDM efficiency in substation switchgear, D. Song et al. created a multi-sensor integrated real-time PDM system for switchgear. Experimental test outcomes revealed that the system could transmit data in real time and accurately capture discharge events, however, the accuracy of feature extraction in this system still needs to be improved [3]. R. Yang et al. proposed a GIS PD detection robot. Experimental results showed that the robot had stable detection efficiency, but the identification accuracy low rise [4]. To solve the problem of reduced accuracy of PDM in air-insulated substations, R. Wang et al. proposed a substation patrol vehicle positioning technology based on a received signal strength indicator and a vehicle-mounted ultra-high frequency sensor array, however, the computational complexity is relatively high [5]. Y. Wang et al. proposed a substation PD fault diagnosis (FD) method that combines logic, convolutional neural network and transfer learning. According to experimental findings, this approach has a high recognition accuracy, but the real-time performance is not very good [6]. To address the lack of comparability of transformer PD UHF range measurement techniques between different measurement systems and sensors, M. Siegel et al. proposed a PDM system. Example results showed that the system was effective [7]. Yan Y et al. proposed a dual-verification monitoring method combining lightweight Building Information Modeling and real-time laser point cloud matching in order to improve the efficiency of equipment maintenance in the substation base. The experimental results show that this method can accurately monitor the maintenance progress of different equipment in the substation. It is reliable, however, the adaptability in handling nonlinear and non-stationary signals still needs to be verified [8]. Aiming at the problem of partial discharge detection in transformers, the SikorSki W team introduced the design of an ultra-high frequency Hilbert curve fractal antenna suitable for partial discharge monitoring systems. The simulation experiment results show that this antenna has the characteristics of small size and high sensitivity But it has not been put into practical use [9].

To achieve RTM of PD in substations based on Web data flow, a methodical and efficient approach is used. For substations to run smoothly, RTM of PDs is crucial. A time-frequency analysis technique for handling nonlinear and non-stationary data is ensemble empirical mode decomposition (EEMD). It is extensively utilized in domains like large data analysis and signal processing and offers the benefits of anti-noise interference and effective big data processing [10]. Hadoop is a distributed data and computing system that is popular in domains like big data analysis because of its high fault tolerance and dependability [11]. Many experts have done relevant research. For example, to solve the problem of delay estimation accuracy of PD signals under the influence of noise and multipath effects, M. L. Li et al. suggested a PD signal preprocessing method based on dual EEMD and reconstruction. The findings demonstrated that this approach has greater stability and robustness when compared to the comparative method [12]. A prediction model based on EEMD and long short-term memory networks was proposed by N. N. V. Nhat et al. This model outperformed the comparison model, according to comparative experimental results [13]. K. Zheng et al. developed an FD model based on EEMD to increase meter measurement accuracy. Results from experiments demonstrated the usefulness of this paradigm [14]. To solve the problems of large amount of substation inspection data and low degree of resource integration, G. Zhang et al. proposed a substation equipment inspection algorithm model based on Hadoop. The model has good real-time performance and accuracy, according to case study and performance analysis results [15]. A distributed data storage management and parallel verification rule execution solution based on the Hadoop platform was proposed by J. Zhao et al. The outcomes of the experiments demonstrated the effectiveness of this method [16]. In view of the delay and scalability problems caused by the processing of high-speed and large capacity stream data in the traditional

power grid system, J. Wu and h. Li proposed a real-time data processing and optimal scheduling scheme for power system based on Kafka and storm technology. The experimental results show that this scheme has faster data processing speed and shorter system delay than the traditional scheme [17]. In order to solve the problem of increasing power monitoring data, K. Zhang and J. Lazaro introduced the principle and technology of big data analysis technology in power equipment life cycle management and fault prediction. The example results show that big data analysis technology can effectively improve the speed of power data processing [18].

The above research results indicate that the current methods for classifying and identifying local discharge in substations have problems such as low accuracy in feature extraction, poor real-time performance, and low adaptability to non-stationary signals. Therefore, to extract the features of substation PD signal data from Web data streams, this work presents EEMD. The singular value entropy (SVE) approach is presented to address the issue of spurious components produced when the EEMD algorithm breaks down intrinsic mode functions (IMF) of various scales. Subsequently, the dimensionality of the acquired feature data is decreased using principal component analysis (PCA). Meanwhile, the support vector machine (SVM) algorithm optimized by the improved sparrow search algorithm (ISSA) is used to construct a hybrid algorithm (HA) combined with the improved EEMD algorithm to identify GIS PD types. Finally, a HA and Hadoop-based substation PDM system is built to enhance substation PDM accuracy and real-time performance. To increase the accuracy of feature extraction, this research innovates by first combining the EEMD and SVE algorithms. Secondly, the PCA algorithm, ISSA and SVM algorithm are combined to improve the accuracy of identifying PD patterns. Finally, the Hadoop framework is introduced to improve the real-time performance of PDM in substations through distributed computing. It is intended to give research on PD in substations a certain theoretical foundation.

The innovation of this research lies in the following aspects: (1) Combining EEMD and SVE algorithms to enhance the accuracy of feature extraction. EEMD can effectively handle nonlinear and non-stationary signals, while SVE further improves the discrimination of features. (2) Integrating PCA algorithm, ISSA and SVM to improve the recognition accuracy of local discharge types. By optimizing the parameters of SVM, the performance of the model is further enhanced. (3) Introducing the Hadoop framework, which improves the real-time performance of local discharge monitoring in substations through distributed computing. The high fault-tolerance and high reliability of Hadoop enable it to effectively handle large-scale data and meet the monitoring requirements of substations in actual operation.

2. Methods and materials

2.1. Construction of EEMD singular value entropy feature extraction method for substation web data flow

The substation is an essential part of the power grid system. Due to its long-term complex operating environment, problems such as dust and air gaps often occur, causing various PD phenomena in the substation [19]. PD will not only lead to insulation aging and shortened equipment life, but may also develop into a comprehensive insulation breakdown problem, thus causing safety hazards [20]. However, due to the complexity of the operating environment, such as excessive noise, the current substation PDM method has problems of low feature extraction accuracy and low detection accuracy. Therefore, the Web data flow based on power grid monitoring is studied, and EEMD and SVE are combined to extract features of substation GIS PD signals. Subsequently, the PCA-ISSA-SVM algorithm is used to identify the PD type. Finally, a substation GIS PDM system based on EEMD-PCA-ISSA-SVM and Hadoop is constructed. Before building the system, it is necessary to understand the EEMD algorithm. EEMD is a commonly utilized signal decomposition technique in signal processing and other domains because of its excellent anti-interference capabilities and high stability [21]. The steps of signal decomposition by EEMD are shown in Figure 1 [22].

The signal's overall average processing time was initially set to M by the EEMD stages in Figure 1. Second, a new signal is created by adding white noise (WN) that follows the standard normal distribution to the original signal (OS). Equation (1) displays the new signal.

$$x_i(t) = x(t) + n_i(t) \tag{1}$$

In Equation (1), $x_i(t)$is the additional WN signal of the ith test. $n_i(t)$ is the ith additive WN sequence, $x(t)$ is the OS, and $i = 1,2, \dots, M$. Second, empirical mode decomposition (EMD)

is performed on the acquired signals containing WN to obtain the respective IMF sum forms. EMD is an adaptive technique for handling nonlinear and non-stationary signals that uses an iterative screening decomposition process to recover the signal's intrinsic IMF [23]. The EMD decomposition process is shown in Figure 2 [24].

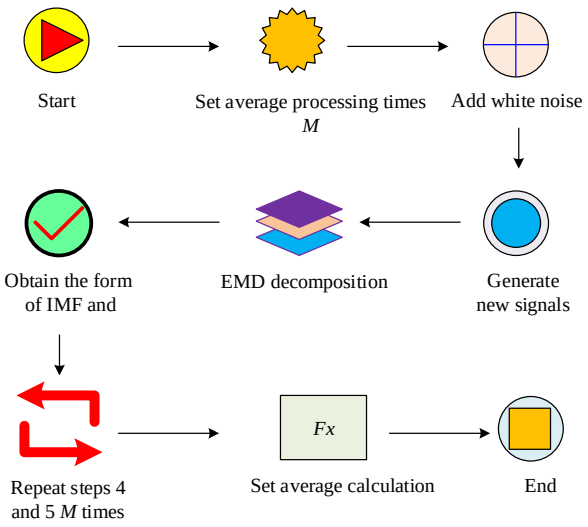


Figure 1. Signal decomposition steps of EEMD.

The EMD decomposition method first locates all of the original time series' maximum and minimum value points, as shown in Figure 2. Additionally, the cubic spline interpolation approach is used to fit the original data's upper and lower envelopes. Next, the upper and lower envelopes of the initial time series are averaged. Meanwhile, the average envelope is subtracted from the original data sequence to form a new data sequence. Finally, it is determined whether there is a local maximum or minimum in the new data sequence. If it exists, the above steps will repeated, otherwise determine whether the residual decomposition is a monotonic function. The result is output if the residual decomposes into a monotonic function; if not, it decomposes once more. Each IMF obtained through EMD decomposition represents an inherent component of the signal with different frequency and amplitude characteristics. IMF can be used to extract the signal's properties. The form of IMF sum is shown in Equation (2).

$$x_i(t) = \sum_{j=1}^{J} c_{i,j}(t) + r_i(t) \tag{2}$$

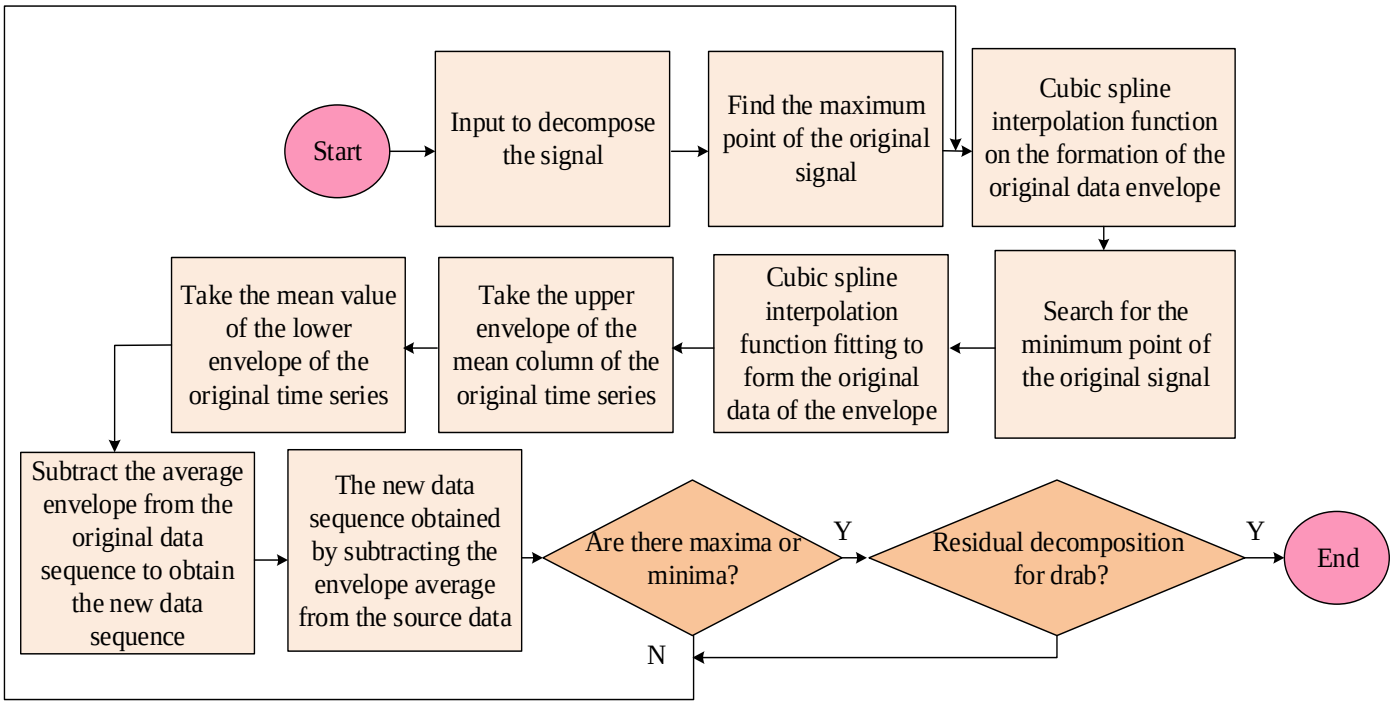


Figure 2. EMD decomposition process.

In Equation (2), $c_{i,j}$ displays the jth IMF obtained from the ith decomposition. $r_i(t)$ displays the residual function. J is the total quantity of IMF. Then, M iterations are performed until the residual no longer contains local maxima or minima. Moreover, Equation (3) illustrates how the IMF set is obtained by adding various WN signals to each decomposition.

$$[\{c_{1,j}(t), \{c_{i,j}(t)\}, \ldots, \{c_{M,j}(t)\}], j = 1,2, \ldots, J \tag{3}$$

Finally, the IMF component of EEMD can be obtained by averaging the IMFs in the set, as shown in Equation (4).

$$c_j(t) = \frac{1}{M}\sum_{i=1}^{M} c_{i,j}(t) \tag{4}$$

SVE is introduced to enhance the EEMD modal

component's accuracy and dependability. Since the multiple IMF components decomposed by EEMD contain signal information of different frequencies, the SVE of the signal can be calculated by performing singular value decomposition (SVD) on the matrix composed of each IMF. This can analyze the changes in the amount of signal information and highlight the signal characteristics more clearly. SVE is a method used to analyze signal characteristics. It is frequently utilized in domains like signal processing and data dimensionality reduction because of its great noise immunity and complicated system characteristic characterization capabilities [25]. The definition of SVE is to first decompose any $m \times n$-order matrix A through SVD, as shown in Equation (5) [26].

$$A = U\sum V^T \tag{5}$$

In Equation (5), U and V is an orthogonal matrix of order $m \times m$ and $n \times n$. $\sum$ is a singular value matrix, and $\sum = diag(\sigma_1, \sigma_2, \ldots, \sigma_n)$. Among them, σ_i is the singular value of A, and $\sigma_1 \geq \sigma_2 \geq \cdots \sigma_n$. Second, the square calculation of the singular values in $\sum$ can obtain σ_1^2, σ_2^2 to σ_n^2, and the sharing degree of each singular value on the matrix A can be obtained. Then, after normalizing each component σ_i^2 so that the sum of all normalized values is 1, the normalization condition of information entropy can be satisfied, as shown in Equation (6).

$$\sum_{i=1}^{n} E_i = 1 \tag{6}$$

In Equation (6), E_i is the normalized value, and $E_i = \frac{\sigma_i^2}{E}, E = \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2$. Finally, combining it with the definition of information entropy, the calculation expression of SVE can be obtained, as shown in Equation (7).

$$H = -\sum_{i=1}^{n} E_i ln\, E_i \tag{7}$$

In Equation (7), H is the singular entropy value. Therefore, based on the above, an EEMD SVE feature extraction method for substation Web data flow is constructed. Figure 3 illustrates the procedure.

In Figure 3, this method first obtains raw signals from the substation through sensors. The raw signal is uploaded to the database via the server. Moreover, it displays the data through the terminal's Web page. Second, the data is decomposed by EEMD to decompose the OS. After decomposition, several IMF components and a residual quantity will be obtained. The kurtosis, mean square error, and Euclidean distance are used to reconstruct the signal of the IMF component containing PD information to generate the original eigenvector eigenmatrix. Subsequently, the SVD algorithm is used to calculate the matrix. The singular values of this matrix are obtained when performing calculations. Finally, the EEMD SVE of the OS can be obtained through Equations (6) and (7).

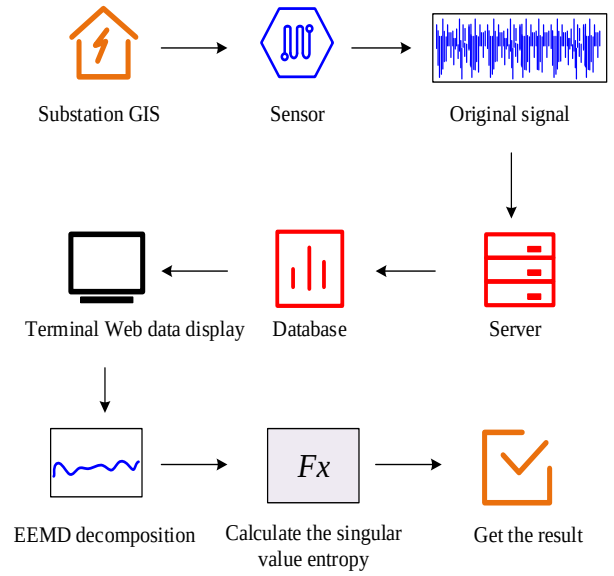


Figure 3. Feature extraction method of EEMD singular value entropy for GIS Web data stream.

2.2. Design of substation PDM system based on EEMD-PCA-ISSA-SVM and hadoop

To increase the effectiveness of monitoring and PD type detection, substation PD signals derived using EEMD and SVE must be dimensionally reduced since their dimensionality is too high. To this end, this study uses the PCA algorithm to perform dimensionality reduction processing. High computational efficiency and the preservation of the global variance structure are two benefits of the PCA method, a data dimensionality reduction technique. It is extensively utilized in domains like data dimensionality reduction and signal processing [27]. Figure 4 illustrates the PCA algorithm's operating steps [28].

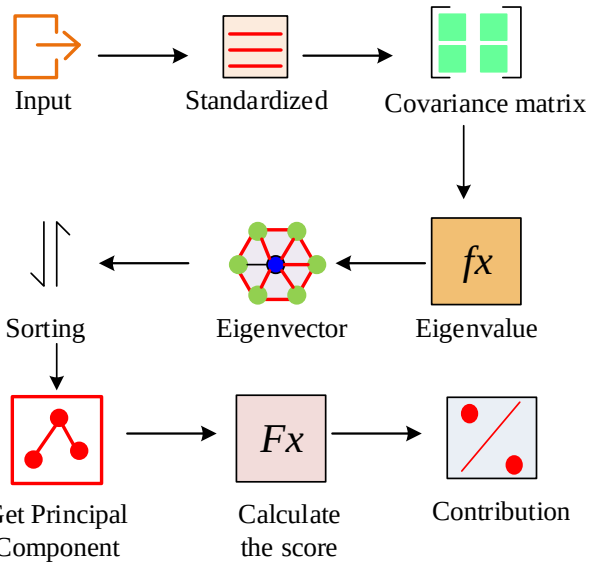


Figure 4. Operation steps of the PCA algorithm.

The PCA algorithm's operation phases in Figure 4 start by

standardizing the input sample data. Next, determine the sample matrix's covariance matrix (CM). Additionally, the CM's eigenvalues and eigenvectors are computed. After the calculation is completed, the eigenvalues are sorted. When sorting, sort from largest to smallest.The primary component is thus the eigenvector that corresponds to the top-ranked eigenvalue. To determine the principal component (PC) of the input data, the eigenvalue that ranks first will be considered the first PC, followed by the second PC, and so on. Then, a number of PCs not exceeding the total number of features are selected for computation. Their scores are evaluated and ranked in descending order. Finally, the cumulative contribution rate (CR) can be obtained by calculating and summing the CRs of the feature values. Among them, Equation (8) displays the ith eigenvalue's CR.

$$D_b = \frac{KT_b}{\sum_{\beta=1}^{l} KT_\beta} \tag{8}$$

In Equation (8), D_b represents the CR of the bth eigenvalue. KT is the eigenvalue of the CM. β displays the index of the sample, and l is the total number of features. After performing dimensionality reduction on the data, the study uses the SVM algorithm to identify types of PDs in substation GIS. SVM is a supervised learning technique that is frequently used in domains like classification and image recognition. It has benefits like a simple model and excellent generalization capacity [29]. The core idea of SVM is to find an optimal hyperplane and use this hyperplane to separate data points of different categories in order to maximize the margin between the two types of data. This margin is called the 'margin boundary', and the data points on the margin boundary are called 'support vectors'. The SVM hyperplane is shown in Figure 5 [30].

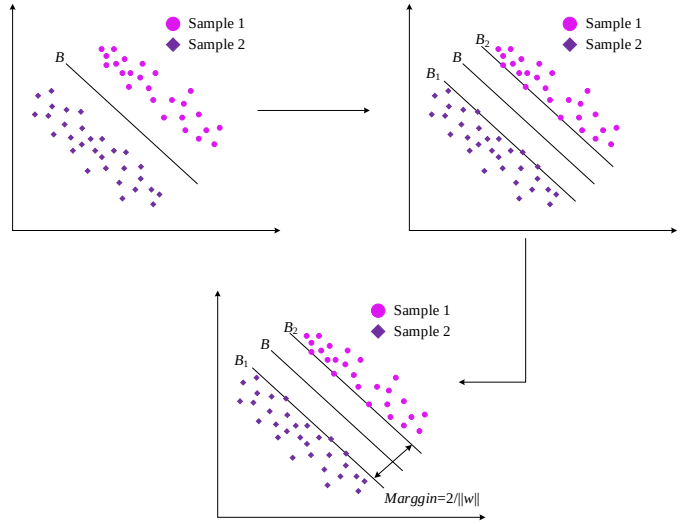


Figure 5. Schematic diagram of the hyperplane of SVM.

In Figure 5, B represents the optimal hyperplane. B_1 and B_2 are the planes closest to B. w is the normal vector of the hyperplane. $M\,arg\,g\,in$ is twice the distance from all support vectors to B. Among them, B_1 is denoted as $w \cdot a + \omega = +1$, and ω is the bias term of the hyperplane. a is the feature vector of the sample. B_2 is denoted as $w \cdot a + \omega = -1$. B is represented as $w \cdot a + \omega$. Equation (9) illustrates the need to build a nonlinear kernel function for classification prediction since the identification of GIS PD signal kinds is a nonlinear high-dimensional challenge.

$$g(x) = sgn\{\sum_{u=1}^{q} \alpha_{ur} y_u K(a_u, a) + c_r\} \tag{9}$$

In Equation (9), c_r is the bias factor. $K(a_u, a)$ is the nonlinear kernel function, which here is the radial basis function. $g(x)$ is the decision function. α_{ur} is the u th Lagrange multiplier. y_u and u display the sample target and index. q is the total quantity of samples. Among them, Equation (10) illustrates how b_{ur} is obtained by minimizing the objective function $L(\alpha)$.

$$minL(\alpha) = \frac{1}{2}\sum_{u,\lambda}^{q} \alpha_u \alpha_\lambda y_u y_\lambda K(a_u, a_\lambda) - \sum_{u=1}^{q} \alpha_u$$

$$s.t. 0 \le \alpha_u \le C \sum_{u=1}^{q} y_u \alpha_u = 0\ u = 1,2,\ldots,q \tag{10}$$

In Equation (10), C is the penalty factor (PF). y_λ is the target of the λth sample. α_u is the Lagrange multiplier. However, when training SVM, there is a difficulty in accurately selecting the PF and the radial basis width of the radial basis function. Choosing an appropriate radial basis width and PF can improve the recognition accuracy of the SVM algorithm. Therefore, the paper presents the sparrow search algorithm (SSA). SSA is a swarm intelligence optimization technique that is frequently used in domains like engineering design optimization and machine learning parameter tuning. Its benefits include a powerful global search capability and a straightforward structure. When performing the search, the SSA first initializes parameters such as the weights of sparrow discoverers (SDs), joiners, and sentinels, as well as the maximum quantity of iterations. Next, it updates the positions of sparrows with three different identities. Among them, the position update of the SD is shown in Equation (11).

$$Y_{\gamma,\eta}^{t+1} = \begin{cases} Y_{\gamma,\eta}^{t} * exp(\frac{-\gamma}{\chi * iter_{max}}), R_2 < ST \\ Y_{\gamma,\eta}^{t} + Q \cdot L, R_2 \ge ST \end{cases} \tag{11}$$

In Equation (11), $Y_{\gamma,\eta}^{t+1}$ and $Y_{\gamma,\eta}^{t}$ represent the positions of the γth discoverer in the ηth dimension at iteration counts $t+1$ and t, respectively. Q and L are random quantities and matrices that follow a normal distribution. R_2 is an early warning value. ST is the safety threshold. $iter_{max}$ is the maximum quantity of iterations. χ is a random quantity in the range of 0 to 1. Equation (12) displays the joiner's position update.

$$Y_{\gamma,\eta}^{t+1} = \begin{cases} Q * exp(\frac{Y_{worst}^{t} - Y_{\gamma,\eta}^{t}}{t^2}), \gamma > \frac{\varphi}{2} \\ Y_p^t + \left|Y_{\gamma,\eta}^{t} - Y_p^t\right| \cdot D^+ \cdot L, \gamma \le \frac{\varphi}{2} \end{cases} \quad (12)$$

In Equation (12), Y_{worst}^{t} represents the worst position in the population. D is a matrix. Y_p^t is the optimal position found by the discoverer, and $\frac{\varphi}{2}$ is the threshold. The position update for the early warning agent is shown in Equation (13).

$$Y_{\gamma,\eta}^{t+1} = \begin{cases} Y_{best}^{t} + \psi \cdot \left|Y_{\gamma,\eta}^{t} - Y_{best}^{t}\right|, f_\gamma \neq f_g \\ Y_{\gamma,\eta}^{t} + U \cdot \left(\frac{\left|Y_{\gamma,\eta}^{t} - Y_{best}^{t}\right|}{(f_\gamma - f_w) + \varepsilon}\right), f_\gamma \neq f_g \end{cases} \quad (13)$$

In Equation (13), f_g and f_g represent the optimal fitness values (FVs) in the entire region and the best FV of the current early warning individual, respectively. U is a random number for directional movement. Y_{best}^{t} is the optimal position in the entire region. ε is an infinitesimal number. ψ is the step size parameter. f_w is the worst FV in the entire region. Then, after calculating the individual FV of each sparrow, the optimal FV and parameter values are updated. Finally, it is checked whether the maximum iteration has been reached. If so, the result is output and the computation ends. Otherwise, the process returns to the second step to continue iteration until the condition is met. The study uses an enhanced technique based on a cosine function for updating the proportion of discoverer persons in an attempt to balance the SSA's global and local search capabilities, as displayed in Equation (14).

$$Per\_iter(t) = \mu cos\left(\frac{\pi(t+h)}{2iter_{max}}\right) + v \quad (14)$$

In Equation (14), Per_iter displays the proportion of discoverers, and μ displays the amplitude of the proportion. h and v are the horizontal and vertical translation factors, respectively. Equation (14) addresses the issue of a fixed proportion of discoverers in the SSA, balancing local and global search capabilities. The cosine function reduces the proportion of discoverers in a "slow steep slow" rhythm. In the early stage, the decline was slow, and more exploration individuals were retained to maintain the global search. Steep drop in the medium term and rapid strengthening of local mining. In the later stage, the algorithm is leveled again to prevent stagnation. This shape naturally staggers the exploration and production peaks of SSA without adding parameters, so as to accelerate convergence and improve accuracy. Equation (15) illustrates how the study optimizes the SSA using the somersault foraging technique of manta ray foraging optimization (MRFO) to keep it from entering local optima.

$$Z_{idis}(t) = Z_\gamma(t) + W(r_1 Z_{best}(t) - r_2 Z_{best}(t)) \quad (15)$$

In Equation (15), $Z_{idis}(t)$ represents the disturbance position, and W is the rolling factor. r_1 and r_2 are random quantities within the interval of 0 to 1. $Z_{best}(t)$ is the optimal disturbance position. According to Equation (15), during the search process of the SSA, as the number of iterations increases, it continuously approaches the current optimal individual, thereby reducing disturbances and enhancing local search capability. Thus, the ISS algorithm is constructed. By combining it with the EEMD singular value algorithm, PCA algorithm, and SVM, a HA aimed at local discharge identification in substation Web data streams is developed, namely the EEMD-PCA-ISSA-SVM algorithm. The computational steps of this algorithm are shown in Figure 6.

The algorithm's computing phases are shown in Figure 6 as follows: First, the sensor acquires the OS from the substation and transmits it to the Web data page of the monitoring system. Second, the EEMD method reduces noise interference and highlights signal qualities by breaking down the OS into a number of IMF components. Then, the SVE algorithm is used to calculate the SVE of the signal, further extracting the feature vectors of the signal. EEMD first adaptively decomposes the discharge signal into several IMFs. For each IMF, its relative energy is computed. The identical IMF is then converted into a Hankel matrix, and the top-ranked singular values are used to calculate Shannon entropy. All entropy values are concatenated into a vector and appended to the energy feature, yielding the final input vector. This two-stage vector is subsequently fed into PCA for dimensionality reduction, after which ISSA-tuned SVM carries out the discharge-type classification, thus improving overall recognition performance. Finally, based on the feature vectors, the ISSA-optimized SVM algorithm classifies and determines the kind of PD and outputs the

recognition results. The computational steps of the ISSA-optimized SVM algorithm first initialize the population size and parameters of the ISSA. Meanwhile, the SVM algorithm's parameters such as the PF and the radial basis width are initialized. Next, the fitness of the individuals in the ISSA's population is calculated. Then, the proportion of discoverers in the ISS algorithm using Equation (14) is updated. After the update is completed, update the entire sparrow population and calculate the individual fitness using Equations (11), (12), and (13). Subsequently, the disturbed positions of each individual using Equation (15) are calculated and then the fitness of the corresponding disturbed individuals is computed. After completing the calculation, the optimal solution for this iteration is computed, and the optimal individual is updated. Then, it is determined whether the maximum iteration count has been reached. If so, the optimal PF and radial basis width are output, thereby obtaining the optimal SVM algorithm. Otherwise, iteration continues until the condition is satisfied. Finally, to achieve RTM of PD in substations, a monitoring system is designed based on Hadoop and the proposed EEMD-PCA-ISSA-SVM. The system is shown in Figure 7.

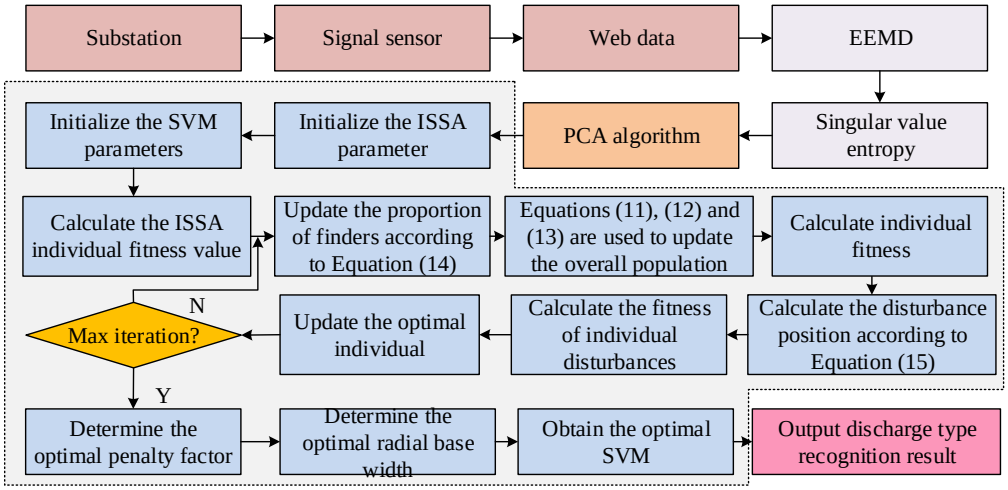


Figure 6. Substation Web data streams hybrid algorithm of recognition of PD.

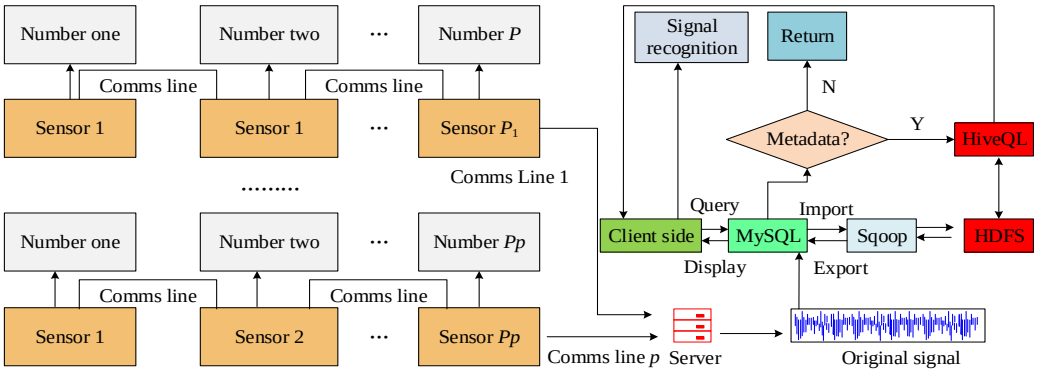


Figure 7. RTM system for PD in substations based on Hadoop and the proposed EEMD-PCA-ISSA-SVM.

In Figure 7, P and p represent the monitoring point numbers and the corresponding sensor numbers. The system monitors the points by arranging sensors in a grouped cascade manner. During transmission, each group of communication lines connects individually to the server, which avoids the complexity of the wiring and reduces signal transmission interference. In this transmission, the sensor communication module is a UT-482 embedded isolated TTL to RS-422 transceiver. The server communication line port is an RS-232/RS-422 UT-202 commercial-grade compact interface converter. After the server obtains the raw signal data, it stores the data in a Structured Query Language (MySQL) relational database. The Sqoop program is then used to import the data into the Hadoop Distributed File System (HDFS). With its enormous storage capacity and great fault tolerance, HDFS is a fundamental part of the Hadoop ecosystem and is extensively utilized in Hadoop

distributed computing. By using HDFS to store data across multiple nodes, the system's data throughput can be improved. Hive Query Language (HiveQL) is used to query commands issued by users. HiveQL queries the command requirements from the data stored in HDFS. If the required data is found, it is returned to the client, otherwise it waits for the next data. When the user wants to check the PD status of the current substation, the signal recognition module identifies and provides feedback on the current PD results based on the client's web data stream. To exploit distributed computing, the entire EEMD-SVE→PCA →ISSA-SVM pipeline is packaged as a Hadoop-Streaming job: each 64 MB HDFS block is processed by a Mapper that extracts the energy-entropy feature vector; the Reducer performs model training and outputs the discharge-type label, which is finally written back to HDFS and exposed to the client through HiveQL. Thus, both data storage and algorithm execution are handled within the Hadoop ecosystem.

3. Results and analysis

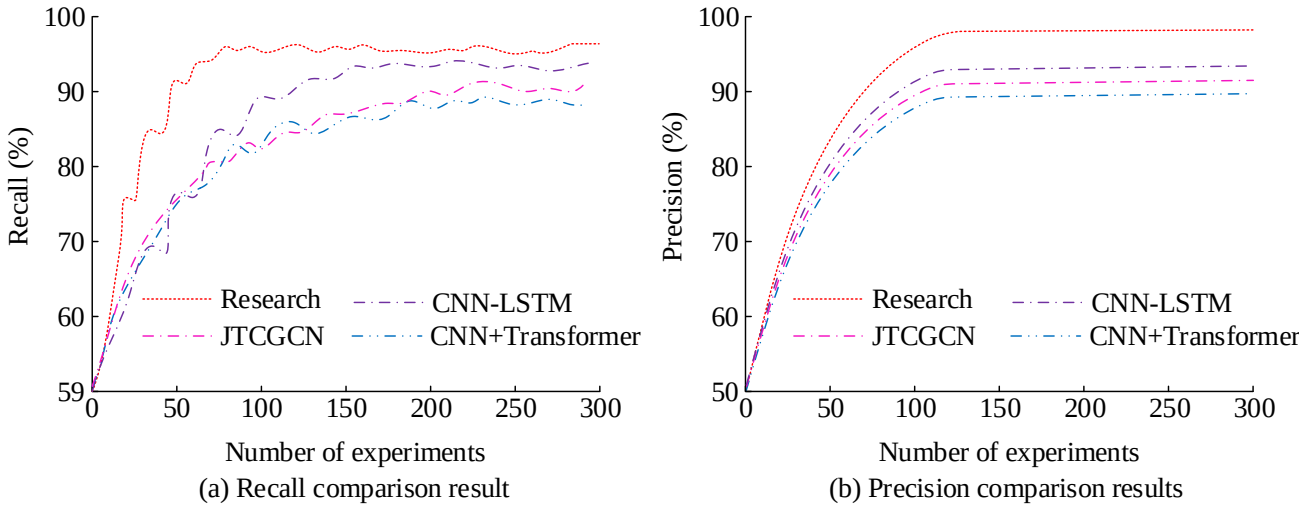
3.1. Performance analysis of hybrid algorithms

After constructing the proposed HA, the study conducts a comparative analysis of its performance. The algorithms used for comparison includes CNN-LSTM, CNN Transformer, and joint temporal and channel graph convolutional network (JTCGCN). Experimental parameter settings: The penalty parameter range for the SVM algorithm is [0.1, 100], and the radial basis width range for the radial basis function is [0.01, 1000]. The maximum iteration for the algorithm is 300. In the SSA, the proportion of explorers is 20%. The ISSA has rolling factors of 0.76, 50, 0.2, and 4, and μ ranges from 0.5 to 0.9. The optimizer used is the SGD optimizer, with a weight decay of 0.01 and an initial learning rate of 1×10-4[31,32]. Parameter tuning uses five-fold cross-validation and employs an early stopping mechanism to prevent overfitting. The data are sourced from the WHTPD dataset and the PDMED dataset, with a total of 1,479 samples randomly selected. Among them, 70% are used as the training set, 20% as the test set, and 10% as the validation set. The selected sample categories include normal conditions, corona discharge, air gap discharge, and surface discharge. The four types of samples are evenly distributed. The WHTPD dataset was collected by the University of Strathclyde and covers common phenomena such as corona discharge, air gap discharge, and surface discharge in complex environments. The dataset is balanced for PD phenomena. Each category contains the same number of data points, totaling 16,000 PRPD patterns, with 600 data points for each pattern. The PDMED dataset was collected under ideal conditions and includes 5,300 images of PRPD patterns in four ideal environments. The data calibration standard is the IEC standard, and the calibration interval is 6 months. During the data preprocessing stage, first, a high-pass filter is employed to reduce noise in the collected partial discharge signals, thereby eliminating low-frequency interference components. subsequently, the maximum-minimum normalization method is used to standardize the signal data, scaling the data to the range of [0,1], in order to eliminate the influence caused by different feature scales and magnitudes, thereby improving the efficiency and effectiveness of the subsequent algorithm processing. Additionally, a low-pass filter is introduced to remove high-frequency noise, a band-pass filter is used to isolate the target signal frequency band, and an adaptive filter is applied to dynamically adjust parameters based on the noise characteristics, further optimizing the noise processing effect. The comparison metrics include precision, recall, F1 score (F1), and classification accuracy. Table 1 depicts the experimental setting.

The study first tests each algorithm's recall and precision in the aforementioned setting. The experimental results are shown in Figure 8.

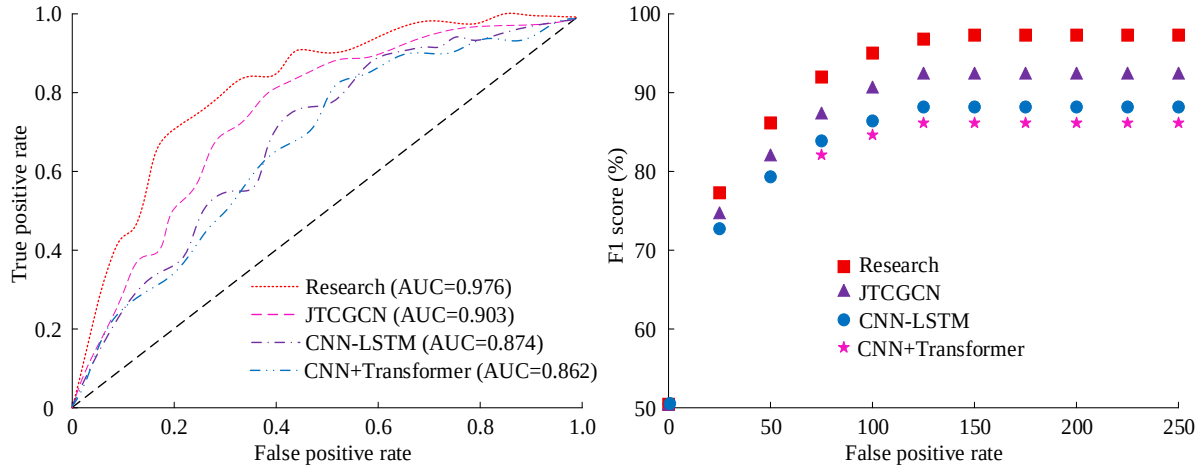
Table 1. Experimental environment.

Parameter names	Parameter
Data analysis software	Spss 26.0
Main frequency	6.0GHz
Random access memory	32GB
Hard disk capacity	500GB
Central processing unit	Intel Core i9-14900K
Matlab version	Matlab 2021b
Operating system	Windows 10 64



(a) Recall comparison result (b) Precision comparison results

Figure 8. Comparison results of precision and recall.

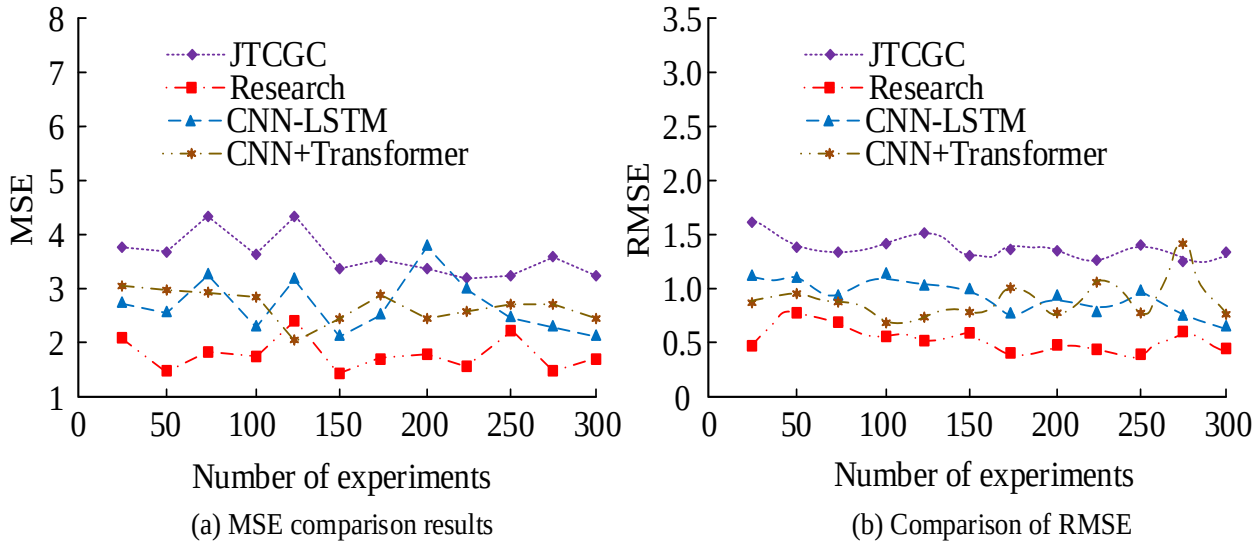


(a) ROC curve comparison results (b) F1 score comparison results

Figure 9. Comparison results of F1s and ROC curves.

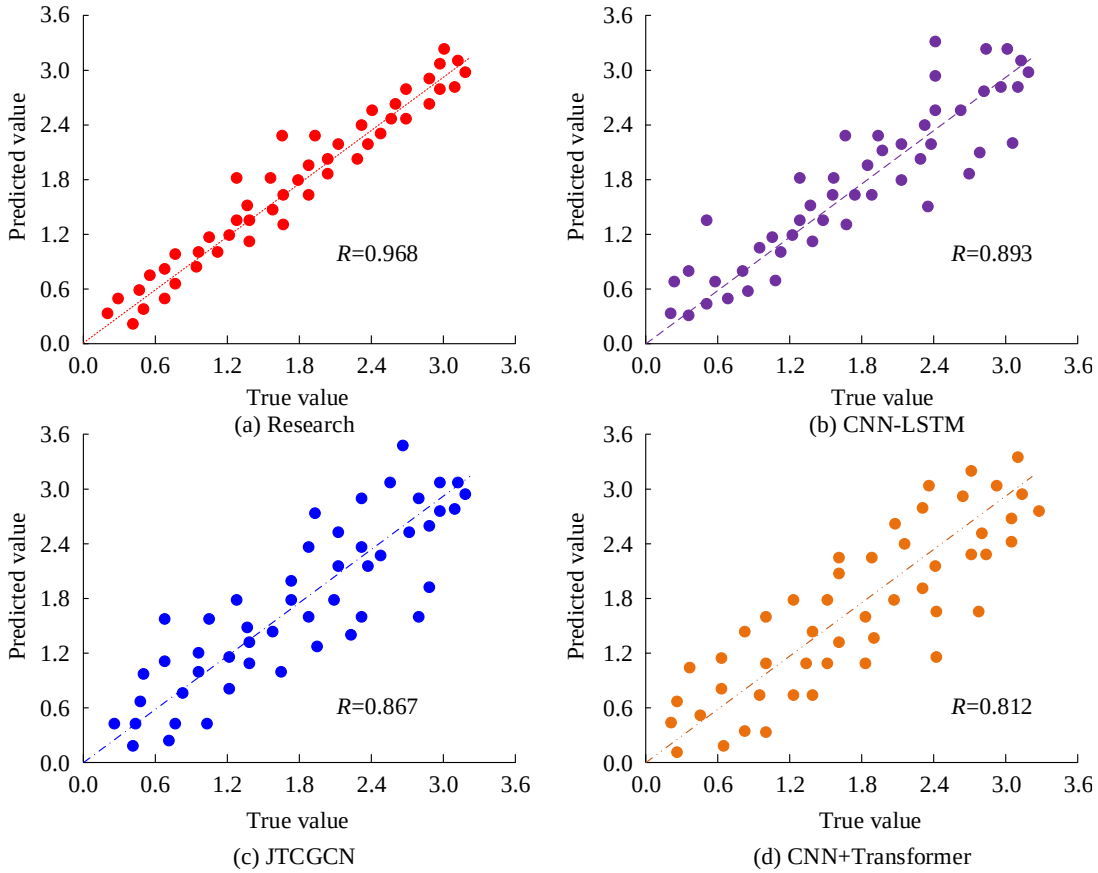
In Figure 8(a), the recall rate of the proposed HA in this study is 97.76%, significantly higher than CNN-LSTM's 91.24%, JTCGCN's 89.98%, and CNN Transformer's 87.64%. In Figure 8(b), the precision of the proposed HA, CNN-LSTM, JTCGCN, and CNN Transformer are 98.67%, 90.92%, 89.26%, and 88.78%. The hybrid method suggested in this study has the highest precision among them. A better recall means that a greater percentage of real positive samples are successfully identified by the algorithm. A greater percentage of samples that are anticipated to be positive are actually positive when precision is higher. The aforementioned findings show that the HA suggested in this study outperforms the contrast methods. The comparison results of the F1s and ROC curves of the various algorithms are shown in Figure 9.

In Figure 9(a), the area under the ROC curve (AUC) of the proposed HA, CNN-LSTM, JTCGCN, and CNN Transformer ROC curve is 0.976, 0.874, 0.903, and 0.862. Among them, the HA proposed in this study has the highest AUC. In Figure 9(b), the F1s of the proposed HA, CNN-LSTM, JTCGCN, and CNN Transformer are 97.88%, 88.92%, 87.67%, and 90.76%. The suggested hybrid method outperforms the others in terms of F1. AUC is a metric that assesses how well a model can differentiate between positive and negative classifications. The algorithm performs better in classification when the AUC value is closer to 1. A higher F1 indicates better recognition and classification performance of the algorithm. In summary, the suggested HA outperforms the comparable algorithms in terms of the ROC curve and F1. Figure 10 displays the mean square error (MSE) and root mean square error (RMSE) comparative findings for each algorithm.



(a) MSE comparison results (b) Comparison of RMSE

Figure 10. Comparison results of MSE and RMSE.



(a) Research (b) CNN-LSTM

(c) JTCGCN (d) CNN+Transformer

Figure 11. Fitting comparison results of the algorithm.

The HA suggested in this work, CNN-LSTM, JTCGCN, and CNN Transformer, have average MSE values of 1.537, 2.641, 3.978, and 2.996, respectively, as shown in Figure 10(a). The hybrid method that this study suggests has the lowest average MSE value among them. In Figure 10(b), the average RMSE value of the HA proposed in this study is 0.462, which is lower than CNN-LSTM's 1.279, JTCGCN's 1.532, and CNN Transformer's 0.876. The lower the MSE and RMSE of an algorithm, the closer its recognition results are to the actual recognition results. In summary, the hybrid method suggested in this study performs better than the comparable algorithms in terms of MSE and RMSE values. The recognition fitting results of each algorithm are shown in Figure 11.

In Figure 11(a), the scatter points of the proposed HA are the

most concentrated, with a fitting coefficient R of 0.968. This is significantly higher than the CNN-LSTM algorithm in Figure 11(b) at 0.893, the JTCGCN algorithm in Figure 11(c) at 0.867, and the CNN Transformer algorithm in Figure 11(d) at 0.812. A higher fitting coefficient indicates that the algorithm has greater accuracy and reliability in classifying and identifying local discharges in substations. In summary, compared with the other algorithms, the proposed HA outperforms them in terms of fitting coefficient. In order to further verify the universality of the proposed HA, the study classified and identified the local discharge conditions caused by different insulating materials. The classification and identification results are shown in Table 2.

Table 2. Classification and Identification Results of Partial Discharge Caused by Different Local Discharge Sources.

Type of local discharge source	Indicator	Research	CNN-LSTM	JTCGCN	CNN+Transformer
Solid insulation	Accuracy	98.03%	90.14%	87.85%	86.76%
	Recall	97.26%	89.88%	87.88%	85.67%
Liquid insulation	Accuracy	97.14%	91.04%	89.97%	87.17%
	Recall	98.38%	88.96%	88.04%	85.65%
Gas insulation	Accuracy	97.68%	92.02%	86.43%	86.68%
	Recall	98.32%	87.98%	86.45%	84.86%

As can be seen from Table 2, the classification recognition accuracy and recall rate of the proposed HA in local discharges caused by solid insulators, liquid insulators and gas insulators are all above 97%, significantly higher than those of the comparison algorithm. This result indicates that the proposed HA in the study has good universality and can identify local discharges caused by various insulating sources. To further verify the impact of EEMD-SVE and PCA-ISSA-SVM on the performance of the HA, ablation experiments were conducted. The experimental results are shown in Table 3.

Table 3. Results of the ablation experiment.

Algorithm	F1	Rcall rate
EEMD-SVE	74.27%	75.32%
PCA-ISSA-SVM	86.89%	85.14%
EEMD-SVE-PCA-ISSA-SVM	97.88%	97.76%

As can be seen from Table 3, after combining EEMD-SVE and PCA-ISSA-SVM, the F1 score and recall rate increased from 74.27% and 75.32% to 97.88% and 97.76% respectively. This result indicates that the combination of the two significantly improves the classification and recognition accuracy of local discharge in substations. In order to verify the performance of the partial discharge method proposed in the study, the performance of the method is compared with that of the Hilbert Huang transform or multi sensor fusion for ultra sonic/uhf signals. the comparison results are shown in Table 4.

It can be seen from Table 4 that when using the HA proposed in the study, the precision rate and recall rate were 98.67% and 97.76%, respectively, which were much higher than those of the Hilbert Huang transform method (92.17% and 93.23%), and the multi sensor fusion for ultra sonic/uhf signals method (94.26% and 93.87%). The above results show that the performance of the proposed method is better than that of the comparison method in terms of precision rate and recall rate. In order to verify the robustness of the proposed method in different noise environments, the low-frequency noise (LFN), high-frequency noise (HFN) and compound noise (CN) noise environments were selected for performance evaluation, and the fault prediction accuracy was used as the index to verify the external effectiveness. The analysis results are shown in Table 5.

Table 4. Comparison results.

Method	Precision	Rcall rate
Rsearch	98.67%	97.76%
Hilbert-Huang transform	92.17%	93.23%
multi-sensor fusion for ultrasonic/UHF signals	94.26%	93.87%

As can be seen from Table 5, the accuracy and recall rate of the proposed HA in len, HFN and CN environments are all above 92%, much higher than CNN transformer, jtcgcn and cnn-lstm. And the accuracy of the proposed HA in three kinds of noise environment is more than 90%, which is much higher than the comparison method. The results show that the proposed ha algorithm has good environmental adaptability and external effectiveness. In order to analyze the differences of the proposed HA in the characteristics of different types of partial discharge signals, 200 samples of each of the four types of discharge were selected for classification and recognition, and the results are shown in Table 6.

Table 5. Analysis results under different noise environments.

Algorithm	Noise type	Accuracy	Recall	Fault prediction accuracy
CNN-LSTM	LFN	88.51%	87.72%	85.63%
	HFN	86.31%	85.53%	83.45%
	CN	84.02%	83.21%	81.32%
JTCGCN	LFN	91.21%	90.42%	88.74%
	HFN	89.01%	88.22%	86.56%
	CN	86.81%	86.01%	84.37%
CNN Transformer	LFN	90.41%	89.62%	87.89%
	HFN	88.21%	87.42%	85.68%
	CN	85.91%	85.12%	83.49%
Research	LFN	94.71%	93.92%	92.15%
	HFN	93.52%	92.71%	90.98%
	CN	92.31%	93.53%	94.76%

Table 6. Classification and identification results of different discharge types.

True \ predicted	Normal	Corona	Air gap	Surface	Recall (%)
Normal	196	2	1	1	98.38
Corona	1	194	3	2	97.47
Air gap	1	2	196	1	98.25
Surface	2	6	2	190	95.36

It can be seen from Table 6 that the recall rate of the proposed HA on the four types of balanced samples showed an obvious gradient: the normal state was 98.38%, and only 3 cases were wrongly divided into corona or air gap. The fundamental reason was that the background noise occasionally produced sporadic pulses, and its phase just fell in the corona concentration area, causing miscarriage of justice. The recall rate of corona discharge was 97.47%. The core problem was that the pulse amplitude dropped sharply during high humidity period, and the statistical characteristics almost overlapped with the sparse small pulse of air gap discharge, resulting in 3 cases being wrongly labeled as air gap. The air gap discharge achieved 98.25%, and 2 cases were still misclassified as corona. The reason was that the pulse interval of small air gap discharge was lengthened and the phase distribution became wider, which was highly similar to the "wide phase tail" of corona. The surface discharge was the lowest (95.36%), and 6 of the 9 false detections were classified as corona. The fundamental reason was that its pulse "tail length" and low-frequency component almost coincided with the envelope shape of corona in the noise environment below 10 dB, which made it difficult to distinguish the boundaries of the network.

3.2. Analysis of the application effectiveness of the monitoring system

After performing a performance analysis of the proposed HA, the study conducts an analysis of the application effects of a substation online monitoring system based on this algorithm when applied to a certain substation. The transformer signal recognition results are shown in Figure 12.

In Figure 12, the proposed system is able to effectively identify substation PD signals based on Web data and can accurately recognize the corona discharge type. This outcome suggests that the suggested substation PDM system has good usability and practical application. To confirm the suggested substation PDM system's performance, the study conducts a comparative performance analysis with substation PDM systems based on CNN-LSTM, JTCGCN, and CNN Transformer. The data collection is carried out using RS-422 sensors. The sensor model is RS-422, with a sampling rate of 5kHz, a range of 0-500 units, an accuracy of ±0.2%, and a resolution of 0.02 units. During the collection process, these sensors are placed at the monitoring points of the transformer and transmitted to the system via a wireless network. After preprocessing, the data is classified and identified using algorithms. The recognition results of each system for normal conditions (A), corona discharge (B), air gap discharge (C), and surface discharge (D) are displayed in Figure 13.

In Figure 13(a), the study shows that the suggested monitoring system achieves identification accuracy of 98.6%, 97.3%, 96.9%, and 97.8% for normal conditions, corona discharge, air-gap discharge, and surface discharge, respectively. These results are significantly higher than those of Figure 13(b) CNN-LSTM at 88.6%, 87.4%, 87.8%, and 84.2%, Figure 13(c) JTCGCN at 89.3%, 82.4%, 89.7%, and 84.9%, and Figure 13(d) CNN Transformer at 90.3%, 89.7%, 86.5%, and 87.7%. In

summary, the substation PDM system proposed in this study can effectively identify various types of discharges, and its accuracy is higher than that of the comparison systems, demonstrating its effectiveness. The comparison results of recognition time, CPU usage, throughput, error rate, response time, and space complexity of each system are illustrated in Table 7.

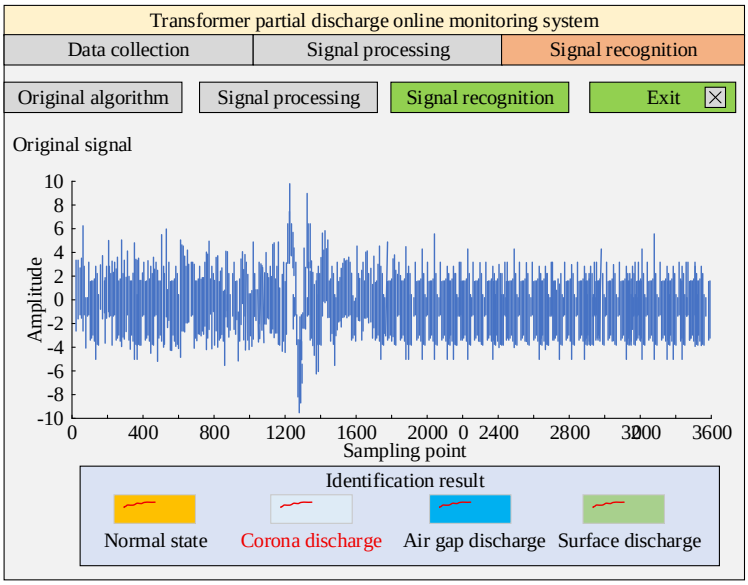


Figure 12. System recognition result.

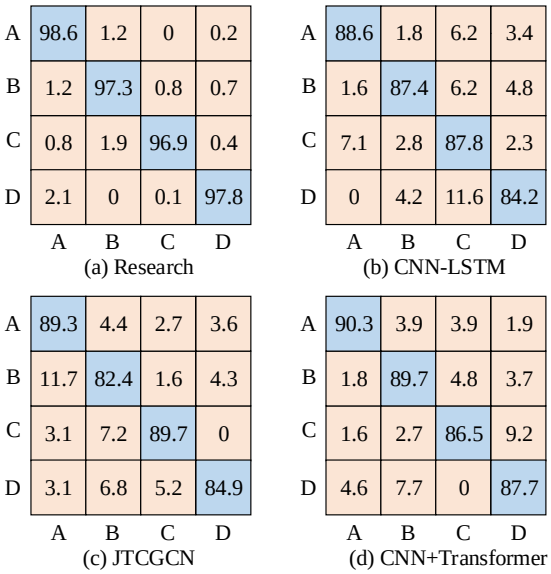


Figure 13. Each system the identification results of PD type.

In Table 7, the proposed monitoring system in the study has a recognition time, CPU usage, response time, and space complexity of 0.46s, 42.67%, 154ms, and O(n), respectively, all lower than those of CNN-LSTM, JTCGCN, and CNN Transformer. Additionally, the error rate and throughput are 0.14% and 1132 req/s, respectively, both better than the comparison systems. In summary, from multiple aspects, the substation PDM system proposed in this study outperforms the comparison systems. From various perspectives, the proposed substation partial discharge monitoring system outperforms the comparison system in terms of performance. This indicates that the system has significant advantages in terms of scalability at the edge deployment cost and under high load conditions. For instance, the lower CPU usage and space complexity mean that the system has better adaptability on resource-constrained edge devices, effectively reducing hardware costs and resource consumption; while the higher throughput and lower error rate indicate that the system can maintain high performance and reliability when handling large amounts of data and high concurrent requests, possessing good scalability and being able to meet the local partial discharge monitoring requirements of substations in actual operation.

Table 7. Experimental results.

Indicator	Research	CNN-LSTM	JTCGCN	CNN+Transformer
Recognition time	0.46s	1.21s	1.32s	0.98s
CPU usage rate	42.67%	56.32%	48.26%	60.03%
Expert satisfaction	1132 req/s	769 req/s	856 req/s	967 req/s
Error rate	0.14%	0.21%	0.34%	0.18%
Response time	154ms	198ms	256ms	342ms
Space complexity	$O(n)$	$O(n^2)$	$O(n)$	$O(n \log n)$

4. Discussion

The suggested HA's performance was compared, and the suggested monitoring system's application effect was examined. The experimental outcomes showed that the suggested HA demonstrated good performance in terms of recall rate, precision, and AUC, all outperforming the comparison algorithms. In the recall rate comparison experiment, the recall rates of the proposed HA, CNN-LSTM, JTCGCN, and CNN Transformer were 97.76%, 91.24%, 89.98%, and 87.64%, with the proposed algorithm achieving the highest recall rate. The results indicated that the introduction of EEMD and SVE enhanced the feature extraction capability, thereby improving the recognition ability of the algorithm. This result was similar to the related findings of W. Zou and others [33]. In the precision comparison experiments, the precision of the proposed HA, CNN-LSTM, JTCGCN, and CNN Transformer were 98.67%, 90.92%, 89.26%, and 88.78%. The suggested HA outperformed the others in terms of precision. The results indicated that the introduction of the PCA algorithm further revealed the signal features, improving the accuracy of the algorithm. This result was consistent with the findings of O. Muñoz et al. in related studies [34]. In the AUC comparison experiment, the AUC values of the proposed HA, CNN-LSTM, JTCGCN, and CNN Transformer were 0.976, 0.874, 0.903, and 0.862. The suggested HA outperformed the others in terms of AUC. This outcome showed that the ISSA's adoption improved classification performance by assisting the algorithm in determining the ideal parameters. This conclusion was in line with research done in 2024 by T. Li et al [35]. In addition, the HA proposed in the study achieved an F1 of 97.88%, an average MSE of 1.537, an average RMSE of 0.462, and a coefficient of determination R of 0.968. These outcomes provided more evidence of the suggested algorithm's improved performance.

Subsequently, the proposed substation PDM system was analyzed for its application effectiveness. In practical use, the system was found to accurately display signals and provide identification results, demonstrating practical value. In the comparative analysis of application effectiveness, the study showed that the monitoring system had identification accuracies of 98.6%, 97.3%, 96.9%, and 97.8% for normal conditions, corona discharge, air gap discharge, and surface discharge, respectively. This is significantly higher than CNN-LSTM, which achieved 88.6%, 87.4%, 87.8%, and 84.2%, and CNN Transformer, which achieved 90.3%, 89.7%, 86.5%, and 87.7%. The results indicated that the combination of EEMD, SVE, PCA algorithm, ISSA, and SVM algorithm enhanced the accuracy of identifying different types of PDs and is effective. These results were consistent with the related findings of Y. Wu et al. in 2023 [36]. The proposed system study indicated that the monitoring system had a recognition time, CPU usage, response time, and space complexity of 0.46s, 42.67%, 154ms, and O(n), respectively, all of which were lower than those of CNN-LSTM, JTCGCN, and CNN Transformer. Moreover, the error rate and throughput were 0.14% and 1132 req/s, respectively, both outperforming the comparison systems. The findings indicated that the system's performance was improved with the implementation of Hadoop distributed computing. In conclusion, the suggested system has good practical usefulness, and the suggested HA is efficient and performs better than the comparative algorithms. The limitation of this study is that, in real-world environments, substations not only face PD issues but also defects in equipment such as flaws and rust. Combining object detection algorithms, such as the YOLO series, with computer vision technology for real-time detection of these issues is a direction for further research. In addition, this research mainly focuses on the local discharge problem. It has not fully considered other important non-PD issues in the substation, such as hardware failures and rust. These non-PD issues also affect the operation safety and reliability of the substation, limiting the wide application of the research results. Future research plans to implement a YOLO-based visual system. It will collect a dataset of substation images containing hardware failures, rust, and normal conditions through visual sensors. And use this dataset to train the YOLO model to detect and classify faults and states. The trained model will be integrated into the existing monitoring system to achieve real-time detection and alarm functions.

5. Summary

To address the issues of low accuracy in signal feature extraction and low precision in PD type recognition in current substation web-based RTM systems, this study introduced the EEMD algorithm for signal feature extraction. To enhance the

decomposition capability of the EEMD algorithm, the SVE algorithm was used for improvement. Subsequently, after dimensionality reduction of the data using the PCA algorithm, the SVM algorithm was employed to classify and identify the types of PD. Meanwhile, an improved ISSA was introduced to optimize the parameters of the SVM algorithm, constructing an EEMD-PCA-ISSA-SVM HA. Finally, based on the HA and Hadoop, a RTM system for substation PDs was proposed. The suggested HA's performance was contrasted with that of alternative algorithms. The findings demonstrated that in terms of recall, precision, F1, and MSE values, the suggested HA performed better than the compared algorithms. Subsequently, the proposed substation PDM system was analyzed for performance in comparison with other systems. The findings suggest that four different types of PDs may be reliably monitored and identified by the suggested substation PDM system, with recognition accuracy surpassing that of the comparator systems. It can also accurately display the OSs and recognition results on the interface. In addition, the proposed substation PD system excels the comparison systems in recognition time, CPU usage, response time, and spatial complexity. In summary, the HA and monitoring system proposed in this study demonstrate superior performance and practical value.

Fundings

The research is supported by: State Grid Qinghai Electric Power Company Science and Technology Project, Research on Rapid Detection and Location Technology for Equipment Discharge Based on Stereo Space in Substations, No.: 777234318.

References

1. Hussain G A, Hassan W, Mahmood F, Shafiq M, Rehman H, Kay J A. Review on partial discharge diagnostic techniques for high voltage equipment in power systems. IEEE Access 2023;11(1): 51382–51394. DOI: 10.1109/access.2023.3279355.
2. Gheisari M, Hamidpour H, Liu Y, Saedi P, Raza A, Jalili A, Amin R. Data mining techniques for web mining: a survey. Artificial Intelligence and Applications 2022;1(1): 3–10. DOI: 10.47852/bonviewAIA2202290.
3. Song D, Zhao L, Gu H, Zhong Y, Zhang Z, Wang H, Zheng Q. Design of intelligent sensor real-time monitoring system for partial discharge in switchgear. International Conference on Mechatronic Engineering and Artificial Intelligence (MEAI 2024) SPIE 2025; 13555(1): 267–273. DOI: 10.1117/12.3064803.
4. Yang R, Li J, Zhu T, Yang W, Hu D, Dong E. An automatic robot for ultrasonic partial discharge detection of gas-insulated switchgear. Industrial Robot: the international journal of robotics research and application 2024; 51(6): 908–921. DOI: 10.1108/ir-01-2024-0005.
5. Wang R, Yin B, Yuan L, Wang S, Ding C, Lv X. Partial discharge positioning method in air-insulated substation with vehicle-mounted UHF sensor array based on RSSI and regularization. IEEE Sensors Journal 2024; 24(11): 18267–18278. DOI: 10.1109/jsen.2024.3387665.
6. Wang Y, Kong L, Cao L, Li Y, Gao X, Chen Y. Partial discharge fault diagnosis in substations using a combined logic and GAMF-CNN network model with transfer learning. Fifth International Conference on Green Energy, Environment, and Sustainable Development (GEESD 2024) SPIE 2024; 13279(1): 197–211. DOI: 10.1117/12.3044425.
7. Siegel M, Kattmann C, Beura C P, Beltle M, Tenbohlen S. Partial discharge monitoring of power transformers by calibrated UHF measurements. Journal of Energy: Energija 2023; 72(4): 19–27. DOI: 10.37798/2023724514.
8. Yan Y, Zhou F, Cai Z, Lu H. Visualization and monitoring algorithms for multi-device parallel overhaul progress at substation bases. Eksploatacja i Niezawodność – Maintenance and Reliability 2025; 28(1). DOI: 10.17531/ein/208485.
9. Sikorski W, Szymczak C, Siodła K, Polak F. Hilbert curve fractal antenna for detection and on-line monitoring of partial discharges in power transformers. Eksploatacja i Niezawodność 2018; 20(3): 343–351. DOI: 10.17531/ein.2018.3.1.
10. Sezen C. Pan evaporation forecasting using empirical and ensemble empirical mode decomposition (EEMD) based data-driven models in the Euphrates sub-basin, Turkey. Earth Science Informatics 2023; 16(4): 3077–3095. DOI: 10.1007/s12145-023-01078-5.
11. Ma C, Zhao M, Zhao Y. An overview of Hadoop applications in transportation big data. Journal of traffic and transportation engineering (English edition) 2023; 10(5): 900–917. DOI: 10.1016/j.jtte.2023.05.003.
12. Li MJ, Chen D W, Wang T, Liu JC, Liu W D. Time delay estimation of partial discharge based on dual EEMD and reconstruction. Chinese Journal of Radio Science 2024; 39(4): 760–768. DOI: 10.12265/j.cjors.2023223.

13. Nhat N N V, Huu D N, Hoai T N T. Evaluating the EEMD-LSTM model for short-term forecasting of industrial power load: A case study in Vietnam. International Journal of Renewable Energy Development 2023; 12(5): 881–890. DOI: 10.14710/ijred.2023.55078.
14. Zheng K, Zhou M, Zhang Y. Status evaluation of electric energy metering device (EEMD) based on artificial intelligence technology. Fifth International Conference on Mechatronics and Computer Technology Engineering (MCTE 2022) SPIE 2022; 12500(1): 1541-1546. DOI: 10.1117/12.2662643.
15. Zhang G, Du Z, Zhang Y, Chen J, Zhang X. Real-time performance evaluation and optimization of electrical substation equipment inspection algorithm based on distributed computing. International Journal of Low-Carbon Technologies 2024; 19(1): 1878–1887. DOI: 10.1093/ijlct/ctae136.
16. Zhao J, Li C, Wang L. Hadoop-based power grid data quality verification and monitoring method. Journal of Electrical Engineering Technology 2023; 18(1): 89–97. DOI: 10.1007/s42835-022-01171-7.
17. Wu J, Li H. Real-time data flow processing and optimization scheduling scheme for power system based on Kafka. Fifth International Conference on Telecommunications, Optics, and Computer Science (TOCS 2024) SPIE 2025; 13629(1): 461–471. DOI: 10.1117/12.3067553.
18. Zhang K, Lazaro J. Failure prediction and life cycle management of power equipment based on big data analysis. Journal of Computer, Signal, and System Research 2025; 2(1): 70–79. DOI: 10.71222/4ptjr081.
19. Escurra CM, Khamlichi A, Dalstein M, Vidal JR, Garnacho F, Mor AR, Vu-Cong T. Methods for partial discharge calibration in gas-insulated substations for HVDC power grids and charge evaluation uncertainty. IEEE Sensors Journal 2023; 23(19): 23486–23493. DOI: 10.1109/JSEN.2023.3302871.
20. Azam S M K, Othman M, Illias H A, Latef T A, Mahadi W N L W, Fahmi D, Ain M F. Planar sensor for noise cancellation during partial discharge detection in open substation. IEEE Sensors Journal 2023; 23(14): 15552–15562. DOI: 10.1109/JSEN.2023.3279555.
21. Zhan C, Ju Z, Xie B, Chen J, Ma Q, Li M. Signal processing for miniature mass spectrometer based on LSTM-EEMD feature digging. Talanta 2025; 281(1): 126904. DOI: 10.1016/j.talanta.2024.126904.
22. Rezaiy R, Shabri A. Enhancing drought prediction precision with EEMD-ARIMA modeling based on standardized precipitation index. Water Science Technology 2024; 89(3): 745–770. DOI: 10.2166/wst.2024.028.
23. Ali M, Khan DM, Saeed I, Alshanbari HM. A new approach to empirical mode decomposition based on akima spline interpolation technique. IEEE Access 2023; 11(1): 67370–67384. DOI: 10.1109/ACCESS.2023.3253279.
24. Samal P, Hashmi MF. Ensemble median empirical mode decomposition for emotion recognition using EEG signal. IEEE Sensors Letters 2023; 7(5): 1–4. DOI: 10.1109/LSENS.2023.3265682.
25. Chen S, Lin Y, Yuan Y, Li X, Hou L, Zhang S. Suppressive interference suppression for airborne SAR using BSS for singular value and eigenvalue decomposition based on information entropy. IEEE Transactions on Geoscience and Remote Sensing 2023; 61(1): 1–11. DOI: 10.1109/TGRS.2023.3263218.
26. Dutt MI, Saadeh W. Monitoring level of hypnosis using stationary wavelet transform and singular value decomposition entropy with feedforward neural network. IEEE Transactions on Neural Systems and Rehabilitation Engineering 2023; 31(1): 1963–1973. DOI: 10.1109/TNSRE.2023.3264797.
27. Marukatat S. Tutorial on PCA and approximate PCA and approximate kernel PCA. Artificial Intelligence Review 2023; 56(6): 5445–5477. DOI: 10.1007/s10462-022-10297-z.
28. Laghmati S, Hamida S, Hicham K, Cherradi B, Tmiri A. An improved breast cancer disease prediction system using ML and PCA. Multimedia Tools and Applications 2024; 83(11): 33785–33821. DOI: 10.1007/s11042-023-16874-w.
29. Kurani A, Doshi P, Vakharia A, Shah M. A comprehensive comparative study of artificial neural network (ANN) and support vector machines (SVM) on stock forecasting. Annals of Data Science 2023; 10(1): 183–208. DOI: 10.1007/s40745-021-00344-x.
30. Rabbani S, Safitri D, Rahmadhani N, Sani AAF, Anam MK. Perbandingan evaluasi kernel SVM untuk klasifikasi sentimen dalam analisis kenaikan harga BBM: Comparative evaluation of SVM kernels for sentiment classification in fuel price increase analysis. MALCOM: Indonesian Journal of Machine Learning and Computer Science 2023; 3(2): 153–160. DOI: 10.57152/malcom.v3i2.897.
31. Zhang H, Zhang Y. An improved sparrow search algorithm for optimizing support vector machines. IEEE Access 2023; 11(1): 8199–8206. DOI: 10.1109/ACCESS.2023.3234579.

32. Li J, Lei Y, Yang S. Mid-long term load forecasting model based on support vector machine optimized by improved sparrow search algorithm. Energy Reports 2022; 8(1): 491–497. DOI: 10.1016/j.egyr.2022.02.188.
33. Zou W, Zhao L, Sun K, Liu P. Denoising method of cable partial discharge signal based on improved VMD and permutation entropy. International Core Journal of Engineering 2023; 9(1): 56–63. DOI: 10.6919/ICJE.202304_9(4).0009.
34. Muñoz O, Schurch R, Ardila-Rey JA. Electrical tree growth identification by means of discrete wavelet transform (DWT) and principal component analysis (PCA). IEEE Transactions on Instrumentation and Measurement 2023; 72(1): 1–9. DOI: 10.1109/tim.2023.3284922.
35. Li T, Wang X, Zeng S, Gu C, Dong C, Zhen L, Liu H. Gas insulated switchgear partial discharge type identification based on frequency division features and improved long short-term memory neural networks. Journal of Nanoelectronics and Optoelectronics 2024; 19(2): 169–177. DOI: 10.1166/jno.2024.3579.
36. Wu Y, Cong P, Wang Y. Charging load forecasting of electric vehicles based on VMD-SSA-SVR. IEEE Transactions on Transportation Electrification 2023; 10(2): 3349–3362. DOI: 10.1109/tte.2023.3299417.